

Probabilistic Modeling of Commercial Building Occupancy Patterns Using Location-Based Map Data

Rawad El Kontar¹, Jing Wang¹, Xin Jin¹, Jennifer King¹, and Tarek Rakha²

¹ National Renewable Energy Laboratory and ² Georgia Institute of Technology

ABSTRACT

Considering occupancy patterns is crucial to simulate buildings' energy use. Current energy models use inputs that simplify the actual diversity in occupancy into static occupancy patterns and are not able to represent the numerous variations in occupancy patterns between buildings and across different locations. Recently, inferring occupancy schedules from metered electricity consumption data was used to model occupancy in commercial buildings. However, the translation from metered data to occupancy schedules requires many assumptions that might not capture the reality, and the process is hindered by the availability of data from advanced metering infrastructure. With the development of information technologies, occupancy modeling should not be limited to traditional approaches. The prevalence of social networks and location services with real-time user feedback provides publicly accessible data via Maps Application Programming Interfaces (APIs) such as Google Maps, SafeGraph, Mapbox, Foursquare, etc. This paper presents an automated framework for modeling parametric occupancy patterns using such APIs to calibrate commercial district buildings' energy models. This process includes three main steps: data extraction and processing, parametric schedules generation, and schedules integration. We demonstrated this framework in districts where we used maps API to generate more accurate behavioral patterns for operations and electric vehicle charging events. We used these patterns to determine differences in energy use across key sociodemographic and spatial parameters. The presented method has the potential for worldwide applications. Users can utilize this framework to extract data for selected locations of interest to create more realistic behavioral patterns for commercial facilities across different districts.

Introduction

People are moving to live in cities more than ever before; more than half of the world's population lives in urban areas, and this is expected to increase to 68 % by 2050 (United Nations, 2019). With this unprecedented urban development, cities now consume approximately two-thirds of the world's energy and consequently account for 70% of the global Carbon emissions (United Nations, 2020). In the U.S., buildings account for 40 % of carbon production (U.S. Energy Information Administration, 2018), and the role of Urban districts design is very influential in facilitating the implementation of building efficiency policies and carbon emission reduction plans (Data-Driven EnviroLab & New Climate Institute, 2020). There is an urgent need to inform design decisions to improve energy efficiency while enhancing the lives of individuals and districts that consume and produce energy.

In the meantime, we face rapid advancements in technologies that are immersed more than ever in our daily lives and are constantly changing our daily activities and behavioral trends. With all the technological advancements, it is necessary to capture a variety of behavioral trends to accurately model occupant activities and behavior. These behavioral trends involve various choices that impact urban life and, specifically, buildings' energy use. Such decisions can include

work travel commute preferences and personal preferences for visiting different establishments such as restaurants, shopping stores, and other commercial buildings. These choices highly affect occupancy patterns, as well as occupants' interaction with building lighting, equipment, heating, and cooling systems that determine the temporal energy use behavior of buildings.

Consequently, a vast amount of data is being collected to capture these behaviors and activities; robust information technologies can capture different types of information enabled by advances in sensors, ubiquitous communications, distributed computing, and deep learning algorithms. For instance, information can be measured data such as energy use and temperature or can be retrieved from surveys like demographic, economic, and categorical information. Moreover, the prevalence of social networks and location services with real-time user feedback also provides data for individuals' activity-related choices, including mobility and building occupancy patterns. With more access to data, buildings energy modelers can develop detailed models and perform a wide range of analyses. Many open-source simulation engines have been developed to perform building and urban scale energy simulations (Crawley et al., 2008).

To simulate how a building's electricity demand varies by time of day, current energy urban energy models (UBEMs) estimate the load patterns of building systems based on inputs that characterize the building properties. The generation of such energy models requires data inputs for building geometries and non-geometric inputs that include occupancy schedules, which significantly shape internal loads. Therefore, identifying building occupancy patterns is essential for modeling accurate occupant energy use behavior and performing building energy analyses. It is challenging to predict occupancy schedules due to their stochastic nature and variability based on different locations and building types. Currently, many energy modelers utilize ASHRAE 90.1 defined profiles for occupancy-related input. These inputs simplify the actual diversity in occupancy into static occupancy patterns and cannot represent the numerous variations in occupancy patterns between buildings, leading to unrealistic energy simulation, especially at the urban district scale (Wilke et al. 2013). This results in simulations where all occupants perform identical actions, leading to erroneous hourly demand peaks (He et al. 2015) and ultimately misrepresenting urban energy demands. Hence, uncertainty in defining behavioral and occupancy patterns is a major reason for discrepancies in simulated energy when compared to measured data. At the district scale, unreal coincident peak occupancy hours caused by static schedules will overestimate the peak demand and, consequently, misinform decision-making when designing district energy systems.

Many methods have been used in the past to extract the occupancy for the building energy model. One of these methods is utilizing sensors that detect occupancy in buildings, such as movement and sound-based sensors (Agarwal et al. 2010) (Dong et al. 2019). Unfortunately, such sensors can only provide the occupancy presence information without the count of people at a specific time of the day. Therefore, they fall short of informing the inputs for building energy modeling and analysis (Pang et al. 2020).

In response to this shortcoming, vision-based technologies such as RGB and thermal cameras (Jazizadeh and Jung 2018) and indoor air technologies such as CO₂ sensors (Jin et al. 2018) were developed. These technologies support the detection sensors to determine the number of people in a specific location. However, such coupled sensors have critical shortcomings such as cost and privacy that hinder the scalability of installing such technologies to gather data at a district and urban/ nation scale.

In more recent approaches, extrapolating operational hours from metered electricity consumption data was used to model occupancy in commercial buildings (Bianchi et al., 2020).

However, the translation from metered data to occupancy schedules requires many assumptions that might not capture the reality. In addition, this process is also hindered by the availability of advanced metering infrastructure data.

The studies mentioned above do show potential to extract building occupancy patterns; however, their shortcomings are significant. Therefore, new data sources for building occupant behavior extraction should be explored. For example, with the release of the American Time of Use data (ATUS) for residential buildings; parametric and stochastic occupancy models have been developed and integrated into residential building modeling frameworks (Chen et al., 2021). These models prove to help add realistic variability in assumed home occupancy profiles rather than smoothed/averaged profiles, leading to more realistic variability in residential load profile predictions. This is important for achieving realistic diversity factors when building loads are aggregated to predict overall load profiles, peak loads, etc. On the other hand, similar data sets for commercial building are currently not available and is more challenging to acquire, and only very few buildings' sensor data can be found with the limited feasibility of installing advanced metering infrastructure and sensors in commercial building all over the nation.

With the development of information technologies in the internet-of-things era, occupancy modeling should not be limited to traditional approaches. As more than a billion people have come to rely on Maps to go from one place to another. Many apps have been built on top of such data that people use and rely on nowadays. Individuals use such map apps to drive to a certain place, such information is anonymously collected and used to determine how busy that place is. The companies that provide location services and Map Application Programming Interfaces (APIs) know the location of the users at all times. Individuals are considered as a phone; if an individual GPS service is enabled then, the phone is constantly tracking their location. Information from different phones is summarized to determine the occupancy of an establishment. These phone-based application uses anonymized location data, real-time searches, and past statistics to determine how busy a place is. If there's not enough data, the system uses past statistics to predict busy times. In this process, machine learning algorithms help extrapolate past and incomplete data to arrive at more accurate estimates. This data is updated constantly; within short time intervals, applications are collecting new information about the world in real time and updating the predictions of occupancy levels. With the significant number of users, hundreds of unique information contributions every second, and robust machine learning techniques, the derived statistics are significant and can scale across our communities.

The prevalence of social networks and location services with real-time user feedback can be utilized to a) provide accessible data either publicly or through purchased licenses via Map APIs such as Google Maps, Mapbox, Foursquare, Bing Maps, etc. b) create new opportunities for defining variable commercial building occupancy probability distributions based on such data across different building types and locations and then using the distributions to generate variability in commercial building occupancy across a district. To address this gap, this paper presents an automated framework for modeling parametric occupancy schedules using Location-based data from Foursquare TomTom API that utilizes Foursquare data (<https://foursquare.com/>) and SafeGraph API foot traffic data (<https://www.safegraph.com/>) to calibrate commercial district buildings' energy models. We demonstrated this framework by extracting such data from different areas in Colorado, USA, and deriving probability distribution models that summarize the operation hours and occupancy patterns across the commercial building in Colorado. The results include comparing different summarized occupancy patterns across the key location. We also present a case study of modeling a district in Denver and

utilizing the derived probability distribution models to model more accurate occupancy behaviors and achieve variability in energy demand profiles.

Methodology

An automated framework is developed for modeling parametric occupancy schedules and integrating such schedules in commercial district buildings' energy models. This process includes three main steps: data extraction and processing, parametric schedules generation, and schedules integration (Figure 1). In the data extraction and processing step, we run automated HTTP API requests to retrieve opening hours data for various commercial building types and specific locations selected by the user. The schedules are then generated by processing the data and creating a temporal probability distribution for each building type representing the deviation of occupancy throughout each day of the week. In the last step, we integrate the schedules as inputs to the developed urban energy model, where occupancy schedules are stochastically assigned to each building model to represent the diversity of occupancy patterns more accurately in the district.

This framework is applied and using two different types of data. 1) opening hours data derived from Foursquare data and TomTom API 2) hourly occupancy data derived from SafeGraph API – foot traffic patterns data.

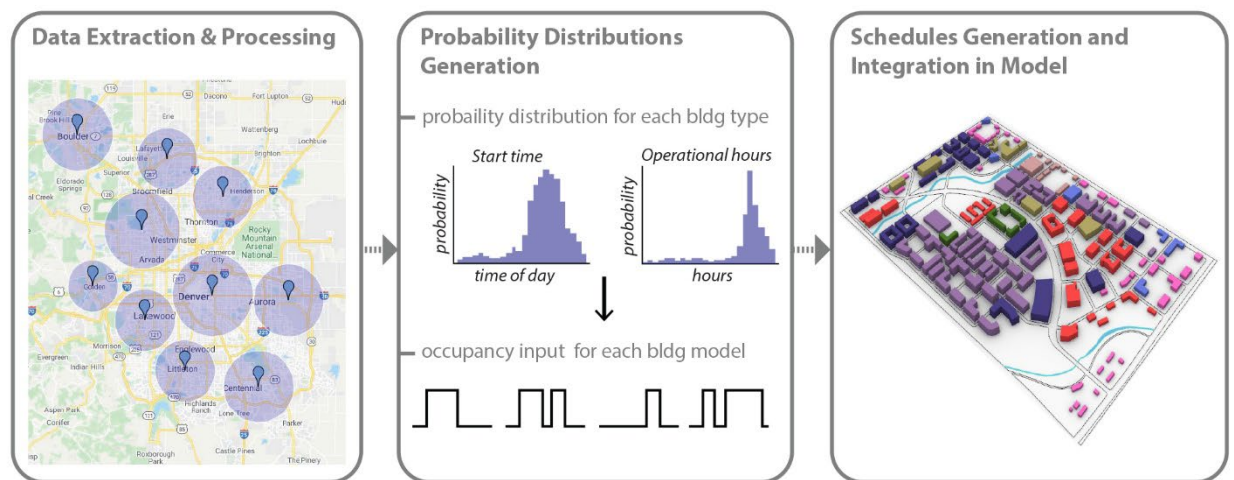


Figure 1. Proposed automated framework

Data extraction and processing

For data extraction, users can define a set of locations by specifying a point coordinate following the longitude and latitude coordinate system and a radius corresponding to each defined point. The output of this process is described in Figure 2, where different areas are defined, preparing the data extraction process for 10 locations of interest in Colorado.

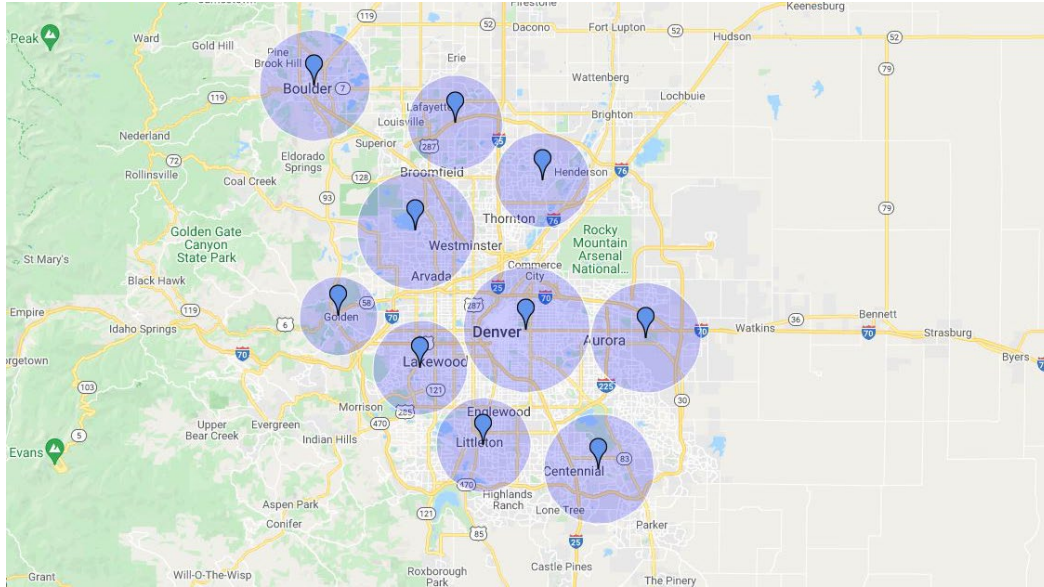


Figure 2. Selecting area of interest defined by longitude, latitude coordinates, and radius.

Then an HTTP API request is sent to the targeted Maps API. To retrieve a response with data from the specified area for building operational hours, we used the TomTom API request to retrieve opening hours for different building types. The opening hours include start and end times for each week of the day. The data source is Foursquare data. For retrieving the hourly occupancy profile, we used the SafeGraph data and API to retrieve patterns data of occupancy for different weeks of the day. This raw data includes information for all the establishments that can be found within this radius. We then extract opening hours data and group it by defined commercial building type.

Probability distributions generation

The schedules are generated by processing the extracted data and creating a temporal probability distribution for each building type representing the deviation of occupancy throughout each day of the week. When modeling urban districts, daily building hours of operation times and occupancy levels vary from building to building and from day to day. By aggregating operation time and occupancy levels for all analogous building types, it is possible to represent the building variation as a whole, grouping all hours of operation or hourly occupancy configurations to a single probability distribution per building type. This approach allows for more flexible and adaptable manipulation of building operation and hourly occupancy for a large group of buildings while preserving its diversity.

The probability distributions are computed by calculating the relative frequencies of the unique values. The relative frequency is calculated with the following formula:

$$\text{Relative Frequency} = \frac{f}{n} \quad (1)$$

Where f is the subgroup frequency, the number of times the data occurred in an observation; n is the total frequency.

For operational hours, we grouped the sample of opening hours data (defined by a start time and operation duration) for each building type. The probability for each start time and hours of operation is computed, forming a probability distribution of the hours of operation for each building type for a typical weekday and a weekend. Figure 3 shows the extracted probability distributions of restaurants' weekday start times and operational hours derived from a sample of ~560 restaurants located in the defined areas of interest.

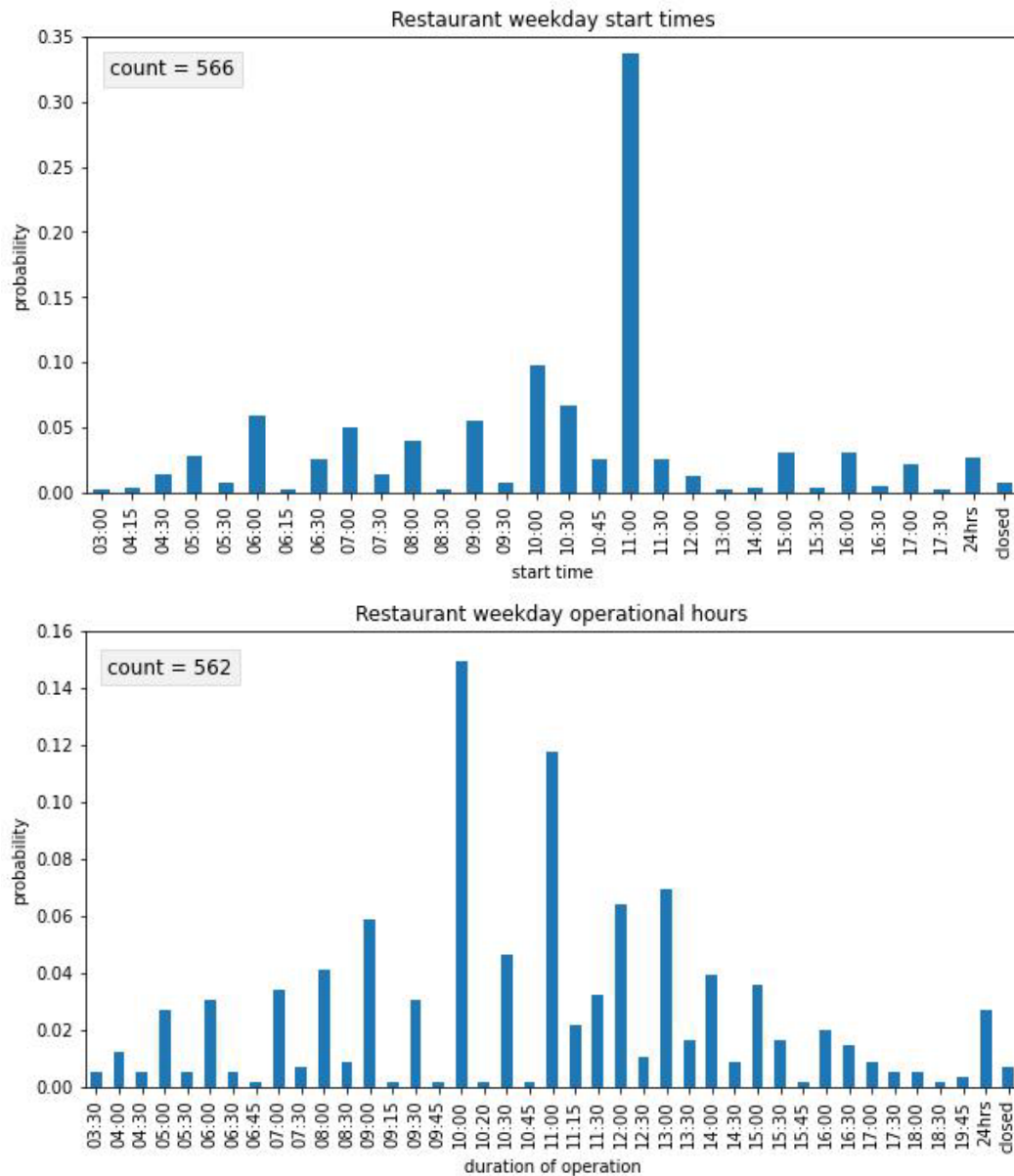


Figure 3. Restaurant weekdays start times and operational hours probability distribution

For extracting the hourly occupancy profile, a probability of occupancy is computed for each hour of a typical weekday and weekend across the commercial building types. Figure 4 shows restaurants' occupancy probability distribution over the hours of a typical weekday derived from the popular time data for all the restaurants in the Denver downtown area.

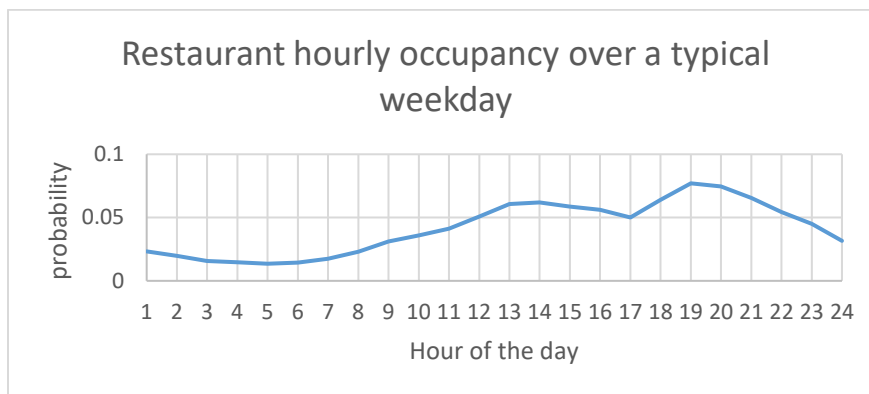


Figure 4. Hourly occupancy probability distribution for a typical weekday

Schedules generation and integration in energy models

Utilizing the derived probability distributions, we can now generate stochastic occupancy profile inputs for urban energy models. In this paper, we streamline the process of generating inputs for URBANopt™ (El Kontar, et al., 2020) energy models using OpenStudio workflows and measures. These inputs include the building occupancy profiles and building hours of operation inputs generated from the probability distribution we computed from the previous step. Equipment use schedules are also derived using methods defined in the OpenStudio standards (<https://github.com/NREL/openstudio-standards>) and described in (Bianchi et al., 2020).

The derived schedules are implemented by updating OpenStudio ScheduleDay objects. The generated occupancy schedules were fed into OpenStudio, and in this process, Baseline schedules for equipment, HVAC, and lighting use schedules are modified as a function of the fed hours of operation and occupancy. Since URBANopt utilizes OpenStudio workflows, we were able to automate updating the schedules for all buildings of the district we are modeling via the URBANopt platform.

Results and Discussions

This section is divided into two parts:

- 1) Part 1: Presents a case study where we compared a baseline scenario that represents default occupancy inputs for the different building types with an Updated Scenario where we utilized the process of generating probability distribution of commercial building hours of operation, generating parametric schedules of operation, and integrating them into a mixed-used district designed in Denver, Colorado to analyze the resulting energy use results.
- 2) Part 2: Presents the results of investigating different approaches when applying steps 1 and 2 in the methodology for extracting and generating probability distribution of occupancy patterns. The generated probability distribution results are compared across different locations in Colorado and various commercial building types.

Case Study: Framework implementation in a district designed in Denver, Colorado, US

We demonstrated this framework on a large, mixed-use district under development in Denver, Colorado, USA. We selected a sample of 100 buildings for each commercial building type from 10 nearby cities in the same metropolitan area. This sums up to a total sample of approximately 1000 buildings for each building type. Users can specify the sample number and select the data extraction areas, defined by a list of longitude/latitude points and radiuses. Subsequently, a query script is executed to send API requests and retrieve buildings' operational hours for each hour of the day and all days of the week. This data is processed to generate occupancy temporal probability distributions across different hours of the day and days of the week. We then developed a district energy model using URBANopt and fed the probability distribution into the model to assign the parametric occupancy schedules to each building model in the district. These steps are automated via a developed python module, which can output updated district energy packaged inputs for specific building archetypes that can seamlessly integrate into district energy analysis platforms like the URBANopt platform via the OpenStudio workflows described in step 3 of the Methodology section.

The community is currently under construction and will have 148 large commercial buildings, including offices, shopping malls, retail stores, hotels, schools, hospitals, etc.



Figure 5. Three-dimensional rendering map of the mixed-use community.

We adopted URBANoptTM (El Kontar, et al., 2020) to build a high-fidelity physics-based model for all the buildings in this community. Then we created two Scenarios a **Baseline Scenario** where we used the default static schedules defined by ASHRA 90.1 and an **Updated Scenario** where we utilized our framework to generate variability in the hours of operation between the buildings.

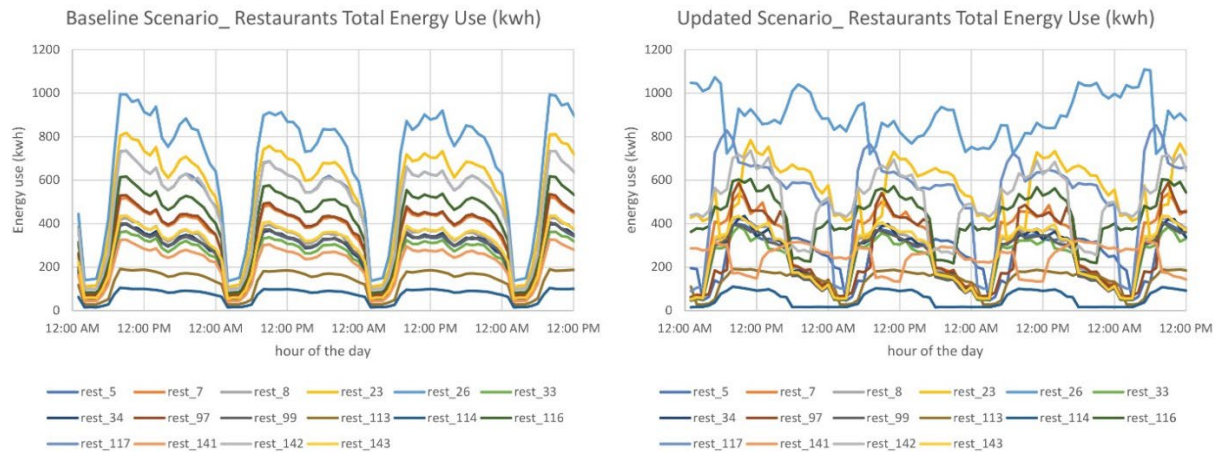


Figure 6. Comparing Baseline Scenario vs Updated Scenario

Figure 6 compares the energy use results of the restaurant in the district across the two scenarios and over the first three days of January. The Baseline Scenario shows identical energy use patterns with coincident peaks of energy demands. In contrast, the Updated Scenario shows that our methodology was successful in introducing variability in energy use demands based on more realistic operational hours.

Figure 7 compares aggregate and averaged energy use results across the two scenarios, where the 16 restaurant buildings in our district are aggregated.

These results demonstrated that energy demand in the updated scenario is more evenly distributed over the course of the day, and this reflects the variability in the restaurants' schedules within the district. The updated scenario shows a similar peak during mid-day to the baseline scenario. However, the baseline scenario shows a significant dip in energy use in the late-night hours and higher peak demand in the late afternoon hours. These extremities are caused by the identical representation of occupancy schedules modeled in the baseline scenario.

Buildings in the baseline scenario have coincident afternoon peaks and morning dips in energy use. When aggregated, such profiles misestimate the energy demand during critical times of the day. For example, when the energy cost is highest, the over-estimated energy demand in the afternoon misleads engineers to making suitable design decisions. A more realistic representation of these profiles helps better inform the design of building systems and the sizing of distributed energy resources and their operation. Moreover, they inform the deployment of more accurate load-shifting/control strategies for the buildings in the district.

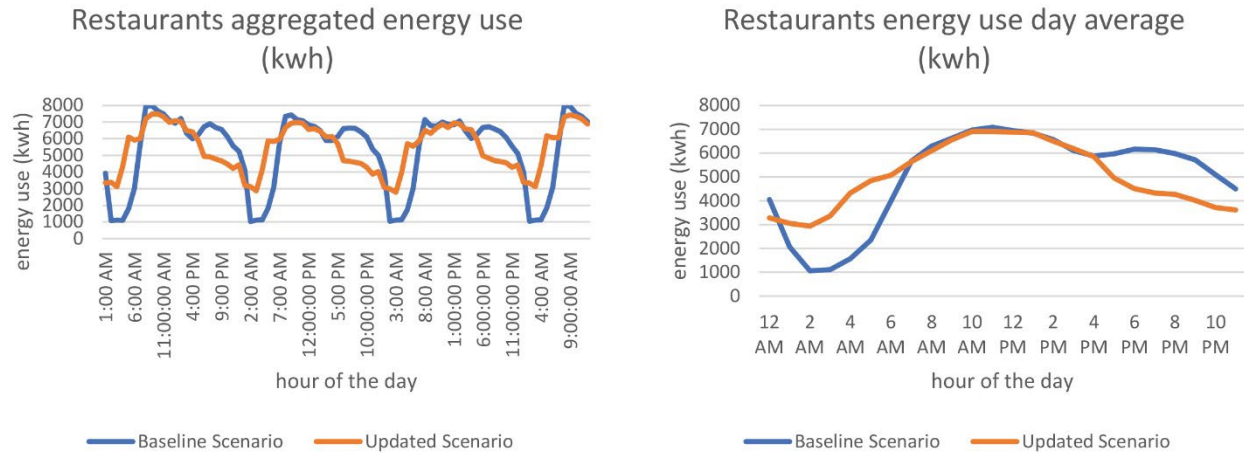


Figure 7. Aggregate and average energy use results comparison between the 2 scenarios

Occupancy probability distributions investigation in CO, USA

This section compares different probability distributions of occupancy and building operation hours across different commercial building types and different data extraction strategies.

In the developed framework, users have the flexibility to extract data based on the location of their preference. In our case study, we choose to extract data from 10 locations near our site, as shown in Figure 2. However, the users can investigate different approaches to data extraction locations. For example, suppose the urban energy modeler is designing a downtown district with dense urban living characteristics. In that case, they might choose sample occupancy schedules from areas with such characteristics. On the hand, users might want to model occupancy schedules that represent the behavior and patterns of rural areas.

Figure 8 shows the occupancy probability extracted from downtowns versus rural areas in Colorado and compared across three different building types (restaurants, offices, and food stores). Based on Figure 8, we can see differences in occupancy patterns when comparing downtown building occupancy with occupancy patterns in rural areas. The restaurants' occupancy in downtown areas is more evenly distributed throughout the day and shows a higher occupancy rate during night hours. In contrast, the occupancy patterns for restaurants in rural areas have significant peak hours of occupancy at lunch and after a work time. Similarly, for offices' occupancy distribution, downtown offices have more variability and include offices that operate during the night while rural areas operate mainly during the day. On the other hand, food store occupancy patterns are similar, with a small occupancy spike in downtown areas compared to rural areas. These results show that this method can be used to capture occupancy preferences and can be used to characterize urban energy models with more accurate occupancy and behavioral schedules.

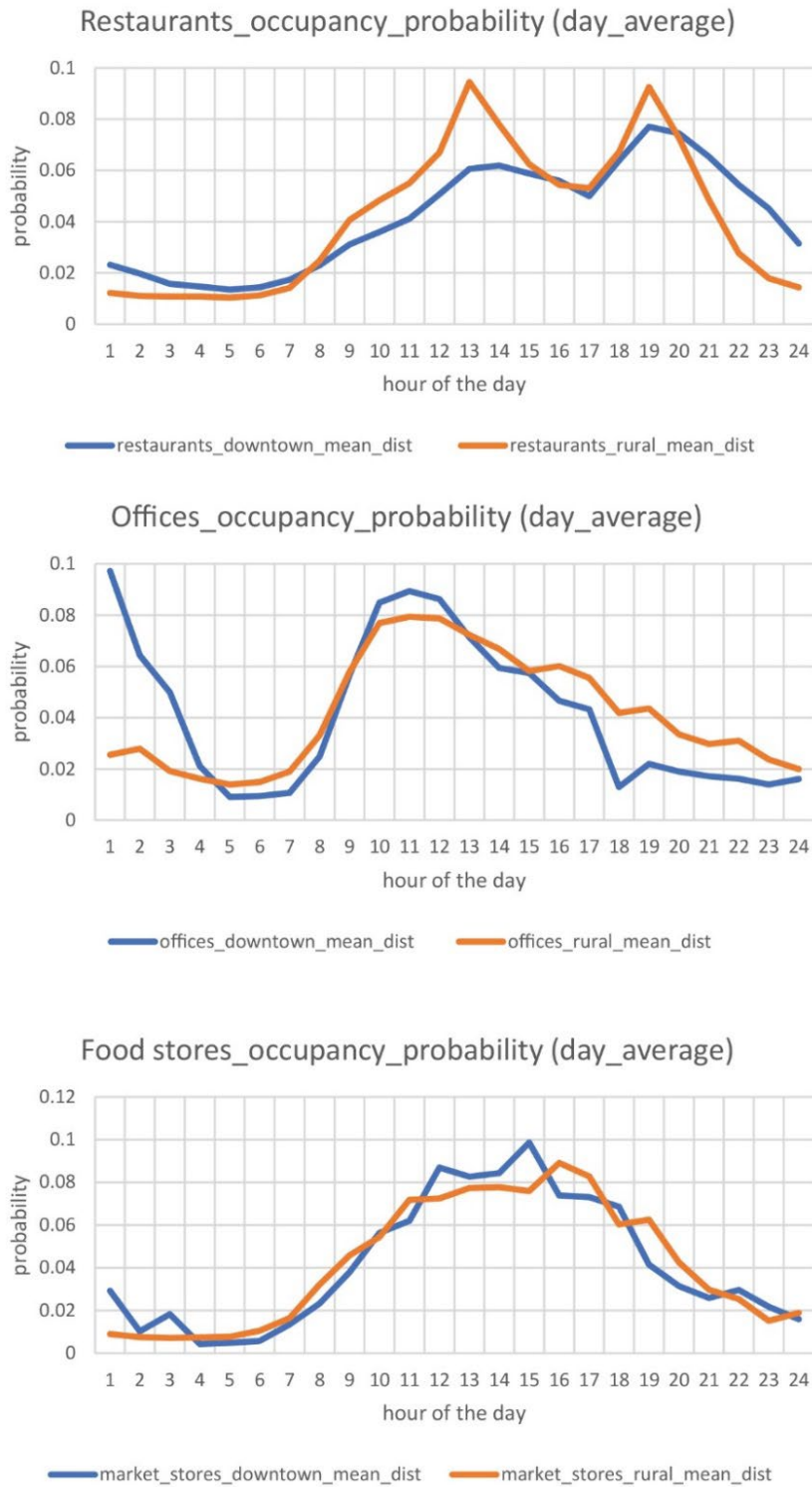


Figure 8 - Occupancy probability distribution comparison between the downtown and rural areas

Conclusion

To analyze the impact of the proposed framework, a scenario with traditional occupancy modeling approaches (following reference profiles from ASHRAE Standards) is constructed and compared to a scenario where we implemented our proposed framework. The generated occupancy profiles from both scenarios are compared, and the impact on energy simulation results is evaluated. Results reveal significant differences in energy load profiles and load distributions with the updated heterogeneous occupancy profiles. The study showcases the effect of occupancy on energy consumption and illustrates the importance of capturing the variability of occupancy schedules in simulation tools to enhance the accuracy of outcomes.

The presented contribution helps model more accurate occupancy and building operational hour inputs and has the potential for worldwide applications. Private and public establishments data and location data are now collected and visualized on a global scale. Users can now utilize this framework to extract data for selected locations of interest to create more realistic occupancy schedules for commercial facilities and calibrate their urban energy models. This innovative approach will benefit utility companies, grid operators, urban planners, and energy modelers, seeking to improve the accuracy of district energy models and better inform the design decisions at a district scale.

Acknowledgment

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding is provided by NREL's Laboratory Directed Research and Development program. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work or allow others to do so, for U.S. Government purposes.

References

- United Nations. World Urbanization Prospects: The 2018 Revision. Technical report, New York, NY: United Nations, 2019.
- United Nations Human Settlements Programme. World Cities Report 2020: The Value of Sustainable Urbanization. Technical report, Nairobi, Kenya: United Nations, 2020.
- U.S. Energy Information Administration. 2019. International Energy Outlook 2019: With Projections to 2050. Technical report.
- Data-Driven EnviroLab & NewClimate Institute. Accelerating Net Zero: Exploring Cities, Regions, and Companies' Pledges to Decarbonise. Technical report, Data-Driven EnviroLab & NewClimate Institute, 2020.

- Crawley, Drury B., Linda K. Lawrie, Frederick C. Winkelmann, Walter F. Buhl, Y. Joe Huang, Curtis O. Pedersen, Richard K. Strand et al. "EnergyPlus: creating a new-generation building energy simulation program." *Energy and buildings* 33, no. 4 (2001): 319-331.
- Wilke, Urs, Frédéric Haldi, Jean-Louis Scartezzini, and Darren Robinson. "A bottom-up stochastic model to predict building occupants' time-dependent activities." *Building and Environment* 60 (2013): 254-264.
- He, Miaomiao, Timothy Lee, Simon Taylor, Steven K. Firth, and Kevin J. Lomas. "Coupling a stochastic occupancy model to EnergyPlus to predict hourly thermal demand of a neighbourhood." (2015).
- Agarwal Y, Balaji B, Gupta R, Lyles J, Wei M, Weng T (2010).Occupancy-driven energy management for smart building automation. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building—BuildSys'10, Zurich, Switzerland.
- Dong B, Prakash V, Feng F, O'Neill Z (2019). A review of smart building sensing system for better indoor environment control. *Energy and Buildings*, 199: 29–46.
- Pang Z, Chen Y, Zhang J, O'Neill Z, Cheng H, Dong B (2020). Nationwide energy saving potential evaluation for office buildings with occupant-based building controls. Paper presented in 2020 ASHRAE Winter Meeting, Orlando, FL, USA.
- Jazizadeh F, Jung W (2018). Personalized thermal comfort inference using RGB video images for distributed HVAC control. *Applied Energy*, 220: 829–841.
- Jin M, Bekiaris-Liberis N, Weekly K, Spanos CJ, Bayen AM (2018). Occupancy detection via environmental sensing. *IEEE Transactions on Automation Science and Engineering*, 15: 443–455.
- Bianchi, Carlo et al. Modeling occupancy-driven building loads for large and diversified building stocks through the use of parametric schedules. U.S.: N. p., 2020.
- Chen, J., Adhikari, R., Wilson, E., Robertson, J., Fontanini, A., Polly, B. and Olawale, O., 2021. Stochastic simulation of residential building occupant-driven energy use in a bottom-up model of the US housing stock. arXiv preprint arXiv:2111.01881.
- El Kontar, Rawad, et al. "URBANopt: An Open-Source Software Development Kit for Community and Urban District Energy Modeling" ASHRAE Topical Conference Proceedings. American, Inc., 2020.