Socioeconomic determinants of public health and residential building energy use in Chicago

Narjes Abbasabadi
Assistant Professor, University of Texas at Arlington

Mehdi Ashayeri
Assistant Professor, Southern Illinois University at Carbondale

Abstract
Limited studies on energy consumption in cities have been done to explore the effects of socioeconomic determinants of public health on urban energy use. This article examines the associations between these factors, including characteristics of occupied housing unit, household income, employment, education, dependency, poverty level, and crowded housing with urban building energy use. The available empirical data from the City of Chicago on socioeconomic indicators of public health 2008-2012 and Chicago energy benchmarking 2016 and Chicago energy Usage 2010 were used. This research applied a machine learning approach based on the Artificial Neural Network (ANN) algorithm to predict the energy use intensity across Chicago, and several explanatory methods were extended to the model to help facilitate interpreting the results. And a cross-validation technique was employed to confirm the results. Findings suggest that all these socioeconomic determinants were strongly associated with energy use of residential buildings. Household income as the highest influential variable among them has a positive relationship with residential operational energy use. Further, urban building energy use was associated with urban form and building characteristics as well as various dimensions of socioeconomic determinants. Endeavors for reducing energy consumption in cities need to consider different dimensions of urban spatial patterns and socioeconomic status of public health.

Keywords: Urban energy modeling, residential building energy use, public health, socioeconomic indicators

Introduction
Many cities across the world have started setting energy reduction goals and moving towards more sustainable and low carbon cities. In doing so, understanding energy consumption patterns and the effects of different factors on energy use is needed to address the energy efficiency targets and tackle the consequences of an increase in energy demand due to unprecedented population growth and uneven urbanization. The previous studies on energy consumption in cities tend to examine how urban attributes influence energy use with a focus on urban form, morphology, and sprawl characteristic [1], building characteristics [2]. However, there is limited research that studies occupancy characteristics and human behavior [3] and households' socioeconomic status of public health on energy consumption in cities [4] . This happens mainly due to the complex nature of human-related factors in urban systems and the fact that multifaceted models need to be developed that consider many factors [5]. As data is becoming more available and advances in artificial intelligence and machine learning techniques provide opportunities for illuminating the complex associations in energy consumption patterns [6].

To address the gap, a machine learning approach with big data analytics was used to develop a predictive urban building energy model and examine the association between urban socioeconomic determinants of public health and building energy use. A large variety of public data representing building stock, urban spatial patterns, socioeconomic status, and building energy use were used. The occupancy in energy context is often quantified through measuring various socioeconomic factors such as employment, education, household size, income, ownership, along with other socio-demographic and behavior factors [7]. In this article, the socioeconomic and occupancy variables include housing crowd, poverty level, unemployment, education, dependency, per capita income, household size, percentage of occupied units, and total occupants. Herein we provide a quantitative analysis of the complex interplay between the socioeconomic and occupancy patterns on one of the main components of urban energy use, residential building operational energy consumption. Including socioeconomic and occupancy indicators can help to model urban building energy use more in-depth, which provides a more comprehensive view of households' energy demand patterns. Chicago, IL, which has committed to long-term emission reduction goals with a focus on energy strategies [8], was chosen as a case study to test the framework.

Methodology

In the present article, energy use is defined as residential building operational energy use at the city scale within Chicago neighborhoods. The model incorporates the key urban attributes in the model include building characteristics, urban spatial patterns, occupancy characteristics, socioeconomic indicators. The model uses several available datasets including Chicago building footprints (CBF) dataset [9], Property tax data from the Assessor's Office [10], Urban Sprawl data for the Unites States [11], Socioeconomic indicators dataset [12], Chicago Energy Benchmarking dataset [13] (2,717 buildings greater than 50,000 ft2), and Chicago Energy Usage dataset [14] (65,378 buildings of all sizes) to predict building energy use for out of sample for all buildings in Chicago where their energy information is not available. Then, the model is used to explain the association between variables. It explains the relative contribution of each variable and quantifies how key urban attributes affect urban energy use with a focus on socioeconomic indicators controlling for other urban attributes and building characteristics.

The Artificial Neural Network (ANN) [15] algorithm is believed as a versatile machine learning algorithm [16] among other algorithms were tested to enable capturing the complex and non-linear relationships between socioeconomic factors of public health and energy dynamics across neighborhoods in the city. The ANN model, as a robust technique, enables capturing non-linear patterns of the complex data. Structuring the topology of the ANN networks and training highly efficient model; therefore, is a challenging task and is usually based on the rule of thumb approach. In the present article, we developed an automation script using nested for-loop (loop inside a loop) approach in R programming language to capture the best model in terms of performance. We used the cross-validation method, the 5-fold, which according to the previous studies is known as an effective method [17].

Furthermore, ANN is known as a "black-box" algorithm that makes the interpretability of the model challenging [18]. There are methods that can be extended to ANN's trained model to increase its explanatory capabilities and allow for quantifying the relative contributions of each variable [19]. The Partial Dependence (PaD) plots was employed to implement such computational tasks. PaD calculates the partial derivatives of the dependent variable based on the independent variables, while keeping other peer variables at their constant values (e.g., mean) [20]. PaD can be plotted as profile (y~x) or heat-map ($y \sim x_1, x_2$) on trained models for all ML algorithms. Lek's Profile [21] as a specific type of profile approach was employed in this research which enables exploring the explanatory capabilities of ANN algorithms. In Lek's Profile, originally, each explanatory variable is investigated in which all the peer explanatory variables are kept at their constant levels (e.g., minimum, median, quartiles, and maximum). Then, each input variable is divided into equal intervals, sequenced between its minimum to maximum values with respect to the assigned range of observations. This research employed Lek's Profile and heat-map methods for exploring relationship between explanatory and response variables.

Results

Table 1 illustrates the results of the ANN model, a MAPE of 4.1 indicates the predictive power for the urban energy use modeling for residential buildings in Chicago. The results show an R^2 value of 0.41, which suggests that the model explains 41% of the variance of building EUI in the model. Thus, considering the urban socio-spatial context is essential for understanding energy consumption patterns in cities in order to address the energy reduction goals. It should be noted that many other influential factors (e.g., occupant behavior factors) would be needed to be incorporated into the model in order to provide a more comprehensive explanation for variations of energy use in cities because relying on limited factors fails to capture all the variance.

The results suggest that the occupancy and socioeconomic variables are strong predictors for urban energy use modeling, which their relative contribution to energy use is varied related to the distance to CBD, representing the building location. Figure 1 shows heat-map of the relationship between socioeconomic factors and distance to CBD in the urban residential building energy context. Buildings located between 5 through 10 miles away from the business center occupied by households with income ranging from $50,000 to $75,000 account for the highest impacts on building site EUI. Further, the unemployment factor for the buildings located between 7.5 to 10 miles away from CBD is associated with the highest impact on building EUI. Regarding the education factor, the building site EUI spike was found to be at the range of 9 through 12 miles based on the distance to CBD. Moreover, the minimum building site EUI was found to be associated with the highest percentage of low-educated building occupants. The highest building site EUI regarding the household dependency factor and distance to CBD was found to be at 7 through 10 miles and 30 through 40 percent of the dependency variable. The maximum impact of the poverty variable was captured for between 5 to 10 miles of distance to CBD, considering building site EUI. And the relationship between the crowded housing variable and distance to the CBD based on building site EUI shows that the higher percentage of crowded housing is associated with higher building EUI, particularly those buildings located between 3 to 10 miles away from the CBD.

Figure 2, the Lek's Profile result, shows how household per capita income (PHI) impacts building EUIs through considering other variables constant on their constant values. There are three ranges on the profile for building site EUI, which are significant including the per capita income ranged from $0 through $64,000, $64,000 through $75,000 (positive), and over $75,000 in which the most variation occurs within the second range. Hence, households with $75,000 per capita income represent the highest contributors on building EUI. The overall trend shows that high-income households tend to consume energy more than average income households. This result confirms the findings from previous studies on the impacts of income on building [22] energy use. Percentage of households below poverty (PHBP) variable is another factor in the model. Poverty variable has an overall negative trend with building EUI, meaning that with increasing the percentage of households' poverty level, its impacts on the building EUI decreases.

The results suggest that the impacts of unemployment on building EUI is a polynomial curve (Figure 2). The impacts is positive, and after a certain level, it becomes negative. This pattern can be interpreted that the household unemployment impacts building occupancy level and consequently it impacts the presence of occupants in building [23]. Thus, this increases the presence at home, and leads to more energy demand for heating, cooling, lighting, and application of appliances. The literature suggests that the impact of the employment factor on building energy demand is a positive relationship [24]. Here the captured non-linear pattern suggests a positive relationship at a certain point, and after that, it becomes negative. It should be noted that this result is specific to the case of Chicago. In [25], six common occupancy scenarios were modeled for building energy use including 1) occupants with full-time jobs, 2) retired occupants or families with young children, 3) occupants who spend their afternoon outside, 4) occupants with a part-time job in the morning, 5) occupants with a part-time job in the afternoon, and 6) occupants with a child that goes to school. The findings of this study suggest that the occupancy pattern has a significant effect on the annual building energy use and occupancy scenarios with higher unoccupied periods show lower energy demand.

The results indicate that the total occupants (TOC) have a positive relationship with building site EUI (Figure 2). This result can be interpreted as increasing TOC may increase presence at home, and accordingly leads to higher energy consumption. Crowded housing (PHC) variable is another occupancy related variable in the model that shows the percentage of occupying a room with more than one person. The results show a positive linear trend with the building EUI. Therefore, like the former socioeconomic indicators, the housing PHC is an essential variable in developing models for urban building energy use. The result also suggests that the household dependency variable (P1864) has the most robust positive trend with the building EUI (between 40 through 43). This variable has a non-linear relationship with the building site EUI. Thus, the household dependency is a strong predictor to determine urban building energy use patterns.

The results obtained herein confirm the importance of occupancy and socioeconomic variables in previous studies such as [26] which examines occupancy patterns based on five factors

Table 1. Performance evaluation of the selected integrated building and transpiration energy use model.

Energy model	MSE	RMSE	MAE	MAPE (%)	R^2
Building	0.052	0.3	0.166	4.10	0.41

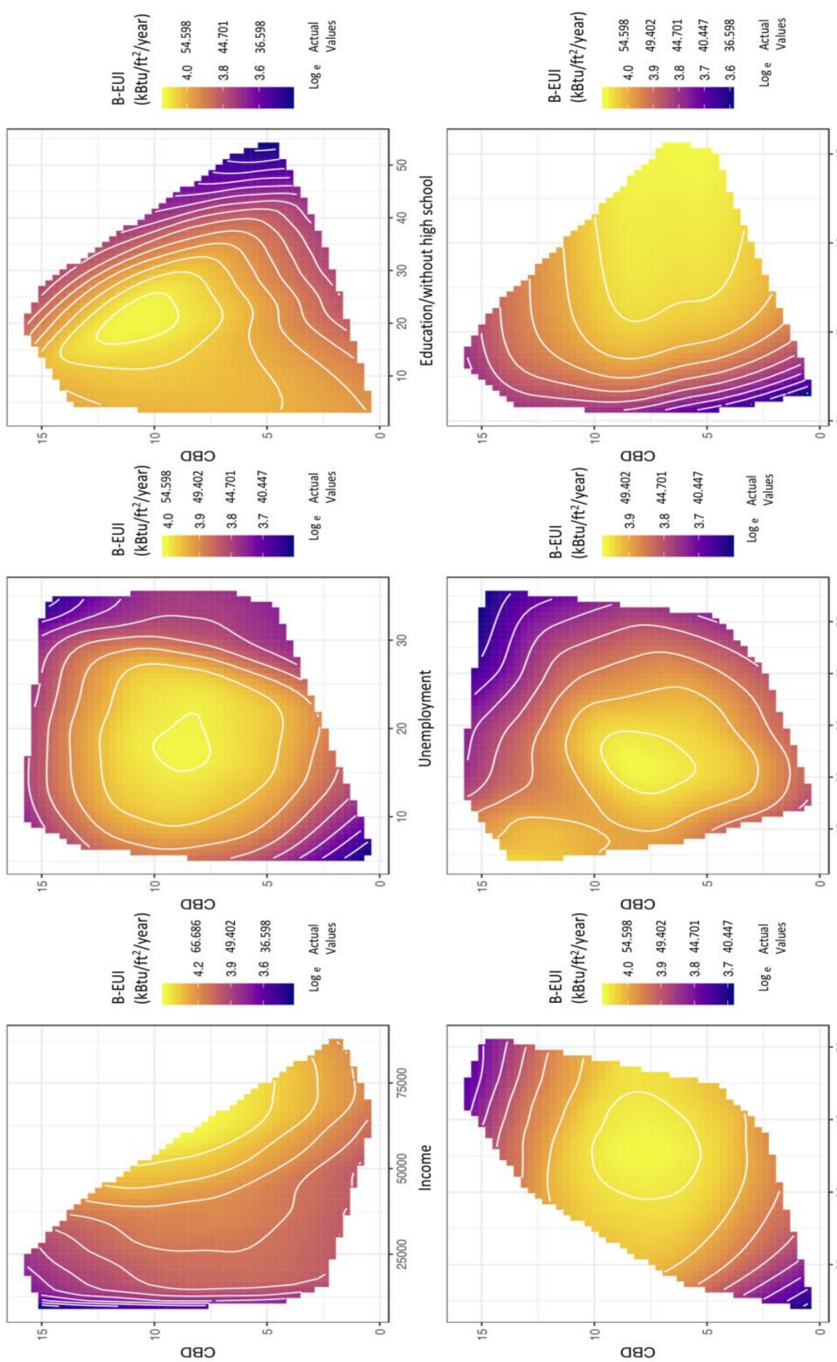

Figure 1. Heat-map plot of 6 socioeconomic factors against distance to CBD based on building EUI.

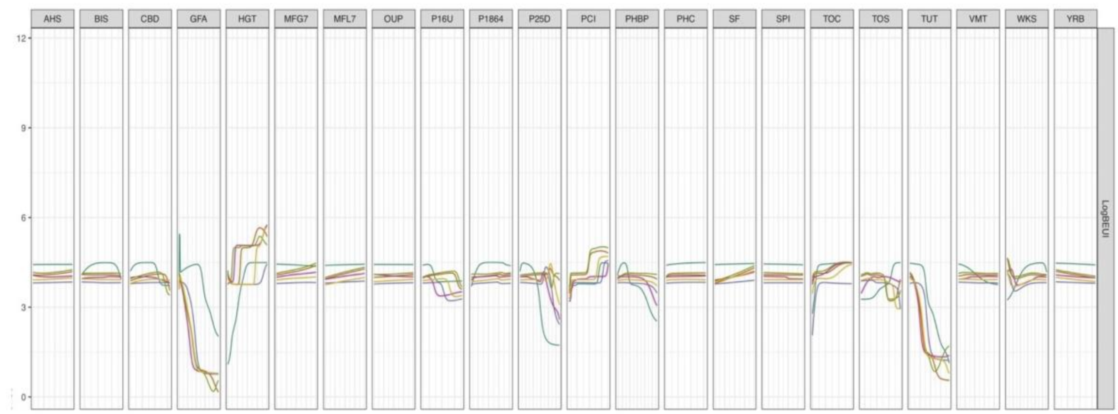

Figure 2. Lek's Profile of all variables with six number of clusters against Log building site EUI.

including 1) household size, 2) employment, 3) education, 4) age, and 5) income parameters according to three occupancy classes of high, medium, and low. Their results show that low occupancy class has the lowest average energy use. In another previous study [27], the occupancy factor was incorporated in the energy use model of a housing tower in London for retrofit purposes, and their results suggest the importance of occupants' socio-demographic characteristics on energy use. The reasons for how occupancy and socioeconomic factors can impact energy use were examined in previous studies (e.g., [28]), which explains the impacts through determining the presence and number of occupants that impact the building energy demand. The occupancy characteristics also impact the use of building systems (such as lighting and appliances), and heat gain, because of occupants' metabolisms and activities, and HVAC loads due to interactions with building components (such as adjusting openings, windows, and doors) [29]. However, a number previous studies, for example, [30], argue that the relative contribution of occupancy, socio-demographics, occupants' behavior, and attribute are significantly lower than building characteristics. The findings of these studies suggest that building factors (e.g., size, household size, ownership, etc.) explain most of variability in residential energy use, and socio-demographic factors show relatively slight explanatory power and are not significant predictors.

Conclusion

This article provides insight on the impacts of socioeconomic indicators on residential building energy use at urban scale. The results from PaD and Lek's Profile methods applied in our research suggest that occupancy and socioeconomic patterns are among the important determinants of urban building energy use. The results show that urban energy use of residential buildings is affected considerably by urban spatial patterns and the socioeconomic factors. The novelty of this research is the inclusion of these factors into building energy use modeling, and applying the artificial intelligence-based approach, ANN method extended with an explanatory model by means of PaD and Lek's Profile. This enables developing an effective predictive model and illuminating complex patterns to explain the contribution of each variable to the model.

Endnotes

1. [1] Reid Ewing and Fang Rong, "The Impact of Urban Form on U.S. Residential Energy Use," *Housing Policy Debate* 19, no. 1 (2008): 37–41, https://doi.org/10.1080/10511482.2008.9521624; Peter W. G. Newman and Jeffrey R. Kenworthy, "Gasoline Consumption and Cities," *Journal of the American Planning Association* 55, no. 1 (March 1989): 24–37, https://doi.org/10.1080/01944368908975398; Carlo Ratti, Nick Baker, and Koen Steemers, "Energy Consumption and Urban Texture," *Energy and Buildings* 37, no. 7 (July 2005): 762–76, https://doi.org/10.1016/J.ENBUILD.2004.10.010.

2. [2] Jonathan Norman et al., "Comparing High and Low Residential Density: Life-Cycle Analysis of Energy Use and Greenhouse Gas Emissions," *Journal of Urban Planning and Development*, no. March 2006 (2006): 10–21, https://doi.org/10.1061/(ASCE)0733-9488(2006)132; Koen Steemers, "Energy and the City: Density, Buildings and Transport," *Energy and Buildings* 35, no. 1 (January 2003): 3–14, https://doi.org/10.1016/S0378-7788(02)00075-0; Giuliano Dall'O', Annalisa Galante, and Marco Torri, "A Methodology for the Energy Performance Classification of Residential Building Stock on an Urban Scale," *Energy and Buildings* 48 (May 1, 2012): 211–19, https://doi.org/10.1016/j.enbuild.2012.01.034.

3. [3] Gabriel Happle, Jimeno A. Fonseca, and Arno Schlueter, "A Review on Occupant Behavior in Urban Building Energy Models," *Energy and Buildings* 174 (September 2018): 276–92, https://doi.org/10.1016/J.ENBUILD.2018.06.030; Mengda Jia, Ravi S. Srinivasan, and Adeeba A. Raheem, "From Occupancy to Occupant Behavior: An Analytical Survey of Data Acquisition Technologies, Modeling Methodologies and Simulation Coupling Mechanisms for Building Energy Efficiency," *Renewable and Sustainable Energy Reviews* 68 (February 1, 2017): 525–40, https://doi.org/10.1016/j.rser.2016.10.011; Zhun Yu et al., "A Systematic Procedure to Study the Influence of Occupant Behavior on Building Energy Consumption," *Energy and Buildings* 43, no. 6 (June 1, 2011): 1409–17, https://doi.org/10.1016/j.enbuild.2011.02.002.

4. [4] Dominik Wiedenhofer, Manfred Lenzen, and Julia K. Steinberger, "Energy Requirements of Consumption: Urban Form, Climatic and Socio-Economic Factors, Rebounds and Their Policy Implications," *Energy Policy* 63 (December 1, 2013): 696–707, https://doi.org/10.1016/j.enpol.2013.07.035; John R. Jensen and Dave C. Cowen, "Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes," in *The Map Reader*, ed. Martin Dodge, Rob Kitchin, and Chris Perkins (Chichester, UK: John Wiley & Sons, Ltd, 2011), 153–63, https://doi.org/10.1002/9780470979587.ch22; Geun Young Yun and Koen Steemers, "Behavioural, Physical and Socio-Economic Factors in Household Cooling Energy Consumption," *Applied Energy* 88, no. 6 (June 1, 2011): 2191–2200, https://doi.org/10.1016/j.apenergy.2011.01.010; Milena Büchs and Sylke V. Schnepf, "Who Emits Most? Associations between Socio-Economic Factors and UK Households' Home Energy, Transport, Indirect and Total CO2 Emissions," *Ecological Economics* 90 (June 1, 2013): 114–23, https://doi.org/10.1016/j.ecolecon.2013.03.007.

5. [5] Narjes Abbasabadi and Mehdi Ashayeri, "Urban Energy Use Modeling Methods and Tools; A Review and an Outlook for Future Tools," *Building and Environment*, July 6, 2019, 106270, https://doi.org/10.1016/j.buildenv.2019.106270.

6. [6] Narjes Abbasabadi et al., "An Integrated Data-Driven Framework for Urban Energy Use Modeling (UEUM)," *Applied Energy* 253 (November 2019): 113550, https://doi.org/10.1016/j.apenergy.2019.113550; Abbasabadi and Ashayeri, "Urban Energy Use Modeling Methods and Tools; A Review and an Outlook for Future Tools."

7. [7] Gesche Huebner et al., "Understanding Electricity Consumption: A Comparative Contribution of Building Factors, Socio-Demographics, Appliances, Behaviours and Attitudes," *Applied Energy* 177 (September 1, 2016): 692–702, https://doi.org/10.1016/j.apenergy.2016.04.075; Gesche M. Huebner et al., "Explaining Domestic Energy Consumption – The Comparative Contribution of Building Factors, Socio-Demographics, Behaviours and Attitudes," *Applied Energy* 159 (December 1, 2015): 589–600, https://doi.org/10.1016/j.apenergy.2015.09.028.

8. [8] "City of Chicago Climate Action Plan," City of Chicago Climate Action Plan, 2019, http://www.chicagoclimateaction.org/.

9. [9] "Building Footprints (Current)," Chicago Data Portal, 2015, https://data.cityofchicago.org/Buildings/Building-Footprints-current-/hz9b-7nh8?category=Buildings&view_name=Building-Footprints-current-.

10. [10] "Cook County Assessor Data," Cook County Government, Open Data, 2019, https://datacatalog.cookcountyil.gov/.

11. [11] "Updated Urban Sprawl Data for the United States," National Cancer Institute, Geographic Information Systems & Science for Cancer Control, 2010, https://gis.cancer.gov/tools/urban-sprawl/.

12. [12] "Census Data-Selected Socioeconomic Indicators in Chicago, 2008 – 2012," Chicago Data Portal, 2019, https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c- c2s2.

13. [13] City of Chicago, "Chicago Energy Benchmarking-2016 Data Reported in 2017, City of Chicago, Data Portal," City of Chicago, Environment & Sustainable Development, 2016, https://data.cityofchicago.org/Environment-Sustainable-Development/Chicago-Energy-Benchmarking-2016-Data-Reported-in-/fpwt-snya.

14. [14] City of Chicago, "Energy Usage 2010, City of Chicago, Data Portal," City of Chicago, Environment & Sustainable Development, 2010, https://data.cityofchicago.org/Environment-Sustainable-Development/Energy- Usage-2010/8yq3-m6wp.

15. [15] F Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.," *Psychological Review* 65, no. 6 (January 1958): 386–408, https://doi.org/10.1037/h0042519.

16. [16] Hengfang Deng, David Fannon, and Matthew J. Eckelman, "Predictive Modeling for US Commercial Building Energy Use: A Comparison of Existing Statistical and Machine Learning Algorithms Using CBECS Microdata," *Energy and Buildings* 163 (2018): 34–43, https://doi.org/10.1016/j.enbuild.2017.12.031; Abbasabadi and Ashayeri, "Urban Energy Use Modeling Methods and Tools; A Review and an Outlook for Future Tools."

17. [17] Sara Torabi Moghadam et al., "A GIS-Statistical Approach for Assessing Built Environment Energy Use at Urban Scale," *Sustainable Cities and Society* 37 (February 2018): 70–84, https://doi.org/10.1016/J.SCS.2017.10.002; Kadir Amasyali and Nora M. El-Gohary, "A Review of Data-Driven Building Energy Consumption Prediction Studies," *Renewable and Sustainable Energy Reviews* 81 (2018): 1192–1205, https://doi.org/10.1016/j.rser.2017.04.095.

18. [18] Julian D Olden and Donald A Jackson, "Illuminating the 'Black Box': A Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks," *Ecological Modelling* 154, no. 1 (August 15, 2002): 135–50, https://doi.org/10.1016/S0304-3800(02)00064-9.

19. [19] Muriel Gevrey, Ioannis Dimopoulos, and Sovan Lek, "Review and Comparison of Methods to Study the Contribution of Variables in Artificial Neural Network Models," *Ecological Modelling* 160, no. 3 (February 2003): 249–64, https://doi.org/10.1016/S0304-3800(02)00257-0.

20. [20] Gevrey, Dimopoulos, and Lek.

21. [21] Sovan Lek et al., "Application of Neural Networks to Modelling Nonlinear Relationships in Ecology," *Ecological Modelling* 90, no. 1 (September 1996): 39–52, https://doi.org/10.1016/0304-3800(95)00142-5.

22. [22] Javad Jazaeri, Robert L. Gordon, and Tansu Alpcan, "Influence of Building Envelopes, Climates, and Occupancy Patterns on Residential HVAC Demand,"

Journal of Building Engineering 22 (March 1, 2019): 33–47, https://doi.org/10.1016/j.jobe.2018.11.011.

23. [23] Zheng Yang and Burcin Becerik-Gerber, "How Does Building Occupancy Influence Energy Efficiency of HVAC Systems?," *Energy Procedia*, CUE 2015 - Applied Energy Symposium and Summit 2015: Low carbon cities and urban energy systems, 88 (June 1, 2016): 775–80, https://doi.org/10.1016/j.egypro.2016.06.111.

24. [24] Jazaeri, Gordon, and Alpcan, "Influence of Building Envelopes, Climates, and Occupancy Patterns on Residential HVAC Demand."

25. [25] Jazaeri, Gordon, and Alpcan.

26. [26] Cristina Carpino et al., "Application of Survey on Energy Consumption and Occupancy in Residential Buildings. An Experience in Southern Italy," *Energy Procedia*, ATI 2018 - 73rd Conference of the Italian Thermal Machines Engineering Association, 148 (August 1, 2018): 1082–89, https://doi.org/10.1016/j.egypro.2018.08.051.

27. [27] Sahar Zahiri and Heba Elsharkawy, "Towards Energy-Efficient Retrofit of Council Housing in London: Assessing the Impact of Occupancy and Energy-Use Patterns on Building Performance," *Energy and Buildings* 174 (September 1, 2018): 672–81, https://doi.org/10.1016/j.enbuild.2018.07.010.

28. [28] Yang and Becerik-Gerber, "How Does Building Occupancy Influence Energy Efficiency of HVAC Systems?"

29. [29] Yang and Becerik-Gerber.

30. [30] Huebner et al., "Understanding Electricity Consumption"; Huebner et al., "Explaining Domestic Energy Consumption – The Comparative Contribution of Building Factors, Socio-Demographics, Behaviours and Attitudes."