

Manuscript Number:

Title: A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings

Article Type: Review Article

Section/Category: Markets, Applications and Industry

Keywords: building performance analysis, data mining, unsupervised learning, visual analytics, clustering, novelty detection, smart meter analysis, portfolio analysis, review, building controls and optimization

Corresponding Author: Mr. Clayton Carl Miller,

Corresponding Author's Institution: ETH Zürich

First Author: Clayton Carl Miller

Order of Authors: Clayton Carl Miller; Zoltan Nagy; Arno Schlueter

Abstract: Measured and simulated data sources from the built environment are increasing rapidly. It is becoming normal to analyze data from hundreds, or even thousands of buildings at once. Mechanistic, manual analysis of such data sets is time-consuming and not realistic using conventional techniques. Thus, a significant body of literature has been generated using unsupervised statistical learning techniques designed to uncover structure and information quickly with fewer input parameters or meta data about the buildings collected. Further, visual analytics techniques are developed as aids in this process for a human analyst to utilize and interpret the results. This paper reviews publications that include the use of unsupervised machine learning techniques as applied to non-residential building performance control and analysis. The categories of techniques covered include clustering, novelty detection, motif and discord detection, rule extraction, and visual analytics. The publications apply these technologies in the domains of smart meters, portfolio analysis, operations and controls optimization, and anomaly detection. A discussion is included of key challenges resulting from this review, such as the need for better collaboration between several, disparate research communities and the lack of open, benchmarking data sets. Opportunities for improvement are presented including methods of reproducible research and suggestions for cross-disciplinary cooperation.

Suggested Reviewers: Joern Ploennings  
IBM Research  
Joern.Ploennings@ie.ibm.com

Fu Xiao  
The Hong Kong Polytechnic University  
linda.xiao@polyu.edu.hk

David Lehrer

UC Berkeley  
lehrer@berkeley.edu

Opposed Reviewers:



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

**Institute of Technology in Architecture**

ETH Zurich  
Mr. Clayton Miller  
Researcher  
Future Cities Laboratory  
1 CREATE Way #06-01  
Singapore

Lawrence Kazmerski  
Editor-in-Chief, Renewable & Sustainable Energy Reviews  
National Renewable Energy Laboratory (NREL)  
Golden, Colorado, USA

Phone (Worldwide Contact Number):  
+1 402 403 0090  
miller.clayton@arch.ethz.ch  
www.systems.arch.ch

Singapore, May 25, 2016

**Journal Paper Submission**

Dear Dr. Kazmerski:

The attached manuscript entitled "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings" is an original research article that I would like to submit for consideration in Renewable & Sustainable Energy Reviews. This paper gives an overview of data mining techniques from various domains as applied to building performance research. This paper was funded through a Fellowship at the Institute of Technology in Architecture (ITA) at the ETH Zürich. My co-authors and I have included a few of our own publications from the past as part of this review.

Any questions or comments can be directed to myself as the corresponding author. I am available via email at miller.clayton@arch.ethz.ch and by phone at 1-402-403-0090.

Kind Regards,

A handwritten signature in black ink, appearing to read 'Clayton Miller', with a stylized flourish at the end.

Clayton Miller

# A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings

Clayton Miller, Zoltán Nagy, Arno Schlueter

*ETH Zürich, Institute of Technology in Architecture (ITA), Architecture and Building Systems (A/S)  
John-von-Neumann-Weg 9, Zürich, Switzerland*

---

## Abstract

Measured and simulated data sources from the built environment are increasing rapidly. It is becoming normal to analyze data from hundreds, or even thousands of buildings at once. Mechanistic, manual analysis of such data sets is time-consuming and not realistic using conventional techniques. Thus, a significant body of literature has been generated using unsupervised statistical learning techniques designed to uncover structure and information quickly with fewer input parameters or meta data about the buildings collected. Further, visual analytics techniques are developed as aids in this process for a human analyst to utilize and interpret the results. This paper reviews publications that include the use of unsupervised machine learning techniques as applied to non-residential building performance control and analysis. The categories of techniques covered include clustering, novelty detection, motif and discord detection, rule extraction, and visual analytics. The publications apply these technologies in the domains of smart meters, portfolio analysis, operations and controls optimization, and anomaly detection. A discussion is included of key challenges resulting from this review, such as the need for better collaboration between several, disparate research communities and the lack of open, benchmarking data sets. Opportunities for improvement are presented including methods of reproducible research and suggestions for cross-disciplinary cooperation.

*Keywords:* building performance analysis, data mining, unsupervised learning, visual analytics, clustering, novelty detection, smart meter analysis, portfolio analysis, review, building controls and optimization

---

## 1. Introduction

The creation and consolidation of measured sensor sources from the built environment and its occupants is occurring on an unprecedented scale. The Green Button Ecosystem

---

*Email address:* miller.clayton@arch.ethz.ch (Clayton Miller)  
*URL:* <http://systems.arch.ethz.ch/> (Clayton Miller)  
*Preprint submitted to Renewable & Sustainable Energy Reviews* *May 25, 2016*

now enables the easy extraction of data from over 60 million buildings<sup>1</sup>. Advanced metering infrastructure (AMI), or smart meters, have been installed on over 58.5 million buildings in the US alone [1]. The Internet-of-Things (IoT) movement provides an array of low-cost sensors, data acquisition devices, and cloud storage. A recent study has predicted a \$70-150 billion impact of IoT in offices and \$200-350 billion in homes [2]. A recent press release from the White House summarizes the impact of utilities and cities in unlocking these data [3]. It announces that 18 utilities, serving more than 2.6 million customers, will provide detailed energy data by 2017. This study also suggests that such accessibility will enable improvement of energy performance in buildings by 20% by 2020. A vast majority of these raw data being generated are sub-hourly temporal data from meters and sensors.

Ruparathna et al. created a contemporary review of building performance analysis techniques for commercial and institutional buildings [4]. This review was comprehensive in capturing approaches related to technical, organizational, and behavioral changes. The majority of publications considered fall within the domains of automated fault detection and diagnostics, retrofit analysis, building benchmarking, and energy auditing. These traditional techniques focus on one building or a small, related collection of buildings, such as a campus. Many require complex characteristic data about each building, such as its geometric dimensions, building materials, the age of type of mechanical systems, and other metadata, to execute the process.

Thus, a critical issue facing the building industry is how these traditional performance analysis techniques can utilize the current explosion of detailed, temporal sources. *If one has access to raw data from thousands, or even millions, of buildings, how can analysis be scaled in a reasonable way?* Someone tasked with this type of analysis needs to extract information with less known meta-data about each building and fewer inputs into the process. In response to this question, researchers from several domains have developed methods of extracting insight from raw, unlabeled data from the built environment. Often these methods fall into the category of unsupervised statistical learning. Methods from this sub-domain of machine learning are advantageous due to their ability to characterize measured or simulated performance data quickly with less analyst intervention, metadata, and ground truth labeled data. They are often used in conjunction with visual analytics approaches as the first pass in the analysis of a large dataset.

A major challenge in utilizing this body of knowledge is that unsupervised machine learning techniques have been proposed, tested, and applied concurrently in several research domains and sub-domains. This situation has resulted in parallel, and mostly non-overlapping, collections of work with similar objectives, yet different publication venues and terminology. This review seeks to quantify this situation through an overview of the use of unsupervised learning techniques developed for the building industry. This task is done more holistically than previous reviews by combining key publications from a larger set of category applications and research domains. This overview is the first to specifically target unsupervised learning and visual analytics techniques for this context. The research sectors of building energy analysis, building simulation, computer science and electrical engineering are analyzed.

---

<sup>1</sup>According to: <http://www.greenbuttondata.org/>

This paper compiles a set of 100 publications released since 2005 that focus on the use of unsupervised statistical learning techniques on temporal or spatial data from non-residential buildings.

The paper is structured as follows, Section 2 gives background on the unsupervised machine learning used in the publications. Section 3 provides an overview of various trends and statistics about the publications, including specific methods, the yearly timeline of release, and which research domains and specific journal and conference publications contain them. Sections 4-7 evaluate each of the publication categories in-depth. Finally, Section 8 provides a discussion of the challenges observed and potential opportunities discovered through the review process.

### 1.1. Previous Reviews

Various reviews have been completed that overlap with a part of the scope of this paper. Most of them are designed to focus on a single core domain of research; the main two areas are building operations analysis and smart grid optimization. One of the earliest reviews of artificial intelligence techniques for buildings was completed in 2003 by Krarti and covered both supervised and unsupervised methods [5]. Dounis updated this work and focused on outlining specific techniques in detail [6]. Reddy's seminal book about a large variety of analysis techniques for energy engineers includes chapters on clustering and unsupervised methods specifically [7]. Lee et al. describe a variety of retrofit analysis toolkits which incorporate unsupervised and visual analytics approaches in a practical sense [8]. Ioannidis et al. created a large ontology of data mining and visual analytics for building performance analysis, however with a strong focus on the techniques and not examples of works using them [9]. From the utility and power grid side, Morias et al. created a general overview of various data mining techniques as focused on power distribution systems [10]. Chicco covered clustering methods specifically focused on load profiling tasks [11]. Zhou et al. included the concept of customer load classification [12].

## 2. Background of Unsupervised Learning and Visual Analytics Techniques

This section gives an overview of the five categories of unsupervised learning and visual analytics techniques discussed in this paper: clustering, novelty detection, motif and discord detection, rule extraction, and visual analytics. The general nature of the techniques is discussed, and the specific methods are briefly listed along with the abbreviations that are used in Tables 1-4.

Statistical learning can be divided into two major categories: supervised learning and unsupervised learning. According to the formative text on the subject by James et al. and Hastie et al., supervised statistical learning techniques involve a set of observations  $x_i, i = 1, \dots, n$  and an associated response variable  $y_i$  [13, 14]. The goal of a supervised process is to predict  $y_i$  using the features generated from  $x_i$ ; this objective is straightforward to verify according to the algorithms accuracy in predicting that response. Unsupervised learning, in contrast, lacks the associated response  $y_i$  because it seeks merely to understand

the relationship between the observations in  $x_i$ , generally in an exploratory fashion. These references go on to explain that unsupervised techniques are, thus, often more subjective in their application and are considered more challenging in their utilization in practical applications [15, 13, 14]. Despite the challenge, various instances of their implementation are found in the literature with relation to building performance data as outlined in the subsequent sub-sections.

## 2.1. Clustering

Clustering is the most common general unsupervised approach applied to building performance data. It is used automatically to generate *subgroups* of similar types of observations. James et al. describe this process as the grouping of  $n$  observations into  $K$  groups, or clusters, according to a set of generated  $p$  features [13]. The two most common types of clustering are K-means and Hierarchical clustering. A wider array of techniques has been developed to optimize the objectives of separating subgroups in more effective and efficient ways. An example of the clustering of daily energy consumption profiles can be seen in Figure 1. This example shows three and half years of hourly energy data for a building in which the diurnal patterns are grouped according to similarity and plotted across the time range of the data set, in addition to the characteristic clustered profiles plotted across a 24 hour period. This application of clustering is known as load profiling, and it is covered in Section 4. Certain types of clustering techniques found in publications in this review are K-means, Principal Component Analysis (PCA), Maximum Likelihood (ML), Hierarchical, Self-Organizing Maps (SOM), Empirical Mode Decomposition (EMC), Fuzzy clustering, Support Vector Machines (SVM), and Ant Colony Clustering.

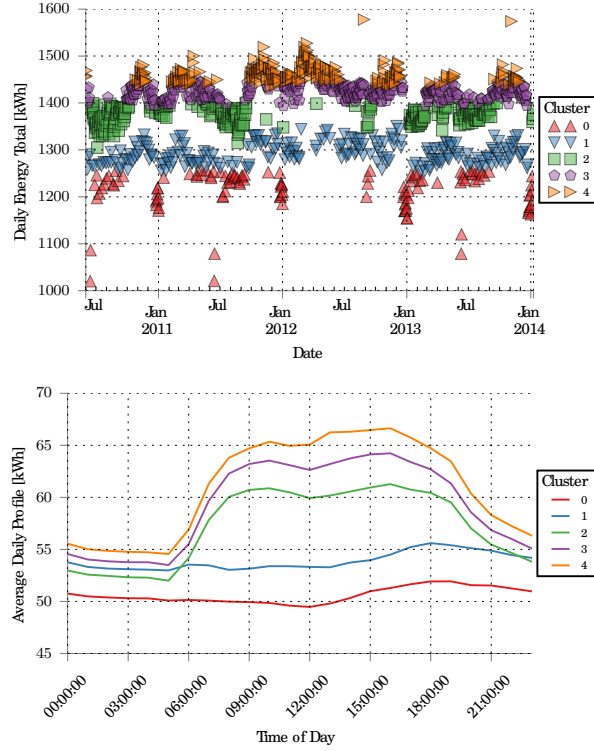


Figure 1: Load profiles clusters from a building electricity consumption data set (adapted from [16])

## 2.2. Novelty Detection

Novelty detection is a category of techniques that finds behavior or observations within a data set that are unique and often do not conform to expected behavior [17]. Concepts such as outlier and anomaly detection fit within this category. An example of an outliers detection process can be seen in Figure 2 in which a set of *quality analysis metrics* were implemented on a single data stream including an outliers detection component that is designated as quality level 2. Certain types of novelty detection techniques in this review include Principal Component Analysis (PCA), Generalized Additive Models (GAM), Classification and Regression Trees (CART), Wavelets, Fourier Transforms, Linear Discriminant Analysis (LDA), and Nearest Neighbor methods.



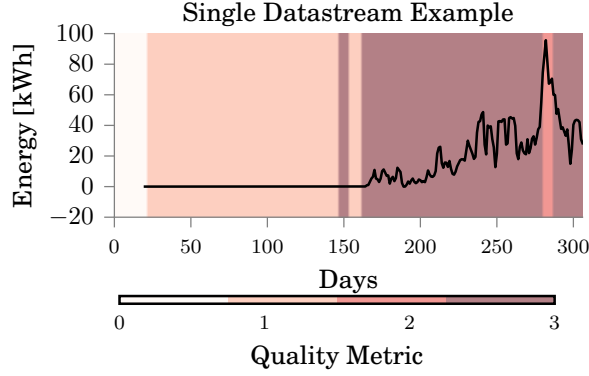


Figure 2: Example of novelty detection through data quality metrics in which a value of 2 signifies and statistical outlier (adapted from [16])

### 2.3. Motif and Discord Detection

Motif and discord detection is a sub-domain unique to the investigation of time-series data sets [18]. A motif is a subsequence of data that exhibits a pattern that frequently occurs in a data stream. A discord is a subsequence that occurs rarely and is considered anomalous amongst the rest of the data set. Figure 3 illustrates the concept of motif and discord candidate identification on a two-week sample data set. This example is of thirteen days of electrical consumption where the diurnal patterns were extracted using Symbolic Aggregate Approximation (SAX) and divided into motif and discord candidates according to their frequency amongst the sample set. Other approaches found in the literature related to motif and discord detection include Markov Models, Maximum Likelihood, Dirichlet Process Gaussian Mixture Models (DPGMM), and Incremental Summarization and Pattern Characterization (ISPC).

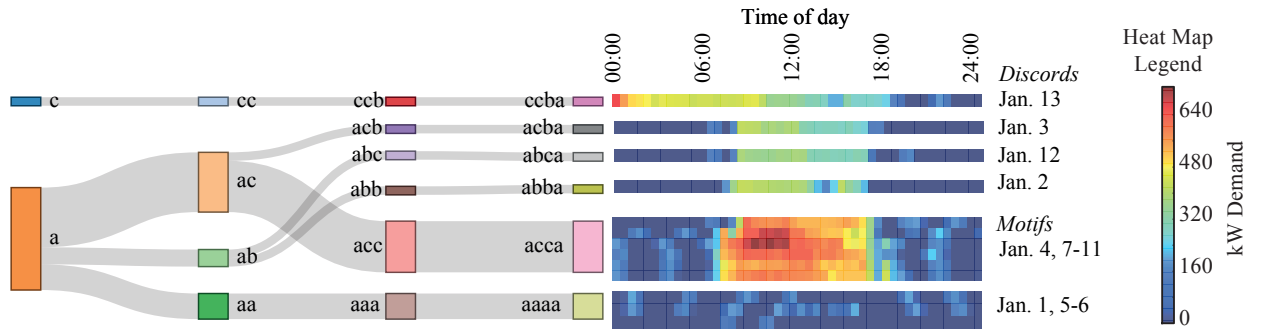


Figure 3: Example of motif and discord candidate extraction from a two week example data set (adapted from [19])

### 2.4. Rule Extraction

Rule extraction techniques focus on the ability automatically to find relationships between variables within a data set. Hastie et al. describe the goal of these processes as

attempting to find shared values of the variables  $X = (X_1, X_2, \dots, X_p)$  that appear most frequently in the database [14]. These rules are often used to support control sequences and fault detection approaches at the component level. The general field of creating Data Association Rules (DAC) is the most common form of this technique. Other methods found in this review include Markov Models and Support Vector Machines (SVM).

## 2.5. Visual Analytics

The domain of visual analytics is relatively new and rapidly developing. Keim et al. define the field as a medium of a semi-automated analytical process, where humans and machines cooperate using their respective distinct capabilities for the most useful results [20]. This domain focuses on the use of visualization and various human interactions such as the Schneiderman mantra of Overview First, Filter and Zoom, and Details-on-Demand [21]. Visual analytics is developing in the building industry through the proliferation of energy management system dashboards. This review divides the visual analytics for building energy into two categories: approaches focused on finding patterns in energy data and those on more general energy management with focuses on performance indicators.

Many of the publications in this review indicate a dominant technique that is determined and included in Sections 4-7 and their associated tables. Some of the publications don't specify a single technique or implement numerous techniques for comparison; these publications are classified under the category of *Multiple Techniques*.

## 3. Methodology and Overview of Publications

This review was started through a selection of unsupervised analytics categories outlined by authoritative sources from the machine learning community [14, 13, 22, 15]. The categories selected are clustering, novelty detection, motif and discord detection, and rule extraction. The field of visual analytics was added to these groups to cover the presentation layer of many of these types of techniques. An initial search of publications was then selected for inclusion through a Google Scholar search of the combination of the method categories and the terms "building energy", "building performance analysis", and "building energy analysis". From this initial list of publications, a set of application categories and sub-categories was developed as seen in Figure 4. A more detailed search of each application class was then completed to account for the unique analytics techniques used in those domains. Only publications with a majority of the focus on utilization of unsupervised techniques and with a focus only on non-residential buildings are reviewed. Only works completed since 2005 are included to discuss only the most contemporary work and due to the relatively recent development of most of the techniques examined. A cutoff date of April 1, 2016, is applied for inclusions of publications in this review.

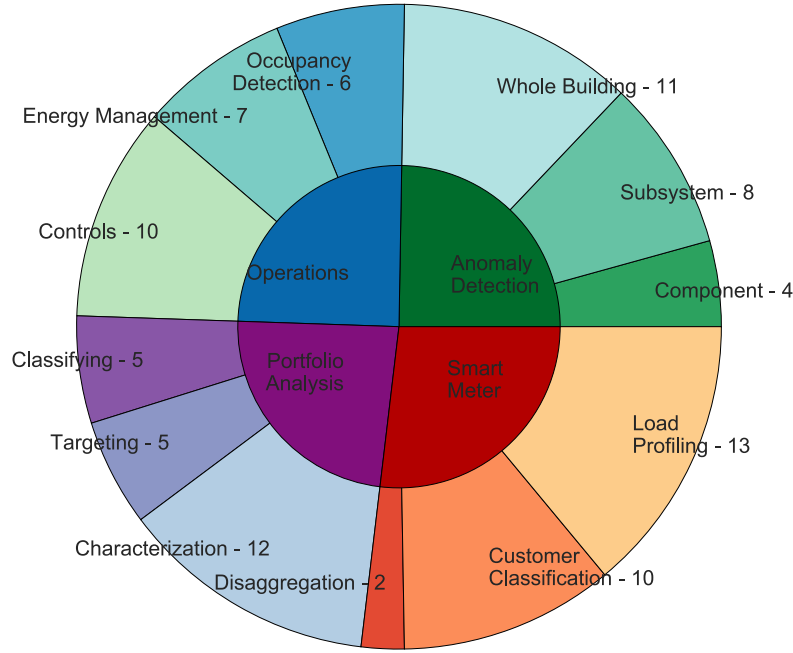


Figure 4: Categories and sub-categories (including number of publications) of building performance analysis applications of unsupervised learning and visual analytics

Analysis of the publications discussed in this paper at a high level reveals insights regarding the frequency and types of techniques used, the number of publications and trend over time, and the popularity of specific publications within the domain.

### 3.1. Specific Techniques

Figure 5 illustrates the frequency of use of various unsupervised and visual analytics methods in the publications reviewed. The largest specific category is papers that use numerous techniques in conjunction to achieve a complex goal. Amongst the papers that have a dominant method, K-Means clustering, data association rules, and visual analytics-based pattern analysis and energy information systems round out the top five.

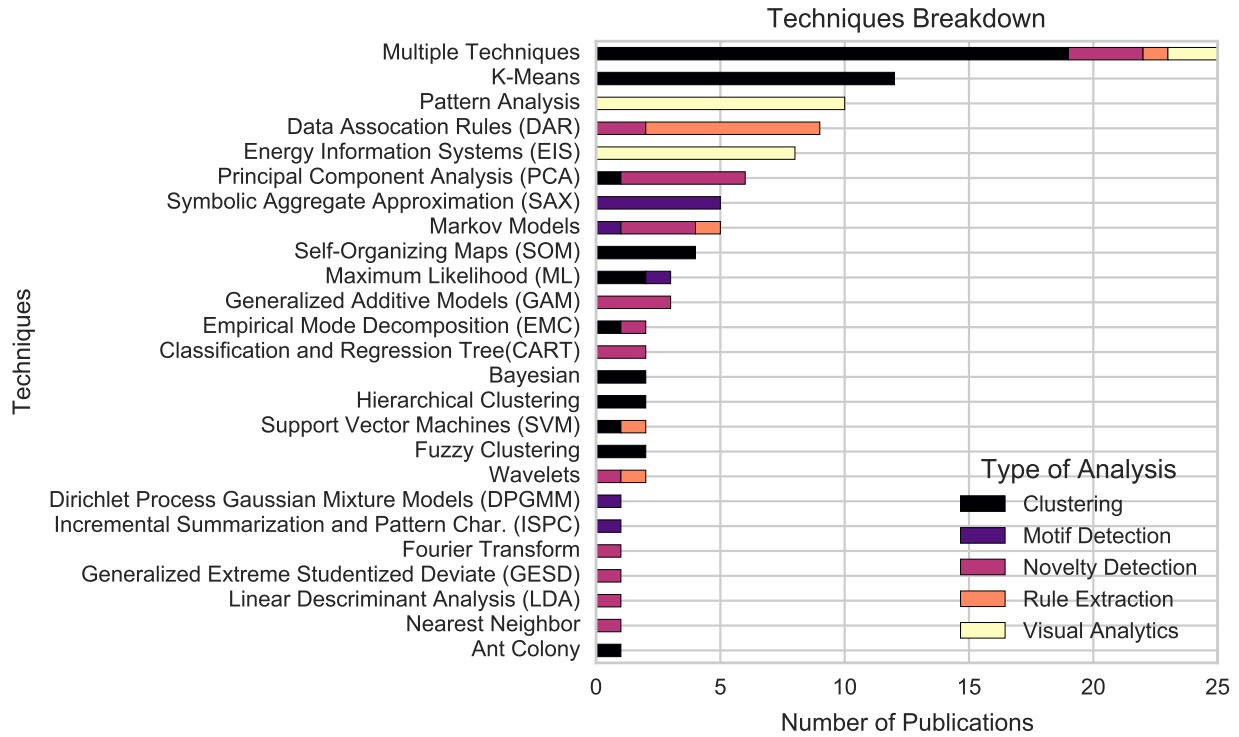


Figure 5: Breakdown of specific and general analytics techniques utilized

### 3.2. Research Sectors

Figure 6 illustrates the breakdown of publications based on the year published since 2005. They are further divided into four broad research domains: building energy analysis, building simulation, computer science and electrical engineering. These research field categories were subjectively determined for each paper through evaluating a combination of which university department the authors were from and in which publication the study was published. Building energy analysis pertains to researchers who predominantly focus on measured data analysis from buildings while simulation experts research forward modeling and simulation of building and urban systems. Both fields of study most often exist within architecture or mechanical engineering departments. Electrical engineering and computer science are two well-established domains and exist in their own departments. It is noticed that there is a gradual increase in the number of publications over the last ten years with electrical engineering and building energy analysis being the most common in the first few years and computer science and building simulation picking up since 2008.

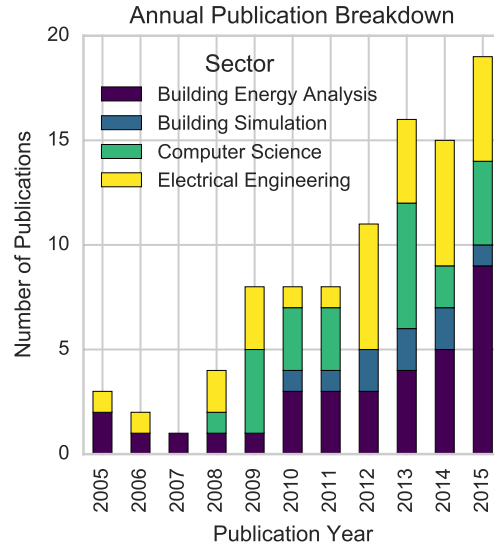


Figure 6: Breakdown of publications by year published and research domain

### 3.3. Publications Venues

This section analyzes the prevalence of certain publication venues within this review. Figure 7 illustrates the breakdown of the publication venues represented. The Energy and Buildings Journal from the building energy analysis domain dominates this list with 17 articles. Building simulation and energy analysis research domains publish most often in this journal as well as Applied Energy and Energy Efficiency. Several IEEE conferences and journals are also dominant as most of the papers from the electrical engineering domain are in these venues.

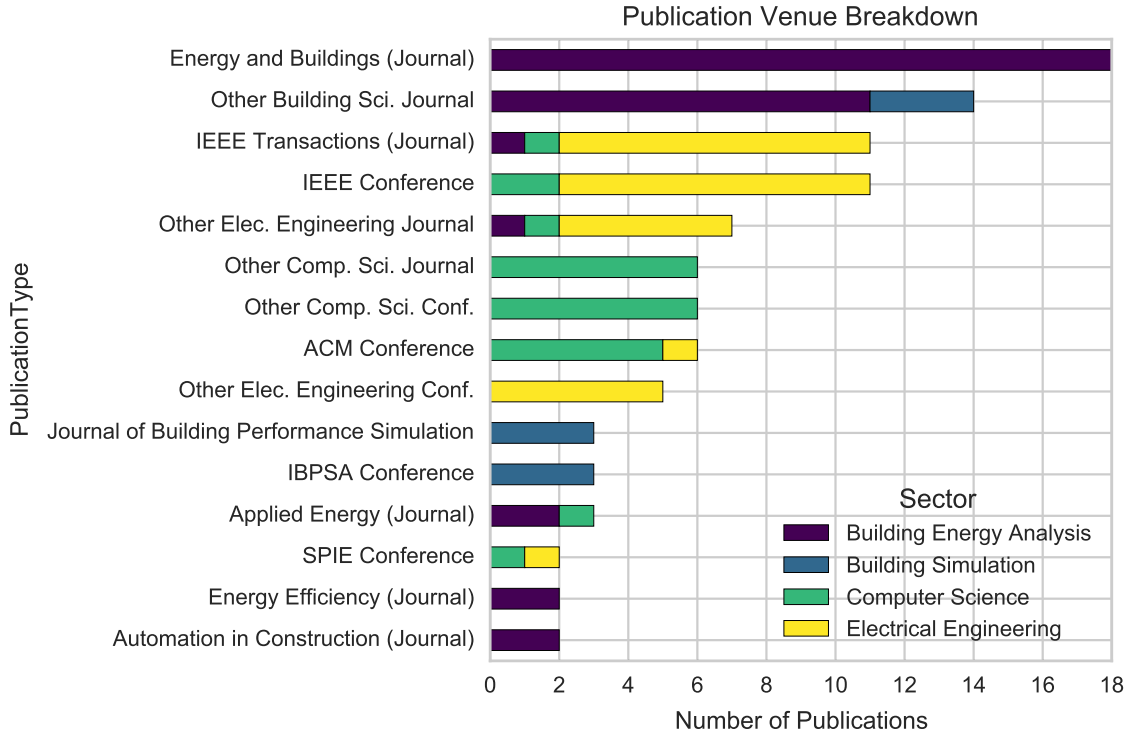


Figure 7: Breakdown of publications by publication type and research domain

The next sections discuss each application category and sub-category of building performance analysis in more detail including a short overview of each publication.

#### 4. Smart Meter Analytics

Advanced Metering Infrastructure (AMI), also known as smart meter systems, is a network of energy meters, most often focused on the electrical power measurement of a whole building. These systems are implemented and utilized by electrical utility providers. The growth of these meters has been rapid in many parts of the world. As mentioned in Section 1, there are over 58 million smart meters alone installed in the United States [1]. Conventional metering infrastructure only facilitates monthly data collection for billing purposes, while the new AMI framework allows for sub-hourly electrical demand readings. These data are primarily used for demand characterization and billing, however, many additional uses are being discovered. A wide-range of studies have been completed in recent years to focus on a range of issues related to automatically extracting information from these data using unsupervised techniques. In this section, three sub-categories of application are discussed: load profiling, account classification, and disaggregation. Table 1 lists these publications and their general attributes.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2016	Customer Class.	Building Energy	Clustering	K-Means	[23]
2015	Load Profiling	Computer Science	Clustering	K-Means	[24]
2015	Customer Class.	Computer Science	Visual Analytics	Patterns	[25]
2015	Load Profiling	Elec. Engineering	Clustering	PCA	[26]
2015	Load Profiling	Elec. Engineering	Clustering	Multiple	[27]
2015	Customer Class.	Computer Science	Novelty Detection	Markov	[28]
2014	Disaggregation	Elec. Engineering	Motif Detection	SAX	[29]
2014	Load Profiling	Building Energy	Clustering	K-Means	[30]
2014	Load Profiling	Elec. Engineering	Clustering	K-Means	[31]
2014	Load Profiling	Computer Science	Visual Analytics	Patterns	[32]
2014	Customer Class.	Elec. Engineering	Visual Analytics	Multiple	[33]
2013	Disaggregation	Computer Science	Motif Detection	DPGMM	[34]
2013	Load Profiling	Elec. Engineering	Clustering	Ant Colony	[35]
2013	Customer Class.	Building Energy	Clustering	Multiple	[36]
2013	Load Profiling	Elec. Engineering	Clustering	Multiple	[37]
2012	Load Profiling	Elec. Engineering	Clustering	Multiple	[38]
2012	Customer Class.	Building Sim.	Clustering	Bayesian	[39]
2011	Load Profiling	Elec. Engineering	Motif Detection	ISPC	[40]
2010	Customer Class.	Elec. Engineering	Clustering	Multiple	[41]
2009	Load Profiling	Computer Science	Clustering	Multiple	[42]
2009	Load Profiling	Elec. Engineering	Clustering	Multiple	[43]
2009	Load Profiling	Elec. Engineering	Clustering	SVM	[44]
2009	Customer Class.	Elec. Engineering	Clustering	Multiple	[45]
2008	Customer Class.	Computer Science	Clustering	SOM	[46]
2006	Customer Class.	Elec. Engineering	Clustering	SOM	[47]
2005	Customer Class.	Elec. Engineering	Clustering	K-Means	[48]

Table 1: Publications from the Smart Meter category

#### 4.1. Load Profiling

Load profiling is the process of grouping temporal subsequences of measured energy data for the purpose of characterizing the typical behavior of an individual customer. It involves time-series clustering and feature extraction. Chicco et al. provide an original example in our review of this process using support vector machine clustering [44]. Gullo et al. and Räsänen et al. took the process further by introducing a framework of various clustering procedures that were implemented on case studies [42, 43]. Ramos et al., Iglesias et al., and Panapakidis et al. tested various conventional and new clustering methods and similarity metrics to determine those most applicable to electrical load profiling [37, 38, 27]. Chicco et al. explored new clustering techniques based on ant colony grouping while Pan et al. discovered the use of kernel PCA for the same purpose [35, 26]. Several groups of researchers such as Lavin and Klabjan and Green et al. have found efficient use in using the core K-Means clustering algorithm for load profiling [30, 31]. Shahzadeh et al. discussed the use of profiling as applied to forecast accuracy of temporal data [24]. Two studies diverge from the standard profile development using clustering paradigm. The first is by De Silva

et al. who uses Incremental Summarization and Pattern Characterization (ISPC) instead of clustering to find load profiles [40]. The other is the visual analytics-based approach of creating a smart meter analytics dashboard by Nezhad et al. to set up and inspect typical load profiles [32].

#### 4.2. Customer Classification

Automated account classification is the next sub-category that utilizes unsupervised learning techniques within the smart meter domain. These methods often employ load profile clustering as a first step but differentiate themselves in using those features to classify accounts, or buildings, that fit within various categories. Therefore, account classification is a type of manual semi-supervised analysis utilizing load profiling as a basis. The study by Figueiredo et al. harnessed K-Means and a labeled sample from accounts in Portugal to showcase this concept [48]. Verdu et al. and Räsänen et al. applied self-organizing maps (SOM) to accomplish a similar study that classifies accounts according to the applicability of several demand response scenarios [47, 46]. Vale et al. give an overview of a general data mining framework focused on characterizing customers [45]. Florita et al. diverge from the use of measured data by creating a massive amount of simulation data of load profiles to quantify energy storage applications for the power grid [39]. Fagiani et al. use Markov Model novelty detection to automatically classify customers who potentially have leakage or waste issues [28]. Cakmak et al. and Liu et al. test new visual analytics techniques within more holistic analysis framework for analyzing customers [33, 25]. Borgeson used various clustering and occupancy detection techniques to analyze a large AMI data set from California [36]. Bidoki et al. tested various clustering techniques to evaluate applicability for customer classification [41]. A recent study in Korea develops a new clustering method for segmenting customers to analyze demand response incentives [23].

#### 4.3. Disaggregation

The last area of smart meter data analysis is the field of meter disaggregation. Disaggregation attempts decompose a measurement signal from a high level reading to the individual loads being measured. This domain is well-researched from a supervised model perspective but recent attempts at unsupervised, pattern-based disaggregation were developed to facilitate implementation on unlabeled smart meter data. Shao et al. use Dirichlet Process Gaussian Mixture Models to find and disaggregate patterns in sub-hourly meter data [34]. Reinhardt and Koessler use a version of symbolic aggregate approximation (SAX) to extract and identify disaggregated patterns for the purpose of prediction [29]. These studies are also unique in that few of the disaggregation studies focus on commercial buildings as opposed to residential buildings.

### 5. Portfolio Analytics

Portfolio analysis is a domain in which a large group of buildings, often located in the same geographical area or owned or managed by the same entity, are analyzed for the purpose of managing or optimizing the group as a whole. Each subsection covers the publications



reviewed in this domain that fall into three categories: characterization, classification, and targeting. Table 2 illustrates an overview of all the publications in this section in addition to their attributes.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2016	Targeting	Building Energy	Clustering	Fuzzy Clust.	[49]
2016	Targeting	Building Energy	Clustering	Multiple	[50]
2015	Characterization	Building Sim.	Clustering	Multiple	[16]
2015	Characterization	Building Energy	Visual Analytics	EIS	[51]
2015	Characterization	Building Energy	Motif Detection	SAX	[19]
2015	Characterization	Elec. Engineering	Visual Analytics	EIS	[52]
2015	Classifying	Building Energy	Clustering	Multiple	[53]
2014	Classifying	Elec. Engineering	Novelty Detection	GAM	[54]
2014	Characterization	Building Sim.	Visual Analytics	Patterns	[55]
2014	Classifying	Building Energy	Clustering	K-Means	[56]
2013	Characterization	Computer Science	Visual Analytics	Patterns	[57]
2013	Targeting	Building Energy	Rule Extraction	DAR	[58]
2012	Classifying	Elec. Engineering	Clustering	Multiple	[59]
2012	Targeting	Building Energy	Clustering	ML	[60]
2012	Characterization	Building Sim.	Clustering	K-Means	[61]
2011	Targeting	Computer Science	Novelty Detection	Markov	[62]
2011	Characterization	Building Energy	Visual Analytics	EIS	[63]
2010	Characterization	Building Energy	Visual Analytics	EIS	[64]
2010	Targeting	Building Energy	Clustering	Multiple	[65]
2009	Targeting	Computer Science	Clustering	ML	[66]
2009	Characterization	Building Energy	Visual Analytics	EIS	[67]
2009	Characterization	Computer Science	Visual Analytics	EIS	[68]
2008	Characterization	Building Energy	Novelty Detection	PCA	[69]
2007	Classifying	Building Energy	Clustering	Fuzzy Clust.	[70]
2005	Characterization	Building Energy	Clustering	Hierarchical	[71]

Table 2: Publications from the Portfolio Analysis category

### 5.1. Characterization

Publications that address the characterization of a portfolio of buildings include unsupervised techniques meant to evaluate and visualize the range of behaviors and performance of the group. A majority of the techniques utilized are either clustering or visual analytics that provide a model of exploratory analysis that enable further steps. Seem produced an influential study that extracts days of the week with similar consumption profiles [71]. Further clustering work was completed by An et al. to estimate thermal parameters of a portfolio of buildings [61]. Lam et al. used Principal Component Analysis to extract information about a group of office buildings [69]. Approaches focused on visual analytics and dashboards were completed by Agarwal et al., Lehrer, and Lehrer and Vasudev [68, 67, 63]. Granderson et al. completed a case study-based evaluation of energy information systems, in which some methods combine some unsupervised approaches with visualization [64]. Diong et al. completed a case study as well focused on a specific energy information system implementation

[52]. Morán et al. and Georgescu and Mezic developed hybrid methods that employed visual continuous maps and Koopman Operator methods respectively to visualize portfolio consumption [57, 55]. Miller et al. completed two studies focused on the use of screening techniques to automatically extract diurnal patterns from performance data and use those patterns to characterize the consumption of a portfolio of buildings [16, 19]. Yarbrough et al. used visual analytics techniques to analyze peak demand on a university campus [51].

## 5.2. Classification

The concept of classifying buildings within a portfolio supplements the characterization techniques by assigning individual buildings to subgroups of relative performance for the purpose of benchmarking or decision-making. Santamouris et al. produced a report using clustering and classification to assign schools in Greece to subgroups of similar performance [70]. Nikolaou et al. and Pieri et al. further extended this type of work to office buildings and hotels [59, 53]. Heidarinejad et al. released an analysis of clustered simulation data to classify LEED-certified office buildings [56]. Ploennigs et al. created a platform for monitoring, diagnosing and classifying buildings and operational behavior within a portfolio to quickly visualizing the outputs [54].

## 5.3. Targeting

Targeting is a concept that builds upon characterization and classification to identify specific buildings or measures to be implemented in a portfolio to improve performance. These publications are differentiated in that specific measures are identified in the analysis. Sedano et al. use Cooperative Maximum-Likelihood amongst other techniques to evaluate the thermal insulation performance of buildings [66]. Gaitani et al. used PCA and clustering to target heating efficiency in school buildings [65]. Bellala et al. used various methods to find lighting energy savings on a campus of a large organization [62]. Petcharat et al. also found lighting energy savings in a group of buildings [60]. Cabrera and Zareipour used data association rules to complete a similar study to find wasteful patterns [58]. Geyer et al. and Schlueter et al. test various clustering strategies to group different buildings within a Swiss alpine village according to their applicability for retrofit interventions [50] and thermal micro-grid feasibility [49].

# 6. Operations, Optimization, and Controls

Unsupervised techniques focused on individual buildings themselves are placed in the category for building operations, optimization, and control. This class contains the largest number of publications, and it incorporates a wider range of applications. It is differentiated from Section 7 in that the applications are not as focused on detecting and fixing the anomalous behavior. This section evaluates publications within the sub-categories of occupancy detection, retrofit analysis, controls, and energy management. Table 3 outlines the publications in this section and their key attributes.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2016	Occupancy Detection	Building Energy	Rule Extraction	Wavelets	[72]
2015	Occupancy Detection	Computer Science	Clustering	K-Means	[73]
2015	Occupancy Detection	Building Energy	Rule Extraction	Markov	[74]
2015	Energy Management	Building Energy	Rule Extraction	DAR	[75]
2015	Energy Management	Building Energy	Visual Analytics	EIS	[76]
2015	Occupancy Detection	Building Energy	Rule Extraction	Multiple	[77]
2015	Controls	Elec. Engineering	Motif Detection	SAX	[78]
2014	Energy Management	Building Energy	Rule Extraction	DAR	[79]
2014	Controls	Elec. Engineering	Rule Extraction	SVM	[80]
2013	Energy Management	Elec. Engineering	Visual Analytics	Patterns	[81]
2013	Controls	Building Energy	Clustering	Multiple	[82]
2013	Energy Management	Building Sim.	Rule Extraction	DAR	[83]
2013	Controls	Building Sim.	Rule Extraction	DAR	[84]
2013	Controls	Building Energy	Clustering	K-Means	[85]
2013	Controls	Computer Science	Clustering	EMC	[86]
2012	Energy Management	Elec. Engineering	Visual Analytics	Patterns	[87]
2012	Occupancy Detection	Elec. Engineering	Motif Detection	ML	[88]
2011	Controls	Building Energy	Rule Extraction	DAR	[89]
2011	Energy Management	Building Energy	Visual Analytics	EIS	[90]
2011	Occupancy Detection	Computer Science	Clustering	Multiple	[91]
2011	Occupancy Detection	Building Sim.	Motif Detection	Markov	[92]
2011	Controls	Computer Science	Visual Analytics	Patterns	[93]
2010	Controls	Computer Science	Motif Detection	SAX	[94]
2009	Controls	Computer Science	Motif Detection	SAX	[95]
2008	Controls	Elec. Engineering	Clustering	Multiple	[96]

Table 3: Publications from the Operations, Optimization, and Controls category

### 6.1. Occupancy Detection

Occupancy detection using unsupervised techniques infers human presence in a non-residential building without a labeled ground truth dataset or as part of a semi-supervised approach using a subset of labeled data. This occupancy detection is then used for analysis or as inputs for control of systems. Augello et al. used multiple techniques to infer occupant presence on a campus in Italy [91]. Dong and Lam used Hidden Markov Models to detect occupancy patterns that were then used in a simulation [92]. Thanayankizil et al. developed a concept called Context Profiling in which occupancy was detected temporally and spatially [88]. Mansur et al. used clustering to detect occupancy patterns from sensor data [73]. The newest studies by Adamopoulou et al. and D'Oca and Hong use a range of techniques to extract rules related to occupancy [74, 77]. A recent study using wavelets illustrates the correlation of occupancy with actual energy consumption [72].

### 6.2. Controls

Controls optimization is an enduring field of study aimed at creating a state of the best operation and energy performance for a building system such as heating, cooling, ventilation

or lighting. Kusiak and Song created a means of optimally controlling a heating plant with clustering as a key step [96]. Patnaik et al. completed studies focused on using motif detection to find modes of chilled water plant operation that proved most optimal [94, 95]. Hao et al. built upon these concepts to create a visual analytics tool to investigate these motifs [93]. May-Ostendorp et al. used rule extraction as a means of enhancing a model-predictive control process of mixed-mode systems [89, 84]. Bogen et al. used clustering to detect usage patterns for building control system evaluation [82]. Fan et al. used clustering to enhance chiller power prediction with the ultimate goal of control optimization [85]. Hong et al. used Empirical Mode Decomposition to spatially optimize the placement of sensors in a building [86]. Domahidi et al. used support vector machines (SVM) to extract optimized rules for supervisory control [80]. Habib and Zucker use SAX to identify common motifs of an absorption chiller for the purpose of characterization and control [78].

### 6.3. Energy Management

Energy management and analysis of an individual building using unsupervised techniques is becoming common due to the increasing amounts of raw building management (BMS) and energy management system (EMS) data. Users of these techniques are often facilities management professionals or consultants who undertake the process to understand how the building is consuming energy. Duarte et al. use visual analytics to process data from an EMS along with various pre-processing techniques [90]. Lange et al. created two overview studies focused spatiotemporal visualization of building performance data and its interpretation in various case studies [87, 81]. Gayeski et al. completed a recent survey of operations professionals on their use of graphical interfaces of BMS and EMS dashboards [76]. Outside of the visual analytics realm, Fan et al., Xiao and Fan, and Yu et al. completed studies of an entire data mining using framework using data association rules to improve operational performance [75, 79, 83].

## 7. Anomaly Detection

Anomaly detection for buildings focuses on the detection and diagnostics of problems occurring within a building, its subsystems, and components. This field is most often focuses on the use of novelty detection or clustering approaches to find anomalous behavior. The sub-categories for this section are divided according to the spatial hierarchy of systems within a building; the highest level is whole building consumption, down to the subsystems such as heating, cooling or lighting and then to the individual components of those systems. Table 4 provides an overview of the publications in this category.

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2015	Whole Building	Building Energy	Novelty Detection	DAR	[97]
2015	Whole Building	Building Energy	Novelty Detection	Multiple	[98]
2015	Subsystem	Elec. Engineering	Novelty Detection	DAR	[99]

Continued on next page

Pub. Year	Sub-Application	Sector	Type of Analysis	Technique	Reference
2014	Subsystem	Building Energy	Novelty Detection	PCA	[100]
2014	Whole Building	Elec. Engineering	Novelty Detection	GAM	[101]
2014	Whole Building	Building Energy	Clustering	K-Means	[102]
2014	Subsystem	Computer Science	Novelty Detection	Multiple	[103]
2014	Subsystem	Building Sim.	Novelty Detection	PCA	[104]
2013	Whole Building	Elec. Engineering	Novelty Detection	GAM	[105]
2013	Whole Building	Computer Science	Visual Analytics	Patterns	[106]
2013	Component	Computer Science	Novelty Detection	EMC	[107]
2013	Whole Building	Computer Science	Novelty Detection	Multiple	[108]
2012	Subsystem	Elec. Engineering	Novelty Detection	Nearest Neighbor	[109]
2012	Component	Building Energy	Novelty Detection	Wavelets	[110]
2012	Whole Building	Elec. Engineering	Novelty Detection	Fourier	[111]
2012	Component	Building Energy	Rule Extraction	DAR	[112]
2010	Whole Building	Computer Science	Novelty Detection	CART	[113]
2010	Whole Building	Building Sim.	Clustering	Hierarchical	[114]
2010	Subsystem	Building Energy	Novelty Detection	PCA	[115]
2010	Subsystem	Computer Science	Visual Analytics	Patterns	[116]
2008	Subsystem	Elec. Engineering	Novelty Detection	LDA	[117]
2006	Whole Building	Building Energy	Novelty Detection	GESD	[118]
2005	Component	Building Energy	Novelty Detection	PCA	[119]

Table 4: Publications from the Anomaly Detection category

### 7.1. Whole Building

Whole building anomaly detection uses the electricity or heating and cooling energy supply in coming to a building to determine sub-sequences of poor performance. This category is complimentary to many of the Smart Meter solutions as they both focus on the use of a single data stream for a building. Seem had an early work again in this category with his work in using novelty detection to find abnormal days of consumption in buildings [118]. Liu et al. used classification and regression trees (CART) [113] and Wrinch et al. use frequency domain analysis for the same purpose [111]. Jacob et al. utilized hierarchical clustering to use as variables in regression models for whole building monitoring [114]. Fontugne et al. created a process known as the *Strip, Bind, and Search* method to automatically uncover misbehavior from the whole building level and subsequently detects the source of the anomaly [108]. Janetzko et al. developed a visual analytics platform to highlight anomalous behavior in power meter data [106]. Chou and Telaga created a hybrid whole building anomaly detection process using K-means [102]. Ploennigs et al. and Chen et al. created similar systems that use generalized additive models (GAM) [105, 101]. In the most recent work, Capozzoli et al. and Fan et al. use various techniques as part of a framework to detect and diagnose performance problems [98, 97].

### 7.2. Subsystems

Subsystem anomaly detection focuses on the use of a broader data set to detect and diagnose faults from a lower level. Yoshida et al. provided a semi-supervised approach that

1  
2  
3 seeks to determine which variables within a building are most influential in contributing  
4 to overall building performance [117]. Wang et al. use PCA to diagnose sensor failures  
5 [115]. Forlines and Wittenberg visualized multi-dimensional data using what they call the  
6 Wakame diagram [116]. Linda et al. and Wijayasekara et al. use various techniques to  
7 diagnose system faults and visualize them spatially [109, 103]. Le Cam et al. use PCA to  
8 create inverse models to detect problems in HVAC systems [104]. Li and Wen created a  
9 similar process using PCA in conjunction with wavelet transform [100]. Sun et al. used  
10 data association rules to create fault detection thresholds for finding anomalies [99].  
11  
12  
13

### 14 7.3. Components 15

16 Component level anomaly detection is a bottom-up fault detection approach that focuses  
17 on determining faults in individual equipment. Wang and Cui use PCA to detect compo-  
18 nent faults in chilled water plants [119]. Yu et al. and Fontugne et al. both compliment  
19 their work at the whole building level to find associated component performance anomalies  
20 automatically [112, 107]. Zhu et al. use wavelets to diagnose issues in air handling units  
21 (AHU) [110].  
22  
23  
24

## 25 8. Discussion 26

### 27 8.1. Challenges 28

29 Several challenges facing the use of unsupervised machine learning in building perfor-  
30 mance were uncovered through this process of review. The first relates to the effect of several  
31 traditional research sectors exploring techniques targeted on the improvement of building  
32 performance. It was found that different sets of terminology are used to describe similar  
33 concepts. For example, in the building energy analysis field, the term *fault* (such as [110])  
34 is used to describe a situation that is similar to what is labeled an *anomaly* in the data  
35 mining domain (such as [107]). Thus, discussions between these fields are restrained and  
36 completing a review of knowledge is difficult.  
37  
38

39 Another critical issue related to differences in domains is the inconsistency of success ob-  
40 jectives. Often individual papers would discuss the accuracy or efficiency of the algorithm  
41 or technical process itself (such as [37]), while others focused exclusively on the end results  
42 of the evaluation such as how much energy was saved (such as [118]). Several examples pub-  
43 lications successfully address both types of issues. For example, Ploennings et al. published  
44 studies which both addressed the applicability of generalized additive models and discussed  
45 their implementation in a platform that is applied to real buildings [105, 54]. Researchers  
46 should strive to optimize in both the theoretical and practical domains to have the most  
47 impact on real buildings.  
48  
49  
50

51 Another observation relates to the lack of easy reproducibility amongst studies. Repro-  
52 ducibility provides the ability for a third-party researcher to easily recreate the results of  
53 a study through a release of the data or code developed. Recent prominent articles have  
54 outlined the importance of reproducibility in science [120] and the sharing of data and code  
55 to enhance this pursuit [121]. The biomedical sciences research community is leading the  
56 way in this effort; editors from over 30 major journals, funding agency representatives, and  
57  
58  
59  
60  
61  
62  
63  
64  
65

scientific leaders from that field created guidelines for the enhancement of reproducibility [120]. Research from the building performance analysis community should follow this lead, specifically on machine learning and other types of empirical analysis.

Another challenge discovered is the lack of clarity regarding which is the optimal technique for each application. For example, a number of studies were completed to test the ability of clustering techniques to group similar daily load profiles [44, 40, 31, 42, 30, 38, 24]. A researcher or analyst who is searching for the best technique can see a survey of implementations through these publications; however, it's hard for them to be compared against each other as each utilizes a different data set and incorporates different methodologies. An explanation of the amount of effort needed to implement a technique is missing in most studies as well. For example, in order to implement a certain algorithm on a potential use-case or data set, an analyst is interested in which parameters need to be tuned, what labeled ground truth data should be gathered, and what expertise is necessary for understanding and implementation. This lack of comparison stifles the ability to make conclusions about the efficiency, interpretability, and appropriateness of use of each algorithm.

## 8.2. Opportunities

To address the challenges related to interdisciplinary collaboration, an opportunity exists for better dialogue between the electrical engineering, computer science, and building performance and simulation analysis research communities through cross-disciplinary venues that encourage submission of works from all domains. Several publication venues were found that publish papers written by academics from diverse backgrounds. IEEE and ACM conferences, in particular, were found to be somewhat diverse as compared to other publication outlets. The further development of these avenues will create a more welcoming environment in which different disciplines can exchange terminology, objectives, and culture.

Concerning reproducibility, the release of specific data sets for data science publications could become the norm, thus facilitating the ability for a third-party to recreate the results. The authors have published several recent papers with this situation [16, 19]. These publications include a sample data set and a set of Jupyter notebooks that can be downloaded and used to replicate the results of those studies easily. The Jupyter notebook website states that it is "an open source, web application-based document that combines live code, equations, visualizations, and explanatory text."<sup>2</sup> The use of these types of formats is an opportunity to enhance the interdisciplinary communication further through the sharing and utilization of publication data.

The final opportunity relates to the ambiguity of algorithm applicability amongst a set of publications. This phenomenon is observed in the wider data mining community as a whole [122]. In their study, Keogh et al. describe a scenario in which "Literally hundreds of papers have introduced new algorithms to index, classify, cluster, and segment time series." They go on to state, "Much of this work has very little utility because the contribution made (speed in the case of indexing, accuracy in the case of classification and clustering, model accuracy in the case of segmentation) offer an amount of improvement that would have been

---

<sup>2</sup><https://jupyter.org/>

completely dwarfed by the variance that would have been observed by testing on many real world datasets, or the variance that would have been observed by changing minor (unstated) implementation details.” They make the case that time series benchmarking data sets should be used to evaluate whether a new proposed algorithm is more beneficial as compared to previous work. The use of benchmark data sets reduces the impact of implementation bias, the disparity in the quality of implementation of a proposed approach versus its competitors, and data bias, the use of a particular set of testing data to confirm a desired finding. These biases were proven common amongst popular data mining publications, and it is suspected that they may be prevalent in the papers in this review. Benchmarking data sets for building performance analysis could be developed and promoted for use in papers similar to what was used in the *Great Building Energy Predictor Shootout* competition that was held in the mid-1990’s [123]. In this competition, standardized training and testing data sets were provided to numerous participants to determine who could create the most accurate model to predict future consumption. A modern-day *energy predictor shootout* could be held to incorporate the numerous advances made in machine learning since then. In addition to the ability to compare accuracy of algorithms, publications should also include more detailed explanations of the effort required to implement the proposed techniques such that a third-party could evaluate whether the effort-to-accuracy balance is right for their application.

## 9. Conclusion

This review provides an overview of publications from several research domains that utilize unsupervised statistical learning and visual analytics as applied to non-residential building performance analysis. The publications are categorized according to machine learning technique and application and sub-application of utilization. Clustering algorithms, and especially the K-Means algorithm, stand out as popular amongst most of the application domains. Visual analytics approaches are also quite popular as generally applied techniques. Almost half of the procedures implemented were only observed once in the publications as a dominant approach; thus, there is an opportunity to explore their potential further. The papers selected increase in number steadily over the last decade with computer science and electrical engineering slowly identifying the non-residential building industry as an excellent target for implementation of developed techniques. The dominant publication venues in this review came from the building energy analysis domain. However, many other publication venues are quickly catching up due to the increase of *big data* sources from the building and urban environments. Several key challenges were identified through this process, including a noticeable divide between several research communities, a diverse set of objectives, and the lack of reproducible studies. The need for open, benchmarking data sets is discussed and potential improvements are suggested.

## 10. Acknowledgments

This research was funded through a Fellowship from the Institute of Technology in Architecture from the ETH Zurich.



## References

- [1] Energy Information Administration, How many smart meters are installed in the United States, and who has them? (May 2015).  
URL <http://www.eia.gov/tools/faqs/faq.cfm?id=108&t=3>
- [2] James Manyika, Michael Chui, Peter Bisson, Jonathan Woetzel, Richard Dobbs, Jacques Bughin, Dan Aharon, Unlocking the potential of the Internet of Things | McKinsey & Company.  
URL [http://www.mckinsey.com/insights/business\\_technology/The\\_Internet\\_of\\_Things\\_The\\_value\\_of\\_digitizing\\_the\\_physical\\_world](http://www.mckinsey.com/insights/business_technology/The_Internet_of_Things_The_value_of_digitizing_the_physical_world)
- [3] The White House, FACT SHEET: Cities, Utilities, and Businesses Commit to Unlocking Access to Energy Data for Building Owners and Improving Energy Efficiency.  
URL <https://www.whitehouse.gov/the-press-office/2016/01/29/fact-sheet-cities-utilities-and-businesses-commit-unlocking-access>
- [4] R. Ruparathna, K. Hewage, R. Sadiq, Improving the energy efficiency of the existing building stock: A critical review of commercial and institutional buildings, *Renewable and Sustainable Energy Reviews* 53 (2016) 1032–1045. doi:10.1016/j.rser.2015.09.084.  
URL <http://www.sciencedirect.com/science/article/pii/S1364032115010540>
- [5] M. Krarti, An Overview of Artificial Intelligence-Based Methods for Building Energy Systems, *Journal of Solar Energy Engineering* 125 (3) (2003) 331. doi:10.1115/1.1592186.  
URL <http://SolarEnergyEngineering.asmedigitalcollection.asme.org/article.aspx?articleid=1456822>
- [6] A. I. Dounis, Artificial intelligence for energy conservation in buildings, *Advances in Building Energy Research* 4 (1) (2010) 267–299. doi:10.3763/aber.2009.0408.  
URL <http://www.tandfonline.com/doi/abs/10.3763/aber.2009.0408>
- [7] T. A. Reddy, *Applied data analysis and modeling for energy engineers and scientists*, Springer Science & Business Media, 2011.  
URL [https://books.google.com/books?hl=en&lr=&id=UkWvWDFiDf4C&oi=fnd&pg=PR9&dq=%22Applied+data+analysis+and+modeling+for+energy+engineers+and+scientists%22+T+Agami+Reddy&ots=tEUJEuedXl&sig=e41JNtpT78gpFI64XB6sNt\\_hQF4](https://books.google.com/books?hl=en&lr=&id=UkWvWDFiDf4C&oi=fnd&pg=PR9&dq=%22Applied+data+analysis+and+modeling+for+energy+engineers+and+scientists%22+T+Agami+Reddy&ots=tEUJEuedXl&sig=e41JNtpT78gpFI64XB6sNt_hQF4)
- [8] S. H. Lee, T. Hong, M. A. Piette, S. C. Taylor-Lange, Energy retrofit analysis toolkits for commercial buildings: A review, *Energy* doi:10.1016/j.energy.2015.06.112.  
URL <http://www.sciencedirect.com/science/article/pii/S0360544215008683>
- [9] D. Ioannidis, A. Fotiadou, S. Krinidis, G. Stavropoulos, D. Tzovaras, S. Likothanassis, Big Data and Visual Analytics for Building Performance Comparison, in: R. Chbeir, Y. Manolopoulos, I. Maglogianis, R. Alhajj (Eds.), *Artificial Intelligence Applications and Innovations*, no. 458 in IFIP Advances in Information and Communication Technology, Springer International Publishing, 2015, pp. 421–430, doi: 10.1007/978-3-319-23868-5\_30.  
URL [http://link.springer.com/chapter/10.1007/978-3-319-23868-5\\_30](http://link.springer.com/chapter/10.1007/978-3-319-23868-5_30)
- [10] J. Morais, Y. Pires, C. Cardoso, A. Klautau, An Overview of Data Mining Techniques Applied to Power Systems, *Data Mining and Knowledge Discovery in Real Life Applications*.  
URL [http://www.intechopen.com/source/pdfs/5962/InTech-An\\_overview\\_of\\_data\\_mining\\_techniques\\_applied\\_to\\_power\\_systems.pdf](http://www.intechopen.com/source/pdfs/5962/InTech-An_overview_of_data_mining_techniques_applied_to_power_systems.pdf)
- [11] G. Chicco, Overview and performance assessment of the clustering methods for electrical load pattern grouping, *Energy* 42 (1) (2012) 68–80. doi:10.1016/j.energy.2011.12.031.  
URL <http://dx.doi.org/10.1016/j.energy.2011.12.031>
- [12] K.-l. Zhou, S.-l. Yang, C. Shen, A review of electric load classification in smart grid environment, *Renewable and Sustainable Energy Reviews* 24 (C) (2013) 103–110. doi:10.1016/j.rser.2013.03.023.  
URL <http://dx.doi.org/10.1016/j.rser.2013.03.023>
- [13] G. James, D. Witten, T. Hastie, R. Tibshirani, *An introduction to statistical learning*, Vol. 112, Springer, 2013.  
URL <http://link.springer.com/content/pdf/10.1007/978-1-4614-7138-7.pdf>

- [14] T. Hastie, R. Tibshirani, J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, 2nd Edition, Springer series in statistics, Springer, New York, NY, 2009.  
URL <http://link.springer.com/article/10.1007%2FBF02985802>
- [15] B. Mirkin, Clustering: a data recovery approach, CRC Press, 2012.  
URL <https://books.google.com.sg/books?hl=en&lr=&id=34-q8geA3yEC&oi=fnd&pg=PP1&dq=clustering+a+data+recovery&ots=XEGQGoSOZY&sig=lq2rKFBq1bpGKQM8gdYS9xI8hrc>
- [16] C. Miller, A. Schlueter, Forensically discovering simulation feedback knowledge from a campus energy information system, in: Proceedings of the 2015 Symposium on Simulation for Architecture and Urban Design (SimAUD 2015), SCS, Washington DC, USA, 2015, pp. 33–40. doi:10.13140/RG.2.1.2286.0964.  
URL <http://dx.doi.org/10.13140/RG.2.1.2286.0964>
- [17] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM computing surveys (CSUR) 41 (3) (2009) 15.  
URL <http://dl.acm.org/citation.cfm?id=1541882>
- [18] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, S. Frankenstein, M. Lerner, GrammarViz 2.0: a tool for grammar-based pattern discovery in time series, Springer, 2014.  
URL [http://link.springer.com/chapter/10.1007/978-3-662-44845-8\\_37](http://link.springer.com/chapter/10.1007/978-3-662-44845-8_37)
- [19] C. Miller, Z. Nagy, A. Schlueter, Automated daily pattern filtering of measured building performance data, Automation in Construction 49, Part A (2015) 1–17. doi:10.1016/j.autcon.2014.09.004.  
URL <http://www.sciencedirect.com/science/article/pii/S0926580514002015>
- [20] D. Keim, G. Andrienko, J.-D. Fekete, C. Grg, J. Kohlhammer, G. Melanon, Visual Analytics: Definition, Process, and Challenges, in: A. Kerren, J. T. Stasko, J.-D. Fekete, C. North (Eds.), Information Visualization, no. 4950 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2008, pp. 154–175, dOI: 10.1007/978-3-540-70956-5\_7.  
URL [http://link.springer.com/chapter/10.1007/978-3-540-70956-5\\_7](http://link.springer.com/chapter/10.1007/978-3-540-70956-5_7)
- [21] B. Shneiderman, The eyes have it: a task by data type taxonomy for information visualizations, in: , IEEE Symposium on Visual Languages, 1996. Proceedings, 1996, pp. 336–343. doi:10.1109/VL.1996.545307.
- [22] R. O. Duda, P. E. Hart, D. G. Stork, Pattern classification, John Wiley & Sons, 2012.  
URL [https://books.google.com.sg/books?hl=en&lr=&id=Br33IRC3PkQC&oi=fnd&pg=PR3&dq=pattern+classification+duda&ots=2wCPIsgbDs&sig=Uq-PAZRxsyIu0Icp7zxDTyV\\_oU](https://books.google.com.sg/books?hl=en&lr=&id=Br33IRC3PkQC&oi=fnd&pg=PR3&dq=pattern+classification+duda&ots=2wCPIsgbDs&sig=Uq-PAZRxsyIu0Icp7zxDTyV_oU)
- [23] D. Jang, J. Eom, M. Jae Park, J. Jeung Rho, Variability of electricity load patterns and its effect on demand response: A critical peak pricing experiment on Korean commercial and industrial customers, Energy Policy 88 (2016) 11–26. doi:10.1016/j.enpol.2015.09.029.  
URL <http://www.sciencedirect.com/science/article/pii/S0301421515301154>
- [24] A. Shahzadeh, A. Khosravi, S. Nahavandi, Improving load forecast accuracy by clustering consumers using smart meter data, in: Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015, pp. 1–7. doi:10.1109/IJCNN.2015.7280393.
- [25] X. Liu, L. Golab, I. Ilyas, SMAS: A smart meter data analytics system, in: Proceedings of the 2015 IEEE 31st International Conference on Data Engineering (ICDE), Seoul, South Korea, 2015, pp. 1476–1479. doi:10.1109/ICDE.2015.7113405.
- [26] E. Pan, H. Li, L. Song, Z. Han, Kernel-based non-parametric clustering for load profiling of big smart meter data, in: Proceedings of the 2015 IEEE Wireless Communications and Networking Conference (WCNC), 2015, pp. 2251–2255. doi:10.1109/WCNC.2015.7127817.
- [27] I. Panapakidis, M. Alexiadis, G. Papagiannis, Evaluation of the performance of clustering algorithms for a high voltage industrial consumer, Engineering Applications of Artificial Intelligence 38 (C) (2015) 1–13. doi:10.1016/j.engappai.2014.10.013.  
URL <http://dx.doi.org/10.1016/j.engappai.2014.10.013>
- [28] M. Fagiani, S. Squartini, M. Severini, F. Piazza, A novelty detection approach to identify the occurrence of leakage in smart gas and water grids, in: Proceedings of the 2015 International Joint

- Conference on Neural Networks (IJCNN), Killarney, Ireland, 2015, pp. 1–8. doi:10.1109/IJCNN.2015.7280473.
- [29] A. Reinhardt, S. Koessler, PowerSAX: Fast motif matching in distributed power meter data using symbolic representations, in: Proceedings of 9th IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp 2014), IEEE, Edmonton, Canada, 2014, pp. 531–538. doi:10.1109/LCNW.2014.6927699.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6927699>
- [30] A. Lavin, D. Klabjan, Clustering time-series energy data from smart meters, *Energy Efficiency* 8 (4) (2014) 681–689. doi:10.1007/s12053-014-9316-0.  
URL <http://link.springer.com/article/10.1007/s12053-014-9316-0>
- [31] R. Green, I. Staffell, N. Vasilakos, Divide and Conquer K-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System, *IEEE Transactions on Engineering Management* 61 (2) (2014) 251–260. doi:10.1109/TEM.2013.2284386.
- [32] A. Jarrah Nezhad, T. K. Wijaya, M. Vasirani, K. Aberer, SmartD: Smart Meter Data Analytics Dashboard, in: Proceedings of the 5th International Conference on Future Energy Systems (e-Energy '14), e-Energy '14, ACM, New York, NY, USA, 2014, pp. 213–214. doi:10.1145/2602044.2602046.  
URL <http://doi.acm.org/10.1145/2602044.2602046>
- [33] H. K. akmak, H. Maa's s, F. Bach, U. G. Khnapfel, A new framework for the analysis of large scale multi-rate power data, Karlsruhe Institute of Technology, 2014.  
URL [http://www.researchgate.net/profile/Hueseyin\\_Cakmak2/publication/264579714\\_A\\_New\\_Framework\\_for\\_the\\_Analysis\\_of\\_Large\\_Scale\\_Multi-Rate\\_Power\\_Data/links/53e8b9820cf25d674ea8552d.pdf](http://www.researchgate.net/profile/Hueseyin_Cakmak2/publication/264579714_A_New_Framework_for_the_Analysis_of_Large_Scale_Multi-Rate_Power_Data/links/53e8b9820cf25d674ea8552d.pdf)
- [34] H. Shao, M. Marwah, N. Ramakrishnan, A Temporal Motif Mining Approach to Unsupervised Energy Disaggregation: Applications to Residential and Commercial Buildings, in: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI 2013), Vol. 250, Bellevue, WA, USA, 2013, p. 2250.  
URL <http://people.cs.vt.edu/~naren/papers/temp-motif-aaai13.pdf>
- [35] G. Chicco, O.-M. Ionel, R. Porumb, Electrical Load Pattern Grouping Based on Centroid Model With Ant Colony Clustering, *IEEE Transactions on Power Systems* 28 (2) (2013) 1706–1715.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6331025](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6331025)
- [36] S. Borgeson, Targeted Efficiency: Using Customer Meter Data to Improve Efficiency Program Outcomes, PhD, University of California, Berkeley, Berkeley, CA, USA (2013).  
URL [https://www.dropbox.com/s/ppi9az3z97mxx3u/Borgeson\\_dissertation.pdf](https://www.dropbox.com/s/ppi9az3z97mxx3u/Borgeson_dissertation.pdf)
- [37] F. Iglesias, W. Kastner, Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns, *Energies* 6 (2) (2013) 579–597. doi:10.3390/en6020579.  
URL <http://www.mdpi.com/1996-1073/6/2/579/>
- [38] S. Ramos, J. Duarte, J. Soares, Z. Vale, F. J. Duarte, Typical load profiles in the smart grid context: A clustering methods comparison, in: Proceedings of the Power and Energy Society General Meeting, 2012 IEEE, IEEE, San Diego, CA, USA, 2012, pp. 1–8.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6345565](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6345565)
- [39] A. R. Florita, L. J. Brackney, T. P. Otanicar, J. Robertson, Classification of Commercial Building Electrical Demand Profiles for Energy Storage Applications, in: Proceedings of ASME 2012 6th International Conference on Energy Sustainability & 10th Fuel Cell Science, Engineering and Technology Conference (ESFuelCell2012), San Diego, CA, USA, 2012.  
URL <http://solarenergyengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1697039>
- [40] D. De Silva, X. Yu, D. Alahakoon, G. Holmes, A Data Mining Framework for Electricity Consumption Analysis From Meter Data, *IEEE Transactions on Industrial Informatics* 7 (3) (2011) 399–407. doi:10.1109/TII.2011.2158844.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5893960>
- [41] S. M. Bidoki, N. Mahmoudi-Kohan, M. H. Sadreddini, M. Zolghadri Jahromi, M. P. Moghaddam,

- Evaluating different clustering techniques for electricity customer classification, in: Transmission and Distribution Conference and Exposition, 2010 IEEE PES, IEEE, New Orleans, LA, USA, 2010, pp. 1–5.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5484234](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5484234)
- [42] F. Gullo, G. Ponti, A. Tagarelli, M. Ruffolo, D. Labate, Low-voltage electricity customer profiling based on load data clustering, in: Proceedings of the 2009 International Database Engineering & Applications Symposium (IDEAS 2009), ACM, Calabria, Italy, 2009, pp. 330–333.  
URL <http://dl.acm.org/citation.cfm?id=1620472>
- [43] T. Rsnen, M. Kolehmainen, Feature-based clustering for electricity use time series data, in: Adaptive and Natural Computing Algorithms: 9th International Conference, ICANNGA 2009, Springer, Kuopio, Finland, 2009, pp. 401–412.  
URL [http://link.springer.com/chapter/10.1007/978-3-642-04921-7\\_41](http://link.springer.com/chapter/10.1007/978-3-642-04921-7_41)
- [44] G. Chicco, I. S. Ilie, Support vector clustering of electrical load pattern data, IEEE Transactions on Power Systems 24 (3) (2009) 1619–1628.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5072232](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5072232)
- [45] Z. A. Vale, C. Ramos, S. Ramos, T. Pinto, Data mining applications in power systems Case-studies and future trends, in: IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific, 2009, IEEE, Seoul, South Korea, 2009, pp. 1–4.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5356830](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5356830)
- [46] T. Rsnen, J. Ruuskanen, M. Kolehmainen, Reducing energy consumption by using self-organizing maps to create more personalized electricity use information, Applied Energy 85 (9) (2008) 830–840. doi:10.1016/j.apenergy.2007.10.012.  
URL <http://www.sciencedirect.com/science/article/pii/S030626190800007X>
- [47] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, F. J. G. Franco, Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps, IEEE Transactions on Power Systems 21 (4) (2006) 1672–1682. doi:10.1109/TPWRS.2006.881133.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1717570>
- [48] V. Figueiredo, F. Rodrigues, Z. Vale, J. B. Gouveia, An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques, IEEE Transactions on Power Systems 20 (2) (2005) 596–602. doi:10.1109/TPWRS.2005.846234.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1425550>
- [49] A. Schluter, P. Geyer, S. Cisar, Analysis of Georeferenced Building Data for the Identification and Evaluation of Thermal Microgrids, Proceedings of the IEEE 104 (4) (2016) 713–725. doi:10.1109/JPROC.2016.2526118.
- [50] P. Geyer, A. Schlter, S. Cisar, Application of clustering for the development of retrofit strategies for large building stocks, Advanced Engineering Informatics.  
URL <http://www.sciencedirect.com/science/article/pii/S1474034616300167>
- [51] I. Yarbrough, Q. Sun, D. C. Reeves, K. Hackman, R. Bennett, D. S. Henshel, Visualizing building energy demand for building peak energy analysis, Energy and Buildings 91 (2015) 10–15. doi:10.1016/j.enbuild.2014.11.052.  
URL <http://www.sciencedirect.com/science/article/pii/S0378778814010081>
- [52] B. Diong, G. Zheng, M. Ginn, Establishing the foundation for energy management on university campuses via data analytics, in: Proceedings of the IEEE SoutheastCon 2015, Fort Lauderdale, FL, USA, 2015, pp. 1–7. doi:10.1109/SECON.2015.7132996.
- [53] S. P. Pieri, IoannisTzouvadakis, M. Santamouris, Identifying energy consumption patterns in the Attica hotel sector using cluster analysis techniques with the aim of reducing hotels CO2 footprint, Energy and Buildings 94 (2015) 252–262. doi:10.1016/j.enbuild.2015.02.017.  
URL <http://www.sciencedirect.com/science/article/pii/S0378778815001127>
- [54] J. Ploennigs, B. Chen, P. Palmes, R. Lloyd, e2-Diagnoser: A System for Monitoring, Forecasting and Diagnosing Energy Usage, in: Data Mining Workshop (ICDMW), 2014 IEEE International Conference on, IEEE, Shenzhen, China, 2014, pp. 1231–1234. doi:10.1109/ICDMW.2014.56.

- URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7022741>
- [55] M. Georgescu, I. Mezic, Site-Level Energy Monitoring and Analysis using Koopman Operator Methods, in: Proceedings of the 2014 ASHRAE/IBPSA-USA Building Simulation Conference (SimBuild 2014), ASHRAE/IBPSA, Atlanta, GA, USA, 2014.  
URL [https://www.ashrae.net/File%20Library/docLib/Events/ASHRAE-IPBSA-USA/Presentations/32\\_Georgescu.pdf](https://www.ashrae.net/File%20Library/docLib/Events/ASHRAE-IPBSA-USA/Presentations/32_Georgescu.pdf)
- [56] M. Heidarinejad, M. Dahlhausen, S. McMahon, C. Pyke, J. Srebric, Cluster analysis of simulated energy use for LEED certified U.S. office buildings, *Energy and Buildings* 85 (2014) 86–97. doi:10.1016/j.enbuild.2014.09.017.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0378778814007087>
- [57] A. Morn, J. J. Fuertes, M. Domnguez, M. A. Prada, S. Alonso, P. Barrientos, Analysis of electricity bills using visual continuous maps, *Neural Computing and Applications* 23 (3-4) (2013) 645–655. doi:10.1007/s00521-013-1409-8.  
URL <http://link.springer.com/10.1007/s00521-013-1409-8>
- [58] D. F. M. Cabrera, H. Zareipour, Data association mining for identifying lighting energy waste patterns in educational institutes, *Energy and Buildings* 62 (2013) 210–216. doi:10.1016/j.enbuild.2013.02.049.  
URL <http://dx.doi.org/10.1016/j.enbuild.2013.02.049>
- [59] T. G. Nikolaou, D. S. Kolokotsa, G. S. Stavrakakis, I. D. Skias, On the Application of Clustering Techniques for Office Buildings’ Energy and Thermal Comfort Classification, *IEEE Transactions on Smart Grid* 3 (4) (2012) 2196–2210. doi:10.1109/TSG.2012.2215059.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6387350>
- [60] S. Petcharat, S. Chungpaibulpatana, P. Rakkwamsuk, Assessment of potential energy saving using cluster analysis: A case study of lighting systems in buildings, *Energy and Buildings* 52 (2012) 145–152. doi:10.1016/j.enbuild.2012.06.006.  
URL <http://dx.doi.org/10.1016/j.enbuild.2012.06.006>
- [61] L. An, R. Horesh, Y. T. Chae, Y. M. Lee, Estimation of Thermal Parameters of Buildings Through Inverse Modeling and Clustering for a Portfolio of Buildings, in: Proceedings of the Fifth National Conference of IBPSA-USA (SimBuild 2012), Madison, WI, USA, 2012.  
URL [http://www.ibpsa.us/simbuild2012/Papers/SB12\\_TS05a\\_4\\_An.pdf](http://www.ibpsa.us/simbuild2012/Papers/SB12_TS05a_4_An.pdf)
- [62] G. Bellala, M. Marwah, M. Arlitt, G. Lyon, C. E. Bash, Towards an understanding of campus-scale power consumption, in: Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys 2011), ACM Request Permissions, Seattle, WA, USA, 2011, p. 73. doi:10.1145/2434020.2434043.  
URL <http://dl.acm.org/citation.cfm?doid=2434020.2434043>
- [63] D. Lehrer, J. Vasudev, Visualizing Energy Information in Commercial Buildings: A Study of Tools, Expert Users, and Building Occupants, Tech. rep., Center for the Built Environment - UC Berkeley (Nov. 2011).  
URL <https://escholarship.org/uc/item/6vp5m5m3.pdf>
- [64] J. Granderson, M. A. Piette, G. Ghatikar, Building energy information systems: user case studies, *Energy Efficiency* 4 (1) (2010) 17–30. doi:10.1007/s12053-010-9084-4.  
URL <http://link.springer.com/article/10.1007/s12053-010-9084-4>
- [65] N. Gaitani, C. Lehmann, M. Santamouris, G. Mihalakakou, P. Patargias, Using principal component and cluster analysis in the heating evaluation of the school building sector, *Applied Energy* 87 (6) (2010) 2079–2086. doi:10.1016/j.apenergy.2009.12.007.  
URL <http://www.sciencedirect.com/science/article/pii/S0306261909005303>
- [66] J. Sedano, J. R. Villar, L. Curiel, E. De La Cal, E. Corchado, Improving energy efficiency in buildings using machine intelligence, in: *Intelligent Data Engineering and Automated Learning-IDEAL 2009*, Springer, 2009, pp. 773–782.  
URL [http://link.springer.com/chapter/10.1007/978-3-642-04394-9\\_95](http://link.springer.com/chapter/10.1007/978-3-642-04394-9_95)
- [67] D. Lehrer, Research Scoping Report: Visualizing Information in Commercial Buildings, Tech. rep.,

- Center for the Built Environment - UC Berkeley (Sep. 2009).
- [68] Y. Agarwal, T. Weng, R. K. Gupta, The energy dashboard: improving the visibility of energy consumption at a campus-wide scale, in: *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys 2009)*, ACM Press, Berkeley, CA, USA, 2009, p. 55. doi:10.1145/1810279.1810292.  
URL <http://portal.acm.org/citation.cfm?doid=1810279.1810292>
- [69] J. C. Lam, K. K. W. Wan, K. L. Cheung, L. Yang, Principal component analysis of electricity use in office buildings, *Energy and Buildings* 40 (5) (2008) 828–836. doi:10.1016/j.enbuild.2007.06.001.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0378778807001764>
- [70] M. Santamouris, G. Mihalakakou, P. Patargias, N. Gaitani, K. Sfakianaki, M. Papaglastra, C. Pavlou, P. Doukas, E. Primikiri, V. Geros, M. N. Assimakopoulos, R. Mitoula, S. Zerefos, Using intelligent clustering techniques to classify the energy performance of school buildings, *Energy and Buildings* 39 (1) (2007) 45–51. doi:10.1016/j.enbuild.2006.04.018.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0378778806001502>
- [71] J. E. Seem, Pattern recognition algorithm for determining days of the week with similar energy consumption profiles, *Energy and Buildings* 37 (2) (2005) 127–139. doi:10.1016/j.enbuild.2004.04.004.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0378778804001434>
- [72] K.-U. Ahn, C.-S. Park, Correlation between occupants and energy consumption, *Energy and Buildings* 116 (2016) 420–433. doi:10.1016/j.enbuild.2016.01.010.  
URL <http://www.sciencedirect.com/science/article/pii/S037877881630010X>
- [73] V. Mansur, P. Carreira, A. Arsenio, A Learning Approach for Energy Efficiency Optimization by Occupancy Detection, in: *Internet of Things. User-Centric IoT*, Vol. 150, Springer International Publishing, Cham, 2015, pp. 9–15.  
URL [http://link.springer.com/10.1007/978-3-319-19656-5\\_2](http://link.springer.com/10.1007/978-3-319-19656-5_2)
- [74] A. A. Adamopoulou, A. M. Tryferidis, D. K. Tzovaras, A context-aware method for building occupancy prediction, *Energy and Buildings* doi:10.1016/j.enbuild.2015.10.003.  
URL <http://www.sciencedirect.com/science/article/pii/S0378778815303212>
- [75] C. Fan, F. Xiao, H. Madsen, D. Wang, Temporal Knowledge Discovery in Big BAS Data for Building Energy Management, *Energy and Buildings* doi:10.1016/j.enbuild.2015.09.060.  
URL <http://www.sciencedirect.com/science/article/pii/S0378778815302991>
- [76] N. Gayeski, S. Kleindienst, J. Gagne, B. Werntz, R. Cruz, S. Samouhos, Data and Interfaces for Advanced Building Operations and Maintenance - RP 1633 Final Report, Tech. rep., ASHRAE (Jun. 2015).
- [77] S. DOca, T. Hong, Occupancy schedules learning process through a data mining framework, *Energy and Buildings* 88 (2015) 395–408. doi:10.1016/j.enbuild.2014.11.065.  
URL <http://www.sciencedirect.com/science/article/pii/S0378778814010329>
- [78] U. Habib, G. Zucker, Finding the Different Patterns in Buildings Data Using Bag of Words Representation with Clustering, in: *2015 13th International Conference on Frontiers of Information Technology (FIT)*, 2015, pp. 303–308. doi:10.1109/FIT.2015.60.
- [79] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, *Energy and Buildings* 75 (2014) 109–118. doi:10.1016/j.enbuild.2014.02.005.  
URL <http://dx.doi.org/10.1016/j.enbuild.2014.02.005>
- [80] A. Domahidi, F. Ullmann, M. Morari, C. N. Jones, Learning decision rules for energy efficient building control, *Journal of Process Control* 24 (6) (2014) 763–772. doi:10.1016/j.jprocont.2014.01.006.  
URL <http://www.sciencedirect.com/science/article/pii/S0959152414000304>
- [81] B. Lange, N. Rodriguez, W. Puech, X. Vasques, Discovering unexpected information using a building energy visualization tool, in: A. M. Baskurt, R. Sitnik (Eds.), *Proceedings of the IS&T/SPIE 2013 Conference on Electronic Imaging*, Vol. 8650, SPIE, San Francisco, CA, USA, 2013, p. 86500Q. doi:10.1117/12.2000273.  
URL <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.>

- 2000273
- [82] A. C. Bogen, M. Rashid, E. W. East, J. Ross, Evaluating a data clustering approach for life-cycle facility control, *Journal of Information Technology in Construction* Vol. 18 (2013) 99–118.  
URL <http://www.itcon.org/2013/6>
  - [83] Z. J. Yu, B. C. M. Fung, F. Haghighat, Extracting knowledge from building-related data A data mining framework, *Building Simulation* 6 (2) (2013) 207–222. doi:10.1007/s12273-013-0117-8.  
URL <http://link.springer.com/10.1007/s12273-013-0117-8>
  - [84] P. T. May-Ostendorp, G. P. Henze, B. Rajagopalan, C. D. Corbin, Extraction of supervisory building control rules from model predictive control of windows in a mixed mode building, *Journal of Building Performance Simulation* 6 (3) (2013) 199–219. doi:10.1080/19401493.2012.665481.  
URL <http://www.tandfonline.com/doi/abs/10.1080/19401493.2012.665481>
  - [85] C. Fan, F. Xiao, W. Shengwei, Prediction of Chiller Power Consumption Using Time Series Analysis and Artificial Neural Networks, in: *Proceedings of the 8th International Conference on Indoor Air Quality, Ventilation and Energy Conservation in Buildings (CLIMA 2013)*, Prague, Czech Republic, 2013.
  - [86] D. Hong, J. Ortiz, K. Whitehouse, D. Culler, Towards Automatic Spatial Verification of Sensor Placement in Buildings, in: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings (BuildSys 2013)*, Roma, Italy, 2013.  
URL <http://www.cs.berkeley.edu/~jortiz/papers/buildsys13.pdf>
  - [87] B. Lange, N. Rodriguez, W. Puech, Energy Consumption Improvement Through a Visualization Software, in: *Energy Efficiency - A Bridge to Low Carbon Economy*, Intech, 2012, pp. 161–184.
  - [88] L. V. Thanayankizil, S. K. Ghai, D. Chakraborty, D. P. Seetharam, Softgreen: Towards energy management of green office buildings with soft sensors, in: *Proceedings of the Fourth International Conference on Communication Systems and Networks (COMSNETS 2012)*, IEEE, Bangalore, India, 2012, pp. 1–6. doi:10.1109/COMSNETS.2012.6151374.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6151374>
  - [89] P. May-Ostendorp, G. P. Henze, C. D. Corbin, B. Rajagopalan, C. Felsmann, Model-predictive control of mixed-mode buildings with rule extraction, *Building and Environment* 46 (2) (2011) 428–437. doi:10.1016/j.buildenv.2010.08.004.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0360132310002477>
  - [90] C. Duarte, B. Acker, R. Grosshans, M. Manic, K. Van Den Wymelenberg, C. Rieger, Prioritizing and Visualizing Energy Management and Control System Data to Provide Actionable Information For Building Operators, in: *Proceedings from 2011 Western Energy Policy Research Conference*, Boise, Idaho, USA, 2011.  
URL [http://mhrq.if.uidaho.edu/papers/2011/WEPC11\\_DuarteAckerKevin\\_EMSDataVisualisation.pdf](http://mhrq.if.uidaho.edu/papers/2011/WEPC11_DuarteAckerKevin_EMSDataVisualisation.pdf)
  - [91] A. Augello, M. Ortolani, G. L. Re, S. Gaglio, Sensor mining for user behavior profiling in intelligent environments, *Advances in Distributed Agent-Based Retrieval Tools* (2011) 143–158.  
URL [http://link.springer.com/chapter/10.1007/978-3-642-21384-7\\_10](http://link.springer.com/chapter/10.1007/978-3-642-21384-7_10)
  - [92] B. Dong, K. P. Lam, Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network, *Journal of Building Performance Simulation* 4 (4) (2011) 359–369. doi:10.1080/19401493.2011.577810.  
URL <http://www.tandfonline.com/doi/abs/10.1080/19401493.2011.577810>
  - [93] M. Hao, M. Marwah, H. Janetzko, R. Sharma, D. A. Keim, U. Dayal, D. Patnaik, N. Ramakrishnan, Visualizing frequent patterns in large multivariate time series, in: *Proceedings of SPIE Conference on Visualization and Data Analysis 2011*, Vol. 7868, SPIE, San Francisco, CA, USA, 2011, pp. 78680J–78680J–10. doi:10.1117/12.872169.  
URL <http://dx.doi.org/10.1117/12.872169>
  - [94] D. Patnaik, M. Marwah, R. K. Sharma, N. Ramakrishnan, Data Mining for Modeling Chiller Systems in Data Centers, in: *Advances in Intelligent Data Analysis IX*, Vol. 6065, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 125–136.

- URL [http://www.springerlink.com/index/10.1007/978-3-642-13062-5\\_13](http://www.springerlink.com/index/10.1007/978-3-642-13062-5_13)
- [95] D. Patnaik, M. Marwah, R. Sharma, N. Ramakrishnan, Sustainable Operation and Management of Data Center Chillers Using Temporal Data Mining, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009), KDD '09, ACM, New York, NY, USA, 2009, pp. 1305–1314. doi:10.1145/1557019.1557159.  
URL <http://doi.acm.org/10.1145/1557019.1557159>
  - [96] A. Kusiak, Z. Song, Clustering-Based Performance Optimization of the Boiler Turbine System, IEEE Transactions on Energy Conversion 23 (2) (2008) 651–658. doi:10.1109/TEC.2007.914183.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4456516>
  - [97] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Automation in Construction 50 (2015) 81–90. doi:10.1016/j.autcon.2014.12.006.  
URL <http://www.sciencedirect.com/science/article/pii/S0926580514002507>
  - [98] A. Capozzoli, F. Lauro, I. Khan, Fault detection analysis using data mining techniques for a cluster of smart office buildings, Expert Systems with Applications 42 (9) (2015) 4324–4338. doi:10.1016/j.eswa.2015.01.010.  
URL <http://www.sciencedirect.com/science/article/pii/S0957417415000251>
  - [99] Y. Sun, T.-Y. Wu, G. Zhao, M. Guizani, Efficient Rule Engine for Smart Building Systems, IEEE Transactions on Computers 64 (6) (2015) 1658–1669. doi:10.1109/TC.2014.2345385.
  - [100] S. Li, J. Wen, A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform, Energy and Buildings 68 (2014) 63–71. doi:10.1016/j.enbuild.2013.08.044.  
URL <http://dx.doi.org/10.1016/j.enbuild.2013.08.044>
  - [101] B. Chen, M. Sinn, J. Ploennigs, A. Schumann, Statistical Anomaly Detection in Mean and Variation of Energy Consumption, in: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR), IEEE, Stockholm, Sweden, 2014, pp. 3570–3575. doi:10.1109/ICPR.2014.614.  
URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6977326>
  - [102] J.-S. Chou, A. S. Telaga, Real-time detection of anomalous power consumption, Renewable and Sustainable Energy Reviews 33 (C) (2014) 400–411. doi:10.1016/j.rser.2014.01.088.  
URL <http://dx.doi.org/10.1016/j.rser.2014.01.088>
  - [103] D. Wijayasekara, O. Linda, M. Manic, C. Rieger, Mining Building Energy Management System Data Using Fuzzy Anomaly Detection and Linguistic Descriptions, IEEE Transactions on Industrial Informatics 10 (3) (2014) 1829–1840. doi:10.1109/TII.2014.2328291.
  - [104] M. Le Cam, R. Zmeureanu, A. Daoud, Application of Data Mining techniques for energy modeling of HVAC sub-systems, in: Proceedings of eSim 2014 - IBPSA-Canada Conference, IBPSA, Ottawa, Canada, 2014.  
URL <http://www.ibpsa.org/proceedings/eSimPapers/2014/7A.3.pdf>
  - [105] J. Ploennigs, B. Chen, A. Schumann, N. Brady, Exploiting Generalized Additive Models for Diagnosing Abnormal Energy Use in Buildings, in: Proceedings of the 5th ACM Workshop On Embedded Systems For Energy-Efficient Buildings (Buildsys '13), Roma, Italy, 2013.  
URL <http://dl.acm.org/citation.cfm?id=2528291>
  - [106] H. Janetzko, F. Stoffel, S. Mittelstdt, D. A. Keim, Anomaly detection for visual analytics of power consumption data, Computers & Graphics 38 (2013) 27–37. doi:10.1016/j.cag.2013.10.006.  
URL <http://dx.doi.org/10.1016/j.cag.2013.10.006>
  - [107] R. Fontugne, N. Tremblay, P. Borgnat, P. Flandrin, H. Esaki, Mining Anomalous Electricity Consumption Using Ensemble Empirical Mode Decomposition, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, Vancouver, BC, Canada, 2013, pp. 5238–5242.  
URL [http://hiroshi1.hongo.wide.ad.jp/hiroshi/papers/2013/romain\\_ICASSP2013.pdf](http://hiroshi1.hongo.wide.ad.jp/hiroshi/papers/2013/romain_ICASSP2013.pdf)
  - [108] R. Fontugne, J. Ortiz, N. Tremblay, P. Borgnat, P. Flandrin, K. Fukuda, D. Culler, H. Esaki, Strip, bind, and search: a method for identifying abnormal energy consumption in buildings, in: Proceedings of the 12th ACM/IEEE Int Conference on Information Processing in Sensor Network (ISPN 2013), ACM, Philadelphia, PA, USA, 2013, pp. 129–140.



- URL <http://dl.acm.org/citation.cfm?id=2461399>
- [109] O. Linda, D. Wijayasekara, M. Manic, C. Rieger, Computational intelligence based anomaly detection for Building Energy Management Systems, in: Proceedings of the 5th International Symposium on Resilient Control Systems (ISRCS 2012), Salt Lake City, UT, USA, 2012, pp. 77–82. doi:10.1109/ISRCS.2012.6309297.  
URL [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6309297&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch\\_All%26queryText%3D%28p\\_Title%3A%22Computational+Intelligence+based+Anomaly+Detection+for+Building+Energy+Management+Systems%22%29](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6309297&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22Computational+Intelligence+based+Anomaly+Detection+for+Building+Energy+Management+Systems%22%29)
- [110] Y. Zhu, X. Jin, Z. Du, Fault diagnosis for sensors in air handling unit based on neural network pre-processed by wavelet and fractal, Energy and Buildings 44 (2012) 7–16. doi:10.1016/j.enbuild.2011.09.043.  
URL <http://dx.doi.org/10.1016/j.enbuild.2011.09.043>
- [111] M. Wrinch, T. H. El-Fouly, S. Wong, Anomaly detection of building systems using energy demand frequency domain analysis, in: Power and Energy Society General Meeting, 2012 IEEE, IEEE, San Diego, CA, USA, 2012, pp. 1–6.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6344790](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6344790)
- [112] Z. J. Yu, F. Haghighat, B. C. M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data, Energy and Buildings 47 (2012) 430–440. doi:10.1016/j.enbuild.2011.12.018.  
URL <http://dx.doi.org/10.1016/j.enbuild.2011.12.018>
- [113] D. Liu, Q. Chen, K. Mori, Y. Kida, A method for detecting abnormal electricity energy consumption in buildings, Journal of Computational Information Systems 14 (2010) 4887–4895.  
URL [http://www.jofcis.com/publishedpapers/2010\\_6\\_14\\_4887\\_4895.pdf](http://www.jofcis.com/publishedpapers/2010_6_14_4887_4895.pdf)
- [114] D. Jacob, S. Dietz, S. Komhard, C. Neumann, S. Herkel, Black-box models for fault detection and performance monitoring of buildings, Journal of Building Performance Simulation 3 (1) (2010) 53–62. doi:10.1080/19401490903414454.  
URL <http://www.tandfonline.com/doi/abs/10.1080/19401490903414454>
- [115] S. Wang, Q. Zhou, F. Xiao, A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults, Energy and Buildings 42 (4) (2010) 477–490. doi:10.1016/j.enbuild.2009.10.017.  
URL <http://dx.doi.org/10.1016/j.enbuild.2009.10.017>
- [116] C. Forlines, K. Wittenburg, Wakame: Sense Making of Multi-dimensional Spatial-temporal Data, in: Proceedings of the International Conference on Advanced Visual Interfaces (AVI 2010), AVI '10, ACM, Roma, Italy, 2010, pp. 33–40. doi:10.1145/1842993.1843000.  
URL <http://doi.acm.org/10.1145/1842993.1843000>
- [117] K. Yoshida, M. Inui, T. Yairi, K. Machida, M. Shioya, Y. Masukawa, Identification of Causal Variables for Building Energy Fault Detection by Semi-supervised LDA and Decision Boundary Analysis, in: Data Mining Workshops, 2008. ICDMW '08. IEEE International Conference on, Pisa, Italy, 2008, pp. 164–173. doi:10.1109/ICDM.W.o2r0k0s8h.o4p4s.2008.46.  
URL [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4733934](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4733934)
- [118] J. E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy and Buildings 39 (1) (2006) 52–58. doi:10.1016/j.enbuild.2006.03.033.  
URL <http://linkinghub.elsevier.com/retrieve/pii/S0378778806001514>
- [119] S. Wang, J. Cui, Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method, Applied Energy.  
URL <http://www.sciencedirect.com/science/article/pii/S0306261904001953>
- [120] Journals unite for reproducibility, Nature 515 (7525) (2014) 7–7. doi:10.1038/515007a.  
URL <http://www.nature.com/doifinder/10.1038/515007a>
- [121] Code share, Nature 514 (7524) (2014) 536–536. doi:10.1038/514536a.  
URL <http://www.nature.com/doifinder/10.1038/514536a>

- 1  
2  
3  
4 [122] E. Keogh, S. Kasetty, On the need for time series data mining benchmarks: A survey and empirical  
5 demonstration, *Data Mining and Knowledge Discovery* 7 (4) (2003) 349–371.  
6 URL <http://link.springer.com/article/10.1023/A:1024988512476>  
7 [123] J. F. Kreider, J. S. Haberl, Predicting hourly building energy use: The great energy predictor  
8 shootout Overview and discussion of results, Tech. rep., American Society of Heating, Refrigerating  
9 and Air-Conditioning Engineers, Inc., Atlanta, GA (United States) (1994).  
10 URL <http://www.osti.gov/scitech/biblio/33315>  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65