DOI: 10.1016/j.enbuild.2018.12.020

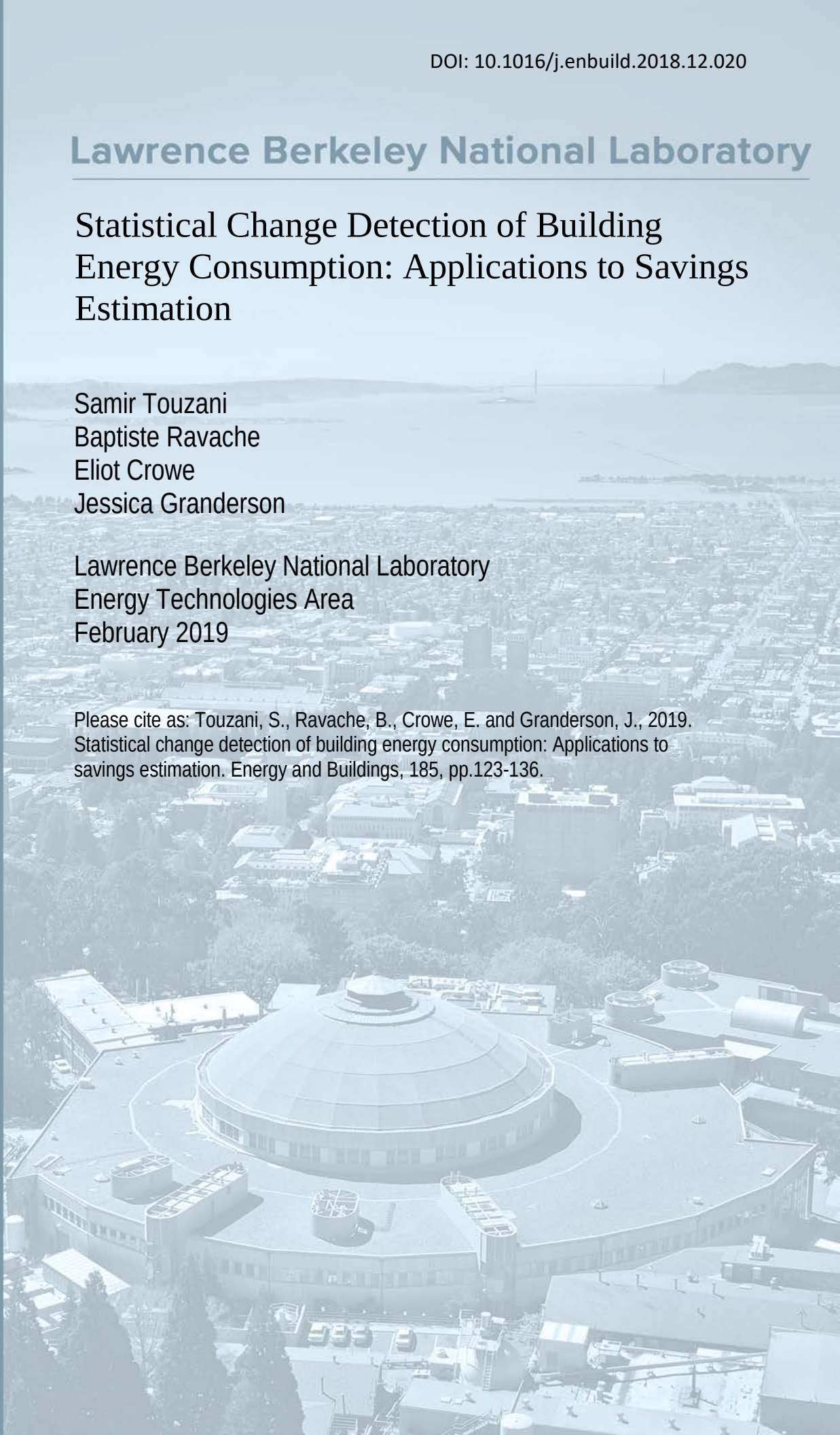
Lawrence Berkeley National Laboratory

Statistical Change Detection of Building Energy Consumption: Applications to Savings Estimation

Samir Touzani
Baptiste Ravache
Eliot Crowe
Jessica Granderson

Lawrence Berkeley National Laboratory
Energy Technologies Area
February 2019

Please cite as: Touzani, S., Ravache, B., Crowe, E. and Granderson, J., 2019. Statistical change detection of building energy consumption: Applications to savings estimation. Energy and Buildings, 185, pp.123-136.



Disclaimer:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Statistical Change Detection of Building Energy Consumption: Applications to Savings Estimation

Samir Touzani, Baptiste Ravache, Eliot Crowe and Jessica Granderson

Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA

Abstract

The surge in interval meter data availability and associated activity in energy data analytics has inspired new interest in advanced methods for building efficiency savings estimation. Statistical and machine learning approaches are being explored to improve the energy baseline models used to measure and verify savings. One outstanding challenge is the ability to identify and account for operational changes that may confound savings estimates. In the measurement and verification (M&V) context, 'non-routine events' (NREs) cause changes in building energy use that are not attributable to installed efficiency measures, and not accounted for in the baseline model's independent variables. In the M&V process NREs must be accounted for as 'adjustments' to appropriately attribute the estimated energy savings to the specific efficiency interventions that were implemented. Currently this is a manual and custom process, conducted using professional judgment and engineering expertise. As such it remains a barrier in scaling and standardizing meter-based savings estimation.

In this work, a data driven methodology was developed to (partially) automate, and therefore streamline the process of detecting NREs in the post-retrofit period and making associated savings adjustments. The proposed NRE detection algorithm is based on a statistical change point detection method and a dissimilarity metric. The dissimilarity metric measures the proximity between the actual time series of the post-retrofit energy consumption and the projected baseline, which is generated using a statistical baseline model. The suggested approach for NRE adjustment involves the NRE detection algorithm, the M&V practitioner, and a regression modeling algorithm. The performance of the detection and adjustment algorithm was evaluated using a simulation-generated test data set, and two benchmark algorithms. Results show a high true positive detection rate (75%-100% across the test cases), higher than ideal false positive detection rates (20%-70%), and low errors in energy adjustment (<0.7%). These results indicate that the algorithm holds for helping M&V practitioners to streamline the process of handling NREs. Moreover, the change point algorithm and underlying statistical principles could prove valuable for other building analytics applications such as anomaly detection and fault diagnostics.

1. Introduction

The growing availability of interval energy use data and the rapid expansion of energy analytics offerings presents tremendous promise to both enable efficiency savings and automate savings quantification. Industry-wide, there is a desire to streamline the measurement & verification (M&V) process for energy-saving projects, in terms of time, costs, and complexity. As organizations work to meet aggressive building energy reduction goals at the national and state levels, there is increasing interest in moving toward performance-based outcomes - whether in codes, incentives, or operational energy goals. In association, there is a growing appetite for measured and verified, as opposed to deemed or calculated savings. At the same time, the

industry is beginning to ask for increased rigor, transparency, and consistency in how savings are determined.

Prior work [1,2] developed and tested promising "M&V 2.0" or "Advanced M&V" approaches to automate the savings estimation process through the analysis of time series meter data. Through statistical test procedures, these articles show that for a large fraction of buildings, automated techniques are accurate in characterizing and predicting building energy use for M&V applications. One critical question that has not yet been fully addressed is how to account for "non-routine events" (NREs). Non-routine events cause changes in energy use that are not attributable to installed efficiency measures and not accounted for in the baseline model's independent variables. They must be accounted for as 'adjustments' [3] to appropriately attribute metered change in consumption to the specific efficiency interventions that were implemented. Commonly encountered non-routine events are associated with fluctuation in occupancy levels, the addition of internal loads such as business-specific equipment, computers and servers, and changes to control operational strategies that are not part of the efficiency measure.

Industry methods to conduct non-routine events adjustments are currently manual and non-standardized; they comprise custom developed estimates based on a project engineer's or an evaluator's professional judgment and expertise, on a case by case basis. Current methods span a diversity of solutions that for detection rely upon site audits, surveys and interviews, and inspection of system-level data should it be available. For quantification, methods used in professional practice span: short-term site measurement, engineering calculations, short-term 'mini' regressions before and after the event, and simulation modeling. A robust statistical approach based on building consumption data has the potential to provide consistency as well as transparency, addressing one of the capability gaps [4] in today's technologies for meter-based automated savings estimation.

A non-routine event detection algorithm must be capable of detecting multiple changes within a given energy time series. Because non-routine events may be temporary, the start and end points of the event must be resolved. In addition, multiple events may occur across a period of interest. However, with high frequency metered energy data, the number of possible solutions to the multiple change point problem increases combinatorially. Hence, in addition to the accuracy in event detection, the algorithm needs to be computationally efficient for scalability across a large number of buildings. Furthermore, commercial building energy consumption time series usually feature non-stationarity (i.e., changes in the statistical properties over time), which make it challenging to distinguish between typical behaviors and NREs. In commercial buildings, the non-stationarity of the energy consumption time series can be due to many different factors, e.g., vacation periods, weather changes, occupancy changes, etc. Thus, unlike the traditional change point detection statistical problem [5], the NRE detection doesn't simply aim to detect changes within a given time series but rather to detect changes that are not attributable to a normal operation of the building. To address this issue the detection algorithm must be able to distinguish between these different behaviors of the energy consumption time series.

In the proposed methodology an approach based on a statistical change point algorithm is used to detect potential NREs. We specifically focus on NREs that occur after the implementation of an efficiency measure (as opposed to those that might occur during the baseline period); this provides a tractable starting point that can be expanded in future investigations. Change points are considered to be the points in the time series where a change in the statistical properties, such as mean and/or variance, is observed. Since NREs vary in nature, they are classified into three categories: (1) short-term NREs, that occur for a limited number of days (usually one or two days); (2) temporary NREs, which may span several weeks to several months; (3) permanent

NREs that result in a lasting change in the energy consumption. This work also introduces a data driven approach to quantify the changes in the energy consumption that are induced by these events, in order to adjust the savings estimate.

2. Non-Routine Event Detection and Adjustment Methodology

2.1 Non-Routine Event Detection

This section proposes an algorithm for detection of potential NRE during the post period (i.e., the period after the implementation of an efficiency measure), which involves four steps: (1) create a building energy use baseline model using pre period data (typically 1 year of data to cover all weather seasons). The baseline model is taken as reflective of the typical behavior of the building; (2) During the post period use the baseline model to generate prediction values representing the energy consumption if the retrofit had not been implemented; (3) For each day of the post period estimate the dissimilarity between the predicted time series and the actual consumption values; (4) Apply change points algorithm to the time series of dissimilarity values.

The proposed approach assumes that if there is no NRE the difference in the behavior of the pre and post energy consumption time series is stable. In addition, it is supposed that there are not significant periods of missing data in the post period (i.e., that there is continuity in the time series). The concept is that when an NRE occurs the difference in the behavior of the time series will change enough to be detected by a change point detection algorithm. This difference in the behavior can be estimated using a proximity measure, also known as "dissimilarity index," which characterizes how close two time-series are to each other in term of values and temporal correlation. The following subsections describe the dissimilarity measure, the change point algorithm and the baseline modeling used in this work. Then the algorithm of the proposed methodology of potential NRE detection is presented.

2.1.1 Dissimilarity measure

In the context of time series data analysis, the definition of dissimilarity (i.e., time series proximity measure) is not straightforward because of the interdependent relationship between observations. Several approaches to define dissimilarity measures of time series have been introduced in literature [6]. The most common approaches consist of evaluating the proximity between time series based on the closeness of the observed values at specific points of time (e.g., Euclidean distance, Minkowski distance and Fréchet distance). However, these measures ignore the temporal structure of the values since the closeness is computed using the differences between the observed values without considering the behavior around these values (i.e., increases or decrease in values). For NRE detection it is important to measure the closeness considering both behavior and value. Figure 1 illustrates time series dissimilarity in value and behavior. In the left figure the dissimilarity between the two time series is only in value, while in the right figure there is a difference in value and in behavior (e.g., from 6 pm to 8 pm one can note a significant decrease in the time series depicted by the solid line while the dashed line time series remains constant).

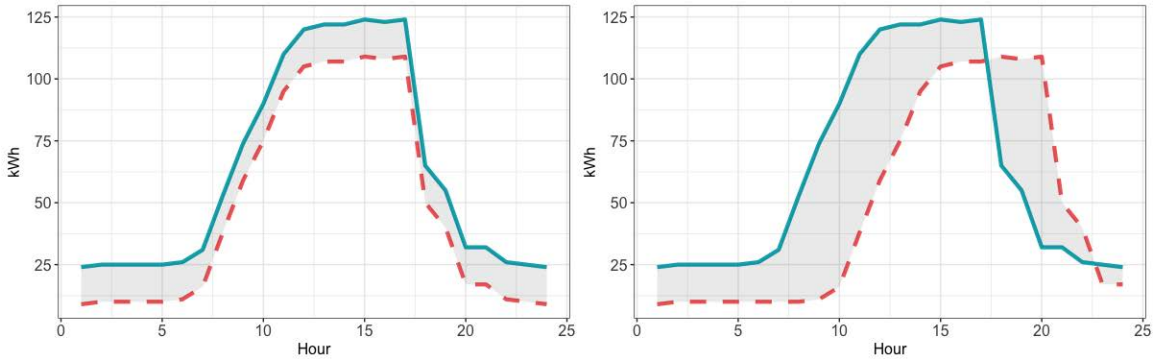


Figure 1. Example of dissimilarity between two time series. The left plot depicts dissimilarity in value, while the right plot depicts dissimilarity in value and behavior.

Chouakria and Nagabhushan [7] introduced a dissimilarity measure that covers both behavior and values. The proximity of the behavior of the time series is measured by computing the first order temporal correlation defined as:

$$CORT(S_1, S_2) = \frac{\sum_{t=1}^{T-1}(S_1(t+1)-S_1(t))(S_2(t+1)-S_2(t))}{\sqrt{\sum_{t=1}^{T-1}(S_1(t+1)-S_1(t))^2}\sqrt{\sum_{t=1}^{T-1}(S_2(t+1)-S_2(t))^2}} \qquad (1)$$

where S_1 and S_2 are two time series of T values.

The $CORT$ measure takes a value in the interval [-1,1]. If $CORT(S_1, S_2) = 1$ then the two time series have similar behavior, i.e., their increase/decrease at any time step is similar in direction and rate. If $CORT(S_1, S_2) = -1$ then the series have an opposite increase/decrease behavior but similar growth in rate. Finally, if $CORT(S_1, S_2) = 0$ then there is no monotonicity between S_1 and S_2 and their growth in rates are stochastically linearly independent. The dissimilarity metric introduced by [7] modulates between the proximity in value and the proximity in behavior. This is done using a modulating function (i.e., $\frac{2}{1+\exp(k\,x)}$) that will increase the value proximity when the temporal correlation (i.e., $CORT$) decreases from 0 to -1, and will decrease the value proximity in value when the temporal correlation increases from 0 to 1. Note that the proximity in value can be calculated using any of the traditional metrics mentioned previously. In this work, Euclidian distance is considered to evaluate the proximity in values. Thus, the dissimilarity proposed by [7] is defined as:

$$diss_{cort}(S_1, S_2) = \frac{2}{1+\exp(k\,CORT(S_1,S_2))}\,diss_E(S_1, S_2) \qquad (2)$$

where $diss_E(S_1, S_2)$ is the Euclidean distance and k is a tuning parameter that modulates the contribution between the proximity in values and the proximity in behavior. In order to have similar contribution of proximity in value and behavior $k=1$ is selected as default in this work. (For further detail see [7]). For example, the CORT dissimilarity (i.e., $diss_{cort}$) for the time series in Figure 1 are ~39.5 for the left scenario and ~180.4 for the right scenario. Note that the sums of the differences between the time series (i.e., savings) are exactly the same (i.e., equal to 360 kWh) for both scenarios.

2.1.2 Change point detection algorithm

Change points are considered to be the location in a time series where a change in the statistical properties of the sequence of observations is observed. Given a time series $S_{1:T} = (y_1, \dots, y_T)$, it is said that there is a change point in the time series $S_{1:T}$ when there exists a time step δ with $\delta \in (1, T-1)$ such that the statistical properties of $S_{1:\delta} = (y_1, \dots, y_\delta)$ and $S_{\delta+1:T} = (y_{\delta+1}, \dots, y_T)$ are different. If there are *n* change points with locations $\delta_{1:n} = (\delta_1, \dots, \delta_n)$ then the time series will be split into *n+1* distinct segments with the *i*th containing observations $S_{(\delta_{i-1}+1):\delta_i} = (y_{\delta_{i-1}}, \dots, y_{\delta_i})$. Therefore, the statistical problem is to estimate the number of change points and their locations. The most common approach is to introduce a cost function for each segment. The sum of costs across the segments can then be defined as the cost of the segmentation process. The change points can be identified by minimizing the segmentation cost. Since *n* the number of change points is unknown, the typical approach to jointly estimate the number of change points and their locations is to solve a penalized segmentation cost function defined as:

$$\sum_{i=1}^{n+1}\left[C\left(S_{(\delta_{i-1}+1):\delta_i}\right)\right] + \beta n \qquad (3)$$

Where $C\left(S_{(\delta_{i-1}+1):\delta_i}\right)$ denotes the segment specific (i.e., $(y_{\delta_{i-1}}, \dots, y_{\delta_i})$) cost function, and βn is a linear penalty term that prevents overfitting the data. The choice of the penalty constant (i.e., β) in (3), can have an important impact on the accuracy of the change points detection and therefore on the non-routine events detection. Several choices for selecting β have been introduced in the literature. Some of the most commonly used are the Akaike's information criteria (AIC), Schwartz's information criteria (SIC), Bayesian information criteria (BIC) and modified Bayesian information criteria (mBIC) [8], which accounts for the length of the segments. In this work mBIC is used for penalty constant selection; this choice was motivated by the fact that some test results showed that in comparison with using the other mentioned criteria, mBIC improved the accuracy of the proposed algorithm.

A standard approach to define the segment specific cost function is to specify a statistical model for the observed data points within a segment. It is therefore possible to define the cost function as proportional to minus the maximum log-likelihood. In this work the observations within a segment are assumed to be independent, identically distributed and generated from a Gaussian distribution with two segment specific parameters mean μ and variance σ^2. Thus, the log-likelihood of the observations $S_{(\delta_{i-1}+1):\delta_i}$ is defined as:

$$l\left(S_{(\delta_{i-1}+1):\delta_i}; \mu, \sigma^2\right) = -\frac{\delta_i - \delta_{i-1}}{2}\log(2\pi) - \frac{\delta_i - \delta_{i-1}}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=\delta_{i-1}+1}^{\delta_i}\left(y_j - \mu\right) \qquad (4)$$

In this work we operate under the assumption that changes may occur in both the mean and the variance. Therefore, computing the segment specific cost function involves using minus twice the log-likelihood function after maximizing it over μ and σ^2. Then the segment cost function is defined as:

$$C\left(S_{(\delta_{i-1}+1):\delta_i}\right) = (\delta_i - \delta_{i-1})\left\{log\left[\frac{1}{\delta_i - \delta_{i-1}}\sum_{j=\delta_{i-1}+1}^{\delta_i}\left(y_j - \bar{y}_i\right)^2\right] + 1\right\} \qquad (5)$$

Where $\bar{y}_i = \frac{1}{\delta_i - \delta_{i-1}}\sum_{k=\delta_{i-1}+1}^{\delta_i} y_k$ is the mean value of the *i*th segment (i.e., $S_{(\delta_{i-1}+1):\delta_i}$).

Commonly used approaches to find the optimal solution for the optimization problem (3) include binary segmentation (BS) [9], the segment neighborhood (SN) [10] and optimal partitioning (OP)

[11]. The BS is computationally fast but at the expense of accuracy, since it provides an approximate minimization of (3). The SN and OP uses dynamic programming to exactly minimize the objective function (3), however the computational cost of this method is too high to scale when a large volume of time series data is analyzed. To avoid this tradeoff between accuracy and speed the pruned exact linear time (PELT) algorithm was introduced [12], which is similar to SN and OP algorithm in that it also computes an exact minimization of the objective function (3). The PELT algorithm explores the data sequentially and exhaustively through the solution space. The computational efficiency of the PELT algorithm is achieved by removing from the minimization, at each iterative step, the candidate location values that are known not to lead to the optimal solutions (see section 3 of [12] for more details). In this article, PELT is used as the change points method for detecting potential NRE locations.

2.1.3 Baseline model

Regression methods are a standard approach used for developing baseline models that aim to model the relationship between the response *y*, which is the pre whole-building energy use and a set of independent variables (also known as explanatory variables) $\boldsymbol{x} = (x^{(1)}, \dots, x^{(d)})$, where *d* is the number of independent variables. For example, in the case of energy use baseline modeling, the input variables can be time of the week and the outdoor air temperature. Mathematically the regression problem can be represented for a given observation set $\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_T, y_T)\}$, as

$$y_t = f(\boldsymbol{x}_t) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \qquad (6)$$

where $\boldsymbol{x}_t = (x^{(1)}, \dots, x^{(d)})$, $t = 1, \dots, T$ are d dimensional vectors of inputs variables, ε_t is independent Gaussian noise with mean 0 and variance σ_ε^2. Building a baseline model consists of approximating the function $f(\boldsymbol{x})$ given a set of *T* observation $\{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_T, y_T)\}$.

In recent years several baseline energy modeling approaches that use interval meter data have been introduced in the academic literature and in the industry. These methods are based on traditional linear regression, nonlinear regression, and machine learning regression methods. In this study a gradient boosting machine (GBM) baseline model is used [13]. The GBM baseline model is an ensemble trees based machine learning method that generates a model of the energy consumption using the following independent variables: time of week, temperature and when the considered building is a school, a vacation indicator. A holiday indicator is not used as an independent variable in the analysis performed in this work in order to use some of holidays as a test case for short-term NRE detection.

2.1.4 Algorithms for non-routine event detection

In this section, three algorithms for detecting potential NREs are described. The first one is the main contribution of this work (using the CORT dissimilarity metric), while the second and third are used as comparison cases. The second algorithm follows the same steps as algorithm 1 but rather than using the CORT dissimilarity the Euclidian dissimilarity was applied. The third algorithm is a naïve comparison case where the change point detection algorithm (i.e., PELT) is applied directly to the post period data.

Algorithm 1
Input: Post period time series and baseline model predictions for the post period

1) Extract the energy consumption time series for each day of the post period. This produces two $d \times T_{day}$ matrices, M_{actual} and $M_{prediction}$, where d is the number of days that are found in the post period and T_{day} is the number of observations in each day (e.g., with hourly data $T_{day} = 24$). This matrix representation of the post period time series is used to compute the dissimilarity metric.
2) For each row of M_{actual} and $M_{prediction}$ compute the daily CORT dissimilarity metric $diss_{cort}\left(M_{actual}[i, 1{:}T_{day}], M_{prediction}[i, 1{:}T_{day}]\right)$, with $i = 1, \dots, d$. The results will be a time series D of length d where each element represent the dissimilarity between the actual and the prediction energy consumption time series for each day.
3) Apply the change point PELT algorithm (described in section 2.1.2) to the D time series.

Output: $\delta_{1:n} = (\delta_1, \dots, \delta_n)$, a vector with locations of *n* change points (i.e., potential NREs)

Algorithm 2
Input: Post period time series and baseline model predictions for the post period.

1) Extract the energy consumption time series for each day of the post period. This produces two $d \times T_{day}$ matrices, M_{actual} and $M_{prediction}$, where d is the number of days that are found in the post period and T_{day} is the number of observations in each day (e.g., with hourly data $T_{day} = 24$). This matrix representation of the post period time series is used to compute the dissimilarity metric.
2) For each row of M_{actual} and $M_{prediction}$ compute the daily Euclidian dissimilarity metric $diss_E\left(M_{actual}[i, 1{:}T_{day}], M_{prediction}[i, 1{:}T_{day}]\right)$, with $i = 1, \dots, d$. The results will be a time series D of length d where each element represent the Euclidian dissimilarity between the actual and the prediction energy consumption time series for each day.
3) Apply the change point PELT algorithm (defined in section 2.1.2) to the D time series.

Output: $\delta_{1:n} = (\delta_1, \dots, \delta_n)$, a vector with locations of *n* change points (i.e., potential NRE)

Algorithm 3
Input: Post period time series

1) Extract the energy consumption time series for each day of the actual post period. This produces a $d \times T_{day}$ matrix M_{actual}, where d is the number of days that are found in the post-period and T_{day} is the number of observations in each day (e.g., with hourly data $T_{day} = 24$).
2) For each row of M_{actual} aggregate the energy consumption. The results will be a time series D of length d where each element represent the daily energy consumption.
3) Apply the change point PELT algorithm (defined in section 2.1.2) to the D time series.

Output: $\delta_{1:n} = (\delta_1, \dots, \delta_n)$, vector with locations of *n* change points (i.e., potential NRE)

Note that the changepoint R package [16] was used for computing the PELT algorithm and the dissimilarities were computed using the Tsclust R package [17].

2.2 Non-Routine Event Adjustment

The proposed methodology for NRE adjustment is based on a statistical approach that involves the NRE detection algorithm, the M&V practitioner's knowledge, and a regression model (i.e., GBM).

The NRE Adjustment methodology follows six steps, some of which can be automated, and some of which must be performed by the M&V practitioner: (1) estimate the locations of potential NREs using the NRE detection algorithm; (2) verify which detected NREs merit an adjustment to the saving calculation using project information or site inquiry, and validate their actual date of occurrence; (3) exclude from the post period time series the data points from the verified NRE periods; (4) train the regression model using the remaining data of the post period time series; (5) use the trained model to predict what the energy consumption would have been within the NRE periods if no NRE was present; (6) use the model predicted values as replacement data for the periods in which the NRE was present. The final step of the methodology produces the adjusted post period time series. Note that in real-world efficiency projects, the question of whether detected NREs merit an adjustment to the savings calculation may be quite nuanced, as further elaborated in the Discussion in Section 5.

3. Methods to assess effectiveness of NRE detection and adjustment algorithms

The proposed methodology for NRE detection and adjustment was tested using simulated hourly energy consumption data generated using EnergyPlus™ models for two types of DOE reference buildings [14]: primary school (PS) and large office (LO). The advantage of using simulation data is that it is possible to generate different testing configurations, i.e., a post period without an energy efficiency measure (and therefore a base case with no savings), a post period with an energy efficiency measure and no NREs, and a post period with an energy efficiency measure and with an NRE. Thus, it is possible to evaluate the accuracy of the detected NRE and the accuracy of the adjusted savings. A dataset comprising eight different combinations were generated (a summary of these eight scenarios is provided in **Table 1**. For each type of building four different TMY weather data were used that correspond to the following cities: Miami, Chicago, San Francisco and Phoenix. Note that in order to have different weather data between the pre and post periods, TMY2 weather data were used for pre period and TMY3 for post period. Depending on the scenario (see Table 1) one of two different type of retrofit was modeled: reducing lighting power density and increasing cooling equipment efficiency. In addition, different NRE were introduced at different time steps in the post period. For the LO buildings four NREs types were applied (one for each scenario that involves LO): (1) change in occupancy start and end time; (2) change in occupancy density; (3) change in the electrical baseload; (4) cooling failure. For the PS three NREs were considered: (1) school days scheduled during the summer vacation; (2) change in the electrical baseload; (3) fault in cooling (cooling production is stopped). In addition to the described NREs, two auxiliary short-term NREs were added in the post period in each scenario. These two NREs correspond to two US federal holidays that occur in January 16th and February 20th (i.e., Martin Luther King Jr. Holiday and Presidents Day). In this study holidays are used as convenient *proxies* for short term NREs - these two days were considered as normal operational days when the baseline models were trained (i.e., there is no independent variable in the input data that state that these two days are holidays). Note that in the EnergyPlus simulation models only these two dates are defined as holidays (i.e., days where the building is considered as unoccupied).

Table 1- Scenarios simulated to test NRE detection and adjustment

Building name	city	Retrofit	NRE
PS_1	Chicago	Reducing Lights Power Density in Classrooms only from 21.52W/m² (2W/ft²) to 13W/m² (1.2W/ft²)	The nominal value of the baseload for electrical plug was multiplied by ~1.83 between March 15th and April 8th (temporary NRE)
PS_2	Miami	Increasing Cooling Equipment Efficiency: Chiller COP from 3.2 to 4.8	In session schedules applied during Summer Vacation: July 13th to August 7th (temporary NRE)
PS_3	San Francisco	Increasing Cooling Equipment Efficiency: Chiller COP from 3.2 to 4.8	School days schedules applied during Summer Vacation: July 1th to August 11th (temporary NRE)
PS_4	Phoenix	Reducing Lights Power Density in Classrooms only from 21.52W/m² (2W/ft²) to 13W/m² (1.2W/ft²)	Cooling failure from June 12th to July 6th (temporary NRE)
LO_1	Miami	Increasing Cooling Equipment Efficiency: Chiller COP from 5.2 to 7.8	Change in occupancy period start and end time: from (8am-5pm) to (6am-6pm) between June 26th and July 31st (temporary NRE)
LO_2	Chicago	Reducing Lights Power Density in Classrooms only from 16.89W/m² (1.6W/ft²) to 8W/m² (0.74W/ft²)	Cooling failure from June 5th through August 9th (temporary NRE)
LO_3	San Francisco	Reducing Lights Power Density in Classrooms only from 16.89W/m² (1.6W/ft²) to 8W/m² (0.74W/ft²)	Two changes in occupant density. First between May 3rd and August 29th (temporary NRE) the occupant density is multiplied by a factor of ~0.57. Then starting from August 30th (permanent NRE) the occupant density is again multiplied by a factor of ~0.64
LO_4	Phoenix	Increasing Cooling Equipment Efficiency: Chiller COP from 5.2 to 7.8	The nominal value of the baseload for electrical plug was multiplied by ~1.71 starting from June 11th (permanent NRE)

Accuracy metrics

The evaluation procedure was performed by applying the proposed algorithms on each combination of building, type of retrofit, and NRE. The accuracy of the NRE detection was assessed by calculating the percentage of identified change points that were true NREs (true positive, or "TP"), and what percentage were not true NREs (false positive, or "FP"). These two metrics are defined as:

$$TP = 100 \times \frac{\sum_{i=1}^{n_{NRE}} I\left(\min_j\{|NRE_i - \delta_j|\} \leq th\right)}{n_{NRE}} \quad (7)$$

$$FP = 100 \times \left[1 - \frac{n_{NRE} \times TP}{n}\right] \quad (8)$$

Where NRE_i is the actual location of the ith NRE, δ_j is the location of the jth detected potential NRE using the proposed algorithms, n_{NRE} is the actual number of locations (i.e., dates) where there is an NRE, n is the number of potential NRE, and th is the threshold that defines an actual NRE as detected by the algorithm if the estimated position is within $\pm th$ number of days. In other words, if the closest detected potential NRE date to the ith actual NRE date is within a range of plus or minus th days, this NRE is considered as successfully detected. In this analysis $th = 2$ was arbitrarily selected. Finally, $I(x)$ is the indicator function defined as

$$I(x) = \begin{cases} 1 & if\ x\ is\ true \\ 0 & else \end{cases}$$

The accuracy of the NRE adjustment were evaluated using the absolute error (AE) of the estimated fractional savings, defined as

$$AE = \left|FS - \widehat{FS}\right| \quad (9)$$

where FS is the actual fractional savings defined as $FS = \frac{E_{no\ retrofit}^{post} - E_{retrofit}^{post}}{E_{no\ retrofit}^{post}}$, with the post period aggregated energy consumption $E_{no\ retrofit}^{post}$ and $E_{retrofit}^{post}$ are respectively corresponding to when no retrofit was applied and when a retrofit was applied. $\widehat{FS}$ is the estimated fractional savings defined as $\widehat{FS} = \frac{\hat{E}_{baseline}^{post} - E_{retrofit}^{post}}{\hat{E}_{baseline}^{post}}$, with $\hat{E}_{baseline}^{post}$ is the post period aggregated energy consumption estimated using GBM baseline model. Note that $\widehat{FS}$ become adjusted fractional savings ($\widehat{FS}_{adjusted}$) if $E_{retrofit}^{post}$ is adjusted using the proposed methodology.

4. Results

4.1 Results of NRE detection algorithm

The accuracy of the proposed algorithm for detecting NRE (i.e., algorithm 1 involving CORT dissimilarity) is compared to algorithm 2 (i.e., involving the Euclidian dissimilarity) and algorithm 3. The metrics of the NRE detection accuracy (i.e., TP and FP) were computed across the full test dataset (i.e., eight scenarios) and are summarized in Table 2. In addition, Figures 2.1-2.8 display the time series of the actual post period energy consumption time series at daily granularity and the corresponding dissimilarity time series (computed as described in algorithm 1). In these plots the normal operation periods are depicted in blue, the NRE period in red (red points for short-term NREs) and in green (if two successive events are involved) and the NRE detected by Algorithm 1 are represented by vertical black lines. Note that the choice of representing the energy consumption time series at daily granularity was motivated by the fact that the NRE detection was performed at the daily level. The reader is reminded that determination of ‘actual’, ‘present’ and ‘absent’ is based on simulated scenarios.

The results in Table 2 shows that algorithm 3 performed poorly in detection of actual NRE in majority of cases, except for PS_1 however for this configuration it had a very significant number of false positives (i.e., 28). Algorithm 2 performed significantly better than algorithm 3 for all the scenarios. In comparison to algorithm 1, algorithm 2 underperformed in 6 scenarios (i.e., PS_1, PS_3, PS_4, LO_1, LO_2 and LO_3), had similar results for LO_3 and had better results in term of detected FP for PS_2. Algorithm 1 performed quite well in term of detecting the actual NRE. For 5 out of 8 buildings configurations Algorithm 1 has detected all the actual NRE. Meanwhile for 3 out of 8 it has detected all except one, however for these 3 building combinations Figure 2.1, 2.6 and 2.7 show that the algorithm has detected a change point relatively close to the NRE but not close enough to be defined as actual NRE, which is due to the restrictive threshold that we have arbitrarily defined in equation 7. Although these results are quite good it is important to note the relatively high number of false positive detected by the algorithm. With a high number of false positives the role of the M&V practitioner in verifying and validating the algorithm outputs

becomes increasingly important (although this 'human in the loop' step must be included regardless of algorithm performance, to ensure rigor, particularly for utility program M&V). Note that to assess some of the impact on the results of using a different value of the threshold *th* in in equation 7, additional results are shown in the appendix for *th = 1* and *th =3.*

Figures 2.1-2.8 show that the false positive are due to a significant variation in the dissimilarity time series which may be due to several reasons, e.g., the interim effect of the implemented retrofit, which generate a change in the energy consumption only during a specific period of the year (i.e., increase in cooling efficiency); a period where the baseline model has lower accuracy, this can increase the dissimilarity metric and such possibly make the change point algorithm detect a change in the statistical properties.

Table 2- Accuracy metrics of the NRE detection for the eight tested scenarios. The numbers of true positives and false positives are shown in parentheses

Building			Number of actual NRE	Algorithm 1			Algorithm 2			Algorithm 3		
Name	Type	Location		Number of detected NRE	True Positive	False Positive	Number of detected NRE	True Positive	False Positive	Number of detected NRE	True Positive	False Positive
PS_1	Primary Schools	Chicago	4	8	75% (3)	63% (5)	7	50% (2)	72% (5)	32	100% (4)	88% (28)
PS_2		Miami	4	7	100% (4)	43% (3)	5	100% (4)	20% (1)	1	0% (0)	100% (1)
PS_3		San Francisco	4	6	100% (4)	33% (2)	8	75% (3)	65% (5)	31	50% (2)	94% (29)
PS_4		Phoenix	4	5	100% (4)	20% (1)	7	75% (3)	57% (4)	9	0% (0)	100% (9)
LO_1	Large Offices	Miami	4	6	100% (4)	33% (2)	8	75% (3)	63% (5)	2	0% (0)	100% (2)
LO_2		Chicago	4	5	75% (3)	40% (2)	2	25% (1)	50% (1)	5	0% (0)	100% (5)
LO_3		San Francisco	4	5	75% (3)	40% (2)	6	50% (2)	67% (4)	1	25% (1)	0% (0)
LO_4		Phoenix	3	9	100% (3)	67% (6)	9	100% (3)	67% (6)	7	33% (1)	86% (6)

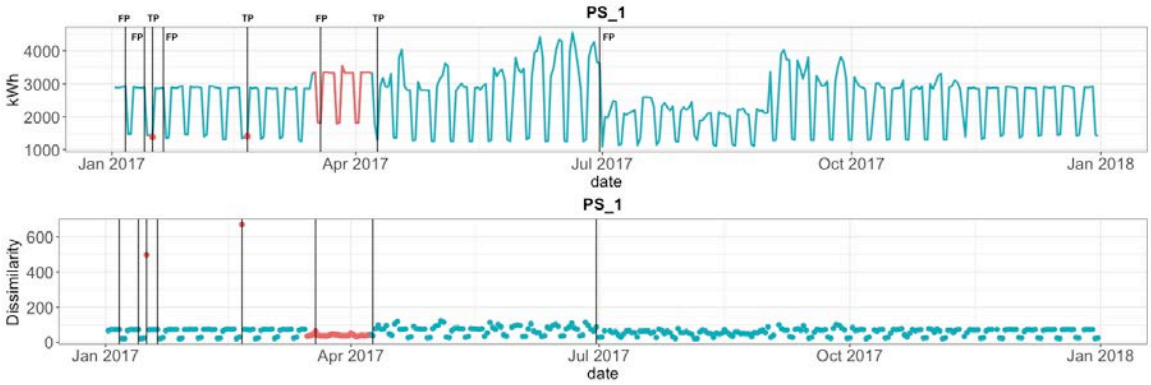


Fig 2.1 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a primary school located in Chicago, with a reduction in lighting power density as the efficiency measure, and a change in the electrical baseload as NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

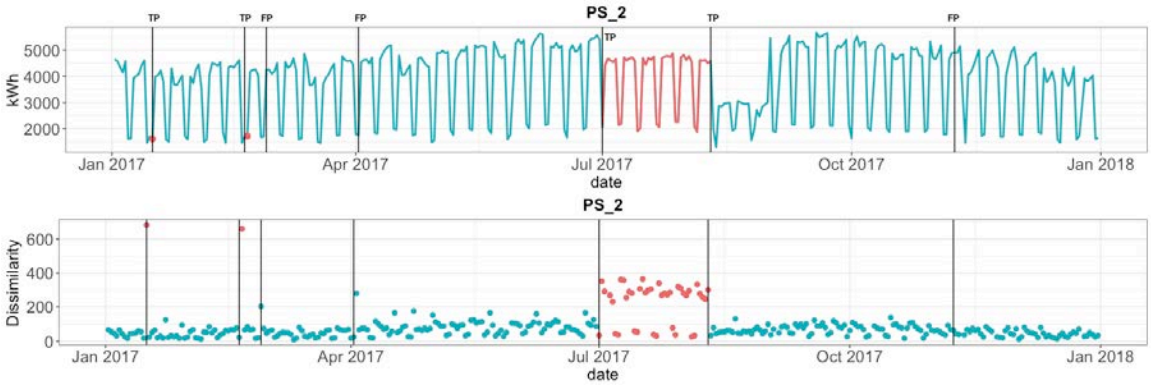


Fig 2.2 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a primary school located in Miami, with an increase in the cooling equipment efficiency as the efficiency measure, and a change in the occupancy density as the NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

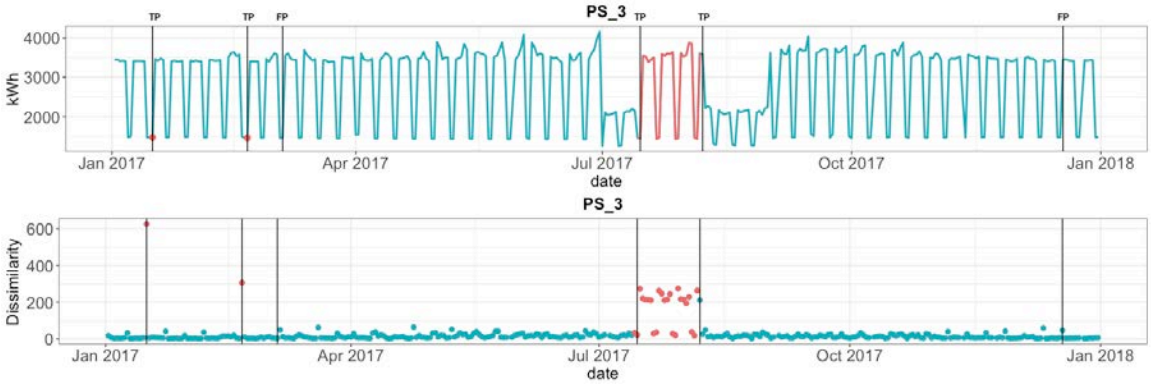


Fig 2.3 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a primary school located in San Francisco, with an increase in the cooling equipment efficiency as the efficiency measure, and a change in the occupancy density as the NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

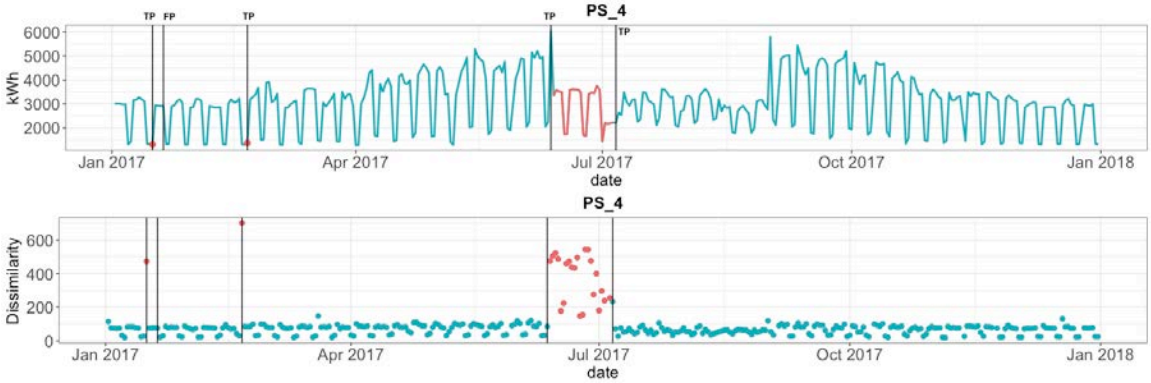


Fig 2.4 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a primary school located in Phoenix, with a reduction in lighting power density as the efficiency measure, and a cooling failure as the NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

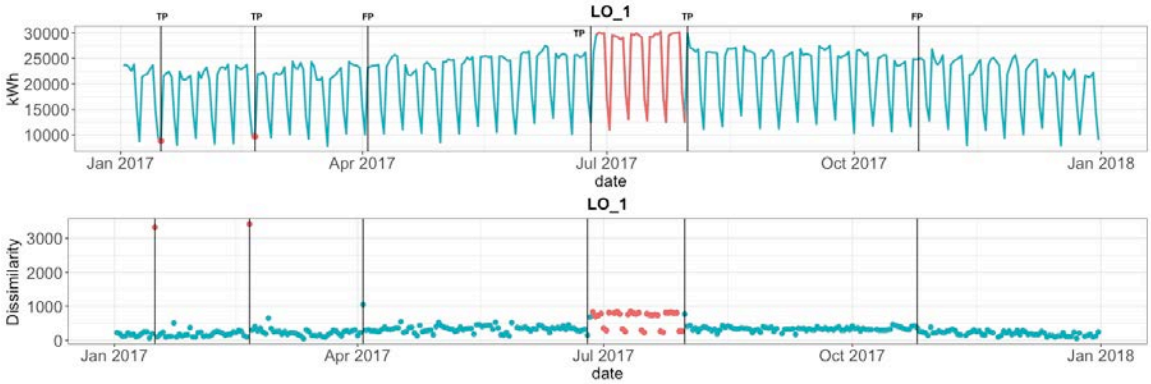


Fig 2.5 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a large office located in Miami, with an increase in the cooling equipment efficiency as the efficiency measure, and a change in the occupancy start and end time as the NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

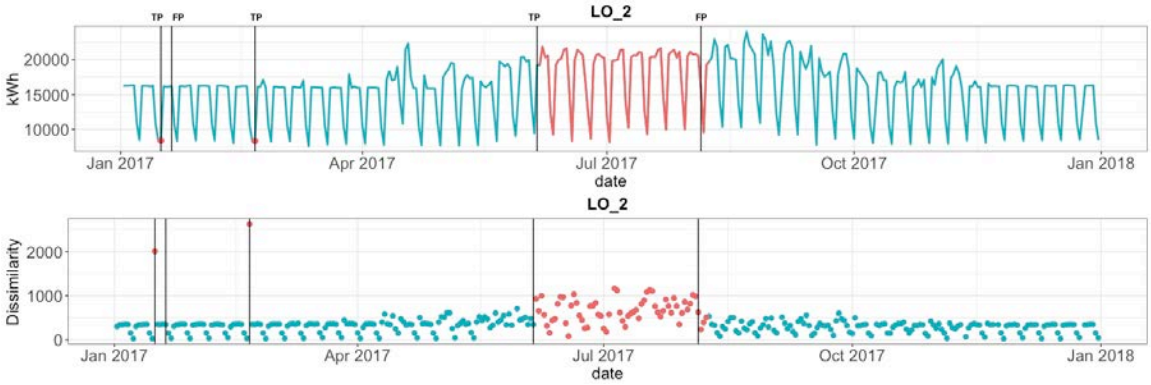


Fig 2.6 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a large office located in Chicago, with a reduction in lighting power density as the efficiency measure, and a cooling failure as the NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

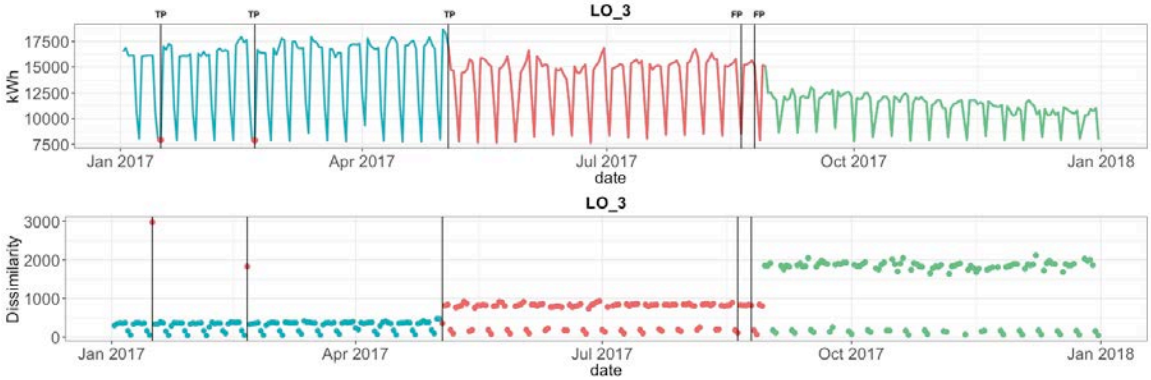


Fig 2.7 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a large office located in San Francisco, with a reduction in lighting power density as the efficiency measure, and changes in the occupancy density as the NRE. Blue depicts periods of time in which NREs were absent; red and green indicate periods of time in which NREs were present.

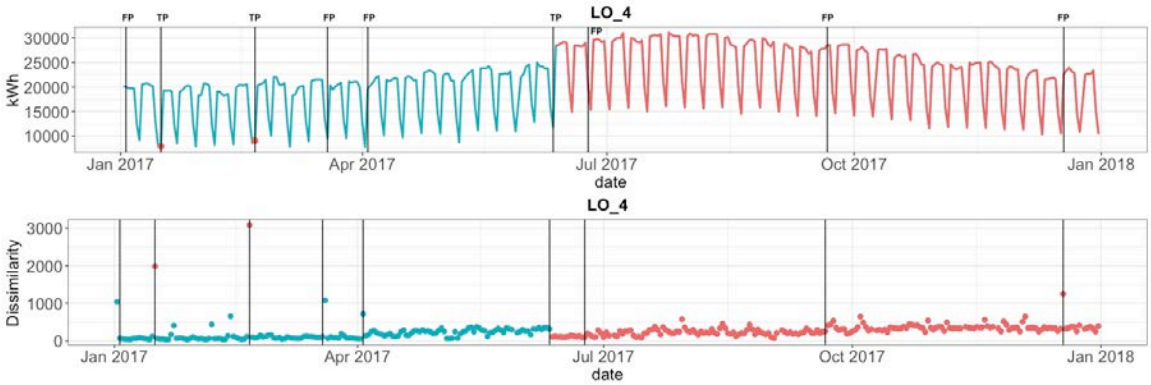


Fig 2.8 Time series of actual energy consumption during the post-period (upper graphic) and time series of the dissimilarity metric (lower graphic) of the scenario for a large office located in Phoenix, with an increase in the cooling equipment efficiency as the efficiency measure, and a change in the electrical baseload as the NRE. Blue depicts periods of time in which NREs were absent; red indicates periods of time in which NREs were present.

4.2 Results of NRE adjustment methodology

Table 3 summarize the results of the accuracy assessment of the proposed approach to adjust the NRE in the post period. The second column (i.e., Actual Fractional Savings) of the table shows the true fractional savings of the retrofit in the absence of any NREs (computed using simulated post-period time series data with no NRE and no retrofit, and subtracting consumption from a time series where the retrofit was implemented but not the NRE). The third column (Non-adjusted) shows the estimated fractional savings and the corresponding absolute error (defined by equation 9). These estimated fractional savings (i.e., $\widehat{\boldsymbol{FS}}$) were computed using the predictions time series provided by the baseline model and the time series where the retrofit and the NRE were implemented (in this scenario the NRE is treated as part of the overall energy savings). In the fourth column (i.e., Adjusted) the adjusted fractional savings (i.e., $\widehat{\boldsymbol{FS}}_{\boldsymbol{adjusted}}$) were calculated using the method described in Section 3.2.

Recall that for the estimated fractional savings the $\widehat{\boldsymbol{FS}}_{\boldsymbol{adjusted}}$ values that are the closest to the actual FS are desired (i.e., lower AE value). From Table 3 one can see that the adjusted estimates of the fractional savings are very close to the actual values, which means that for this test dataset the proposed NRE adjustment methodology successfully reduced the error introduced by the NRE. For all the building configurations the AE of the adjusted FS estimates were smaller or equal than 0.7%, while for the non-adjusted estimates the AE span from 1% to 11.9%. Figures 3.1-3.8 are the graphical representation of the adjustment of the NRE periods during the post period. In each plot a zoom around the NRE period is depicted, i.e., the left graphics show the two NRE that correspond to the two US federal holidays that occurs in January 16th and February 20th (i.e., Birthday of Martin Luther King Jr. and Washington's birthday). The right graphics show a zoom around the NRE periods that correspond to the NRE introduced in the EnergyPlus model and which are described in Table 1. In red are represented the values of the energy consumption before the NRE adjustment and in blue the $\widehat{\boldsymbol{FS}}_{\boldsymbol{adjusted}}$ values generated for the NRE periods. It is important to note that the two holidays were adjusted in this analysis because they are considered as *proxies* for short-term NREs. This is not to suggest that holidays, as a general rule, merit savings adjustments. An appropriate strategy to address holidays is to explicitly include them as independent variables in the baseline model.

Table 3- Accuracy metrics of the NRE adjustment for the eight tested building configurations

Building			Actual Fractional Savings	Non-adjusted		Adjusted	
Name	**Type**	**Location**		**Estimated Fractional Savings**	**Absolute Error**	**Estimated Fractional Savings**	**Absolute Error**
PS_1	Primary School	Chicago	15.3%	14.3%	1%	15.2%	0.1%
PS_2		Miami	8.6%	4.9%	3.7%	8.3%	0.6%
PS_3		San Francisco	1.6%	-1.1%	2.7%	1.1%	0.5%
PS_4		Phoenix	13.7%	16.6%	2.9%	13.1%	0.6%
LO_1	Large Office	Miami	7.8%	6.5%	1.3%	7.3%	0.5%
LO_2		Chicago	10.5%	11.9%	1.4%	10.6%	0.1%
LO_3		San Francisco	10.8%	22.7%	11.9%	10.2%	0.6%
LO_4		Phoenix	5.1%	-3.8%	8.9%	5.8%	0.7%

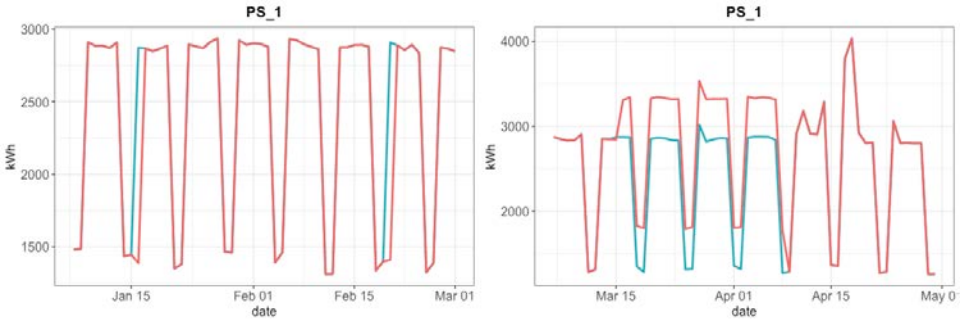


Fig 3.1 Plots of post period data of the scenario for a primary school located in Chicago, with a reduction on lighting power density as the retrofit, and a change in the electrical baseload as NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

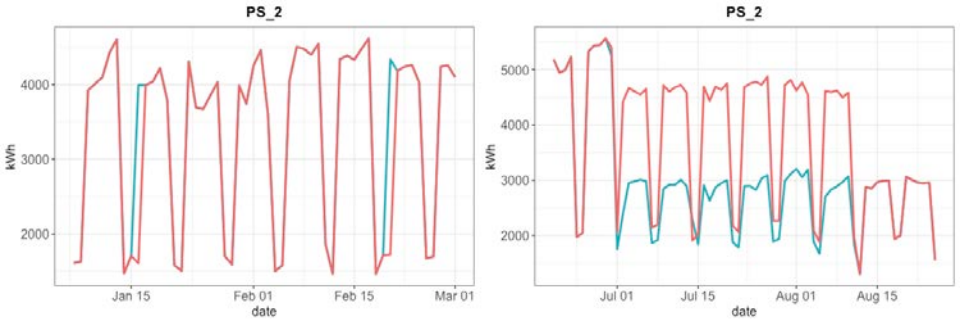


Fig 3.2 Plots of post period data of the scenario for a primary school located in Miami, with an increase in the cooling equipment efficiency as the retrofit and a change in the occupancy density as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

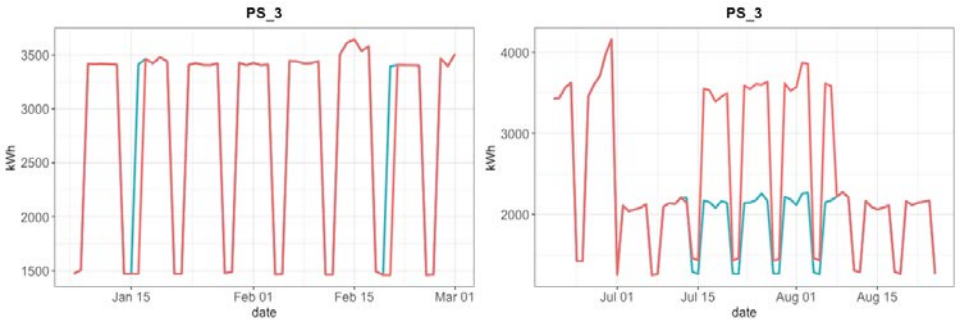


Fig 3.3 Plots of post period data of the scenario for a primary school located in San Francisco, with an increase in the cooling equipment efficiency as the retrofit and a change in the occupancy density as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

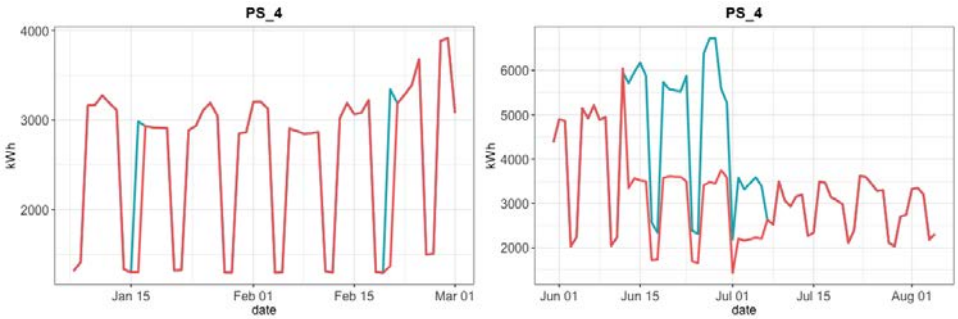


Fig 3.4 Plots of post period data of the scenario for a primary school located in Phoenix, with a reduction on lighting power density as the retrofit and a cooling failure as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

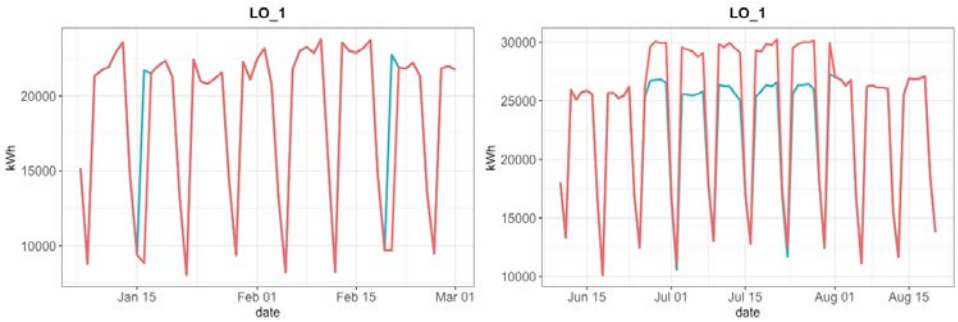


Fig 3.5 Plots of post period data of the scenario for a large office located in Miami, with an increase in the cooling equipment efficiency as the retrofit and a change in the occupancy start and end time as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

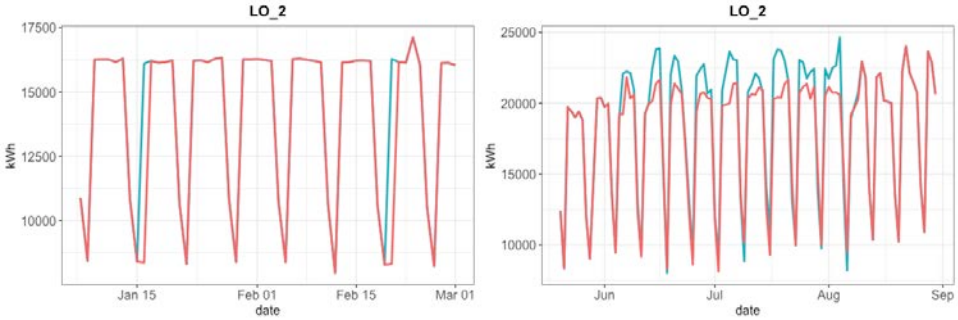


Fig 3.6 Plots of post period data of the scenario for a large office located in Chicago, with a reduction on lighting power density as the retrofit and a cooling failure as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

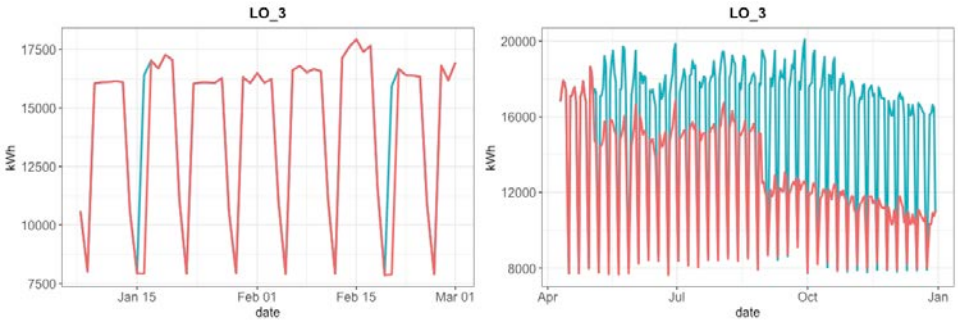


Fig 3.7 Plots of post period data of the scenario for a large office located in San Francisco, with a reduction on lighting power density as the retrofit and changes in the occupancy density as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

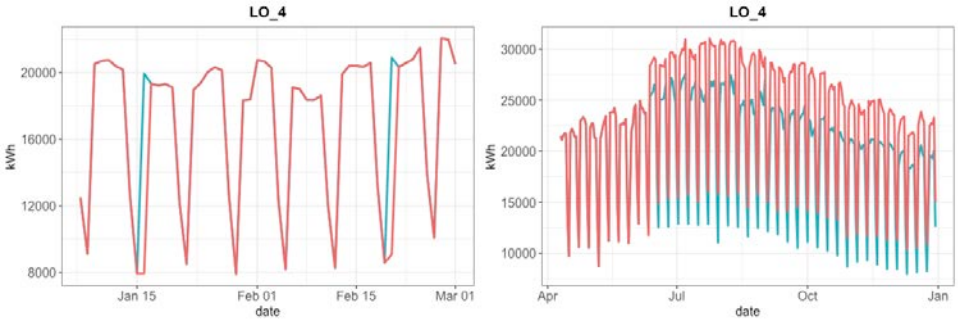


Fig 3.8 Plots of post period data of the scenario for a large office located in Phoenix, with an increase in the cooling equipment efficiency as the retrofit and a change in the electrical baseload as the NRE. Left graphic shows two holidays in January/February, and right graphic shows subsequent NRE periods in each data set. Actual consumption data is shown in red, and algorithm-adjusted values are shown in blue.

5. Discussion

The results showed that the performance of the *detection* portion of the algorithm was quite strong, both in an absolute sense, and versus the two other tested algorithms. The true positive rate ranged from 75%-100%. Given that the threshold used in the true positive metric was set relatively conservative at +/- 2 days, it is possible that the true positive rate could be further increased, however this would come with a tradeoff for what is practical for real-world implementation. Given the relatively limited number of test cases, it is difficult to draw conclusions as to the nature and type of events that were not detected by the algorithm, however this could be explored in future work.

The false positive NRE detection rate was lower than that of the benchmarks in the majority of test cases, however, with a range of 20-70% is likely higher than ideal. The improved performance versus the benchmarks is due to the use of the CORT dissimilarity metric.

The relatively high true positive rate and higher than ideal false positive rate might suggest that there may be room to 'tune' the algorithm so that a high true positive rate is maintained while also improving the false positive rate. Significant effort was indeed spent tuning, and the results shown are the best that could be achieved in the scope of this effort. The challenge associated with false positives is rooted in the inherent variability in building energy consumption – it is

very difficult to differentiate between a non-routine change and the normal behavior of the building.

Savings M&V for incentive-based utility efficiency programs present a high bar for rigor as compared to other M&V applications; in that context it might be acceptable to trade a relatively higher false positive rate for a higher true positive rate. This consideration speaks to the concept of streamlining versus full automation. It is important to emphasize that the goal of advanced M&V or M&V 2.0 is not full automation, but rather to leverage higher resolution data and automated computation to arrive at more timely, accurate results [15]. Thus, current research focuses on analysis methods paired with process guidance for professional practitioners, maintaining a 'human in the loop'.

The design for human in the loop workflows relates to interpretation of the results for the adjustment portion of the algorithm. Once a potential event is detected, the practitioner must verify whether the event was or was not non-routine, using either their knowledge of the project or building, confirmation with site points of contact, or other methods. This verification may also support the practitioner to isolate the true (as opposed to algorithm estimated) dates of occurrence of the event. The adjustment results that were presented showed 0.7% error or less in the final energy savings estimates, versus ground truth, once adjustments were made to account for the non-routine events. This represents a best-case scenario for which the event dates are known with certainty. In real-world application, there will likely be some uncertainty in isolating precise dates of event occurrence. In addition, it likely that adjustments for events of longer duration will incur more error. This is because the statistical models used to make the adjustment must be constructed from a more constrained set of data, while also predicting longer periods of consumption.

More generally, in real-world efficiency projects, the practitioner must confront the question of whether a detected NRE merits an adjustment to the savings calculation. Two examples presented in this work illustrate some of the nuance involved. A short-duration building shut-down (e.g., holiday) that was not expected given the independent variables used to train the model, but that nonetheless represents standard operation of the building would not be a candidates for adjustment, as if it did not occur. Similarly, a school that is assumed to shut-down during summer, yet regularly operates for a portion of the summer is not operating 'non-routinely' and would not require an adjustment. While these examples relate to assessment of anomalous behavior versus behavior that fits an expected or modeled pattern, the industry is also grappling with questions surrounding materiality. That is, what is the threshold at which an adjustment should be made, or is likely to improve the savings estimate, given that the adjustments calculation itself will also carry some uncertainty.

Independent of these nuances in savings estimation, change detection has additional relevance to ensuring persistence in efficient operations. The methods used in this work are equally applicable to, and can be extended to automated energy and equipment fault detection, where it is also necessary to identify deviation from the norm.

6. Conclusions and Future Work

A methodology for NRE detection and adjustment based on data driven methods was introduced and applied on a simulated dataset generated with EnergyPlus. The results of this study show that

the proposed approach holds promise for helping M&V practitioners to streamline the process of detecting NREs and adjusting the corresponding energy consumption to improve the accuracy of the meter-based savings estimates. There are many opportunities for future research, beginning with testing and validation against a larger set of simulated scenarios, and testing against data from real buildings and projects. Further work is needed to explore the impact on algorithm performance of parameters such as NRE duration, magnitude and type, and building type. Use by practitioners will provide valuable insights as to the practical viability of the approach, and suitability of the current balance between true and false positive rates.

Beyond expanded testing there is opportunity to augment the methodology to estimate the uncertainty on the adjustment, and to improve it for use with long-duration NREs. At a more general level, there is value in additional research to characterize how frequently NREs arise in efficiency projects of various types, the magnitude of their impact on savings results.

Finally, we recognize that meter-based savings analyses are often delivered as one of many capabilities in commercial energy analytics tools that are used to identify operational efficiency measures. As such, there is also clear value in future work to explore how the change detection techniques developed in this work would also be applied to energy anomaly and equipment faults detection.

Acknowledgement

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] Granderson, J., S. Touzani, C. Custodio, M. D. Sohn, D. Jump, and S. Fernandes. 2016. "Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings." *Applied Energy*, *173*, pp.296–308.

[2] Granderson, J., Touzani, S., Fernandes, S. and Taylor, C., 2017. "Application of automated measurement and verification to utility energy efficiency program data". *Energy and Buildings*, *142*, pp.191-199.

[3] Efficiency Valuation Organization (EVO). International Performance Measurement and Verification Protocol: Concepts and options for determining energy and water savings, Volume I. January 2012. EVO 10000-1:2012.

[4] Granderson, J. and Fernandes, S., 2017. The state of advanced measurement and verification technology and industry application. *The Electricity Journal*, *30*(8), pp.8-16.

[5] Truong, C., Oudre, L. and Vayatis, N., 2018. A review of change point detection methods. *arXiv preprint arXiv:1801.00718*.

[6] Montero, P. and Vilar, J.A., 2014. TSclust: An R package for time series clustering. *Journal of Statistical Software*, *62*(1), pp.1-43.

[7] Chouakria, A.D. and Nagabhushan, P.N., 2007. Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, *1*(1), pp.5-21.

[8] Zhang, N.R. and Siegmund, D.O., 2007. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, *63*(1), pp.22-32.

[9] Edwards, A.W. and Cavalli-Sforza, L.L., 1965. A method for cluster analysis. *Biometrics*, pp.362-375.

[10] Auger, I.E., and Lawrence, C.E., 1989. Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology*, *51*(1), pp.39-54.

[11] Jackson, B., Scargle, J.D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L. and Tsai, T.T., 2005. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, *12*(2), pp.105-108.

[12] Killick, R., Fearnhead, P. and Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), pp.1590-1598.

[13] Touzani, S., Granderson, J. and Fernandes, S., 2018. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, *158*, pp.1533-1543.

[14] Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., Liu, B., Halverson, M., Winiarski, D., Rosenberg, M. and Yazdanian, M., 2011. US Department of Energy commercial reference building models of the national building stock.

[15] Franconi, E., Gee, M., Goldberg, M., Granderson, J., Guiterman, T., Li, M. and Smith, B., 2017. The Status and Promise of Advanced M&V: An Overview of "M&V 2.0" Methods, Tools, and Applications.

[16] Killick, R. and Eckley, I., 2014. changepoint: An R package for changepoint analysis. *Journal of statistical software*, *58*(3), pp.1-19.

[17] Montero, P. and Vilar, J.A., 2014. Tsclust: An r package for time series clustering. *Journal of Statistical Software*, *62*(1), pp.1-43.

Appendix:

In order to assess the impact of varying the threshold *th* in Equation 7 on the results of the NRE detection algorithms, the accuracy metrics were recalculated using two different values for *th* (i.e., *th=3* and *th=1*). Table A.1 and Table A.2 summarize the results for threshold *th=3* and *th=1* respectively. In comparison to the results with *th=2*, there was no improvement in the results for algorithm 1 when *th=3* was used. For algorithm 2 and algorithm 3 the results were improved for two scenarios (scenarios PS_3 and LO_1 for algorithm 2; scenarios PS_4 and LO_4 for algorithm 3). With *th=1* the results of algorithm 1 were the most impacted in comparison to the two other algorithms. In three out of eight test scenarios (scenarios PS_1, PS_2 and PS_3) the NRE detection accuracy decreased. For algorithm 2 and algorithm 3 the accuracy deteriorated for one test scenario each (i.e., PS_2 for algorithm 2 and PS_1 for algorithm 3). For both values of *th* (i.e., *th=1* and *th=3*) the algorithm 1 outperformed the two others in the majority of test scenarios.

Table A.1- Accuracy of NRE detection for the eight tested scenarios using *th=3* in Equation 7. The numbers of true positives and false positives are shown in parentheses

Building			Number of actual NRE	Algorithm 1			Algorithm 2			Algorithm 3		
Name	Type	Location		Number of detected NRE	True Positive	False Positive	Number of detected NRE	True Positive	False Positive	Number of detected NRE	True Positive	False Positive
PS_1	Primary Schools	Chicago	4	8	75% (3)	63% (5)	7	50% (2)	72% (5)	32	100% (4)	88% (28)
PS_2		Miami	4	7	100% (4)	43% (3)	5	100% (4)	20% (1)	1	0% (0)	100% (1)
PS_3		San Francisco	4	6	100% (4)	33% (2)	8	100% (4)	50% (4)	31	50% (2)	94% (29)
PS_4		Phoenix	4	5	100% (4)	20% (1)	7	75% (3)	57% (4)	9	25% (1)	89% (8)
LO_1	Large Offices	Miami	4	6	100% (4)	33% (2)	8	100% (4)	50% (4)	2	0% (0)	100% (2)
LO_2		Chicago	4	5	75% (3)	40% (2)	2	25% (1)	50% (1)	5	0% (0)	100% (5)
LO_3		San Francisco	4	5	75% (3)	40% (2)	6	50% (2)	67% (4)	1	25% (1)	0% (0)
LO_4		Phoenix	3	9	100% (3)	67% (6)	9	100% (3)	67% (6)	7	66% (2)	66% (5)

Table A.2- Accuracy of NRE detection for the eight tested scenarios using *th=1* in Equation 7. The numbers of true positives and false positives are shown in parentheses

Building			Number of actual NRE	Algorithm 1			Algorithm 2			Algorithm 3		
Name	Type	Location		Number of detected NRE	True Positive	False Positive	Number of detected NRE	True Positive	False Positive	Number of detected NRE	True Positive	False Positive
PS_1	Primary Schools	Chicago	4	8	50% (2)	75% (6)	7	50% (2)	72% (5)	32	50% (2)	94% (30)
PS_2		Miami	4	7	25% (1)	85% (6)	5	50% (2)	20% (3)	1	0% (0)	100% (1)
PS_3		San Francisco	4	6	75% (3)	50% (3)	8	75% (3)	63% (5)	31	50% (2)	94% (29)
PS_4		Phoenix	4	5	100% (4)	20% (1)	7	75% (3)	57% (4)	9	0% (0)	100% (9)
LO_1	Large Offic	Miami	4	6	75% (3)	50% (3)	8	75% (3)	63% (5)	2	0% (0)	100% (2)

LO_2	es	Chicago	4	5	75% (3)	40% (2)	2	25% (1)	50% (1)	5	0% (0)	100% (5)
LO_3		San Francisco	4	5	75% (3)	40% (2)	6	50% (2)	67% (4)	1	25% (1)	0% (0)
LO_4		Phoenix	3	9	100% (3)	67% (6)	9	100% (3)	67% (6)	7	33% (1)	85% (6)