Generating traffic-based building occupancy schedules in Chattanooga, Tennessee from a grid of traffic sensors

Anne S Berres[1], Brett Bass[2], Joshua R New[1], Piljae Im[1], Marie Urban[1], Jibonananda Sanyal[1]
[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2]Bredesen Center, University of Tennessee at Knoxville, Knoxville, TN, USA

Abstract

Building occupancy significantly impacts energy use, timing for demand impacts, and is a significant source of uncertainty in building energy models. There are relatively few sources that define building occupancy schedules and number of occupants per building or space type. More importantly, these sources define traditional schedules that are likely not to reflect the true occupancy of a given building. We construct traffic-based occupancy schedules which are more responsive to changes in mobility patterns, and which can realistically estimate occupant arrivals, departures, and counts in individual buildings.

Key Innovations

- Source of 2020 traffic data for city-scale population movement
- Travel data used to update occupancy for building energy models
- A sample of 600 buildings were simulated using actual meteorological year weather data for 2019
- These simulations are compared to 15-minute data electricity use of each building from 2019
- Reduced the mean Coefficient of Variation of the Root Mean Squared Error (CV(RSME)) by 79.2% (but increased the median CV(RMSE) by 16.4%).

Practical Implications

Assuming reference or prototypical occupancy in buildings can result in similar simulation results for buildings that vary by 3-5x in total monthly energy consumption, and they generally do not account for changes such as increased telework. To address such uncertainty, consider using traffic or other data to better inform occupancy and equipment schedules.

Introduction

Occupancy significantly impacts energy use, timing for demand impacts, and is a significant source of uncertainty in building energy models. While an American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Multidisciplinary Task Group is attempting to extend occupancy information, there are relatively few sources (e.g, Building America protocols by Hendron and Engebrecht (2010), ASHRAE Standard 90.1 by ASHRAE and Illuminating Engineering Society (IES) (1989), and prototype building models by Deru et al. (2011)) that define building occupancy schedules and number of occupants per building or space type. These sources define traditional schedules that not are likely to reflect the true occupancy of a given building.

The population modeling community continues to explore higher resolution research into local population dynamics where population distribution is developed at the building level using residential surveys (Leasure et al. (2020)). Other research has led to the development of a global learning framework to specifically report ambient building occupancy through the use of open source driven observation models (Stewart et al. (2016); Wang et al. (2021)), and probabilistic risk models that incorporate building occupancy (Silva et al. (2020); United States Geological Survey (2020)). On a separate thread, Sparks et al. (2020) used social media Point of Interest to globally determine day and night by clock hour for an economic answer to day and night, and Lu et al. (2020) recently compared Facebook and Google popularity curves for efficacy of development of 24-hour building occupancy.

There have been a number of works around building occupancy in relation to energy use. Gunay et al. (2017) gathered zone level occupancy data from private offices to evaluate Heating, ventilation, and air conditioning (HVAC) impact using EnergyPlus simulations. Ren et al. (2017) evaluated discrepancies between measured energy demand from smart meters compared with simulation predictions and fixed occupancy schedules. Wang et al. (2020) probabilistically derived building occupancy schedules from lighting power consumption times, and compared them with aggregated measurements. Berres et al. (2019) used a transportation simulation based on annual traffic volumes, as well as more detailed data by day of week (Berres et al. (2019)). They determine agent arrivals and departures based on the simulated vehicle trajectories, and assign the vehicle drivers to the surrounding buildings using a quadtree-based approach for faster performance. Similar works were presented by Qu et al. (2020) and Alharin et al. (2020). This provided promising results, however, since these approaches were based on annual travel information, they cannot model traffic impact, or the recent increase in telecommuting behavior.

In this work, we bridge this gap by using measured traffic data for each day, which provides travel information with much higher fidelity, and at a much higher spatial resolution where intersection-level traffic sensing is available.

Methodology

Traditional building occupancy schedules have very harsh transitions from hour to hour, whereas actual occupancy changes more smoothly in large commercial buildings. Local commute differences, traffic impacts, and quarantine work patterns lead to significant variations in occupancy. With a mobility-informed approach to building occupancy, it is possible to not only generate more realistic building occupancy schedules from historic traffic data, but also to update them in real-time as the traffic sensors provide continuous updates on the state of the traffic system, traffic jams on the adjacent highways, travel times, and vehicle load on city roads. In the future of smart buildings, this could be used to project changes in arrival times, and preemptively activate or better manage energy-consuming appliances within a building to match the projected level of occupancy. It could also be used to inform population modeling used by emergency responders about actual impacted populations at a given time during disasters.

Over the past few years, cities have become increasingly smart and well-connected. Modern sensing technologies are becoming more widespread, and cities are starting to see the merit of obtaining detailed information about city traffic at the intersection level without expending personnel to perform tedious and expensive manual traffic counts. With an ever-increasing density of such sensors, this data is becoming relevant beyond the domain of traffic engineering. This new data is a great opportunity for obtaining information about population movements at an aggregate (i.e. anonymized) level. It is also much less expensive to obtain, and less sensitive to handle than Global Positioning System (GPS) traces of phones or vehicles. Figure 1 provides an overview of our workflow, and it serves as a guide through this section. The workflow is split into two parts (indicated by arrows), with one part starting from traffic count data (left column), and the other part starting from building geometries (right column).

Data Sources

Our approach requires a variety of data sources:

- **Traffic Data**: We collect traffic data from GridSmart sensors which are placed on 96 of Chattanooga's 334 signalized intersections. In the downtown area, 45 sensors are forming a well-connected network. Each sensor uses a fish-eye camera to detect and record every vehicle that travels through the intersection.
- **Building metadata**: All building metadata was aggregated and processed to generate building energy models, simulated on high performance computing resources, and analyzed as part of the "Automatic Building Detection and Energy Model Creation" (AutoBEM) software suite (New et al. (2018)).
- **Building electricity**: The Electric Power Board (EPB) of Chattanooga, TN provided 15-minute, whole-building, electricity use data from advanced metering infrastructure for each building during calendar year 2019.
- **Prototype building models**: Prototype building models are a set of models for 16 building types that represent much of the built environment. Current models represent about 75% of US commercial buildings (U.S. Department of Energy (2019)). The prototype building model schedules are used as a baseline for comparison to the traffic-based schedules produced here.
- **Weather files**: The authors purchased (actual) Meteorological Year (MY) weather files from the nearest airport with meteorological variables corresponding to calendar year 2019.

Building data preparation

The EPB service area is modeled as a whole using the AutoBEM framework (New et al. (2018)); a collection of methods, data sources, and algorithms to generate and simulate building energy models. Building geometries were selected from the Chattanooga region from Microsoft's data-set of more than 125 million buildings across the United States, Microsoft (2018). Building heights were found using Light Detection and Ranging (LiDAR) of the region. These two features provide a general shape to the building. Assignment of a prototype building type and vintage to each individual building filled the remaining parameters required for building energy modeling such as HVAC type, insulation, glazing fraction, occupancy (stock), etc. The prototype buildings were assigned to each individual building by comparing simulated electricity results to measured electricity for each building (Garrison et al. (2019)). The buildings are generated using OpenStudio (U.S. Department of Energy (2020b)) and simulated using EnergyPlus (U.S. Department of Energy (2020a)). The EPB service area is substantially bigger than the area covered by traffic lights (especially those with the traffic sensors),

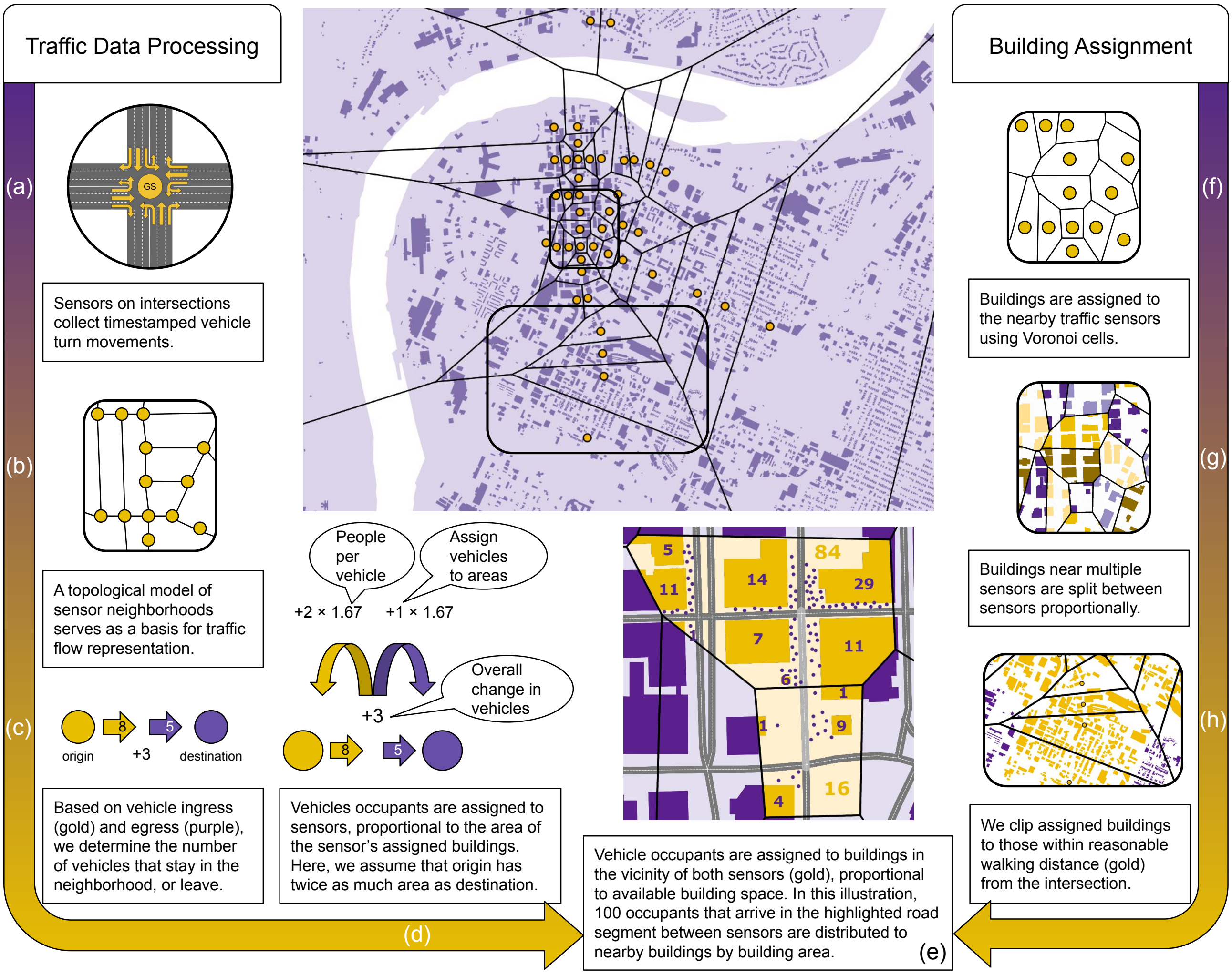

Figure 1: An overview of the methodology presented in this paper, guided by the data workflow. The left-hand side of the figure, (a)-(d), illustrates the traffic data workflow: the data is collected from sensors, and connected through a topological model. Based on vehicle turning movements, we determine the number if vehicles entering the link between two intersections, and the vehicles leaving the link. The difference in vehicles is assigned to the two intersections, relative to their available building space. The vehicle counts are multiplied by 1.67, the average number of occupants per vehicle, to reflect the number of prospective building occupants. The right-hand side of the figure, (f)-(h), illustrates the building workflow: each building is assigned to all nearby intersections, based on their Voronoi cells. If the building footprint intersects with multiple cells, the building area is assigned proportionally to the different intersections based on the amount of overlap between footprint and Voronoi cell. We eliminate buildings which are outside reasonable walking distance. Finally, the two workflows join at the bottom of the figure (e): the prospective building occupants for each intersection are distributed to its assigned buildings (partial or full) based on the available building area.

of more than 178,000 buildings. We clip the extent of the buildings down to those that are broadly near downtown Chattanooga. This step results in a total of 1,893 buildings which are considered in the following steps. These building energy models were then edited to utilize custom occupancy schedules using the *eppy* Python module (Santosh et al. (2016)).

Traffic data preparation

First, we download data for every day of 2020, and process it using an improved version of the official Python Library by GridSmart (2021). This creates tables for each day and intersection, which contain each vehicle's timestamped information on approach and turn direction, as illustrated in Figure 1(a). In traffic engineering terminology, a vehicle approaching an intersection from the South is considered to be "Northbound," meaning that its trajectory is facing North. To obtain a more intuitive nomenclature for further use, we transform the data such that the outcome provides us with origin and destination (O/D) direction. For example, a Northbound that turns left has its origin in the South and destination in the West. Each vehicle's turning movement is represented by an O/D pair. Finally, we compile hourly aggregates of distinct O/D pairs for each intersection.

Topological model for traffic flow

The objective of this work is to determine occupancy data based on traffic. To satisfy this objective, we have to find out how many vehicles that enter an area stay in this area, and how many of them simply travel through. Therefore, we build a topological model which mirrors the sensor neighborhoods. For each sensor in our area of interest (downtown Chattanooga), we add the nearest neighbor in each direction, if such a neighbor exists. These neighbors can be on the neighboring intersection, or a few intersections away if there are no closer sensors. This topological model of the network (Figure 1(b)) serves as a basis to determine traffic flow between sensors, as shown in Figure 1(c). In this example, eight vehicles leave the yellow origin sensor, but only five vehicles arrive at the destination sensor. Three vehicles stayed behind in the area, and the vehicle occupants can be added to nearby buildings. If this number is negative, we remove occupants from buildings accordingly.

Assigning buildings to intersections

Each building is assigned to the closest intersection sensor (Figure 1(f)). Voronoi diagrams are a computational geometry approach (De Berg et al. (1997)) which yields cells that contain all the points which belong to each node (sensor), i.e. points which are closer to their node than to any other node. To assign buildings to intersections, we determine which buildings are part of each intersection's Voronoi cell. As there are many large buildings and relatively small city blocks in downtown Chattanooga, we allow buildings to be associated with multiple intersections. This determination is made in a two-step procedure. First, we check which Voronoi cells the vertices of each building footprint falls into. Then, we determine the intersection of the building footprint with the Voronoi cell, and we compute the area of the building section that is assigned to the cell. A visual representation of this assignment is shown in Figure 1(g).

As some of the Voronoi cells are large due to a decrease in sensor density outside of downtown, we require the buildings to be within a reasonable walking distance. Based on available parking spaces, we chose a distance of 3-4 city blocks as a threshold. This reduced the number of buildings considered in this approach to 1,336. Figure 1(f) illustrates this threshold, using gold for buildings within sensor range, and purple for buildings outside the sensor range.

Finally, we ensure that we do not duplicate any buildings. We address this by assigning appropriate fractions of building area to the intersections they are assigned to, as seen in Figure 1(h). E.g., if 40% of a 1000 m^2 building footprint intersect a cell, we only add 400 m^2 to this cell's available building area.

One of the key advantages of this method is that it suffices to compute the mapping between intersection sensors and buildings once. This step produces multiple outputs: a two-way mapping between buildings and intersections (for fast look-up), available total building area per Voronoi cell, and each building's partial areas for the Voronoi cells it intersects.

Assigning vehicle occupants to intersections

We take this intermediate step to assign the ingress and egress we determined through traffic flow analysis to the intersections. As the traffic flow is determined between pairs of intersections, occupants of incoming or departing vehicles should be distributed between both intersections. We split the number of arriving and departing vehicles according to the available building area at each intersection, as determined in the previous step. This allows us to account for different building sizes and gives us flexibility in building assignments. For example, in Figure 1(d), the origin intersection has twice as much building area as the destination, therefore the three vehicles are split into two for the origin and one for the destination.

Furthermore, we have to account for the number of vehicle occupants. According to a report by the Vehicle Technologies Office (2018), the average vehicle in the United States has 1.67 occupants. While vehicle occupancy varies by demographic, we do not have sufficient data to support differentiation. Therefore, we use a multiplier of 1.67 to convert from the number of vehicles to the number of vehicle occupants.

Assigning occupants to buildings

Within each intersection's Voronoi cell, we distribute vehicle occupants to the surrounding buildings. As in previous steps, we assign vehicle occupants to buildings according to the area of the fraction of the building that lies inside in the Voronoi cell in relation to the overall available building area. Figure 1(e) illustrates both of these steps for the vehicle occupants from one road segment (light gray) between neighboring sensors. The Voronoi cells corresponding to this segment are colored gold, whereas surrounding cells are colored purple. We assume that 100 vehicle occupants arrive on this road segment (purple dots), and enter the surrounding buildings. The southern cell only has about 16% of the available building area between the two cells. Therefore, we assign 16 occupants to the southern cell, and 84 occupants to the northern cell (please note that rounding is only performed for illustrative purposes). This is also indicated by the large gold numbers inside of each cell.

Within each cell, we distribute the occupants between buildings according to the available building area of each building in the cell, relative to the total available building area of the cell. For instance, the top right golden building contains 35% of the available building area in this Voronoi cell (please note that the building footprint does accurately reflect the available area for multi-story buildings). This means that we assign $0.35 \cdot 84 \approx 29$ occupants to this building. We indicate the number of occupants by a purple number on the footprint of each building section. For example, the large building in the right has 11 occupants that are assigned from the northern intersection, and one occupant that is assigned from the southern intersection. Finally, we sum up the occupancy of all building parts in different cells at each time step to obtain the number of occupants for every building at each time.

Creating building occupancy schedules

To create occupancy schedules, we need the timestamped occupancy for each time step. We begin by collecting the hourly occupancy totals for each building. Then, we find the largest value of the timeseries of each building to serve as its maximum occupancy. Then, the hourly totals are divided by the maximum occupancy to obtain relative occupancy (between 0 and 1). Last, we produce one file per building which contains hourly occupancy for each hour of the year. All produced files from this step are used by EnergyPlus to adjust occupancy for the simulation.

Handling data quality issues

There are different causes for data quality issues or incompleteness, such as sensor age (new sensors have less data), power outages, sensor configuration issues (changes or distorted data), or sensor malfunction. We address data gaps by filling in data from the nearest similar days. Due to the variations between different days of the week, this is limited to the same day of the week, and due to seasonality, we further limit this to days that are at most four weeks offset from the missing day. If the missing data cannot be replaced by such days, we skip the corresponding intersection. In these instances, we fall back on a standard occupancy schedule. To address distorted traffic counts, we rescale the incoming and outgoing traffic such that the overall number of vehicles balances out over the course of each day. This was necessary as some intersections have consistently higher traffic counts than others without plausible cause. While this is not optimal for building occupancy, it can be very helpful information for traffic engineers to further optimize the sensor configuration.

Results

In order to evaluate the traffic-based occupancy schedules, we compare their performance in simulating accurate building energy use to that of stock occupancy schedules. To perform this comparison, we set up two simulations which are identical in all aspects except the occupancy schedule. We then compare the simulated energy use for each simulation with measurements for an entire year of data.

Simulations

Previous work had assimilated several sources of building-specific data including building footprints, heights, building types, and other characteristics for the eight-county service territory of EPB. These building energy models were simulated using MY weather data for 2019 from the nearest airport, so this study does not account for microclimate variations. DOE's whole-building flagship simulation engine, EnergyPlus, was used for these simulations which runs primarily single-threaded on desktops without the ability to leverage multi-core Central Processing Units (CPUs) or Graphical Processing Units (GPUs) for most of the physics-oriented computations. Over 150,000 buildings were simulated on Argonne Leadership Computing Facility's (ALCF) Theta supercomputer, currently ranked the world's 39th most powerful supercomputer and one of the top utilizing primarily CPUs, rather than GPUs, for its computing performance. EPB provided 15-minute electricity use for each building for the entire year of 2019. However, many of the traffic sensors we use are new and were not yet in place for large parts of 2019, or were still under calibration. Therefore, we use more recent 2020 data. This choice has the obvious shortcoming that 2020 was an unusual year, however, it is also a chance to explore the resulting differences. To align the data for the two years, we use 365 days of traffic data from December 31, 2019 to December 29, 2020. This corrects the offset in the day of the week, and it accounts for the leap day. We set up EnergyPlus using 2019 climate data. We use stock occupancy schedules for the control simulation, to compare with a simulation using traffic-based occupancy schedules. The traffic-based occupancy schedules regulate the people, lighting, and interior equipment of the building energy models. It is worth noting that the building types assigned to each building were de-

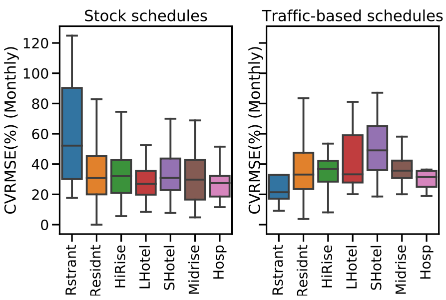

Figure 2: Comparison of errors between simulations using stock occupancy schedules and traffic-based schedules.

termined by classifying buildings based on their electricity use signatures in previous analyses, and they are consistent between the two simulations. This does not always reflect the building's actual building use, but it improves the match of electricity profiles between simulations and measured data.

Comparison metrics

We compare the outputs of both simulations with the measured data. For this comparison, we use the CV(RSME). This is a quantitative metric used for building energy modeling that measures uncertainty in the model compared to real data. The CV(RMSE) for N data points (e.g. number of time steps for each building) is computed as $CV(RMSE) = \frac{1}{\bar{Y}}\sqrt{\frac{\Sigma_{i=1}^{n}(Y_i - \hat{Y})^2}{N}}$ where $\hat{Y}$ is the simulated energy, Y is the measured energy, and $\bar{Y}$ is the mean value of the measurements.

In the following, we will discuss the results of these comparisons. Figure 2 shows a comparison of the CV(RSME) between the simulation using stock occupancy schedules (left) and the simulation using traffic-based occupancy schedules (right). While the median error of the two simulations is quite similar for most building types, the standard deviations (whiskers) and confidence intervals (boxes) change noticeably between the two. Outliers were omitted from the charts for better legibility. The schedules for restaurants shows the most visible difference, with a median of 52 for stock schedules compared to 21 for traffic-based schedules. This constitutes an improvement of 59.6%. This implies that the stock occupancy schedules do not accurately represent the actual schedule of these restaurants. Large and small hotels have noticeably higher errors, which can be explained by a reduction in travel, and therefore reduced hotel visits.

A closer look at the individual scatterplots in Figure 3 reveals more insights about the differences between the two simulations. These scatterplots contrast the errors of the two simulations against each other at the individual building level. Data points above the diagonal indicate a lower error for traffic-based schedules, data points below indicate a higher error for traffic-

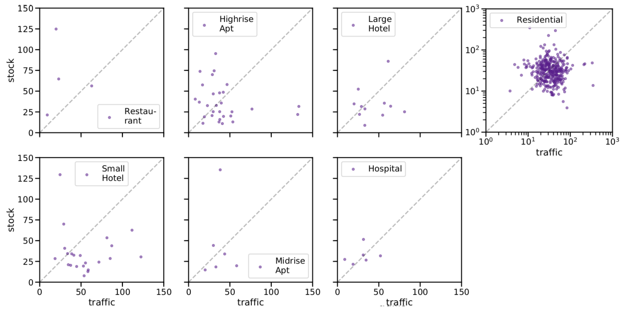

Figure 3: Comparison of errors between simulations using stock occupancy schedules and traffic-based schedules. For better comparability, the axes were limited to the interesting areas.

based schedules. Data points on the diagonal have the same error in both simulations. Some outliers on the diagonal have been cut off to provide a better view of the data points that we are trying to compare.

Traffic-based occupancy schedules generally performed better than stock occupancy for small hotels. For high-rise and mid-rise apartment buildings, both schedules had some buildings in which they outperformed the other. For high-rise apartments, this relationship is balanced, whereas for mid-rise apartments, stock schedules tended to perform better than travel-based occupancy schedules. This can in part be explained by the changes in teleworking across the population due to COVID-19 in 2020, which resulted in a change in commute patterns. Restaurants, hotels, and hospitals all had very few buildings with substantial differences in performance between the simulations. Finally, residential buildings had an extremely wide range in errors. The distribution of errors (and differences between errors of the two simulations) is very even. Overall, we found that while the median error increases by 16.4% for traffic-based schedules (stock: 95.1, traffic 110.7), the mean error decreases by 79.2% (stock: 698.7, traffic 145.5). While it is difficult to decide which of the two is better, the authors believe that a reduction in mean error (fewer large outliers or outliers with smaller error) can still be considered an improvement. We expect that a comparison with 2020 electricity consumption data will yield more favorable results.

The differences in results are not as large as one might expect due to two reasons. First of all, the mismatch in years creates an obvious hurdle, which we have minimized by aligning the data by day of the week. More importantly, the impact of COVID-19 further distorts the results.

COVID-19 impact

In order to gain a better understanding of error, we examine the fluctuation of error over time. As Figure 4 demonstrates, the error fluctuates dramatically during key timeframes on the COVID-19 timeline (Tennessee Office of the Governor (2021)). The first

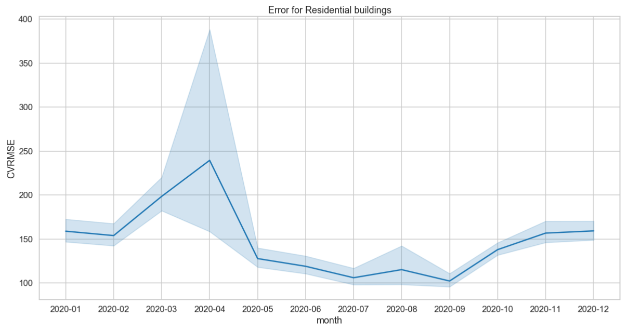

Figure 4: The error for traffic-based schedule in residential buildings peaks during early COVID-19 response in March and April, and returning to similar levels as early in the year. Results for other types of homes and hotels display the same behavior.

COVID-19 case in Tennessee was recorded on March 5, 2020, with the first recorded case in Chattanooga following soon after on March 13, 2020. COVID-response in Tennessee began with recommendations to socially distance (March 13) and close schools (March 17), and led to the prohibitions of gatherings of more than 10 individuals (March 22) and Shelter-in-Place order for Tennessee (April 2). This order was in effect from April 4 until April 30, but businesses were allowed to open on May 1, 2020. This big change in the population's behavior is clearly reflected in the error. The error throughout the remainder of the year remains at or below pre-COVID levels.

Conclusion

We have presented a workflow to create occupancy schedules based on traffic data that was collected through a network of intersection sensors. The schedules were created through a combination of topological, geometric, and geospatial methods, and they were successfully integrated in an EnergyPlus model. The resulting occupancy schedules provide a customized view of building occupancy, which is sensitive to changes at an hourly, monthly, and seasonal scale. With the available traffic data, even shorter timeframes (e.g., 15 minutes) could be explored. As the traffic data is available in real-time, it could even be used to collect real-time occupancy estimates.

Due to data limitations, the comparison in this paper combined 2019 and 2020 data. Under normal circumstances, the impact of this comparison would be less severe, however, the changes in telecommuting behavior through COVID-19 were remarkable. Nevertheless, we were able to produce some promising results which will enable us to further optimize building occupancy schedules. Based on preliminary testing, we expect that a comparison which uses data for an identical time range will yield much better results.

The main objective of this paper was to demonstrate the co-simulation of the occupancy model and EnergyPlus. However, a more detailed analysis of the impact of customized occupancy schedules will be an interesting future study, that is best undertaken for datasets with consistent temporal scopes.

For future work, this methodology can be extended to support different travel modes (such as pedestrians). This will not only introduce additional individuals, but it will also enable a more detailed record of whether drivers visited buildings in the same area, or whether they walked a longer distance to their destination, which will increase applicability of this methodology for areas with is a higher proportion of pedestrians, bicyclists, and public transportation users. Furthermore, point-of-interest data or GPS trace data could serve to further refine the methodology, and serve as a data source for validation.

Finally, it will be interesting to see this workflow applied to a larger area, as more and more sensors are installed in Chattanooga and other smart cities.

Acknowledgments

This work was funded in part by field work proposal CEBT105 under US Department of Energy Building Technology Office Activity Number BT0305000, the Office of Electricity Activity Number TE1103000. This research was also funded in part by the National Geospatial-Intelligence Agency and Approved for Public Release, 21-142.

The authors would like to thank Amir Roth and Madeline Salzman for their support and review of this project. They would also like to thank Mark Adams for his contributions to the AutoBEM software for model generation, and Jibonananda Sanyal for AutoSIM contributions for scalable simulation.

References

Alharin, A., Y. Patel, T.-N. Doan, and M. Sartipi (2020). Data analysis and visualization of traffic in chicago with size and landuse-aware vehicle to buildings assignment. In J. Nichols, B. Verastegui, A. B. Maccabe, O. Hernandez, S. Parete-Koon, and T. Ahearn (Eds), *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*, pp. 518–529. Springer International Publishing.

ASHRAE and Illuminating Engineering Society (IES) (1989). 90.1-1989. *Energy Efficient Design of New Buildings Except Low-Rise Residential Buildings*.

Berres, A., P. Im, K. Kurte, M. Allen, and J. Sanyal (2019). A data-driven urban scale mobility model for estimating building occupancy for energyplus. In *ASHRAE Building Performance Analysis Conference*. Denver, CO (USA).

Berres, A., P. Im, K. Kurte, M. Allen-Dumas, G. Thakur, and J. Sanyal (2019). A mobility-driven approach to modeling building energy. In *Proceed-*

ings of 2019 IEEE International Conference on Big Data, pp. 3887–3895.

De Berg, M., M. Van Kreveld, M. Overmars, and O. Schwarzkopf (1997). Computational geometry. In *Computational geometry*, pp. 1–17. Springer.

Deru, M., K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski, M. Rosenberg, et al. (2011). U.S. department of energy commercial reference building models of the national building stock.

Garrison, E., J. New, and M. Adams (2019). Accuracy of a crude approach to urban multi-scale building energy models compared to 15-min electricity use.

GRIDSMART Technologies, Inc. (2021). *Open Source GridSmart Python Code*. `https://bitbucket.org/GRIDSMART/public-gtipy`.

Gunay, H., W. O'Brien, I. Beausoleil-Morrison, W. Shen, G. Newsham, and I. Macdonald (2017). The effect of zone level occupancy characteristics on adaptive controls. In *Proceedings from Building Simulation 2017*. San Francisco (USA).

Hendron, R. and C. Engebrecht (2010, 10). Building america house simulation protocols (revised).

Leasure, D. R., W. C. Jochem, E. M. Weber, V. Seaman, and A. J. Tatem (2020). National population mapping from sparse survey data: A hierarchical bayesian modeling framework to account for uncertainty. *Proceedings of the National Academy of Sciences 117*(39), 24173–24179.

Lu, X., F. Feng, Z. Pang, T. Yang, and Z. ONeill (2020). Extracting typical occupancy schedules from social media (tossm) and its integration with building energy modeling. In *Building Simulation*, pp. 1–17. Springer.

Microsoft (2018, July). Microsoft building footprints. `https://www.microsoft.com/en-us/maps/building-footprints`.

New, J. R., M. Adams, P. Im, H. Yang, J. Hambrick, W. Copeland, and J. A. Bruce, Lilianand Ingraham (2018). Automatic building energy model creation (AutoBEM) for urban-scale energy modeling and assessment of value propositions for electric utilities.

Qu, A., Y. Wang, Y. Hu, Y. Wang, and H. Baroud (2020). A data-integration analysis on road emissions and traffic patterns. In J. Nichols, B. Verastegui, A. B. Maccabe, O. Hernandez, S. Parete-Koon, and T. Ahearn (Eds), *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*, pp. 503–517. Springer International Publishing.

Ren, G., M. Sunikka-Blank, and X. Zhang (2017). The influence of variation in occupancy pattern on domestic energy simulation prediction: A case study in shanghai. In *Proceedings from Building Simulation 2017*. San Francisco (USA).

Santosh, P., T. Tran, E. Youngson, and J. Bull (2016). eppy: A scripting language for e+, energyplus. `https://pythonhosted.org/eppy`.

Silva, V., D. Amo-Oduro, A. Calderon, C. Costa, J. Dabbeek, V. Despotaki, L. Martins, M. Pagani, A. Rao, M. Simionato, et al. (2020). Development of a global seismic risk model. *Earthquake Spectra 36*(1), 372–394.

Sparks, K., G. Thakur, A. Pasarkar, and M. Urban (2020). A global analysis of cities geosocial temporal signatures for points of interest hours of operation. *International Journal of Geographical Information Science 34*(4), 759–776.

Stewart, R. N., M. L. Urban, S. E. Duchscherer, J. Kaufman, A. M. Morton, G. Thakur, J. Piburn, and J. Moehl (2016, 1). A bayesian machine learning model for estimating building occupancy from open source data. *Natural Hazards 81*(3).

Tennessee Office of the Governor (2021). Covid-19 timeline. `https://www.tn.gov/governor/covid-19/covid19timeline.html`.

United States Geological Survey (2020). PAGER. `https://earthquake.usgs.gov/data/pager/"`.

U.S. Department of Energy (2019). Commercial prototype building models. `https://www.energycodes.gov/development/commercial/prototype_models`.

U.S. Department of Energy (2020a). Energyplus.

U.S. Department of Energy (2020b). Openstudio. `https://www.openstudio.net/`.

U.S. Department of Energy (2018). *Fact of the Week #1040, July 30, 2018: Average Vehicle Occupancy Remains Unchanged From 2009 to 2017*. `https://www.energy.gov/eere/vehicles/articles/fotw-1040-july-30-2018-average-vehicle\-occupancy-remains-unchanged-2009-2017`.

Wang, J., Y. Ye, W. Zuo, J. New, and A. Rose (2021, Feb). City-scale building occupancy prediction using geographic information system data.

Wang, J., W. Zuo, S. Huang, and D. Vrabie (2020, 09). Data-driven prediction of occupant presence and lighting power-a case study for small commercial buildings.