T.C.
IŞIK UNIVERSTIY
SCHOOL OF GRADUATE STUDIES

MASTER THESIS
DEPARTMENT OF CIVIL ENGINEERING
CIVIL ENGINEERING PROGRAM

Hadi MERCHAD

ENHANCING SIMULATION ACCURACY IN BUILDING ENERGY MODELING THROUGH DATA-DRIVEN APPROACHES

SUPERVISOR
Asst. Prof. Dr. Önder UMUT

İSTANBUL, June 2025

T.C.
IŞIK UNIVERSITY
SCHOOL OF GRADUATE STUDIES

MASTER THESIS
DEPARTMENT OF CIVIL ENGINEERING
CIVIL ENGINEERING PROGRAM

Hadi MERCHAD
(23CIVL5001)

ENHANCING SIMULATION ACCURACY IN BUILDING ENERGY MODELING THROUGH DATA-DRIVEN APPROCHES

SUPERVISOR
Asst. Prof. Dr. Önder UMUT

İSTANBUL, June 2025

T.C.
IŞIK UNIVERSITY
SCHOOL OF GRADUATE STUDIES

MASTER'S THESIS
DEPARTMENT OF CIVIL ENGINEERING
CIVIL ENGINEERING PROGRAM

Hadi MERCHAD
(23CIVL5001)

ENHANCING SIMULATION ACCURACY IN BUILDING ENERGY MODELING THROUGH DATA-DRIVEN APPROACHES

Date: 26/06/2025

Thesis Supervisor: Asst. Prof. Dr. Önder UMUT / Işık University

Jury Members:

Dr. Sadık YİĞİT / ZHAW University

Asst. Prof. Dr. Bora AKŞAR / Işık University

Asst. Prof. Dr. Ehsan ETMINAN / Işık University

Assoc. Prof. Dr. Göker AKSOY / Gebze Teknik University

İSTANBUL, June 2025

ÖZET

VERİYE DAYALI YAKLAŞIMLARLA BİNA ENERJİ MODELLEMESİNDE SİMÜLASYON DOĞRULUĞUNUN ARTIRILMASI

Bu tez, konut tipi binalarda enerji tüketimine olan kullanıcı davranışlarının etkisini, deterministik (sabit) ve olasılıksal (değişken) zaman çizelgesi modellerini karşılaştırarak incelemiştir. Türkiye genelinde gerçekleştirilen çalışmada, konutlardaki günlük aktivitelerin 15 dakikalık aralıklarla kaydedildiği 170 detaylı anket yanıtı toplanmıştır. Bu veriler, kullanıcı davranışlarındaki farklılıkları da içerecek şekilde genişletilerek 1000 adet yüksek çözünürlüklü günlük kullanıcı zaman çizelgesine dönüştürülmüş ve enerji simülasyonlarında kullanılmıştır.

EnergyPlus programında, sabit (deterministik) ve değişken (olasılıksal) çizelge yöntemleriyle simülasyonu yapılan iki konut binası modeli bir yüksek katlı ve bir alçak katlı konfigürasyon oluşturulmuştur.

Her iki bina modelinde de aynı kullanıcı davranış çizelgeleri kullanılmıştır; senaryolar arasındaki tek fark bina formudur ve bu farkın enerji tüketimi üzerindeki etkisi analiz edilmiştir.

Kullanılan metodoloji, MATLAB ve Python aracılığıyla gerçekçi olasılıksal çizelgelerin oluşturulması, bu çizelgelerin CSV formatında EnergyPlus'a otomatik aktarımı ve her senaryo için 50 rastgele simülasyonun yürütülmesini içermektedir. Deterministik modeller, literatürde yer alan standart günlük rutinlere dayandırılmış ve yıl boyunca aynı şekilde uygulanmıştır. Simülasyon çıktıları; aydınlatma, HVAC (ısıtma, havalandırma, iklimlendirme), diğer elektrikli cihaz kullanımı, toplam elektrik tüketimi ve toplam enerji tüketimi olmak üzere beş kategoride değerlendirilmiştir.

Sonuçlar, olasılıksal değerlerin genellikle dağılımın ortalamasına yakın gerçekleştiğini ancak sistem boyutlandırması ve pik yüklerin belirlenmesinde önemli olan aşırı davranışları tam olarak yansıtamadığını ortaya koymuştur. Olasılıksal modeller, priz yükleri ve elektrik tüketiminde daha geniş bir değişkenlik göstermiş; HVAC yüklerinde ise daha az değişkenlik gözlemlenmiş ancak kullanıcıların evde bulunma zamanlarındaki değişikliklerden etkilenmiştir. Elde edilen bulgular, bina enerji performansı simülasyonlarında gerçek kullanıcı davranışlarının dikkate alınmasının enerji talebinin daha doğru temsil edilmesi için gerekli olduğunu göstermiştir.

Olasılıksal simülasyonlarda toplam ortalama enerji kullanımı, iki bina modeli arasında 63.9–79.5 kWh/m^2 aralığında gerçekleşmiş, deterministik senaryolarda ise bu değerler sırasıyla 74.2 ve 71.4 kWh/m^2 olarak belirlenmiştir. HVAC yüklerinde değişkenlik sınırlı kalırken, priz yükleri ve aydınlatma tüketimi farklı davranış kalıplarına bağlı olarak önemli ölçüde değişiklik göstermiştir. Bu gözlemler, deterministik modellemenin içsel değişkenliği gizlediğini ve olasılıksal simülasyonların gerçek kullanıcı etkilerini daha iyi yansıttığını bir kez daha teyit etmiştir.

Bu araştırma, Türkiye'deki konut binalarına uygulanabilir, kültürel bağlamı gözeten ve veriye dayalı bir modelleme yaklaşımı sunmuş; olasılıksal simülasyon yöntemlerinin enerji arz-talep analizi, politika değerlendirmesi ve sürdürülebilir tasarım optimizasyonu açısından daha güçlü ve gerçekçi bir platform sunduğunu ortaya koymuştur.

Anahtar Kelimeler: Kullanıcı Davranışı, Olasılıksal Modelleme, Konut Binaları, Enerji Simülasyonu, EnergyPlus

ABSTRACT

ENHANCING SIMULATION ACCURACY IN BUILDING ENERGY MODELING THROUGH DATA-DRIVEN APPROACHES

This thesis investigated the contribution of occupant behavior towards residential building energy consumption by comparing deterministic and probabilistic schedule models. 170 in-depth survey responses were obtained across Türkiye in an effort to record daily residential activities every 15 minutes. These were augmented into 1000 high-resolution daily occupant schedules with the incorporation of variation in behavior into energy simulations.

Two residential building models, a high-rise and a low-rise configuration were simulated using Energy Plus with fixed (deterministic) and variable (probabilistic) schedule methods.

Importantly, the occupant schedules used in both models were identical; the only difference between the two scenarios was the building form, allowing analysis of geometry-driven energy variations.

The methodology used consisted of realistic probabilistic schedule creation using MATLAB and Python, automated interfacing with EnergyPlus as CSV inputs, and simulation of 50 randomized runs per scenario. The deterministic models built on standard daily routines from the literature and duplicated over all days of the year. The outputs of the simulations were evaluated in five categories of energy consumption: lighting, HVAC, other electrical uses, total electricity, and total utility consumption.

The outcomes revealed that probabilistic values tend to occur around the average of probabilistic distributions but could not capture extreme behaviors that play a significant role in system sizing and peak load.

Probabilistic models had wider variability in plug loads and electricity consumption but less varied HVAC loads that still remained influenced by changing patterns of occupant presence. The results highlighted the necessity for real occupant behavior to be included within building performance simulation for better energy demand representation.

Total average energy usage for probabilistic simulation ranged between 63.9–79.5 kWh/m² for the two scenarios, compared to 74.2 and 71.4 kWh/m² under deterministic values. Variability was seen to be restricted for loads under HVAC, but varied considerably for other plug loads and lighting based on different behavior patterns. These observations reinforce the fact that internal variation is hidden under deterministic modeling, and that probabilistic simulation gives better insight into actual occupant impact on energy usage.

The research brought a culturally informed, fact-based modeling approach applicable in Turkish residential buildings and confirmed that probabilistic simulation methods offer a stronger and more realistic platform for analyzing the energy supply and demand, evaluation of policies, and optimization of sustainable designs.

Keywords: Occupant Behavior, Probabilistic Modeling, Residential Buildings, Energy Simulation, EnergyPlus

ACKNOWLEDGEMENT

I am grateful to my advisor, Asst. Prof. Dr. Önder Umut, for his encouragement, wise counseling, and support throughout the process of my master's thesis. His mentorship not only influenced the nature of this thesis but has had a substantial academic impact on me as well. It has been an honor to work under his guidance.

My sincere appreciation goes to my co-adviser, Dr. Sadık Yiğit, for his wise suggestions, helpful comments, and positive encouragement that greatly helped enrich the depth and precision of this research.

I am indebted equally to the thesis committee members for their time, helpful suggestions, and thoughtful criticism, all of which helped in shaping and improving the quality of the work.

Most importantly, I am most grateful for my parents and friends whose unwavering belief in me, emotional support, and encouragement acted as an inspiring element and a source of strength. They not only played a pivotal part throughout this academic journey but have remained supportive in my life as well.

Hadi MERCHAD

TABLE OF CONTENTS

<u>PAGE NO</u>:

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS LIST

AM: Ante Meridiem

ASHRAE: American Society of Heating, Refrigerating and Air-Conditioning Engineers

CAM: Custom Activity Model

CO_2: Carbon Dioxide

CSV: Comma-Separated Values

DHW: Domestic Hot Water

ENERGYPLUS: EnergyPlus Simulation Software

Eppy: EnergyPlus Python Processing Library

EPW: EnergyPlus Weather File

HVAC: Heating, Ventilation, and Air Conditioning

HTML: HyperText Markup Language

IDF: Input Data File

ISO: International Organization for Standardization

kWh: Kilowatt-hour

kWh/m^2: Kilowatt-hour per square meter

MATLAB: Matrix Laboratory

m^2: Square meter

PM: Post Meridiem

TMY: Typical Meteorological Year

TS: Turkish Standards

TV: Television

UK: United Kingdom

CHAPTER 1

1. INTRODUCTION

Building efficiency is now a key area of concern in the battle against climate change. The building sector uses a significant proportion of global energy and emissions, an estimated 30% of final energy is consumed in building operations, with about 26% of related CO_2 emissions (International Energy Agency, n.d.). Improving the energy performance of buildings is therefore necessary in a drive towards climate and sustainability goals. Improved technology and building designs are as necessary as improved building performance modeling for the purposes of decision-making. A greater understanding of energy consumption by buildings can help inform efficiency factors, design low-energy buildings, and eventually lower the built environment's carbon footprint. Indeed, the use of energy within households is in itself a significant contributor to carbon emissions and is significantly affected by the activities of the inhabitants themselves (Richardson et al., 2008).

In a residential building, occupants' actions the manner in which the residents utilize spaces and equipment or systems and when they do it, all have a significant influence on the actual consumption of energy. The usage pattern of electricity and heat in a residential building is most strongly determined by the activities of the inhabitants of a dwelling (Chen et al., 2021; Widén et al., 2009). This is a reflection of the weaknesses of using fixed or deterministic plans such as those provided by ASHRAE that have little variation from one day to another. Deterministic schedules cannot accommodate regional habits, cultural tendencies, nor changes in living patterns like working from the residence. For instance, the work of He et al. (2015) revealed that the incorporation of stochastic models in the building simulation tool EnergyPlus improved conformity with observed UK dwelling thermal demand patterns.

Advancing the simulation accuracy, several probabilistic modeling methods have evolved. Some draw on transition probabilities obtained from time-use surveys (Richardson et al., 2010), and others employ machine learning or hybrid modeling methods for the identification of energy-related patterns (Chen et al., 2022; Ahmed et al., 2023). These models create more realistic and richer profiles of behavior by accounting for occupant action randomness and interactions with systems. Their extension into EnergyPlus has shown significant influence on plug load simulation, heat/cooling demand, and lighting usage (Glasgo et al., 2017; He et al., 2022). Most of the models are however either derived from the non-residential environment or lack local calibration – missing culturally specific usage patterns (Aerts et al., 2014; Cecconi et al., 2017).

The aim of this research is to improve the precision of residential building energy simulations using probabilistic, data-driven occupant schedules in the context of EnergyPlus simulation. In contrast to the conventional deterministic methods that presume a fixed occupant pattern, the aim of this work is to account for real world variation and cultural differences using stochastic scheduling based on real Turkish survey data. Weekday and weekend behaviors are differentiated as the basis for building activity-based profiles using 15-minute resolution data for simulation. In the process, it aims to minimize the difference between simulated and real consumption of energy and make contributions towards more realistic and sustainable building designs for residences. The technique is also encouraging overall sustainability goals by enabling more realistic forecasting of energy demands, which makes it possible for policymakers and building designers both to specify certain periods of high usage of energy and fine-tune systems according to actual needs of the occupants.

Furthermore, the study positions itself among prior research that has emphasized the need for local validation (Dino & Akgül, 2019; Muslim, 2021) and advances recent efforts to capture uncertainty in occupant modeling, including both aleatoric and epistemic uncertainty (Kim & Park, 2025). By using a structured yet flexible probabilistic method tailored to Turkish residential routines, this thesis adds to a growing body of literature advocating for culturally-aware, data-driven modeling in energy simulations.

CHAPTER 2

2. LITERATURE SURVEY

2.1. BUILDING ENERGY PERFORMANCE AND OCCUPANT BEHAVIOR

Understanding the occupant behavior is critical to model the energy consumptions of residential buildings accurately. Occupant behavior and schedules significantly influence heating, cooling, lighting, and appliances, and therefore, understanding behavior is the critical factor in modeling using energy simulations for buildings (Chen et al., 2021; Richardson et al., 2008). Predefined occupancy schedules obtained using typical profiles (e.g., ASHRAE) cannot capture the variability and complexity of real people's behavior. This has prompted attempts to create behaviorally responsive models. Yet, such deterministic approaches, although simple to implement, ignore behavioral heterogeneity and flexibility in time-of-day preferences, cultural routines, or seasonal response factors all important in the real residential environment.

Because internal gains due to occupant drive demand and internal heat gains, such diversity needs to be modeled in simulations to achieve correct estimates of energy use, particularly at peak times. Research like Widén et al. (2009) demonstrates how time-use data can be utilized to create very disaggregated domestic activity patterns, which provide rich information about energy uses and how energy is utilized at specific times. More specifically, sleeping, cooking, appliance uses, and home absence patterns differ from country to country and culture to culture, and therefore it is critical to incorporate appropriate local behavior in energy simulations. In Türkiye, for instance, there are many households that might have uses like late-night appliance uses and nightly family time, which don't appear in standard models.

Furthermore, comfort attitudes, awareness, and access to resources all play significant roles in occupants' behavior with heating and cooling systems and vary considerably within demographic segments. Simulation tools will misestimate the patterns of energy use unless they simulate real behavior, and will most likely predict lower than actual space heating and hot-water energy needs. As energy performance becomes increasingly central to sustainability policy and action to limit climate change, ever-more modeling will be needed not only of the envelope and the systems, but of the occupants themselves.

This insight is the foundation of the growing trend in the literature toward context-specific modeling, reinforcing the fact that the construction sector cannot decarbonize without consideration of how occupants behave within dwellings. When energy-efficient technologies are installed in buildings with unrealistic patterns of use, the difference between forecast and real performance increases, undermining long-term energy savings.

2.2. PROBABILISTIC OCCUPANT BEHAVIOR MODELING

Probabilistic approaches represent a valuable alternative by simulating behavior using statistical variability instead of predetermined schedules. Richardson et al. (2008) constructed one of the earliest large stochastic residential occupancy models from UK time-use data. It represents minute-by-minute occupant state using transition probabilities. Richardson et al. (2010) then incorporated appliance use and advanced scheduling methods. Widén et al. (2009) constructed a model from Swedish time-use data, validating its activity-based electricity and water demand profiles. Widén and Wäckelgård (2010) executed a Markov-based simulation well predictive of Swedish occupants' state transitions. The models allowed researchers to transition from homogeneous schedules to typical presence-absence patterns.

Improved probabilistic models have appeared in the past decade. Chen et al. (2022) integrated duration modeling with Markov chains to distinguish short and long absences. Cecconi et al. (2017) utilized Monte Carlo methods for probabilistic modeling, and Aerts et al. (2014) employed clustering to specify different types of households with different schedules. Kim and Park (2025) developed a model integrating both aleatoric (randomness) and epistemic (knowledge-related) uncertainty.

These studies distinguish themselves in methodology. For example, Richardson et al. (2010) focus on time-use probability transitions and are heavily weighted toward nationally gathered diary data. Conversely, Widén et al. (2009) are interested in high-resolution construction of demands using time-series synthesis and empirical measurement validation. Chen et al. (2022) add temporal dynamics to their Markov methodology using residence duration modeling allowing it to predict time in each activity state. Cecconi et al. (2017) add to the literature utilizing a Monte Carlo methodology to enable modeling of broad behavior combinations by sampling from input probability distributions. Aerts et al. (2014) branch out further by aggregating into household types and establishing schedules in relation to these, as opposed to using global transition probabilities. Each of the methods has its specific strengths: whereas Markov chains are well-suited to represent sequential patterns, clustering segments out different types of households.

In addition, some methods are more adaptable to simulation, while others are oriented toward realism. For instance, Markov models are easier to incorporate into simulation tools such as EnergyPlus, whereas data-driven clustering methods may take longer to process but deliver patterns of occupant heterogeneity that are closer to reality. Some are tailored to individual house level, whereas others are scalable to entire groupings of houses, even neighborhoods.

In spite of their dissimilarities, all the models have one common objective, i.e., to substitute oversimplified assumptions with statistically founded behavioral variability. This facilitates the modeling of occupant-driven loads in house energy simulations to be meaningfully representative and context-sensitive, especially where they are integrated with simulation tools such as EnergyPlus.

2.3. PROBABILISTIC VS. DETERMINISTIC APPROACHES

Comparisons of probabilistic and deterministic modeling have consistently revealed that the former provide results with higher accuracy and flexibility. He et al. (2015) validated a stochastic model within EnergyPlus to demonstrate greater accuracy than using deterministic schedules. He et al. (2022) modeled three behavioral types, normal, austerity, and wasteful with up to ±30% variations in energy consumption. Al-Saegh et al. (2024) have compared data-driven, stochastic, and deterministic methods and concluded probabilistic approaches to significantly better forecast domestic energy demand than the use of deterministic methods. Moreover, deterministic profiles cannot adequately represent special behavior types like night shift employees, pensioners, and households with children whose behavior is different from usual.

Although deterministic methods provide certainty and conformance to standards, they cannot monitor how actual users use light, HVAC, and plug loads during the course of a day. A variety of studies have shown that, not only might deterministic modeling potentially underestimate demand at peak hours but even ignore behavioral irregularities that might lead to energy inefficiency or wastage. For instance, where occupants are doing unplanned high-energy behavior like prolonged appliance usage or unplanned hot water uses deterministic methods might well leave them out of consideration altogether, and thus deliver suboptimal forecast performance.

A further issue is the fact that deterministic schedules produce smoothed load profiles artificially. There are actually variability and dips due to variable or sporadic behavior in real life. Probabilistic approaches allow for variability by utilizing statistically derived behaviors from survey data and time-use diaries. This produces energy simulations representative of real behavior and allows for better building performance evaluation both during design and retrofitting.

Furthermore, research has revealed that even within the same city or metropolitan area, deterministic assumptions will not always reflect the diversity of behavior within social groups. For instance, young adults, students, and teleworkers all have different rhythms to the day. In the absence of such distinctions, simulation construction will risk generating misleading results that do not pick up genuine inefficiencies, nor optimization potential. As user-centric design becomes increasingly prevalent in buildings, and ultimately the development of performance-based evaluation within energy codes, probabilistic methods will play a key role in delivering realistic outputs and uncovering particular energy-saving opportunities.

2.4. SUSTAINABILITY AND CARBON REDUCTION IMPLICATIONS

Occupant behavior plays a critical role in shaping the operational performance of buildings and, consequently, their environmental impact. Chen et al. (2022), and similarly Chen et al. (2023), emphasized the role of behavior-sensitive simulation in achieving net-zero goals. These efforts are key for integrating user diversity into low-carbon retrofit planning. He et al. (2022) and Kim & Park (2025) bring attention to the layered uncertainties in occupant modeling, which become critical in scenario planning. Dino and Akgül (2019) analyzed Turkish housing under climate change scenarios, showing that using general profiles can misrepresent regional seasonal energy demand shifts. In addition to peak load misalignment, such generalized assumptions may also underestimate the carbon reduction potential of behaviorally-informed control strategies, such as adaptive comfort or smart thermostat use. Incorporating actual behavioral diversity into models allows planners and policy-makers to design better incentives and technologies that support both comfort and sustainability.

These insights are particularly relevant for the Turkish residential context. Türkiye has a mix of climate zones, socioeconomic diversity, and daily routines that differ from Western behavioral templates. However, most residential modeling studies still rely heavily on foreign or generic data, meaning local variation is poorly captured. Existing work such as Dino and Akgül (2019) highlights the need to include real occupant routines to understand how residents respond to climate or design interventions. Furthermore, household routines in Türkiye are shaped by specific cultural habits such as late-night activity patterns, communal family use of living spaces, and less frequent but longer appliance usage sessions that differ markedly from profiles used in European or North American simulations.

While these studies show how occupant models are crucial to sustainability efforts, many of them lack country-specific data or are focused on commercial rather than residential buildings. There is also a gap in capturing how household structure (e.g., single-person versus multi-family homes) affects energy use patterns, as well as how socioeconomic differences may influence energy behaviors and choices. For example, lower-income households may reduce heating use regardless of comfort due to affordability constraints—a nuance not captured by standard profiles. This thesis aims to address these overlooked factors by incorporating detailed activity-based data into simulations calibrated for Turkish homes.

2.5. ENERGYPLUS FOR OCCUPANT-CENTRIC SIMULATION

One of the most popular simulation engines for building energy performance is the EnergyPlus, which accommodates extensive input about internal loads, HVAC, and outdoor conditions in detail. Its functionalities enable the inclusion of user-specified schedules of occupants and appliances at high temporal resolution, and thus it is very well-suited for occupant-focused modeling.

Hopkins et al. (2011) demonstrated the application of EnergyPlus to simulate a representative U.S. stock of houses using schedules at high levels of detail. Muslim (2021) examined different simulation tools and validated that EnergyPlus is one of the only ones capable of applying minute-level occupancy schedules. Glasgo et al. (2017) validated simulated electricity demand in houses using circuit-level metering data, underpinning the significance of reconciling the schedules in EnergyPlus with actual occupant behavior. Ahmed et al. (2023) noted that modeling from occupants enhances simulation accuracy but is very reliant on context-specific data and suitable integration. Widén et al. (2009) showed how to utilize time-use data to create domestic hot water and electricity load profiles and validated the same using EnergyPlus.

Even with these capabilities, limitations still apply. Default occupancy assumptions, generic thermostat logic, and simplified presence profiles are still used by many users of EnergyPlus. The tool itself, although very flexible, does not come with probabilistic scheduling capabilities out of the box. This implies that high-resolution data for occupants has to be pre-processed externally, formatted right, and imported via IDF or CSV-compatible schedules. In practice, this adds one level of complexity which discourages deeper behavioral data integration, and in resource-constrained studies, this is no triviality.

In Turkish studies, most work applying the EnergyPlus has been targeting envelope enhancements or system efficiency, not behaviorally realistic scheduling. One of the largest obstacles to behavioral realism development in Turkish simulations is still the absence of standardized, localized schedule data available in the suitable format to input into EnergyPlus. In addition, surveys with 15-minute granularity for activities are very uncommon, making it labor intensive but necessary to build stochastic residential models.

EnergyPlus supports as many as 8760 hourly schedule points annually, or even 15-minute resolution if correctly set up. Yet, specifying multiple schedule input objects, associating them with the appropriate zones, and having all the transitions match control logic (heating set points, appliance loads, light usage) is tedious scripting or using any special scheduling software. This poses technical barriers to adoption and helps to limit the use of probabilistic occupant schedules in realistic simulations.

To fill in such gaps, this thesis produces probabilistic schedules for occupants from survey data structured specifically for input into EnergyPlus, and incorporates them into simulation templates for various configurations of the building. Subsequently, the simulations derived from such data better match the actual diversity and timing of occupants and provide better accuracy in analyzing energy uses, load profiles, and comfort levels indoors. This illustrates the manner in which probabilistic schedules, if incorporated within a rich simulation platform such as EnergyPlus, will allow richer insight into the reality of how well buildings perform and enable better-informed design and policy strategies.

CHAPTER 3

3. METHODOLOGY

A Figure 3.1 presents the general flow of the study. It is initiated with the collection of people's daily activities using surveys. The data is processed and used to develop the probability tables that form the basis of daily schedules. Schedules generated are readied for Energy Plus and combined with building model (IDF file) and weather file (EPW file). The inputs are all used in running EnergyPlus simulations using an automated process. Lastly, the simulation output is obtained and examined in order to determine the impact of occupant behavior on energy consumption in residential buildings.

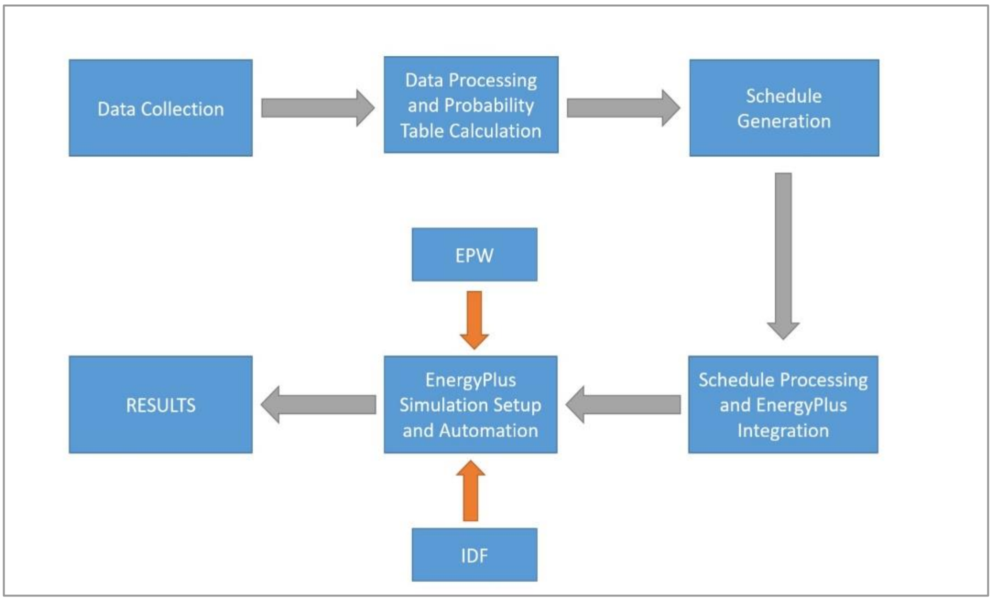

Figure 3.1 Overall workflow of the methodology

3.1. DATA COLLECTION METHODOLOGY

A key component of this research is comprehending how residents use their living areas and energy consumption during the day. A survey-based methodology was used to collect empirical data on home energy usage in Türkiye. The goal of the survey was to capture people's everyday activities with a high degree of temporal precision so that occupancy trends and energy-using habits could be more precisely understood.

The approach was created to gather data from people in the real world who had a variety of habits, jobs, and demographics. In addition to capturing individuals' general behavior, the study included a distinction between weekday and weekend routines, acknowledging that social habits and work schedules might have a substantial impact on energy use.

3.1.1. Survey Design and Structure

In consideration of clarity, completeness, and user-friendliness, a systematically planned questionnaire was developed as an initial instrument for obtaining time-specific occupancy and energy consumption data. It was built through the use of Google Forms, an openly available and accessible platform, allowing for streamlined data acquisition, live monitoring of responses, and direct export for further analysis. A checkbox grid approach was chosen because of its flexibility in capturing time-based data in detail; respondents were instructed to submit an account of their activities at every 15-minute increment within 24 consecutive hours. Such fine temporal resolution was necessary in the study because it facilitated accurate monitoring of occupant activity and related energy actions within and across different time points. For purposes of inclusivity and high overall participation, respondents were offered the questionnaire in two forms using different languages: another in Turkish for convenient understanding and participation among native Turkish speakers, and one in English for participants who were more familiar with English, among them foreign inhabitants resident in Türkiye. Both forms of the questionnaire had exactly the same structure and content, thus minimizing any language-based bias and enabling direct equivalence of responses for purposes of comparison.

The questionnaire consisted of various major sections focused on capturing contextual as well as behavioral information. The initial part was to capture demographic information to facilitate segmentation and group-level analysis. Participants were requested to state their gender so gender-based differences in energy behavior can be examined; choose their age bracket so analysis can be done across different generations; and state their occupation so understanding professional context and daily routines' impact on time spent at home and energy consumption can be done. Demographic information was necessary not just for categorization purposes but for understanding subgroups' uses, for instance, for students, working adults, remote workers, home stayers, and retirees.

The second large component of the survey was activity-based occupancy schedule data. Participants were asked to keep account of what they were doing for an entire day using the 15-minute segment grid, providing rich and fine-grained data for examination. The activity list was extensive and spanned a broad spectrum of in-residence and out-of-residence behaviors. The full list of activities included: sleep, outside home, shower, cook, hair dryer/iron, phone, TV, laptop, washing machine, dishwasher, using sport machines (indoor), cleaning machines, and being at home without using energy. Such detail permitted not just determination of what activity was occurring, but also exactly when and for how long, allowing highly accurate energy use profiles to be developed for specific lifestyles.

Last, an essential aspect of the survey was its differentiation between weekday and weekend activity. Recognizing the possibility that daily activity frequently varies considerably depending upon the day of the week, each respondent was requested to complete the questionnaire twice: once for an average weekday and again for an average weekend day. This format permitted variations in behavior such as earlier rises and more organized routines during weekdays, contrasted to late mornings and increased leisure-based activities at weekends. Furthermore, differences in energy-based actions such as greater evening use of electrical devices and increased home-cooked meals at weekends were captured in the data. Both weekday and weekend entries were essential in gaining an overall view of home life and enabled two independent accurate probability tables for daily and weekend activity to be produced, which were subsequently inputted into simulation modelling. In overall design, the survey guaranteed highly structured, rich, and diverse data were collected, which would serve to enable accurate and meaningful energy behavior modelling in future research phases.

3.1.2. Survey Implementation and Distribution

After testing for clarity and finalization, the questionnaire was shared in a multi-level manner such that it reached a diverse population of people with different lifestyles and backgrounds. This provided heterogeneous, balanced, and truly representative data. Sharing in a formulated manner among academic and professional networks was one of the primary methods of doing this. Professors, academic staff, and other professionals were approached and asked to disseminate among their colleagues, staff members, and professional networks. This academic networking particularly helped in obtaining replies from people with systematic daily routines such as university-based students living in residence halls or shared apartments and academic professionals with regular work routines. Since academics have a wide range of social and professional contacts, this helped extension towards a diverse group of participants based on various professions and daily routines.

The second method employed involved peer-to-peer sharing that is commonly referred to as “snowball sampling.” In that process, the questionnaire spread among friends, family members, and colleagues and these individuals encouraged the next level of people in their own networks. It assisted in the more natural spread of the questionnaire and included responses that might not be directly related to academic networks. It assisted in reaching professionals from diverse industries, stay-at-home individuals with a variation of energy usage patterns different from people going out for work, and even older participants with a different daily routine that might differ from younger individuals. This snowballing greatly increased the scope of the participant base and assisted in the enhancement of the reliability as well as the representativeness of the questionnaire information.

Along with academic and peer sharing, the survey was distributed across digital platforms as a way of further expanding its reach. It was shared within WhatsApp groups, posted within professional networks on LinkedIn, and attached in academic and work-related email newsletters. These digital channels ensured that participants had access and convenience in filling out the survey as they did not need to attend any meeting or physical sessions. The combination of all three methods of distribution academic networks, peer sharing, and digital outreach was able to reach a broad and diverse audience within a limited timeframe. Therefore, the data received was diverse and comprehensive, providing a good base for the rest of the research.

3.1.3. Response Collection and Maintaining Data Integrity

During the period of data collection, a total of 170 participants successfully completed the survey. Since each person was asked to fill it out twice, once for a typical weekday and once for a typical weekend, the final dataset contained 340 individual daily records. This provided a rich and detailed set of information for analysis, covering a wide variety of household routines and energy usage habits across different days of the week. However, gathering responses was only part of the process; ensuring that the data was high-quality and reliable was equally important. To maintain data integrity, several careful steps were taken after the survey responses were collected. First, every submission was reviewed to check for errors, missing time slots, or any activities that seemed unrealistic in duration or frequency. For example, if a participant marked “sleep” across an entire day or used too many activities at once, those entries were flagged for review. Any responses that were incomplete or clearly incorrect were excluded from the final dataset to make sure the results were not skewed by faulty data. Additionally, duplicate submissions were removed so that each person was counted only once in the analysis. Another crucial stage for validating data was to ensure weekday and weekend answers for each participant were reasonable. For instance, an individual might have an early rising time and set working schedule during weekdays, while weekend activities might include extended sleeping periods and greater leisure time. Ensuring such discrepancies came through in the answers was one aspect of verifying that participants had thoroughly understood and filled in the survey. This double-checking allowed for enhanced credibility in the final dataset and for guaranteed authenticity in the schedules rather than arbitrary answers in haste. Due to such stringent quality checks, the final dataset was not merely voluminous but clean and dependable, ready to be used in further data handling and interpretation.

3.1.4. Ethical Considerations

Throughout the process of data collection, research strictly adhered to ethical standards in safeguarding participant privacy and handling information in an ethical manner. One of the initial and fundamental steps involved obtaining consent. Participants were adequately informed about why they were participating in the research, what information was being collected, and how this information would be utilized. They were also informed that participation was voluntary, and they have every right to discontinue at any stage without incurring any repercussions. This process helped ensure each and every individual who undertook the questionnaire had complete knowledge of what they were doing and willingly decided to participate.

In order to maintain participant privacy and ensure anonymity, the questionnaire never collected any information that would identify individuals by name, email address, phone numbers, or home address. This guaranteed that all answers could never be traced back to an individual participant so they were free to give answers without worrying about privacy. Although the questionnaire was about daily routines and energy consumption instead of intrusive personal information, the requirement for answers to be anonymous contributed an added sense of responsibility. All the responses were safely kept and used solely for research purposes in an academic context. They were stored in secure computer files, accessible to the researcher (the author of the thesis) and supervising instructor alone.

3.2. DATA PROCESSING AND PROBABILITY TABLE CALCULATION

Following the success with the collection of 170 responses from participants throughout Türkiye, the subsequent step of importance involved the processing of the data in preparation for analysis. The responses obtained, with comprehensive 15-minute interval activity records, were extracted from Google Forms into a CSV (Comma-Separated Value) file format for systematic and orderly processing.

3.2.1. Organizing and Structuring the Data

After the CSV file of all the responses received in the survey was exported from Google Forms, the process of preparation began by structuring the content in a manner that enabled efficient analysis with a meaningful context. Each entry in the file corresponded with a person's daily activity schedule spanning a complete 24-hour duration with 96 intervals of 15 minutes each. Since each respondent filled in two distinct answers, one for a week day and one for a weekend,the dataset was divided into two separate sets. The division had to be carried out separately since the behavior pattern differences with workday responsibilities compared with the freedom of weekends had to be captured accordingly. Proper labeling of each entry with regard to its weekday and weekend status ensured the distinct processing of the weekday and weekend records in the subsequent phases. The raw responses were cleaned and formatted as well as verified for any formatting issues at this stage to ensure each entry remained correctly aligned with its time slot as well as activity type. The formatted data was prepared in a format ready for direct consideration of the probability calculations, enabling the analysis process to continue with a robust, well-organized set of data.

3.2.2. Probability Table Calculation

Following the successful structuring of the dataset into weekday and weekend entries, the subsequent step involved the computation of the likelihood of each activity taking place in each 15-minute interval over a day. This step played a significant role in the creation of the subsequent probability tables that would be used later for the simulation of realistic occupant activity patterns. The tables were formulated with a view of identifying how likely each activity such as sleep, cooking, the utilization of electronics, or spending time away from the dwelling is likely to happen whenever in the course of the day. To make this possible, a simple but-efficient mathematical equation was employed to determine the likelihood of occurrence for each of the activities in each of the time intervals as per the responses in the survey. The equation used for the same purposes of measurement included:

$$P(At) = \frac{\text{Number of participants performing activity A at time t}}{\text{Total number of responds}} \qquad (3.1)$$

In this equation, P(At) is the likelihood of activity A taking place within a given time slot t. The number of people who reported doing that specific activity at that time is in the numerator, and the number of respondents in the dataset is in the denominator. This process was carried out separately by each 15-minute interval of the day and done separately for each type of activity. Thus, for each time slot, the entire probability distribution across all possible activities was created. Therefore, it was possible to ascertain, for instance, the likelihood of a person sleeping at 02:00 AM, preparing a meal at 06:30 PM, or using a laptop at 10:15 PM.

In practice, this was realized by scanning the data row by row in order to find out which activities each of the participants reported for a certain time period. For a given 15-minute interval, a count of how many participants were doing each of the activities was taken. These counts were divided by the number of participants in order to find the corresponding probabilities. This procedure was separately done for weekday responses and weekend responses, each yielding a different table of probabilities: one table for a standard workday and one table for weekends. Each table contained 96 rows (each of the 96 time intervals within a day), each of them holding the probabilities for all the activities observed in that particular time interval.

Following the initial calculations, the information was double-checked for consistency. The sum of all the probabilities of any individual time slot should have been approaching 1.0 (or 100%), as each time slot should constitute a complete array of activities during that time. Any discrepancy missing values, inconsistencies, or responses not aligned with each other was looked into and rectified by referring back against the original raw data. This verification process served the dual purpose of ensuring the final tables of probabilities being accurate, balanced, and true reflections of human behavior patterns. Having finished the calculations, these tables served as the basis of the next phase of the study when they were used within simulation tools to produce realistic, randomized daily routines.

3.2.3. Handling Overlapping Activities: Probability Adjustments

In data processing, one of the difficulties was presented in the manner in which some of the activities were categorized in the initial design of the survey. In some cases, multiple devices or appliances were listed in one checkbox, so their separate probabilities of use had to be isolated and equally distributed in generating the probability tables. In an effort to have final probabilities represent actual day-to-day behavior, an adjustment process was used to split those collected activities into separate components.

The initial large categorization included device use, in which phones, television (TV), and laptops were grouped under one common checkbox. Since respondents who used this option might have used any one of the three devices, and one could not determine which one, the only reasonable resolution was to divide the overall probability among them. That is, for any time period for which this combined category was chosen, its probability was distributed equally among the three devices. That formula utilized was:

$$P(\text{Phone at t})=P(\text{TV at t}) = P(\text{Laptop at t}) = P(\text{Device Usage at t})/3 \quad (3.2)$$

It guaranteed that all the devices had an even proportion of the overall probability, with fairness and evenness of the dataset maintained.

An adjustment of this type was required for the other classified category consisting of the hairdryer and the iron. Unlike the previous scenario, however, the two machines are not used at the same rate. Hairdryers appear in usage more frequently than the iron does in the course of daily life, specifically in the household scenario. To accommodate this daily usage pattern, an unequal but not an equal division was employed. Based on an educated guess and from a practical usage perspective, a division of two-thirds for the hairdryer and one-third for the iron was used. This was expressed in the following formulae:

$$P(\text{Hairdryer at t}) = (2/3)P(\text{Hairdryer-Iron at t}) \quad (3.3)$$

$$P(\text{Iron at t}) = (1/3)P(\text{Hairdryer-Iron at t}) \quad (3.4)$$

3.2.4. Observations and Trends Noticed During Processing

During data processing and probability table calculations, various apparent trends in behavior started to surface that were consistent with regular daily routines and electrical consumption habits in homes. One of the strongest trends was an extremely high likelihood of sleeping at nighttime, primarily between 00:00 (midnight) and 07:00 AM. Here, an overwhelming proportion of subjects reported being in a sleeping state, validating expected nighttime behavior and offering a consistent baseline for simulating evening electrical demand. Another apparent trend was in weekday work and school hours, in the time block 08:45 to 16:45, when an overwhelming proportion of subjects reported leaving home. This trend strongly corresponded to the standard Turkish working and schooling schedule and reinforced validation of weekday dataset credibility. It also showed that at this time block, home electrical consumption would also decline, as electrical appliance activity would be less.

In contrast, weekend activity was more variable and flexible. As contrasted with the fixed pattern of weekdays, activity distribution in weekends was less concentrated in specific hours of the day and involved fewer individuals leaving home early in the morning. Rather, individuals stayed home for longer in the morning, and activities were distributed over different periods. This absence of fixed structure in weekends extended further to larger segments of home-based activity, leading to an independent pattern of energy demand during weekends. A second observation was an evening peak in device use, particularly for viewing television, using laptops, and charging phones. These were clearly distinguishable in probability level in terms of peaks in use between 20:00 and 23:00, of which the peak was even stronger during weekends, when individuals had more leisure time to spend at home. These trends highlighted the significance of an accurate weekend evening energy demand model, particularly for leisure use.

Collaboratively, these trends validated the reliability and consistency of the dataset and also helped determine various key behavioral assumptions for future simulation models. Identifying such trends in processing enabled research to construct accurate probability distributions and verify that modeled energy consumption closely replicated actual day-to-day and time-of-day behaviors.

3.2.5. Data Verification and Preparation for Further Analysis

Before we progressed to the simulation and modeling phase, it was necessary to ensure the validity and completeness of the processed data. Having performed this last step, we ensured that the derived activity probabilities from the responses of the survey could confidently be used in energy performance analysis. To start the verification process, the dataset underwent a rigorous check for missing entries or invalid activity records. Any response containing empty time periods, repeat rows, or conflicting activities, a person listed as both "sleeping" and "outside of the house" simultaneously was tagged and eliminated. It wasn't just necessary for these quality checks to prevent logical flaws; it also helped ensure that the calculations of the probabilities remained consistent and valid.

After the dataset cleaning process, the activity data was structured into a matrix format such that each row represented a 15-minute time slot throughout the day and each column included the calculated probability of a given activity within that time. Having the activity data in a matrix format enabled easy incorporation within simulation tools and the capacity for the comparison of activities within various time intervals in an orderly manner. It also provided a full and complete set of probabilities for every row that added up close to 1.0, reflecting a well-distributed set of all the activities within each time slot.

3.3. SCHEDULE GENERATION USING MATLAB SIMULATION

Following the creation of both the weekday and weekend probability tables, the subsequent step was to develop realistic daily activity schedules with the assistance of MATLAB. This process played an integral role in the simulation of occupant behavior and energy consumption patterns as it enabled the creation of time-stamped activity profiles from the calculated probabilities. The simulation served the purpose of mimicking real-life variability in how individuals construct their daily routines within realistic constraints such as minimum activity durations and reasonable transitions of activities.

3.3.1. Purpose of the Schedule Generation

After weekday and weekend probability tables had been completed, the next step in the methodology was creating realistic daily activity schedules based on these distributions. This step was central for the simulation of the way individuals behave in a residential building within the course of 24 hours. Simulating according to the probability-driven behavior observed in the surveys but with the incorporation of rational rules along with realistic time-use tendencies was the target. A simulation approach was taken with the help of MATLAB such that one is able to create rich occupant schedules by apportioning specific activities within every 15-minute interval of the day.

The generation of the schedule employed probabilistic logic, and instead of filling each time slot with the most likely activity, the simulation introduced the random elements weighted by likelihood, and thus various different schedules were generated each simulation and even from the same probability table. This increased the diversity of the outcomes and provided them with a more realistic texture, more accurately reflecting how people's habits diverge day by day. In addition to this, certain constraints on action had to be added in an effort to eliminate impossible transitions as well as in an effort to respect common habits. A person would not suddenly transition from “sleep” into “outside of the home,” for example, and activities like sleep or cooking required a minimum duration in real-time.

By mixing rule-based logic with probability-driven information, the resultant schedules provided a realistic and dynamic simulation of daily life. It ensured that each activity was selected against true survey patterns and enabled the creation of varied daily routines that imitate real people's behaviors. 700 weekday and 300 weekend schedules were generated and stored, creating a pool of ready-to-use activity patterns that would be used in the subsequent step of the methodology. These enable a fundamental basis for allocating varied probabilistic occupant behavior to people in subsequent energy simulation phases.

3.3.2. Overview of the MATLAB Code

For creating activity schedules, two different MATLAB programs were written, one for weekdays and one for weekends. Although similar in structure and principles, each of them incorporated separate assumptions and time-based rules to represent different workday and non-workday-based behavior.

For weekday code, we began by loading the weekday probability table from the CSV file. The table had 96 rows (for every 15-minute period in a day) and multiple columns for each activity. Normalization was done for each set of time slot probabilities so that they added to 1. Having loaded, code extracted the list of activities and corresponding time slots, and used a weighted random choice approach like in Markov Chain to pick an activity using the probabilities of the current time. A Markov approach in this context is choosing the next activity purely based on current time slot's distribution without reference to activity history. This meant every time slot was still statically independent while generating coherent-appearing schedules when blended with behavioral rules.

In an effort to prevent unfeasible agendas, rational constraints were used. These were in the form of minimum durations for some activities (e.g., 4 hours for sleeping, etc.) and time-based considerations that shaped the flow of daily routines. Realistic transitions were also enforced through code. For instance, if an agent had been sleeping, they were not allowed to directly move to "outside home" in the following time slot—they would have to do an intermediate activity, such as showering or "at home without energy use." Post 20:00, relaxing activities such as watching television or phone use were encouraged to mimic an evening lifestyle. The weekend variant of the code had the identical structure yet greater flexibility in its actions, for instance, permitting an 11:00 AM wake-up time. It had max duration caps for activities such as watching television or working on laptops to prevent excessively recurrent patterns. Additionally, weekend logic permitted longer consecutive home activities and incorporated a late-hour routine in which sleeping was only allowed after an unwinding activity.

In each code, the completed schedule was constructed slot by slot following the prescribed rules and probability weights. When completed, the activity schedule was printed and saved as a table of what activity was performed at each point in the day. These schedules generated by MATLAB were used as the key behavioral input for energy use simulation and served as the basis for the ensuing simulations.

3.3.3. Algorithm Structure

The schedule generation phase followed a predefined algorithm coded in MATLAB with the weekday or weekend probability table as the input. The tables had the probability of occurrence of each activity for each 15-minute interval of the day. The data was loaded and the variables initialized such as the current schedule array, the previous activity, and the duration rule enforcing timers.

It iterated over all 96 time intervals, choosing activities via weighted random selection according to current probabilities. A running activity that had not yet fulfilled its minimum duration continued if already selected. In all other situations, there were logical rules that governed transitions. E.g., a subject cannot transition directly from sleeping into leaving the house, and for times 20:00 and later, relaxing activities such as TV watching or phone usage were favored in an attempt to reproduce realistic evening rituals. Minimum durations avoided unrealistic activity switching.

The weekend schedule employed the same format but with increased flexibility, such as permitting later rises and less structured activity. It also added limits on the maximum duration of each activity to avoid repetitive behavior, such as watching television or using a laptop. These limits kept the calendars flexible but behaviorally plausible.

As the loop ended, the last schedule of one activity per time slot was saved in tabular format with the corresponding time intervals. It was readily reviewable and ready for utilization in subsequent phases of the research. It was programmed in a repeatable manner such that it produced distinct realistic schedules whenever it was invoked, all the while adhering to the underlying behavior patterns as well as the restrictions set in the code.

3.3.4. Challenges and Refinements

Some of the challenges that occurred in the simulation creation and testing that required meticulous improvement in the logic included the overly long extension of sleep late into the morning, mainly on weekends. It was rectified by adding additional logic that forced the sleep transition after a specific period of time to render the schedules realistic and as per popular daily habits.

Another problem involved over-clustering of a given activity, particularly with respect to the use of electronic equipment. Absent any other restrictions, the model occasionally allowed repetitive behaviors like the use of television or laptop for impossibly long times. To counteract that, the CAM included limits at weekends on the duration of some activities, along with checks for finishing long stretches of repetitive activity.

The transitions among activities also had to be calibrated. Direct transitions between sleep activities and "out of the house" were originally permissible, but proved not realistic. These transitioned into intermediary activities like waking up and getting ready in place of them in an attempt to capture the natural process before going out of the house. These changes in combination served the cause of creating generated schedules not merely data-driven but behavioral plausible.

3.3.5. Final Output and Integration

The outcome of this schedule generation procedure was a dataset of 1000 complete daily activity schedules of which 700 were created for weekdays and 300 for weekends. This division was specifically chosen for the replication of the natural distribution of weekdays and weekends in a standard week. Since a standard week is made up of five weekdays and two weekend days, that 70:30 ratio (or approximately 5:2) is realistic and guarantees that simulations covering multi-day or weekly periods have balance and validity.

Each of the daily schedules occupies a complete 24-hour period, partitioned into 96 15-minute intervals and holding a comprehensive list of the occupant activities for every interval. Probabilistic variation is included in these schedules along with behavioral logic, yielding a wide range of realistic daily routines that reflect patterns that have been observed in the original survey. It is this dataset that then forms the basis of the next phases of simulation when daily energy consumption is modeled based on occupant behavior.

3.4. PYTHON-BASED SCHEDULE PROCESSING AND ENERGYPLUS INTEGRATION

Following the generation of the daily activity schedules in MATLAB from the probability values came the step of transforming them into formal numerical schedules that could then be readily read in by EnergyPlus. This constituted several Python scripts wherein a script would loop through an individual's activities, generate variations within units of apartments, and then aggregate them finally in order to create the building level overall energy schedule. The ultimate goal was to show how different individuals occupy houses on an average day and translate that in the building energy use consumption format for input in engineering-based building usage simulations. It bridged the missing step between human behavior as inferred from surveys and advanced engineering-based simulations.

3.4.1. Purpose and Role of the Python Workflow

The Python scripts acted as an interface between survey inputs and the energy analysis tool. Their function, in essence, was to quantify people's self-stated activity patterns in describing activity in the context of energy. These included being-at-home, appliance use, hot water or lighting use, and the like. This was in order for lifestyle information to be translated from qualitative to something that could be read by EnergyPlus. The workflow of the scripts wasn't just for processing individual behavior, however, but for deriving multiple, diverse sets of apartment-level schedules. Randomly sampling different sets of individuals from the survey database, the workflow generated several different profiles of apartments. The diversity of the profiles enabled us to consider the impact of diverse patterns of living on energy use, resulting in results that were precise as well as realistic. The scripts were developed in order to perform all of these functions automatically, for the purpose of uniformity of output as well as ease in terms of time cost of preparation of input for simulation.

3.4.2. Translating Daily Activity into Numeric Schedules

To carry out this process, the workflow began with selecting four of the survey respondents at random from 1000 respondents, representing a group of four individuals living together in the same household. These respondents had recorded activities in 15-minute blocks over the course of a workday, creating in-depth knowledge of the way in which individuals spent their time in the home. Activities included sleeping, cooking, use of electricity, showering, or leaving the home, all of which have power implications.

Each 15-minute interval was then processed in order to estimate values for electricity demand. Occupancy was determined through observation of how many of the four selected individuals, assumed to be living in the same household, were actually home during each time interval. Appliance usage was estimated through determination of activity such as the watching of TV or work on the laptop. Lighting demand was estimated from observation of individuals being home and awake, and hot water use as being then associated with shower activity. One of the key calculations in the calculations made was electric appliance usage fraction. This was determined through observation of the number of individuals using appliances and an added bonus for the fridge, which was assumed always to be continuously operating. The total was then divided by the number of appliance types being accounted for. This made sure that there always existed a small, reasonable amount of electricity available, even when other appliances were not being used.

With these behavior-based calculations, some values were fixed throughout the day. Internal heat gains from human presence were set to 125.28 W per person, representing the total sensible and latent heat (radiative, convective, and evaporative) emitted by individuals engaged in light activity. This assumption corresponds to a metabolic rate of approximately 1.2 met, as defined in ISO 7730:2005 for sedentary or light office work and widely adopted in thermal comfort and building energy simulation standards (International Organization for Standardization [ISO], 2005). In addition, the heating and cooling set points were kept constant at 19°C and 26°C, respectively. While 20°C is defined in TS 825:2024 as the standard winter indoor temperature, the 19°C value reflects the earlier TS 825:2008 version and remains technically valid. The cooling temperature of 26°C is consistent with both past and current versions of TS 825 as the standard summer indoor design temperature (Turkish Standards Institute, 2024).

These variables combined to create an 96-entry daily schedule of activities with one for each 15-minute interval. The schedule took into account when individuals were at home, which appliances they were most likely to be using, and how their activities affected hot water and light usage. It was a whole day of activity related to energy that resulted from one particular combination of actual individuals' activities.

3.4.3. Conversion of Daily Schedules to Annual Hourly Format

Although the 15-minute schedule proved useful detail, it did have to be translated in order to supply the hourly resolution used with EnergyPlus simulations. To do so, the daily schedule was duplicated for all 365 days of the year in order to generate an annual repeating year-round pattern. The four 15-minute values within each hour were then averaged in order to generate an individual hourly value of every variable. That way, the overall structure of the schedule could be maintained while attaining the time resolution used for simulation.

The conversion produced the resulting final schedule with 8,760 rows for every hour of the year. The key variables of occupancy, lighting, appliance uses, hot water, and system settings were included and formatted in an EnergyPlus compatible way. This included having discrete column headers, standard variable order, and time step resolution correctly defined. EnergyPlus compatibility enabled the output file being input for simulation directly.

3.4.4. Automating the Generation of Multi-Apartments

After processing for an individual flat, its annual calendar of hours would be generated. The same process would then be done multiple times for the different variations in an entire housing complex. For every subsequent flat, another set of four random individuals would be selected, with the same process being used as for generating another calendar. The resulting documents for every flat would be saved under its individual title for the purpose of management.

After generating schedules for all the apartments, the generation of the building-level average schedule was the subsequent step. This was done through the summation of values of all the apartments' hours of operation for an average value per time step. The generated building schedule captured both the typical patterns as well as the variations within households thus creating a realistic outcome of the individuals' electricity consumption for the entire building. Organization and structure of the average schedule were as per EnergyPlus specifications so that the generated average schedule could directly be used in the simulations.

3.4.5. Summary and Integration within the Simulation Process

Briefly, the workflow of Python's schedule processing played an important role in reducing raw behavioral input data to structured, usable inputs for simulation. The workflow preserved both day-to-day pattern and variation from house-to-house through survey-based activity analysis, aggregating schedules from multiple apartments as well as through building-level average calculation. The workflow accounted for a large set of energy-related activities including occupancy, lighting, appliance use, and hot water demand as well as uniform modeling assumptions such as temperature set points as well as infiltration rates. The clean, hourly schedule produced for the entire building's energy behavior was suitable for direct input for EnergyPlus as well as the foundation for the remainder of work for this research effort: simulating building performance for energy as well as estimating the impact of occupant behavior on results at the building scale.

3.5. ENERGYPLUS SIMULATION SETUP AND AUTOMATION

After creating behavior-based occupant schedules using Python, the next step of the methodology involved testing how these schedules affect building energy performance. To do this, simulations were carried out using EnergyPlus, a widely used building energy modeling software. Since the study involved a large number of behavioral profiles, the simulations were automated using a Python script, which allowed each schedule to be tested under two different building models. This setup provided a structured and consistent way to understand how daily routines impact energy use in both energy-efficient and less efficient homes.

3.5.1. IDF Templates: High and Low Building Models

Two different building models were designed and utilized within the simulations. Both were written in EnergyPlus using the IDF format to define different residential building types. The IDF (Input Data File) is the core format used by EnergyPlus to describe a building’s physical and operational characteristics. It contains definitions for geometry, materials, schedules, internal loads, and simulation settings, which EnergyPlus reads to simulate energy demand over time. The first model, the High-Rise Building IDF, is representative of a high-rise, multi-storey apartment dwelling. This model has zones denoted from Storey 0 to Storey 8, indicating it simulates at least a nine-storey structure. It was set up to simulate a high-rise residential building with stacked zones and a tight vertical profile. This model does not have any particular HVAC system incorporated, but it does utilize the Ideal Loads Air System to simulate heating and cooling thermal demands for occupants and zone conditions.

The Low-Rise Building IDF model is the second model and is smaller and low-rise, with only four floors and zones from Storey 0 to Storey 3. It is representative of an even spread-out, low-rise building with even more surface area to volume ratio. Similar to the high-rise model, it has the Ideal Loads Air System in lieu of a full HVAC system. It uses similar construction naming conventions and material layers, so thermal envelope properties of the building are not extremely different.

In spite of all the similarities, the disparity in form, configuration, and thermal conditioning still causes measurable differences in energy performance under similar internal and weather conditions. Charting the same occupant schedules in both types of buildings, the study focuses on monitoring how the shape and complexity of the building influence energy use. In each of the models, external schedule files were employed to specify variables like occupancy, activity levels, light utilization, appliance usage, hot water demand, and thermostat levels. All the rest of the components of the building models internal heat gains, wall constructions, and system type were identical to enable a level playing ground in the comparison.

Both models were simulated using the same occupant behavior schedules, internal loads, and thermostat settings. The only variable was the building form, enabling a focused comparison of geometric influence on energy demand under identical usage conditions.

3.5.2. Weather File

To simulate real-world conditions, every run utilized a weather file specific to Istanbul, Türkiye. All runs utilized the TUR_IB_Istanbul-Ataturk.AP.170600_TMYx.2009–2023.epw file, which has hourly data for one full year for outdoor temperature, humidity, solar radiation, and wind (One Building, n.d.). This one was chosen because it is the most current typical meteorological year accessible for Istanbul and dictates weather conditions from 2009 to 2023. Simulations are made more accurate and contextually relevant by utilizing weather data from the very same geographic area where survey data were gathered.

3.5.3. Python Automation and the Use of the Eppy Library

Running hundreds of simulations manually would have been time-consuming and error-prone. To solve this, a Python automation script was created using the Eppy library. This library was selected because it allows direct reading and editing of EnergyPlus model files using Python. In this study, it was especially useful for automatically inserting the correct occupant schedule into each simulation run. By doing this through code, the process became much faster, more accurate, and fully repeatable.

The script identified all generated behavior schedule files, each representing a full year of activity for one apartment. For each file, two simulations were launched one using the High Building IDF and one using the Low Building IDF. Before each run, the IDF file was copied and updated with the correct behavioral input. Then, EnergyPlus was executed with the Istanbul weather file and the necessary simulation settings. Output reports were saved in clearly labeled folders, and summary HTML files were moved to a central location for easy access.

Although the simulations generated several types of outputs, this study focused mainly on one key result: The Utility Use Per Conditioned Floor Area (kWh/m²), which gives a clear picture of how much energy is consumed relative to the size of the building. This value was extracted from each summary report and used to compare simulation outcomes.

To analyze these results, an additional Python script was developed to automate the post-processing stage. This script scanned the summary reports for each of the 50 simulations and extracted the relevant table. It then compiled these values and generated histograms, showing how energy use varied across all behavior scenarios and between the two building models. These visual results are presented and discussed in the following chapter.

3.6. DETERMINISTIC OCCUPANT SCHEDULE IMPLEMENTATION

To supplement the probabilistic modeling framework and have a baseline for comparison, a deterministic occupant schedule was formulated and incorporated into the simulation procedure. The schedule was built based on constant values from literature and not stochastic or survey information. More specifically, it aligned with the daily routine pattern described in a study by Yiğit (2021), which provides a basic residential occupancy model that has been widely applied for surrogate modeling in Turkish households (Yiğit, 2021).

The deterministic approach made a four-person family assumption based on a predetermined pattern whereby time slots were allocated for sleeping, housework, relaxation, and other in-home activities. The time from 00:00 through 07:00 was allocated as sleeping time with 100% occupancy. One of the four people, or 25% occupancy, was considered available in the home for housework from 07:30 through 12:30. The same was applied for other time slots. The deterministic activity slots and the occupation rates based on the study mentioned are given in Table 3.1 below.

These time blocks were also converted into numerically equivalent occupancy fractions and other scheduling inputs for use in EnergyPlus. The day was split into 96 segments of 15-minute intervals, which were averaged into hourly values by summing every four consecutive entries.

Occupancy scheduling modeled the percentage of residents in the house for each hour based directly on the activity timeline from literature. The electric appliance usage scheduling was arrived at under an assumption that residents operated electrical appliances whenever they were active and not just sleeping or quietly sitting. On the same basis as in the probabilistic model, people operating electrical appliances per hour were divided by the number of electrical appliances in consideration for the study. Lastly, one was always added in consideration of the refrigerator being in continuous use.

Table 3.1 Daily activity timeline and occupancy assumptions for deterministic schedule (Yiğit 2021).

Time Range	Activity Description	Assumed Occupancy Fraction
00:00–07:00	Sleeping	1.00
07:00–07:30	Breakfast	1.00
07:30–12:30	Housework	0.25
12:30–15:30	Resting	0.25
15:30–16:30	Housework	0.25
16:30–19:00	Housework and resting	0.75
19:00–20:00	Housework and resting	1.00
20:00–20:30	Dinner	1.00
20:30–23:00	Sitting quietly	1.00
23:00–24:00	Sleeping	1.00

Lighting schedules were designed under the premise that whenever any individual was in the building and not asleep, lights were on. The same logic of normalized scaling was applied for its correspondence with the balance of the rest of the methodology. The hot water demand in households was modeled in order to capture usual showering and washing habits in which demand for hot water was high in morning and evening hours in correspondence with literature specifications. The heating and cooling set points were fixed at 20°C and 24°C throughout the day similar to what was carried out during probabilistic modeling. The infiltration and ventilation schedules were set based on occupant pattern in which ventilation functioned while people were inside and infiltration was treated as constant. The deterministic schedule spanned one day in hourly resolution and was repeated in order to run through all 8760 hours of the year using an identical daily pattern. The layout guaranteed a repeatable and stable profile that was directly comparable against the behaviorally diverse results of the probabilistic runs. The deterministic schedule was imported into an identical EnergyPlus simulation environment utilizing the same building model and weather file except for occupant-schedule differences. The High and Low template building models were simulated in order to determine energy use for this stationary behavioral event. The results were then compared and visualized against those from the probabilistic runs in order to determine the impact of scheduling variability on building energy performance.

CHAPTER 4

4. RESULTS AND DISCUSSION

4.1. PROBABILISTIC SIMULATION OUTPUT

The simulation output based on probabilities, achieved by running 50 randomly generated occupancy schedules for both high and low usage scenarios, illustrates the significant impact that people's behavior has on overall and category-based energy consumption. The variability throughout simulations confirms what is commonly known, that even with identical building conditions, various usage patterns and daily activities could demand a remarkably different level of energy.

4.1.1. High Scenario

Under the high-rise building case, overall utility usage (HVAC, electricity, and others) varied from about 63.9 to 76.1 kWh/m². As (Figure 4.1) illustrates, it is fairly normal, with a high concentration of runs around 70–72 kWh/m². This would mean that the majority of occupancy patterns for high-rise building case scenarios lean towards driving overall usage toward the upper-middle tier of outcomes.

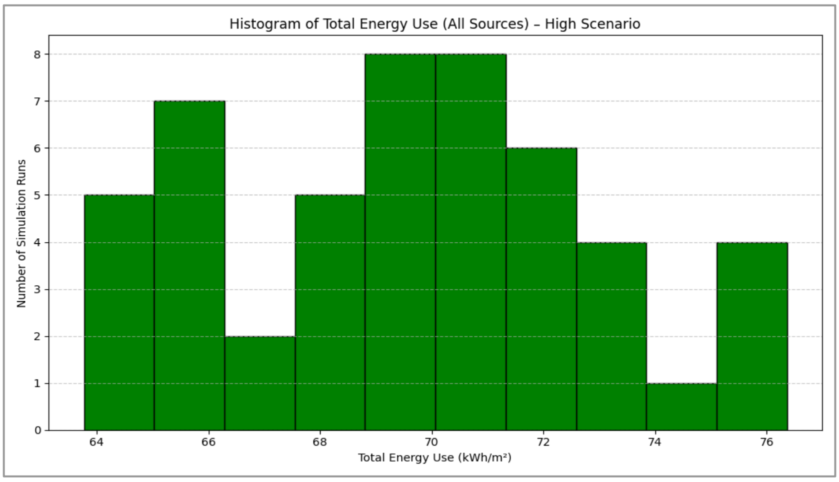

Figure 4.1 Total Utility Energy Use (All Sources) – High-Rise Building

Where only electricity consumption is considered, the values varied from 31.7 to 42.5 kWh/m², with a majority of values being 36–39 kWh/m² (Figure 4.2). The use of electricity for lighting, however, varied from 13.1 to 19.8 kWh/m² with a peak concentration at 16–18 kWh/m² (Figure 4.3). This is an indication of a high sensitivity of lighting demand to occupancy presence and usage type.

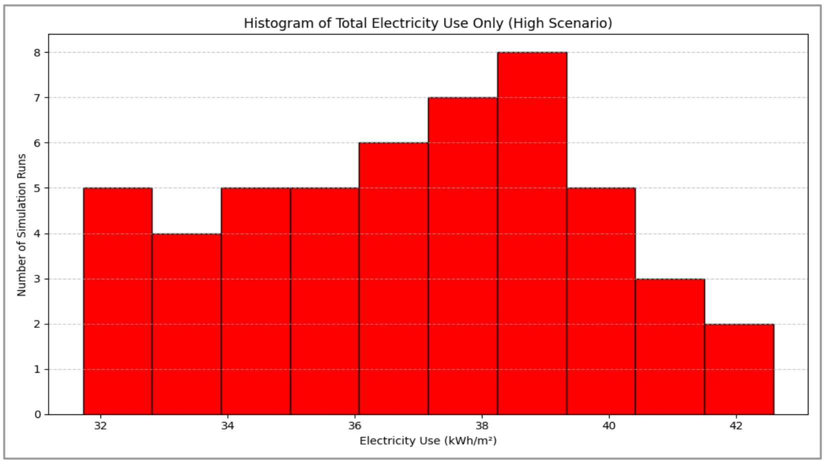

Figure 4.2 Total Electricity Use Only – High-Rise Building

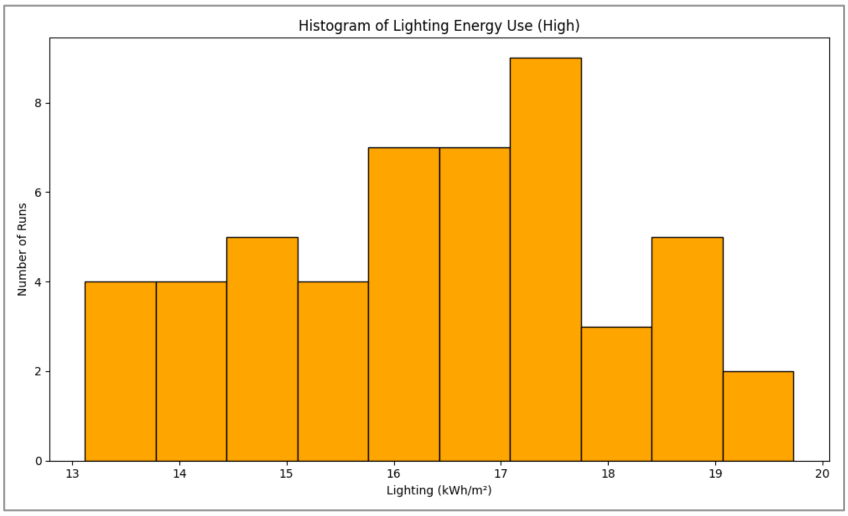

Figure 4.3 Lighting Energy Use – High-Rise Building

The energy usage of the HVAC had a smaller spread, with a range of 29.6–32.1 kWh/m² with a peak frequency at 30.5–31.5 kWh/m² (Figure 4.4). This relatively stable range of HVAC values indicates that even though stochastic behavior influences heating and cooling, its variability is less compared to plug end uses.

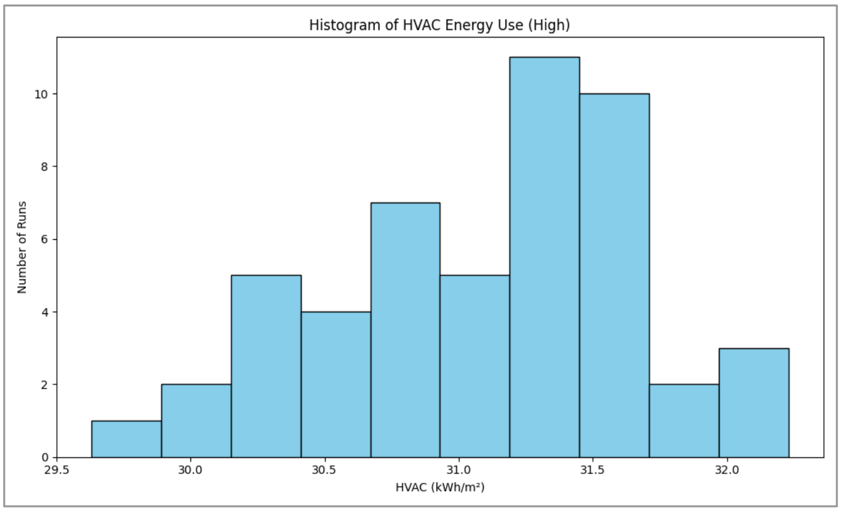

Figure 4.4 HVAC Energy Use – High-Rise Building

For "Other" electric usage appliances and domestic hot water values varied between 19.4 and 25.3 kWh/m² (Figure 4.5). This category showed the broadest range, illustrating the variability of activities, including cooking, showering, or using electronic devices, that impact miscellaneous end-use loads.

(Figure 4.6) is a stacked bar chart of the contribution of lighting, HVAC, and others towards the overall consumption over all 50 stochastic runs for the high scenario. Even while HVAC is fairly stable, there's considerable variability within the category 'Other', further illustrating stochastic occupant behavior's influence on plug load consumption. The graphical breakdown places what is seen within the histogram ranges into perspective and highlights the cumulative influence of small variations over several categories.

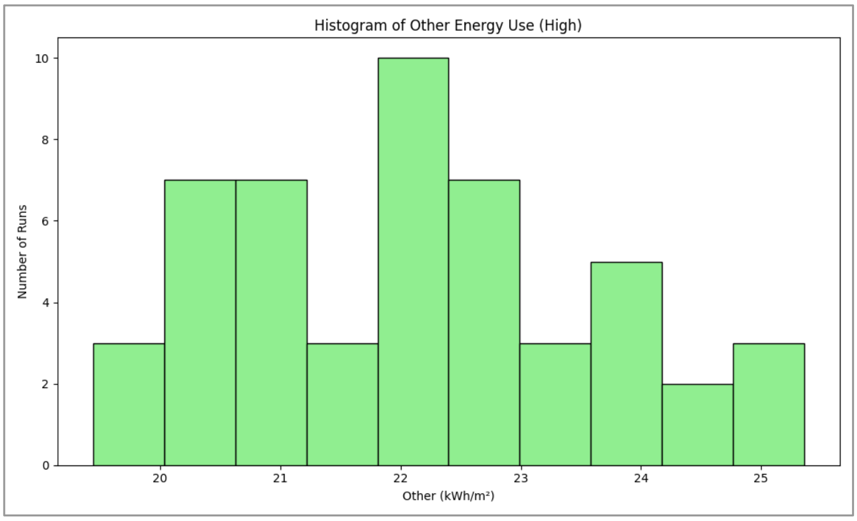

Figure 4.5 Other Energy Use (Appliances + DHW) – High-Rise Building

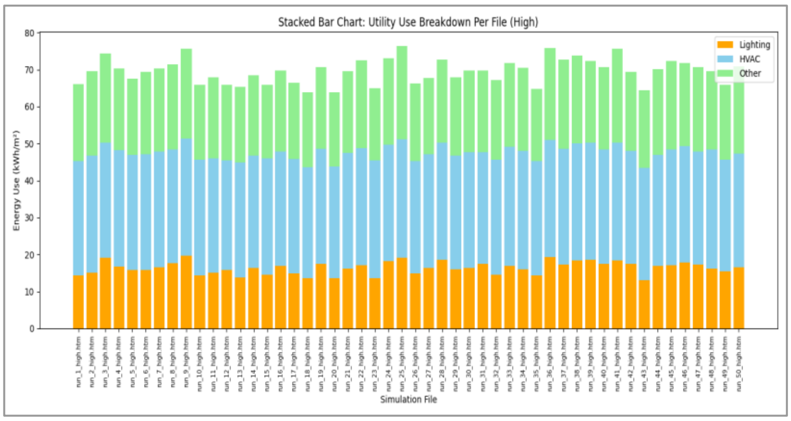

Figure 4.6 Stacked Bar Chart – Utility Use Breakdown Per File (High)

4.1.2. Low Scenario

The low-rise building case also showed a wide spread of outcomes. The overall use of utility energy varied from 66.3 to 79.5 kWh/m², which surprisingly actually exceeded that of the high-rise building case because of stochastic variability (Figure 4.7). The counterintuitive finding is a characteristic of stochastic modeling's internal variability, by which even based on "low" profiles, occupant schedules can at times generate a greater usage pattern because of clustering of active activities.

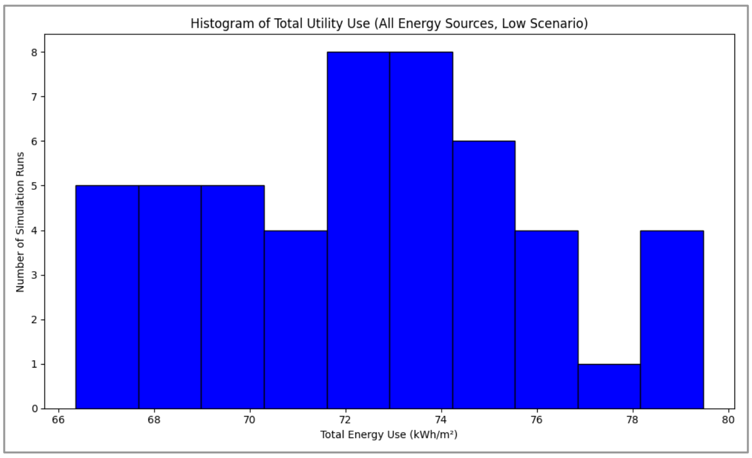

Figure 4.7 Total Utility Energy Use (All Sources) – Low-Rise Building

Electricity usage by itself, even for the low-rise building case, turned out to be within a comparable general range to that of the high case, with a range of 31.7 to 42.5 kWh/m² (Figure 4.8). Lighting use of energy also had a comparable distribution, with a range of 13.1 to 19.8 kWh/m² and with values peaking at 17 kWh/m² (Figure 4.9). These findings once again confirm that lighting is strongly sensitive to presence timing and presence duration, independent of a given scenario type.

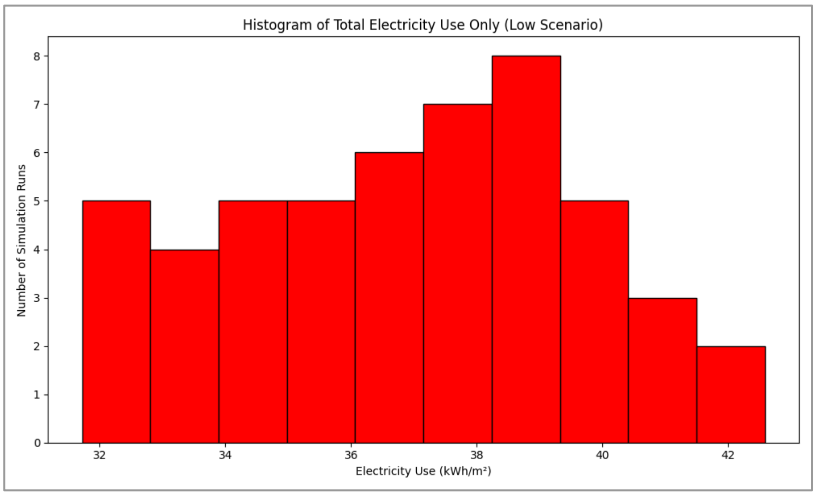

Figure 4.8 Total Electricity Use Only – Low-Rise Building

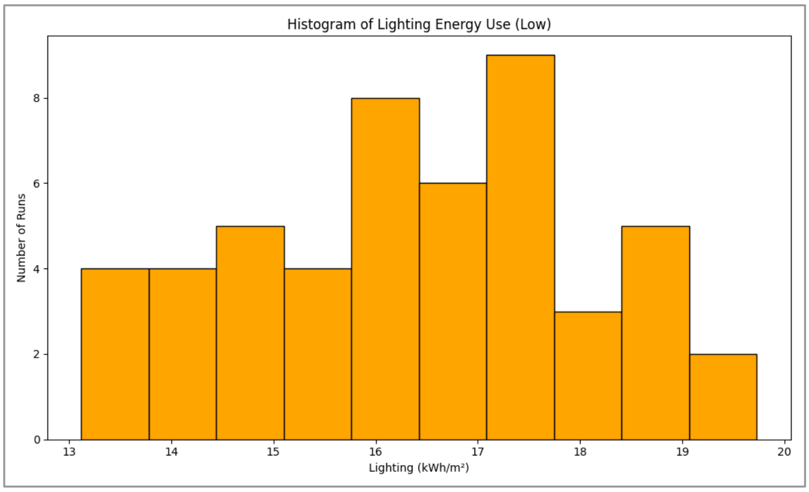

Figure 4.9 Low Energy Use – Low-Rise Building

HVAC usage had an overall slightly greater trend than for the high case, with usage ranging from 31.9 to 35.3 kWh/m² (Figure 4.10). This once more points toward probabilistic influences such as coincidental indoor occupancy during peak heating or cooling periods contributing greater demand even with otherwise conservative estimates.

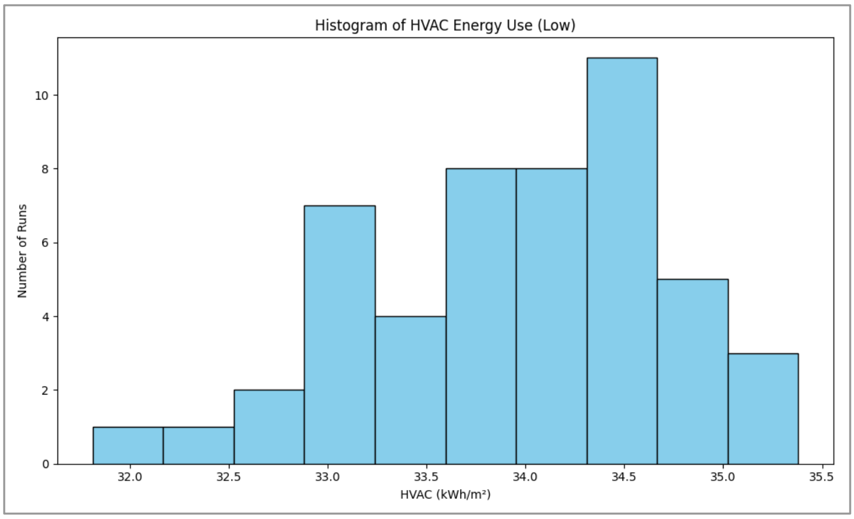

Figure 4.10 HVAC Energy Use – Low-Rise Building

Last, appliance and hot water usage, which are other uses for electricity varied from 19.4 to 25.3 kWh/m² (Figure 4.11). This is consistent with variability of the high-demand case and reinforces the significant impact of device operation by users on overall building energy demand.

(Figure 4.12) is the low-scenario stacked bar chart, which provides a breakdown over the 50 runs. As with the high scenario, there are consistent values for HVAC while there are variations for lights and others. This comparison supports that there is non-trivial variation introduced by occupant behavior on energy use profiles, especially for those categories that are based on individual and household-level behaviors.

Overall, the stochastic findings for high and low scenarios affirm the impact of building occupants on performance. Even with the identical building, envelope, and system assumptions, variations in everyday habits result in large differences in lighting, HVAC, and plug load demand affirming the importance of stochastic methods for contemporary energy modeling.

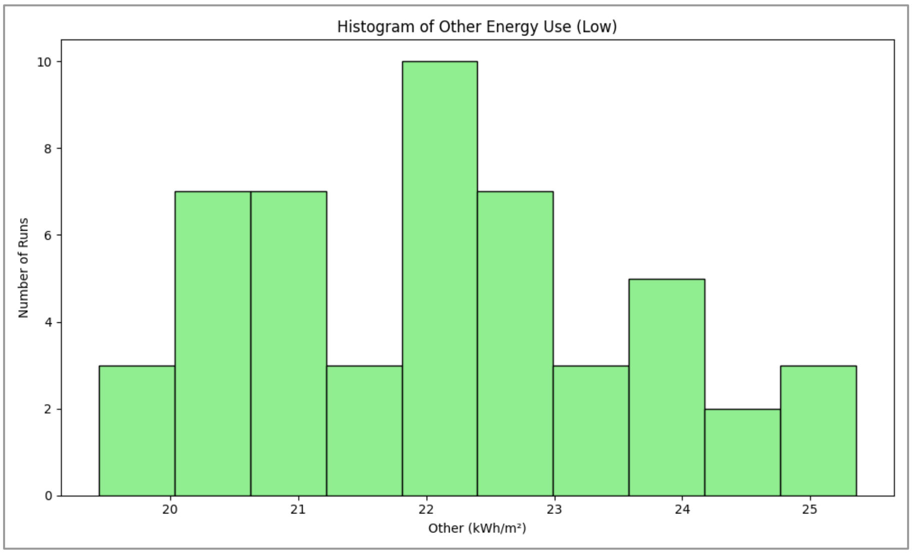

Figure 4.11 Other Energy Use (Appliances + DHW) – Low-Rise Building

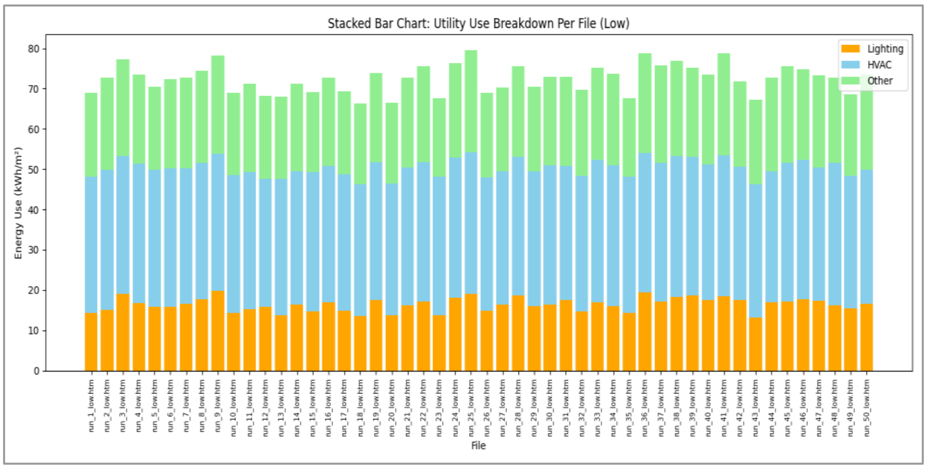

Figure 4.12 Stacked Bar Chart – Utility Use Breakdown Per File (Low)

4.2. DETERMINISTIC SIMULATION RESULTS

This section presents the results of the deterministic simulations conducted using fixed occupant behavior patterns derived from literature-based daily activity schedules. A single set of deterministic occupancy schedules was used for both high and low energy scenarios, replicated across all 365 days of the year. The results reflect the total and component energy demands for lighting, HVAC, and other electricity uses.

occupant schedule was used in both cases, the building templates used for the High and Low scenarios produced different energy consumption outcomes. The deterministic results, shown in Table 4.1, illustrate the energy performance outcomes of a building modeled under a static, average behavior profile. These values serve as a reference baseline to be compared against the probabilistic cases in the following section.

Table 4.1 Deterministic schedule-based simulation results for High-Rise and Low-Rise Building Scenarios

Energy Category	High-Rise Building (kWh/m²)	Low-Rise Building (kWh/m²)
Lighting	15.8	15.8
HVAC	36.4	33.6
Other Uses	22.0	22.0
Total Electricity Use	34.4	34.4
Total Utility Use	74.2	71.4

4.3. COMPARISON BETWEEN DETERMINISTIC AND PROBABILISTIC SIMULATION RESULTS

This section presents an extensive comparison of simulation output generated based on deterministic and probabilistic occupant behavior models. The deterministic model uses a set fixed daily routine developed from literature-related schedules and the probabilistic model infuses behavior variability with 50 randomized runs across 1000 daily activity schedules drawn from 170 original survey responses. Through the comparative analysis, the manner in which the selection of the occupant modeling technique affects building energy performance outcomes across the primary end-use sectors is showcased.

4.3.1. Lighting Energy Use

In the deterministic case, lighting demand remained constant at 15.8 kWh/m² for both high and low scenarios. For the probabilistic simulations, however, lighting energy consumption fluctuated between 13.2 and 19.7 kWh/m² for high and low scenarios. This spread accounts for the high sensitivity of lighting demand on occupant usage patterns, time-of-day activities, and indoor vs. outdoor occupancy behavior. Though the determinist value is close to the probabilistic range mean value, it does not capture such behavioral edge cases as early risers and low lighting usage from habits or preference-driven factors. The probabilistic method does capture variability correctly for seasonal routines, number of people, and irregular daily routines.

4.3.2. HVAC Energy Consumption

In deterministic simulations, HVAC energy use was 36.4 kWh/m² for the high scenario and 33.6 kWh/m² for the low scenario. The probabilistic model produced HVAC consumption ranging from 29.6 to 32.1 kWh/m² (high) and 32.0 to 35.3 kWh/m² (low). Interestingly, the deterministic values were consistently above the upper bounds of the probabilistic ranges. This indicates that the fixed schedule in deterministic modeling tends to overestimate heating and cooling demands by assuming continuous occupancy and full HVAC usage throughout the day. On the other hand, probabilistic models consider factors such as time spent outside the home, partial or adaptive HVAC use, and behavioral adjustments (e.g., changing clothing or using fans), all of which reduce thermal loads. This divergence highlights a key limitation of assuming static thermal demand.

4.3.3. Other Electricity Use

The 'Other' category is the miscellaneous end-uses such as appliance usage, hot water usage for household purposes (i.e., bathing or washing), and electronic entertainment. The deterministic outcome for both the scenarios remained 22.0 kWh/m². In contrast, the probabilistic results ranged from 19.4 to 25.3 kWh/m² in the high scenario and 19.3 to 25.3 kWh/m² in the low scenario. This broader range reflects occupant diversity in both activity types and timing such as clustered usage during morning or evening routines. Deterministic modeling compresses this complexity into a single averaged profile, which overlooks simultaneous use among household members and inactive periods.

4.3.4. Total Electricity Use

Total electricity consumption (including lighting and other plug loads) was 34.4 kWh/m² in both deterministic scenarios. Probabilistic results ranged from 31.7 to 42.5 kWh/m² (high) and 31.8 to 42.5 kWh/m² (low), indicating significant behavioral variability. Some profiles depicted energy-conscious lifestyles with minimal plug load and lighting use, while others reflected extended indoor presence and greater use of electrical devices. The deterministic model provides a central estimate but does not account for behavioral peaks, demand clustering, or occupancy-driven unpredictability factors that can influence both system sizing and grid load profiles.

4.3.5. Total Utility Energy Consumption

When aggregating all energy sources, deterministic total utility use reached 74.2 kWh/m² in the high scenario and 71.4 kWh/m² in the low scenario. The probabilistic ranges were 63.9 to 76.1 kWh/m² for high, and 66.3 to 79.5 kWh/m² for low. These results demonstrate that occupant-driven variability can push total energy demand above or below what a single deterministic simulation might predict. Even in an energy-efficient building, factors such as activity timing, occupant absences, and load clustering influence total demand considerably.

A consolidated comparison of deterministic values versus the probabilistic ranges across all key energy categories is presented in Table 4.2.

Table 4.2 Comparison of Deterministic and Probabilistic Simulation Results

Energy Category	Deterministic (High)	Probabilistic Range (High)	Deterministic (Low)	Probabilistic Range (Low)
Lighting	15.8	13.1 – 19.8	15.8	13.1 – 19.8
HVAC	36.4	29.6 – 32.1	33.6	31.9 – 35.3
Other Uses	22.0	19.4 – 25.3	22.0	19.4 – 25.3
Total Electricity Use	34.4	31.7 – 42.5	34.4	31.7 – 42.5
Total Utility Use	74.2	63.9 – 76.1	71.4	66.3 – 79.5

4.3.6. Stacked Energy Breakdown Visualization

Figure 4.6 presents a stacked bar chart of the lighting, HVAC, and other loads split across the 50 probabilistic simulations of the high scenario. HVAC is fairly consistent, but lighting and particularly other loads exhibit great variability due to varied occupancy schedule and usage patterns.

Likewise, Figure 4.7 shows the same plot for the low scenario. Even though the levels of energy are generally lower across the board, the same pattern holds: HVAC has less variability because of thermostatic control and fixed set points, but plug and hot water loads vary considerably. Combined, these graphs highlight that not all categories of energy respond equally to behavior some are considerably more influenced by human variation than others.

4.3.7. Discussion and Implications

The comparison shows a clear and consistent pattern: deterministic outputs bunch around averages of probabilistic distributions but miss the wider range of behavioral diversity. The insensitivity is precisely the problem that arises in applications in which extremes matter system sizing, estimation of peak demand, demand-response planning, resiliency modeling. Although deterministic models offer interpretability and transparency, they lack crucial edge cases and behavioral dynamics.

Conversely, probabilistic modeling, being more computation-intensive, presents a more realistic and informative account of true building occupancy. It is probabilistic and accommodates uncertainty, describes a range of behavioral profiles, and best accounts for the risk and variability that influence energy performance in the real world. Such behavioral uncertainty is required for energy-efficient building design objectives, retrofit optimization, or policy evaluation.

Overall, probabilistic modeling presents a more realistic and richer description of how occupant behavior influences energy consumption in residential buildings. Probabilistic methods accommodate the variability, randomness, and diversity of the way people occupy, utilize, and engage with building systems in real life, rather than using fixed daily patterns as in deterministic approaches. More realistic probabilistic models become increasingly suitable for applications involving high behavioral uncertainty, system precision for sizing, or testing policies under a broad variety of conditions.

Despite the benefit of offering a good reference or baseline against which more realistic loads might compare, deterministic schedules lack consideration of edge cases such as infrequent occupiers, irregular workers, and households that have active evening appliance use. Although these edge cases occur infrequently, they have a disproportionate influence on peak demand, energy performance scores, and resilience in extreme situations (i.e., heatwaves, occupancy shifts, or demand response events).

The evidence shows that even with the same building and system assumptions, stochasticity will yield ranges of utility consumption broad enough to affect building decision-making, comfort predictions, and carbon assessments. Thus, for simulations used to inform real decisions, develop policies, or optimize retrofit measures, probabilistic occupant models should not merely be helpful but necessary.

CONCLUSION AND SUGGESTIONS

In this thesis we aimed to determine the way in which probabilistic, behavior-driven occupant schedules enhance the reality and precision of residential building energy simulations within Türkiye. Through the creation of a novel dataset of 170 rich daily activity surveys and converting them into 1000 high-resolution occupancy schedules, the research effectively illustrated the extent to which variation in occupant behavior drives energy consumption trends in the categories of lighting, HVAC, and plug loads.

By comparative simulation with EnergyPlus, both probabilistic and deterministic models were run within two energy performance scenarios, a High and a Low template building. While deterministic schedules offered an uncomplicated, repeatable baseline, they never fully captured the range of day-to-day variation. Compared with the probabilistic schedules based on real behavior data, these picked up the temporal variability, activity grouping, and cultural context of Turkish occupant lifestyles.

The outputs found that deterministic simulations tended toward the mean of probabilistic outputs but failed to capture significant peaks and troughs in energy use due to occupant behavior variation, both high and low that might have significant implications on system sizing, grid planning, and comfort modeling. For example, plug loads and lighting varied more widely under probabilistic modeling as a result of the variability of home presence and usage patterns, but HVAC loads proved more stable but not unaffected by subtle shifts in occupancy.

The conclusions remind that energy performance is not just a function of envelope and equipment selection but is strongly influenced by how people live and utilize spaces. It is therefore imperative that this thesis emphasizes the need for a transition towards occupant-based modeling practices that account for cultural habits and temporal irregularity, particularly in situations when default profiles that have been imported from other regions cannot capture local conditions.

In addition, the thesis offered a reproducible process for creating and incorporating stochastic schedules into EnergyPlus, with the assistance of Python automation and importing schedules using CSVs. This process not only applies to other regions and building qualities but is also scalable for uses in subsequent studies aimed at testing policies, loads forecasting, or the development of behavioral interventions.

In summary, this work adds a culturally responsive, evidence-based framework for probabilistic occupant scheduling in residential building simulation. It presents a convincing argument for the wider acceptance of stochastic approaches in simulation practice not simply as a matter of academic rigor but as a way of capturing the way that people actually live. With building design more and more involved with occupant health, energy consumption, and climate sustainability, the inclusion of real human behavior in the core of simulation is not simply an enhancement, it is a requirement.

Future research can build upon this thesis by expanding the dataset to include seasonal variation and regional differences across Türkiye, which may further enhance the representativeness of probabilistic schedules. Additionally, integrating occupant comfort feedback or real-time sensor data could refine behavior modeling and increase simulation accuracy. Researchers are also encouraged to explore the application of this approach to other building types, such as commercial or educational facilities, to evaluate the scalability of behavior-based modeling across the built environment. Finally, policymakers and design practitioners should consider adopting culturally-aware occupant models as part of building codes and energy policy development to bridge the gap between real-world usage and modeled performance.

REFERENCES

Aerts, D., Minnen, J., Glorieux, I., Wouters, I., & Descamps, F. (2014). A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. *Building and Environment, 75*, 67–78. https://doi.org/10.1016/j.buildenv.2014.01.021

Ahmed, O., Sezer, N., Ouf, M., Wang, L., & Hassan, I. G. (2023). State-of-the-art review of occupant behavior modeling and implementation in building performance simulation. *Renewable and Sustainable Energy Reviews, 185*, 113558. https://doi.org/10.1016/j.rser.2023.113558

Al-Saegh, S., Tahmasebi, F., Tang, R., & Mumovic, D. (2024). Comparison of deterministic, stochastic, and energy-data-driven occupancy models for building stock energy simulation. *Buildings, 14*(9), 2933. https://doi.org/10.3390/buildings14092933

Cecconi, F. R., Manfren, M., Tagliabue, L. C., Ciribini, A. L. C., & De Angelis, E. (2017). Probabilistic behavioral modeling in building performance simulation: A Monte Carlo approach. *Energy and Buildings, 148*, 128–141. https://doi.org/10.1016/j.enbuild.2017.05.013

Chen, J., Adhikari, R., Wilson, E., Robertson, J., Fontanini, A., Polly, B., & Olawale, O. (2022). Stochastic simulation of occupant-driven energy use in a bottom-up residential building stock model. *Applied Energy, 325*, 119890. https://doi.org/10.1016/j.apenergy.2022.119890

Chen, S., Zhang, G., Xia, X., Chen, Y., Setunge, S., & Shi, L. (2022). The impacts of occupant behavior on building energy consumption: A review. *Sustainable Energy Technologies and Assessments, 45*, 101212. https://doi.org/10.1016/j.seta.2021.101212

Dino, I. G., & Akgül, C. M. (2019). Impact of climate change on the existing residential building stock in Turkey: An analysis on energy use, greenhouse gas emissions and occupant comfort. *Renewable Energy, 141*, 828–846. https://doi.org/10.1016/j.renene.2019.03.150

Glasgo, B., Hendrickson, C., & Azevedo, I. L. (2017). Assessing the value of information in residential building simulation: Comparing simulated and actual building loads at the circuit level. *Applied Energy, 203*, 348–363. https://doi.org/10.1016/j.apenergy.2017.05.164

He, M., Lee, T., Taylor, S., Firth, S. K., & Lomas, K. (2015). Coupling a stochastic occupancy model to EnergyPlus to predict hourly thermal demand of a neighbourhood. In *Proceedings of BS2015: 14th Conference of the International Building Performance Simulation Association* (2101–2108), Hyderabad, India, Dec. 7–9.

He, Y., Chen, Y., Chen, Z., Deng, Z., & Yuan, Y. (2022). Impacts of occupant behavior on building energy consumption and energy savings analysis of upgrading ASHRAE 90.1 energy efficiency standards. *Buildings, 12*(8), 1108. https://doi.org/10.3390/buildings12081108

Hopkins, A. S., Lekov, A., Lutz, J., Rosenquist, G., & Gu, L. (2011). Simulating a nationally representative housing sample using EnergyPlus. *Lawrence Berkeley National Laboratory*. https://eta-publications.lbl.gov/sites/default/files/lbnl-5079e.pdf

International Energy Agency. (n.d.). *Buildings – Energy System*. Retrieved May 4, 2025, from https://www.iea.org/energy-system/buildings

International Organization for Standardization. (2005). *ISO 7730:2005 – Ergonomics of the thermal environment: Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria*. International Organization for Standardization.

Kim, S., & Park, C. S. (2025). Quantification of aleatoric and epistemic uncertainty in data-driven occupant behavior model for building performance simulation. *Energy and Buildings*. Advance online publication. https://doi.org/10.1016/j.enbuild.2025.115818

Muslim, S. A. (2021). EnergyPlus—Towards the selection of right simulation tool for building energy and power systems research. *Journal of Energy and Power Technology*, *3*(3). https://doi.org/10.21926/jept.2103034

One Building. (n.d.). *TUR_IB_Istanbul-Ataturk.AP.170600_TMYx.2009–2023 [EPW file]*. Climate.OneBuilding.org. Retrieved April 15, 2025, from https://climate.onebuilding.org/

Richardson, I., Thomson, M., & Infield, D. (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, *40*(8), 1560–1566. https://doi.org/10.1016/j.enbuild.2008.02.006

Richardson, I., Thomson, M., Infield, D., & Clifford, C. (2010). Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, *42*(10), 1878–1887. https://doi.org/10.1016/j.enbuild.2010.05.023

Turkish Standards Institute. (2024). *TS 825:2024 – Thermal insulation requirements for buildings (Binalarda Isı Yalıtım Kuralları)*. Turkish Standards Institute.

Widén, J., & Wäckelgård, E. (2010). A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy, 87*(6), 1880–1892. https://doi.org/10.1016/j.apenergy.2009.11.006

Widén, J., Lundh, M., Vassileva, I., Dahlquist, E., Ellegård, K., & Wäckelgård, E. (2009). Constructing load profiles for household electricity and hot water from time-use data—Modelling approach and validation. *Energy and Buildings*, *41*(7), 753–768. https://doi.org/10.1016/j.enbuild.2009.02.013

Yiğit, A. (2021). Surrogate model-based multi-objective optimization of residential building envelopes: A case study for Istanbul. *Sustainable Energy Technologies and Assessments*, *47*, 101512. https://doi.org/10.1016/j.seta.2021.101512

CURRICULUM VITAE