COMPILING OPEN DATASETS TO IMPROVE URBAN BUILDING ENERGY MODELS WITH OCCUPANCY AND LAYOUT DATA

AYCA DURAN[1], ORCUN KORAL ISERI[2], CAGLA MERAL AKGUL[3], SINAN KALKAN[4] and IPEK GURSEL DINO[5]
[1,2,5] Department of Architecture, Middle East Technical University
[3] Department of Civil Eng., Middle East Technical University
[4] Department of Computer Eng., Middle East Technical University
[1]aycad@metu.edu.tr, 0000-0001-6027-2962
[2]koral.iseri@metu.edu.tr, 0000-0001-7735-3363
[3]cmeral@metu.edu.tr, 0000-0001-8720-1216
[4]skalkan@metu.edu.tr, 0000-0003-0915-5917
[5]ipekg@metu.edu.tr, 0000-0003-2216-9192

Abstract. Urban building energy modelling (UBEM) has great potential for assessing the energy performance of the existing building stock and exploring various actions targeting energy efficiency. However, the precision and completeness of UBEM models can be challenged due to the lack of available and reliable datasets related to building occupant and layout information. This study presents an approach that aims to augment UBEM with open-data sources. Data collected from open data sources are integrated into UBEM in three steps. Step (1) involves the generation of occupant profiles from census data collected from governmental institutions. Step (2) relates to the automated generation of building plan layouts by extracting data on building area and number of rooms from an online real-estate website. Results of Steps (1) and (2) are incorporated into Step (3) to generate residential units with layouts and corresponding occupant profiles. Finally, we make a comparative analysis between data-augmented and standard UBEM based on building energy use and occupant thermal comfort. The initial results point to the importance of detailed, precise energy models for reliable performance analysis of buildings at the urban scale.

Keywords. Urban Building Energy Modelling; Occupancy; Residential Building Stock; Unit Layout Information; Open-source datasets; Energy Demand; Indoor Thermal Comfort; SDG 11.

1. Introduction

Buildings alone account for 40% of energy consumption and 36% of greenhouse gas

POST-CARBON, Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022, Volume 2, 669-678. © 2022 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

emissions (European Commission, 2020). The global population might reach 11 billion by 2050, with the majority of the additional population residing in cities, bringing the total urban population to 6.5 billion (United Nations, 2019). Cities must be able to assess their energy use and investigate methods for reducing energy consumption and environmental effect to contribute to achieving one of the Sustainable Development Goals (SDGs) of the United Nations, SDG11, focusing on "Sustainable cities and communities" (United Nations, 2015). In order to estimate, compare, rank, and contrast the energy used in cities by building stock, Urban Building Energy Modelling (UBEM) has gained an increasing research interest in past decades. Varying between scales of a city block to an entire city, UBEM has been extensively used to guide energy-efficient design, ensure code compliance, obtain performance rating credits, evaluate retrofit options, and optimise building operations (Hong et al., 2020).

UBEM requires several parameters for building characteristics as inputs, such as building geometry, location, climate, use type, energy systems and occupancy data. Modelling occupancy, which is a complex problem even for a single building since occupants interact with buildings in many different ways and cause uncertainty in building energy use estimations, has been a challenge for UBEM. Studies have shown that up to 30% of the variation in building energy performance can be attributed to occupants (Tian et al., 2018). Therefore, modelling occupancy has been one of the critical factors for UBEM to estimate building performance indicators accurately.

Accurate estimations in urban scale modelling are significantly affected by the availability and adaptability of large datasets to UBEM. The availability and cost of the occupancy data, along with privacy concerns, obstructed the process to obtain data for energy models (Putra et al., 2021). In most UBEM attempts, building occupancy information is rather simplified due to the lack of necessary data in district or urban scales (Mosteiro-Romero et al., 2020). When data directly linked to occupancy presence or activity is unavailable, synthetic population generation based on various data resources become one of the key solutions.

Researchers adopted several methods to model occupancy from outside resources, such as surveys, data-mining techniques, and sensors to generate synthetic occupancy data in recent decades. For instance, a recent study examined occupant presence and characteristics based on 12 years of survey data representing occupant presence in buildings and household characteristics to generate annual occupancy schedules (Mitra et al., 2020). Generated occupancy schedules relied on age, day of the week, number of household members and the age distribution of occupants in households. Researchers have found a 41% difference between commonly accepted default occupancy schedules and the schedules they generated for residential buildings, although both schedules share similar patterns. Similarly, another study relied on a time use survey and census data to assess occupancy and behavioural information (Jeong et al., 2021). Researchers obtained information for household composition and occupant activities during different periods of the year. Although similar studies have demonstrated that the generation of synthetic occupancy data is not new, implemention of the approach in the UBEM framework is yet to be explored (Happle et al., 2018).

Estimating the occupant density within a thermal zone is critical. However, estimating the exact occupant presence for a large number of thermal zones is hard to specify when data is not available. Collecting data for each residential unit of a neighbourhood is not practical and, in most cases, data is not available due to privacy concerns. Generally, standardised measures for occupancy, a fixed value for people per floor area, is supplied to energy models which do not represent the actual occupant presence and systematically lower heating energy loads (Tahmasebi & Mahdavi, 2017). In this respect, previous studies relied on the relationship between tenant units and the number of occupants (Sun & Erath, 2015). Unit layout information has the potential to reflect the probable household size. For residential buildings, the number of bedrooms can imply the number of people that could occupy the unit. Although exact modelling of interior unit division is not possible, occupant presence in units can be inferred from the number of rooms and floor areas available in real-estate service databases. Therefore, this study contributes to previous methods for synthetic occupancy generation based on census data by integrating the data obtained from real-estate advertisements to associate generated occupant profiles with building units.

This study aims to support UBEM with occupancy and layout data by compiling publicly available datasets. The proposed occupancy modelling approach contributes to the knowledge in synthetic occupancy generation for UBEM when data directly linked to occupancy is not present. The proposed approach is applied to UBEM of a neighbourhood in Ankara, Turkey, with 599 residential buildings. To contrast the results of the proposed occupancy generation approach with the default midrise apartment occupancy, two sets of simulations are compared based on performance indicators of heating energy demand (QH) and indoor overheating degrees (IOD).

2. Methodology

This paper presents a methodology for occupancy generation and residential unit division planning to increase the resolution of the UBEM framework based on real-world data sets. Generated occupant profiles were associated with residential unit layouts in terms of total area and room number (e.g., 1+1, 2+1). The building stock performance of the generated occupancy and default occupancy schedules were compared (Figure 1).

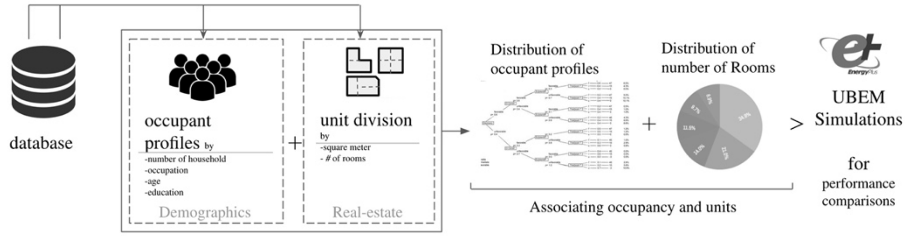


Figure 1. The proposed methodology.

2.1. CASE STUDY

The proposed methodology is applied to the Bahcelievler neighbourhood, Turkey (Figure 2). The analysis region is home to a dense mass of buildings since it is located in Turkey's second largest city, Ankara. Three to four-story buildings occupy 574,353 square meters (sqm) floor areas in the city. The residential buildings represent 93% of the building stock (599 residential buildings out of 642). The building typology of the region consists of primarily residential buildings with retail and office units on the ground floors (approximately 10% of the total units per building). Ankara is in ASHRAE 4B climate region; thus, heating energy demand is dominated, and cooling demand was not calculated in the scope of this study.

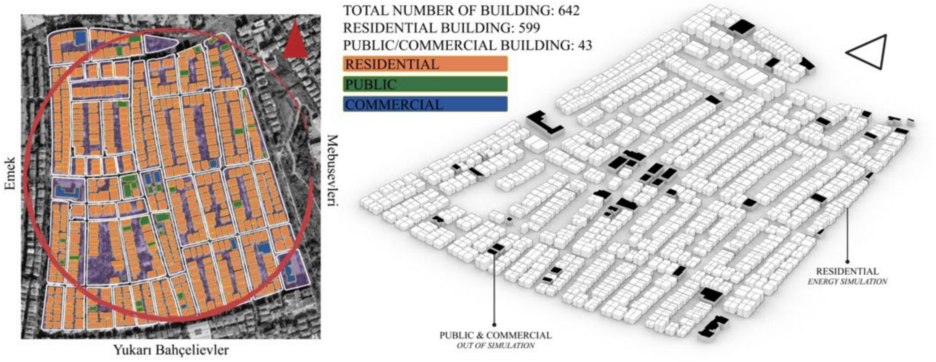


Figure 2. Bahcelievler neighbourhood.

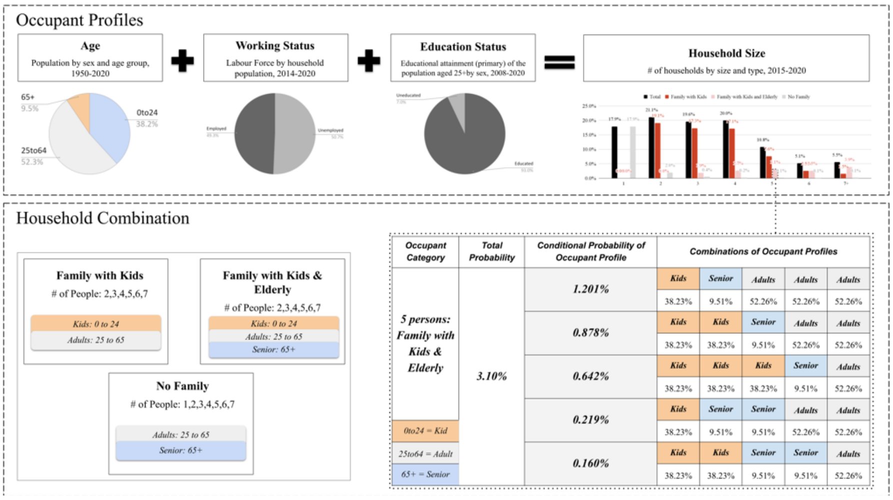
2.2. OCCUPANT PROFILE GENERATION

Detailed household information for residential building stock is not available for the studied region. Therefore, the proposed occupancy modelling was produced based on the analysis of publicly available census data obtained from the Turkish Statistical Institute (TUIK). The proposed occupancy modelling consists of two phases: the formation of occupant profiles and the modelling of the household combination according to the conditional probability methodology. Several assumptions were made during the various steps of this exploratory modelling study. In the first step, occupant profiles were formed according to the age ranges of the occupants (TUIK, 2021b), and divided into three subgroups: 0to24, 25to64, and 65+. As a next step, age groups were separated into three classes: work, school, and home (Table 1) according to their education, employment status and the information of whether residents were present at home or not (TUIK, 2021c). Occupant profiles were created by modelling their presence at home at different rates within 24 hours on weekdays. For instance (Table 1), children aged 0-5 who are at home, children aged 6-24 who goes to school and children between the ages of 15-24 who do not study or work were gathered under the same group of kids aged 0to24. Independent of the profiles, all occupants are assumed to be at home on weekends.

Occupant profile	Kids, 0to24 (38.23% of the population)	Adult, 25to64 (52.26% in all population)	Senior, 65+ (9.51% in all population)
Subgroups	Home (e.g., infant) School (e.g., elementary, kindergarten) Work	Work Home (e.g., unemployed, retired)	Home (i.e., retired)
Weekdays	Home: 8.96% School: 51.94% Work: 39.1%	Home: 39.6% Work: 60.4% (65+ employment status excluded)	Home (in): 100%
Weekend	Home: 100%		

Table 1. Occupant profiles and subgroups for weekdays and weekends.

In the second step, occupant profiles were divided into seven groups according to the size of the household, e.g., one-person household, two-person household (TUIK, 2021a). Densities of different household combinations were defined based on conditional probability in terms of the number of households and the characteristics of the members of the family. Three different family types were defined: family with children, family with children and elderly, and no family. Children, adults, and elderly groups were assigned to these groups according to their proportion in the total population. These characteristics were determined according to census data obtained from the TUIK, e.g., two-person family with kids, four-person family kids and elderly (Figure 3).



Occupant Category	Total Probability	Conditional Probability of Occupant Profile	Combinations of Occupant Profiles				
5 persons: Family with Kids & Elderly	3.10%	1.201%	Kids	Senior	Adults	Adults	Adults
			38.23%	9.51%	52.26%	52.26%	52.26%
		0.878%	Kids	Kids	Senior	Adults	Adults
			38.23%	38.23%	9.51%	52.26%	52.26%
		0.642%	Kids	Kids	Kids	Senior	Adults
			38.23%	38.23%	38.23%	9.51%	52.26%
		0.219%	Kids	Senior	Senior	Adults	Adults
0to24 = Kid			38.23%	9.51%	9.51%	52.26%	52.26%
25to64 = Adult		0.160%	Kids	Kids	Senior	Senior	Adults
65+ = Senior			38.23%	38.23%	9.51%	9.51%	52.26%

Figure 3. Occupant profile generation

For instance, the proportion of a family with five children in the total population was 7.6%. This ratio is supported by the conditional distribution of child, elderly and adult groups. Specifically, a combination of 3 adults, one child, and one elderly person is 2.327%, according to census data. The ratio found was also multiplied by the total housing unit of the neighbourhood to determine how many times this household occupant combination is repeated in the model. On the other hand, it is noticed that while some combinations are theoretically possible, they are probabilistic unlikely. For example, the proportion of households consisting of five members of non-family in the total population is 0.10%. The probability of a five-person household combination consisting of five elderlies in this combination set is minimal so that it can be ignored.

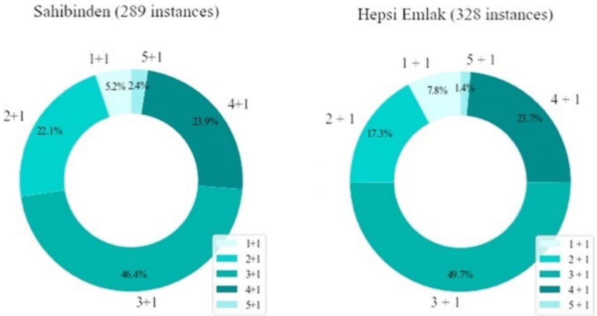


Figure 4. Frequencies of unity types based on number of rooms.

2.3. UNIT LAYOUT GENERATION

The level of detail of the building energy modelling process affects the accuracy of the results (Biljecki et al., 2014). Initially, a total number of units in the buildings was obtained through the address inquiry system (Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü, n.d.). However, the total number of units for each building obtained from the address inquiry system does not relate to household information. Therefore, the number of occupants and their likelihood of being present at home for each unit is unknown. At this stage, real-estate advertisements belonging to the studied district is interpreted. Initially, floor area and the number of rooms belonging to building units in Bahcelievler were obtained from two of the most accessed online real-estate web services in Turkey, namely, *Hepsi Emlak* (*Hepsiemlak*, 2006) *and Sahibinden* (*Sahibinden*, 2000). Advertisements available on October 8, 2021, for the Bahcelievler neighbourhood were recorded for each resource. After data preprocessing of two datasets and removal of outliers, a total number of 671 data points, 382 from *Hepsi Emlak* and 289 from *Sahibinden,* are analysed individually due to the possibility of the same advertisement taking place in both resources.

According to both the data sets, slightly less than half of the buildings in Bahcelievler have three rooms and a living room (3+1) with 46% and 49% in the examined resources (Figure 4). 2+1 and 4+1 houses occupy around 22% of the datasets, following the most common 3+1 houses. The relationship between the number of rooms and the unit floor areas (Figure 5) is also investigated to identify the number of rooms in the units in the digital model. The floor area of modelled units' is used to find the number of rooms. The probability of a unit with a specific floor area having a particular number of rooms is calculated based on the floor area and reflected in the number of rooms in units.

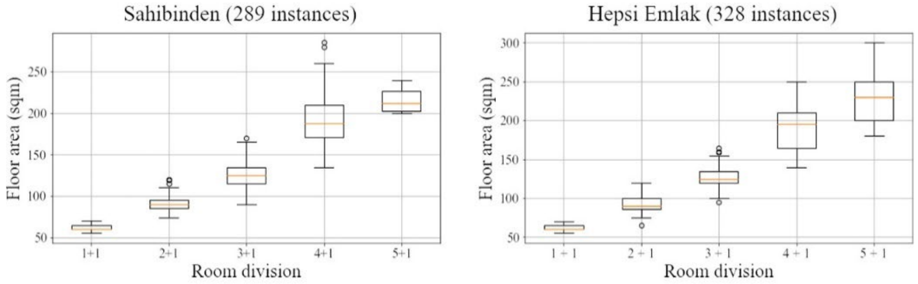


Figure 5. The relationship between number of rooms and unit floor areas.

2.4. ASSOCIATING OCCUPANCY AND UNIT LAYOUT INFORMATION

Detailed occupancy modelling for residential units is critical since internal space conditioning and electrical equipment power usage is affected by occupants. First, residential units were grouped according to floor areas and the number of rooms obtained from open-source real-estate advertisements. The number of rooms can vary according to architectural layout design; even the floor areas are the same. For instance, units between 100 and 110 sqm can contain 2+1, 3+1, 4+1. During the calculations, overlaps in the floor area and the number of rooms was also considered. Secondly, generated occupancy is mapped into the units based on the number of rooms based on the assumption that the number of people is strongly correlated with the number of bedrooms of a residential unit. With this assumption, unlikely assignments, such as, a family of 7 people living in 1+1 units, were eliminated.

2.5. UBEM PROPERTIES AND SIMULATION PROCESS

Each residential building in the case study was modelled in detail down to the size of a unit (Figure 6). For the non-residential buildings, neither floor nor unit partitioning was made, and these buildings were only included as context buildings in the building energy simulations. The original drawings of the layout of buildings were converted to simple four-corner rectangles to reduce the computational cost of the simulations. 10% of the total area of a parcel was given to the vertical circulation area. Each building and its neighbours in close proximity is parametrically modelled and simulated. Since the large number of inputs and outputs of UBEM increase computation time and modelling difficulty, each building was simulated with the residential and office units it contains together with the buildings nearby, impacting insolation as context geometries.

Building energy simulations are conducted using the EnergyPlus (Crawley et al., 2000) engine through the Ladybug Tools (Sadeghipour Roudsari & Pak, 2013). Materials were selected following TS-825-2013 (i.e., heat insulation rules in buildings) (T.C. Çevre Bakanlığı, 2013) and ASHRAE (ASHRAE, 2013) standards. A representative climate data for the typical meteorological year was used for simulations. Energy model parameters other than selected occupancy-dependent parameters, such as u-values, infiltration rate, window to wall ratio so on, are kept constant for simulations with the default and generated occupant profiles.

Two sets of simulations are conducted. The total number of floors of the buildings and the number of units per floor were determined with the data collected from the address inquiry system and real-estate agencies. Only the occupancy schedules, the number of people per sqm and the equipment density are determined as occupant profile dependent variables (Heydarian et al., 2020). The equipment use has been determined based on the age ranges of the occupants (Table 1). 65+ is assumed to have an equipment density of 2W/sqm, while the other profiles have 3W/sqm.

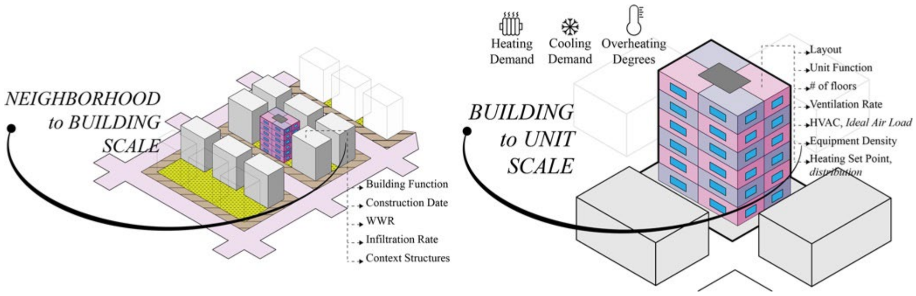


Figure 6. Properties of building energy model

Simulations with the default and generated occupancy schedules are compared based on QH (kWh/sqm) and IOD (°C). IOD estimation reflects the impact of design parameters in summers when the cooling system is not actively used. The calculation of IOD is the annual summation of the indoor operative temperature above 28°C for the whole residential units (CIBSE, 2006).

3. Results

599 residential buildings of the studied neighbourhood were simulated to observe the difference between generated and default occupancy profiles. While the QH is significantly greater, with the mean difference of 17.77 kWh/sqm, in simulations with generated occupant profiles, IOD values are similar in both cases.

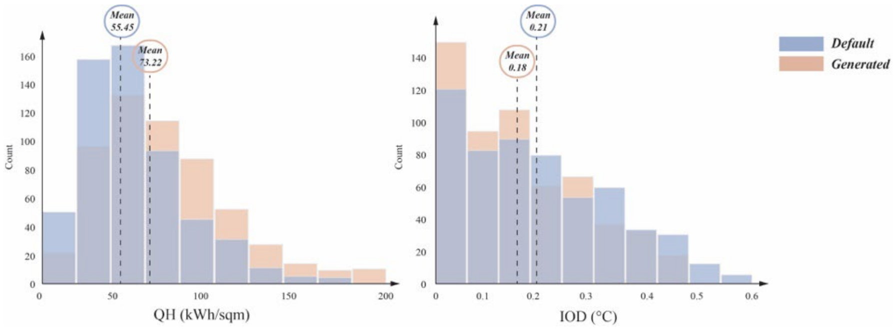


Figure 7. Default vs. generated occupancy schedule for energy demand (kWh/sqm)

Simulation outputs for QH imply that generated occupancy profiles increase instances located towards the upper ends of the data set, shifting the distribution mean to the right (Figure 7). More instances close to the median value are observed with the standardised schedules. On the other hand, IOD for generated occupancy schedules does not differ considerably from the default schedules.

4. Discussion and Conclusion

We have proposed a methodology to infer occupant profiles by compiling census and real-estate data when a direct source of occupancy information is not available. In addition to widely used census data for occupancy profile generation, we analysed real-estate advertisements to obtain data for building layouts and the possible combination of their inhabitants. Performance simulations with the proposed occupancy profile generation approach have increased the number of observations above the mean value of the QH. In contrast, the use of default occupancy profiles resulted in the accumulation of data points around the mean value of the complete dataset. QH with the generated occupancy profiles was around 32% greater than the simulation with default profiles. Unsurprisingly, default occupancy produced standardized results, while the proposed approach increased the observation of less common data points. On the other hand, the observed difference in IOD was much smaller. This result could stem from the use of mechanical ventilation during the summer period.

Results of this study rely on several assumptions and generalizations that were made and explained in the methodology chapter of this paper. The findings of this exploratory study aligns with the previous research reporting that the default occupancy underestimates the heating energy demand (Tahmasebi & Mahdavi, 2017; Tian et al., 2018). To increase the reliability of the results, multiple simulations with different yet random household profile allocation will be tested in the following study. Future studies should also consider validating the results of the proposed methodology with data directly relating to occupancy.

Acknowledgements

The study was conducted within the scope of a project titled "Tool development for urban building energy modelling (UBEM) against climate change supported by machine learning and building simulations." (No. 120M997). The project is granted by the Scientific and Technological Research Council of Turkey (TUBITAK).

References

ASHRAE. (2013). *ASHRAE Standard 90.1-2013 -- Energy Standard For Buildings Except Low-rise Residential Buildings*.

Biljecki, F., Ledoux, H., Stoter, J., & Zhao, J. (2014). Formalisation of the level of detail in 3D city modelling. *Computers, Environment and Urban Systems,* 48, 1–15. https://doi.org/10.1016/j.compenvurbsys.2014.05.004

CIBSE. (2006). *Guide A, Environmental Design*.

Crawley, D. B., Lawrie, L. K., Pedersen, C. O., & Winkelmann, F. C. (2000). EnergyPlus: Energy Simulation Program. *ASHRAE Journal, 42.*

European Commission. (2020). *Energy efficiency in buildings.*

Happle, G., Fonseca, J. A., & Schlueter, A. (2018). A review on occupant behavior in urban building energy models. *Energy and Buildings*, 174, 276–292. Elsevier Ltd. https://doi.org/10.1016/j.enbuild.2018.06.030

Hepsiemlak. (2006). Retrieved October 8, 2021, https://www.hepsiemlak.com/

Heydarian, A., McIlvennie, C., Arpan, L., Yousefi, S., Syndicus, M., Schweiker, M., Jazizadeh, F., Rissetto, R., Pisello, A. L., Piselli, C., Berger, C., Yan, Z., & Mahdavi, A. (2020). What drives our behaviors in buildings? A review on occupant interactions with building systems from the lens of behavioral theories. *Building and Environment,* 179. https://doi.org/10.1016/j.buildenv.2020.106928

Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment,* 168. https://doi.org/10.1016/j.buildenv.2019.106508

Jeong, B., Kim, J., & de Dear, R. (2021). Creating household occupancy and energy behavioural profiles using national time use survey data. *Energy and Buildings,* 252. https://doi.org/10.1016/j.enbuild.2021.111440

Mitra, D., Steinmetz, N., Chu, Y., & Cetin, K. S. (2020). Typical occupancy profiles and behaviors in residential buildings in the United States. *Energy and Buildings*, 210. https://doi.org/10.1016/j.enbuild.2019.109713

Mosteiro-Romero, M., Hischier, I., Fonseca, J. A., & Schlueter, A. (2020). A novel population-based occupancy modeling approach for district-scale simulations compared to standard-based methods. *Building and Environment*, 181. https://doi.org/10.1016/j.buildenv.2020.107084

Nüfus ve Vatandaşlık İşleri Genel Müdürlüğü. (n.d.). *Adres Kayıt Sistemi (Address inquiry system).* Retrieved December 20, 2021, from https://adres.nvi.gov.tr/Home

Putra, H. C., Andrews, C., & Hong, T. (2021). Generating synthetic occupants for use in building performance simulation. *Journal of Building Performance Simulation*, 14(6), 712–729. https://doi.org/10.1080/19401493.2021.2000029

Sadeghipour Roudsari, M., & Pak, M. (2013). Ladybug: A parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design. *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association* (pp. 3128–3135).

Sahibinden. (2000). Retrieved October 8, 2021, https://www.sahibinden.com/

Sun, L., & Erath, A. (2015). A Bayesian network approach for population synthesis. *Transportation Research Part C: Emerging Technologies,* 61, 49–62. https://doi.org/10.1016/j.trc.2015.10.010

Tahmasebi, F., & Mahdavi, A. (2017). The sensitivity of building performance simulation results to the choice of occupants' presence models: a case study. *Journal of Building Performance Simulation,* 10(5–6), 625–635. https://doi.org/10.1080/19401493.2015.1117528

T.C. Çevre Bakanlığı. (2013). TS-825-2013, *Binalarda ısı Yalıtım Kuralları.*

Tian, W., Heo, Y., de Wilde, P., Li, Z., Yan, D., Park, C. S., Feng, X., & Augenbroe, G. (2018). A review of uncertainty analysis in building energy assessment. *Renewable and Sustainable Energy Reviews*, 93, 285–301. Elsevier Ltd. https://doi.org/10.1016/j.rser.2018.05.029

TUIK. (2021a). *Number of households by size and type*, 2015-2020. Retrieved October 20, 2021, from https://data.tuik.gov.tr

TUIK. (2021b). *Population by sex and age group*, 1950-2020. Retrieved October 20, 2021, https://data.tuik.gov.tr

TUIK. (2021c). *Population by sex and age group*, 1950-2020. Retrieved October 20, 2021, https://data.tuik.gov.tr

United Nations. (2015). *#Envision2030 Goal 11: Sustainable Cities and Communities.*

United Nations. (2019). *World Population Prospects 2019.*