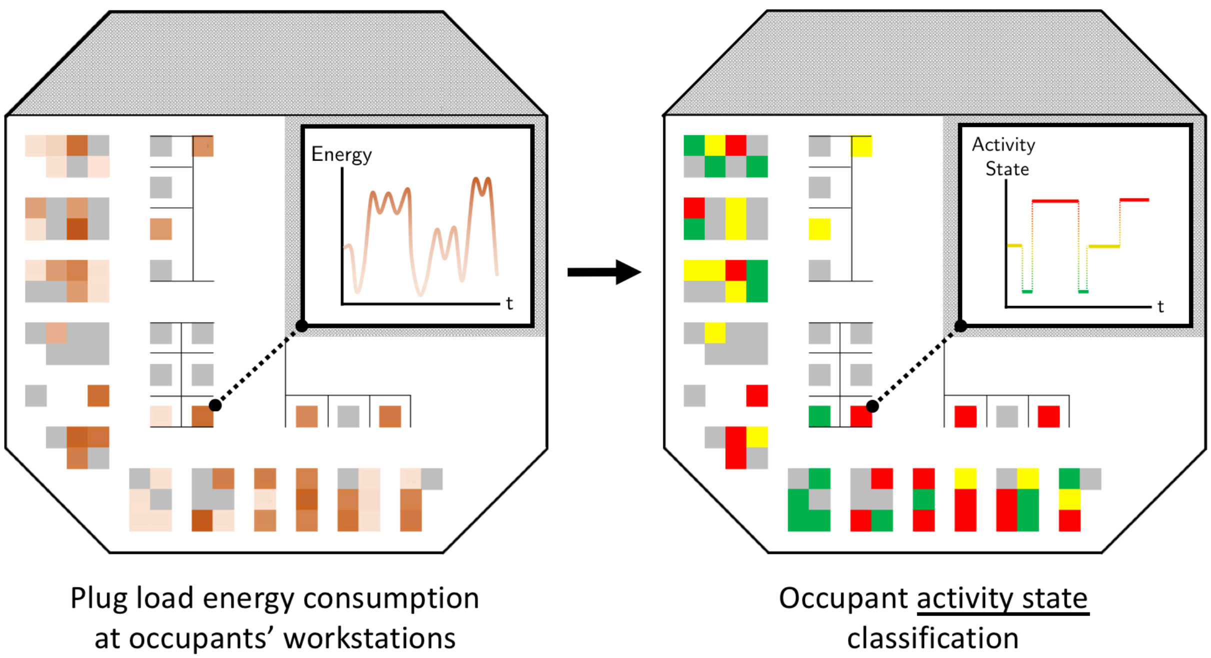
NOTICE: This is the authors' version of a manuscript that was accepted for publication in *Advanced Engineering Informatics*. Changes from the publishing process, such as editing, corrections, and formatting may not be reflected in this manuscript. A definitive version can be accessed via the following link: https://doi.org/10.1016/j.aei.2018.04.009

Please cite this research as follows:
Sonta, A. J., Simmons, P. E., and Jain, R. K. (2018). *Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach.* Advanced Engineering Informatics, 37, 1-13.

Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach

Andrew J. Sonta[1], Perry E. Simmons[2], Rishee K. Jain[3]

ABSTRACT



[1] Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Room 269B, Stanford, CA 94305; asonta@stanford.edu
[2] Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Room 269B, Stanford, CA 94305; perry.simmons@stanford.edu
[3*] Urban Informatics Lab, Department of Civil & Environmental Engineering, Stanford University, 473 Via Ortega, Room 269A, Stanford, CA 94305; rishee.jain@stanford.edu (corresponding author)

Buildings are our homes and our workplaces. They directly affect our well-being, and they impact the natural global environment primarily through the energy they consume. Understanding the behavior of occupants in buildings has vital implications for improving the energy efficiency of building systems and for providing knowledge to designers about how occupants will utilize the spaces they create. However, current methods for inferring building occupant activity patterns are limited in two primary areas: First, they lack adaptability to new spaces and scalability to larger spaces due to the time and cost intensity of collecting ground truth data for training the embedded algorithms. Second, they do not incorporate explicit knowledge about occupant dynamics in their implementation, limiting their ability to uncover deep insights about activity patterns in the data. In this paper, we develop a methodology for classifying occupant activity patterns from plug load sensor data at the desk level. Our method makes us of a common unsupervised learning algorithm—the Gaussian mixture model—and, in addition, it incorporates explicit knowledge about occupant presence and absence in order to preserve adaptability and effectiveness. We validate our method using a pilot study in an academic office building and demonstrate its potential for scalability through a case study of an open-office building in San Francisco, CA. Our method offers key insights into spatially and temporally granular occupancy states and space utilization that could not otherwise be obtained.
Keywords: Building energy; Data-driven; Gaussian mixture model; Knowledge-based; Occupant activities; Occupant dynamics; Space utilization; Zero-training

1. INTRODUCTION

Buildings are integral to our daily lives. People spend an estimated 87% of their time indoors [1], and researchers have shown that buildings directly affect our well-being [2]. Moreover, buildings worldwide account for over 19% of energy-related CO_2 emissions and 51% of global energy consumption [3], making them an integral part of our sustainable energy future. Fundamentally, buildings consume this energy to provide their occupants with services, including thermal and visual comfort, access to water, and power for electronic devices. As a result, understanding the relationship between buildings and their occupants is central to designing buildings that enhance occupant well-being, improve service delivery, and reduce energy usage.

We define occupant dynamics as the complex interactions between buildings and humans, encompassing occupant presence, occupant behavior (*i.e.*, the specific actions that occupants take in buildings, such as working at a workstation, taking a break, or even interacting with lighting or heating, ventilation, and air conditioning (HVAC) controls), occupant activity states (*i.e.*, abstracted and categorized information about occupant behavior), and the impact occupant behavior has on building operations. These dynamics are challenging to understand due to the increasing complexity of our building systems and the socio-technical complexities of occupant behavior. Even gaining a clear picture of the spatial and temporal activity patterns of occupants within a building is a non-trivial task [4]. While new types of sensors have facilitated more data-driven approaches to understanding occupant-building dynamics, they suffer from a few key limitations. Sensors designed to directly detect occupancy often mischaracterize the spaces they are sensing due to the complexity associated with various building spaces [5]. New statistical and data mining techniques that have been proposed to infer occupancy patterns from emerging high-fidelity data streams such as light levels [6], energy use [7], sound [8], and video [9] typically require a significant amount of ground truth training data that is cumbersome and often cost prohibitive to collect, thereby limiting their applicability and feasibility at scales beyond small pilot studies. Conversely, knowledge-based approaches to understanding occupant dynamics in buildings (*e.g.*, surveys, on-site engineering audits) can yield insights on occupant dynamics [10,11] but suffer from common reliability and scalability issues associated with indirect collection instruments [12].

In various aspects of building design, construction, and management involving human activity, researchers have shown that combining expert knowledge about buildings with automated computing techniques can vastly improve the effectiveness of the embedded methods. In the context of augmented reality within buildings, researchers have shown that integrating explicit engineering knowledge about building layout and operator movement into the automated augmented reality framework can improve the accuracy of the overall system [13]. In construction management, the process of extracting meaningful information about the activities of construction workers from raw cellphone data can be enhanced by incorporating explicit engineering knowledge about the necessary levels of detail required for improving the effectiveness of construction activity simulations [14]. These studies and others like them emphasize the point that automated methods can be made more accurate and effective by integrating knowledge about the specific domain in the design of the overall methodology.

In this paper, we present a new methodology that integrates knowledge-based and data-driven approaches to understanding occupant activities in buildings with the goal of informing enhanced building design and energy efficient operations. Our method infers activity states for individual occupants using time-series data from low-cost, off-the-shelf plug load sensors. It incorporates explicit domain knowledge about how occupant activities impact plug load data into a common unsupervised learning algorithm—the Gaussian mixture model—to characterize the data into abstracted levels of activity. We design our method to be able to automatically analyze the highly variable data associated with occupant presence separately from the less variable data associated with occupant absence. This design decision in our method allows it to more deeply characterize the data while maintaining adaptability to new spaces, potential for scalability to larger spaces, and high accuracy. We validate and demonstrate that our method is able to determine individual occupancy states with a high-level of accuracy on a small control study, and we demonstrate the merits and applicability of our approach on a case study of a real 47-person open office in San Francisco, CA, USA.

2. BACKGROUND

Building designers and managers are increasingly utilizing sensors and the data they collect to make decisions about how buildings are designed, built, and operated [15]. These sensors measure properties such as air temperature and humidity, lighting levels, sound, movement, and plug load energy use [16–20]. Each of these types of sensors produces time-series data that provides information about the changing state of the building. In many cases, data produced within a building can be utilized to make decisions that can improve the energy efficiency of that building: for example, a lighting sensor may provide feedback to lighting controls that can dim the overhead lighting if the building is receiving enough light from outdoors. In others, data can be used to understand characteristics of existing buildings so that the design of future buildings can be improved: for example, data describing existing building occupancy can be linked with predictive energy models to increase the accuracy of energy models [21].

This explosion of data has created an opportunity to provide new knowledge to engineers, designers, and building managers. In particular, previously unavailable information about the state of occupancy in buildings—the presence or absence of occupants as well as their activities—can be useful both for efficient building control of existing buildings and for improved space planning of future buildings [22]. Along with other data streams specific to each building system, the detection of occupant activities has been shown to be significant in addressing all forms of energy use in buildings, from lighting control [23–25] to HVAC control [26,27]. In addition, as knowledge about space use becomes more widely available to designers, the integration of design heuristics with occupancy models will be integral to designing spaces that better suit the needs of occupants [28]. In this section, we discuss the state of data-driven decision making in buildings for energy efficient building operations and improved building design, as well as the importance of occupancy and the state of the art for detecting occupant presence and occupant behavior in buildings. We elucidate the need for a robust, adaptable method for determining the activity states of occupants in buildings.

2.1. Data-driven & occupant-driven energy efficiency

Over recent years, the analysis of building energy data with statistical and data mining techniques has been shown to be helpful in improving energy efficient management of building systems. Within buildings, researchers have worked toward achieving a condition in which building systems—such as lighting, heating, and cooling—are provided only as much as they are needed, and only where and when they are needed. Matching these building systems with occupancy information has been shown to lead to significant energy savings [17,18]. Recently in commercial buildings, energy use data collected through power strips installed at the individual outlet level have been used for multiple approaches to save energy in buildings: to show that energy is wasted due to inefficient occupant behavior, such as leaving lights or other systems on during non-occupied hours [29]; to calibrate and improve the accuracy of building energy models in conjunction with other building data sources [30]; and to describe the behavior of occupants and improve schedule modeling in buildings [31].

Many studies have noted the high impact occupant presence and behavior has on building energy use [32–34]. Jia et al. [35] has noted that occupant behavior (as distinct from occupancy) relates to more than just the presence or absence of occupants in buildings—that is, the *activities* of occupants within the building have a large impact on building energy performance. However, this human element, which is responsible for much of building energy use, is often difficult to characterize. One reason is because it is multidimensional, requiring a fundamental understanding of spatial, temporal, and social dimensions of occupant behavior [36].

Understanding each of these dimensions and reconciling their effects on occupant behavior is critical to gaining a broad understanding of occupant behavior and its impact on building energy use. Furthermore, the structure and type of the social network of occupants has been shown to be highly influential when it comes to how occupants behave and adapt to information in buildings [37,38]. Researchers have shown that providing the right information to occupants can lead to changes in behavior that reduce buildings' energy consumption [39–43]. Due to the energy-consumption impact, complexity, and ever-changing nature of occupant dynamics in buildings, there remains a pressing need to better understand them.

2.2. Occupancy data & space utilization

While whole-building data and occupancy data have typically been studied in the context of energy efficient management of existing buildings, they also have the potential to be tremendously useful in providing knowledge to designers in the early stages of building design. Previous research has utilized model-based optimization in the design of buildings [44], and more specifically, in the planning of space layouts in buildings [45–47]. Recent work has conceptualized models that utilize computing in the assessment of the functional properties of designed spaces [48]. Specifically, analyzing designs for their ability to perform their function—for example, the ability for a proposed office space to promote a productive work environment—depends on knowledge from empirically based methods (*e.g.*, surveys) [49].

Architects have traditionally used personal perceptions of how occupants will use the spaces they design in their planning process. Formalized integration of human-centered knowledge into the building design process has previously been focused on perceptions of space [50] and heuristics for improved layouts [51], among others. More recent work has underscored the notion that it is difficult to quantify and optimize the function of spaces due to a lack of information about how occupants utilize spaces designed for them. Dzeng et al. [46] found that function space assignment optimizations that are based on user activities can increase the prescribed function objectives significantly (*e.g.*, improving overall space use by optimizing prescribed building assignments in a remodeling effort). However, these methods typically use occupant activity simulation models that are built on occupant activity data obtained through onerous methods such as defining heuristics from previous spaces and predetermined schedules [46] or from specialized occupant movement sensors [52]. With the potential to accurately and granularly detect occupant activities in existing buildings from more ubiquitous sensors, new design-knowledge integration approaches will have greater opportunity to incorporate empirically grounded occupant activity patterns into new design heuristics.

As engineers and designers continue developing tools to aid in the design of buildings that more appropriately meet the needs of their occupants, analysis of the *utilization* of spaces has become increasingly important. Space-use analysis helps designers determine how appropriately the spaces within a building are serving their occupants. Spaces that are *properly utilized* fulfill their design intentions by having a certain level of occupancy at predetermined times and by not inhibiting occupants from performing predetermined activities. Spaces that are not properly utilized can either be underutilized (in which case they are inefficient in their use of space), or they can be too crowded (in which case they inhibit occupants from performing the activities they were meant to be able to perform), with proper utilization rates depending on the nature of the space being analyzed [53,54]. Recently, researchers have proposed frameworks that can be helpful for architects working on space-utilization in the programming phase of their design process, but those frameworks depend on a detailed understanding of how occupants use the spaces designed for them [54]. There remains significant opportunity to analyze the activities of occupants in existing spaces for greater understanding of occupant dynamics in planned spaces [55]. Furthermore, carefully representing the information gained from these analyses can provide useful knowledge to key decision-makers such as building designers or managers [56].

Further improvement of the accuracy of models that help designers understand how occupants will use the spaces they plan depends on a solid understanding of how occupants utilize spaces in existing buildings. Only by monitoring and understanding the dynamics of occupant dynamics in existing buildings can we hope to imagine how occupants will behave in the new buildings we design [57].

2.3. Detecting occupant presence and activities in buildings

Because occupant behavior is so important to the energy use and space-use planning of buildings, there is a need for better tools to detect, model, and understand levels of occupancy and the activities of occupants. Melfi et al. [58] discusses the need for understanding occupancy at a high level of granularity in terms of temporal resolution, spatial resolution, and resolution of occupancy (activities versus presence/absence). Specifically, as the level of resolution increases, more information can be gained from the sensors, elucidating the need for sensors that can determine the activities of occupants at the spatial resolution of individual workstations and the temporal resolution of minutes.

Recent work has utilized sensors or combinations of sensors—including infrared [59], video [60], and acoustic [61]—to estimate the occupancy of rooms in buildings. These studies have shown that intelligently controlling systems such as lighting through use of occupancy sensors can save significant amounts of energy in buildings. Other recent work has used computer vision algorithms to characterize the movement of occupants in building spaces, noting that understanding the activities of building occupants leads toward a better understanding of how spaces can be designed for improved spatial efficiency (*i.e.*, more properly utilized spaces) and better user experiences [57,62].

More recent work has utilized plug load energy data collected at the desk level as an additional input for algorithms that estimate the true occupancy levels of buildings [63–65]. Zhao et al. [31] has shown that plug load data of computers and task lights at the desk level can be utilized to determine occupants' activities rather than just the level of occupancy. Due to the fact that plug load sensors are relatively inexpensive and often already installed in commercial office buildings for investigations into plug load management [66], they can be considered a low-cost alternative to sensors designed specifically for occupancy detection, such as infrared sensors. Moreover, many sensors that are designed specifically for occupant presence detection—such as infrared, acoustic, and CO2 sensors—require large time lags up to 60 minutes for high accuracy, while plug load sensors have been shown to be useful in determining occupant presence at time scales on the order of 5-15 minutes [33,38].

The analysis of plug load data for the detection of occupant activities typically involves data mining techniques and classification algorithms such as decision trees [63,64] and hidden Markov models [65]. These classification techniques are used to map the collected plug load energy use data to levels of occupancy, or in more sophisticated algorithms, to the types of activities occupants perform in buildings, such as working at a workstation. This recent work using plug load sensors has high value in advancing our methodologies for determining occupant activities, however, it requires training the classification models on ground truth data that is often onerous and cost-prohibitive to collect. While such previous work demonstrates how energy use data can be utilized to gain an understanding of occupant activities, it is limited in its ability to naturally adapt to quick changes in the appliances used at the desk or to additional workers in the space. Scaling to large buildings with many individuals would require large, coordinated efforts to collect ground truth data for training individual models. As a result, it is necessary to have a method that is robust and that can easily adapt to individual consumption patterns in order to glean a deeper understanding of occupant dynamics within a building. For this reason, this paper aims to contribute an effective methodology for detecting individual occupant activity patterns across a building. By combining knowledge-based and data-driven approaches, we are able to uncover occupant activity patterns from plug load energy use data and utilize such insights to inform energy efficient operational and space utilization strategies in a commercial building.

3. METHODOLOGY

Our occupant activity state classification method maps occupants' energy use, collected through plug load energy use sensors over a 5-20-minute period, to activity states. Based on major ideas from Section 2, our method is designed with a few key motivations in mind. The method should be able to adapt to different situations, such as when occupants in different buildings or floorplans tend to have different types of appliances at their workstations. For design of the algorithm, this adaptability would mean that parameters would not need to be tuned and set before the method is applied. In addition, the method should not require the use of training data in order to infer activity states, as we believe that a cost-effective and adaptable method should be able to be applied to new situations (such as new buildings or floorplans) without requiring onerous data collection procedures. In order to accomplish these tasks, our method is designed to combine a common unsupervised learning approach—the Gaussian mixture model—with explicit engineering knowledge about the typical structure of plug load energy signatures.

Various algorithms have been applied to time-series data—like that captured through plug load sensors—for the purposes of uncovering the underlying clusters in the data. In the context of analyzing plug load data, previous work has used supervised learning algorithms that train models on ground truth data and utilize these models to predict the correct classification of new data. In particular, the naïve Bayes and support vector machine algorithms have been shown to be successful in determining whether or not occupant workstations are occupied [31]. However, these supervised learning methods require the collection of ground truth data for training. Because we are interested in designing a method that does not require training on ground truth data, we looked toward unsupervised clustering algorithms that recognize patterns in the data and produce clusters with similar characteristics. Previous work has utilized the k-means clustering algorithm to disaggregate occupancy presence data into typical patterns of occupancy schedules [67]. The k-means algorithm avoids the need to train underlying models on ground truth data, but it requires setting the number of clusters—"k"—before running the model. Therefore, the user must either know ahead of time how many clusters are present in the data, or tune the parameter by running the algorithm with various values of "k" and picking the model based on some goodness-of-fit test. To avoid the need to train on ground truth data and/or tune model parameters, we chose to build our method around variational Bayesian inference.

Our method requires collecting continuous time series energy use data from each occupant's workstation using a plug load sensor at a time granularity of 5-20 minutes, as this time scale adequately captures changes in occupant activities [31,36]. A component selection process that utilizes a Variational Bayesian Gaussian Mixture Model (VB-GMM) determines the number of activity states present in the data. For the component selection process, a VB-GMM is applied for each occupant for each day separately, then a single number of components (M) is chosen for all occupants. After the component selection process, new Gaussian Mixture Models (GMM) are fit to the data for each occupant for each day separately, and the energy use data is classified based on the model fits.

The component selection process is based on a variational Bayesian inference method that utilizes the GMM as its basis. We employ this component selection process as a means of alleviating the need to make any a priori assumptions as to the number of activity states the occupant have. The component selection process makes use of engineering domain knowledge about occupant dynamics and plug load energy consumption in order to glean more compelling insights about occupant activities. In particular, we allow our method to recognize when data corresponding to occupant presence is classified as distinct from data corresponding to occupant absence, and we allow the component selection process to analyze the presence data separately. This flexibility, which would not be possible without incorporating knowledge about occupant dynamics, building energy use, and plug load energy data, allows our method more fully analyze the data. After the component selection process, new GMMs are fit for each occupant and for each day using the inferred number of components. This step ensures that all periods of occupant energy use are being classified with the same number of components, enabling effective comparison across occupants. This avoids situations in which, for example, one occupant tends to have 3 components, and another tends to have 4 components, and it enables extraction of insights across the entire occupant population in a building. Figure 1 shows a conceptual outline of our method, as applied to a sample dataset of three occupants over three days. For the purposes of this example, the number of components is arbitrarily chosen as 3.

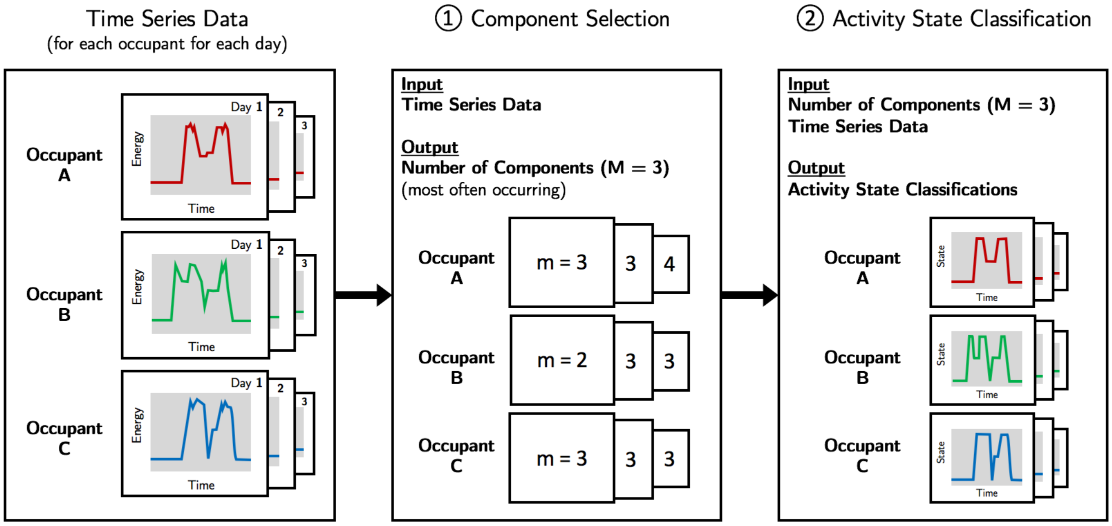


Figure 1: Conceptual outline of our classification method

3.1. Gaussian Mixture Model

We define a time series of plug load energy use measurements collected at the desk-level for each occupant:

$$\mathbf{X}_{i,d} = \{x_1, \dots, x_T\} \tag{1}$$

where i is the occupant index (for all occupants $1, \dots, I$), d is the day index (for all days $1, \dots, D$), and T is the number of time steps in the period of study (T depends on the amount of time between measurements, which can be set as required by the user). For the purposes of this study, we utilize a time step of 15 minutes, resulting in $T = 96$ if the full day is analyzed. Data collected over one day for one occupant, as indexed by i and d (*i.e.*, $\mathbf{X}_{i,d}$) defines a full sub-dataset. Figure 2 shows a histogram of a typical full sub-dataset. Many algorithms can be applied to the classification or clustering of granular (sub-hourly) time series datasets, such as k-means, naive Bayes, and support vector machines [31]. However, for these classifiers to be effective in determining occupant activity states, either they need to be trained on data that describes the true state of occupancy for each occupant in the study, or their parameters need to be tuned on existing data by the user of the method. Collecting this ground truth data for training is possible in a small experimental setting, but as previously stated, it becomes quickly intractable as we move to the analysis of large commercial buildings or even portfolios of buildings.

To derive a method to effectively classify energy states without training data, we looked to domain research on analyzing human activity states and dynamics. Previous work analyzing human heart-rate data has utilized a GMM to classify time series data into states describing some physical activity phenomena [68]. Heart rate measurements taken over the course of a day exhibit multimodality due to the various physiological processes that affect heart rates. We have found plug load energy use data to exhibit similar multimodality, as can be seen in Figure 2, and thus we adopted the GMM as a basis of our method.

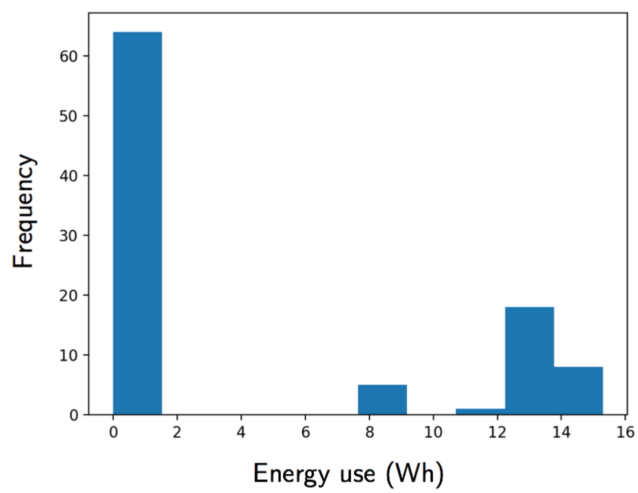


Figure 2: Typical plug load energy use histogram for one occupant over one day.

A GMM is based on the notion that the unimodal Gaussian distribution is useful and common in modeling real world data, and when observed data is clustered around multiple peaks—as is the case both with human heart rate variability and with occupant energy use—this multimodal data can be effectively modeled as a mixture of multiple unimodal Gaussian distributions that may or may not be independent. In a GMM, the likelihood function for observation x is given by:

$$p(x) = \sum_{k=1}^{K} \phi_k \mathcal{N}(x \mid \mu_k, \sigma_k) \quad (2)$$

where K is the number of mixture components and ϕ_k is the weight of each component. Each component follows a normal distribution with mean μ_k and standard deviation σ_k:

$$\mathcal{N}(x \mid \mu_k, \sigma_k) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_k)^2}{2{\sigma_k}^2}\right) \tag{3}$$

The values of each of the component weights (ϕ_k) sum to one so that the total probability distribution is normalized. Each component of the GMM corresponds to a *state* present in the data, with each state being described by the component's Gaussian distribution.

3.2. Variational Bayesian Gaussian Mixture Model

It is common to use Expectation Maximization (EM) to fit each component of a GMM, but this method requires the user of the model to make an *a priori* assumption about the number of components that are used in the GMM, and therefore if adapted to the occupant energy use setting, it would require the user to already know the number of energy use states naturally occurring in the office setting. However, requiring this prior greatly inhibits the extensibility of the model as it may be difficult to discern the number of energy use states prior to collecting data. As a result, we aimed for our model to be adaptable to the natural number of energy use states embedded in the data. We use a variation of the GMM called the Variational Bayesian GMM (VB-GMM) to allow the model to use Bayesian inference to choose the number of components. Recent developments in statistical inference have led to the accessibility of such models [68]. We consider the GMM as described above, and for each observation x_t we include a corresponding latent variable $\mathbf{z}_i$ comprising a 1-of-K binary vector with elements z_{tk} for $k = 1, \dots, K$. The observed dataset for each occupant can be denoted by $\mathbf{X} = \{x_1, \dots, x_T\}$ and the latent variables by $\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_T\}$. Given the component weights $\boldsymbol{\phi}$, we can write the conditional distribution of $\mathbf{Z}$:

$$p(\mathbf{Z}|\boldsymbol{\phi}) = \prod_{t=1}^{T} \prod_{k=1}^{K} \phi_k^{z_{tk}} \tag{4}$$

We can also write the conditional distribution of the observed data, given the latent variables and component weights:

$$p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{t=1}^{T} \prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Lambda_k^{-1})^{z_{tk}} \tag{5}$$

where $\boldsymbol{\mu}$ is the set of component means and $\boldsymbol{\Lambda}$ is the set of component precisions defined as the inverses of the standard deviations. When solving the model, we introduce priors over the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\phi}$. Following common Bayesian statistical practices, we use a Dirichlet distribution over the mixing coefficients $\boldsymbol{\phi}$ and a Gaussian-Wishart prior governing the mean and precision of each component. We utilize the Python scripting language and the Scikit learn package [69] to estimate the parameters of the Variational Bayesian GMM. The model flexibly chooses the number of final components that best describe the data out of a given possible number of components. In addition to the classification results, a key output of a VB-GMM fit is the number of components (n) that are used to classify at least one data point. A full theoretical description of the Variational Bayesian GMM can be found in [70].

3.3. Component selection

The component selection process utilizes the Variational Bayesian GMM applied to time-series plug load data. The overall component selection process is applied to all of the collected plug load data across the space being analyzed. The process utilizes the VB-GMM model by applying it to each *full sub-dataset* independently. Figure 3 shows the flow of the component selection process. First, the set of data points in the period of analysis is defined. Then, a Variational Bayesian GMM is fit to each sub-dataset—that is, to each occupant and each day independently—inferring the number of components (n) present in the sub-dataset for one occupant for one day. As long as there are more days to analyze for an occupant and more occupants in the space being analyzed (more sub-datasets), this process is repeated, until all occupants and days are analyzed. After all occupants and days are analyzed, the number of components most often chosen (the mode of the full set) is determined as M.

In this first step, all of the data points in each day are analyzed for component selection, and we refer to this process as the *primary component selection process*. Additionally, we refer to the number of components chosen by the method in the primary component selection process as M_1. We define three possibilities:

1. $M_1 = 1$. If only one state occurs most often, only one state is present in the data and no further analysis is done.
2. $M_1 = 2$. Often, the VB-GMM selects two states most often in this primary component selection process. Figure 4 shows a typical occurrence of the VB-GMM choosing two components for a full sub-dataset. By observing the data, we note that this classification essentially splits plug load energy use data for each day and occupant into two states: (1) a

state in which very little or no energy is being used at the workstation, and (2) a state in which at least some energy is being used at the workstation. As can be inferred from Figure 4, the higher energy state as classified by the GMM has much more variability than the lower energy state. This typically occurs because it is separating data associated with presence from data associated with absence. Previous analyses of occupant activities have adopted the practice of first separating the data corresponding with occupancy from that corresponding with absence and analyzing the occupancy data separately [65]. Therefore, we integrate this domain knowledge on occupant dynamics into our method: when exactly two components are chosen in the *primary component selection process*, the data classified in the higher energy component for each occupant is separated and a new VB-GMM is applied to this separated data (as long as two or more data points are in the higher component). We define this process as the *secondary component selection process*, and the process in Figure 3 is repeated, this time just for the data classified in the higher energy component. Again, we determine the number of components (n) chosen most often across all occupants and all days, and refer to this value as M_2. After the component selection process, in the subsequent classification step, this two-step process is once again applied.

3. $M_1 > 2$. If three or more components occur most often in the primary VB-GMM, this number of components is used in the classification step.

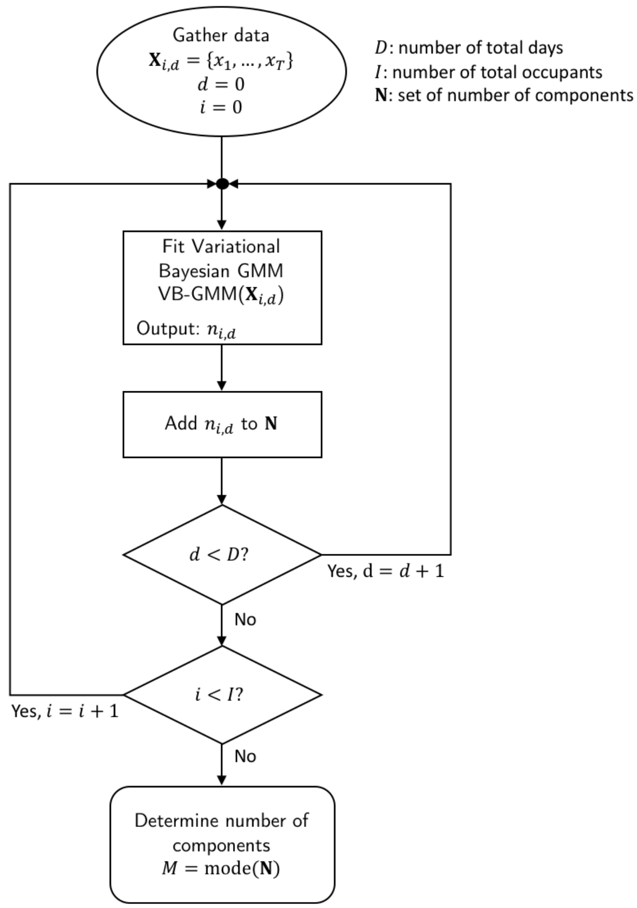


Figure 3: Component selection process using the Variational Bayesian GMM.

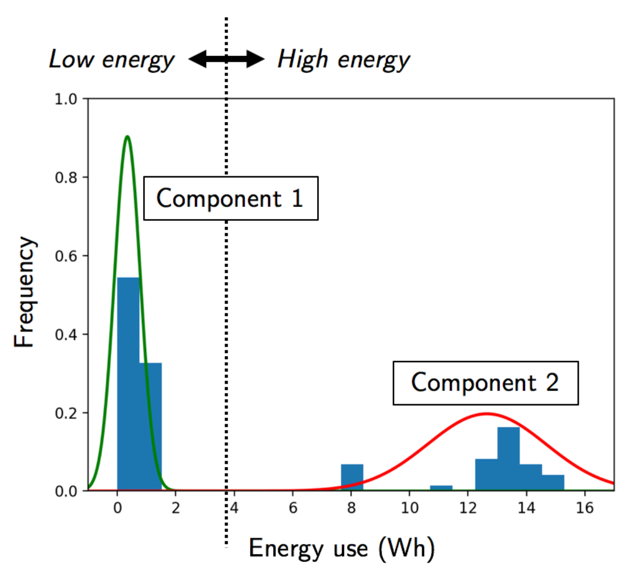


Figure 4: Result of fitting 2-component GMM to all energy use data for one occupant over one day, with dashed line showing where the "Low energy" *and the* "High energy" *components cross.*

3.4. Activity state classification

Once the number of components is inferred through the component selection process, we fit new GMMs to the data so that all occupants are classified into one of the same number of activity states. GMMs are fit independently for each occupant for each day, but all GMMs use the same number of components, as determined in the component selection process. Once a GMM is fit to data, the data points **X** are classified into components: $\{\mathbf{x}_1, \ldots, \mathbf{x}_M\} \in \mathbf{X}$**.** Figure 5 shows the flow for the activity state classification process applied to data from one occupant for one day. If the *primary component selection process* resulted in exactly two components ($M_1 = 2$)—one corresponding to occupant absence ($\mathbf{x}_{low}$) and the other corresponding to occupant presence ($\mathbf{x}_{high}$)—then a GMM is fit with two components to the data and the higher energy component is separated. Then a new GMM is fit to the higher energy data based on the number of components selected in the *secondary component selection process* (M_2). If the *primary component selection process* resulted in three or more components ($M_1 > 2$), a GMM is fit to the original data with that number of components. Once the GMMs have been fit to the data for each occupant for each day separately, the data points are classified into their respective activity states as determined by the GMM fits.

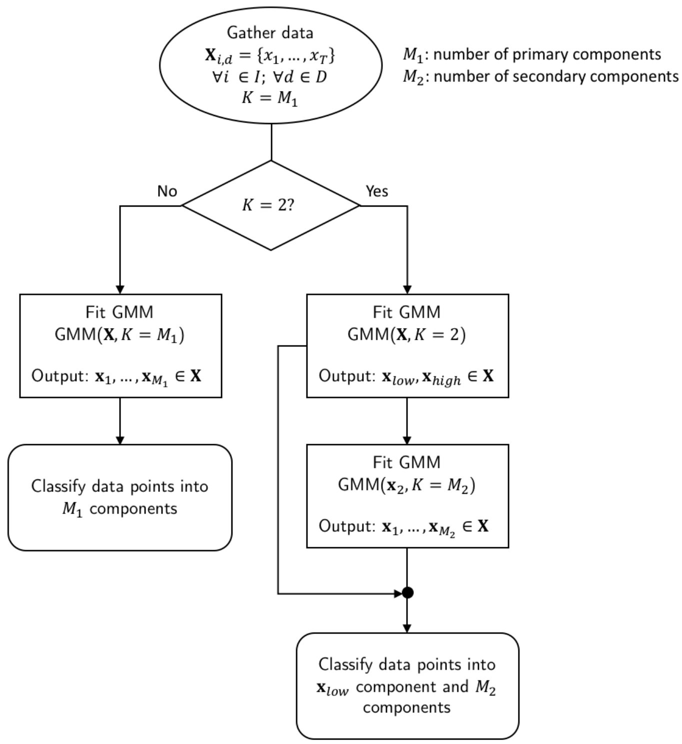


Figure 5: Activity state classification process using the GMM for one occupant and one day.

Figure 6 summarizes the classification process for one occupant and one day in the case where the *primary component selection process* resulted in two components, and the *secondary component selection process* also resulted in two components. In this case, a total of three activity states are utilized to describe the data. Figure 6-a shows a fit of a 2-component GMM to an occupant's energy use data over one day. Any energy use value can be assigned probabilities that it is associated with each component of the GMM. The point between the means of the two components where the probabilities are equal—the point in Figure 6-a at which the probability density function lines cross—becomes the cutoff point between the *lower energy* ($\mathbf{x}_{low}$) and *higher energy* ($\mathbf{x}_{high}$) states: any value below this point is part of the *lower energy* state and any value above this point is part of the *higher energy* state. Each of the occupant's energy use values observed throughout the day is classified according to this system. In the secondary step, the *higher energy* state data is separated out, and another 2-component GMM is fit to just this data, as shown in Figure 6-b. Once again here, the point at which the probability density function lines cross becomes the cutoff point between the two states. As a result of this specific process, three states are determined ($\mathbf{x}_{low}$, $\mathbf{x}_1$, $\mathbf{x}_2$), which can be visualized by the green, yellow, and red density functions in Figure 6.

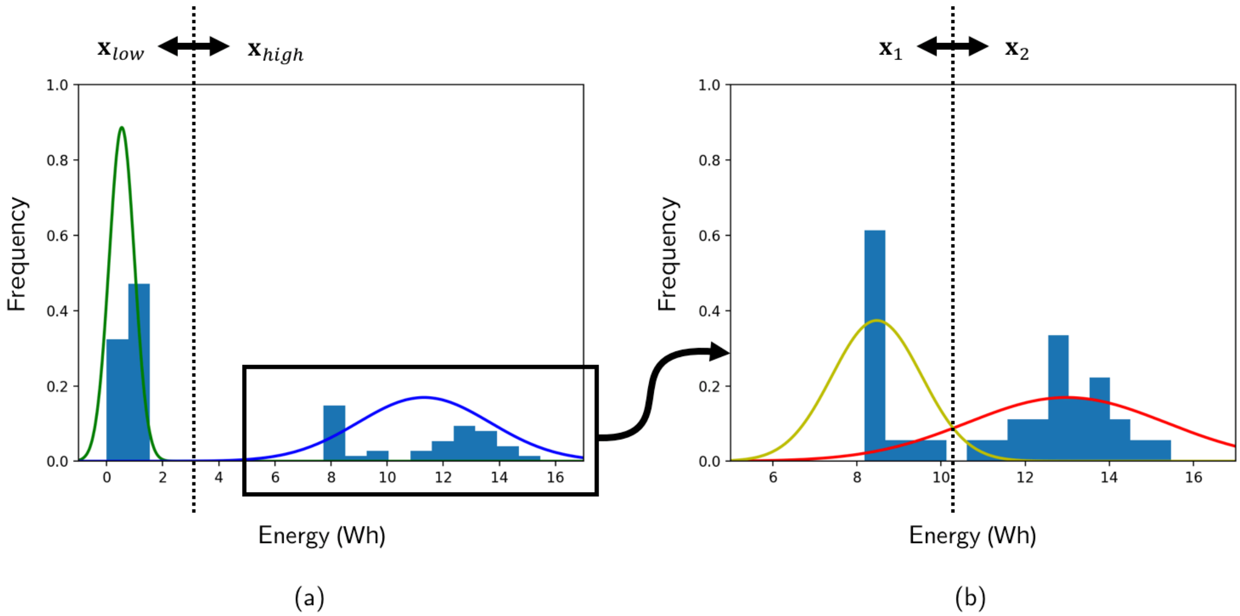


Figure 6: Activity classification process when primary component selection process results in two components. (a) GMM fit to all energy use data from one day. (b) A second GMM fit to just the "higher energy" data from (a).

We design our method to automatically determine when the *secondary component selection process* is necessary in order to provide deeper insights into the data associated with occupant presence. We do this by incorporating explicit domain knowledge into the method. Specifically, the method recognizes when the GMM is only forming two clusters—one associated with absence, and one associated with presence—and it then analyzes the more variable presence data in a second step. If we did not design our method with this functionality to recognize when two steps are necessary, the analysis of the data would be much less insightful, limiting the final number of activity states to two—presence and absence—in some situations. Without incorporating domain knowledge about the typical structure of plug load data signatures, our method would be less effective in describing behavioral patterns of occupants and in turn be limited in its applicability for informing building design and operations.

4. VALIDATION STUDY

In order to validate and benchmark the performance of our proposed energy state classification method, we conducted a two-week study involving seven occupants in an academic office setting. The occupants were five graduate students, a postdoctoral research scholar, and a professor in a total of three offices. Each occupant had his or her own desk. Three graduate students and the postdoctoral research scholar shared one office. Another two graduate students shared another office, and the professor had their own office. We note that validation studies of this size are common in the field of occupant activity recognition in office settings [71]. Typical occupant activities included working on a computer at a desk, going to classes, holding meetings in one of the three offices, and taking lunch, coffee, or social breaks.

4.1. Data collection & analysis

Plug load energy use over 15-minute intervals was monitored and stored using HOBO data loggers connected to a power strip at each occupant's workstation. Each station consisted of a computer plugged into the power strip, and five of the seven workstations had a monitor plugged into the power strips. Participants also were allowed to include some miscellaneous but relatively small power loads, such as cell phone chargers.

To determine how well our classification model captures activity states of occupants, the participants of the validation study manually recorded their activities and the times of their activities over the course of each day. Occupants kept track of the times they arrived and left their desks both at the start/end of their workday and during breaks and meetings throughout the day. This recording of occupant activities formed the "ground truth" data of occupant activity states, with noted activities indicating shifts between the states of occupancy defined in our methodology. To compare the results of our activity classification method with the ground truth data, we counted each instance of an occupant arriving at or leaving his or her desk as an "event" associated with a transition between states. For example, if the occupant notes an arrival at his or her desk at the start of the workday, this event is associated with a shift from a lower energy state to a higher energy state. Similarly, if the occupant notes that he or she has left the desk, this event is associated with a shift from a higher energy state to a lower energy state. Over the course of the two-week study, 345 events were recorded by the seven occupants.

We input the collected plug load energy use values to our classification method and compared each day's state classification results with the ground truth activity data. (The classification process resulted in a total of three states.) If the occupant's recorded event corresponded with a shift between states that aligns with the activity (*e.g.*, a shift from a higher state to a lower state if the occupant leaves his or her workstation), then we consider this event to be correctly classified (*i.e.*, a true positive). If the method shows a shift between energy states, but no event was recorded, we consider this a false positive. And if the occupant records an event but no shift between states was detected by the method, we consider this a false negative. Figure 7 shows the raw data and the occupant activity detection results for one occupant over a workday, with the two most common of these scenarios—true positives and false negatives—depicted by annotation. At 9:07 a.m., the occupant noted a break, but the method did not indicate a shift in occupancy states at this time—hence, we annotate this as a false negative. At 5:03 p.m., the occupant notes that he or she left for the day, and the method indicated a corresponding shift from a higher state to a lower state—hence, we annotate this as a true positive.

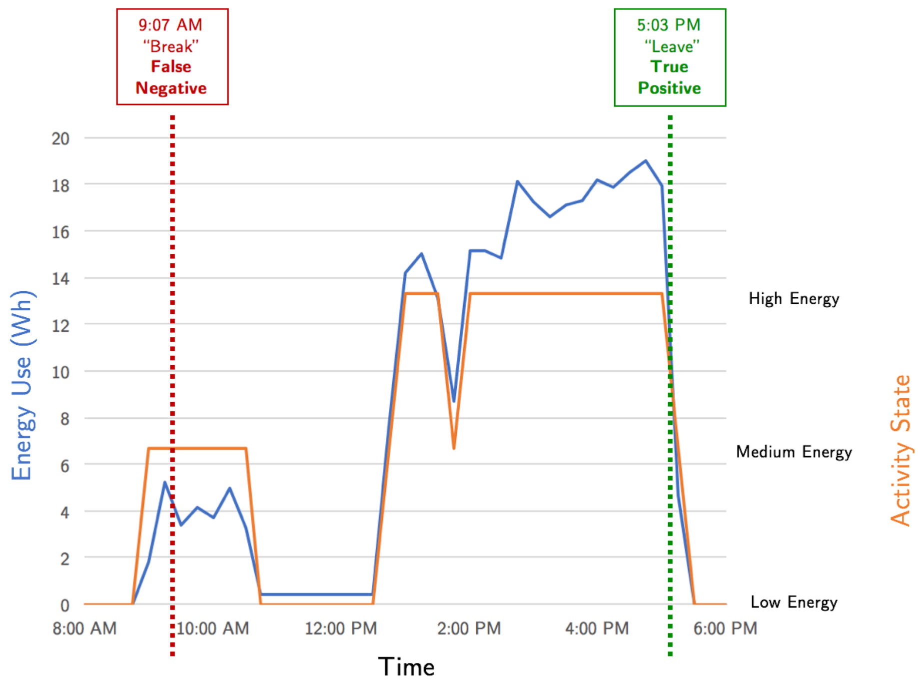


Figure 7: Raw energy use values and classification results, shown with an example of a False Negative as well as a True Positive.

4.2. Validation Results and Discussion

We repeated the analysis shown in Figure 7 for each occupant for each day to understand the effectiveness of our classification method. For building management systems to be able to effectively implement information about the behavior of occupants, a high level of accuracy is required. However, not all measures of accuracy have the same implications. As indicated in [4], situations in which detection sensors or methods determine a state of absence in the building when in fact the building is occupied are more problematic than when they determine a state of occupancy when in fact the building is unoccupied. These situations can lead to lights switching off or a lack of HVAC service to occupants in a room, which can not only cause discomfort, but it can also lead to occupants overriding the intelligent building systems that make use of the information about occupant presence [72]. Consequently, it is extremely important to minimize the rate at which these sensors determine false absence. At the same time, however, if the rate at which sensors determine false occupancy is too high, building managers will not be able to effectively make use of the information, leading to missed opportunities for energy efficiency. Recent work [5,9] has suggested that in order to be effective this false occupancy rate can be only as high as 20%.

To measure the performance of our method in capturing occupant activity shifts, we utilize the precision and recall metrics. The precision metric for the method—the ratio of true positives to all positives (including true and false)—is calculated to be 92.7%. The recall metric—the ratio of true positives to the sum of true positives and false negatives—is calculated to be 73.9%. Both of these metrics suggest that our method is accurate in capturing changes in occupant activity states.

It is important to note the rates at which our method leads to false measures of absence versus false measures of occupancy. We can infer from Figure 7 that our definition of false negatives typically occurs when an occupant indicates a change in state associated with a break from working at their workstation, but the method does not recognize this change in state. (It is also possible that a false negative could be associated with a missed state transition associated with starting the workday or leaving the workday, but these situations occur very rarely.) Therefore, a false negative is typically associated with a false measure of occupancy at the workstation (*i.e.,* the occupant has taken a break but the method has not recognized this break). Alternatively, a false positive indicates that the method has detected a change in state when no state change has been recorded by the occupant. We note that this is not the same type of error as the false negative that is indicated in Figure 7, where the occupant did record an activity, but the algorithm did not detect a change in state. Most often, false positives are associated with the method detecting that the occupant has taken a break, when in fact the occupant is still working at the workstation. Therefore, it is very important to minimize these false positives, because false measures of absence are more problematic than false measures of occupancy, as discussed above. Our results show a very high precision rate of 92.7%, indicating that false measures of the absence of occupants are minimized. At the same time, our recall rate of 73.9% suggests that our method is still providing valuable information about the true state of occupant activities in the building. We also determined the overall false occupancy rate by comparing the output of the method with the occupant notes. We find the overall false occupancy rate to be 5.8%—well below the 20% threshold suggested by the literature [5,9].

Table 1 shows the precision and recall metrics for each occupant in the validation study. As shown in the table, Occupant 7 has the lowest values for both precision (84%) and recall (44%). We hypothesize the precision and recall discrepancies occur because Occupant 7 has atypically low variability in actual energy use. The standard deviation of energy use for Occupant 7 for medium and high energy states was 2.82 Wh, whereas the standard deviation of energy use for the other occupants ranged between 4.41 and 9.74 Wh. Therefore, in raw energy consumption terms, the energy consumed by Occupant 7 was less likely to change significantly in correlation with a noted event. On days where the method performed abnormally poorly (*i.e.*, more than 50% of noted events were false positives), the average standard deviation of energy use values was 3.56 Wh. On days where the method performed relatively well (*i.e.*, less than 25% of events were false positives), the average standard deviation of energy use values was 6.79 Wh. This indicates that higher variability in energy use values is correlated with better performance of the model.

Table 1: Precision and Recall metrics for each occupant in the validation study.

Occupant	Precision	Recall	False Occupancy Rate
1	100%	77%	5.8%
2	93%	85%	2.6%
3	93%	80%	1.6%
4	91%	82%	4.2%
5	99%	60%	17.1%
6	97%	81%	2.1%
7	84%	44%	19.8%
All Occupants	*92.7%*	*73.9%*	*5.8%*

5. CASE STUDY: SAN FRANCISCO OFFICE BUILDING

We applied our method to data collected in a typical office building in San Francisco, CA, USA in order to demonstrate the merits of our methodology for informing energy efficient operations and space utilization of a commercial building. The occupants were employees of an organization headquartered in San Francisco. Based on observations during a site-visit, we found that activities in the space were comprised of typical office work at a computer workstation, meetings, and breaks from working (including lunch and social breaks).

5.1. Data collection

We utilized plug load data collected for 47 occupants in an open-office building in San Francisco, CA, USA using Enmetric plug load sensors [73]. The sensors continuously collected and reported total energy use over 15-minute periods at the individual desk level. By manual inspection of the data, we found that some of the sensors had connectivity issues, and that on certain days, these sensors did not report any energy consumption at all. We therefore limited our analysis of this data to a clean segment of the data spanning 41 continuous days. We note that our method could be applied to data outside this time span, but that more occupants would simply be classified into a very low energy state throughout the day. For the sake of brevity and clarity, we choose to concentrate our analysis and discussion the 41-day period that had consistent data.

5.2. Case study results discussion

Following the methodology described in section 3 above, we first performed the component selection step and then used the results to classify the data into activity states. During the component selection process, each Variational Bayesian GMM was limited to using up to 5 components to classify the sub-dataset. With a total of 1,927 sub-datasets, the Variational Bayesian GMM was applied to each sub-dataset, and the majority of the models (51.3%) chose 2 components to classify the data, as shown in Figure 8. None of the models chose more than three components to classify the data. After determining that the Variational Bayesian GMM most often chose 2 components for this dataset, we fit another GMM to the plug load energy use data for each occupant and for each day, using exactly 2 components for each GMM. This process allows for the classification of each 15-minute period of occupant activity into one of 2 states: *low energy* and *high energy*. Because exactly two components were chosen in the primary component selection process, the data classified into the *high energy* state was separated and the secondary component selection process was performed.

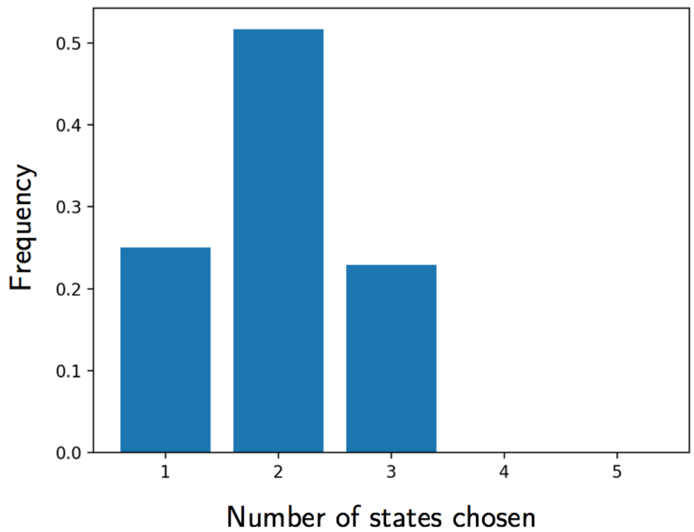


Figure 8: Number of states chosen (out of 5) by Variational Bayesian GMM for all energy use values for each occupant for each day.

Another Variational Bayesian GMM was fit to each day's *high energy* data for each occupant, and once again the majority of the models (58.7%) chose 2 components, as shown in Figure 9. As long as the initial *low/high* energy state classification results in at least two data points classified as *high energy*, we can then fit another 2-component GMM to the data points initially classified as *high energy*, reclassifying the *high energy* state into two states: *medium energy* and *high energy.* If one or fewer data points is initially classified as *high energy*, our method does not perform the second classification into *high* and *medium* energy states.

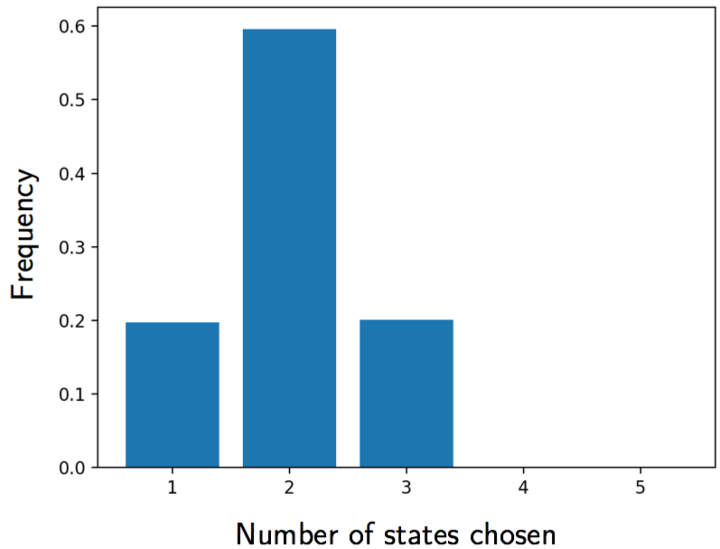


Figure 9: Number of states chosen (out of 5) by Variational Bayesian GMM for just the "high energy" *energy use values for each occupant for each day.*

Our method resulted in a total of three possible states for each 15-minute period for each day for each occupant, which we delineate as *low energy*, *medium energy*, and *high energy* states. By allowing the Variational Bayesian GMM to adaptively fit to the plug load energy use data for each day and occupant, we infer the natural number of states that describe the overall dataset. The result of this analysis is a mapping from raw energy use values to states of occupant activities in the building. Figure 10 illustrates this mapping for each occupant over the floor plan of the building. As Figure 10-a shows, it can be difficult to understand the meaning of the raw energy use values, since a value of, for example, 5Wh over the 15-minute period can mean very different things for two different occupants. However, as Figure 10-b shows, the mapping into activity states abstracts information about these energy use values, providing deeper insights into the activities of occupants across the floorplan of the building.

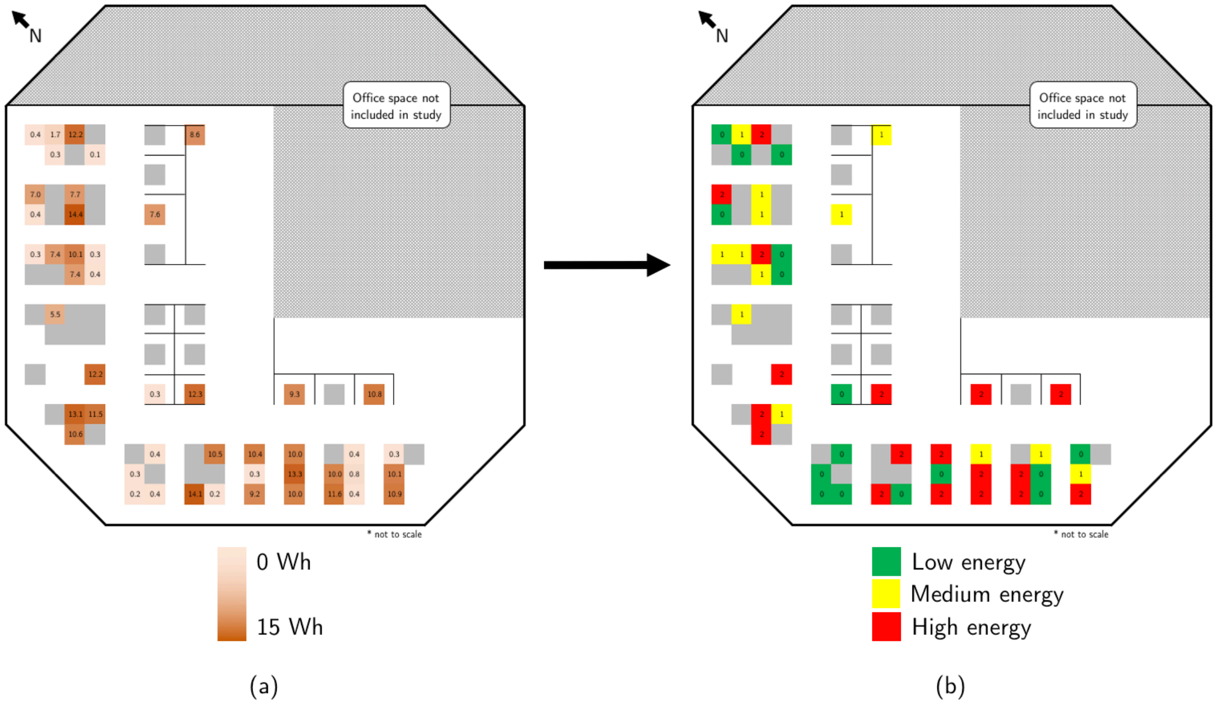


Figure 10: Mapping from raw energy use values (a) to occupant activity states (b) for each occupant over one 15-minute period

A potentially useful application of this methodology is to aid in understanding nuanced aspects of the energy efficient operations and space utilization of buildings to which it is applied. Figure 11 shows the progression of occupancy states for each of the 47 occupants over one afternoon/evening in the study period, from the 15-minute period ending at 5:00pm to the 15-minute period ending at 7:15pm (10 time-steps in total). Monitoring this progression provides insight into how the space utilization in this office setting changes over time. This information provides value for building managers who seek to ensure that spaces are being utilized efficiently. Furthermore, as building designers and engineers embrace the notion of *performance based design*, they are beginning to adapt the programming phase of the design process to include more nuanced understandings of space utilization. Monitoring and visualizing occupant activity states in buildings can help in the development of space use frameworks that more closely align with the true states of occupant activities in buildings.

As building lighting and HVAC systems continue to be installed with more granular control over spatial and temporal dimensions, building managers can also use this information to optimize the control of these building systems to reduce the amount of energy required to provide services to their occupants. Visualizing this information also provides for the opportunity to make recommendations for co-optimization between occupants, space, and building systems. For example, if groups of occupants who do not have desks near each other regularly shift to *low energy* states at the same time, these occupants could potentially be relocated to be physically near each other so that building systems can reduce services like lighting and HVAC in the space they occupy. An example of such a realignment strategy is depicted in Figure 12. Here, occupants within the blue circles are identified as occupants that shift from a higher state to a lower state from 2:00pm to 2:30pm on a specific day in the study period, perhaps for a meeting or to take a break at the same time (Figure 12-a). If this pattern recurs commonly in building, one potential strategy would be for these occupants to move to workstations that are physically near one another Figure 12-b). After realignment, lighting and HVAC systems could adjust to the change in occupancy states at the identified workstations.

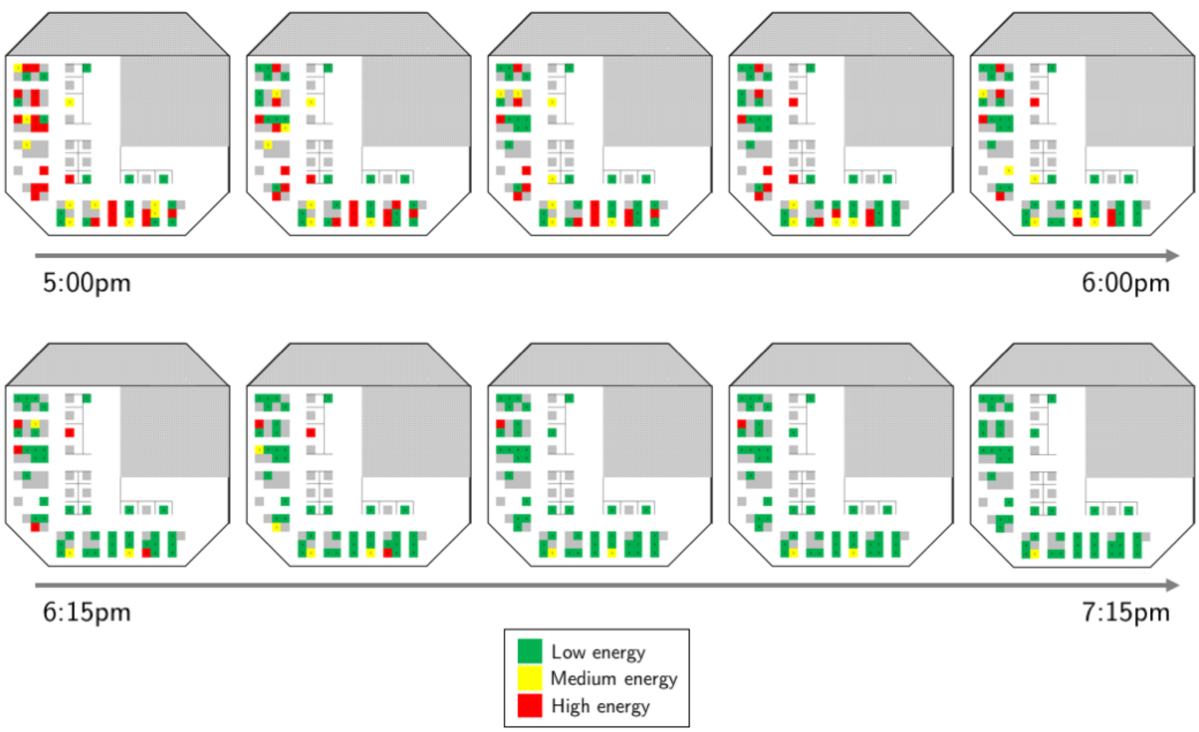


Figure 11: Visualization of the changing space-use levels over time using activity state classification method.

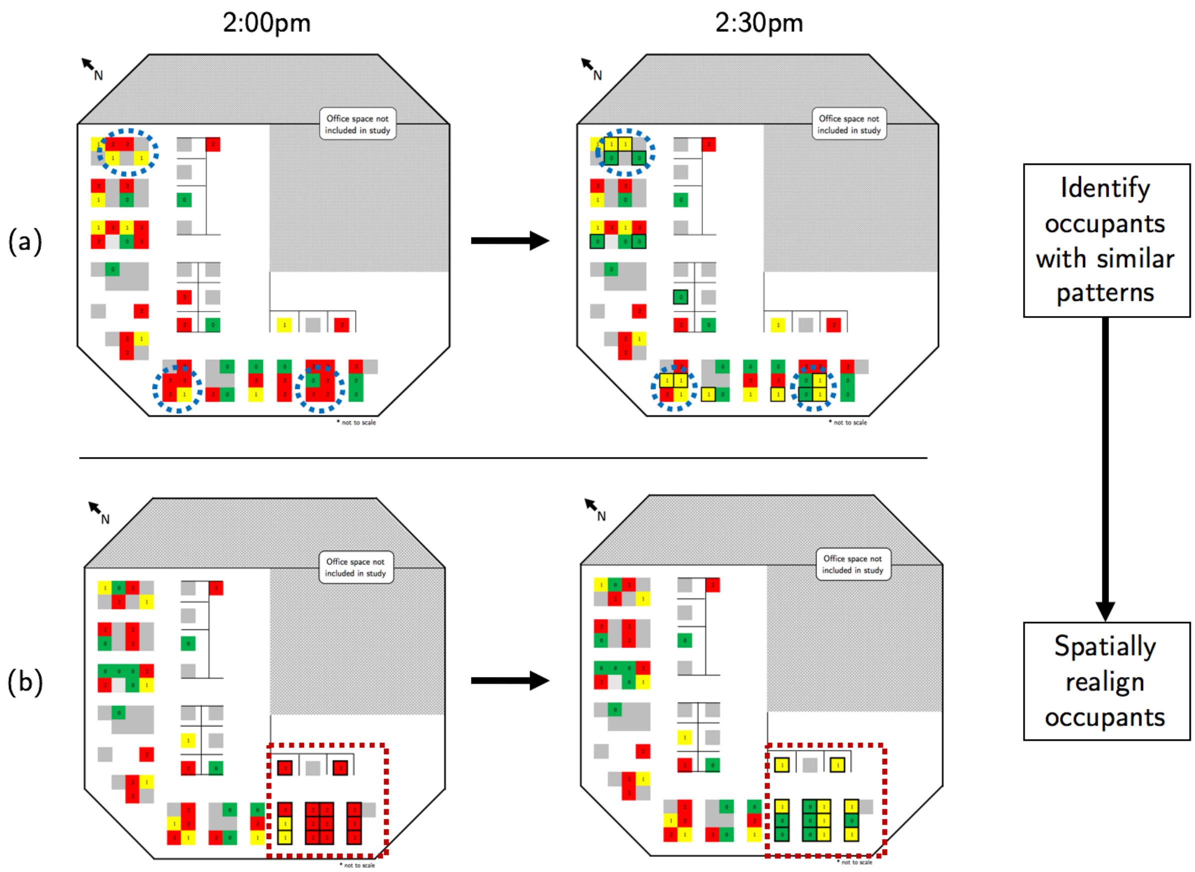


Figure 12: Potential occupant realignment strategy. (a) Occupants with similar patterns are identified. (b) After realignment, intelligent building systems can take advantage of activity state shifts.

The analyses presented in Figure 11 and Figure 12 demonstrate the ability of our model to provide new knowledge to building managers and designers. Without requiring training data or an *a priori* assumption about the number of activity states in the plug load data, we are able to glean effective insights about occupant activities in the space. Furthermore, the visual representation of the occupancy states across the floorplan offers new insights that could not otherwise be interpreted from the raw plug load energy use values. As Akbas et al. [55] notes, effective visual representation of new spatio-temporal information can be an effective decision-support tool for managers and designers. As a result, our method has the ability to help building operators make decisions for energy efficient system management and to help designers build models of occupant activities for improved design of future spaces.

6. LIMITATIONS AND FUTURE WORK

The main limitations of our method stem from the inherent constraints of plug load energy use data to capture activities of occupants in buildings. While the method performs well enough to provide valuable information to building engineers, operators, and designers—based on suggested precision metrics from the literature—there is opportunity to further improve the precision. For example, while plug load energy use data typically changes when occupants take extended breaks from their workstations, there are situations in which plug loads stay high while occupants take short breaks from their workstations. Future work could incorporate the use of other sources of data, such as infrared sensors, in order to complement the plug load data collected for our method. A composite data stream that includes multiple sources could lead to more precise detection of occupant activities.

In addition to possible improvements in accuracy, future work could consider identifying occupant activities that are not associated with plug load energy use. Plug load sensors are cost-effective for this task and often easily accessible, since they are commonly installed in office buildings for various purposes beyond inferring activity patterns, such as for monitoring the energy consumption of miscellaneous equipment. While their data provide a good proxy for occupant activities, a more robust understanding of occupant behavior should include non-energy-intensive activities. Again, such activities could be recognized using data streams that are complementary to the plug load energy use data. Similarly, the method developed in this paper could be extended for the analysis of analogous data produced from other sensors, in particular when time series data exhibits multimodality and domain knowledge about the states associated with the components of the distribution is known.

It is important to note that while the validation study discussed in Section 4 shows reasonable reliability and internal validity, claims about its external validity must be made with caution. While our inference method is designed to be able to adapt to individualized settings, where different occupants have different baseline, average, and maximum plug load energy consumption, further studies are needed to determine the validity of the classification results in settings beyond this internal validation study. Our internal validation study demonstrates the robustness and adaptability of our method within the setting of the study, but future studies with large-scale ground truth data collection are needed for broader claims about the true scalability of our method.

One exciting area of future work involves utilizing this data to gain insight into the natural structure of the occupant network in the building. Our method provides information about states of occupant activities in the building, which can be useful in understanding not just individual activity states, but also the relationships among occupants. For example, two occupants that have very similar patterns of activities in buildings could be highly related socially or organizationally. Building managers and designers could make use of this information by potentially suggesting shifts in the occupant layout in the building, allowing building systems to be more closely aligned with the states of occupancy across the building. Furthermore, by gaining an understanding of the structure of the network of occupants, eco-feedback systems that attempt to convince occupants to adjust their behavior could become more effective, as the network structure of occupants has been shown to have high importance in these strategies [37,43].

7. CONCLUSIONS

The main contributions of this work are twofold: first, to introduce a new adaptable method that integrates knowledge-based and data-driven approaches for inferring occupant dynamics in building; and second, to demonstrate how our proposed method can be utilized to infer the occupant dynamics occurring in a building and inform intelligent optimization strategies for energy efficiency and space utilization. By integrating a variational Bayesian version of the Gaussian mixture model with explicit domain knowledge about occupant dynamics and plug load data signatures, we designed our method to require no ground truth data to perform with a high level of accuracy. These methodological design decisions allow our method to be more easily applied to situations where ground truth data is difficult to collect, such as when there are many occupants across one or more buildings.

In analyzing newly accessible plug load data streams to infer occupant activity states in buildings, we can gain a deeper understanding of the complexity of occupant dynamics at a high level of spatial and temporal granularity. In turn, this deeper understanding translates to new knowledge about occupant dynamics that can help building designers, engineers, and managers better understand how occupants respond to the spaces they occupy. These decision-makers will now be armed with the knowledge that can enable them to intelligently manage building operations and design to enhance energy efficiency, space utilization, and occupant satisfaction. Buildings will inevitably continue to play a crucial role in each of our lives as occupants and in the world’s sustainable energy future. New methods that combine the extant knowledge of occupant dynamics and building systems with emerging data-driven methods could provide us with the necessary insights to design, operate, and manage the next generation of high-performance buildings.

ACKNOWLEDGEMENTS

The material presented in this manuscript is based in part upon work supported by a Stanford Graduate Fellowship, Terman Faculty Fellowship, the Center for Integrated Facility Engineering and the US National Science Foundation (NSF) under Grant No. 1461549. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

REFERENCES

[1] N.E. Klepeis, W.C. Nelson, W.R. Ott, J.P. Robinson, A.M. Tsang, P. Switzer, J. V Behar, S.C. Hern, W.H. Engelmann, The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants, J. Expo. Anal. Environ. Epidemiol. 11 (2001) 231–252. doi:10.1038/sj.jea.7500165.

[2] T. Keeling, D. Clements-Croome, R. Luck, P. Pointer, How the sensory experience of buildings can contribute to wellbeing and productivity, in: Proc. 7th Wind. Conf. Chang. Context Comf. an Unpredictable World, Network for Comfort and Energy Use in Buildings, Windsor, UK, 2012. http://nceub.org.uk (accessed August 30, 2017).

[3] O. Lucon, D. Ürge-Vorsatz, A. Zain Ahmed, H. Akbari, P. Bertoldi, L.F. Cabeza, N. Eyre, A. Gadgil, L.D.D. Harvey, Y. Jiang, E. Liphoto, S. Mirasgedis, S. Murakami, J. Parikh, C. Pyke, M.V. Vilariño, Buildings, 2014. doi:10.2753/JES1097-203X330403.

[4] W. Shen, G. Newsham, B. Gunay, Leveraging existing occupancy-related data for optimal control of commercial office buildings: A review, Adv. Eng. Informatics. (2017). doi:10.1016/j.aei.2016.12.008.

[5] H.B. Gunay, A. Fuller, W. O'Brien, I. Beausoleil-Morrison, Detecting Occupants' Presence in Office Spaces: A Case Study, in: eSIM Conf. Proc., Hamilton, ON, 2016.

[6] Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz, A Multi-sensor Based Occupancy Estimation Model for Supporting Demand Driven HVAC Operations, Proc. 2012 Symp. Simul. Archit. Urban Des. (2012) 2:1--2:8. doi:10.1061/9780784412329.146.

[7] M. Jin, R. Jia, Z. Kang, I.C. Konstantakopoulos, C.J. Spanos, PresenceSense: zero-

training algorithm for individual presence detection based on power monitoring, in: Proc. 1st ACM Conf. Embed. Syst. Energy-Efficient Build. - BuildSys ’14, ACM Press, New York, New York, USA, 2014: pp. 1–10. doi:10.1145/2674061.2674073.

[8] A. Khan, J. Nicholson, S. Mellor, D. Jackson, K. Ladha, C. Ladha, J. Hand, J. Clarke, P. Olivier, T. Plötz, Occupancy monitoring using environmental & context sensors and a hierarchical analysis framework, in: Proc. 1st ACM Conf. Embed. Syst. Energy-Efficient Build., ACM, 2014: pp. 90–99. doi:10.1145/2674061.2674080.

[9] V.L. Erickson, Y. Lin, A. Kamthe, R. Brahme, A. Surana, A.E. Cerpa, M.D. Sohn, S. Narayanan, Energy efficient building environment control strategies using real-time occupancy measurements, in: Proc. First ACM Work. Embed. Sens. Syst. Energy-Efficiency Build. - BuildSys ’09, ACM Press, New York, New York, USA, 2009: p. 19. doi:10.1145/1810279.1810284.

[10] R.V. Andersen, J. Toftum, K.K. Andersen, B.W. Olesen, Survey of occupant behaviour and control of indoor environment in Danish dwellings, Energy Build. 41 (2009) 11–16. doi:10.1016/j.enbuild.2008.07.004.

[11] A. Ingle, M. Moezzi, L. Lutzenhiser, R. Diamond, Better home energy audit modelling: incorporating inhabitant behaviours, Build. Res. Inf. 42 (2014) 409–421. doi:10.1080/09613218.2014.890776.

[12] H.B. Gunay, W. O’Brien, I. Beausoleil-Morrison, A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices, Build. Environ. 70 (2013) 31–47. doi:10.1016/j.buildenv.2013.07.020.

[13] M. Neges, C. Koch, M. König, M. Abramovici, Combining visual natural markers and IMU for improved AR based indoor navigation, Adv. Eng. Informatics. 31 (2017) 18–31. doi:10.1016/J.AEI.2015.10.005.

[14] R. Akhavian, A.H. Behzadan, Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers, Adv. Eng. Informatics. 29 (2015) 867–877. doi:10.1016/J.AEI.2015.03.001.

[15] T. Weng, Y. Agarwal, From Buildings to Smart Buildings—Sensing and Actuation to Improve Energy Efficiency, IEEE Des. Test Comput. 29 (2012) 36–44. doi:10.1109/MDT.2012.2211855.

[16] H. Doukas, K.D. Patlitzianas, K. Iatropoulos, J. Psarras, Intelligent building energy

management system using rule sets, Build. Environ. 42 (2007) 3562–3569. doi:10.1016/j.buildenv.2006.10.024.

[17] V. Singhvi, A. Krause, C. Guestrin, J.H. Garrett, H.S. Matthews, Intelligent light control using sensor networks, in: Proc. 3rd Int. Conf. Embed. Networked Sens. Syst. - SenSys ’05, ACM Press, New York, New York, USA, 2005: p. 218. doi:10.1145/1098918.1098942.

[18] D.H.W. Li, K.L. Cheung, S.L. Wong, T.N.T. Lam, An analysis of energy-efficient light fittings and lighting controls, Appl. Energy. 87 (2010) 558–567. doi:10.1016/j.apenergy.2009.07.002.

[19] X. Guo, D. Tiller, G. Henze, C. Waters, The performance of occupancy-based lighting control systems: A review, Light. Res. Technol. 42 (2010) 415–431. doi:10.1177/1477153510376225.

[20] T. Weng, B. Balaji, S. Dutta, R. Gupta, Y. Agarwal, Managing plug-loads for demand response within buildings, in: Proc. Third ACM Work. Embed. Sens. Syst. Energy-Efficiency Build. - BuildSys ’11, ACM Press, New York, New York, USA, 2011: p. 13. doi:10.1145/2434020.2434024.

[21] A.C. Menezes, A. Cripps, D. Bouchlaghem, R. Buswell, Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap, Appl. Energy. 97 (2012) 355–364. doi:10.1016/j.apenergy.2011.11.075.

[22] A. Roetzel, A. Tsangrassoulis, U. Dietrich, Impact of building design and occupancy on office comfort and energy performance in different climates, Build. Environ. 71 (2014) 165–175. doi:10.1016/j.buildenv.2013.10.001.

[23] A.D. Galasiu, Energy saving lighting control systems for open-plan offices: a field study, Leukos. 4 (2007). http://nparc.cisti-icist.nrc-cnrc.gc.ca/eng/view/object/?id=b23dfd7d-3280-4740-b2fa-c57cd48806e9 (accessed July 18, 2017).

[24] B. Roisin, M. Bodart, A. Deneyer, P. D ’herdt, Lighting energy savings in offices using different control systems and their real consumption, Energy Build. 40 (2008) 514–523. doi:10.1016/j.enbuild.2007.04.006.

[25] A. Williams, B. Atkinson Pe, K.G. Phd, E.P. Pe, F. Rubinstein Fies, Lighting Controls in Commercial Buildings, Leukos. 8 (2012) 161–180.

doi:10.1582/LEUKOS.2012.08.03.001.

[26] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings, in: Proc. 11th ACM Conf. Embed. Networked Sens. Syst. - SenSys ’13, ACM Press, Rome, Italy, 2013: pp. 1–14. doi:10.1145/2517351.2517370.

[27] B. Dong, K.P. Lam, Building energy and comfort management through occupant behaviour pattern detection based on a large-scale environmental sensor network, J. Build. Perform. Simul. 4 (2011) 359–369. doi:10.1080/19401493.2011.577810.

[28] A. Roetzel, Occupant behaviour simulation for cellular offices in early design stages—Architectural and modelling considerations, Build. Simul. 8 (2015) 211–224. doi:10.1007/s12273-014-0203-6.

[29] O.T. Masoso, L.J. Grobler, The dark side of occupants’ behaviour on building energy use, Energy Build. 42 (2010) 173–177. doi:10.1016/j.enbuild.2009.08.009.

[30] P. Raftery, M. Keane, A. Costa, Calibrating whole building energy models: Detailed case study using hourly measured data, Energy Build. 43 (2011) 3666–3679. doi:10.1016/j.enbuild.2011.09.039.

[31] J. Zhao, B. Lasternas, K.P. Lam, R. Yun, V. Loftness, Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining, Energy Build. 82 (2014) 341–355. doi:10.1016/j.enbuild.2014.07.033.

[32] M. Bonte, F. Thellier, B. Lartigue, Impact of occupant’s actions on energy building performance and thermal sensation, Energy Build. 76 (2014) 219–227. doi:10.1016/j.enbuild.2014.02.068.

[33] L.K. Norford, R.H. Socolow, E.S. Hsieh, G. V. Spadaro, Two-to-one discrepancy between measured and predicted performance of a “low-energy” office building: insights from a reconciliation based on the DOE-2 model, Energy Build. 21 (1994) 121–131. doi:10.1016/0378-7788(94)90005-1.

[34] S. Taherian, M. Pias, G. Coulouris, J. Crowcroft, Profiling energy use in households and office spaces, in: Proc. 1st Int. Conf. Energy-Efficient Comput. Netw., ACM, 2010: pp. 21–30. http://dl.acm.org/citation.cfm?id=1791318 (accessed February 7, 2012).

[35] M. Jia, R.S. Srinivasan, A.A. Raheem, From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation

coupling mechanisms for building energy efficiency, Renew. Sustain. Energy Rev. 68 (2017) 525–540. doi:10.1016/j.rser.2016.10.011.

[36] A.J. Sonta, R.K. Jain, R. Gulbinas, J.M.F. Moura, J.E. Taylor, OESPG: Computational Framework for Multidimensional Analysis of Occupant Energy Use Data in Commercial Buildings, J. Comput. Civ. Eng. 31 (2017) 4017017. doi:10.1061/(ASCE)CP.1943-5487.0000663.

[37] K. Anderson, S. Lee, C.C. Menassa, Impact of Social Network Type and Structure on Modeling Normative Energy Use Behavior Interventions, J. Comput. Civ. Eng. 28 (2014) 30–39. doi:10.1061/(ASCE)CP.1943-5487.0000314.

[38] X. Xu, J.E. Taylor, A.L. Pisello, Network synergy effect: Establishing a synergy between building network and peer network energy conservation effects, Energy Build. 68 (2014) 312–320. doi:10.1016/j.enbuild.2013.09.017.

[39] H. Staats, E. van Leeuwen, A. Wit, A longitudinal study of informational interventions to save energy in an office building., J. Appl. Behav. Anal. 33 (2000) 101–104. doi:10.1901/jaba.2000.33-101.

[40] H. Staats, P. Harland, H.A.M. Wilke, Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household, Environ. Behav. 36 (2004) 341–367. doi:10.1177/0013916503260163.

[41] S.C. Staddon, C. Cycil, M. Goulden, C. Leygue, A. Spence, Intervening to change behaviour and save energy in the workplace: A systematic review of available evidence, Energy Res. Soc. Sci. 17 (2016) 30–51. doi:10.1016/j.erss.2016.03.027.

[42] A. Khosrowpour, Y. Xie, J.E. Taylor, Y. Hong, One size does not fit all: Establishing the need for targeted eco-feedback, Appl. Energy. 184 (2016) 523–530. doi:10.1016/j.apenergy.2016.10.036.

[43] R.K. Jain, J.E. Taylor, G. Peschiera, Assessing eco-feedback interface usage and design to drive energy efficiency in buildings, Energy Build. 48 (2012) 8–17. doi:10.1016/j.enbuild.2011.12.033.

[44] M. Mourshed, S. Shikder, A.D.F. Price, Phi-array: A novel method for fitness visualization and decision making in evolutionary design optimization, Adv. Eng. Informatics. 25 (2011) 676–687. doi:10.1016/j.aei.2011.07.005.

[45] J.H. Jo, J.S. Gero, Space layout planning using an evolutionary approach, Artif. Intell.

Eng. 12 (1998) 149–162. http://ac.els-cdn.com/S095418109700037X/1-s2.0-S095418109700037X-main.pdf?_tid=a2a8c5a8-931a-11e7-871b-00000aacb362&acdnat=1504713009_a3be26a0231c1d29633a827b88e4de22 (accessed September 6, 2017).

[46] R.-J. Dzeng, W.-C. Wang, F.-Y. Hsiao, Y.-Q. Xie, An Activity-Based Simulation Model for Assessing Function Space Assignment for Buildings: A Service Performance Perspective, Comput. Civ. Infrastruct. Eng. 30 (2015) 935–950. doi:10.1111/mice.12177.

[47] G. Suter, F. Petrushevski, M. Šipetić, Operations on network-based space layouts for modeling multiple space views of buildings, Adv. Eng. Informatics. 28 (2014) 395–411. doi:10.1016/j.aei.2014.06.004.

[48] M. Bhatt, C. Schultz, Minqian Huang, The shape of empty space: Human-centred cognitive foundations in computing for spatial design, in: 2012 IEEE Symp. Vis. Lang. Human-Centric Comput., IEEE, 2012: pp. 33–40. doi:10.1109/VLHCC.2012.6344477.

[49] M. Bhatt, C. Schultz, People-centered spatial design: On visuospatial cognition in the built environment, in: Int. Conf. Hum. Behav. Des., Ascona, Switzerland, 2014. doi:10.13140/2.1.4184.8327.

[50] C. Schultz, M. Bhatt, A. Borrmann, Bridging qualitative spatial constraints and feature-based parametric modelling: Expressing visibility and movement constraints, Adv. Eng. Informatics. 31 (2017) 2–17. doi:10.1016/j.aei.2015.10.004.

[51] M. Bhatt, J. Hois, O. Kutz, F. Dylla, Modelling Functional Requirements in Spatial Design, in: Proc. 29th Int. Conf. Concept. Model., Springer-Verlag, Berlin, Heidelberg, 2010: pp. 464–470. http://dl.acm.org/citation.cfm?id=1929757.1929804.

[52] R.-J. Dzeng, C.-W. Lin, F.-Y. Hsiao, Application of RFID tracking to the optimization of function-space assignment in buildings, Autom. Constr. 40 (2014) 68–83. doi:10.1016/j.autcon.2013.12.011.

[53] J.C. Vischer, Towards a user-centred theory of the built environment, Build. Res. Inf. 36 (2008) 231–240. doi:10.1080/09613210801936472.

[54] T.W. Kim, R. Rajagopal, M. Fischer, C. Kam, A knowledge-based framework for automated space-use analysis, Autom. Constr. 32 (2013) 165–176. doi:10.1016/j.autcon.2012.08.002.

[55] R. Akbas, C. Clevenger, J. Haymaker, Temporal Visualization of Building Occupancy

Phase, in: Proc. Int. Work. Comput. Civ. Eng., American Society of Civil Engineers, Reston, VA, 2007: pp. 208–215. doi:10.1061/40937(261)26.

[56] J. Dijkstra, H. Timmermans, Towards a multi-agent model for visualizing simulated user behavior to support the assessment of design performance, Autom. Constr. 11 (2002) 135–145. doi:10.1016/S0926-5805(00)00093-5.

[57] A. Tomé, M. Kuipers, T. Pinheiro, M. Nunes, T. Heitor, Space–use analysis through computer vision, Autom. Constr. 57 (2015) 80–97. doi:10.1016/j.autcon.2015.04.013.

[58] R. Melfi, B. Rosenblum, B. Nordman, K. Christensen, Measuring building occupancy using existing network infrastructure, in: 2011 Int. Green Comput. Conf. Work. IGCC 2011, 2011. doi:10.1109/IGCC.2011.6008560.

[59] S. Meyn, A. Surana, Y. Lin, S.M. Oggianu, S. Narayanan, T.A. Frewen, A sensor-utility-network method for estimation of occupancy in buildings, in: Proc. 48h IEEE Conf. Decis. Control Held Jointly with 2009 28th Chinese Control Conf., IEEE, 2009: pp. 1494–1500. doi:10.1109/CDC.2009.5400442.

[60] M. Trivedi, K. Huang, I. Mikic, Intelligent environments and active camera networks, IEEE Int. Conf. Syst. Man Cybern. 2 (2000) 804–809. doi:10.1109/ICSMC.2000.885948.

[61] C. Kleine-Cosack, M.H. Hennecke, S. Vajda, G.A. Fink, Exploiting Acoustic Source Localization for Context Classification in Smart Environments, in: Springer, Berlin, Heidelberg, 2010: pp. 157–166. doi:10.1007/978-3-642-16917-5_16.

[62] J. Krukar, R.C. Dalton, C. Hölscher, Applying HCI Methods and Concepts to Architectural Design (Or Why Architects Could Use HCI Even If They Don’t Know It), in: Springer, Cham, 2016: pp. 17–35. doi:10.1007/978-3-319-30028-3_2.

[63] M. Amayri, A. Arora, S. Ploix, S. Bandhyopadyay, Q.-D. Ngo, V.R. Badarla, Estimating occupancy in heterogeneous sensor environment, Energy Build. 129 (2016) 46–58. doi:10.1016/j.enbuild.2016.07.026.

[64] A. Arora, M. Amayri, V. Badarla, S. Ploix, S. Bandyopadhyay, Occupancy Estimation Using Non Intrusive Sensors in Energy Efficient Buildings, in: Proc. 14th Conf. Int. Build. Perform. Simul. Assoc., Hyderabad, India, 2015: pp. 1441–1448. http://www.ibpsa.org/proceedings/BS2015/p2460.pdf (accessed July 17, 2017).

[65] M. Milenkovic, O. Amft, An opportunistic activity-sensing approach to save energy in office buildings, Proc. Fourth Int. Conf. Futur. Energy Syst. - E-Energy ’13. (2013) 247.

doi:10.1145/2487166.2487194.

[66] S. Poll, C. Teubert, Pilot Study of a Plug Load Management System: Preparing for Sustainability Base, in: 2012 IEEE Green Technol. Conf., IEEE, 2012: pp. 1–6. doi:10.1109/GREEN.2012.6200945.

[67] S. D’Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408. doi:10.1016/j.enbuild.2014.11.065.

[68] T. Costa, G. Boccignone, M. Ferraro, Gaussian mixture model of heart rate variability, PLoS One. 7 (2012) e37731. doi:10.1371/journal.pone.0037731.

[69] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830. http://www.jmlr.org/papers/v12/pedregosa11a.html (accessed September 1, 2017).

[70] C.M. Bishop, Probability Distributions, in: Pattern Recognit. Mach. Learn., 2006: pp. 67–127. doi:10.1117/1.2819119.

[71] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, L. Soibelman, Real-time activity recognition for energy efficiency in buildings, Appl. Energy. 211 (2018) 146–160. doi:10.1016/J.APENERGY.2017.11.055.

[72] W. O’Brien, H.B. Gunay, The contextual factors contributing to occupants’ adaptive comfort behaviors in offices - A review and proposed modeling framework, Build. Environ. 77 (2014) 77–88. doi:10.1016/j.buildenv.2014.03.024.

[73] Enmetric Systems, (n.d.). https://www.enmetric.com/ (accessed September 19, 2017).