

Data Analytics for Managing Power in Commercial Buildings

GOWTHAM BELLALA, Hewlett Packard Labs
 MANISH MARWAH, Hewlett Packard Labs
 MARTIN ARLITT, Hewlett Packard Labs
 GEOFF LYON, Hewlett Packard Labs
 CULLEN BASH, Hewlett Packard Labs
 AMIP SHAH, Hewlett Packard Labs

Commercial buildings are significant consumers of electricity. The first step towards better energy management in commercial buildings is monitoring consumption. However, instrumenting every electrical panel in a large commercial building is expensive and unnecessary. In this paper, we propose a greedy meter (sensor) placement algorithm based on maximization of information gain subject to a cost constraint. The algorithm provides a near-optimal solution guarantee, and our empirical results demonstrate a 15% improvement in prediction power over conventional methods. Furthermore, to identify power saving opportunities, we use an unsupervised anomaly detection technique based on a low-dimensional embedding. Further, to enable a building manager to effectively plan for demand response programs, we evaluate several solutions for fine-grained, short-term load forecasting. Our investigation reveals an interesting relation between time-series variability and the optimal forecast method, with support vector regression and an Ensemble model being the best method for time-series with high variability, while a simple method such as a linear regression is equally effective on time-series with low variability. Finally, to better manage resources such as lighting and HVAC, we propose a semi-supervised approach combining hidden Markov models (HMM) and a standard classifier to model occupancy based on readily available port-level network statistics. We show that the proposed two step approach simplifies the occupancy model while achieving good accuracy. The experimental results demonstrate an average occupancy estimation error of 9.3% with a potential reduction of 9.5% in lighting load using our occupancy models.

CCS Concepts: •Information systems → Information systems applications; Data mining; •Social and professional topics → Sustainability; •Theory of computation → Theory and algorithms for application domains;

Additional Key Words and Phrases: Commercial buildings, power management, meter placement, anomaly detection, short term load forecast, occupancy modeling

ACM Reference Format:

Gowtham Bellala, Manish Marwah, Martin Arlitt, Geoff Lyon, Cullen Bash, and Amip Shah, 2016. Data Analytics for Managing Power in Commercial Buildings. *ACM Trans. Cyber-Phys. Syst.* 0, 0, Article 0 (0), 25 pages.
 DOI: 0000001.0000001

1. INTRODUCTION

In the United States alone, there are an estimated five million commercial buildings. In 2015, these building consumed roughly one third of the electricity generated in the country. The energy costs for commercial buildings exceeds \$100 billion annually. Due to recent economic turmoil, and increased awareness of environmental concerns (e.g., global climate change), many companies want to understand how resources such as electricity are used in their buildings, so that steps can be taken to reduce consump-

Author's addresses: Hewlett Packard Labs, 1501 Page Mill Road, Palo Alto, CA 94304.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 0 ACM. 2378-962X/0/-ART0 \$15.00

DOI: 0000001.0000001

tion. Often, they turn to consulting firms for services like building energy efficiency analyses. The work described in this paper addresses shortcomings of existing analyses of this sort. Preliminary versions of this work appeared in [Bellala et al. 2011; 2012]. In this paper, we combine and extend the results of our earlier work.

A first challenge for improving power use in a building is understanding how much power each appliance or device in the building uses. One option is to install power meters on all electrical panels, to collect usage data near the consumers, and get a (approximate) per-appliance breakdown of power use. The main disadvantage of this approach is the cost to buy and install the meters. For companies that own a lot of buildings (e.g., Walmart has more than 10,000 stores globally), this cost becomes prohibitively expensive. Thus, one research question we address is where to place a limited set of meters in a building, while minimizing the information loss. We propose an efficient greedy algorithm that provides a near-optimal solution.

A second challenge we investigate is how to systematically monitor building energy use and automatically detect problems that arise over time. A limitation of manual consulting services is that they can only identify issues that are occurring at the time the analysis is conducted, and typically only a limited number of panels are monitored, identified by an expert based on the likelihood of energy savings. Having a consultant repeat the study on a regular (e.g., daily) basis is not cost effective, so an automated technique is highly desired. We present results from applying our unsupervised anomaly (fault) detection and ranking methods for monitoring tens of meters over a six month period.

A third challenge relates to providing fine-grained, short-term consumption forecasts. Accurate short-term forecasts enable the implementation of demand response programs in concert with a utility provider. Utilities are interested in lowering peak power usage, since that lowers their infrastructure costs (typically based on the peak load). Utilities, through demand response programs, provide incentives to customers for reducing their consumption during peak load periods (as specified by the utility). Knowledge of expected power consumption in different parts of a campus or building allows a building operations manager to effectively plan for such an event. We evaluate several methods for fine-grained short-term load forecasting, and discover that support vector regression and an ensemble model work best overall, while simple methods such as linear regression and gradient boosted machines also tend to be effective on time-series with low variability.

Lastly, consulting studies will typically recommend static solutions to reduce building energy use. For example, turn on all lights only during work hours (e.g., 8am to 6pm), and turn most off otherwise. While such techniques do help reduce power use in a building, further savings are possible. One approach is to only turn on lights (or HVAC systems) in areas where people are currently in, and to turn them off when the people leave. To facilitate such dynamic resource management, we developed a semi-supervised method for occupancy modeling.

Our group has instrumented three large commercial buildings on the Hewlett Packard Labs campus in Palo Alto, CA to enable an understanding of how, when and where power is consumed in a commercial campus. This instrumentation provides extensive data on the campus power use, which establishes the “ground truth” against which we evaluate our power management methods.

The paper makes the following contributions.

- we propose a greedy algorithm for determining meter placement in a building’s electrical infrastructure, to maximize mutual information while minimizing the cost of metering. Besides being computationally efficient, we also show that the proposed greedy algorithm guarantees a near-optimal solution. In particular, we show that

- mutual information becomes submodular under a special graphical structure that arises in distribution networks such as power, water and gas.
- we propose an unsupervised technique to identify anomalous usage periods in power consumption time series. From over six months of data from a large test bed (three buildings totaling 300,000 sq. ft.), we identify several power saving opportunities. For some of the observed anomalies, potential energy savings ranged from 22% to 28% per anomaly.
 - we evaluate several models for short-term power consumption forecast on 12 months of data and several power meters. Our investigation reveals an interesting relation between time-series variability and optimal forecast method, with support vector regression and an Ensemble model being the best method for time-series with high variability, while a simple method such as linear regression is equally effective for time-series with low variability.
 - we propose a novel semi-supervised approach combining a hidden Markov model (HMM) with a classifier to model human occupancy levels in our test bed. Such models can be used for more efficient management policies for lighting and HVAC. Based on our occupancy models, we estimate that modifying the lighting schedule can save an additional 9.5% of lighting power beyond the static schedule.

The remainder of the paper is organized as follows. Section 2 highlights related work. Section 3 provides an overview of our campus, the power delivery and measurement infrastructure, and the campus power usage characteristics. Section 4 introduces our overall framework. Sections 4.1, 4.2, 4.3 and 4.4 describe our meter placement, anomaly detection, short-term load forecasting and occupancy modeling methods, respectively. Section 5 evaluates our methods. Section 6 summarizes our work and future directions.

2. PRIOR AND RELATED WORK

A recent study by the United States Energy Information Administration [USEIA 2015] estimated commercial buildings to consume roughly one third of the total electricity generated in the US. These estimates motivated research to improve building energy efficiency. A large body of recent work addressed different aspects of building energy management in commercial buildings such as diagnosis and anomaly detection [Wang et al. 2016; Teraoko et al. 2014], building energy modeling [Yin et al. 2015; Masuda and Claridge 2014], control of HVAC and lighting [Brooks et al. 2015; Haq et al. 2014], and building occupancy modeling [Balaji et al. 2013; Ghai et al. 2012]. Our work approaches these problems from a data mining perspective, and proposes novel methods in some of these areas.

Meter Placement: Obtaining detailed, per device power usage information tends to be intrusive and quite expensive, which includes both the cost of the meter and its installation (often a significant part). This raises an interesting question as to where and how many power meters must be installed. This problem of selecting optimal locations for meter placement can be framed as a budgeted optimization problem. It has been well studied in the literature in different contexts, such as in the case of sensor placement in a water distribution network [Leskovec et al. 2007] or observation selection in an autonomous robotic exploration [Krause and Guestrin 2007]. Several different criteria have been proposed in the literature for selecting these optimal locations [Krause and Guestrin 2007]. The most popular among them being mutual information [Krause et al. 2006; Krause et al. 2008]. However, as we discuss in Section 4.1, this constrained optimization problem with mutual information as an objective function is known to be NP-hard, and hence requires use of greedy, near-optimal strategies to solve the optimization problem.

Anomaly Detection: Examining data for anomalies is a known approach for identifying abnormal system behavior. Catterson et al. [2010] used this approach to monitor old power transformers. Their goal was to proactively search for abnormal behavior that may indicate the transformer is about to fail. Teraoko et al. [2014] propose a web service based fault management framework for HVAC systems and uses building management system data to detect anomalies. Li et al. [2010] looked for anomalies in building power consumption, by employing simple statistical tests such as the Q-test to detect time points with abnormal power usage. On the other hand, Jakkula and Cook [2010] demonstrate that a clustering based approach can be more efficient in identifying abnormal activities in power consumption data, over such simple statistical tests. Our work on anomaly detection is built on these observations.

Short-term Load Forecasting: Load forecasting has been studied in the literature for a long time. Many traditional approaches such as ARMA, ARIMA and linear regression that were proposed in the statistics community have been applied for short-term load forecast at an aggregate scale such as a Utility or regional level [Huang and Shih 2003]. However, these methods are not powerful enough to forecast load at finer spatial scales where the consumption patterns tend to be highly irregular.

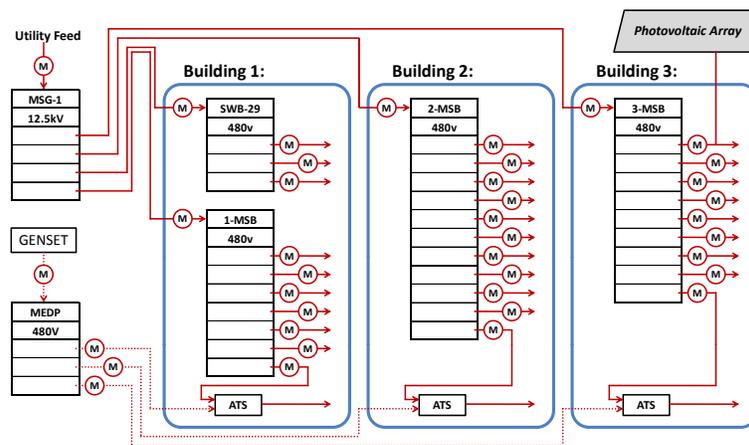
To address this limitation, more recent work on short-term forecasting used machine learning algorithms such as support vector machines [Pai and Hong 2005; Sapankevych and Sankar 2009], multi-layer perceptron [Hippert et al. 2001] and neural networks [An et al. 2013], and incorporated external factors such as weather to demonstrate improved forecast accuracy over simple regression models. At the same time, other recent studies in the context of smart grid to obtain household level short-term load forecast have shown a fundamental limitation to the predictability of individual customers [Edwards et al. 2012; Sevlian and Rajagopal 2013; 2014]. In this paper, we study fine-grained and short-term load forecasting in commercial buildings.

Occupancy Modeling: Occupancy information can be used directly by building control systems to reduce the energy consumption of air conditioning, lighting, IT infrastructure, and other building systems [CEC 1993]. Erickson and Cerpa [2010] show that an occupancy based HVAC control can achieve significant reductions in overall energy consumption while conforming to the American Society of Heating, Refrigerating and Air-conditioning [ASHRAE 2004; 2007] comfort standards. Unfortunately, tracking the number of people in a building is often more difficult than one might expect. While data from badging or existing security cameras may help estimate the total number of occupants in a commercial building, they do not provide fine-grained (cube/zone-level) occupancy estimates required to enable fine-grained control.

Most previous work on occupancy modeling used specific sensors such as CO₂ sensors [Newsham and Birt 2010], PIR sensors [Agarwal et al. 2011], or a host of camera sensor nodes [Erickson and Cerpa 2010] to estimate the number of occupants. However, these approaches require installation and maintenance of additional infrastructure, and is often not a viable solution, especially in large commercial buildings. An alternative approach is to use proxies that can provide reasonable estimates of the occupancy information. In this context, Melfi et al. [2011] proposed the use of existing network infrastructure to estimate occupancy. They studied the use of Dynamic Host Control Protocol (DHCP) logs and other explicit ways such as monitoring PC activity to estimate occupancy. Balaji et al. [2013] propose a less intrusive approach that leverage WiFi connectivity logs to determine occupancy. However, their zone of detection is often large covering up to 10 HVAC zones. In this work, we use a similar less intrusive approach that uses wired network connections to provide fine-grained (cube-level) occupancy information.



(a)



(b)

Fig. 1. (a) Aerial view of the campus, showing the six main buildings, three of which were instrumented for resource monitoring (b) Power distribution topology, the top-level distribution panels within each building and the power meter placements.

3. CAMPUS OVERVIEW

We use our campus as a test bed to investigate the monitoring and management of resources such as power, gas, water, and waste, with our primary focus on power. The campus contains six main buildings with a total footprint of 700,000 sq. ft. Our efforts are focused on three two-storey buildings, as highlighted in Figure 1(a). These three buildings have a footprint of 300,000 sq. ft, and host about 500 occupants.

3.1. Power Distribution Topology

Figure 1(b) shows the power distribution topology for Buildings 1, 2 and 3, which are all fed by a single utility feed (3-phase 12.5kV). An emergency generator (3-phase 480V) maintains critical loads in the event of a utility failure. Automatic transfer switches (ATS) are used to revert from the utility to backup power. Each building has a main distribution panel (3-phase 480V) that branches to about 10 major sub-loads or sub-panels. Building 1 also has a secondary top-level power feed and distribution panel. In addition, Building 3 has a 135kW photo-voltaic array on top of it, that offsets power demand during daylight hours.

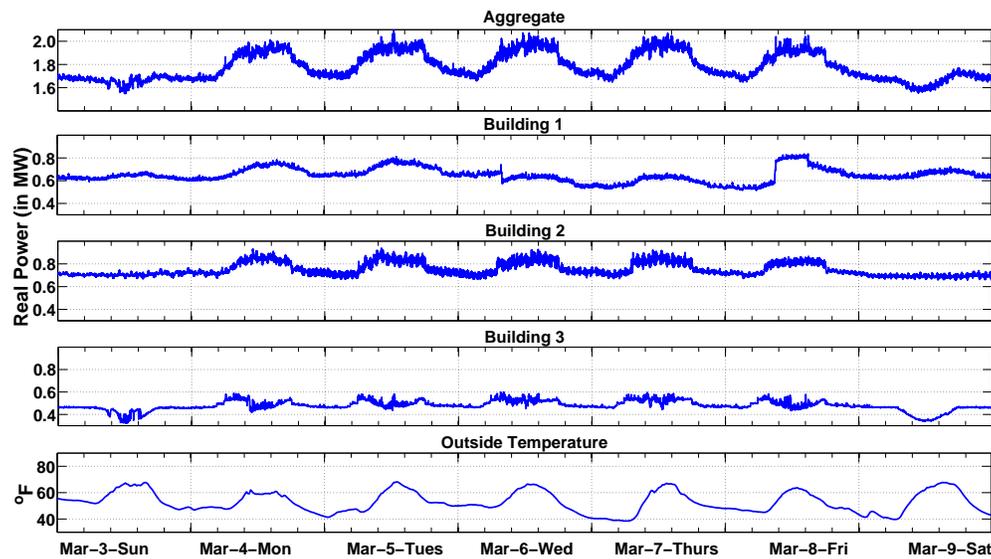


Fig. 2. Campus power use.

3.2. Power Data Collection

We have deployed 137 power meters on our campus. These include meters for building, top-level load distribution panels in Buildings 1-3, second-tier distribution panels within each building, and a few third-tier distribution panels in Building 1. The installed electrical meters are commercial (3-phase) devices from Schneider Electric (www.schneider-electric.com). Data is retrieved from each meter every 10 seconds using the MODBUS over Ethernet protocol. The data includes metrics such as individual line voltage, real and apparent power, power factor, current and frequency. The data is stored in a PI-Server from OSIsoft (www.osisoft.com).

3.3. Power Usage Characteristics

This section provides a brief summary of several properties of power use on the campus. The top plot in Figure 2 shows the aggregate power use for Buildings 1-3 over a one week period (from March 3 through March 9, 2013). This plot reveals several key characteristics. First, the demand has both a constant (base) and variable load components. The base load is quite significant, at 1.6 MW. The variable component adds another 0.4 MW of demand on work day afternoons. Second, there is a distinct time of day pattern. Power demand is lowest during the night and early morning, and highest during the late morning and afternoon. Third, there is a pronounced day of week behavior, with weekends (and non-work days in general) consisting primarily of the base load, and workdays having a noticeable variable load.

An implication of a significant base load is that very little insight as to what is responsible for campus power consumption, can be gleaned from the aggregate power. Besides, known disaggregation techniques (such as in Hart [1992]) do not work, as they cannot be scaled to handle hundreds to thousands of loads that are present in commercial buildings. This means that more meters must be installed. Our meter placement algorithm addresses the issue of how many meters are needed and where are they needed, to minimize the cost while maximizing the information obtained.

The middle three plots in Figure 2 show the total power demand for Buildings 1, 2, and 3, respectively. The bottom plot shows the outside temperature. Comparing

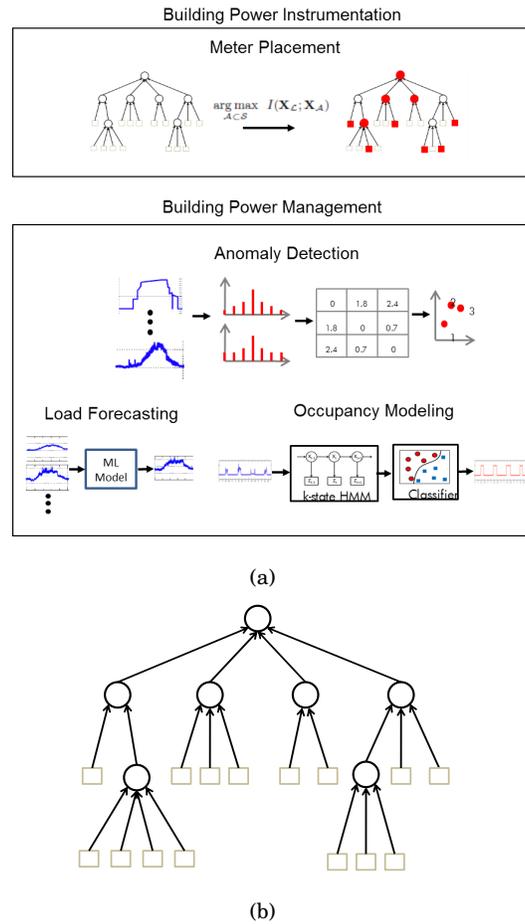


Fig. 3. (a) Overview of the methods used (b) Example topology of building electrical panels where meters can be installed.

this plot to the others reveals a correlation between outside temperature, occupancy (i.e., work hours) and power use. This motivates our investigation in Section 4.4 of occupancy modeling, to reduce the use of heating or cooling in areas of the building that people are not actively using.

4. METHODS

Figure 3(a) shows the overall framework of our approach. The meter placement algorithm forms the basis for instrumenting a building power infrastructure. For building power management, we propose an unsupervised anomaly detection and ranking method based on low dimensional embedding and k-nearest neighbors. Short-term load forecasting is used for managing demand for demand response events. For dynamic control of lighting and HVAC resources, we describe a semi-supervised approach that uses network port statistics to model occupancy.

4.1. Meter Placement

As noted in Section 3.3, the significant base load in the power demand of these buildings and the sheer number of devices in a large commercial building make known load disaggregation methods unreliable, thus requiring extensive metering of different electrical panels in each of these buildings for fine-grained power monitoring. However, one issue with this approach is that the total number of panels that could potentially be monitored can be very large to the extent that meter deployment at all these locations is not economical, given that the cost of each power meter can range anywhere between \$900 to \$3,000. In addition, the cost of installing these meters can be prohibitively expensive, especially in legacy buildings. This raises an interesting research question as to how and which panels should be selected for power meter deployment.

There are several criteria one could use to choose the panels for meter deployment. They include the total energy consumption of a panel, variability in the energy consumption, number of sub-panels/loads, predictability of panel power demand, or other information-theoretic measures. We choose mutual information, an information-theoretic measure that in a loose sense chooses panels that are highly unpredictable in terms of their power demands. As we show in Section 5.1, the panels selected using this criterion are superior to those selected using criterion such as the total energy consumption or variability in energy consumption.

Next, we demonstrate how this problem can be formulated as an optimization problem with the goal of choosing the set of panels with maximum information content.

4.1.1. Problem Formulation. Before we formulate the problem, we need to introduce some notation. The panels at different locations on the site are related in a topological manner that can be represented by a tree, as shown in Figure 3(b). Each node in this tree denotes a panel, where the leaf nodes (denoted by squares) correspond to panels that directly feed either a single load (for example, a chiller or a compressor) or a set of loads (for example, lighting load). The remaining nodes in the tree topology (denoted by circles) correspond to panels that feed other panels given by their child nodes in the tree.

Let S denote the entire set of panels, i.e., all the nodes in a given tree, and let $\mathcal{L} \subset S$ denote the set of leaf nodes. For any node $i \in S$, let X_i be a random variable denoting the power consumption recorded at panel i . Then, for any set of nodes $\mathcal{A} \subset S$, we denote by $\mathbf{X}_{\mathcal{A}}$ the random variables associated with the nodes in \mathcal{A} .

Given a constraint on the number of meters that can be afforded (k), we use mutual information as a criteria to choose the best set of panels for meter deployment, which can be formulated as an optimization problem shown below.

$$\begin{aligned} \arg \max_{\mathcal{A} \subset S} I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}}) \\ \text{s.t. } |\mathcal{A}| \leq k, \end{aligned} \quad (1)$$

where $I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}})$ denotes the amount of information conveyed about the power consumption at panels in \mathcal{L} by monitoring power consumption at panels in \mathcal{A} . Note that when $k \geq |\mathcal{L}|$, i.e., in the scenario where one could afford to deploy a power meter at each of the leaf nodes, the Mutual Information is maximized by choosing \mathcal{A} to be the set of all panels in \mathcal{L} . On the other hand when $k < |\mathcal{L}|$, the above optimization problem attempts to find the best set of panels that provide maximum information about power consumption at each of these leaf nodes.

4.1.2. Solution. Mutual Information is given by

$$I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}}) = H(\mathbf{X}_{\mathcal{L}}) - H(\mathbf{X}_{\mathcal{L}} | \mathbf{X}_{\mathcal{A}}),$$

which corresponds to the reduction in the uncertainty of power consumption at panels in \mathcal{L} given the power consumption information at panels in \mathcal{A} , where

$$H(\mathbf{X}_{\mathcal{L}}) = - \sum_{\mathbf{x}_{\mathcal{L}}} \Pr(\mathbf{X}_{\mathcal{L}} = \mathbf{x}_{\mathcal{L}}) \log_2 (\Pr(\mathbf{X}_{\mathcal{L}} = \mathbf{x}_{\mathcal{L}})), \quad \text{and}$$

$$H(\mathbf{X}_{\mathcal{L}}|\mathbf{X}_{\mathcal{A}}) = - \sum_{\mathbf{x}_{\mathcal{L}}, \mathbf{x}_{\mathcal{A}}} \Pr(\mathbf{X}_{\mathcal{L}} = \mathbf{x}_{\mathcal{L}}, \mathbf{X}_{\mathcal{A}} = \mathbf{x}_{\mathcal{A}}) \log_2 (\Pr(\mathbf{X}_{\mathcal{L}} = \mathbf{x}_{\mathcal{L}}|\mathbf{X}_{\mathcal{A}} = \mathbf{x}_{\mathcal{A}})).$$

Unfortunately, the optimization problem in (1) is NP-hard. Hence, we propose a greedy approach to optimize the given problem. The greedy algorithm chooses panels for meter deployment in a sequential manner, where given the set of panels that have already been chosen by the algorithm (denoted by \mathcal{A}), the next best panel is chosen to be the one that maximizes the gain in mutual information, i.e.,

$$j^* = \arg \max_{j \notin \mathcal{A}} I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A} \cup j}) - I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}}).$$

The solution obtained using the above greedy algorithm is not necessarily an optimal solution for the optimization problem in (1). However, we show below that the obtained greedy solution is guaranteed to be near-optimal.

4.1.3. Near-optimality of greedy solution. To show that the solution obtained using the above greedy approach is near-optimal, we rely on the theory of sub-modularity that was initially introduced by Nemhauser et al. [1978], and recently popularized by the work of [Krause and Guestrin 2005a; 2005b]. Krause and Guestrin [2005b] study budgeted maximization problems of the form

$$\begin{aligned} & \arg \max_{\mathcal{A} \subseteq \mathcal{S}} F(\mathcal{A}) \\ & \text{s.t. } |\mathcal{A}| \leq k, \end{aligned}$$

where $\mathcal{S} = \{1, \dots, N\}$ is a set of elements and $F : \mathcal{A} \rightarrow \mathbb{R}$ is a function that maps the set of elements to the real line. A greedy solution to this problem is to select elements sequentially according to

$$j^* = \arg \max_{j \notin \mathcal{A}} F(\mathcal{A} \cup j) - F(\mathcal{A}).$$

They further show that the solution obtained using this greedy approach will be near-optimal in the following sense

$$F_{\text{greedy}} \geq \left(1 - \frac{1}{e}\right) F_{\text{opt}},$$

iff the objective function F is submodular, where submodularity is defined below.

Definition 4.1. (Submodularity) Let F be a function that maps from a set of elements \mathcal{S} to the real line \mathbb{R} . Then, F is said to be submodular iff $\forall \mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{S}$ and for any $j \notin \mathcal{B}$,

$$F(\mathcal{A} \cup j) - F(\mathcal{A}) \geq F(\mathcal{B} \cup j) - F(\mathcal{B}).$$

In our optimization problem, the objective function is mutual information, which unfortunately is not submodular, except for some known special cases [Krause and Guestrin 2005a; Krause et al. 2006; Krause et al. 2008]. However, as we show in the following lemma, mutual information turns out to be submodular in our problem setting, thus guaranteeing near-optimality of the greedy algorithm described in Section 4.1.2.

LEMMA 4.2. *Given the tree topology described in Section 4.1, let S denote the set of nodes in the tree and \mathcal{L} the set of leaf nodes. Then, $\forall \mathcal{A} \subseteq \mathcal{B} \subseteq S$, and for any $j \notin \mathcal{B}$,*

$$I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A} \cup j}) - I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}}) \geq I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{B} \cup j}) - I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{B}}) \quad (2)$$

PROOF. From the definition of mutual information, we have

$$\begin{aligned} I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A} \cup j}) - I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}}) &= H(\mathbf{X}_{\mathcal{L}} | \mathbf{X}_{\mathcal{A}}) - H(\mathbf{X}_{\mathcal{L}} | \mathbf{X}_{\mathcal{A} \cup j}) \\ &= H(X_j | \mathbf{X}_{\mathcal{A}}) - H(X_j | \mathbf{X}_{\mathcal{L}}, \mathbf{X}_{\mathcal{A}}), \end{aligned} \quad (3)$$

where the second equality follows from the first by expanding the entropy terms and by simple manipulation of the resulting terms. Note that the second term in (3) is equal to 0, i.e., $H(X_j | \mathbf{X}_{\mathcal{L}}, \mathbf{X}_{\mathcal{A}}) = 0$, as given the power consumption at all of the leaf nodes, the power consumption at any panel upstream is completely deterministic.

Hence, the relation in (2) reduces to showing

$$H(X_j | \mathbf{X}_{\mathcal{A}}) \geq H(X_j | \mathbf{X}_{\mathcal{B}}),$$

which follows from the principle of ‘‘information never hurts’’ in information theory [Cover and Thomas 1991]. Thus, proving the submodularity of mutual information under the given tree topology. \square

From the above lemma, it follows that the information content captured by the greedy set of panels ($\mathcal{A}_{\text{greedy}}$), is at least 63% of the information content that would be captured by the optimal set of panels (\mathcal{A}_{opt}),

$$I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}_{\text{greedy}}}) \geq (1 - \frac{1}{e})I(\mathbf{X}_{\mathcal{L}}; \mathbf{X}_{\mathcal{A}_{\text{opt}}}).$$

4.1.4. *Use of Granger Causality.* An alternative strategy to select meters would be to apply Granger causality, which unlike mutual information, also considers the *direction* of the flow of information. Note that this could be remedied by the use of Transfer entropy, which is a version of mutual information that can detect the direction of information flow [Schreiber 2000], however, transfer entropy is currently restricted to bivariate situations.

Granger Causality (or G-causality) test, which was initially introduced in the field of economics [Granger 1969], is a statistical hypothesis test for determining whether one time series is useful in forecasting another. It is normally tested in the context of linear regression models. For example, let $X(t)$ and $Y(t)$ be two time series. Consider the following two auto-regressive models for predicting $X(t)$

$$X(t) = \sum_{j=1}^p a_j X(t-j) + e_1(t) \quad \text{and} \quad X(t) = \sum_{j=1}^p a_j X(t-j) + \sum_{j=1}^p b_j Y(t-j) + e_2(t),$$

where p is the maximum number of lagged observations included in the model, and $e_1(t), e_2(t)$ are the prediction errors (residuals) for the two regression models. If the variance in the prediction error is reduced by the inclusion of $Y(t)$ in the model, then Y is said to G-cause X . In other words, Y is said to G-cause X if the coefficients in $\{b_j\}_{j=1}^p$ are jointly significantly different from zero.

This test could potentially be used to reveal any hidden causal relationships between the loads (leaf nodes in the tree topology). Incorporating these relationships could further lead to a better choice of panels for meter deployment.

4.2. Anomaly Detection

Anomaly detection is useful in understanding and managing power consumption of a large campus. The primary goal of anomaly detection here is to detect any abnormal

behavior in the power usage time series. Note that an anomaly indicates an irregular usage pattern and may not always correspond to a component failure or faulty operation; anomalies include irregular power usage resulting in high power consumption.

There are two main challenges for performing anomaly detection. The first challenge is the lack of labeled data to train an algorithm for detecting anomalous behavior. Obtaining labeled data is an expensive procedure as it requires a human (usually a building administrator) to meticulously go through the vast amount of power data. In addition, it might also necessitate injecting faults to obtain a good representation of anomalies in the training data. The second challenge is the high dimensionality of the power data. As mentioned in Section 3, the power consumption data is collected every 10 seconds resulting in around 8,640 samples per meter per day.

In order to deal with these two problems, we propose a novel cluster-based unsupervised algorithm that detects anomalous points via a low-dimensional embedding of the power data. This algorithm takes as input the power time series observed by a meter over multiple days, and outputs the probability of the power consumption behavior being anomalous on each of these days. These probability scores can then be used to generate a ranked list of the data in the decreasing order of the data point being anomalous. This ranked list is useful to a building administrator in prioritizing the data points that need to be further inspected. The algorithm is described in more detail in the following section.

4.2.1. Method. First, we need to introduce some convention. We refer to power data measured by a single meter over a 24 hour period (i.e., one single day) as one observation or as a single power-time curve. However, the proposed algorithm is oblivious to the time resolution of an observation, and hence can be varied. For example, one can consider a finer time resolution such as 4 hours or a longer time period such as 1 week, as one observation. As mentioned above, due to the lack of labeled data, we would need to resort to an unsupervised approach where we cluster the power-time curves of each meter. The intuition behind this approach is that the data points that exhibit normal behavior form a tight cluster and all those points that lie outside this cluster are highly likely to correspond to an anomalous behavior.

In order to compare two power-time curves, we would need a good measure to quantify the dissimilarity between two observations. We propose the use of standard Euclidean distance measure or the l_2 distance between the frequency spectrum of two power-time curves as a measure of dissimilarity. Note that the frequency spectrum consists of two components - magnitude and phase. We restrict our attention to the magnitude of the frequency spectrum as it contains all of the necessary information regarding the power consumption behavior. Figure 4(a) outlines the proposed algorithm, which consists of five steps. We will now describe each step of the algorithm in detail.

Step 1: Missing Value Imputation. A power-time curve may have some missing values that could have been caused either due to a hardware or a software failure. Treating these missing values as zeros will lead to unnecessarily high frequencies in the frequency spectrum. We adopt a weighted global average strategy to impute the missing values. This method can be used to impute blocks of missing values, while preserving the local structure. Specifically, let $x[n]$, $n = 1, \dots, N$ denote a power-time curve where N denotes the number of time samples. For any time index $1 \leq m \leq N$ with $x[m]$ missing, we impute its value by

$$x[m] = \frac{\sum_{k=1}^N w[k]x[k]}{\sum_{k=1}^N w[k]},$$

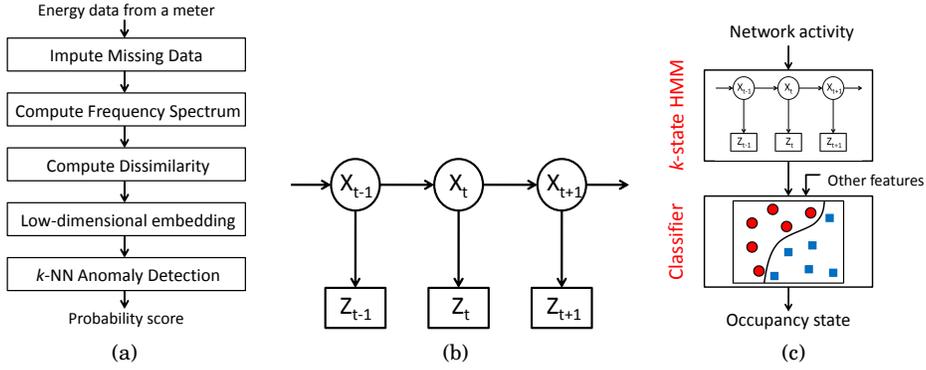


Fig. 4. (a) Proposed Anomaly Detection Algorithm (b) A Hidden Markov model with hidden variable X_t and observed variables Z_t (c) Proposed 2-stage approach for Occupancy Estimation

where the weights $w[k]$ are chosen such that they decrease as a function of their distance from the missing value. For example, the weight function can be chosen to be $w[k] = 1/|m - k|^2$.

The above imputation strategy can be considered as a temporal smoothing technique. Alternatively, one could perform spatial smoothing via nearest neighbors, where the missing values are replaced with values obtained from other power-time curves that have a similar profile in the non-missing region. In our data, on average less than 3% of values were missing.

Step 2: Frequency Spectrum Computation. In this step, we compute the frequency spectrum of the power-time curve obtained after imputing its missing values. Given a sequence $x[n]$, for $n = 1, \dots, N$, its frequency spectrum is computed as

$$X[k] = \sum_{n=1}^N x[n] * \exp\left(-j2\pi(k-1)\frac{n-1}{N}\right), \quad 1 \leq k \leq N.$$

We denote the magnitude of the frequency spectrum by $Y[k]$, where $Y[k] = |X[k]|$, $k = 1, \dots, N$. We are interested only in the magnitude of the frequency spectrum as it contains all the information regarding the total power consumption.

Step 3: Dissimilarity Matrix Computation. Given M different observations corresponding to power-time curves on M different days, we compute the dissimilarity between the power consumption profiles for any two observations using the standard Euclidean (or l_2) distance measure between their frequency spectrums as

$$\delta_{ij} = \left[\sum_{k=1}^N (Y_i[k] - Y_j[k])^2 \right]^{\frac{1}{2}}.$$

The $M \times M$ dissimilarity matrix Δ is obtained by computing the above distance measure for all pairs of observations. The resulting dissimilarity matrix Δ should be symmetric, i.e., $\Delta = \Delta^T$.

Step 4: Low-dimensional Embedding. Given the $M \times M$ dissimilarity matrix Δ , we use a dimensionality reduction algorithm such as MDS (Multi-dimensional scaling) to obtain a low-dimensional Euclidean embedding of the M observations in a $d \ll N$ dimensional Euclidean space (i.e., \mathbb{R}^d). An Euclidean embedding algorithm finds a set of M points

in \mathbb{R}^d such that the pairwise distances between these points are close to the values given in the matrix Δ .

Figure 9 demonstrates a low-dimensional Euclidean embedding of 33 power-time curves where $d = 2$. Note that each point in the resulting euclidean embedding corresponds to a power-time curve.

Step 5: k -NN Anomaly Detection. Given this low dimensional embedding, the last step is to compute the probability score of each observation being anomalous. We compute these values through a k -NN (nearest neighbor) density estimation algorithm. Note that a low-dimensional embedding of the power data is crucial for this step, as density estimation is known to perform poorly in a high dimensional space.

For every point $\mathbf{y} \in \mathbb{R}^d$ in the low dimensional space, the local density at that point can be estimated as

$$\hat{f}(\mathbf{y}) = \frac{k}{\text{Vol. of smallest hyper-sphere containing } k \text{ NNs of } \mathbf{y}}$$

where k is chosen to be $O(M^{\frac{1}{d}})$. Given the local densities at each of the M observations, the probability of an observation being an anomaly is computed as

$$\Pr(\mathbf{y}_i \text{ is anomalous}) = 1 - \frac{\hat{f}(\mathbf{y}_i)}{\max_{j=1, \dots, M} \hat{f}(\mathbf{y}_j)}.$$

Intuitively, observations that are in a high density region are less likely to be anomalous and those in low density regions are more likely to be anomalous.

4.3. Short-Term Load Forecasting

With the advent of advanced metering infrastructure, many commercial buildings are now being fitted with smart meters that can record electricity consumption every 15 minutes or less. Mining these large amounts of electricity consumption data will provide valuable insights into peak demand periods for buildings and provide accurate short term load forecasts. Such short term load forecasts (e.g., 1 hour ahead, 1 day ahead) and at finer spatial scale (e.g., floor level, zone level, load level) is crucial for many applications, such as frequency and voltage regulation, demand response participation, and micro grid management [Moslehi and Kumar 2010].

4.3.1. Methods. We study several machine learning algorithms and compare their performance for fine-grained, short-term load forecasting. Particularly, we compare six different algorithms - linear regression (LM), gradient boosted machines (GBM), support vector regression using a radial basis function (SVM), Gaussian process regression using a radial basis function (GAUSSPR), random forest (RF), and an Ensemble model that averages the output of the above 5 algorithms. Below, we describe the forecast operation more formally.

Let y_t denote a consumption time-series, where the time index t is typically discretized to hourly intervals. Given historical consumption data (i.e., $\{y_t : t \leq T\}$), our goal is to make a 24 hour ahead forecast, i.e., to estimate the consumption values y_t for $T \leq t \leq T + 24$. To make such a short-term forecast, we use features related to historical data, contextual features related to time, and other external factors such as weather forecast.

For historical features, we use the load consumption at the same time over the past 3 days (i.e., $\{y_{t-24}, y_{t-48}, y_{t-72}\}$) as well as the load consumption at the same time and same day of the week over the past 3 weeks (i.e., $y_{t-24*7}, y_{t-24*7*2}, y_{t-24*7*3}$). For contextual features related to time, we use hour of the day (0 to 23), day of the week (Sunday to Saturday), date (1 to 31) and season (Summer, Fall, Winter, Spring). In

addition, we use external factors corresponding to weather forecast such as ambient temperature and ambient humidity level.

For each machine learning algorithm, we optimize the model hyper-parameters using a k -fold time-series cross validation which works as follows. Given a time-series $\{y_t : 1 \leq t \leq T\}$, let $0 < T_1 < T_2 < \dots < T_k < T$ be k ordered time instants that are chosen either randomly or uniformly over the entire time period. In the i th fold, the forecast model is trained using the consumption data in the time interval $t \in [0, T_i]$, and cross-validated on data belonging to the time interval $t \in (T_i, T]$.

4.4. Occupancy Modeling

Occupancy modeling forms another important component for efficient power management in buildings. Many commercial buildings employ either a fixed time L-HVAC (Lighting, Heating ventilation and air conditioning) schedule or a fixed temperature set point schedule. This often leads to unnecessary conditioning of the building, especially when the actual occupancy is low. Hence, some recent work has suggested occupancy-based L-HVAC scheduling for efficient power management. However, most of this work assumes the availability of occupancy sensors, whose installation and maintenance may be prohibitive on a large campus.

Alternatively, Melfi et al. [2011] proposed the use of existing network infrastructure to estimate occupancy. They studied the use of Dynamic Host Control Protocol (DHCP) logs and other explicit ways such as monitoring PC activity in estimating occupancy. We instead develop an implicit occupancy sensing procedure, where we use traffic statistics associated with network ports in each cubicle to build occupancy models. Network switches typically maintain per port counters for the amount of traffic flowing in and out. We retrieve these statistics from the switches in the buildings every 30 minutes. The estimated occupancies at the cube level are then used to estimate occupancy of a zone (e.g., multiple cubicles), which can further be used for occupancy-based L-HVAC scheduling of that zone.

A primary challenge with using network data to estimate occupancy is the lack of labeled data. Obtaining labeled data from each occupant is not only expensive but also raises issues related to privacy. To address this, we consider two different approaches. The first is an unsupervised approach where we use a Hidden Markov Model (HMM) to estimate occupancy from network data. The second is a two-stage semi-supervised approach. Its first stage involves unsupervised learning using HMM, while the second stage trains a classifier using minimal labeled data. The two approaches are described in more detail in the following section.

4.4.1. Methods. We first propose an unsupervised approach where we model the problem of occupancy estimation from network data as a Hidden Markov Model (HMM). A HMM is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. Unlike a regular Markov model where the state transition probabilities are the only unknown parameters, in an HMM, the state is not directly visible, but output dependent on the state is visible. Each state has a probability distribution over the possible output tokens, referred to as emission probabilities. The observed output tokens provide some information about the hidden states. HMM has two phases, a learning phase where the state transition probabilities and the emission probabilities are estimated, and a decoding phase to estimate the hidden states from the observed tokens using the estimated probability parameters.

We model the problem of occupancy estimation from network data as an HMM, where the hidden variables X_t correspond to the binary occupancy states, and the observed variables Z_t correspond to the port level network traffic data, as shown in Figure 4(b). We consider a simple model where we assume that the transition prob-

abilities and the emission probabilities do not vary with time or any other external parameters. As we show in Section 5.4, this model performs fairly well.

However, the above assumptions may not be entirely true. In fact, it seems more appropriate to model these probabilities as a function of other features such as time of the day, day of the week, etc. However, the dependence of these parameters on such external features is known to significantly increase the complexity of HMMs, to the extent of making them intractable on large datasets. Alternatively, we propose a novel two stage semi-supervised approach that can efficiently incorporate the effect of external features, as shown in Figure 4(c).

In the first stage of this two stage approach, we model the network data using HMM with k underlying states for the hidden variable, where we choose the value of k that optimizes the log-likelihood function,

$$\log\left(\prod_{t=0}^T \Pr(Z_t|X_t)\right).$$

In the second stage, we train a classifier whose input is a feature vector consisting of the output state of k -HMM along with other external parameters such as time of the day, day of the week, etc. This approach remains tractable even on large datasets while efficiently incorporating the effect of any external parameters on the occupancy.

Besides, there are two other key advantages of using this first stage over just a supervised algorithm with network activity as one of the inputs. The first advantage is that it significantly reduces the labeling effort of an occupant during the training phase, where an occupant can now provide binary labels to the k output states of the HMM rather than providing their occupancy logs over time. This also addresses the issue of privacy to some extent. The other advantage of using the k -HMM is that it significantly reduces the size of the input feature space as k is usually very small compared to the total number of possible states for the network data. The result of a smaller feature space is that it requires less training data to efficiently train a classifier.

5. EXPERIMENTAL RESULTS

5.1. Meter Placement

To assess our meter placement method, we begin by using the greedy algorithm described in Section 4.1 to select the most informative meters among the 137 power meters installed on our campus. The underlying tree topology corresponding to these power meters is shown in Figure 5. The figure also shows the average power consumption values for the panels in the top three tiers. For simplicity, we assume in our experiments that the random variables corresponding to the power consumption at different panels are Gaussian.

Next, we greedily select the meters in a sequential manner using the greedy algorithm described in Section 4.1 under three different criteria: mutual information, total power consumption and variability in power consumption. Table I shows a ranked list of the meters in the order in which they are selected using these three criterion. Figure 6(a) shows plots comparing the three ranked lists. The diagonal line (dotted red line) in these plots correspond to the scenario where the two ranked lists are exactly equal.

Given this ranked list and a budget ($t < 137$) on the number of panels that can be metered, the top t panels from the ranked list can be chosen to be metered. The power consumption measured at these metered panels can then be used to predict the power consumption at the remaining panels. We use this predictive ability as a measure of goodness of the selected meters. We now compare the above three criteria based

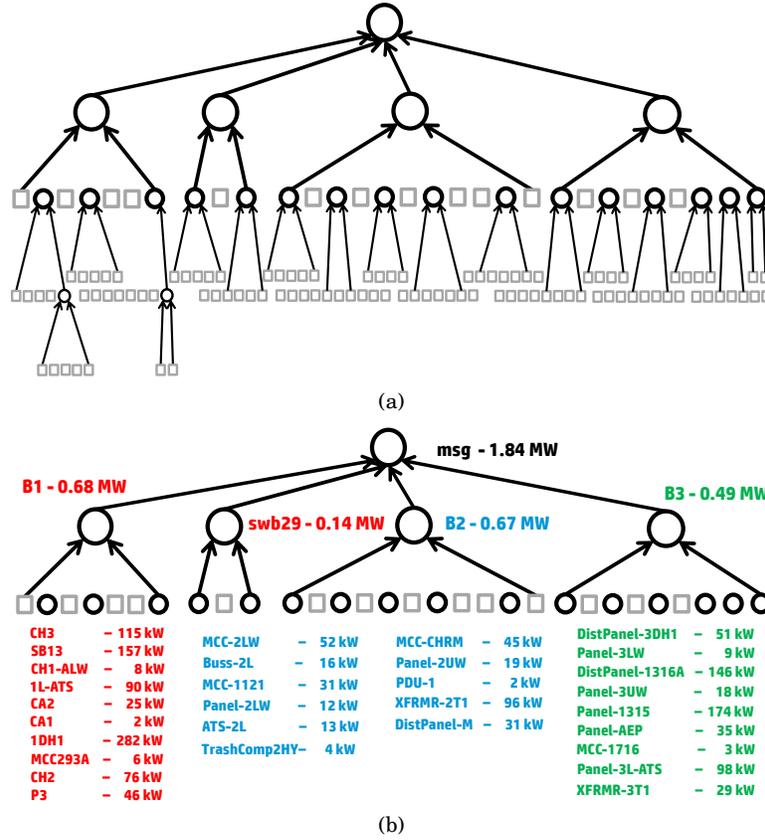


Fig. 5. (a) Tree topology of the 137 panels (b) Average power consumption of the top tier panels

Table I. Selection order of first 12 meters

#	Total Power Consumption	Variability in Power Consumption	Mutual Information
1	msg	msg	msg
2	b1-main	b1-main	swb29-CH2
3	b2-main	swb29-main	b1-CH3
4	b3-main	b2-main	b2-main
5	b1-1DH1	swb29-CH2	b3-main
6	b3-1315	b3-AEP	b3-AEP
7	b1-SB13	b1-CH3	swb29-main
8	b3-DistPanel-1316A	b1-SB13	b1-CH1ALW
9	b1-CH3	b3-3DH1	b2-DistPanelM
10	b3-ATS3L	b2-DistPanelM	b1-1DH1-2328
11	b2-XFRMR2T1	b3-3DH1-DPA1504	b2-MCC2LW
12	b1-ATS1L	swb29-P3	b3-3DH1

on the predictive ability of the corresponding panels selected. A random selection is also included as a baseline. Figure 7(a) demonstrates the average RMS (root mean squared) prediction error over all non-metered panels as a function of the number of panels metered (t). Similarly, Figure 7(b) demonstrates the average normalized RMS prediction error, where the normalization is based on the average power consumption of a meter. Note from these two figures that the proposed mutual information based

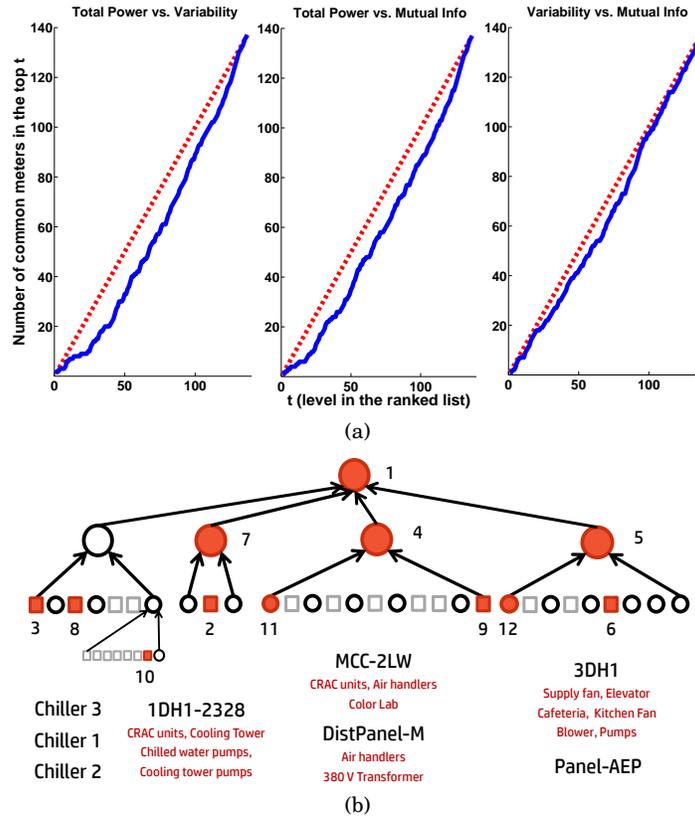


Fig. 6. (a) Comparison of ranked list similarity (b) The top 12 panels selected by the greedy algorithm with mutual information criterion

meter deployment outperforms those based on the total power consumption and the variability in power consumption. The curve corresponding to the random selection is averaged over 100 different random orderings of the 137 meters.

Figure 6(b) demonstrates the top 12 panels selected using mutual information. The validity of the obtained solution can also be verified intuitively. For example, panel b1-main is not selected as it is completely deterministic given the aggregate meter (msg) and the other three main meters (swb29-main, b2-main and b3-main). Similarly, most of the tier 3 panels that were selected consist of loads that are less predictable given the others. For example, b3-Panel-AEP corresponds to the panel that directly measures the power generated by the photo-voltaic array installed on Building 3. This meter is less predictable and hence can be considered as a good choice for meter deployment.

5.1.1. Discussion and Extensions. One limitation of our current implementation is the Gaussian assumption on the distribution of the random variables. This is not a strictly valid assumption, as most panels have at least two distinct operating states, one with higher power consumption during business hours and the other with a baseload power consumption during non-business hours. Hence, a more accurate approach would be to model these random variables as a mixture of Gaussian. However, one limitation of using a Gaussian mixture model is that there is no closed form expression for entropy of a Gaussian mixture density, and hence one would need to approximate it.

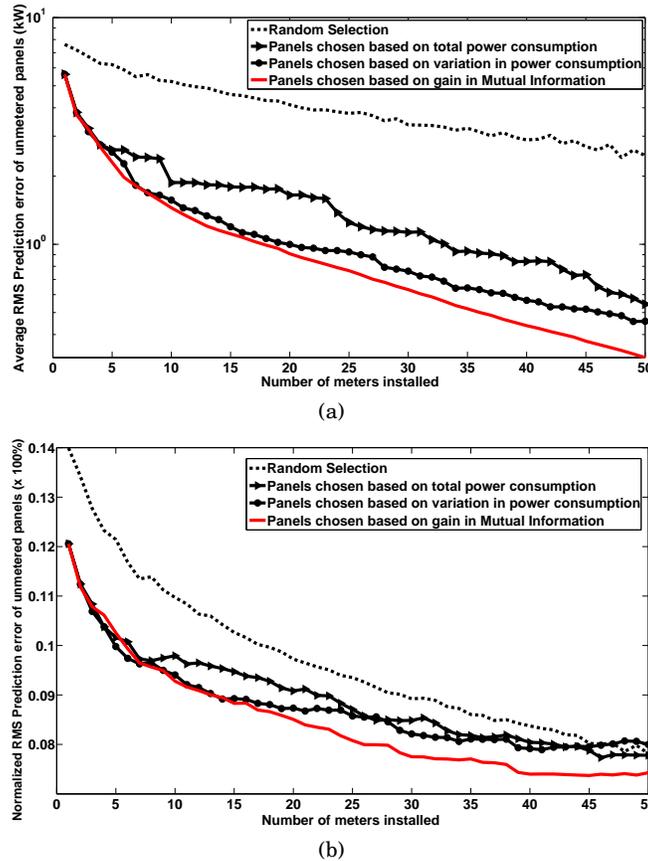


Fig. 7. Comparison of the predictive ability of the panels selected for metering in estimating power consumption at non-metered panels: (a) Average RMS prediction error (b) Average Normalized RMS prediction error

5.1.2. Use of Granger Causality. As described in Section 4.1.4, we investigate the use of Granger causality in selecting panels for meter placement. We use the G-causal test to rank the panels based on their predictive ability (results omitted for brevity). However, the mutual information based ranking performed better than the G-causal ranking with respect to the RMS prediction error on the non-metered panels. This could be due to the fact that the G-causal test is currently limited to linear regression models, whereas information theoretic measures are also sensitive to any non-linear relationships. Although extensions of the G-causal test to non-linear models exist, they are computationally less efficient and their statistical properties are not well studied¹.

5.2. Anomaly detection

We performed anomaly detection on six months of data from 35 of the 137 power meters, corresponding to the panels in the top three tiers of the tree topology (Figure 5). To validate our results, for three of these meters (b1-main, b2-main and b3-main), we obtained the ground truth by consulting with the building administrator, who looked at the entire time series data and marked days with potential anomalous regions.

¹http://www.scholarpedia.org/article/Granger_causality

Date	Score			Category	# of Anomalies	
Jul 6	0.99	Meter name	AUC	1	High power usage	66
Jul 7	0.97			2	Low power usage	65
Jun 28	0.80			3	Irregular Shutdown	6
Jun 20	0.75	b1-main	0.87	4	Irregular (time) usage	9
Jul 8	0.64	b2-main	0.96	5	Oscillatory behavior	28
⋮	⋮	b3-main	0.99	6	Abnormal drop/rise	29
(a)		(b)		(c)		

Fig. 8. (a) A sample ranked list of anomalies (b) Accuracy of the proposed algorithm in identifying anomalies as measured by AUC (c) Anomaly categories.

As described in Section 4.2, our algorithm assigns a probability score to each day, which can be used to obtain a ranked list of days in decreasing order of them being anomalous, as shown in Figure 8(a).

Given this ranked list, a building administrator could choose a threshold k and declare the top k points as anomalies for further inspection, and the remaining as normal, where k could vary from 0 to the maximum number of points in the input data. Each choice of k results in a certain number of false positives and false negatives. For example, when $k = 0$, i.e., when all the points are declared as normal, the false positive rate (FPR) is 0 while the false negative rate (FNR) will be 1. On the other hand, when k is the number of points, the associated FPR is 1 and FNR is 0. Varying this threshold k results in different values of FPR and FNR, leading to a receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) defines the quality of the obtained ranking. In the ideal case, where all the anomalous points are ranked at the top followed by normal points, the AUC takes the maximum value of 1. On the other hand, a random ranking achieves an AUC value of 0.5. We use AUC as a performance metric for our algorithm. Figure 8(b) shows the AUC values for the ranked list obtained using our algorithm on three meters.

Further, we applied the anomaly detection algorithm on the remaining 32 meters of the top three tiers, where we obtained a ranked list of anomalous days for each meter. We then manually characterized the top k anomalies in these ranked lists by assigning them categories, as shown in Figure 8(c). Note that a particular anomaly could belong to multiple categories. Detecting these anomalies could potentially offer several benefits such as energy savings, detecting faulty equipment resulting in savings in maintenance costs, etc. Potential power savings in the high power usage and irregular time usage anomalies varied from around 50 kWh to 2,000 kWh per anomaly. In Figure 9, we demonstrate four of these six categories.

The first three examples shown in this figure provide opportunity for potential energy savings. Figure 9(a) corresponds to a meter whose load consists of overhead lighting on a floor in one of the buildings. The low-dimensional embedding obtained using MDS shows a tight cluster of days with normal behavior, and two points (circled) that were detected as anomalous. The first anomaly corresponds to an abnormal low power usage (Category 2), which turned out to be July 4th, a holiday; while the other anomaly (June 17th) corresponds to high and irregular time usage (categories 1 and 4), where the lights remained on all night. This is an example of a potential anomaly (e.g., caused by an error in the lighting control system) that if fixed could have saved about 180 kWh of electricity. Figure 9(b) demonstrates a similar anomaly (June 23rd) with respect to the air handling units (AHU), where the air handlers were operating at full capacity until late in the night. Correcting this anomaly could have saved about 450 kWh of electricity. Note that the AHU load has multiple normal modes of operation depending on the utilization level of the air handling capacity. Similarly, Figure 9(c) demonstrates

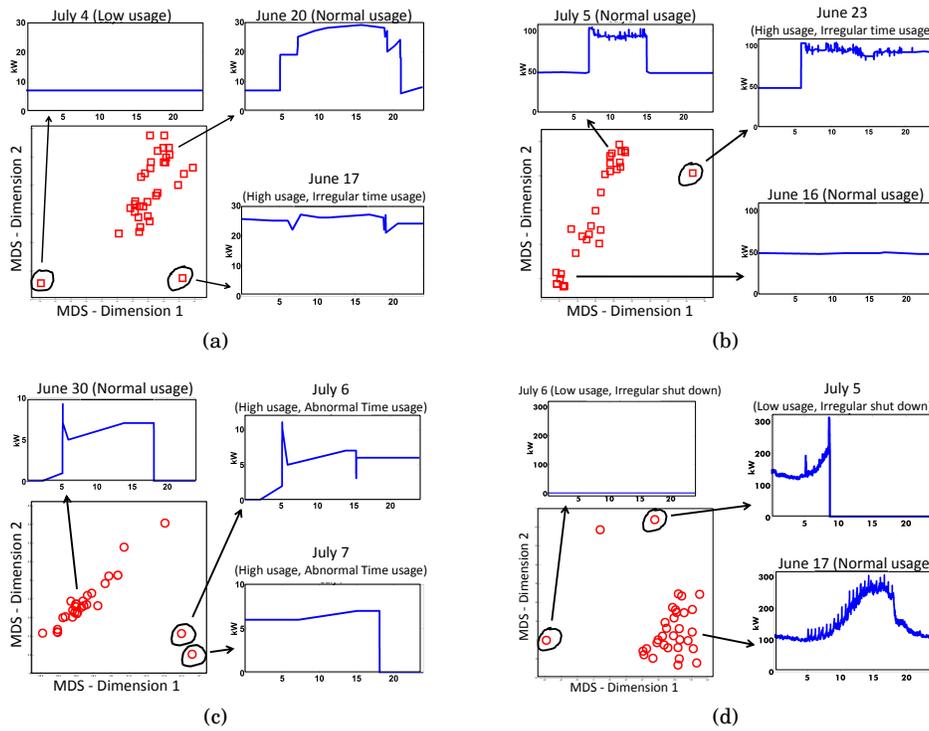


Fig. 9. Low dimensional embedding of power data corresponding to weekdays for four different meters. The figure also shows power profile during normal operation and the power profiles of abnormal events as detected by the algorithm.

two anomalies (July 6th and 7th) with respect to a fan load, where the motor fans were operating all through the night.

Finally, the example shown in Figure 9(d) corresponds to a chiller load. In this case, we detected three anomalous points corresponding to three consecutive days (July 5th and 6th shown in the figure), where the chiller was abruptly shut down (Categories 2 and 3) during business hours. If this was not caused due to a maintenance schedule, it could potentially correspond to a failed component.

5.3. Short-term Forecasting Results

We tested short-term load forecasting on 10 power meters corresponding to panels selected from the top three tiers of the tree topology. Table II lists these meters along with their load description and the variability in their consumption as measured by the normalized standard deviation (i.e., standard deviation/mean). We used 1 year of historical data with a four fold cross-validation to train the models. The consumption data was aggregated to hourly intervals, and the results are based on a 24 hour ahead forecast.

We evaluate the performance of our algorithms in terms of forecast accuracy. There are several accuracy measures such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Coefficient of Variation (CV) that have been proposed in the literature for time-series analysis. However, each of these measures have some drawbacks. For example, RMSE and MAE are not scale invariant. While MAPE and CV are normalized metrics and hence scale invari-

Table II. Description of the meters selected for short-term forecasting.

	Meter name	Normalized standard deviation	Load description
1	msg	0.24	Aggregate: Site
2	b1-main	0.14	Aggregate: Building 1
3	swb29-main	1.48	Aggregate: Building 1
4	b3-main	0.10	Aggregate: Building 3
5	b1-CH3	0.61	Single Load: Chiller
6	b2-MCC1121	0.14	Multiple Loads: Fans
7	b2-2UW	0.56	Multiple Loads: Lighting
8	b2-2HY	0.56	Multiple Loads: Outdoor
9	b3-3UW	0.62	Multiple Loads: Lighting
10	b3-3T1	0.20	Multiple Loads: Office space

ant, they do not have a lower or upper bound on the percentage error. They also tend to be very sensitive to outliers in the data. To overcome these problems, Bandyopadhyay et al. [2015] suggested the use of Symmetric Mean Absolute Percentage Error (SMAPE). SMAPE is a bounded measure that ranges between 0% and 100%, and can be used to compare the forecast accuracy across models and across meters. SMAPE is defined as

$$\text{SMAPE} = \frac{\sum_{t=1}^{24} |\hat{y}_t - y_t|}{\sum_{t=1}^{24} (|\hat{y}_t| + |y_t|)} * 100,$$

where y_1, y_2, \dots, y_{24} and $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{24}$ correspond to the original and forecasted values of the time-series over a 24 hour period.

Figure 10 shows the forecasting results for the 10 meters using the six algorithms described in Section 4.3.1. The forecast error measured using SMAPE is averaged over 4 test days. The meters are ordered based on their normalized standard deviation (NSD) from low to high. One would expect the forecast performance to depend on the aggregation level of the meter, i.e., single load vs. aggregate building. Instead, our results demonstrate that the forecast performance tends to depend on the normalized standard deviation or the variability of the time-series.

Particularly, note from Figure 10 that meters with a low normalized standard deviation aka low variability ($\text{NSD} < 0.3$) have a low forecast error, while meters with high normalized standard deviation ($\text{NSD} > 0.3$) tend to have a higher forecast error. Moreover, for meters with low variability, the difference between the performance of the various algorithms is insignificant, and hence a simple algorithm such as linear regression may be used for short-term forecasting. On the other hand, for meters with high variability, the forecast performance tends to vary highly among the different algorithms, with SVM and Ensemble achieving best overall performance.

5.4. Occupancy Modeling

In this section, we first quantify the performance of the proposed algorithms for occupancy estimation by comparing the estimated occupancy states to the ground truth. We then use the estimated occupancy to forecast potential energy savings that could be achieved by implementing an occupancy based lighting schedule.

We compare the performance between four different algorithms. The first algorithm is based on HMM with two underlying states, which is a completely unsupervised approach. The second and the third algorithms are semi-supervised approaches using a k -state HMM as described in Section 4.4.1, followed by a Naive Bayes classifier or a support vector machine (SVM) based classifier. The last algorithm is a supervised approach where the classifier is trained directly based on the port-level network statistics, and other features such as time of day.

To quantify the estimation accuracy of these algorithms, we collected the ground truth data from ten different occupants over a period of 16 weekdays. The occupants

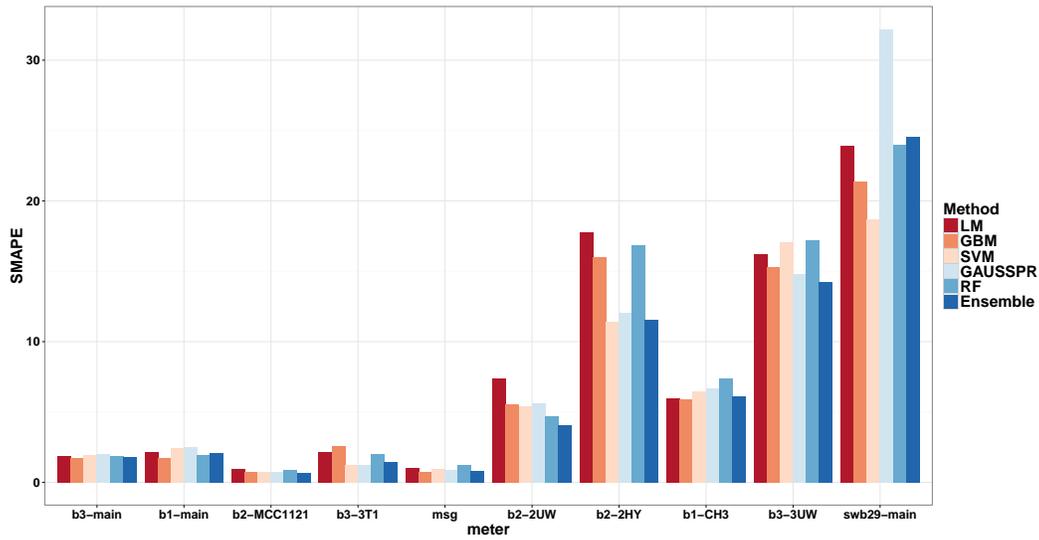


Fig. 10. Average error in short-term forecasting.

maintained their occupancy logs by marking their presence in their cube at a time resolution of 30 minutes, where an occupant marks their presence only if they are present for a majority of that 30 minute period. The in and out flowing network switch statistics are also collected every 30 minutes. The results of the experimental evaluation are shown in Figure 11.

Figure 11(a) compares the average error (or misclassification) rate in the estimated cube occupancy using the four different algorithms. The error rate is averaged over 16 different test cases. In each test case, the last three algorithms are trained using 15 days of occupancy data and tested on the left out day. Note from this figure that the two-state HMM in spite of being an unsupervised approach performs well, with an error rate less than 15% in eight of the ten cubes. In addition, k -state HMM along with SVM does marginally to significantly better in eight of the ten cubes, than a supervised approach that is directly based on the network statistics.

In Figure 11(b), we compare the quality of the two features (k -state HMM output and network statistics) in terms of estimating the occupancy states. The feature quality is measured using normalized mutual information between the feature states and the true occupancy labels. This figure shows that the k -state HMM output has a better feature quality than the port-level network statistics, further validating that the pre-processing step resulted in an improved classification accuracy.

The estimated occupancy states for each cube are then aggregated to estimate the occupancy of a zone. Figure 11(c) compares the true occupancy of a zone comprised of 10 occupants with that estimated using the 4 different approaches. Note from this figure that k -HMM+SVM performs the best, and 2-HMM, in spite of being an unsupervised approach, does well.

Our next experiment estimates the energy savings that can be obtained by using an occupancy-based lighting schedule. The current lighting schedule in our building is such that the lights are switched on as occupants arrive in the morning and the lights of all zones on one floor (about 159 cubicles) are set to turn off automatically at 9 pm. One of the 137 power meters captures the lighting load on this floor. We demonstrate the potential energy savings by implementing an occupancy-based schedule for

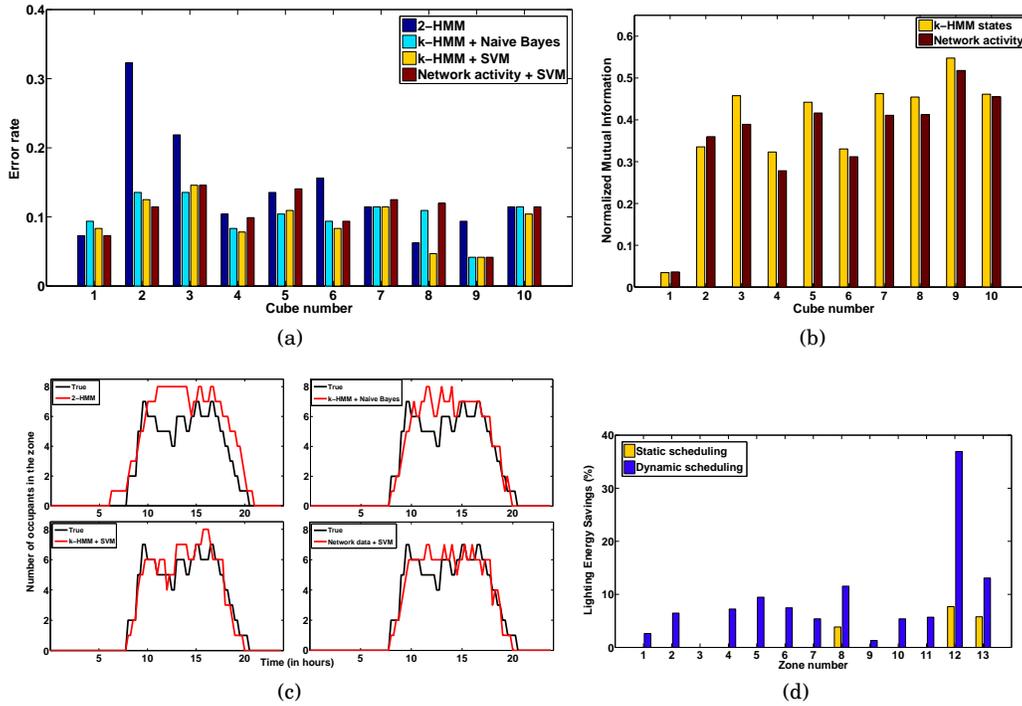


Fig. 11. (a) Average error rates of models (b) Feature quality (c) True zone occupancy vs. estimated values (d) Projected energy savings based on occupancy-based lighting schedule

switching off the lights at night, where the occupancy is estimated using 2-HMM. The use of this completely unsupervised approach is justified since 2-HMM usually overestimates the leaving time of an occupant, thus providing a conservative estimate of the time when all the occupants in a zone would have left. Figure 11(d) demonstrates the average energy savings for 13 different zones in one of the buildings, using two approaches. The first is static scheduling where for each zone, the lights are re-scheduled to turn off based on the worst case scenario observed over all days for that zone. This approach does not offer much savings as there could be a rare incident where an occupant of a zone stays till 9 pm or later. The second approach is to dynamically schedule the switch off time for lights in a zone based on the estimated occupancy for that zone. That is, the lights in a zone are turned off when the estimated occupancy for that zone becomes zero. This time may vary each day. Note from Figure 11(d) that this approach provides significant savings in the lighting energy. Overall, the proposed approach provides around 9.53% in savings for the building in consideration whose peak lighting load is 45 kW.

6. CONCLUSIONS

Commercial buildings consume significant amounts of energy. Concerns over energy prices and global climate change are motivating building operators to reduce energy consumption. In this paper, we proposed and evaluated four methods to aid in this effort. Our meter placement algorithm is both efficient and effective, guaranteeing a near optimal solution to information maximization by exploiting submodularity. In comparisons with other methods, the ability of the meter set selected using our algorithm to predict the measurements of the unselected meter set were found to be superior (by

an average of about 15%). Our anomaly detection method is shown to identify numerous types of unexpected consumption patterns. Our investigation on fine-grained, short-term load forecasting revealed an interesting relation between time-series variability and the optimal forecast method, with support vector regression and an Ensemble model being the best method for time-series with high variability, while a simple method such as linear regression is equally effective for time-series with low variability. Lastly, our occupancy modeling approach can be used to dynamically control lighting or HVAC resources, thereby reducing their energy consumption.

We plan to extend our work in numerous ways. We would like to further integrate our power management module by aiding anomaly detection through occupancy modeling. We would also like to automate the anomaly characterization task, and extend our anomaly detection algorithm to be able to incorporate feedback obtained from a building administrator.

ACKNOWLEDGMENTS

The authors would like to thank John Vogler for his valuable support and expert advise as a building administrator, and Ken Burden for gracefully providing us with the network log data.

REFERENCES

- Y. Agarwal, B. Balaji, S. Dutta, R. K. Gupta, and T. Weng. 2011. Duty-Cycling Buildings Aggressively: The Next Frontier in HVAC Control. *IPSN* (2011).
- N. An, W. Zhao, J. Wang, D. Shang, and E. Zhao. 2013. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy* 49, 0 (2013), 279–288.
- ASHRAE. 2004. *ASHRAE Standard 55: Thermal Environmental conditions for human occupancy*. Technical Report. ASHRAE, Inc.
- ASHRAE. 2007. *ASHRAE Standard 62.1: Ventilation for acceptable indoor air quality*. Technical Report. ASHRAE, Inc.
- B. Balaji, J. Xu, A. Nwokafor, R. Gupta, and Y. Agarwal. 2013. Sentinel: Occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. Rome, Italy.
- S. Bandyopadhyay, T. Ganu, H. Khadilkar, and V. Arya. 2015. Individual and Aggregate Electrical Load Forecasting: One for All and All for One. In *ACM Sixth International Conference on Future Energy Systems, e-Energy*. Bangalore, India.
- G. Bellala, M. Marwah, M. Arlitt, G. Lyon, and C. Bash. 2011. Towards an Understanding of Campus-scale Power Consumption. *3rd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (Buildsys)* (2011).
- G. Bellala, M. Marwah, M. Arlitt, G. Lyon, and C. Bash. 2012. Following the electrons: Methods for power management in commercial buildings. *18th ACM SIGKDD International conference on Knowledge discovery and data mining (KDD)* (2012).
- J. Brooks, S. Kumar, S. Goyal, R. Subramany, and P. Barooah. 2015. Energy-efficient control of under-actuated HVAC zones in commercial buildings. *Energy and Buildings* 93 (2015), 160–168.
- V. Catterson, S. McArthur, and G. Moss. 2010. Online conditional anomaly detection in multivariate data for transformer monitoring. *IEEE Transactions on Power Delivery* 25, 4 (2010), 2556–2564.
- CEC. 1993. *Advanced Lighting Guidelines*. Technical Report CEC Publication 400-93-014. California Energy Commission.
- T. M. Cover and J. A. Thomas. 1991. *Elements of Information Theory*. Wiley Interscience.
- R. E. Edwards, J. New, and L. E. Parker. 2012. Predicting future hourly residential electrical consumption. *Energy and Buildings* 49 (2012), 591–603.
- V. Erickson and A. Cerpa. 2010. Occupancy based demand response HVAC control strategy. In *ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*. Zurich, Switzerland.
- S. K. Ghai, L. V. Thanayankizil, D. P. Seetharam, and D. Chakraborty. 2012. Occupancy detection in commercial buildings using opportunistic context sources. In *IEEE International Conference on Pervasive Computing and Communications Workshop (PerCom)*. Lugano.

- C. W. J. Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* (1969).
- M. A. Hag, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, F. Hussin, and D. M. Said. 2014. A review on lighting control technologies in commercial buildings, their performance and affecting factors. *Renewable and Sustainable Energy Reviews* 33 (2014), 268–279.
- G. Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80, 2 (1992), 1870–1891.
- H. S. Hippert, C. E. Pedreira, and R. C. Souza. 2001. Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems* 16, 1 (2001), 44–55.
- S. J. Huang and K. R. Shih. 2003. Short-term load forecasting via ARMA model identification including non-gaussian process considerations. *IEEE Transactions on Power Systems* 18, 2 (2003), 673–679.
- V. Jakkula and D. Cook. 2010. Outlier detection in smart environment structured power datasets. In *IEEE Intelligent Systems*. London, UK.
- A. Krause and C. Guestrin. 2005a. Near-optimal Nonmyopic Value of Information in Graphical Models. *Uncertainty in Artificial Intelligence (UAI)* (2005).
- A. Krause and C. Guestrin. 2005b. *A note on the budgeted maximization of submodular functions*. Technical Report. CMU-CALD-05-103.
- A. Krause and C. Guestrin. 2007. Near-optimal Observation Selection using Submodular Functions. *National Conference on Artificial Intelligence (AAAI)* (2007).
- A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg. 2006. Near-optimal Sensor Placements: Maximizing Information while Minimizing Communication Cost. *International Symposium on Information Processing in Sensor Networks (IPSN)* (2006).
- A. Krause, A. Singh, and C. Guestrin. 2008. Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *Journal of Machine Learning Research* 9 (2008), 235–284.
- J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. 2007. Cost-effective Outbreak Detection in Networks. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (2007), 420–429.
- X. Li, C. Bowers, and T. Schnier. 2010. Classification of energy consumption of a building with outlier detection. *IEEE Transactions on Industrial Electronics* 57, 11 (2010), 3639–3644.
- H. Masuda and D. E. Claridge. 2014. Statistical modeling of the building energy balance variable for screening of metered energy use in large commercial buildings. *Energy and Buildings* 77 (2014), 292–303.
- R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen. 2011. Measuring Building Occupancy Using Existing Network Infrastructure. *International Green Computing Conference* (2011).
- K. Moslehi and R. Kumar. 2010. A reliability perspective of the Smart grid. *IEEE Transactions on Smart Grid* 1, 1 (2010), 57–64.
- G. Nemhauser, L. Wolsey, and M. Fisher. 1978. An analysis of the approximations for maximizing submodular set functions. *Mathematical Programming* (1978).
- G. Newsham and B. Birt. 2010. Building-level occupancy data to improve ARIMA-based electricity use forecasts. In *ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys)*. Zurich, Switzerland.
- P. F. Pai and W. C. Hong. 2005. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Research Systems* 74, 3 (2005), 417–425.
- N. Sapankevych and R. Sankar. 2009. Time series prediction using support vector machines: A survey. *IEEE Computational Intelligence Magazine* 4, 2 (2009), 24–38.
- T. Schreiber. 2000. Measuring Information transfer. *Physical Review Letters* 85 (2000), 461–464.
- R. Sevlian and R. Rajagopal. 2013. Value of aggregation in smart grids. *IEEE International Conference on Smart Grid Communications* (2013), 714–719.
- R. Sevlian and R. Rajagopal. 2014. A scaling law of short term electricity load forecasting on varying levels of aggregation. *arXiv preprint arXiv:1404.0058* (2014).
- H. Teraoko, B. Balaji, R. Zhang, A. Nwokafor, B. Narayanaswamy, and Y. Agarwal. 2014. *BuildingSherlock: Fault Management Framework for HVAC Systems in Commercial Buildings*. Technical Report. UCSD-CS2014-1007.
- USEIA. 2015. [http://www.eia.gov/forecasts/aeo/pdf/0383\(2015\).pdf](http://www.eia.gov/forecasts/aeo/pdf/0383(2015).pdf). (2015).
- H. Wang, P. Xu, X. Lu, and D. Yuan. 2016. Methodology of comprehensive building energy performance diagnosis for large commercial buildings at multiple levels. *Applied Energy* 169 (2016), 14–27.
- R. Yin, S. Kilicote, and M. A. Piette. 2015. Linking measurements and models in commercial buildings: A case study for model calibration and demand response strategy evaluation. *Energy and Buildings* (2015).