Technology|Architecture + Design

ISSN: 2475-1448 (Print) 2475-143X (Online) Journal homepage: http://www.tandfonline.com/loi/utad20

Profiling Occupancy Patterns to Calibrate Urban Building Energy Models (UBEMs) Using Measured Data Clustering

Rawad El Kontar & Tarek Rakha

To cite this article: Rawad El Kontar & Tarek Rakha (2018) Profiling Occupancy Patterns to Calibrate Urban Building Energy Models (UBEMs) Using Measured Data Clustering, Technology|Architecture + Design, 2:2, 206-217, DOI: 10.1080/24751448.2018.1497369

To link to this article: https://doi.org/10.1080/24751448.2018.1497369

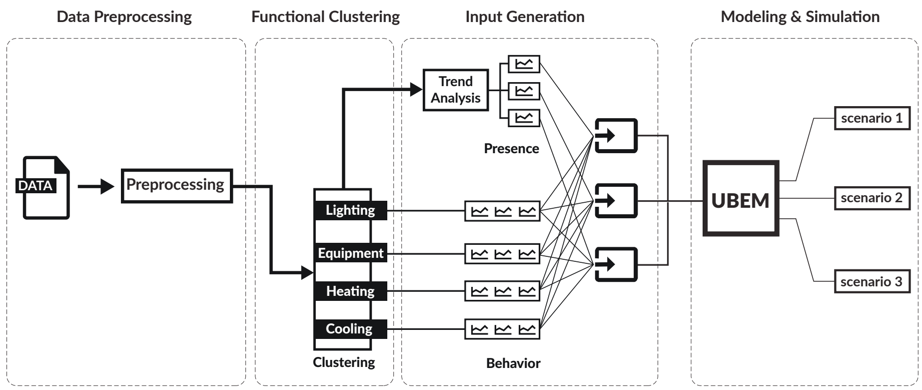
Published online: 29 Nov 2018.

Rawad El Kontar
Syracuse University

Tarek Rakha
Syracuse University

Profiling Occupancy Patterns to Calibrate Urban Building Energy Models (UBEMs) Using Measured Data Clustering

Uncertainty in predicting occupancy patterns leads to discrepancies in simulated building energy when compared to measured data. Typical simulation models represent occupants through identical schedules and repetitive behavior. However, users' activity patterns comprise numerous variations, especially when focusing on interactions in buildings on the neighborhood scale. Urban-scale simulations inform design decisions, and one of the major challenges is identifying representable inputs for occupancy behavior. This paper presents a framework for modeling occupancy and consequent energy loads in residential buildings using measured data for calibration; it employs a functional clustering approach to profile energy use, which generates inputs for Urban Energy Models (UBEMs). The framework is demonstrated on a residential neighborhood and reveals that the generated inputs can more accurately predict community energy load patterns.



◁ Figure 1. Research framework for the generation of inputs to calibrate UBEM simulation.

Introduction

Buildings in the United States are responsible for the largest share of all energy consumption in the country. Consequently, due to global climate change, energy conservation in buildings has been receiving significant attention. Energy use has also influenced urban planning and the development of energy policies at an urban scale. Therefore, the roles urban planners and designers play in reducing the negative impacts on the built environment are critical. Urban-scale building modeling was developed as a way to represent the state of urban energy consumption and predict its future evolution.[1,2]

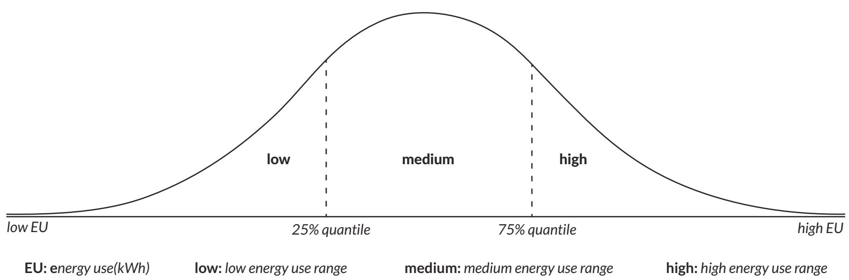
Urban Building Energy Model (UBEM) simulates various performance measures, including operational energy use, in order to inform designers, urban planners, and policy makers in their evaluation of energy demand and supply strategies, design decisions, and performance of urban energy systems.[3] Similar to building models, the generation of a UBEM requires the definition of data inputs for building geometries in addition to a large set of non-geometric parameters that includes usage schedules and behaviors, which affect internal loads. These non-geometric inputs are usually reduced by deterministically simplifying the real diversity of occupant behavior into defined archetypes.[4] This results in simulations where all occupants perform identical actions, leading to erroneous hourly demand peaks and ultimately to the misrepresentation of urban energy demands.[5] Therefore, uncertainty in defining occupancy patterns and behavior is a major cause of discrepancies in simulated building energy when compared to real measured data.

The International Energy Agency (IEA), Energy in Buildings and Communities Program (EBC), Annex 53, has identified occupant behavior as one of the main driving factors of energy use in buildings.[6] Human behavior entails the occupants' interaction with building equipment, lighting, heating, and cooling systems. These systems determine the energy use of a building; therefore, the behavioral factor in building performance simulation is of significant value.[7, 8, 9] In brief, occupancy schedules and behavior are a necessary input for simulation models to accurately predict energy use, and models should be able to generate simulations of temporal behavioral patterns to better inform users regarding their energy use behaviors.

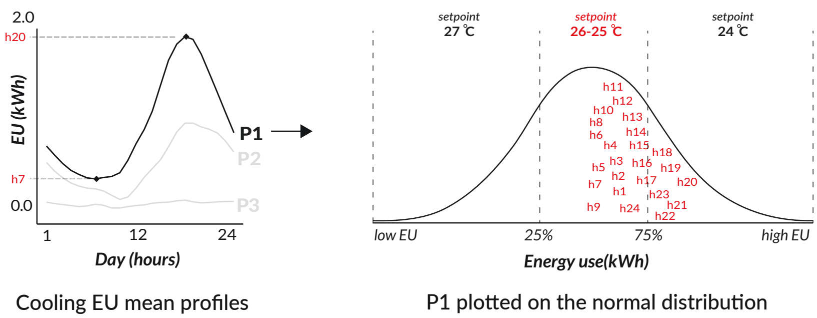
Previous literature devoted to this issue focused on energy load prediction and pattern profiling to represent occupant behavior. Profiling the energy load of occupants has been applied qualitatively and quantitatively to improve load prediction.[10-19] When attempting to profile energy, data clustering was utilized to classify and analyze the energy consumption behavior in buildings.[12-17] Common clustering methods in determining energy load versus time are: K-means, the self-organizing map, and the minimum variance criterion; the fuzzy C-means and combinations of these methods can also be found.[12, 17, 18] Other more robust approaches are model-based, including cluster-wise regressions and mixture models.[20] The effectiveness of clustering methods varies when applied to different data sets, and there is an inherent tradeoff when comparing clustering methods.[20] Therefore, defining a clear analytical purpose aids in choosing an appropriate clustering method.[20]

Identifying patterns of occupant presence and predicting occupancy schedules is another essential step that will allow for the more accurate modeling of occupancy energy use behaviors. However, it is difficult to observe and predict occupancy schedules, as this involves stochastic variables and requires a significant amount of data. Three typical methods were used in previous literature to model occupant presence. The first method consists of representing occupants as groups with fixed schedules.[21] These groups are combined afterward to represent the schedule of the whole building. In the second method, occupant schedules are represented as a probability distribution.[22, 23, 24] The third method consists of analyzing observed behavior.[25, 26] While these methods improved the modeling of occupant schedules, they are limited in presenting accurate schedules, because (1) occupant schedules are highly stochastic; therefore, it is inappropriate to simply label occupants to belong to a certain schedule or certain distribution; (2) results are not practical, as they only conclude with summarizing rules of occupant presence rather than workflows that can present occupancy schedules in future-case scenarios; (3) the results lack validation with measured data; and (4) observed data attained just from a portion of a building cannot be generalized to represent the whole building.[27]

In order to represent occupant behavior and operational energy use accurately, simulation models should be calibrated. Calibration is widely used at a building scale and has proven to be successful in improving simulation accuracy. In an attempt to define the improvement level of calibration; Samuelson et al. analyzed each of the standard calibration tasks systematically. These tasks included inputting actual weather data, adding unregulated loads, revising process loads, and updating a small number of inputs. The results showed that the bulk of this improvement came from revising process loads using sub-metered data.[28] Unfortunately, such calibration tasks are challenging to apply to UBEMs due to the time and computational power



◁ Figure 2.Parametrizing the range of set points through a normal distribution of cooling energy uses.



◁ Figure 3.Process of assigning a cooling set point to a mean profile deduced from clustering analysis, where h is the energy use value at a specific hour, e.g., h20 = the energy use value at hour 20.

associated with the modeling and calibration process. In UBEMs, profiling energy use and behaviors could be represented in various ways. In previous limited literature, researchers associated energy use with household income.[29, 30] Filogamo et al. estimated several occupancy parameters from national statistics as a function of average income.[31] However, this deterministic approach does not represent the variety of behaviors accurately. In a different approach that addresses uncertainty, Bayesian calibration has been used to adjust urban energy models.[32] However, in practice, this method relies on excessive computational effort, a high level of expertise, and significant labor time. Therefore, there is a gap in the literature on how to effectively utilize metered data to calibrate energy models to accurately represent energy use patterns through a practical process.

The majority of previous occupant behavior research focused on behavior in offices rather than residential behavior, since the latter is hindered by scarce data availability and privacy concerns. However, residential behavior accounts for the largest amount of variations and randomness.[33] In the meantime, smart meter installation in residential buildings has increased greatly, and it is expected to continue to grow due to the recent interest in residential energy consumption of newly built neighborhoods.[34] Occupant behavior parameters are among the most uncertain in energy modeling, yet behavior is one of the main drivers of energy use in the residential sector.[35, 36] A detailed survey of electricity consumption in the United Kingdom monitored 250 homes, and Godoy-Shimizu et al. found significant variations between electrical base loads of households, especially in lighting, in which there was a vast difference between the lightest and heaviest users.[37] Urban models only validate the averaged results on an annual basis; therefore, discrepancies in occupant schedules might not be apparent due to the aggregation of results. However, when generating hourly energy use at a neighborhood scale, representing every building with identical behavior is expected to be erroneous. Modeling techniques that deal with unknown parameters have been extensively used to address individual building energy simulations.[38, 39] However, it is unclear how to apply these methods at an urban scale where the process is constrained by the over-parameterized and high computational cost of simulation. While some statistical models for individual parameters can reduce computation time, parameter data is not available at an urban scale.[40] As a result, most UBEMs have so far used deterministic characterization at a detailed level supported with the available data.

This paper focuses on addressing the research gap of load profiling in residential urban neighborhoods by connecting occupant schedules to behavioral profiles and creating a practical workflow to represent patterns of energy loads in a systematic way. This methodology allows results to be applied as input data to calibrate UBEMs. A framework is presented for modeling occupancy presence and consequent energy loads in residential buildings on the urban scale. The work employs a computational clustering approach to data from a residential community in Austin, Texas, where energy is continuously measured using smart meters. The paper's goal is to determine energy use profiles that represent occupant presence, as well as activity patterns, by clustering available hourly energy use data. First, energy use behaviors are defined using functional clustering of hourly data for each of the building energy use variables (lighting, equipment, cooling and heating). The resultant energy use behavior for all the variables is then utilized to derive occupancy presence profiles. Profiles from the twofold process are then matched to associate occupancy presence with behavioral patterns on a daily basis. Finally, the outcome is translated for use as an input for UBEM simulation software. The clustering and input generation process is automated using R, a code-driven application, and a UBEM for the residential community is developed using the Urban Modeling Interface (UMI) plugin for Rhino3D CAD software.[41, 42] The UBEM is constructed with initial inputs from existing construction and presence/behavioral inputs that are generated from the clustering process. The study concludes by defining the effect of occupancy on energy consumption and illustrating the importance of hourly changes into simulation tools to enhance the accuracy of outcomes.

Data Description

The analysis considers single-family households with metered electricity use data and energy audits. These datasets were collected as part of the Pecan Street Smart Grid Demonstration Project at the Mueller development, located in Austin, Texas, and operated by the Pecan Street Research Institute.[43] The Mueller development, planned to include up to 5700 households, currently has 750 single-family homes built after 2007. These households use smart meter monitoring that measures energy consumption at a 1-minute interval for the whole home, and 6 to 22 subcircuits and major appliances.[44] Exploring the data and identifying key audits of the buildings is important in the analysis. Building sizes and type have a major effect on building energy use, and since the aim is to define behavioral profiles that represent a cluster of buildings, the analyzed buildings were grouped according to their size and type prior to the clustering process. This can be done through a multi-clustering analysis to identify important groups within a heterogeneous neighborhood. As a result, 67 single-family homes constructed between 2007 and 2010 with similar sizes and identical construction materials were selected for analysis.

Methodology

The proposed framework is illustrated in Figure 1. Data preprocessing is described, followed by the functional clustering approach. The results from clustering are then translated as inputs for the model. Finally, the last section discuss the development of a calibrated UBEM.

Data Preprocessing

Firstly, the data is processed in three steps: the dataset is organized and cleaned of corrupt and missing data. Secondly, energy use measures collected from sensors are restructured into hourly energy consumption from 0:00 till 23:00 for each component of a household to reduce data dimensions. Finally, hourly data from each household is grouped into four categories of lighting, equipment, heating, and cooling, described in the table below. These steps make the data less challenging to analyze and therefore helps to accelerate the proposed workflow.

Table 1. Description of utility type and relevant appliances.

Lighting	Lighting energy use from all rooms in the house
Equipment	Washing machines, dryers, house fans, dishwashers, freezers, ice makers, jacuzzi, kitchen appliances, microwaves, ovens, pool, pool pump, refrigerator and security equipment
Heating	Furnace, air handler and stand-alone heaters
Cooling	Air compressors and window unit air conditioners

Functional Data Clustering

If collecting sensor energy use readings from N different households, for the i th household, the history of the observed energy use pattern is denoted as $\boldsymbol{y}_i = \{y_i(t_{i1}), \ldots, y_i(t_{ip_i})\}^{\mathrm{T}}$ where T denotes a transpose, represents the number of observations, i.e., sensor measurements, for unit $i \in \{1, \ldots, N\}$and $\{t_{iq}, q = 1, \ldots, p_i\} \subset \mathcal{R}$ (the real line) represents the observation time points for unit i, the time at which energy use is measured. For instance, $t \in \{0{:}00, 1{:}00, \ldots, 23{:}00\}$ may represent the daily energy consumption structured into hourly use for each household. Note that the measurements represent accumulated readings of a sensor value over an hour. Based on the energy use profiles collected from N different households, functional clustering is utilized to define energy use behavior patterns. To address the functional nature of the data and identify common energy use patterns, the analytical approach utilizes model-based clustering, proposed by Bouveyron et al.[45] This approach is based on extensive research on discriminative analysis, which has been used for clustering in univariate, multivariate, and functional settings.[46]

Let $\{y_1(t), \ldots, y_N(t)\}$ denote the function that needs clustering. The first steps are based on recovering the functional nature of the data through a finite basis expansion,

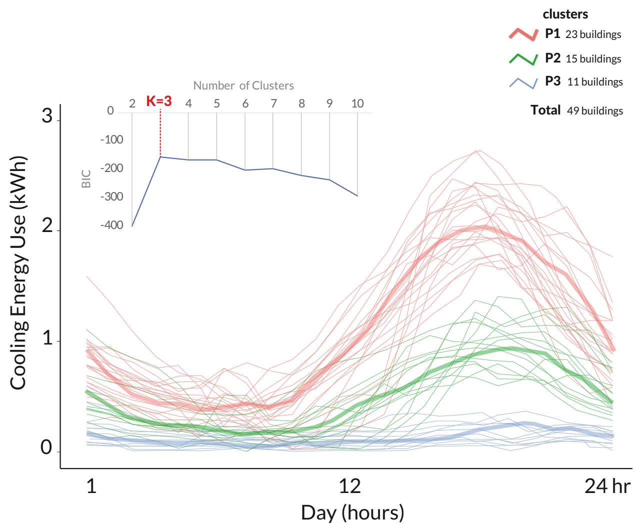
$$y_i(t) = \sum_{j=1}^{z} \omega_{ij} \varphi_j(t),$$

where $\varphi_j(t)$ are a set of basis functions with coefficients ω_{ij}. The coefficients $\omega_i = (\omega_{i1}, \ldots, \omega_{iz})$ for function $y_i(t)$ are then assumed to belong to a mixture Gaussian distribution, where clustering of the time series data is performed in a discriminative functional subspace.

$$P(\omega) = \sum_{k=1}^{K} \pi_k \phi(\omega; \eta\mu_k, \eta^t \Sigma_k \eta + \Lambda),$$

where P denotes probability, π_k is the mixing probability, ϕ is the standard Gaussian density function, μ_k and Σ_k are the mean and covariance matrix of the k^{th} cluster for the mapping of ω into the discriminative subspace, where η is a matrix representing the mapping to the discriminative subspace. Additionally, Λ denotes the covariance matrix related to measurement noise. Then the optimal number of clusters K is selected using a Bayesian Information Criterion (BIC).

In this clustering analysis, hourly data of each building is plotted over 24 hours. Continuous functions of 365 days are plotted for each building as repetitive measures. To simplify the clustering analysis in the following step, the 365 functions for each day of the year of each building are presented as a mean profile. Subsequently, building profiles are clustered into K number of groups (see Figure 4), and a mean behavioral profile with a confidence interval is plotted as a representation of each cluster. From the clustering results, the user can identify the group each building belongs to and determine the number of buildings in every clustered group. The user can visualize this functional clustering process. Clustering errors can be sometimes distinguished by the user who identifies the meaningful patterns that computers cannot see, that is, reinterpreting through reasoning. The benefit of this method is that it takes advantage of the temporal dynamic of the data and graphically models it. In addition, the low



△ Figure 4. Clustering analysis using cooling energy use data.

computational complexity and efficient visualization of the clustered systems adds practicality to real applications.

To assess the effectiveness of the clustering analysis, a number of cluster validity indices can be used. In our study, the Mean Index Adequacy (MIA) is calculated for each clustering analysis and presented in the Results section. The generated value relies on the compactness of each cluster; if the members in the cluster are close together, the MIA is low. Refer to Dent et al. for further explanation on the mechanisms underlying the MIA.[47]

The functional data clustering is applied to determine mean behavioral profiles for each of lighting, equipment, heating, and cooling energy use on weekdays and weekends separately. These profiles are then translated to define usage schedules and behavior inputs.

Generation of Usage Schedules and Behavior Inputs

The translation of behavioral energy use profiles is achieved using the formula below. Let $F_c(t_1, t_2)$ be the fraction to be applied to a load between time instances t_1 and t_2 for a specific variable cluster vc, where $v \in \{l, e, c, h\}$ where l, e, c, h are lighting, equipment, cooling, and heating respectively.

$$F_c(t_1, t_2) = \int_{t_1}^{t_2} y_{vc}(t).dt / Max\ Load$$

The maximum load is user-defined and represents the maximum load that should be applied to a specific building type. However, the maximum load can be approximated from the clusters as the total energy use at peak times. Deduced schedules from weekdays and weekends are combined to form weekly and annual schedules of occupancy presence and usage behaviors. The combination of a weekday and weekend schedule for specific buildings should be informed by identifying the cluster that the building belongs to in both the weekend and weekday clustering results.

While modifying schedule settings aims to more precisely represent the energy use behaviors on an hourly basis, illuminance targets and heating/cooling set points also have a significant effect on defining the overall intensity of the lighting, cooling, and heating energy use. For heating and cooling, a set point temperature defines the temperature range the occupants will likely try to maintain; the temperature below which heating is turned on or the temperature above which cooling is turned on. The authors defined a range of set points based on ASHRAE 90.1 as a standard, parametrized for energy efficiency: heating set points (18°–22°) with 20 ° being the mean; cooling set points (24°–27°) with 25.5° being the mean.[48] For lighting, illuminance targets are set based on Illuminating Engineering Society of North America (IESNA) guidelines, where 150 lux is required for bedroom and storage, while for kitchens and study areas, 300 lux is preferred.[49] Therefore, the authors set a range of 150 lux to 300 lux for illuminance targets. Note that set points significantly vary in residential buildings and are relatively challenging to determine, especially at a individual building scale. However, the defined set points can be represented when simulating energy use at the urban scale and help in calibrating the UBEMs. Further research should incorporate metered gas data to more accurately define heating set points, and lighting standards need defining for overall building use rather than just detached building spaces.

Using the defined range of set points, each behavioral profile will be assigned a set point or illuminance target that correspond to its energy use intensity. The framework for assigning heating/cooling set points and illuminance targets for different clusters is illustrated in Figures 2 and 3 and detailed in the steps below. Without loss of generalizability, in the steps below it is assumed that the outcome of the behavioral clustering analysis was three clusters for heating, cooling, and lighting variables.

Step 1: Define set points/illuminance targets as low, medium, and high

Using the energy data from each of the analyzed variables across all different hours and building IDs, build the empirical distribution and find the corresponding ranges in which cooling energy is used as an example, as shown in Figure 2. Note that the quantiles are user-specified. In this case 0.25, 0.75 and 1 are chosen, which is a natural quantile specification.

Step 2: Assign set points/illuminance targets to clusters

At each hour, determine where each reading of energy use for each cluster falls compared to the parametrized specifications (low, medium, high) illustrated in Figure 2 above. This procedure is demonstrated in Figure 3. The resultant graph shows that most energy use points fall in the middle zone and a set point of 25° is assigned to the corresponding profile.

Occupancy Presence Input Generation

The use of lighting, equipment, and conditioning appliances are typically linked to the presence of the users. While most researchers have mainly focused on studying the importance of occupant interaction with lighting appliances to model the randomness of occupancy presence, the relationship of lighting energy use and occupancy schedules is more likely dominated by daylighting effects.[50, 51, 52, 53] Therefore, all energy variables are considered as a proxy for occupancy presence in our framework, and occupants' schedules are represented by defining different behaviors of heating, cooling, equipment, and lighting energy use. For this reason, clustering analysis results for each variable are analyzed to determine presence schedules. The process is applied by conducting trend analyses to group different occupancy presence variations. In the case study, only one identified trend was translated empirically

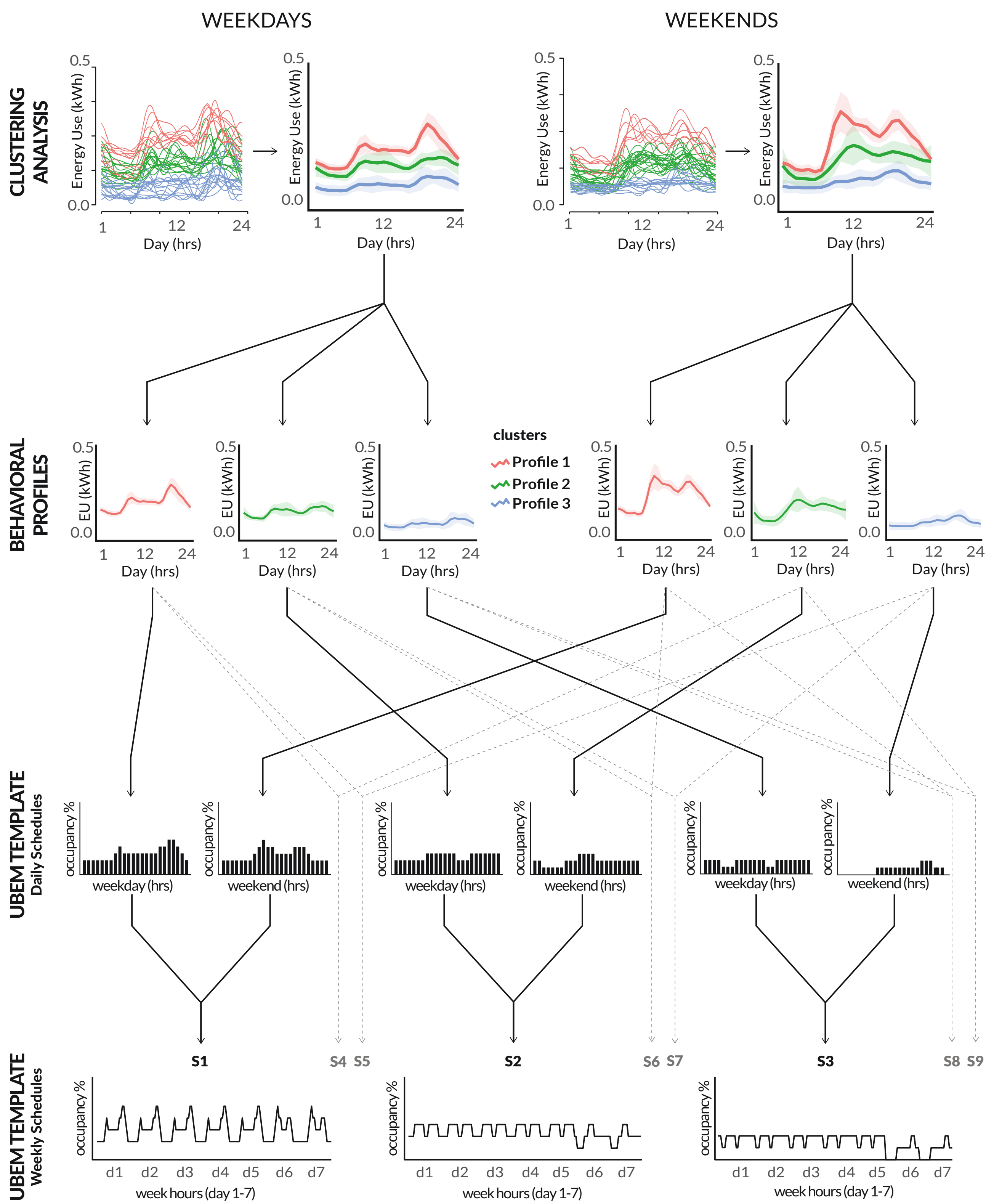
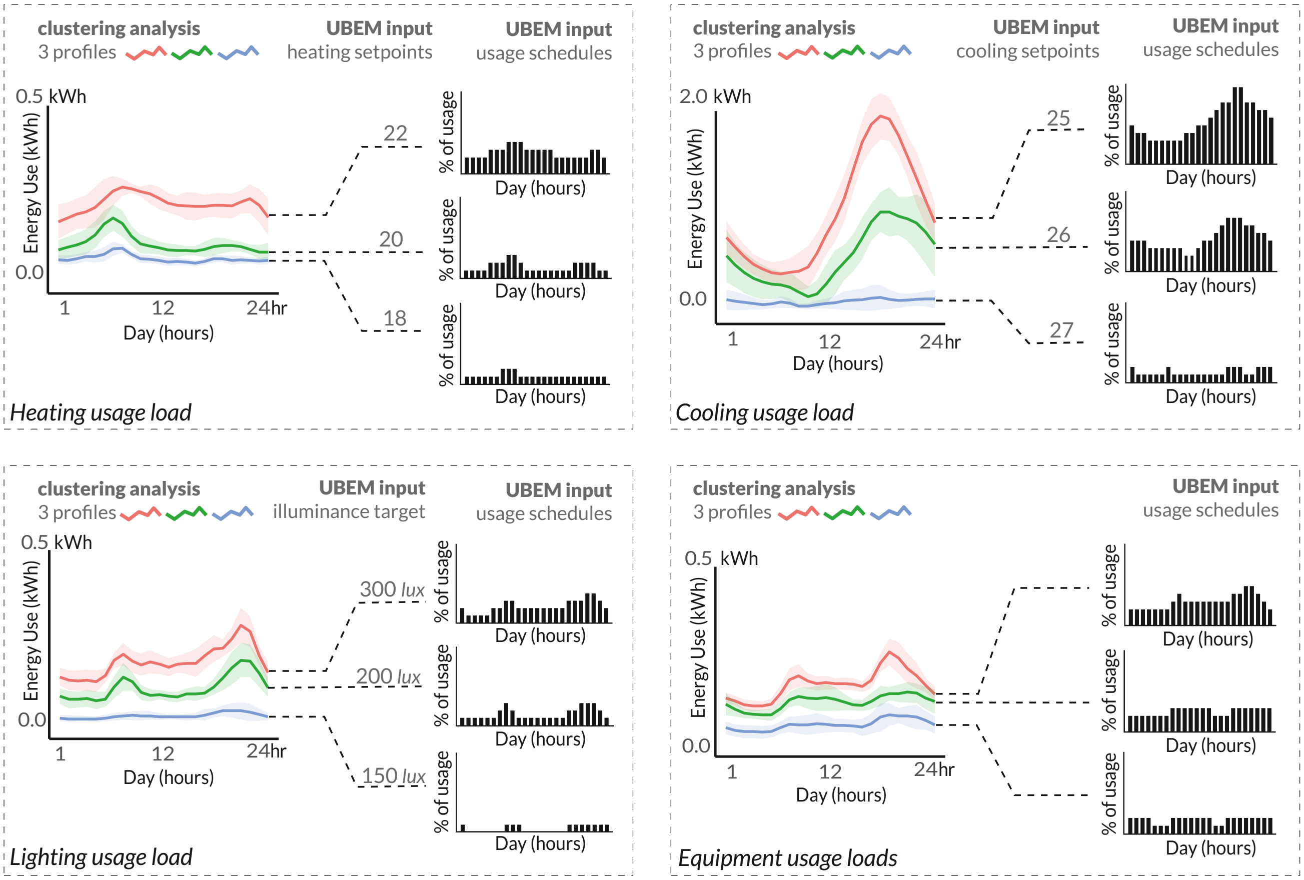
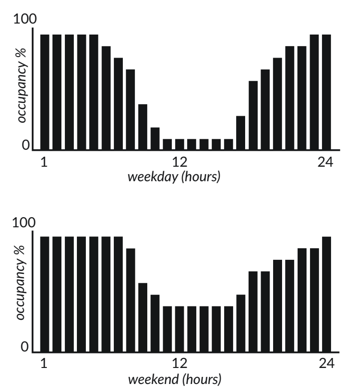


Figure 5. Generating usage schedule inputs from profiling equipment energy use.



△ Figure 6. Mean behavioral profiles deduced from clustering analysis, translated into usage schedules.



◁ Figure 7. Generated presence schedules.

into a schedule of occupancy (Figure 7); this example is explained in the Results section. While the community population was represented with only one dominating trend, others may exist that would help to deduce more presence profiles. Future research should review and further develop this approach.

Template Generation

In this phase, occupancy presence schedules are matched with all possible variations of behavioral/usage schedules along with the associated illuminance targets and heating/cooling set points to create an input for a UBEM; this input is characteristic of the whole population.

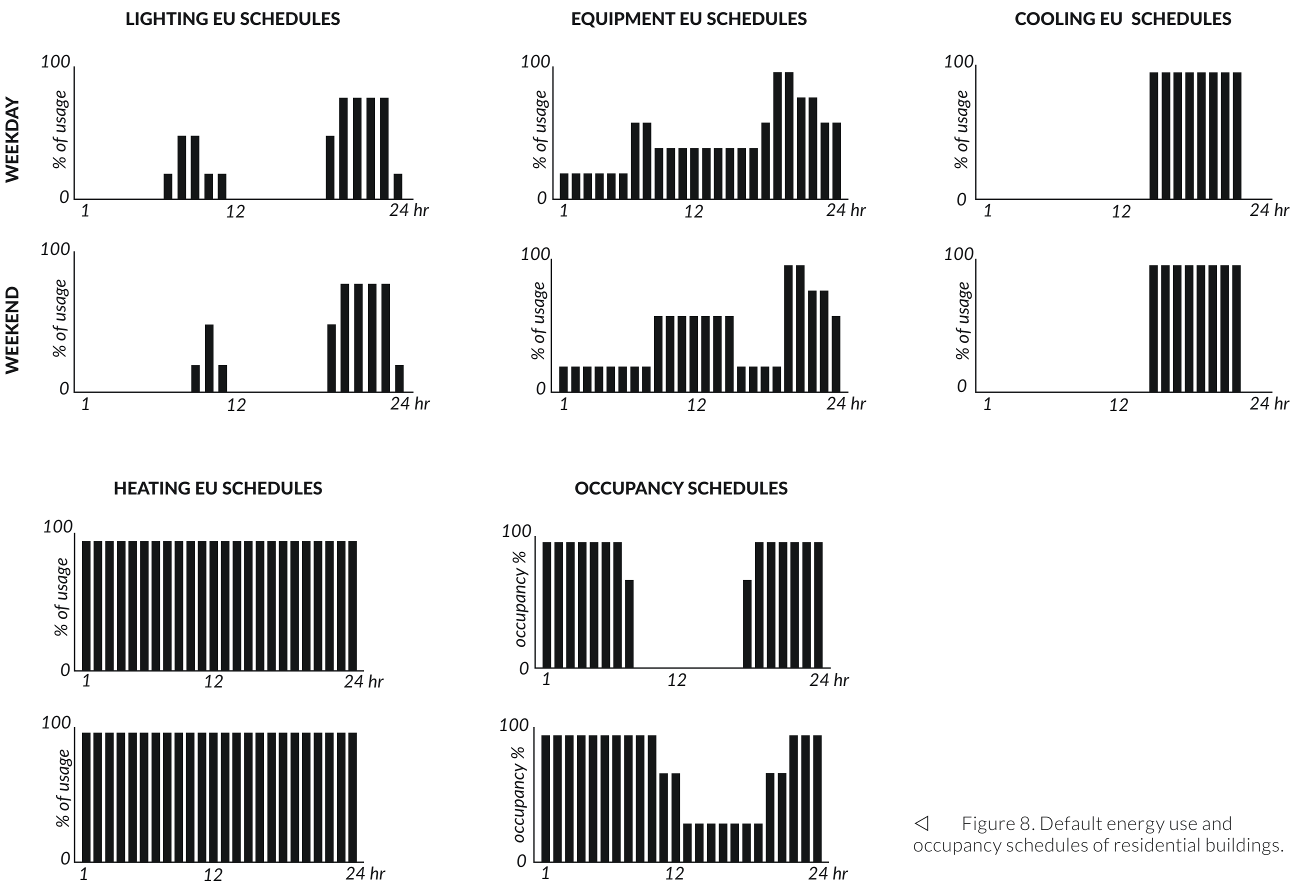
UBEM inputs in the form of template libraries are used to include the schedules and behaviors. The templates are then assigned to the corresponding building's ID in the UBEM. The use of the template contributes the following:

1. Occupancy schedules represent occupancy density and presence.
2. Heating behavioral profiles represent the conditioning use schedules for heating and are used to specify heating set points.
3. Cooling behavioral profiles represent the conditioning use schedules for cooling and are used to specify cooling set points.
4. Lighting behavioral profiles represent lighting usage schedules and are used to specify illuminance target values.
5. Equipment behavioral profiles represent plug load usage schedules.

Calibration of the model is achieved by using the generated inputs that represent behaviors of the neighborhood more accurately. This calibration process is more effective in comparison to previous calibration methods that required high computational power and time.

UBEM Development and Simulation

In this final phase, building and developing a UBEM model is achieved through three steps: characterization, generation, and simulation. First, weather information, buildings and context



◁ Figure 8. Default energy use and occupancy schedules of residential buildings.

geometry, and non-geometric properties should be specified. Data inputs for building geometry, construction, material properties, and window-to-wall ratio (WWR) are usually provided from a survey of existing construction or from municipal archives. The building and context geometries information are used to build a three-dimensional model in Rhino3D. This digital massing model provides volumetric information of the built environment and is used to calculate orientations and areas. The inputs of occupancy schedules, behavioral schedules, heating/cooling set points, and illuminance targets, along with material properties and WWR information, should be used to refine geometrical and non-geometric settings for the templates. After assigning the 3-D massing with all the templates, simulations are performed. Users can simulate different scenarios and iterations of their proposals. These can be design proposals in which the user adjusts geometric parameters, the choice of materials, heating/cooling set points, and illuminance targets, or the user can study load shifting strategies by comparing different occupancy and usage schedules to the base parameters defined by the clustering method.

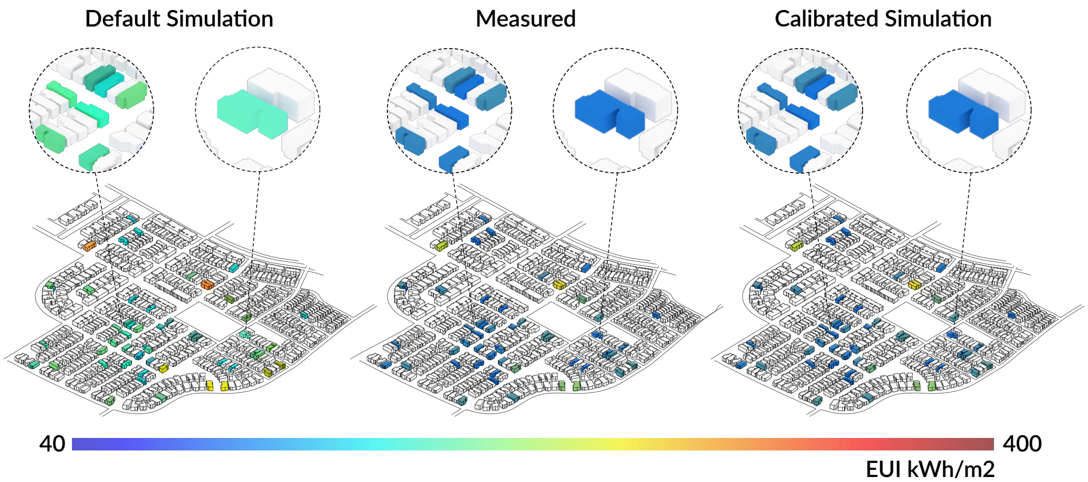
Results

A residential community in Austin, Texas, where energy is continuously measured using smart meters, is used as a case study to demonstrate this methodology. Following the methodology steps, the authors started the preprocessing phase with 67 building data sets. After data cleaning, 18 buildings were removed due to data corruption and absence of data points, leaving 49 houses suitable for further analysis.

After preprocessing the data, behavioral/usage profiles are determined by applying a functional clustering analysis to cluster each of the variables on weekdays and weekends separately. Figure 4 shows the clustering process of cooling energy use; the optimal number of clusters K is determined to be 3 according to the BIC versus K graph.The MIA results for each of cooling, heating, lighting, and equipment clustering analysis are 0.22, 0.28, 0.12, and 0.20 respectively; these results indicate that the clusters generated for all variables have low variability within each cluster.

Following the methodology of usage schedules, the three mean profiles that represent behavioral patterns are translated into schedules that represent fractions of energy use for every hour of the day. The inputs for weekdays and weekends are then matched and combined to create an annual schedule for every possible combination. This process is illustrated in Figure 5, which presents the generation of inputs from profiling equipment energy use data; S1–S9 represent all the possible combinations of weekday and weekend schedules.

Usage schedules are generated by separately profiling lighting, heating, cooling, and equipment energy use. Lighting and HVAC behavioral profiles are then characterized by the illuminance target or heating/cooling set point, which is formulated according to steps 1 and 2. Figure 6 shows the generated inputs in correspondence with their mean profiles.



▷ Figure 9. Comparing measured data with default simulation and calibrated simulation models. (Credit: Rawad El Kontar)

In the next step, an occupancy schedule is deduced from a trend analysis of the profiled behavioral patterns. Through analysis of the clustered profiles of equipment and lighting energy use, the authors identified peak hours of energy use to be in the morning hours from 7 a.m. until 10 a.m., and in the afternoon from 6 p.m. until 11 p.m. The majority of the occupants are inferred to be workers and students who leave the house in the morning and return in the evening. Therefore, weekday and weekend schedules are constructed representing this trend (Figure 7) as follows: during weekdays, full occupancy decreases gradually from 6 a.m. to 9 a.m., when occupant presence stabilizes at a low level during the day and then gradually increases from 5 p.m. onward, when occupants return from work or school. During the weekend, waking hours and sleeping hours are delayed gradually, while the level of occupancy during the day tends to be higher. Daily schedules are combined into an annual schedule that represents the occupancy of the whole neighborhood.

Next, the occupancy presence schedule is combined with each of the usage schedules and behaviors to form several input templates, whereby each template corresponds to a cluster of buildings. It is important to note that the number of behavioral and presence profiles one generates will determine the number of combinations that can be included in the general template library. This methodology characterizes the urban built environment more accurately, since the average usage patterns and occupancy schedules summarize the behavior of the whole population.

Finally, a baseline UBEM for this case study is developed using the Urban Modeling Interface (UMI) plugin for Rhino3D CAD software. UMI's operational energy simulation is based on an algorithm that abstracts an arbitrarily shaped set of building volumes into a group of simplified "shoebox" building energy models and uses EnergyPlus as an underlying simulation engine.[55] The Austin-Camp Mabry Actual Meteorological Year (AMY) weather file, containing data from 2014 (January 1 to December 31) and purchased from White Box Technologies, was used for the study.[56] Inputs for building geometric data, construction, and materials are defined by extracting information from the existing construction of the community.

Starting with this baseline UBEM, two files are created. The first is a default file in which all the buildings are assigned the same template that includes default occupancy and usage schedules of UMI residential default set points and illuminance targets. These selected defaults are based on ASHRAE and IESNA: a heating set point of 20°, cooling set point of 25°, illuminance target of 200 lux.[48, 49] The selected schedules are based on the Swiss Society of Engineers and Architects (SIA) and are shown in the Figure 8.[57]

The second file is calibrated so that buildings are grouped according to the clustering analysis. Each cluster of buildings is assigned with a generated template that corresponds to its occupancy, behavioral energy use patterns, and energy use intensity.

In order to test our designed inputs, simulations from the files were generated and compared. Figure 9 shows false color simulation results for total operational energy use in Energy Use Intensity (EUI) from the default and calibrated models compared with a false-colored 3-D model that shows a real representation of the EUI obtained from measured data.

Discussion

Hourly results of cooling, equipment and lighting energy use were extracted from the simulated models and plotted as averaged profiles over a day. To test the validity of the model, the simulation results of one building are plotted (Figure 10). These results include both the UBEM simulated results from the default and calibrated files, which were then compared with plots from the measured data.

Based on outcomes illustrated in Figure 10, the relevance of using measured data in calibrating UBEMs can be discussed. First and most importantly, the model based on functional clustering significantly improved the performance of the simulation model to match real performance. As shown in Figure 10, the simulated patterns (denoted as y_{sim}) in cooling, equipment, and lighting energy use are notably closer to the measured data (denoted as y_{mea}), within a 10% margin of error, compared to the default model results.

These results are validated through the Root Mean Square Error (RMSE), which is a statistical measure to describe the similarity of two data sets. It characterizes the average variance of the elements of the simulated profile with respect to the measured profile.

A small RMSE indicates smaller variance between the compared data series, defined as:

$$rel.RMSE = \frac{1}{24}\sqrt{\sum_{i=1}^{n=24}\left(\frac{(y_{sim,i} - y_{mea,i})}{y_{mea.i}}\right)^2}$$

According to the RMSE and through comparison with the measured data, the authors identified the following improvements resulting from the calibration method:

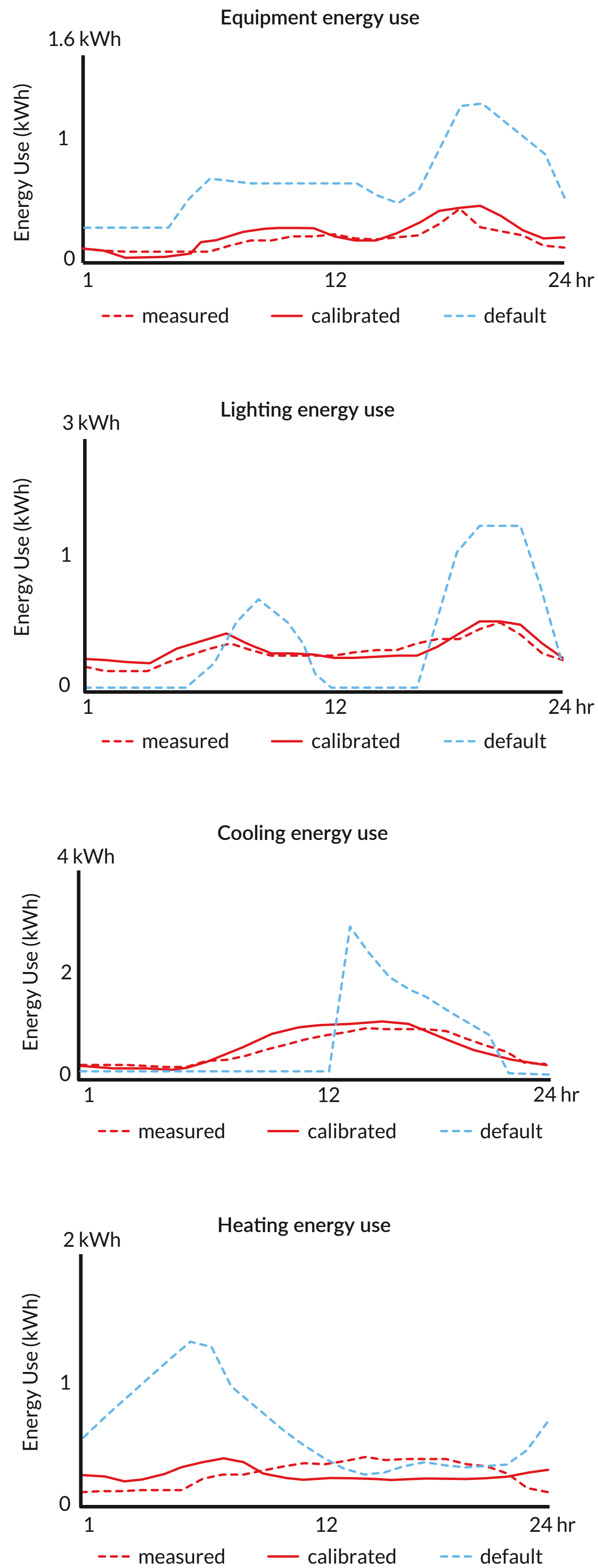
- In lighting energy use, the error decreased from 32% to 4%.
- In equipment energy use, the error decreased from 60% to 8%.
- In cooling energy use, the error decreased from 30% to 7%

In the results, heating energy use was disregarded since the use of heating is negligible in the warm climate of Austin, Texas. As this example demonstrates, the performance of the clustering approach is dependent on data availability and the size of the sample data from the community, as well as observations of contextual issues.

The observation of our study validates that the accuracy of simulation models can be improved by identifying behavioral patterns using a measured data-driven approach. Second, the results confirm that the proposed modeling framework is able to scale to an urban level due to the prediction accuracy provided through summarizing the population behavior into a small number of clusters. The resulting daily (24-hour) profiles of occupant energy use behavior are the main outcomes of this study. At an urban scale, these simulated hourly profiles allow users to more accurately inform occupants about their hourly energy use behavior in relation to community energy performance, and this consequently informs load shifting strategies. The simulated results more accurately visualized the building energy demand peaks and primary energy load patterns. Thus, urban designers and policy makers can more accurately test the performance of potential future case scenarios in relationship to existing conditions.

Conclusion

Occupancy presence and behavior have a significant impact on building energy consumption. With the growing use of simulation tools to support built environment research and practice, misrepresentation of occupant presence and behavior can misinform design decisions. The field of UBEMs is still emerging, and simulation errors at the scale of multiple buildings can be momentous. This study demonstrates how the use of measured data can develop more accurate UBEMs that are particularly useful in creating design cases. The method's robustness is shown through the scenario-based approach, which is accessible for researchers exploring speculative designs based on numerous inputs that follow community trends without relying on excessive computational power, an extraordinary level of expertise, or substantial labor time. The impact of the proposed framework is demonstrated through the use of functional data clustering that creates occupancy-based inputs, which calibrated UBEMs within a 10% maximum margin of error. Due to the limitations of the process that focused on only one community occupancy trend, future research should investigate the impact of multiple occupancy presence trends and their effect on overall community performance. Ultimately, this work could aid in the development of UBEMs calibrated in real time to assist users, designers, utilities



△ Figure 10. Comparing calibrated simulation results with measured data and default simulations.

companies, and others in making informed decisions that reduce the environmental impact of communities, neighborhoods, and cities using measured data.

Acknowledgments
This publication is based on work funded in part by the National Science Foundation (NSF), under the Smart and Connected Communities program grant 1737550, and the Syracuse Center of Excellence Faculty Fellows program. The authors would like to thank the Pecan Street Institute for providing access to measured energy data for the Mueller community in Austin, Texas. The authors are also grateful for the work that students Elena Echarri, Yu Qian Wang, and Rutuja Ganoo provided to support this manuscript.

Notes

1. Reinhart, C. F., and C. C. Davila. 2016. "Urban Building Energy Modeling–A Review of a Nascent Field." *Building and Environment* 97: 196–202.
2. Bourdic, L., and S. Salat. 2012. "Building Energy Models and Assessment Systems at the District and City Scales: A Review." *Building Research & Information* 40(4): 518–526.
3. Allegrini, J., K. Orehounig, G. Mavromatidis, F. Ruesch, V. Dorer, and R. Evins. 2015. "A Review of Modelling Approaches and Tools for the Simulation of District-Scale Energy Systems." *Renewable and Sustainable Energy Reviews* 52: 1391–1404.
4. Wilke, U., F. Haldi, J.L. Scartezzini, and D. Robinson. 2013. "A Bottom-Up Stochastic Model to Predict Building Occupants' Time-Dependent Activities." *Building and Environment* 60: 254–264.
5. He, M., T. Lee, S. Taylor, S. K. Firth, and K. J. Lomas. 2015. "Coupling a Stochastic Occupancy Model to EnergyPlus to Predict Hourly Thermal Demand of a Neighbourhood." *Proceedings of Building Simulation 2015*, December 7-9, 2015, Hyderabad.
6. Yoshino, H., T. Hong, and N. Nord. "IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods." *Energy and Buildings* 152 (2017): 124-136.
7. Mahdavi, A. 2011. "People in Building Performance Simulation." *Building Performance Simulation for Design and Operation*, edited by R. Hensen (London: Spon Press, 2011), 56-83.
8. Zhang, W., S. Tan, Y. Lei, and S. Wang. 2014. "Life Cycle Assessment of a Single-Family Residential Building in Canada: A Case Study." *Building Simulation* 7(4): 429–438.
9. Attia, S., A. De Herde, E. Gratia, and J. L. M. Hensen. 2013. "Achieving Informed Decision-Making for Net Zero Energy Buildings Design Using Building Performance Simulation Tools." *Building Simulation* 6(1): 3–21.
10. Ayodele, T. R., A. S. O. Ogunjuyigbe, and I. A. Atiba. 2017. "Assessment of the Impact of Information Feedback of Prepaid Meter on Energy Consumption of City Residential Buildings Using Bottom-Up Load Modeling Approach." *Sustainable Cities and Society* 30: 171–183.
11. Chicco, G., and I.-S. Ilie. 2009. "Support Vector Clustering of Electrical Load Pattern Data." *IEEE Transactions on Power Systems* 24(3): 1619–1628.
12. Panapakidis, I. P., T. A. Papadopoulos, G. C. Christoforidis, and G. K. Papagiannis. 2014. "Pattern Recognition Algorithms for Electricity Load Curve Analysis of Buildings." *Energy and Buildings* 73: 137–145.
13. Kazas, G., E. Fabrizio, and M. Perino. 2017. "Energy Demand Profile Generation with Detailed Time Resolution at an Urban District Scale: A Reference Building Approach and Case Study." *Applied Energy* 193: 243–262.
14. López, J. J., J. A. Aguado, F. Martín, F. Munoz, A. Rodríguez, and J. E. Ruiz. 2011. "Hopfield–K-Means Clustering Algorithm: A Proposal for the Segmentation of Electricity Customers." *Electric Power Systems Research* 81(2): 716–724.
15. Marszal-Pomianowska, A., P. Heiselberg, and O. K. Larsen. 2016. "Household Electricity Demand Profiles–A High-Resolution Load Model to Facilitate Modelling of Energy Flexible Buildings." *Energy* 103: 487–501.
16. Mutanen, A., M. Ruska, S. Repo, and P. Jarventausta. 2011. "Customer Classification and Load Profiling Method for Distribution Systems." *IEEE Transactions on Power Delivery* 26(3): 1755–1763.
17. Tsekouras, G. J., N. D. Hatziargyriou, and E. N. Dialynas. 2007. "Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers." *IEEE Transactions on Power Systems* 22(3): 1120–1128.
18. Chicco, G., R. Napoli, and F. Piglione. 2006. "Comparisons among Clustering Techniques for Electricity Customer Classification." *IEEE Transactions on Power Systems* 21(2): 933–940.
19. Nikolaou, T. G., D. S. Kolokotsa, G. S. Stavrakakis, and I. D. Skias. "On the application of clustering techniques for office buildings' energy and thermal comfort classification." *IEEE Transactions on Smart Grid 3*, no. 4 (2012): 2196-2210.
20. Hsu, D. "Comparison of Integrated Clustering Methods for Accurate and Stable Prediction of Building Energy Consumption Data." 2015. *Applied Energy* 160: 153–163.
21. Zhang, T., P.-O. Siebers, and U. Aickelin. 2011. "Modelling Electricity Consumption in Office Buildings: An Agent Based Approach." *Energy and Buildings* 43(10): 2882–2892.
22. Zhou, X., D. Yan, T. Hong, and X. Ren. 2015. "Data Analysis and Stochastic Modeling of Lighting Energy Use in Large Office Buildings in China." *Energy and Buildings* 86: 275–287.
23. Liang, X., T. Hong, and G. Q. Shen. 2016. "Occupancy Data Analytics and Prediction: A Case Study." *Building and Environment* 102: 179–192.
24. Sun, K., D. Yan, T. Hong, and S. Guo. 2014. "Stochastic Modeling of Overtime Ooccupancy and Its Application in Building Energy Simulation and Calibration." *Building and Environment* 79: 1–12.
25. Ryan, T. and J. S. Vipperman. 2013. "Incorporation of Scheduling and Adaptive Historical Data in the Sensor-Utility-Network Method for Occupancy Estimation." *Energy and Buildings* 61: 88–92.
26. D'Oca, S., and T. Hong. 2015. "Occupancy Schedules Learning Process through a Data Mining Framework." *Energy and Buildings* 88: 395–408.
27. Wang, D., C. C. Federspiel, and F. Rubinstein. 2005. "Modeling Occupancy in Single Person Offices." *Energy and Buildings* 37(2):121–126.
28. Samuelson, H. W., A. Ghorayshi, and C. Reinhart. 2016. "Post-Occupancy Evaluation and Partial-Calibration of 18 Design-Phase Energy Models." In P*roceedings ASHRAE/IBPSA-USA Building Simulation Conference 2014*, Atlanta, September 10 -12 2014; pp. 168-176
29. Aydinalp, M., V. I. Ugursal, and A. S. Fung. 2002. "Modeling of the Appliance, Lighting, and Space-Cooling Energy

Consumptions in the Residential Sector Using Neural Networks." *Applied Energy* 71(2): 87–110.

30. Santamouris, M., K. Kapsis, D. Korres, I. Livada, C. Pavlou, and M. N. Assimakopoulos. 2007. "On the Relation between the Energy and Social Characteristics of the Residential Sector." *Energy and Buildings* 39(8): 893–905.
31. Filogamo, L., G. Peri, G. Rizzo, and A. Giaccone. 2014. "On the Classification of Large Residential Buildings Stocks by Sample Typologies for Energy Planning Purposes." *Applied Energy* 135: 825–835.
32. Sokol, J., C. C. Davila, and C. F. Reinhart. 2017. "Validation of a Bayesian-Based Method for Defining Residential Archetypes in Urban Building Energy Models." *Energy and Buildings* 134: 11–24.
33. Peng, C., D. Yan, R. Wu, C. Wang, X. Zhou, and Y. Jiang. 2012. "Quantitative Description and Simulation of Human Behavior in Residential Buildings." *Building Simulation* 5(2): 85–94.
34. Pan, S., X. Wang, Y. Wei, X. Zhang, C. Gal, G. Ren, D. Yan et al. 2017. "Cluster Analysis for Occupant-Behavior Based Electricity Load Patterns in Buildings: A Case Study in Shanghai Residences." *Building Simulation*10(6): 889–898.
35. Majcen, D., L. C. M. Itard, and H. Visscher. 2013. "Theoretical vs. Actual Energy Consumption of Labelled Dwellings in the Netherlands: Discrepancies and Policy Implications." *Energy Policy* 54: 125–136.
36. Branco, G., B. Lachal, P. Gallinelli, and W. Weber. 2004. "Predicted versus Observed Heat Consumption of a Low Energy Multifamily Complex in Switzerland Based on Long-Term Experimental Data." *Energy and Buildings* 36(6): 543–555.
37. Godoy-Shimizu, D., J. Palmer, and N. Terry. 2014. "What Can We Learn from the Household Electricity Survey?" *Buildings* 4(4): 737–761.
38. De Wit, S., and G. Augenbroe. 2002. "Analysis of Uncertainty in Building Design Evaluations and its Implications." *Energy and Buildings* 34(9): 951–958.
39. Hopfe, C. J., G. Augenbroe, and J. Hensen. 2013. "Multi-criteria Decision Making under Uncertainty in Building Performance Assessment." *Building and Environment* 69: 81–90.
40. Machairas, V., A. Tsangrassoulis, and K. Axarli. 2014. "Algorithms for Optimization of Building Design: A Review." *Renewable and Sustainable Energy Reviews* 31: 101–112.
41. Ihaka, Ross, and R. Gentleman. "R: a language for data analysis and graphics." *Journal of Computational and Graphical Statistics 5*, no. 3 (1996): 299-314.
42. Reinhart, C., T. Dogan, J. A. Jakubiec, T. Rakha, and A. Sang. 2013. "Umi-an Urban Simulation Environment for Building Energy Use, Daylighting and Walkability." In 13th Conference of International Building Performance Simulation Association, 476-483. Chambery, France, August 26, 2013.
43. What is Pecan Street; 2014. http://www.pecanstreet.org/about/what-ispecan-street-inc. Retrieved from Ihaka, Ross, and R. Gentleman. "R: a language for data analysis and graphics." *Journal of Computational and Graphical Statistics 5*, no. 3 (1996): 299-314.
44. Rhodes, J.D., C. R. Upshaw, C. B. Harris, C. M. Meehan, D. A. Walling, P. A. Navrátil et al. 2014. "Experimental and Data Collection Methods for a Large-Scale Smart Grid Deployment: Methods and First Results." *Energy* 65: 462–71.
45. Bouveyron, C., E. Côme, and J. Jacques. 2015. "The Discriminative Functional Mixture Model for a Comparative Analysis of Bike Sharing Systems." *The Annals of Applied Statistics* 9(4): 1726–1760.
46. Bishop, C. M. *Pattern Recognition and Machine Learning.* (New York: Springer, 2006)
47. Dent, I., T. Craig, U. Aickelin, and T. Rodden. 2012. "An Approach for Assessing Clustering of Households by Electricity Usage." In *Proceedings 12th Annual Workshop Computational Intelligence (UKCI)*, pp. 1-4, Sep. 2012.
48. ASHRAE (American Society of Heating, Refrigerating and Air Conditioning Engineers). 1989. *90.1: Energy Efficient Design of New Buildings Except New Low Rise Residential Buildings*. Atlanta, GA: ASHRAE.
49. Rea, M. S. (ed). *The IESNA Lighting Handbook: Reference and Application*, 9th. (New York: Illuminating Engineering Society of North America, 2000).
50. Hunt, D. R. G. 1980. "Predicting Artificial Lighting Use—A Method Based upon Observed Patterns of Behaviour." *Lighting Research & Technology* 1(1): 7–14.
51. Haves, P., and P. J. Littlefair. 1988. "Daylight in Dynamic Thermal Modelling Programs: Case Study." *Building Services Engineering Research and Technology* 9(4): 183–188.
52. Reinhart, C. F. 2004. "Lightswitch-2002: A Model for Manual and Automated Control of Electric Lighting and Blinds." *Solar Energy* 77(1): 15–28.
53. Newsham, G. R., A. Mahdavi, and I. Beausoleil-Morrison. 1995. "Lightswitch: A Stochastic Model for Predicting Office Lighting Energy Consumption." In Proceedings of Right Light Three, the Third European Conference on Energy Efficient Lighting, Newcastle-upon-Tyne, Country, June 1995, 60-66.
54. Dogan, T., and C. Reinhart. 2017. "Shoeboxer: An Algorithm for Abstracted Rapid Multi-zone Urban Building Energy Model Generation and Simulation." *Energy and Buildings* 140: 140–153.
55. Crawley, D. B., L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, and R. K. Strand, et al. 2001. "EnergyPlus: Creating a New-Generation Building Energy Simulation Program." *Energy and Buildings* 33(4): 319–331.
56. Retrieved from http://weather.whiteboxtechnologies.com
57. Merkblatt, S. I. A. "2024: Standard-Nutzungsbedingungen für die Energie-und Gebäudetechnik." Zürich: Swiss Society of Engineers and Architects (2006).

Rawad El Kontar is a Master of Science in Architecture student and Research Assistant in the Performative Praxis Lab (PPL) in the School of Architecture at Syracuse University. He received his Bachelor of Architecture degree from the Lebanese American University in 2015 and is a practicing architect. His research interests include performance-based design and building performance simulation.

Tarek Rakha is an Assistant Professor of Architecture at Syracuse University. He directs the PPL, which focuses on three areas of research: sustainable urban mobility and outdoor thermal comfort, daylighting and energy efficiency in buildings, and building envelope diagnostics using drones. Rakha holds a PhD in Building Technology from the Massachusetts Institute of Technology.