Two unsupervised learning algorithms for detecting abnormal inactivity within a household based on smart meter data

Zhenyu Zhou[a], Yanchao Liu[a], Tingli Hu[b], Caisheng Wang[b]

[a]*Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth Street, Detroit, MI 48201, USA. e-mail: {zyzhou,yanchaoliu}@wayne.edu*
[b]*Department of Electrical and Computer Engineering, Wayne State University. e-mail: {tinglih,cwang}@wayne.edu*

Abstract

We study the problem of detecting abnormal inactivities within a single-occupied household based on smart meter readings. Such abnormal events include immobilizing medical conditions or sudden deaths of elderly or disabled occupants who live alone, the delayed discovery of which poses realistic social concerns as the population ages. Two novel unsupervised learning algorithms are developed and compared: one is based on nested dynamic time warping (DTW) distances and the other based on Mahalanobis distance with problem-specific features. Both algorithms are able to cold-start from limited historical data and perform well without extended parameter tuning. In addition, the algorithms are small profile in terms of data usage and computational need, and thus are suitable for implementation on smart meter hardware. The proposed methods have been thoroughly validated against real data sets with simulated target scenarios and have exhibited satisfactory performance. An implementation scheme on smart meter hardware is also discussed.

Keywords: Smart meter data analysis, anomaly detection, nested dynamic time warping, unsupervised learning

1. Introduction

1.1. Background

Population aging, the increase of median age in society due to declining fertility and rising life expectancy, is an irreversible global trend with enormous economic and social consequences. In 2015, the World Health Organization (WHO) estimated that elderly populations will nearly double from 12% to 22% by 2050 (Kekade et al., 2018). Among the many alarming issues that come along with population aging, the unattended death of elderly adults who live alone is becoming a pressing humanitarian concern worldwide. The U.S. Census (2014) reports that 18 million U.S. residents over the age of 65 are unmarried, and a majority of them live alone (11 million, by Census report in 2011). Many who died during the 1995 Chicago heat wave were poor, elderly residents, and some deaths were unnoticed until days later. In China, the one-child policy coupled with the rural-to-urban migration of the younger generation has left tens of millions of elderly people living in solitude in rural and suburban homes. The urbanization movement also lodged more and more people in pigeon-hole apartments neighboring strangers, which exacerbated the social isolation and

the lack of mutual attention in the aged group. If the urgency and manageability of such conditions are misjudged, severe consequences will follow. In Japan, the concern and reality of Kodokushi, referring to people dying alone and remaining undiscovered for a long period of time, is not uncommon among senior citizens, a great proportion of whom live alone. It was reported that in Tokyo more than two thousand senior citizens died unattended and unnoticed in 2008.

Today's wireless and communication technology enables a wide selection of methods to track an individual's activity. For instance, GPS-enabled smartphones can accurately pinpoint a person's whereabouts and with consent the data can be transmitted to the concerned party in real-time via wireless networks (Whipple et al., 2009; Brown et al., 2007); wearable devices such as smart watches can analyze a person's activities traces and biomedical conditions (Wu and Luo, 2019; Reeder and David, 2016); indoor environmental sensors such as motion, acoustic, CO2 sensors as well as surveillance cameras can capture a home's liveliness status with high accuracy (Guerra-Santin and Tweed, 2015; Anderson et al., 2009). However, most these methods are to some extent intrusive. In other words, they either require installation and maintenance of instruments which incurs sizable one-time and running costs or require close participation of those being monitored. These prerequisites are hardly realistic among the needed population. Many seniors who live alone cannot afford a high-level of home care enabled by sensor technology, and many choose not to have sensors installed for privacy preferences. For the smart device option, if the person being monitored takes off or forgets to put on the required wearable device, the monitoring will be interrupted at best, and a false alarm may even be generated which can be costly to respond to. In comparison, integrating emergency detection functions into the building's standard fixtures, such as the electricity or water meters, would provide a low-cost, non-intrusive and passive participation option to enhance the well-being of the building's occupants.

1.2. Contribution

In this paper, our primary focus revolves around utilizing smart meter data to identify irregular energy consumption patterns within a household occupied by a single individual. Within this context, detecting abnormal patterns becomes crucial as they may indicate alarming events such as occupant immobilization or even death. Therefore, achieving timely and accurate detection of these patterns holds significant societal importance. Unlike numerous machine learning applications where labeled training data is readily available, our scenario presents a challenge in obtaining abnormal patterns from real-life situations. To tackle this obstacle, we propose the development of two unsupervised, distance-based learning algorithms specifically tailored to this task. These algorithms incorporate empirical insights into feature generation and algorithmic parameters. Both methods rely on the comparison of transformed subsequences derived from raw measurements obtained from the aggregated main circuit of the household. However, each algorithm adopts a distinct distance metric. The inclusion of problem-specific features plays a pivotal role in the effectiveness of our detection algorithms. We will delve into a thorough discussion of these features and provide extensive simulation experiments to demonstrate their efficacy.

Our contribution can be summarized to include: (1) a socially responsible definition of household anomaly with regard to the health and well-being of single-living senior citizens; (2) the identification of several effective yet easy-to-compute feature sequences and their

application in the Mahalanobis distance based Grubb's test for detecting anomalies in smart meter data streams; (3) a nested dynamic time warping (DTW) algorithm for the anomaly detection task; (4) simulation experiments, numerical validation and implementation schemes of the proposed algorithms.

The remainder of the paper is organized as follows. Section 2 reviews the related literature and highlights our contribution. Section 3 develops the main methods, and Section 4 presents simulation experiments and result analyses. Our rationale for algorithm choice and implementation schemes are discussed in Section 5 and Section 6 concludes the paper and discusses possible future work.

2. Related Literature

Energy consumption data from residential smart meters have been increasingly leveraged in areas of electric load analysis, forecasting, and management (Wang et al., 2018). Anomaly detection is a sub-area of load analysis and has received much attention in recent years (Himeur et al., 2021). Prominent applications include detection of abnormal behavior of end-users (Himeur et al., 2020), detection of faulty appliances (Yip et al., 2018), household occupancy detection (Laaroussi et al., 2020), non-technical loss detection (Yip et al., 2018; Aziz et al., 2020), and at-home elderly monitoring (Visconti et al., 2019; Patrono et al., 2018). Analytical methods applied to smart meter data grossly fall into several distinct categories, including empirical and observational methods, machine learning and statistical methods, and distance-based methods. We will provide a brief review of recent works on smart meter data analysis by methodology category, and then focus on reviewing the occupancy monitoring applications of smart meter data to highlight our contribution.

2.1. Empirical and Observational Methods

Because the effect of human activity on power meter readings is relatively complex, a wealth of work in these applications has used empirical and observational methods. Patrono et al. (2018) proposed an approach for monitoring elderly behavior by detecting home appliance's usage. A smart meter was used to constantly measures the overall energy consumption, and transmit the collected values to a cloud server. The energy disaggregation, which attempts to estimate which appliances are in use at the moment, would be performed on the server. However, the server-side algorithms were not provided in the paper. Visconti et al. (2019) proposed a sensors system for monitoring energy consumption and verifying the behavior of the less-sufficient people during their everyday home life. While the authors focused on hardware architecting and cellphone App prototyping, analytics and anomaly detection algorithms were not the focus of the paper. Chen et al. (2014) proposed a statistical predictive method for detecting anomalies in the mean and variance of energy consumption. In their method, the prediction interval of the baseline consumption was estimated using a generalized additive model, and the variation of the baseline was estimated by an autoregressive conditional heteroscedastic model. Zyabkina et al. (2018) focused on detecting anomalies in the time series of power quality parameters, such as the voltage and current values. They extracted indices of the original time series and classified high indices within a sliding window as being abnormal. The authors noted the dependence of the classification performance on the window size, and suggested that a two-week window size best fit the application. Liu and

Nielsen (2018) adopted a prediction-based detection approach aimed at detecting anomalies in the daily energy consumption pattern of a residential building monitored by a smart meter. In this approach, a periodic autoregression model with exogenous variables (PARX) was used for predicting the daily pattern, and the actual consumption pattern was deemed anomalous if it was sufficiently different (as determined by a distance metric and a Gaussian probability model) from what was predicted. The study was based on data sets with hourly resolution and the aim was to assist decision making in smart energy management. For the purpose of better load profiling, Shamim and Rihan (2020) proposed a dimension reduction method to first simplify the high-dimensional observational data by extracting features and then cluster the load profiles based on the selected features using the K-means algorithm. The features consisted of singular values of the pre-processed smart meter data matrix as well as the wavelet energy entropy. In comparison, the features used in our anomaly detection approach are identified through domain knowledge (i.e., the fundamental reasons for anomalous events) and insights from exploratory analysis.

2.2. Machine Learning and Statistical Methods

Machine learning methods or a combination of statistical methods and machine learning methods have also been commonly used in the smart meter analytics literature. For instance, Yan et al. (2020) used machine learning methods to help utilities carry out household characteristics classification based on smart meter data. Both time- and frequency-domain features were initially input into a random forest for feature selection, and then the selected features were fed into a support vector machine to perform the final classification tasks. The approach was primarily data-driven, and limited domain knowledge was necessitated. Jakkula and Cook (2010) detected outliers in power consumption data sets using a statistical approach and a clustering approach. The statistical approach was to analyze the t-score in various sized data windows, and the clustering approach was based on the k-nearest neighbors algorithms in which the discrete time warping distance was used. Gu et al. (2019) introduced an anomaly detection framework based on the Huber contamination model, in which a distance-to-measure based nearest neighbor method performed the separation of the anomalous distribution from the normal distribution. Various geometric and analytics properties of the underlying data distribution were leveraged in the theoretical development. Qiu et al. (2018) proposed a power anomaly detection and alarming system based on log analysis. They trained a model by the fault log data from the collection and billing system, and the model was used to detect frauds and anomalies in real-time. Gao et al. (2020) proposed a robust time series anomaly detection algorithm by integrating robust time series decomposition and convolutional neural network having an encoder-decoder architecture with skip connections. The authors showed that the proposed algorithm, called RobustTAD, was able to outperform several forecasting-based, decomposition-based, and statistics-based algorithms on certain benchmark data sets. Himeur et al. (2020) introduced a method to detect abnormalities in building energy consumption based on a deep neural network for micro-movement classification, i.e., classifying movements as good use, no change, turn on, turn off, excessive power consumption and outdoor consumption, etc. The authors nonetheless recognized that the requirement for an annotated data set for model training pose some practical limitation to supervised learning methods, which we tend to concur. Liu et al. (2016) proposed an anomaly detection algorithm by combining supervised learning and statistical

methods trained upon one's historical consumption patterns. They implemented the algorithm into a lambda architecture, and validated the effectiveness experimentally. In a similar vein, Araya et al. (2017) proposed an ensemble framework that combined pattern-based and prediction-based anomaly classifiers via a majority voting mechanism. In experiments that aimed to detect anomalous building energy consumption, the ensemble approach was shown to outperform the method that relied on the pattern-based anomaly classifier alone. Devlin and Hayes (2019) presented a multi-layer, feedforward neural network to identify the power consumption status of common household appliances via decomposing the household's smart meter measurements, which could then be used to infer occupants' activities. To increase the amount of training data available, the authors had to duplicate (i.e., over-sample) the limited number of true appliance profiles in the original data set. We comment that this is a typical compromise while applying deep learning methods to an application with data scarcity.

2.3. Distance-based Methods

Distance-based methods have also been investigated in the related literature. For example, Tran et al. (2016) evaluated algorithms for distance-based outlier detection in data streams under various stream settings and outlier rates. Their approach is not deliberately applied to power meter monitoring, but the outlier detection method is suitable for the application for power meter anomaly detection. Yijia and Hang (2016) proposed a waveform feature extraction model for power consumption anomaly detection. The authors first extracted the edge point sequence to identify the normal patterns, and then measured the similarity distance between the loss changing pattern and the targeted feeder. In this way, users who were suspected of power theft could be quickly pinpointed.

2.4. Literature on Occupancy Monitoring

While energy consumption is only weakly correlated to occupancy in large non-domestic buildings (Martani et al., 2012), a household's electricity consumption is highly correlated with its occupants' activities, and such a correlation has been widely exploited from the perspective of usage forecasting, energy conservation and home energy management systems (HEMS). For instance, knowing the occupancy level of a building can improve energy use forecasts (Newsham and Birt, 2010); furthermore, a building's energy consumption such as the heating, ventilation and air conditioning (HVAC) setpoints and lighting brightness can be dynamically adjusted according to the occupancy status (Agarwal et al., 2010; Erickson and Cerpa, 2010; Ardakanian et al., 2018). The reverse relationship, i.e., using electricity usage to infer occupancy status, is relatively less studied. Zou et al. (2017) proposed an occupancy sensing system to detect, count and locate occupancy through the WiFi traffic. They implemented the system in a 1500 m^2 built environment and validated the performance. Rafsanjani and Ahn (2016) experimentally analyzed the relation between the occupant's entry and departure events and a building's energy usage. In a follow-on paper (Rafsanjani et al., 2018), the authors developed a clustering algorithm to efficiently monitor electric load of a building based on occupants' entry and departure events. Liu et al. (2018) proposed a probabilistic graphical modeling approach to utilize multivariate time-series data for non-intrusive load monitoring (NILM). Their approach applied stochastic model for pattern learning that can discover specific patterns of energy usage or generation and particularly focused on accurate energy disaggregation. Kleiminger et al. (2015) implemented several

supervised machine learning algorithms as classifiers: Support vector machine (SVM), K-nearest neighbors (KNN), Gaussian mixture models (GMM), hidden Markov model (HMM), and a thresholding approach to detect occupancy information (at home or away) using aggregate electricity consumption data from smart meters. However, the classifiers may fail when there are small temporal changes in occupant's behavior. In (Chen et al., 2013a), a non-intrusive occupancy monitoring (NIOM) algorithm was proposed to produce a continuous trace of a home's occupancy status based on the smart meter's readings. By noting that when a home was occupied the power usage was higher and more variable than when it was unoccupied, the authors developed a threshold-based approach to detect occupancy changes. Specifically, an occupancy change event was retro-predicted to occur at a given time point if both the average and standard deviation of power usage over a fixed window in the immediate past exceeded a corresponding baseline threshold, whereas the baseline threshold was estimated by nighttime usage with certain assumptions. Occupancy events that were close to each other along the time line were then clustered together to produce a reasonable trace of binary occupancy status. The NIOM problem has been increasingly investigated in recent research, though different researchers used different data sets and basic assumptions. For instance, Becker and Kleiminger (2018) based their analysis on several data sets, including an electricity and occupancy data set (Tang et al., 2015), the ECO data set (Beckel et al., 2014), and part of the Smart* data set (Chen et al., 2013b), and proposed three unsupervised learning methods, i.e., a hidden Markov model, a geometric moving average algorithm and a Page-Hinkley test based algorithm, to monitor occupancy based on smart meter readings. The authors concluded that unsupervised learning methods could outperform supervised learning methods in certain cases. Razavi et al. (2019) analyzed the electricity consumption behavior of more than 5000 households over an 18-month period and deployed an array of machine learning methods aimed at detecting and predicting the occupancy status of the households. In this work a genetic programming approach was proposed to extract useful features for training the predictive models. Hattori and Shinohara (2017) considered novel applications of NIOM including ambient assisted living, sales promotions, and peak load shifting. To overcome the difficulty in utilizing the low-resolution electric consumption data, the authors proposed a data generation step to enrich the data set with estimations that reflect the household characteristics, and used the enriched data to predict the household's occupancy status.

Compared to the existing NIOM algorithm in the literature, our methods boast several new features. First, our methods operate in a multivariate feature space whereas NIOM links different criteria by simple rules. Second, our algorithms can be deployed for a cold-start at any home of a similar setting, as long as a limited set of past meter readings are accessible in a rolling window. No long-term memory of data is necessary. Third, abnormal inactivity encompasses richer information (or more uncertainty) than the binary occupancy status. For instance, an immobilizing medical condition or a death event might occur when the occupant is using an appliance such as the electric stove or bedroom light. Hence the power consumption would remain largely unchanged after the event. Unlike moving-window based simple occupancy detection algorithms, our method is able to detect such happenings since more features of the time series under study can be revealed.

3. Methods

3.1. Assumptions and Problem Statements

Factors that influence residential power consumption can be categorized into four groups: external conditions (e.g., location and weather), physical characteristics of dwelling, household electric appliance, and occupants (Kavousian et al., 2013). If we examine the temporal consumption pattern for a given home, the main factors are then reduced to weather, appliance and occupants' activity. And we have made the following assumptions for the goal of detecting abnormal inactivity:

- We target the simple home environment (small houses or apartments) with single-occupied households.
- We assume that heating is sourced from natural gas and air-conditioning, if a unit exists, runs on a separate circuit.
- We assume a simple indoor lifestyle of the occupant in which lighting and cooking constitute the main source of electric consumption.
- Only the main circuit power is measured by the smart meter.

The second assumption eliminates the need to strip the weather effect from smart meter readings, which otherwise could be performed as a preprocessing step for our algorithm. Specifically, we constructed the targeted home environment by subsetting the Smart* Home Dataset (Barker et al., 2012), which provides more than three years of 30-minute-level and sub-circuit-level energy usage data accompanied by detailed home environment description for three homes in Western Massachusetts. We composed our data set for study using four sub-circuits from Home A. The data set consists of the electricity consumption incurred by electric range, fridge range, bedroom lights and kitchen lights. Figure 1 plots the power consumption (in Watts) of the four sub-circuits in the home over a 24-hour period sampled every 30 minutes. The occupant's normal activity throughout the day is clearly displayed by her intermittent usage in lighting and cooking. In addition, we can see that no single category is in dominance in terms of the peak power level. Note that in the fourth assumption for a typical home, only the main circuit power is measured by the smart meter, so the meter reading (input to our algorithm) only reveals the overall consumption level averaged in the sampling interval, in this case, a 30-minute interval.

With the above assumptions, we are ready to state the main research questions to be investigated in the sequel. We define *anomaly*, or *abnormal inactivity*, as any concerning deviation of a home's electric power consumption pattern from what is expected out of the occupant's long-held routine, that is indicative of an immobilizing health event (e.g., falling down, passing out, drug overdose or death) otherwise gone unnoticed. We will develop data analytic algorithms that can detect such anomalous events as timely as possible, based on the household's aggregated energy consumption observed every 30 minutes by a smart meter device. For practical implementation, the anomaly detection models must be portable and adaptable to different household environments. To this end, we focus on developing unsupervised learning models that are able to establish a nominal baseline from historical data

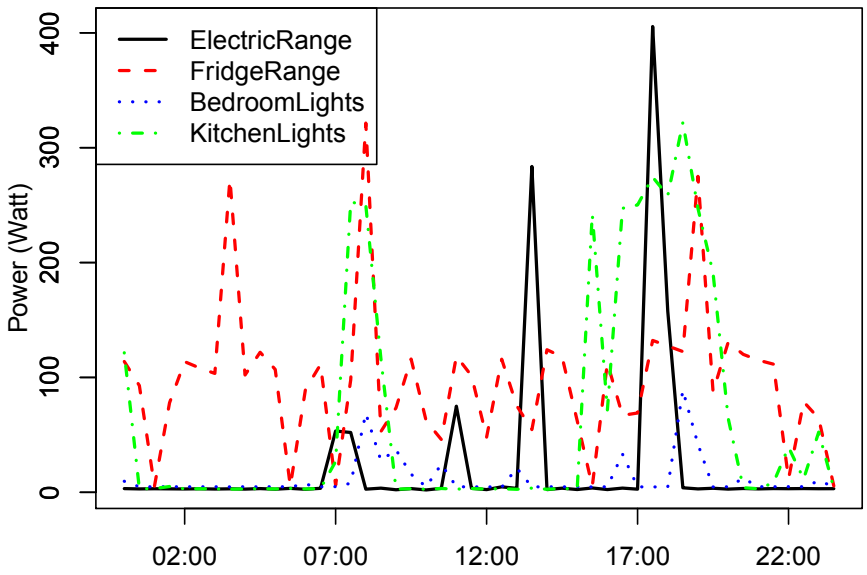

Figure 1: A simple home's electric consumption by appliance.

and extract features from it, without relying on labeled training data. Specifically, to tell apart normal and abnormal energy consumption patterns, we measure the distance between a recent sequence of smart meter readings (e.g., in the past 6 hours) and a typical sequence that would be expected under nominal conditions. We present a Mahalanobis distance based method in Sections 3.2 and 3.3, where we derive several features (i.e., summarizing statistics) from the raw consumption sequence, and compute the distance between the feature vector value and the distribution of the feature vector as estimated from historical meter readings which were assumed anomaly-free[1]. The Mahalanobis distance is a measure of the distance between a point (i.e., the feature vector value) and a distribution (i.e., the feature vector distribution under normal conditions), therefore it is suitable for the purpose. Our contribution resides in deriving the features and their correlation coefficients that are effective at telling apart normal and abnormal sequences using the Mahalanobis distance. In Sections 3.4 and 3.5, we develop a nested dynamic time warping (DTW) method. In this method, anomaly is detected by measuring the DTW distance between two sequences which themselves are DTW distances extracted from the raw consumption sequence, thus the detection algorithm is able to pick up pattern changes caused by the target event regardless of the time of occurrence. Furthermore, in Section 3.6, we describe a prediction-based baseline approach for performance comparison purposes. To validate the effectiveness of the anomaly detection algorithms, we simulate test cases with both normal and abnormal instances and test the algorithms in a binary classification setting, even though the algorithmic parameters are learned in the unsupervised learning framework.

[1]This assumption is reasonable, as the anomalies we consider in this paper are terminal events of the household's occupancy, such as the sudden death of the occupant

Table 1: Nomenclature

Data	Description
t, s	Time interval index
x_t	Actual Watt at time t
X_t	Watt at time t under the normal case, a random variable whose distribution is estimated from past samples
$\hat{\mu}_{X_t,D}$	Mean of X_t estimated by the D-day same-time samples
$\hat{\sigma}_{X_t,D}$	StdDev of X_t estimated by the D-day same-time samples
$\sigma_{T,t}$	StdDev of the population $\{x_{t-T+1}, \ldots, x_t\}$
$\Sigma_{T,t}$	Random process of $\{\sigma_{T,t}\}$ under the normal case
$\hat{\mu}_{\Sigma_{T,t},D}$	Mean of $\Sigma_{T,t}$ estimated by the D-day same-time samples
$\hat{\sigma}_{\Sigma_{T,t},D}$	StdDev of $\Sigma_{T,t}$ estimated by the D-day same-time samples
f_t	Feature vector computed at time t
w	Weight vector for features
T	Number of look-back intervals
y_t	Actual Watt at time t
Y_t	Consumption (Watt) at time t in nominal condition
d_t	The DTW distance between X_t and Y_t at time t
	Other symbols will be self-explanatory in the text.

3.2. Feature Extraction

In this section, we will first analyze several hypothesized cases that are representative of the types of anomalous events in focus, and derive several summative statistics series from the smart meter's reading series that are sensitive to relevant changes in the consumption pattern. These statistics will then be used as *features* (i.e., input values calculable at the time of inference) in the distance-based anomaly detection algorithm, to be presented in Section 3.3. The nomenclature used in the analysis and algorithm development is given in Table 1.

Figure 2 plots the normal power consumption over a two-day period (i.e., Sunday, Feb 1 to Monday, Feb 2, 2014), together with the would-be power levels for three anomaly cases that were set to occur at different times. Anomaly 1 occurs at 2 am on Sunday, which mimics a death during sleep. Realistically, there is no way to detect such a quiet happening in the first few hours until the next morning, when the usual morning peak is gone amiss. If Anomaly 1 did occur, the Normal case (the black solid curve) would stop at Sunday 2 am hence our visual comparison is really against the gray area, which depicts the 95% confidence interval for X_t, the (uncertain) electricity consumption at time t whose distribution is estimated by the same-time meter readings in the past D days. Here and for the subsequent numerical analysis we took $D = 30$ to use the past 30 days of data. Given that the time line of each day is evenly divided into c metering intervals, and a meter reading x_s is available for the past D days, where the subscript s indicates the sth interval over the whole period under study, the expected consumption level μ_{X_t} at any time t can be estimated by

$$\hat{\mu}_{X_t,D} = \frac{1}{D} \sum_{d=1}^{D} x_{t-c \cdot d}$$

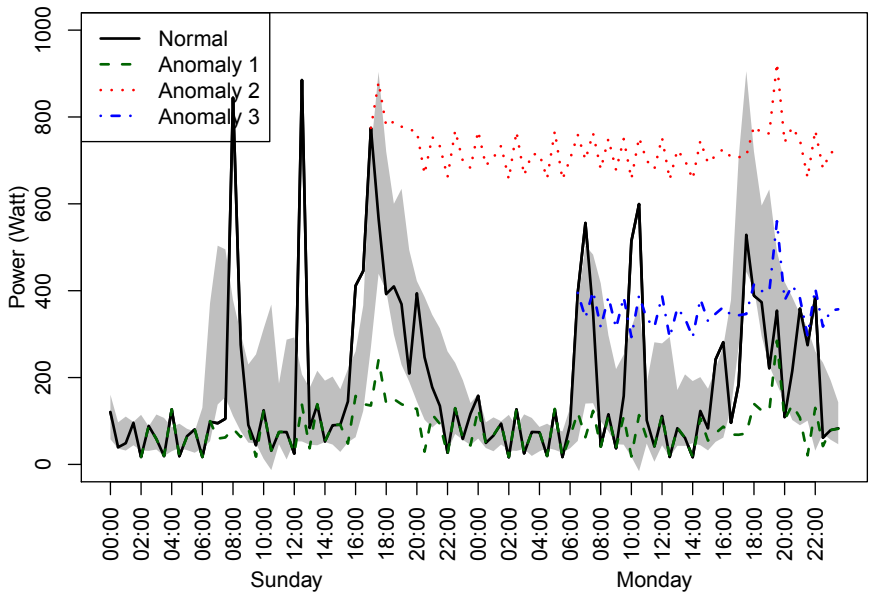

Figure 2: Three instances of anomaly: power levels.

and the standard deviation σ_{X_t} can be estimated by

$$\hat{\sigma}_{X_t,D} = \left(\frac{1}{D-1} \sum_{d=1}^{D} (x_{t-c\cdot d} - \hat{\mu}_{X_t,D})^2 \right)^{1/2}.$$

Note that both parameter estimates are calculated using the information observable as of time t. Here the "hat" symbol on top of the corresponding parameter symbol indicates that the quantity is an estimate, rather than the parameter itself (which is unknown). It is reasonable to assume that under the normal case (i.e., no Anomaly) the consumption X_t follows a Gaussian distribution with mean $\hat{\mu}_{X_t,D}$ and standard deviation $\hat{\sigma}_{X_t,D}$. As the actual x_t is observed, we can then compute the likelihood of this observation using the Gaussian density function. If mildly low-likelihood observations accumulate for several consecutive intervals such that the probability of observing the sequence is extremely low, an alarm should be raised.

Anomaly 2 occurred at 5 pm Sunday when the occupant supposedly passed out while cooking in the kitchen. The power level sustained an unusually high level (as the Electric Range and Kitchen Lights remained on after the incidence) for the next few intervals. Such an anomaly should be detected more quickly given that the subsequent meter readings digressed from the predicted distribution too far too soon.

The third hypothesized anomaly occurred at Monday 6:30 am when the power usage was at a mid-level, for instance some (but not all) kitchen lights were turned on or the lights were at a dimmed brightness.

Across the three hypothesized instances of anomaly, we can generate the following insights.

- Unlike an occupancy event in which the average consumption will register a noticeable change compared to the past few hours, an anomaly event does not necessarily change

the average power consumption. Post an anomaly incidence, the average power level can stay at any level in the full range of total consumption capacity, as demonstrated in Figure 1. This suggests that NIOM methods that monitor the changes in the average consumption, e.g., in Chen et al. (2013a), are not applicable in our context.

- The post-anomaly variance of the power, however, will immediately reduce and rest at the level of the background load. Changes in the variance provide a timely indication of an anomaly.

- The power level itself, via comparing with its expected distribution learned from the day-to-day pattern in the past, can also be used as a reliable signal to indicate abnormal deviation of consumption.

Let us further explore the second point above. First, we will derive a new series $\{\sigma_T\}_t$ from the Watt series $(x_{t-T+1}, \ldots, x_t)$, to represent the observed variance in the past T Watt readings. It is defined as

$$\sigma_{T,t} = \left(\frac{1}{T} \sum_{s=t-T+1}^{t} (x_s - \bar{x}_{T,t})^2 \right)^{(1/2)} \tag{1}$$

where $\bar{x}_{T,t} = (\sum_{s=t-T+1}^{t} x_s)/T$. Note that the series $\{\sigma_T\}_t$ is what is actually observed up to time t, regardless of whether an anomaly has occurred or not. We denote the Watt series under normal conditions by X_t, which is a stochastic process. The variance series $\Sigma_{T,t} = (\sum_{s=t-T+1}^{t}(X_s - \bar{X}_{T,t})^2/T)^{(1/2)}$ is also a stochastic process. Therefore, we can form (and continuously refine) an idea for what the distribution of $\Sigma_{T,t}$ should look like under the hypothesis that no anomaly has occurred up to time t, and check if the observed $\sigma_{T,t}$ conforms well to this distribution, i.e., by evaluating the likelihood of $\sigma_{T,t}$. A low-likelihood observation, or a sequence of low- and decreasing-likelihood observations, would indicate that an anomaly had occurred. Similar to the treatment for X_t, we assume that $\Sigma_{T,t}$ follows a Gaussian distribution for each t, whose mean $\mu_{\Sigma_{T,t}}$ and standard deviation $\sigma_{\Sigma_{T,t}}$ are estimated from the past D-day same-time observations of $\sigma_{T,t}$. Specifically,

$$\hat{\mu}_{\Sigma_{T,t},D} = \frac{1}{D} \sum_{d=1}^{D} \sigma_{T,t-c\cdot d} \tag{2}$$

$$\hat{\sigma}_{\Sigma_{T,t},D} = \left(\frac{1}{D-1} \sum_{d=1}^{D} (\sigma_{T,t-c\cdot d} - \hat{\mu}_{\Sigma_{T,t},D})^2 \right)^{(1/2)} \tag{3}$$

Figures 3 and 4 illustrate the series $\{\sigma_T\}_t$ for $T = 48$ (24 hours) and $T = 12$ (6 hours), respectively, for the normal and anomaly cases. The 95% confidence interval at each time t is constructed using equations (2) and (3) with $D = 30$ and $t_{0.975,30} = 2.04$, the appropriate quantile from the t-distribution. We can see from the leveled confidence intervals in Figure 3 that the variance of power consumption in any 24-hour interval is nearly constant, which is simply because the interval length is 24 hours, the natural period for human being's everyday activities. For instance, if the occupant has dinner-cooking and bedtime-reading habits, the power peaks incurred by these events can be captured once in about any 24-hour interval

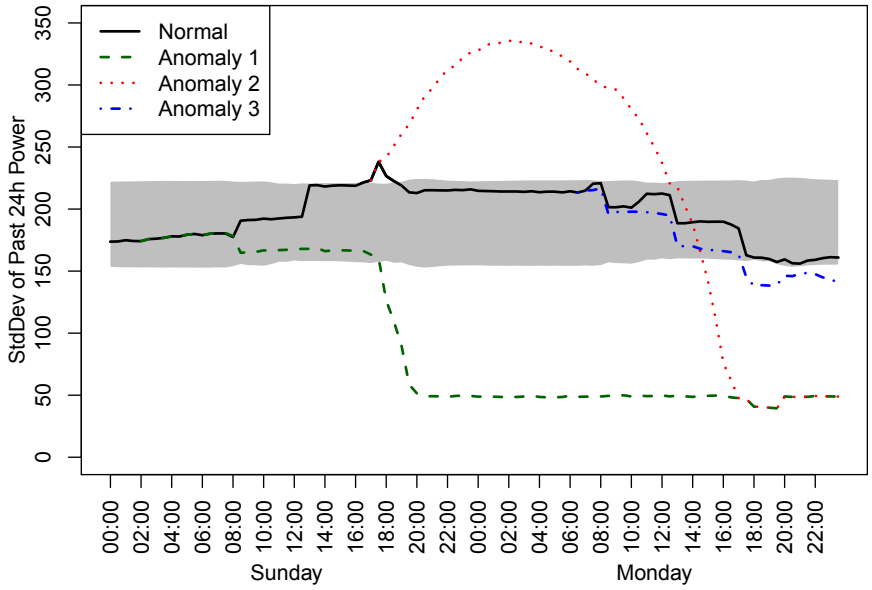

Figure 3: Three instances of anomaly: 24-hour variation.

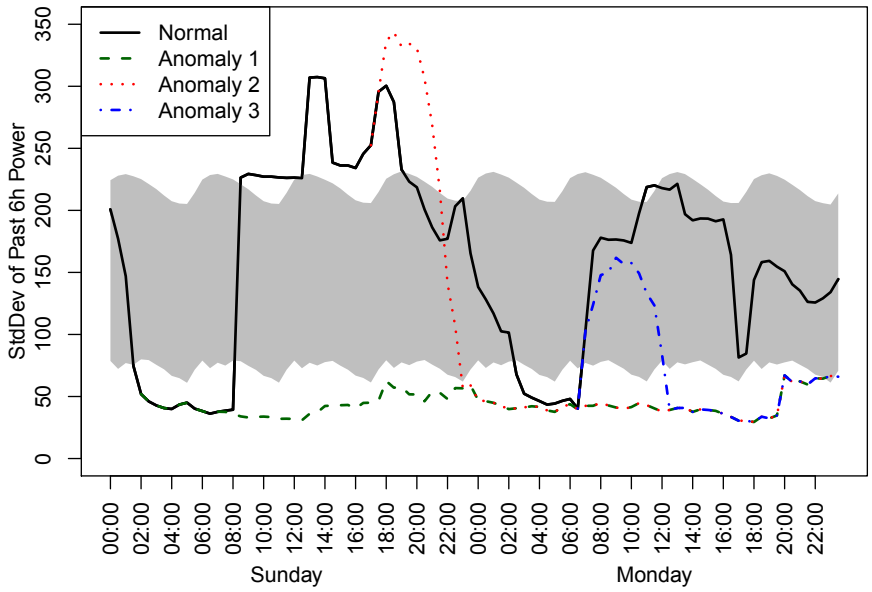

Figure 4: Three instances of anomaly: 6-hour variation.

Table 2: Simple detection rule by variance drop.

	Occurred	Detected ($T = 24$h)	Detected ($T = 6$h)
Anomaly 1	Sun 02:00	Sun 16:00	Sun 02:00
Anomaly 2	Sun 17:00	Mon 15:00	Sun 22:00
Anomaly 3	Mon 06:30	Mon 17:30	Mon 12:00

and where such peaks happen to locate within the interval does not affect the variance at all. This is a nice feature, in the sense that when a significant deviation from the typical (nearly constant) variance is detected, we can reliably conclude that something abnormal has happened. Coming with the high reliability is its insensitivity - the gap between the occurrence and detection of an anomaly can be up to half a day. Fortunately, the reliability-sensitivity tradeoff is adjustable via the parameter T. We can achieve a quicker detection by choosing a smaller T, at the cost of a higher risk of false positive claims, i.e., falsely claiming a normal case to be abnormal. For illustration let us assume a simple detection rule: an anomaly is said to have occurred by time t if $\sigma_{T,t}$ is below the lower limit of the confidence interval. Note that this rule does not retro-predict *when* the anomaly occurred, which will be discussed in future work. By this rule, we can compare the performances of $T = 24h$ versus $T = 6h$ by comparing Figure 3 and Figure 4. Apparently, the latter setting led to much quicker detections for all three anomaly cases. The specific time of detection is summarized in Table 2.

3.3. Mahalanobis Distance Method

For a new observation x_t, we have discussed how to compute different features based on the deviation of x_t and $\sigma_{T,t}$ from the hypothesized distributions of X_t and $\Sigma_{T,t}$ under the normal condition. In particular, we are able to evaluate the likelihood of the new observation by each feature. At this point, based on the important features we have identified so far there is a great deal of freedom in designing an anomaly detection algorithm. For instance,

1. We can independently evaluate the likelihood of each feature and compare against a preset threshold, and then employ a voting-based decision rule to determine whether the new observation is anomalous or not.

2. We can construct a joint Gaussian distribution of all features and make an inference by evaluating the likelihood in the feature space.

3. We can compute the Mahalanobis distance of the feature instance and apply the Grubb's test (Chandola et al., 2009; Laurikkala et al., 2000) to determine anomaly.

In this section, we will apply the third option and leave a comprehensive study of algorithm design for future work. Let $f_t \in \mathbb{R}^p$ be the value vector of p features for the observation at time t and in this paper we use three features $f_t = [\sigma_{48,t}, \sigma_{12,t}, x_t]$, where the first two elements follow the definition in equation (1). Let $w_i, i = 1, \ldots, p$ be a pre-assigned weight for feature i, with $\sum_{i=1}^{p} w_i = 1$, reflecting the relative importance (as a trade-off of reliability and sensitivity) of feature i. The weight assignment can be based on ad-hoc knowledge or

belief about the feature space (as is the case here) or can be optimally learned via cross-validation on a training data set. To the authors' knowledge, weighting the features as in equation (4) in the context of multivariate Grubb's test is a novel invention that has not been discussed in the literature. In the algorithm to be tested below, we arbitrarily select a weight vector $w = [0.6, 0.3, 0.1]$, to reflect the insights gained from Figure 2 to 4 that the 24-hour variance is a more reliable indicator of anomaly than the other two. We then compute the weighted Mahalanobis distance between f_t and its distribution, as follows

$$\delta_t = \sqrt{(f_t - \bar{f}_t)^T W S^{-1} W (f_t - \bar{f}_t)} \tag{4}$$

where $\bar{f}$ is the mean vector estimated by the past D-day same-time observations, in particular $\bar{f}_t = [\hat{\mu}_{\Sigma_{48,t,D}}, \hat{\mu}_{\Sigma_{12,t,D}}, \hat{\mu}_{X_t,D}]$; W is a p-by-p diagonal matrix whose diagonal entries are w_i, $i = 1, \ldots, p$ and whose off-diagonal entries are zero; S is the covariance matrix of the D-day same-time historical records of the feature columns weighted by w. Note that calculating the feature vector f_t at any time t involves only the past 48 meter readings as of t. However, the calculation of the mean vector $\bar{f}_t$ requires $48 \cdot D$ (e.g., 1440, if $D = 30$ days) historical meter readings, according to equations (2) and (3).

The new observation x_t is determined to be anomalous if the following condition (Grubb's test) holds,

$$\delta_t > \frac{D-1}{\sqrt{D}} \sqrt{\frac{t^2_{\alpha/(2D),D-2}}{D-2+t^2_{\alpha/(2D),D-2}}} \tag{5}$$

where D is the number of data points used to estimate the parameters, and we have $D = 30$ here since we use the past 30-day same-time data to estimate parameters; α is a significance level and we choose $\alpha = 0.9$ for the experiments. Note that a single observation can appear anomalous by chance when no abnormal activity has really occurred. We further safeguard the algorithm against making frequent false positive claims by requiring that an anomaly alarm is raised only when the past K observations all pass the Grubb's test. In other words, we claim an anomaly has occurred by time t when inequality (5) holds simultaneously for δ_s, $s = t - K, \ldots, t$. The proposed detection process is summarized graphically in Figure 5.

3.4. Dynamic Time Warping Distances

This subsection reviews the basic concepts and formulation of the dynamic time warping (DTW) distances, and the next subsection presents a novel anomaly detection approach based on a nested application of DTW on the original data series. DTW is a well-known technique to find optimal alignments between two time-series sequences of different lengths. It has been successfully applied in a wide spectrum of applications, including word and speech recognition (Myers and Rabiner, 1981b,a; Myers et al., 1980), signature verification (Munich and Perona, 1999), fingerprint verification (Kovacs-Vajna, 2000) and aligning gene expressions (Aach and Church, 2001). The gist of the DTW method is to measure the similarity between two time series by computing the least cumulative distance between them. The definition of the DTW distance can be found in (Müller, 2007) (pp 69-71), and we reproduce it below for completeness. Given two time series $X = (x_1, x_2, ..., x_m)$ with length of m and $Y = (y_1, y_2, ..., y_n)$ with length of n, one can build a distance matrix D, i.e., $D(i, j) = |x_i - y_j|$, and define $\Phi = \{\phi(1), \phi(2), ..., \phi(K)\}$ as an (m, n)-warping path with length K, in which each

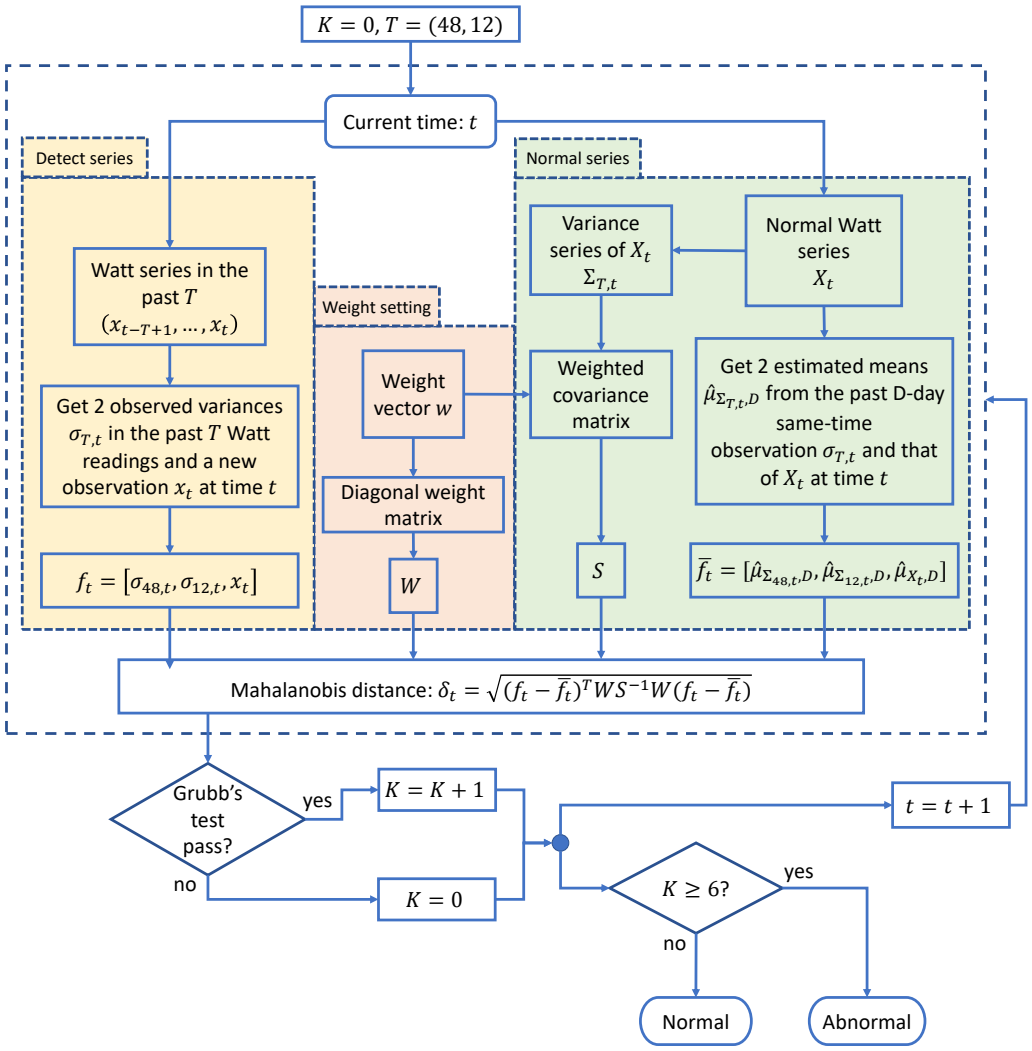

Figure 5: Proposed Mahalanobis distance method flow chart. The algorithm is executed at each time point t to detect the target types of anomaly that has possibly occurred.

element $\phi(k)$ represents an index pair $(i_k, j_k) \in \{1, \dots, m\} \times \{1, \dots, n\}$, for $k \in \{1, \dots, K\}$. The path Φ has the following properties:

1. $\phi(1) = (1, 1)_1,\ \phi(K) = (m, n)_K$
2. $i_1 \leq i_2 \leq \dots \leq i_K,\ j_1 \leq j_2 \leq \dots \leq j_K$
3. $\phi(k+1) - \phi(k) \in \{(1, 0),\ (0, 1),\ (1, 1)\},\ \forall k$

The cost $c_\Phi(X, Y)$ of a warping path ϕ between X and Y is given by

$$c_\Phi(X, Y) := \sum_{k=1}^{K} D(i_k, j_k) \tag{6}$$

An optimal warping path between X and Y is a warping path Φ^* having minimal cost among all possible warping paths. The DTW distance between X and Y is then defined as the cost

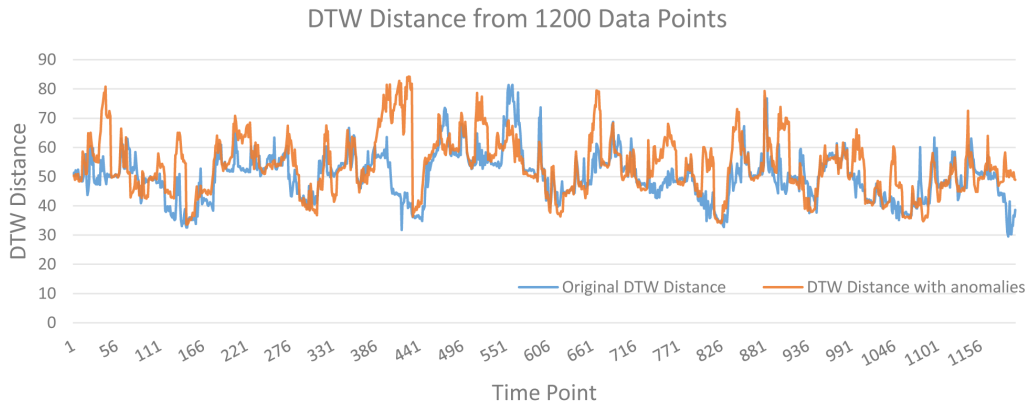

Figure 6: Demonstration of a wide range of DTW distances across different periods of time.

of Φ^*, i.e., $\mathrm{DTW}(X, Y) := c_{\Phi^*}$. In our DTW-based algorithm, we used the R package "dtw" (Giorgino, 2009) to calculate the DTW distance for any given pair of sequences X and Y.

3.5. A Nested DTW-based Detection Algorithm

For a given series of data, at time t we take the latest N time points ($N = 48$ is used subsequently to include the past 24 hours' condition), denoted by X_t, $X_t = (x_{t-47}, x_{t-46}, ..., x_t)$ comparing with another 24-hour sequence Y_t from 30 days ago, that is, $Y_t = X_{t-T} = (x_{t-T-47}, x_{t-T-46}, ..., x_{t-T})$, where $T = 30 \times 48$. Using DTW method, we obtain a DTW distance d_t between X_t and Y_t at each time t. As time advances, we can get more DTW distances, i.e., $d_{t+1}, d_{t+2}, \ldots, d_{t+n}$, for n time periods into the future. These distances show a quantitative difference between the current information on the residential electricity consumption and the nominal consumption level as exhibited by historical data. One would expect that if an anomaly were to occur at time t, the DTW distances $d_t, d_{t+1}, ..., d_{t+n}$ would exhibit a notable increase due to dissimilarity with what the sequence would look like under nominal conditions. However, as shown in Figure 6, the DTW distances vary substantially over time, and the ranges of the DTW distances for the original and abnormal periods are mostly overlapped. Thus, on the DTW distance alone, it is difficult to set a fixed threshold value based on which we could separate nominal values and anomalies very well.

To overcome this challenge, we further calculate the DTW distances of the distance series (hence, it is a nested DTW method). Specifically, let $P = (d_{t-47}, d_{t-46}, ..., d_t)$ and $Q = (d_{t-T-47}, d_{t-T-46}, ..., d_{t-T})$, we can calculate the DTW distance between P and Q and denote it as $d_t^{'}$. In this way, the range of DTW distances is narrowed and the interesting signals are strengthened, which makes setting a detection threshold easier.

We are especially interested in one particular case of anomaly, in which the resident experiences sudden death or accident that would leave the bedroom light, kitchen light or the electric range unchanging after the accident's occurrence, while the fridge keeps working as usual. We can add a filter to the DTW distance vector to single out this particular type of instance of interest. In particular, after examining the historical data, we found an exploitable fact: the mean and median values of the fridge range consumption are almost the same, while for other appliances the mean consumption values are greater than the median values, and this difference is up to 65 in terms of the overall power consumption. See the

Table 3: Statistics of historical data 02/02/2014 - 11/30/2014.

Watts	FR*	KL	BL	ER	Total
Max	498.8	761.1	338.5	825.5	1547.9
Mean	76.0	74.8	14.8	17.8	183.3
Median	75.6	3.6	4.9	3.4	118.8
1st Quartile	33.3	3.3	4.8	2.9	62.7

* FR: Fridge Range; KL: Kitchen Lights; BL: Bedroom Lights; ER: Electric Range.

summaries in Table 3 for demonstration.

The rationale behind taking the difference between the mean and median among these appliances is that lighting and cooking are high-power and low-duration consumption, whereas the fridge is of an opposite nature. Therefore, after an anomaly occurs, i.e., lights and oven remain in their pre-anomaly states, over time the mean-median gap will noticeably diminish. Such a winding-off pattern provides a strong signal for the anomaly event we are interested in.

The detection process is listed in Algorithm 1 in the form of a function, and is summarized graphically in Figure 7. Input parameters include the current time t, the major look-back period T and the minor look-back period N. In our context, the nested DTW algorithm is configured to look back 30 days and 24 hours, that is, $T = 1440$ and $N = 48$. We use the following shorthand notation for sequence $(x_s)_{s=k}^{n} := (x_k, x_{k+1}, ..., x_n)$ for any k, $n \in \mathbb{Z}$, and $k < n$. The function DTW(X, Y) for any two sequences X and Y returns the cost (or distance) of the warping path as given in equation 6 along with the other defined operations leading to that equation. Overall, the DetectAnomaly function is called at each time point t and a Boolean anomaly indicator is returned.

3.6. Prediction-based method - a baseline

Both the Mahalanobis Distance method and the nested DTW method leverage some problem-specific insights which are key to their effectiveness. For comparison purposes, we present a generic, prediction-based anomaly detection approach. Prediction-based approaches have been commonly used in the literature, see, e.g., Araya et al. (2017), although algorithmic designs are different for different applications. As a comparison baseline we implement a straightforward idea here. Specifically, whenever a new meter reading becomes available (i.e., every 30 minutes), an ARIMA model is constructed using the time series starting from the same-time observation 31 days ago and ending at the same-time observation one day ago (i.e., 30 days of data), and is applied to "forecast" the energy consumption for the past 24 hours (48 data points to forecast). If the actual energy consumption pattern in the past 24-hours is sufficiently different from what is forecast by the model, then an anomaly is deemed to have occurred in the past 24 hours. For simplicity, the Euclidean distance between the two 48-point sequences (i.e., the actual and the forecast) will be calculated and compared to a decision threshold. We used the `auto.arima` function in the 'forecast' package (Hyndman and Khandakar, 2008) in R to determine the parameters for ARIMA. The selected model is ARIMA$(0, 1, 1) \times (2, 0, 0)_{48}$, in which the first vector and the second vector represent the (auto-regressive, difference, moving-average) coefficient values, for the non-seasonal and

Algorithm 1 Nested DTW method for anomaly detection at current time t

function DETECTANOMALY(t, T, N)
  for $i = t - N + 1$ to t **do**
    $X_i \leftarrow (x_s)_{s=i-N+1}^{i}$
    $X_{i-T} \leftarrow (x_s)_{s=i-N+1-T}^{i-T}$
    $X_{i-2T} \leftarrow (x_s)_{s=i-N+1-2T}^{i-2T}$
    $d_i \leftarrow DTW(X_i, X_{i-T})$
    $d_{i-T} \leftarrow DTW(X_{i-T}, X_{i-2T})$
  $P_t \leftarrow (d_s)_{s=t-N+1}^{t}$
  $Q_t \leftarrow (d_s)_{s=t-N+1-T}^{t-T}$
  $d'_t = DTW(P_t, Q_t)$
  if $d'_t < TH_{DTW}$ **then**
    result $\leftarrow X_t$ is normal
  else
    $m_t \leftarrow \text{mean}(X_t) - \text{median}(X_t)$
    if $m_t > TH_{Stat}$ **then**
      result $\leftarrow X_t$ is normal
    else
      result $\leftarrow$ Anomaly detected in X_t
  return result

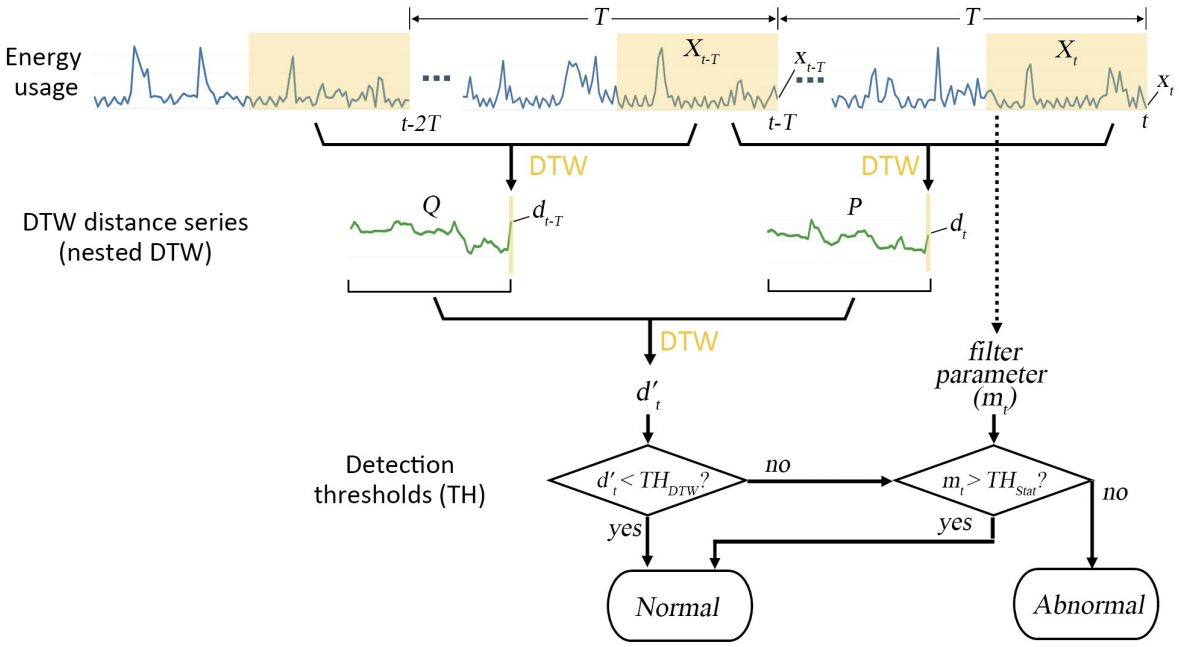

Figure 7: Illustration of the nested DTW method for anomaly detection.

seasonal (with period 48) components, respectively. We have experimented with different decision thresholds on the Euclidean distance and adopted the value 3 since it performs the best.

4. Experiments and Results

4.1. Validation of the Mahalanobis distance method

We adopted Home A's meter data in the year 2014 (Barker et al., 2012) as the normal case, in which a meter reading is available every 30 minutes for the whole year. We have made two minor modifications on the data set: (1) The data in December were removed, because the meter readings in that month indicate that the appliances were off most of the time, making this period inappropriate to serve as the normal case; (2) The misaligned time stamps due to daylight saving were corrected, which happened at 2014-03-09 02:00 and 2014-11-02 01:00.

Since the data set was collected from a real household setting and there were times during the year when the house was indeed vacant for a prolonged period, e.g., up to tens of hours, we would expect such incidences to be detected by our algorithm as such occupancy changes constitute a deviation from the norm (though it probably meant that the occupants went for a vacation). Therefore, our first order of business is to execute the algorithm for the normal cases. Specifically, we ran the detection algorithm for each 30-minute interval from Feb 2 to Nov 29, and identified 88 anomaly cases. First, many of these cases indeed appeared to be anomalous, such as exhibiting an unconventional fluctuation of power. However, we have no knowledge about what caused such anomalies since the ground-truth occupant's activity data was unavailable. We will publish the full result data set and interested readers can mine details there. Second, even though some misclassified cases are present, a false alarm can be avoided by implementing a validity check in practice. For instance, an automatic phone call can be initiated in the next daytime hour to verify the occupant's safety status.

Next, we tested the algorithm on artificially generated anomaly cases. Given the base data set, we created an anomaly case by the following steps: (1) Randomly pick a time point, e.g., 2014-08-11 11:30:00 as the occurrence time of the anomaly; (2) Freeze the power level of Bedroom Lights, Kitchen Lights and Electric Range (but not the Fridge Range) at the current level for all subsequent time points; (3) Add up all appliances' power by t for each t, to serve as the total consumption x_t read from the smart meter.

We also set a 24-hour timeout limit: if the algorithm is unable to identify the anomaly within 24 hours of its actual occurrence, it is considered a failure case; otherwise, it is a success case and the time to detection (TTD) is recorded. We tested 1000 randomly sampled anomaly cases, among which 767 were successfully detected by the algorithm and the TTD distribution is summarized in Figure 8. Since we used $K = 6$, the TTD is lower bounded at 3 hours, which is also the most common TTD among all success cases. It is worth noting that none of the algorithmic parameters, including the weight vector w, minimum sequence length K and the choices for T, etc., has undergone any systematic tuning to achieve the current performance. This is evidence for the effectiveness of the features as well as the algorithm's cold-start capability.

4.2. Validation of the nested DTW method

We conduct two sets of experiments to test the algorithm's ability to (1) identify the anomaly when one exists, and (2) mark a sequence as normal when no anomaly exists. The data generation and experiment processes are described below.

For the first set of experiments, we simulated 11233 single-anomaly test sequences. In the test sequence for which the anomaly is planted at time t, the detection algorithm that is

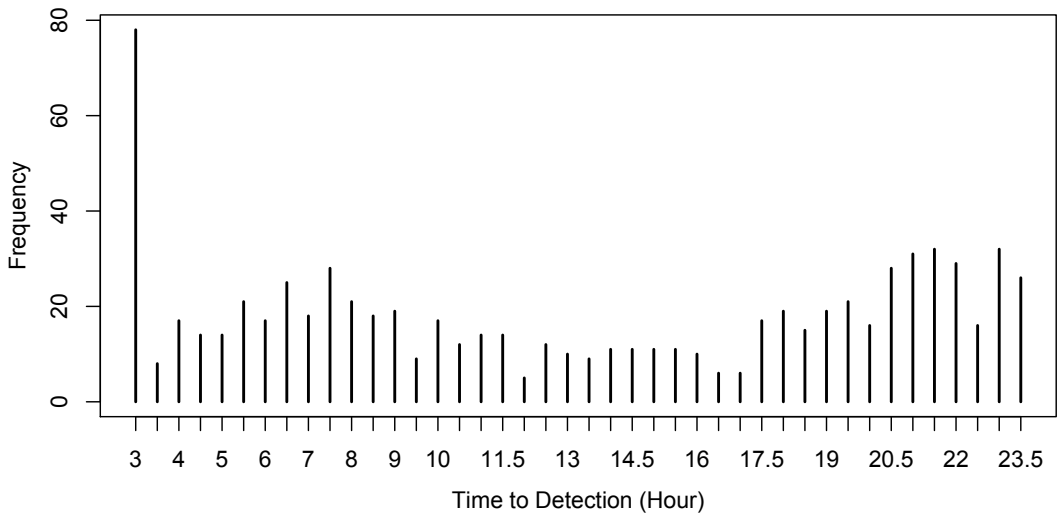

Figure 8: Distribution of the time to detection (TTD) for success cases.

run at time point s, $t \leq s \leq t + 48$, is fed with a data sequence $(x_{s-2T-94}, \ldots, x_s)$ covering two months plus two days, where we adopt $T = 30 \times 48$, i.e., looking back 30 days. For each run at time s, $s \in \{t, \ldots, t + 48\}$, the detection algorithm will be queried to make a decision about whether an anomaly has occurred. If the algorithm answers "yes" in run s and if $s - t \leq h$, then this test sequence will be counted as a true positive (TP) case for the performance category with time-to-detection threshold H, for all $H \geq h$. The test sequence that is not counted as TP for a given performance category will be counted as the false negative (FN) case for this performance category. In the end, we have marked all the test sequences as either FN or TP in this way.

For the second sets of experiments, the algorithm is fed with normal data, and true negatives (TN) and false positives (FP) will be recorded by the same rule as described above. In each experiment, the threshold of the difference between the mean and median of the DTW distances are set as a parameter. Specifically, we experimented 10, 20, 30 and 40 to understand its effect on detection accuracy. The results are summarized in Figure 9 and detailed in Table 4.

Figure 9 shows that the algorithm's performance increases steadily with the increase in detection lag time. At around H25 (meaning that the current time is 12.5 hours after the occurrence of the anomaly), the accuracy increase starts to plateau. The filter value of 40 is also the best one among the four values tested, although its advantage winds off as the lag time increases. To evaluate the performance of the algorithm, we calculate the precision, recall and accuracy, defined as follows. Precision $= (TP)/(TP + FP)$; Recall $= (TP)/(TP + FN)$; Accuracy $= (TP + TN)/(TP + TN + FP + FN)$. We compare the performances under different time window widths allowed for detection, as shown in Figure 10. The results suggest that the algorithm performs well when the detection time window is over 24 (i.e, 12 hours). So we choose H24 as the baseline time window, and perform subsequent comparison experiments under different parameters based on this time window. On this baseline, we find the threshold value (for the difference between mean and median of DTW distances) of 20 performs the best in distinguishing the sudden death event.

Table 4: Proportion of anomalies detected under different lag time using DTW distance method.

Parameter	10		20		30		40	
Lag Time	Detected (%)	Cumulative Detected (%)	Detected (%)	Cumulative Detected (%)	Detected (%)	Cumulative Detected (%)	Detected (%)	Cumulative Detected (%)
H1	6.77	6.77	12.45	12.45	20.10	20.10	29.30	29.30
H2	1.38	8.15	2.08	14.54	3.11	23.21	3.63	32.93
H3	1.39	9.54	2.03	16.57	2.86	26.07	3.16	36.09
H4	1.53	11.07	2.21	18.78	2.83	28.90	3.43	39.52
H5	1.71	12.78	2.42	21.20	3.17	32.07	3.37	42.88
H6	2.05	14.83	2.88	24.07	3.35	35.41	3.46	46.35
H7	2.30	17.13	3.20	27.27	3.48	38.89	3.34	49.68
H8	2.83	19.96	3.53	30.80	3.77	42.67	3.52	53.20
H9	3.11	23.07	3.73	34.53	3.77	46.44	3.46	56.66
H10	3.47	26.54	4.01	38.54	4.04	50.49	3.41	60.07
H11	3.72	30.26	3.77	42.30	3.60	54.08	3.26	63.33
H12	3.47	33.73	3.78	46.09	3.46	57.54	3.04	66.37
H13	3.43	37.16	3.41	49.50	3.08	60.62	2.72	69.09
H14	3.44	40.59	3.19	52.68	2.98	63.61	2.41	71.50
H15	3.40	44.00	3.48	56.16	2.92	66.53	2.49	74.00
H16	3.45	47.45	3.52	59.68	2.97	69.50	2.36	76.36
H17	3.62	51.07	3.55	63.23	2.92	72.42	2.26	78.62
H18	3.69	54.77	3.58	66.81	2.93	75.35	2.27	80.89
H19	3.85	58.61	3.71	70.52	2.67	78.02	2.12	83.01
H20	4.04	62.65	3.45	73.98	2.73	80.75	1.98	84.98
H21	4.25	66.90	3.68	77.66	2.58	83.33	1.82	86.80
H22	4.40	71.30	3.49	81.14	2.48	85.82	1.81	88.61
H23	4.18	75.48	3.21	84.36	2.34	88.16	1.47	90.07
H24	4.28	79.76	3.01	87.37	1.87	90.03	1.24	91.31
H25	3.07	82.84	1.98	89.34	0.95	90.98	0.45	91.77
H26	1.09	83.92	0.62	89.97	0.46	91.44	0.39	92.16
H27	0.85	84.78	0.48	90.45	0.33	91.77	0.28	92.44
H28	0.75	85.52	0.46	90.91	0.35	92.12	0.30	92.74
H29	0.65	86.17	0.37	91.28	0.30	92.42	0.27	93.01
H30	0.66	86.83	0.45	91.74	0.36	92.78	0.32	93.33
H31	0.57	87.40	0.33	92.07	0.24	93.02	0.22	93.55
H32	0.60	88.00	0.28	92.34	0.24	93.26	0.23	93.79
H33	0.53	88.53	0.31	92.66	0.25	93.51	0.23	94.02
H34	0.41	88.94	0.28	92.93	0.20	93.71	0.20	94.22
H35	0.44	89.38	0.23	93.16	0.20	93.91	0.19	94.41
H36	0.40	89.78	0.23	93.39	0.19	94.10	0.19	94.60
H37	0.50	90.28	0.28	93.68	0.22	94.32	0.21	94.81
H38	0.44	90.71	0.28	93.96	0.22	94.54	0.20	95.01
H39	0.29	91.01	0.23	94.19	0.21	94.76	0.20	95.21
H40	0.41	91.42	0.23	94.42	0.20	94.96	0.20	95.41
H41	0.43	91.85	0.24	94.66	0.22	95.18	0.22	95.63
H42	0.36	92.20	0.28	94.93	0.26	95.44	0.24	95.87
H43	0.34	92.54	0.24	95.17	0.19	95.63	0.17	96.04
H44	0.34	92.88	0.24	95.42	0.20	95.82	0.16	96.20
H45	0.34	93.22	0.22	95.64	0.18	96.00	0.16	96.36
H46	0.26	93.47	0.17	95.81	0.16	96.16	0.13	96.49
H47	0.21	93.69	0.12	95.93	0.12	96.29	0.12	96.61
H48	0.27	93.96	0.18	96.11	0.14	96.43	0.12	96.73
Failed	6.04	-	3.89	-	3.57	-	3.27	-

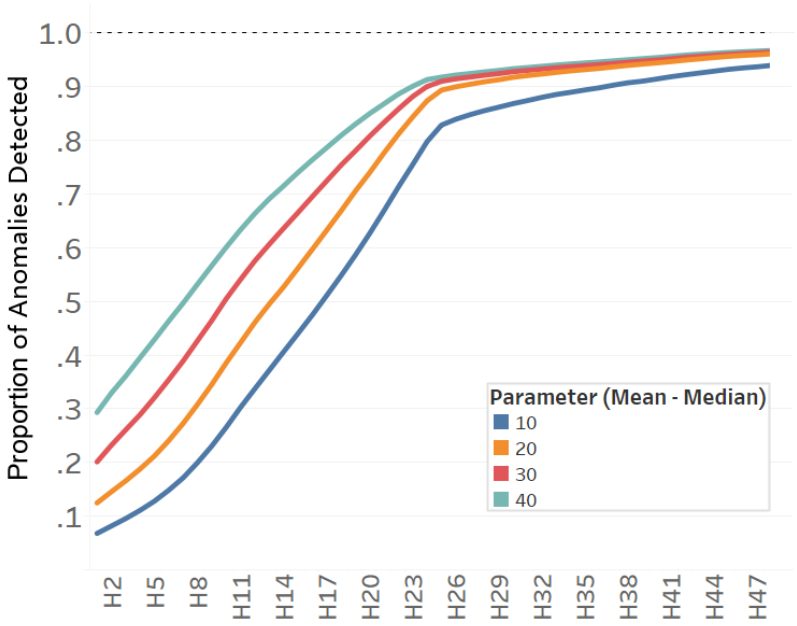

Figure 9: Proportion of anomalies detected under different lag time and four different parameter values.

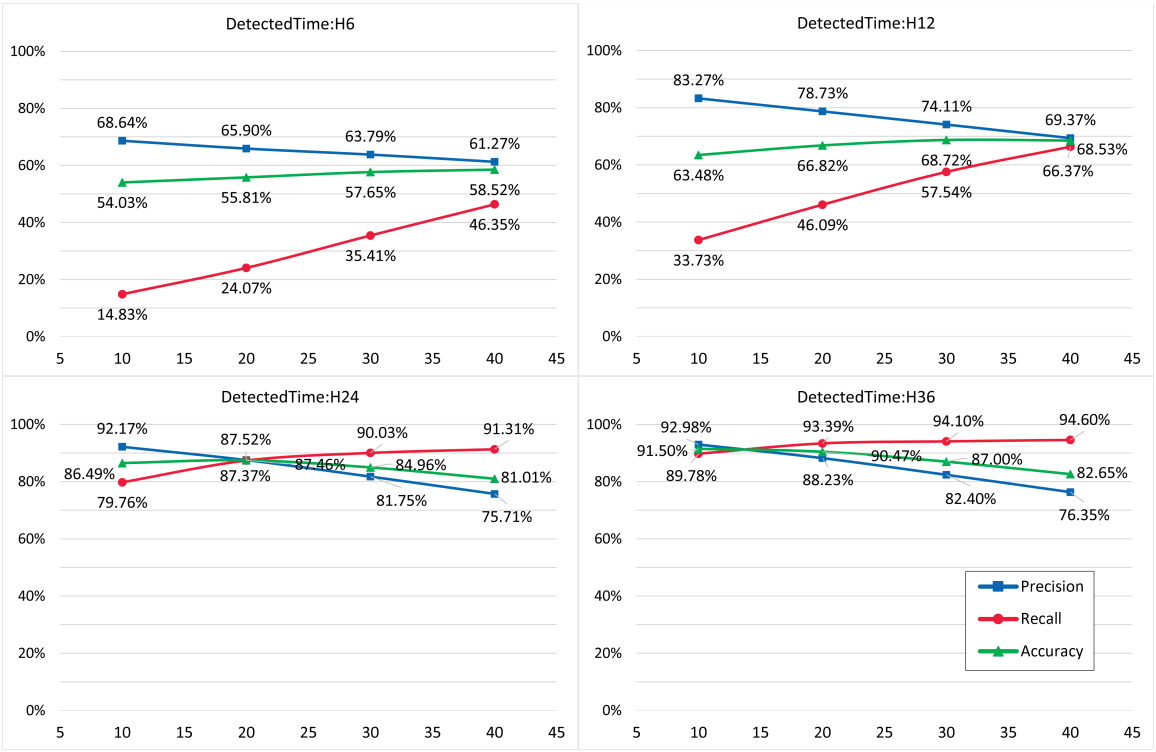

Figure 10: Performance of DTW method under different time-to-detection and filter parameter threshold values. In each subplot, the horizontal axis represents the filter parameter threshold and the vertical axis represents the Precision, Recall and Accuracy value in percentage.

The preceding method set $T = 30 \times 48$, meaning that the DTW distance is tracked for the past 30 days. Now, we experiment with different values for T in search for a more

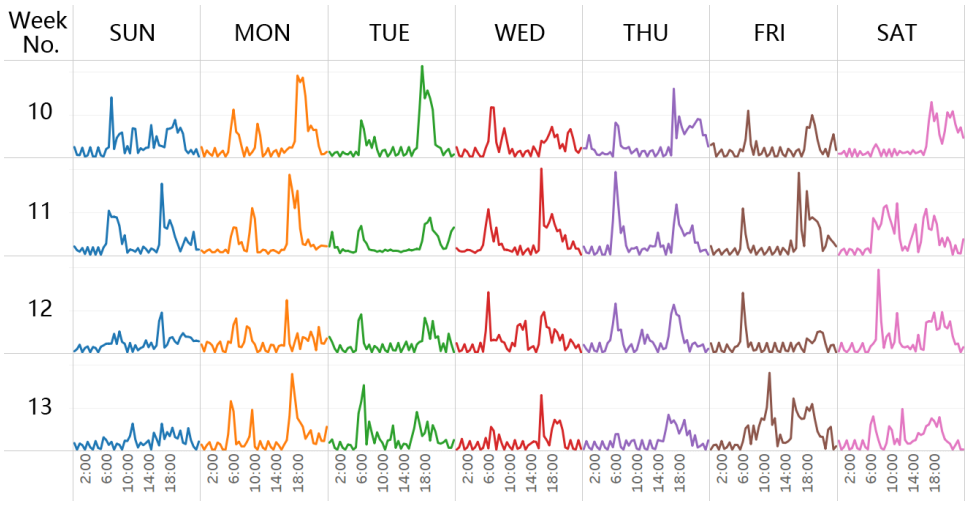

Figure 11: Weekly and daily patterns of total electricity consumption.

effective "look-back" period. Figure 11 shows the electricity consumption in 28 days of March, beginning from the 10th week of the year and ending on the Saturday of the 13th week of the year. We notice that patterns on weekdays are quite similar with comparable peaks, while the peaks on weekends appear less uniform. So, looking back seven days is able to make normal cases match better, and hence reduce the false negative rate. For this observation, we choose $T = 7 \times 48$ and repeat the experiments. Table 5 summarizes the comparison results.

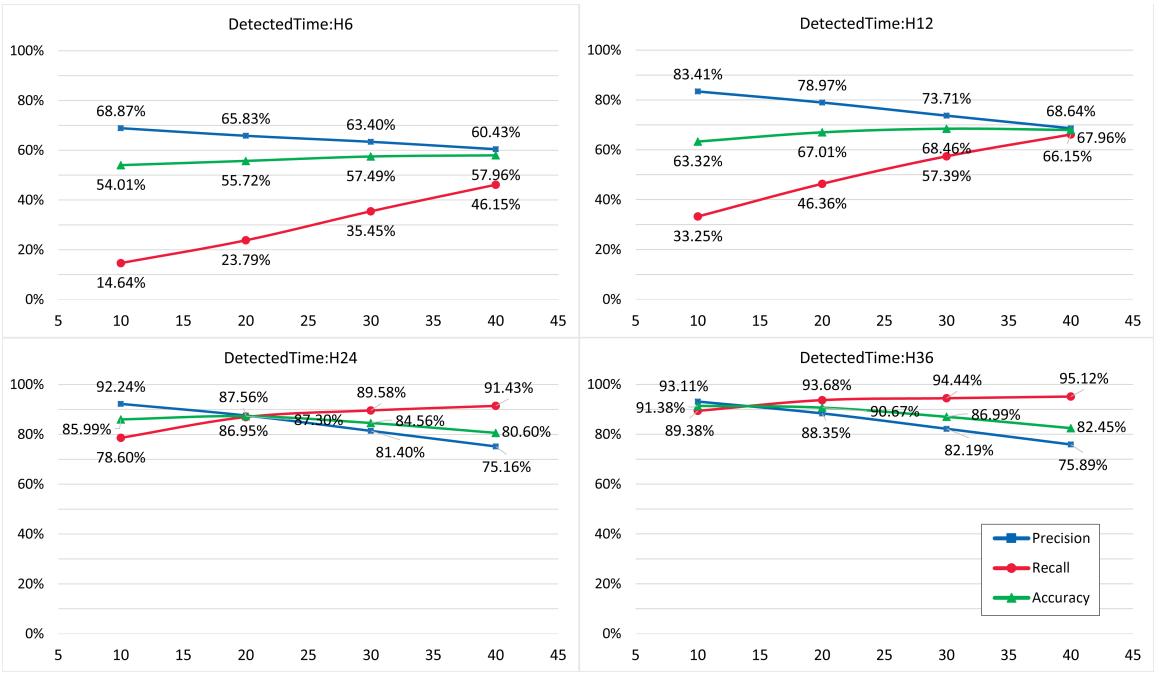

Figure 12: Performance of the nested DTW method under different parameter settings with a 7-day look-back period.

The comparisons in Table 5 and in Figure 12 show that the seven-day-look-back method

Table 5: Summary statistics of normal case data.

Parameter	Case	$T = 30 \times 48$		$T = 7 \times 48$	
		Count	Percentage	Count	Percentage
10	FP	761	6.78%	743	6.62%
	TN	10471	93.22%	10489	93.38%
20	FP	1399	12.46%	1387	12.35%
	TN	9833	87.54%	9845	87.65%
30	FP	2258	20.10%	2299	20.47%
	TN	8974	79.90%	8933	79.53%
40	FP	3291	29.30%	3395	30.23%
	TN	7941	70.70%	7837	69.77%

* FP: False Positive; TN: True Negative

gives limited improvement compared to the previous method, especially when the detection time window is short. It may be due to the following reasons.

- Well matched patterns only help improve the accuracy for normal cases. In other words, they only help reduce the false negative rate. A "Sudden death pattern" is very different from a normal pattern, which does not ostensibly enlarge the DTW distances.
- The threshold is small enough to filter positive cases, and some false positive cases due to threshold can be corrected by the "mean and median" filter. Thus, the method does not contribute much in reducing the false positive rate.
- Patterns on weekends are not well matched, which also affects the actual performance.

4.3. Comparison between the two algorithms and with the baseline approach

We design experiments to compare the performances of the proposed algorithms with the baseline approach. To ensure comparability, we run the Mahalanobis distance method and the prediction-based baseline method on the same data sets (the 11233 single-anomaly test sequences and the nominal data) with the same slide window width i.e., $T = 30 \times 48$, as were used in testing the nested DTW method. For the Mahalanobis distance method we experiment different levels of weight, that is, taking f_t at values $[0.1, 0.5, 0.4]$, $[0.5, 0.1, 0.4]$, $[0.5, 0.4, 0.1]$ and $[0.6, 0.3, 0.1]$, and label these settings as 154, 514, 541 and 631, respectively. Notice that the last setting is the same as that used in Section 3.3. The results are highlighted in Figures 13 and 15 and detailed in Table 7.

From Figure 13, we can see that the performance of the Mahalanobis distance method varies with weight settings. The precision, recall and accuracy are high when the detection window is more than 24 periods (12 hours) under the weight setting of $[0.1, 0.5, 0.4]$, while the performance under other weight settings is less attractive. Figure 15 more clearly reveals that the parameter setting $[0.1, 0.5, 0.4]$ generates the dominating performance among the four settings tested, regardless of the detection lag time. This observation suggests that the variance in energy consumption over the past 6 hours is a more important anomaly indicator

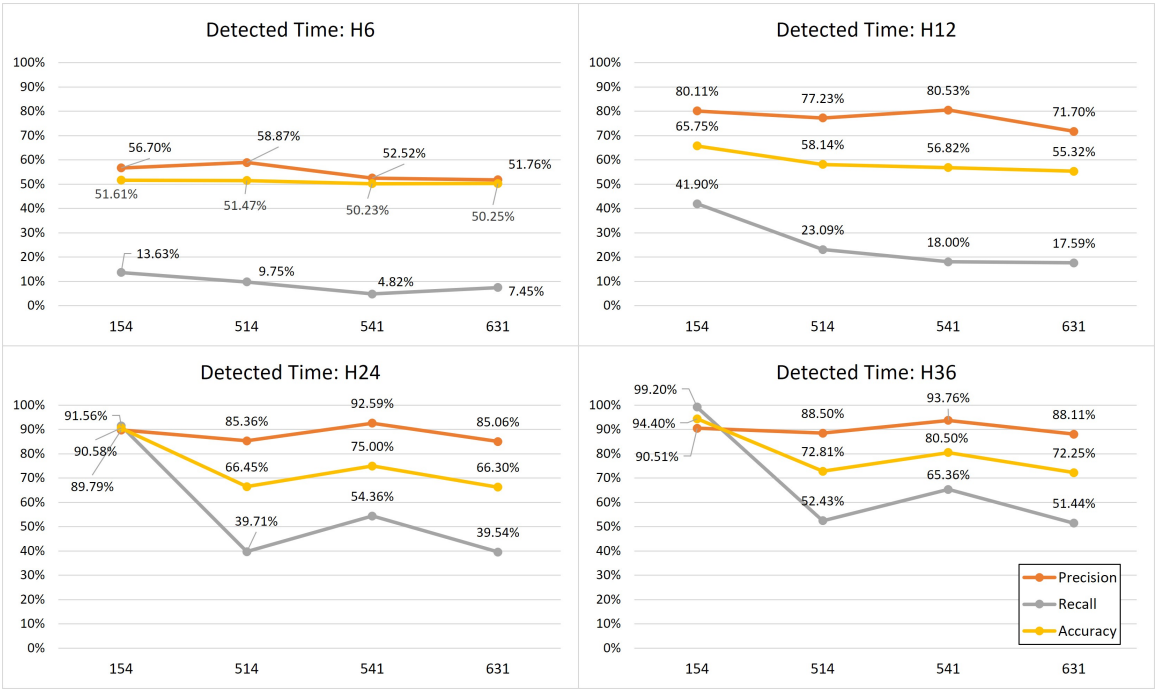

Figure 13: Performances of Mahalanobis distance method under different time to detection.

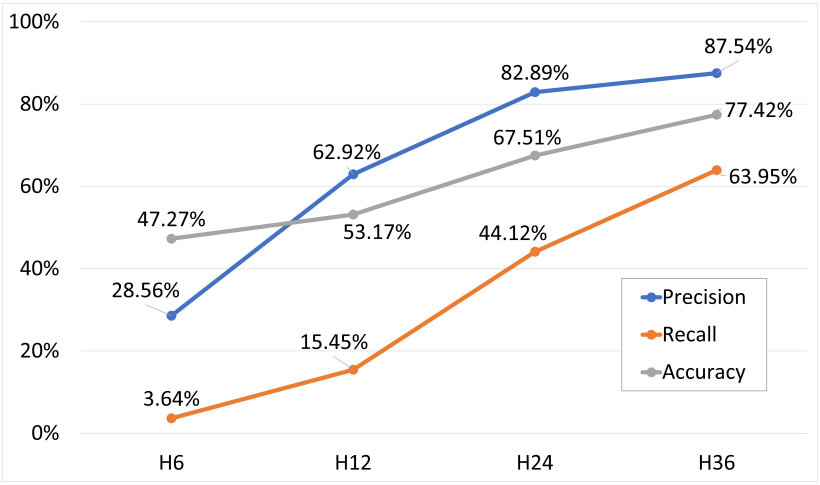

Figure 14: Performance of the prediction-based baseline method under different time-to-detection values.

than the variance over the past 24 hours. Comparing Figure 10 and Figure 13, we can see that the DTW method performs slightly better than the Mahalanobis distance method at an early stage in H12, that is, within the first 6 hours after the anomaly occurs. On the other hand, the Mahalanobis distance method performs better in H24 and H36 under the wight setting of $[0.1, 0.5, 0.4]$.

The performance of the baseline approach at different time-to-detection values is plotted in Figure 14. We can see that, as expected, the detection becomes more reliable as more time is allowed, though the practical usefulness of the detection would decrease over time. Moreover, the precision tends to plateau at below 90%, approaching an intrinsic upper limit imposed

Table 6: Performance comparison among three methods.

Methods		$\text{ARIMA}_{(0,1,1)\times(2,0,0)_{48}}$	$\text{Mahalanobis}_{(154)}$	$\text{Nested DTW}_{(20)}$
Avg. Detect Time (Hrs.)		12.4	7.4	7.5
Precision (%)	H6	28.56	56.70	65.90
	H12	62.92	80.11	78.73
	H24	82.89	89.79	87.52
	H36	87.54	90.51	88.23
Recall (%)	H6	3.64	13.63	24.07
	H12	15.45	41.90	46.09
	H24	44.12	91.56	87.37
	H36	63.95	99.20	93.39
Accuracy (%)	H6	47.27	51.61	55.81
	H12	53.17	65.75	66.82
	H24	67.51	90.58	87.46
	H36	77.42	94.40	90.47

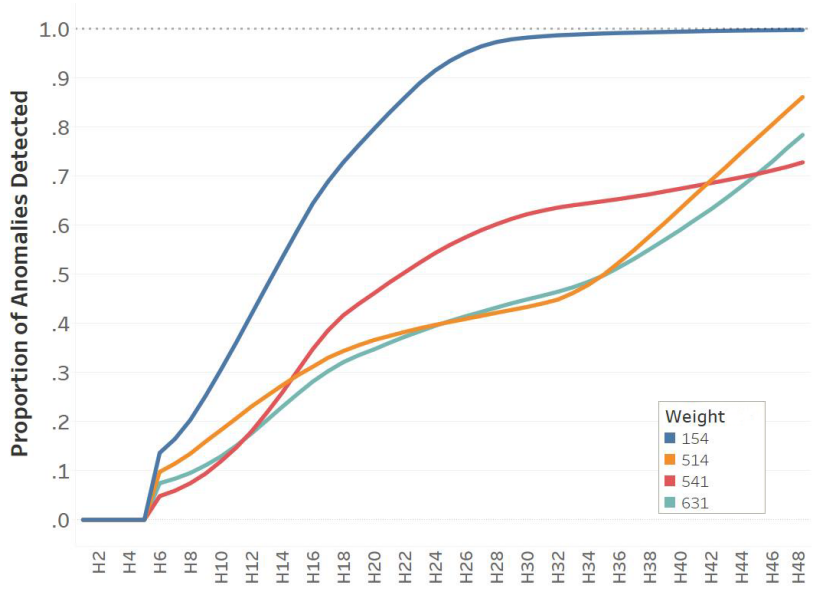

Figure 15: Proportion of anomalies detected by the Mahalanobis distance method under different lag time and different parameter settings.

by the time-series prediction model. The comparison of all three methods is listed in Table 6. It is apparent that the Mahalanobis method and the nested DTW method both outperform the baseline method in all aspects. Part of the reason is that the proposed methods not only model the distribution of the past observations, but also leverage the characteristics of the target events which help guide the parameter setting. Compared to the prediction-based method, our unsupervised learning approaches are more adaptable to complex and fast-changing patterns, apart from their ease of implementation on smart devices.

Table 7: Proportion of anomalies detected under different lag time using Mahalanobis distance method.

Weight	154		514		541		631	
Lag Time	Detected (%)	Cumulative Detected (%)	Detected (%)	Cumulative Detected (%)	Detected (%)	Cumulative Detected (%)	Detected (%)	Cumulative Detected (%)
H1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
H6	13.63	13.63	9.75	9.75	4.82	4.82	7.45	7.45
H7	2.85	16.48	1.71	11.46	1.11	5.93	0.93	8.38
H8	3.85	20.32	2.01	13.47	1.51	7.44	1.17	9.54
H9	4.91	25.24	2.47	15.94	1.97	9.41	1.56	11.10
H10	5.34	30.58	2.34	18.28	2.50	11.91	1.80	12.90
H11	5.54	36.12	2.38	20.65	2.80	14.71	2.19	15.09
H12	5.79	41.90	2.44	23.09	3.29	18.00	2.50	17.59
H13	5.76	47.66	2.15	25.25	3.82	21.82	2.69	20.28
H14	5.71	53.37	2.14	27.38	4.06	25.88	2.72	23.00
H15	5.62	58.99	2.02	29.40	4.47	30.35	2.63	25.63
H16	5.45	64.44	1.77	31.18	4.42	34.77	2.53	28.16
H17	4.50	68.93	1.84	33.02	3.82	38.59	2.13	30.29
H18	3.92	72.85	1.37	34.39	3.09	41.68	1.84	32.13
H19	3.47	76.32	1.18	35.57	2.31	44.00	1.41	33.54
H20	3.35	79.67	1.05	36.62	2.15	46.14	1.19	34.73
H21	3.27	82.93	0.83	37.45	2.21	48.35	1.32	36.05
H22	3.09	86.02	0.82	38.27	2.04	50.39	1.25	37.29
H23	3.02	89.04	0.76	39.03	2.04	52.43	1.12	38.41
H24	2.52	91.56	0.69	39.71	1.93	54.36	1.13	39.54
H25	2.02	93.58	0.62	40.34	1.72	56.08	1.00	40.54
H26	1.61	95.19	0.61	40.94	1.51	57.59	0.94	41.48
H27	1.27	96.47	0.61	41.56	1.42	59.00	0.86	42.35
H28	0.89	97.36	0.63	42.19	1.22	60.22	0.90	43.25
H29	0.56	97.92	0.58	42.77	1.10	61.33	0.87	44.12
H30	0.35	98.26	0.61	43.37	0.94	62.27	0.79	44.91
H31	0.25	98.51	0.69	44.07	0.71	62.98	0.75	45.66
H32	0.21	98.73	0.79	44.86	0.63	63.62	0.80	46.46
H33	0.12	98.85	1.35	46.21	0.47	64.09	0.92	47.38
H34	0.13	98.99	1.67	47.89	0.41	64.50	1.11	48.49
H35	0.12	99.11	2.05	49.93	0.44	64.93	1.30	49.79
H36	0.09	99.20	2.50	52.43	0.43	65.36	1.65	51.44
H37	0.06	99.26	2.55	54.98	0.49	65.85	1.75	53.19
H38	0.07	99.33	2.78	57.76	0.48	66.33	1.92	55.11
H39	0.07	99.40	2.76	60.52	0.58	66.91	1.95	57.06
H40	0.06	99.47	2.88	63.40	0.55	67.46	1.99	59.06
H41	0.06	99.53	2.89	66.30	0.57	68.03	2.12	61.18
H42	0.07	99.60	2.84	69.14	0.58	68.61	2.07	63.25
H43	0.05	99.65	2.78	71.91	0.59	69.20	2.27	65.52
H44	0.04	99.70	2.93	74.84	0.61	69.80	2.39	67.91
H45	0.04	99.73	2.84	77.68	0.63	70.44	2.48	70.39
H46	0.03	99.76	2.81	80.49	0.74	71.17	2.55	72.94
H47	0.04	99.80	2.86	83.35	0.76	71.93	2.84	75.78
H48	0.03	99.82	2.78	86.13	0.90	72.83	2.64	78.42
Failed	0.18	-	13.87	-	27.17	-	21.58	-

5. Discussions

5.1. Rationale for unsupervised learning

The proposed methods are both unsupervised learning methods that are based purely on the historical data with limited preprocessing or artificial treatments. In contrast, supervised learning would require training a model using labeled data. There are some challenges in preparing labeled data for this application. Due to the great variability of the instantaneous power consumption, it is impossible to tell whether a single data/time point is anomalous or not based on observations leading to this point. Instead, an anomalous event always reveals itself as a sequence of data points whose pattern is significantly different from what is expected. Since the ground-truth anomalous event occurs at a specific time point but its effect unfolds over the next several time points, it is difficult to select and label an anomalous sequence without imposing subjectivity, i.e., decisions such as how many time points after the event will be considered as part of the anomalous sequence. Furthermore, supervised learning methods such as logistics regression, decision trees and neural networks would require a model (i.e., coefficients, rule sets, or parameters) to be pre-trained offline, which would lack flexibility in dealing with occupant's behavioral change and routine differences in different households. More frequent model updates would be required which might pose additional challenges in device deployment and maintenance. In comparison, the data-driven unsupervised learning methods are relatively immune to these nuisances. Therefore, we believe that the distance based unsupervised learning methods proposed herein are novel and useful contributions.

5.2. Rationale for distance-based method

The baseline method, ARIMA, is a statistical model that combines autoregressive and moving average components to learn patterns in time sequence data. The model is parametric and assumes that the pattern learned from historical data do not change over time. However, in the problem of detecting abnormal inactivity based on smart meter readings, the mean, variance, and covariance of electricity consumption within any practical time window are not constant, rendering ARIMA less effective at characterizing the relevant features. In comparison, our distance-based unsupervised learning methods are specifically designed to capture the complex and changing patterns in household electricity consumption, thus are more capable of detecting anomalous events.

5.3. Notes on implementation

Both algorithms developed in this paper have been coded into computer programs (Python) that involve only basic arithmetic operations independent of any third-party library. The programs are highly portable and can be directly transcribed into C or MicroPython programs suitable for deployment on a microcontroller unit (MCU). At any time, at most 4500 single-precision input values (i.e., three months' of meter readings) need to be stored, corresponding to a RAM requirement well under the capacity of most MCUs. The anomaly detection computation is carried out once every 30 minutes (the actual computation only takes no more than a couple of seconds) and in most of the idle time, the MCU can either enter a low-power sleep mode or handle other smart meter functions, such as taking energy measurements and aggregation. Figure 16 illustrates the schematic overview of our proposed implementation method. In this scheme, an Espressif ESP32 MCU equipped with a 240 MHz single-core

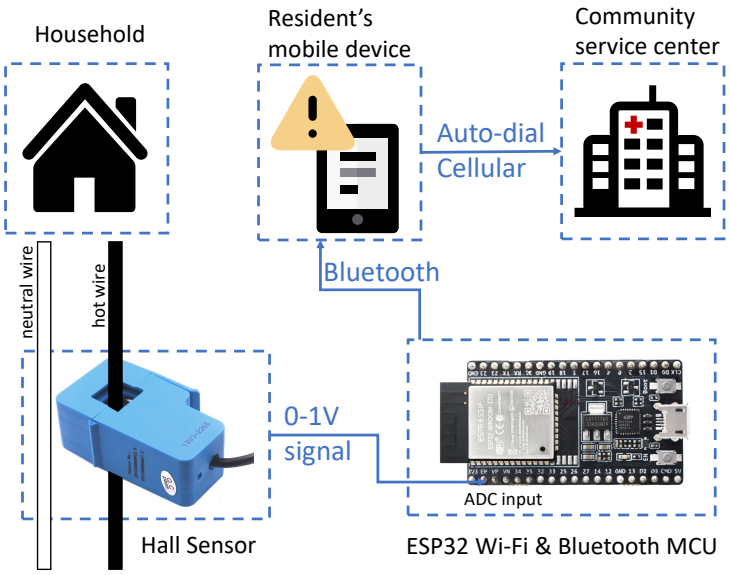

Figure 16: System architecture for practical deployment of the proposed algorithms.

CPU, a 2.4GHz Wi-Fi radio and a Bluetooth lower energy (BLE) radio is used to handle all of the data aggregation, anomaly detection and communication functions. Specifically, the electric current in the household's main circuit is measured via a Hall sensor and the sampled measurements (voltage signals in the 0 to 1 V range) are read using the analog-to-digital converter (ADC) peripheral of the MCU. The MCU then aggregates these samples into Watts for the 30-min interval at the nominal voltage level, e.g., 110 V. The Watt values are stored in a first-in-first-out (FIFO) queue in the RAM. Every 30 minutes, the anomaly detection algorithm is executed based on Watt values in the queue. If an anomaly is detected, the MCU will ping the resident's cellphone via a pre-installed App through BLE (or Wi-Fi) and request a simple response, such as a tap on the cellphone screen. The resident, if safe and healthy, can simply tap the phone screen to cancel the false alarm. On the other hand, if the pings are not responded to for contiguous periods, outside assistance will be called automatically and a pre-recorded voice message will be played to request an emergency home visit.

This implementation scheme offloads as much computation as possible to the edge device (i.e., the ESP32 MCU) instead of sending raw meter readings to a server for central processing. Compared to the alternative, see, e.g., Patrono et al. (2018), this scheme is more efficient in the use of communication bandwidth and potentially more scalable. Yet another implementation approach is to inject the anomaly detection functions into the onboard software suite of an existing smart meter brand, possibly as a third-party add-in module, although this approach would depend on the manufacturer's willingness to foster an open hardware interface along with a software marketplace. We believe that the market is evolving in the promising direction as the IoT and edge AI technologies mature.

6. Conclusion

Smart meter data analytics provides a promising pathway to non-intrusive household anomaly monitoring. The electricity consumption data aggregated every 30 minutes do

not reveal privacy-leaking activities in the house, yet, the data can be leveraged to detect major abnormal events, such as the unattended death of the occupant in a single-person home. Timely detection of such events is important both socially and economically. In this paper, we have proposed and compared two effective anomaly detection algorithms operating on electricity consumption data recorded by residential smart meters. Effective features were extracted from the time series data and two methods based on the weighted Mahalanobis distance and the nested DTW distance have been developed to identify single-point anomalies. A sequence smoothing technique was used to effectively hedge against false alarms. The algorithms have demonstrated remarkable efficacy on test scenarios from real data sets. Comparison results have suggested that the Mahalanobis distance based method is more reliable at a longer lag time, while the nested DTW method performs better when the lag time is shorter. Both methods do not require offline training or parameter tuning to work well, and thus can be programmed directly into the smart meter hardware for practical implementation.

While the current work operates on an individual household, follow-on research could extend the scope to cover multiple households in a neighborhood, which would introduce more exploitable features, correlations and patterns. The system implementation and integration challenges, including the hardware and software design, device networking and outbound communication protocols, etc., need also be investigated in future work. More observational study and data analysis based on real-life implementation are also worthwhile topics for future investigation.

Acknowledgement

The first and second authors are supported by the National Science Foundation under grant number CMMI-1944068. The authors would like to thank the associate editor and two anonymous reviewers whose comments helped improve the quality of the paper.

References

Aach, J. and Church, G. M. (2001). Aligning gene expression time series with time warping algorithms. *Bioinformatics*, 17(6):495–508.

Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., and Weng, T. (2010). Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 1–6, New York, NY, USA. ACM.

Anderson, D., Luke, R. H., Keller, J. M., Skubic, M., Rantz, M., and Aud, M. (2009). Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer vision and image understanding*, 113(1):80–89.

Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. A., and Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings*, 144:191–206.

Ardakanian, O., Bhattacharya, A., and Culler, D. (2018). Non-intrusive occupancy monitoring for energy conservation in commercial buildings. *Energy and Buildings*, 179:311–323.

Aziz, S., Naqvi, S. Z. H., Khan, M. U., and Aslam, T. (2020). Electricity theft detection using empirical mode decomposition and k-nearest neighbors. In *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, pages 1–5. IEEE.

Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., and Albrecht, J. (2012). Smart*: An open data set and tools for enabling research in sustainable homes. In *Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012)*.

Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., and Santini, S. (2014). The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM conference on embedded systems for energy-efficient buildings*, pages 80–89.

Becker, V. and Kleiminger, W. (2018). Exploring zero-training algorithms for occupancy detection based on smart meter measurements. *Computer Science-Research and Development*, 33(1):25–36.

Brown, T. M., McCabe, S. A., and Wellford, C. F. (2007). *Global positioning system (GPS) technology for community supervision: Lessons learned*. Number 219376. Noblis.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58.

Chen, B., Sinn, M., Ploennigs, J., and Schumann, A. (2014). Statistical anomaly detection in mean and variation of energy consumption. In *2014 22nd International Conference on Pattern Recognition*, pages 3570–3575. IEEE.

Chen, D., Barker, S., Subbaswamy, A., Irwin, D., and Shenoy, P. (2013a). Non-intrusive occupancy monitoring using smart meters. In *Proceedings of ACM BuildSys 2013*.

Chen, D., Barker, S., Subbaswamy, A., Irwin, D., and Shenoy, P. (2013b). Non-intrusive occupancy monitoring using smart meters. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, pages 1–8.

Devlin, M. A. and Hayes, B. P. (2019). Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data. *IEEE Transactions on Consumer Electronics*, 65(3):339–348.

Erickson, V. L. and Cerpa, A. E. (2010). Occupancy based demand response hvac control strategy. In *Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 7–12, New York, NY, USA. ACM.

Gao, J., Song, X., Wen, Q., Wang, P., Sun, L., and Xu, H. (2020). RobustTAD: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint arXiv:2002.09545*.

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, 31(i07).

Gu, X., Akoglu, L., and Rinaldo, A. (2019). Statistical analysis of nearest neighbor methods for anomaly detection. *arXiv preprint arXiv:1907.03813*.

Guerra-Santin, O. and Tweed, C. A. (2015). In-use monitoring of buildings: An overview of data collection methods. *Energy and Buildings*, 93:189–207.

Hattori, S. and Shinohara, Y. (2017). Actual consumption estimation algorithm for occupancy detection using low resolution smart meter data. In *SENSORNETS*, pages 39–48.

Himeur, Y., Alsalemi, A., Bensaali, F., and Amira, A. (2020). A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cognitive Computation*, 12(6):1381–1401.

Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., and Amira, A. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, 287:116601.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, 27:1–22.

Jakkula, V. and Cook, D. (2010). Outlier detection in smart environment structured power datasets. In *2010 sixth international conference on intelligent environments*, pages 29–33. IEEE.

Kavousian, A., Rajagopal, R., and Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, 55(Supplement C):184 – 194.

Kekade, S., Hseieh, C.-H., Islam, M. M., Atique, S., Mohammed Khalfan, A., Li, Y.-C., and Abdul, S. S. (2018). The usefulness and actual use of wearable devices among the elderly population. *Computer Methods and Programs in Biomedicine*, 153:137–159.

Kleiminger, W., Beckel, C., and Santini, S. (2015). Household occupancy monitoring using electricity meters. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 975–986.

Kovacs-Vajna, Z. M. (2000). A fingerprint verification system based on triangular matching and dynamic time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1266–1276.

Laaroussi, Y., Bahrar, M., El Mankibi, M., Draoui, A., and Si-Larbi, A. (2020). Occupant presence and behavior: A major issue for building energy performance simulation and assessment. *Sustainable Cities and Society*, 63:102420.

Laurikkala, J., Juhola, M., and Kentala, E. (2000). Informal identification of outliers in medical data. In *Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology*, pages 20–24.

Liu, C., Akintayo, A., Jiang, Z., Henze, G. P., and Sarkar, S. (2018). Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network. *Applied Energy*, 211:1106–1122.

Liu, X., Iftikhar, N., Nielsen, P. S., and Heller, A. (2016). Online anomaly energy consumption detection using lambda architecture. In *International Conference on Big Data Analytics and Knowledge Discovery*, pages 193–209. Springer.

Liu, X. and Nielsen, P. S. (2018). Scalable prediction-based online anomaly detection for smart meter data. *Information Systems*, 77:34–47.

Martani, C., Lee, D., Robinson, P., Britter, R., and Ratti, C. (2012). Enernet: Studying the dynamic relationship between building occupancy and energy consumption. *Energy and Buildings*, 47(Supplement C):584 – 591.

Müller, M. (2007). *Dynamic Time Warping*, pages 69–84. Springer Berlin Heidelberg, Berlin, Heidelberg.

Munich, M. E. and Perona, P. (1999). Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 108–115. IEEE.

Myers, C. and Rabiner, L. (1981a). A level building dynamic time warping algorithm for connected word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(2):284–297.

Myers, C., Rabiner, L., and Rosenberg, A. (1980). Performance tradeoffs in dynamic time warping algorithms for isolated word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(6):623–635.

Myers, C. S. and Rabiner, L. R. (1981b). A comparative study of several dynamic time-warping algorithms for connected-word recognition. *Bell System Technical Journal*, 60(7):1389–1409.

Newsham, G. R. and Birt, B. J. (2010). Building-level occupancy data to improve arima-based electricity use forecasts. In *Proceedings of the 2Nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 13–18, New York, NY, USA. ACM.

Patrono, L., Rametta, P., and Meis, J. (2018). Unobtrusive detection of home appliance's usage for elderly monitoring. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6. IEEE.

Qiu, H., Tu, Y., and Zhang, Y. (2018). Anomaly detection for power consumption patterns in electricity early warning system. In *2018 tenth international conference on advanced computational intelligence (ICACI)*, pages 867–873. IEEE.

Rafsanjani, H. N. and Ahn, C. (2016). Linking building energy-load variations with occupants' energy-use behaviors in commercial buildings: non-intrusive occupant load monitoring (niolm). *Procedia Engineering*, 145:532–539.

Rafsanjani, H. N., Ahn, C. R., and Chen, J. (2018). Linking building energy consumption with occupants' energy-consuming behaviors in commercial buildings: Non-intrusive occupant load monitoring (niolm). *Energy and Buildings*, 172:317–327.

Razavi, R., Gharipour, A., Fleury, M., and Akpan, I. J. (2019). Occupancy detection of residential buildings using smart meter data: A large-scale study. *Energy and Buildings*, 183:195–208.

Reeder, B. and David, A. (2016). Health at hand: A systematic review of smart watch uses for health and wellness. *Journal of biomedical informatics*, 63:269–276.

Shamim, G. and Rihan, M. (2020). Multi-domain feature extraction for improved clustering of smart meter data. *Technology and Economics of Smart Grids and Sustainable Energy*, 5(1):1–8.

Tang, G., Wu, K., Lei, J., and Xiao, W. (2015). The meter tells you are at home! non-intrusive occupancy detection via load curve data. In *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 897–902. IEEE.

Tran, L., Fan, L., and Shahabi, C. (2016). Distance-based outlier detection in data streams. *Proceedings of the VLDB Endowment*, 9(12):1089–1100.

Visconti, P., Costantini, P., de Fazio, R., Lay-Ekuakille, A., and Patrono, L. (2019). A sensors-based monitoring system of electrical consumptions and home parameters remotely managed by mobile app for elderly habits' control. In *2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI)*, pages 264–269. IEEE.

Wang, Y., Chen, Q., Hong, T., and Kang, C. (2018). Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid*, 10(3):3125–3148.

Whipple, J., Arensman, W., and Boler, M. S. (2009). A public safety application of gps-enabled smartphones and the android operating system. In *2009 IEEE International Conference on Systems, Man and Cybernetics*, pages 2059–2061. IEEE.

Wu, M. and Luo, J. (2019). Wearable technology applications in healthcare: a literature review. *Online J Nurs Inform*, 23(3).

Yan, S., Li, K., Wang, F., Ge, X., Lu, X., Mi, Z., Chen, H., and Chang, S. (2020). Time–frequency feature combination based household characteristic identification approach using smart meter data. *IEEE Transactions on Industry Applications*, 56(3):2251–2262.

Yijia, T. and Hang, G. (2016). Anomaly detection of power consumption based on waveform feature recognition. In *2016 11th International Conference on Computer Science & Education (ICCSE)*, pages 587–591. IEEE.

Yip, S.-C., Tan, W.-N., Tan, C., Gan, M.-T., and Wong, K. (2018). An anomaly detection framework for identifying energy theft and defective meters in smart grids. *International Journal of Electrical Power & Energy Systems*, 101:189–203.

Zou, H., Jiang, H., Yang, J., Xie, L., and Spanos, C. (2017). Non-intrusive occupancy sensing in commercial buildings. *Energy and Buildings*, 154:633–643.

Zyabkina, O., Domagk, M., Meyer, J., and Schegner, P. (2018). A feature-based method for automatic anomaly identification in power quality measurements. In *2018 IEEE international conference on probabilistic methods applied to power systems (PMAPS)*, pages 1–6. IEEE.