Prototype-Wise Sensitivity Analysis of Urban Building Energy Simulation Surrogate Modeling Accuracy

Xiyu Pan[1]; Neda Mohammadi, Ph.D.[2]; John E. Taylor, Ph.D.[3]; Yujie Xu, Ph.D.[4]; and Tianzhen Hong, Ph.D.[5]

[1]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA. Email: xyp@gatech.edu
[2]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA. Email: nedam@gatech.edu
[3]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA. Email: jet@gatech.edu
[4]Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA. Email: thong@lbl.gov
[5]Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA. Email: yujiex@lbl.gov

ABSTRACT

Urban Building Energy Modeling (UBEM) is an important reference for urban energy-related policymaking. Because of the significant impact of urban microclimates on the energy simulation, UBEM requires simulations of many microclimate-prototype pairs. Surrogate modeling is commonly used to reduce the cost of simulation computations. In UBEM surrogate modeling, it is important to determine the percentage of microclimates related to a prototype used for generating surrogate model training data. This study analyzes the prototype-wise variations and sensitivities of surrogate model estimation accuracy to the microclimate sampling ratios. The results of the study can help determine the number of simulations used for generating surrogate modeling data, avoid redundant simulations, and reduce the computational cost for UBEM surrogate modeling and its time.

INTRODUCTION

Urban Building Energy Modeling (UBEM) includes urban-scale computational modeling and simulation of various metrics such as cooling loads, heating loads, thermal emissions, thermal comfort, etc., and is essential for actions such as energy performance evaluation, energy policymaking, district planning, and retrofitting. UBEM is increasingly gaining attention, and other trends such as smart city digital twins are also contributing to its popularity (Francisco et al. 2020). One of the common approaches to UBEM is to model the energy consumption and thermal performances of a limited number of building prototypes using a simulation model such as EnergyPlus and then scale up the simulation results at the prototype level to the entire urban region based on the metadata of the urban buildings (e.g., the floor area and spatial distribution of each building prototype) (Ang et al. 2020). However, the computational time of the UBEM approach can be long, especially considering the impact of urban microclimates on UBEM.

Urban microclimates can affect the results of building energy simulations by changing the heating and cooling demand, the amount of heat exchange between the building and the environment, and the interactions among the buildings (Pisello et al. 2012). On the other hand, microclimates can be highly heterogeneous in urban areas because of the spatial distribution of the natural and built environments and the climate variation over the years. To account for this heterogeneity, UBEM models use gridded weather data with high spatial resolution. However, the simulation runs scales quadratic with the increase in the spatial resolution of the weather grids. For instance, Los Angeles County's UBEM needs to model nearly 9,000 microclimates if the 1×1 kilometer microclimate grid cell size, suggested by previous studies, is adopted (Sezer et al. 2023), and the number of prototype-microclimate pairs will grow to an unmanageable number, especially for multiple building prototypes and multi-year modeling.

Surrogate modeling is one of the important means to reduce the computational cost of UBEM. The basic idea of surrogate modeling is to train a data-driven machine learning model using the inputs of the simulation model as features and the outputs of the simulation model as targets. Once the machine learning model can estimate the simulation outputs with satisfactory accuracy, given the simulation inputs, it can replace the simulation model for various tasks, such as model calibration (Herbinger et al. 2023), uncertainty analysis (Fu et al. 2020), sensitivity analysis (Tian et al. 2015), and performance optimization (Yigit 2021).

In UBEM surrogate modeling, a key parameter is the sampling ratio of the microclimates used to generate the data for training the surrogate models. While a higher sampling ratio could generate more training data and potentially increase the accuracy of the surrogate modeling, it will increase the computational cost. Different building prototypes may require different sampling ratios to balance accuracy and computational cost. There is a lack of research on the effect of sampling ratios on the surrogate modeling accuracy. Therefore, this study measures the sensitivity of prototype-wise surrogate modeling accuracy to the sampling ratios. It was found that accuracy rises as the sampling rate rises, but the magnitude of the positive effect declines with the growing sampling rate. Increasing the sampling rate from 15% to 25% is of minimal help for the surrogate modeling for most prototypes. The findings of this study are expected to help determining microclimates sampling ratios for UBEM surrogate modeling and better making a balance between computation time and accuracy.

METHOD

Surrogate modeling based UBEM. Figure 1 illustrates the framework of UBEM based on surrogate modeling. First, building prototypes and microclimates in the city are predefined. The building prototypes are defined by Input Data File (IDF) files, and the microclimate is defined by a time series of hourly meteorological features throughout the year, which is stored in EnergyPlus Weather File (EPW) files. In the second step, microclimates are matched to each prototype. In the second step, each building in the city is matched to a building prototype and a microclimate. Each prototype is therefore matched to multiple microclimates. The microclimates for each prototype are randomly selected according to a sampling ratio α, and the EPW files of the selected microclimates together with the IDF file of the prototype are imported into EnergyPlus for simulation. The simulation model outputs the hourly electricity consumption for the whole year. The third step is to train the surrogate models, one for each prototype building. A bidirectional Long Short-Term Memory model was utilized in this study, which uses a time window to take meteorological features within a fixed length period as the input and estimates the electricity consumption at the next hour. After the model is trained, sequences of meteorological features of new microclimates are used as inputs and the surrogate model estimates the hourly electricity consumption for the whole year on a rolling basis. The final step of the method is to scale up the annual hourly electricity consumption estimates for all prototype-microclimate pairs obtained in Step 3 to the whole city through spatial join calculations.

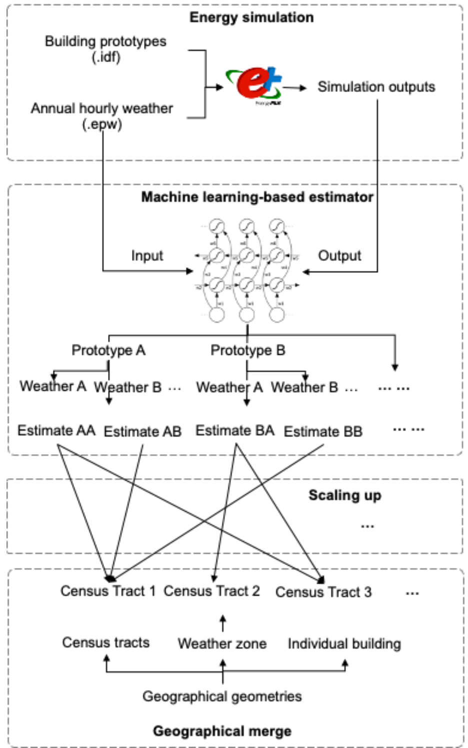

Figure 1. The framework of the surrogate modeling based UBEM.

Sensitivity analysis of accuracy to sampling ratios. In the second step of the surrogate modeling method, the microclimate sampling ratio α was set to be the same for all building prototypes. However, the difficulty of estimating electricity consumption varies across building types, given the selected weather and energy inputs. For example, the electricity consumption of factory buildings depends heavily on production schedules rather than weather variations, but residential buildings have a more complex weather-electricity consumption relationship and may require more training data to train the surrogate model. Therefore, this study compares the prediction accuracy of different sampling ratios for different building prototype surrogate models. In Step 3, the building prototype-wise estimation accuracy is measured and visualized to demonstrate the performance and sensitivity of the surrogate modeling accuracy with different sampling ratios. Normalized Mean Absolute Error (nMAE) is used to measure the accuracy, which is defined as $\frac{MAE}{\frac{1}{N}\sum_{n=1}^{N}|y_n|}$, where MAE is defined as $\frac{\sum_{n=1}^{N}|y_n-\hat{y}_n|}{N}$, $\hat{y}_n$ is the estimation of the n^{th} sample, and y_n is the ground truth of the n^{th} sample.

RESULTS

Data sources. LA County is divided by a grid into 62 12*12 km cells, and each grid has a microclimate, which is represented by 8 meteorological features, such as relative humidity, temperature at 2 meters, etc. LA County has 54 building archetypes, which are formed by building types (e.g., multi-family houses, single-family houses, hotels) and three vintages (i.e., pre-1980, 2004, and 2013). Refer to the work by Xu et. al. for the details of the datasets (Xu et al. 2022) and the datasets and codes of this study are available at the GitHub repository github.com/pppxiyu/SurUBEM.

Sensitivity of estimation accuracy to sampling ratios. Figure 2 demonstrates the accuracy of the UBEM surrogate modeling for different prototypes with different microclimate sampling ratios. For all prototypes, a higher microclimate sampling ratio generates more training data, and the surrogate model's estimation error for electricity consumption is usually lower. It is also noted that the marginal accuracy improvement from increasing the sampling ratio is rapidly decreasing for many prototypes. For example, there is no significant difference in nMAE with 15% and 25% training data for multifamily houses and small offices. In addition to this, some prototypes, such as midrise apartments, small offices, manufacturing buildings, and single- and multi-family houses are sensitive to the microclimate sampling ratio, while colleges, warehouses, hotels, hospitals, etc. are less sensitive to sampling ratio.

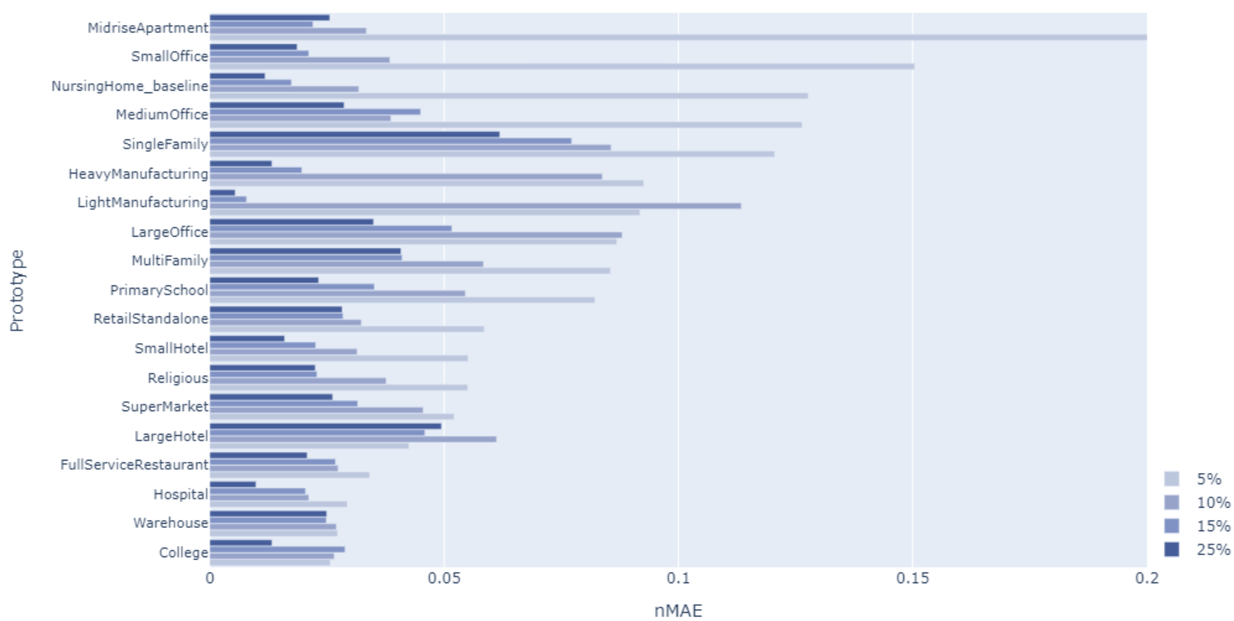

Figure 2. nMAE of hourly electricity estimation for prototypes with 5% to 25% microclimate sampling rate α. Midrise Apartment's nMAE with 5% sampling rate is 0.82, which is not fully shown in the figure.

CONCLUSION

This study analyzes the prototype-wise variations and sensitivities of UBEM surrogate model estimation accuracies to microclimate sampling ratios. It is found that although increasing the sampling ratio within a certain range usually improves the accuracy significantly, the sensitivity of different prototypes to this manipulation is different. The results of this study contribute to the determination of prototype-wise microclimate sampling ratios for UBEM surrogate modeling, which avoids unnecessary simulation computations and further reduces the computational cost of UBEM while ensuring overall accuracy. In addition, appropriate sampling ratios could also potentially reduce the data collection requirements for urban microclimates, which can also be time-consuming and computationally intensive. Despite the contributions, the findings of this study are limited to the prototypes being studied. Although the studied prototypes cover most of the common building types, there are still some other building types or the same building types with different characteristics that are not considered. Future research could dive deeper into the reasons for different appropriate sampling ratios and discover the relationship between specific data distributions and the appropriate sampling ratios to provide more generalized conclusions.

ACKNOWLEDGEMENTS

This paper is supported by the National Science Foundation (NSF) under Grant No. 1837021. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF. This study is based on the first author's work during the internship in Lawrence Berkeley National Laboratory.

REFERENCES

Ang, Y. Q., Z. M. Berzolla, and C. F. Reinhart. (2020). "From concept to application: A review of use cases in urban building energy modeling." *Appl. Energy*, 279: 115738.

Francisco, A., N. Mohammadi, and J. E. Taylor. (2020). "Smart City Digital Twin-Enabled Energy Management: Toward Real-Time Urban Building Energy Benchmarking." *J. Manag. Eng.*, 36 (2).

Fu, X., W. Tian, Y. Sun, C. Zhu, and B. Yin. (2020). "Uncertainty Analysis of Urban Building Energy Based on Two-Dimensional Monte Carlo Method." *Proc. 11th Int. Symp. Heating, Vent. Air Cond. (ISHVAC 2019)*, 1315–1323.

Herbinger, F., C. Vandenhof, and M. Kummert. (2023). "Building energy model calibration using a surrogate neural network." *Energy Build.*, 289 (April).

Pisello, A. L., J. E. Taylor, X. Xu, and F. Cotana. (2012). "Inter-building effect: Simulating the impact of a network of buildings on the accuracy of building energy performance predictions." *Build. Environ.*, 58: 37–45.

Sezer, N., H. Yoonus, D. Zhan, L. L. Wang, I. G. Hassan, and M. A. Rahman. (2023). "Urban microclimate and building energy models: A review of the latest progress in coupling strategies." *Renew. Sustain. Energy Rev.*, 184: 113577.

Tian, W., R. Choudhary, G. Augenbroe, and S. H. Lee. (2015). "Importance analysis and meta-model construction with correlated variables in evaluation of thermal performance of campus buildings." *Build. Environ.*, 92: 61–74. Elsevier Ltd.

Xu, Y., P. Vahmani, A. Jones, and T. Hong. (2022). "A multi-scale time-series dataset of anthropogenic heat from buildings in Los Angeles County."

Yigit, S. (2021). "A machine-learning-based method for thermal design optimization of residential buildings in highly urbanized areas of Turkey." *J. Build. Eng.*, 38 (January): 102225. Elsevier Ltd.