

# An Hourglass Approach to Urban Energy Computing

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

**Doktor/in der technischen Wissenschaften**

by

**Neda Ghiassi**

Registration Number 0928685

to the Faculty of Informatics  
at the Vienna University of Technology

Advisor: Univ.Prof. Dipl.-Ing. Dr. techn. Ardeshir Mahdavi

The dissertation has been reviewed by:

---

(Prof. Dirk Saelens)

---

(Prof. Jessen Page)

Wien, 25.10.2017

---

(Neda Ghiassi)



## **Erklärung zur Verfassung der Arbeit**

Neda Ghiassi

Währinger Straße 51/9, 1090 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder im Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

-----

(Ort, Datum)

-----

(Unterschrift Verfasserin)

*„... I am a part of all that I have met;  
Yet all experience is an arch wherethro'  
Gleams that untravell'd world whose margin fades  
For ever and forever when I move...”*

Alfred, Lord Tennyson, 1842, Ulysses



## Acknowledgments

My eternal gratitude goes out to

my teacher, my mentor, my friend, professor Mahdavi, who knows the best stories;  
the most courageous woman and the most honorable man in my life, mom and dad, to whom I owe all that I am and all that I will ever be;

my center of gravity and my partner in crime, Sam, who knows me and makes me laugh and drives me crazy and keeps me sane;

my best friends, my playmates, my people, Nassim and Kaveh, who are the air that I breathe and the ground that I stand on;

those, who give my world scent and color and glitter and glory, Aghaye Ahani, Kobra jun, Ladan, Ryan, khaleh Massoum, khaleh Fati, Sina, Kimia, Sepideh, Aghaye Vahid, and Kasra;

and those, without whom something would always be missing, Sahari, Samira, Behtash, Koosha, Nazanin, Amoo Amir, Mahnameh, Mehrad, Carmin, Pouyan, and Shayan.

This project was partially funded by the Environmental Informatics Doctoral College. I would like to express my sincere thanks to professor Schahram Dustdar, and professor Ardeshir Mahdavi for the opportunity.

I would like to gratefully acknowledge the contributions of several colleagues to the present research effort. These include Stefan Glawischnig, Kristopher Hammerberg, Ulrich Pont, Christian Steineder, Owat Sunanta, Mahnameh Taheri, Farhang Tahmasebi, and Milena Vuckovic.

I am truly grateful to Renate Weiss, Barbara Wiesböck, the adorable Josef Lechleitner, and the one and only Elisabeth Finz for their sincere support.

And last but not least, I heartily thank my examiners professor Schahram Dustdar, professor Jessen Page, professor Helmut Rechberger, and professor Dirk Saelens, who accepted to dedicate their precious time to evaluating this work.

## Abstract

The development of effective strategies to improve the energy performance of the built environment depends on reliable data on the spatial and temporal distribution of energy demand and supply. As such, the interest in the urban energy computing has been steadily increasing. However, in most efforts, the informational and computational challenges have led to the adoption of simplified computational routines. These models fail to capture the temporal dynamics of load patterns and their dependency on transient phenomena (occupants and climate) with appropriate resolution.

The present contribution reports on developmental activities towards generation of a bottom up urban stock heating demand model, which enables the use of Building Performance Simulation (BPS) tools for urban-level inquiries. For this purpose, a two-step method was adopted and applied to an urban instance in the city of Vienna, Austria. The first step, addresses the challenge of high informational and computational demand of BPS tools based on a systematic reduction of the extent of the required computations through sampling. Toward this end, key energy-relevant features of the buildings are used, along with a well-known datamining technique to classify the urban building stock and select representative buildings. Detailed descriptions of the selected buildings are utilized to generate detailed simulation models. Since loss of diversity is a natural consequence of any sample-based study, to recover part of the lost diversity, in a second step, a re-diversification routine was developed. This routine automatically generates permutations of the simulation models of the sample buildings, with diversified descriptions of non-geometric physical and operational building parameters. To represent operative diversity, stochastic techniques have been employed to model plausible yet diverse representations of occupants' presence and actions. The physical diversity, mainly pertaining to the thermal quality of construction components, has been treated through parametric representation of relevant material properties. As a prerequisite to the suggested method, GIS data and relevant performance assessment standards are utilized to generate an energy-relevant representation of the urban stock, which informs the two-step method. Since this framework reduces the computation domain in a first step and enhances it through the re-diversification process, the term “hourglass model” has been adopted to characterize it.

The suggested method drastically reduces the modeling effort associated with large-scale application of BPS tools through sampling. Preliminary evaluations suggest a promising accord between the predicted and the expected values of heating demand, both at aggregated and disaggregated levels.

## Kurzfassung

Die Entwicklung effektiver Strategien für die Verbesserung der Energieperformanz der bebauten Umgebung ist abhängig von verlässlichen Daten über der räumlichen und temporalen Verteilung des Energiebedarfs und -potenzials. Hierfür nimmt das Interesse an der urbanen Energieberechnung stetig zu. Allerdings, in den meisten vorherigen wissenschaftlichen Bemühungen, wegen der hohen Berechnungs- und Datenaufwände, wurden vereinfachte Rechenprozeduren adoptiert. Diese Modelle können die temporale Dynamik der Energielast und ihre Abhängigkeit von den transienten Phänomenen (z. B. Bewohner und Klima) nicht mit der passenden Auflösung erfassen. Dieser Beitrag berichtet von den Entwicklungsaktivitäten zur Schaffung eines Bottom-up Heizwärmebedarfsmodells für den urbanen Gebäudebestand, welche die Nutzung detaillierten Gebäude-Performanz-Simulationsmethoden (BPS) für energetische Untersuchungen auf urbanen Ebene ermöglicht. Zu diesem Zweck wurde eine Zwei-Schritt-Methode entwickelt und auf eine urbane Instanz in der Stadt Wien, Österreich bezogen. Der erste Schritt befasst sich mit dem hohen Daten- und Rechenaufwand der BPS-Methoden und basiert auf der systematischen Verkleinerung des erforderlichen Rechenausmaßes mittels Sampling. In dieser Hinsicht wurden entscheidende energierelevante Gebäudeeigenschaften zusammen mit wohlbekannten Data-Mining-Methoden verwendet, um den urbanen Gebäudebestand zu klassifizieren und repräsentative Gebäude auszuwählen. Detaillierte Beschreibungen der gewählten Gebäude werden benutzt, um entsprechende Simulationsmodelle zu erstellen. Da der Verlust der Diversität eine natürliche Konsequenz der Sample-basierten Studien ist, wurde zur Teilweisen zurückgewinnung der verlorenen Diversität einen Rediversifizierungsschritt entwickelt. Diese automatisierte Prozedur erstellt Permutationen der Simulationsmodelle der repräsentativen Gebäuden, mit diversifizierten nicht-geometrischen physikalischen und operativen Gebäudeeigenschaften. Zur Repräsentierung der operativen Diversität wurden stochastische Verfahren angewendet, um plausible aber auch diverse Repräsentationen der Präsenz und Aktionen der Bewohner zu erstellen. Die physikalische Diversität - hauptsächlich in Bezug auf die thermische Qualität der Konstruktionselemente - wurde durch parametrische Repräsentierung der relevanten Materialeigenschaften behandelt. Als Voraussetzung der vorgeschlagenen Methode wurde eine energetisch-relevante Repräsentation des urbanen Gebäudebestandes auf Basis GIS-Daten und relevanten Gebäude-Performanz-Evaluierungsstandards erstellt, welche die Zwei-Schritt-Methode informiert und unterstützt. Da das vorgestellte Verfahren im ersten Schritt das Rechenausmaß verkleinert und dieses im zweiten Schritt durch Rediversifizierung ausweitet, wurde den Begriff „Hourglass Model“ zur Beschreibung dessen adoptiert. Die vorgeschlagene Methode reduziert den mit der großflächigen Verwendung der BPS-Methoden verbundenen Modellierungsaufwand stark durch Sampling. Vorversuche deuten eine vielversprechende Übereinstimmung zwischen den vorausgerechneten und erwarteten Werten des Heizbedarfs an, sowohl auf Kumulierten- als auch auf Gebäudeebene.

## Contents

<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1. Motivation and Problem Statement.....	1
1.2. Research Objectives .....	6
1.3. Structure of the Work .....	9
<b>2. STATE OF THE ART .....</b>	<b>11</b>
2.1. Urban Building Energy Modeling Approaches.....	11
2.2. Computation Complexity, Modeling Extent, Resolution.....	14
2.3. Domain Reduction Criteria .....	21
<b>3. GENERAL APPROACH .....</b>	<b>25</b>
3.1. Suggested Framework .....	25
3.2. Selected Urban Instance .....	28
<b>4. THE URBAN BUILDING STOCK REPRESENTATION MODULE .....</b>	<b>30</b>
4.1. Introductory Comments .....	30
4.2. Tools and Materials .....	31
4.3. Building Stock Representation Routine .....	35
<b>5. THE REDUCTIVE MODULE .....</b>	<b>43</b>
5.1. Introductory Comments .....	43
5.2. Classification Criteria.....	44
5.3. Classification Algorithms and Tools.....	58
5.4. Clustering Scenarios .....	67
5.5. Evaluation of Classification and Sampling Scenarios .....	69
5.6. Results and Discussion .....	78
<b>6. THE RE-DIVERSIFICATION MODULE .....</b>	<b>89</b>
6.1. Introductory Comments .....	89

6.2. Tools and Material .....	90
6.3. Development of Simulation Models.....	92
6.4. Diversification Parameters .....	98
6.5. Diversification Method .....	99
6.6. Illustrative Scenarios.....	107
6.7. Results and discussion.....	109
<b>7. CONCLUSION.....</b>	<b>118</b>
7.1. Summary of Contributions.....	118
7.2. Future Research Horizons.....	125
<b>REFERENCES.....</b>	<b>127</b>
Project Related Publications.....	127
Bibliography .....	129
List of Abbreviations.....	139
List of Symbols and Units .....	140
List of Tables (Main text).....	148
List of Tables (Appendix 2).....	149
List of Figures (Main text) .....	150
List of Figures (Appendix 1) .....	153
List of Figures (Appendix 2) .....	154
List of Equations.....	155
<b>APPENDIX 1 .....</b>	<b>157</b>
<b>APPENDIX 2.....</b>	<b>191</b>



CHAPTER 1

# Introduction

## 1.1. Motivation and Problem Statement

The Paris environmental summit of 2015 concluded that global net zero emissions must be achieved before 2070 to avoid catastrophic levels of global warming (Citiscope 2015). Governments across the world have set up ambitious plans for the reduction of emissions and energy use. Urban areas, hosting now more than half of the world's population (The World Bank 2016), are considered the principle human habitat and, therefore, a core topic in the sustainability discourse. The city is an ecosystem composed of intertwined and interacting sub-systems, which may be physical or virtual, static or dynamic, predictable or complex, alive or inanimate. Buildings, climate and natural context, transport systems, media, energy production and distribution networks, financial and legal structures, inhabitants, etc. are all various subsystems forming this ecosystem. The entire system behavior is a product of the various chains of events within its different sub-systems and the interactions between them.

Buildings, as major constituents of the urban ecosystem are responsible for more than 40 percent of global energy use and a third of global greenhouse gas emissions (UNEP 2009). Moreover, the rate of growth in the building related emissions, including through use of electricity has been disconcertingly high over the past few decades. Levine et al. (2007) have documented a CO<sub>2</sub> emission growth rate of 2.5% and 1.7% per annum for commercial and residential building sectors respectively over the period of 1971 through 2004. This significant increase rate is partly due to the rapid population growth, which results in the expansion of urban areas. It is also in part attributable to economic advancements, particularly in the case of developing countries, resulting in elevated living standards and expectations. In addition to the anthropocentric aspects, changes in climate patterns, such

as the steady increase in global temperatures can result in higher energy consumption (i.e. thorough increasing use of air-conditioning) and consequently higher emissions.

On the bright side, buildings also present a high potential for energy saving and emission mitigation. According to UNEP (2009), “with proven and commercially available technologies, the energy consumption in both new and existing buildings can be cut by an estimated 30 to 80 percent with potential net profit during the building life-span”.

Various demand side energy management strategies in this regard include improvement of building envelope and systems, alteration of building operation and use patterns, and shared use of facilities. In the context of European cities where a major share of the existing building stock was built prior to the establishment of stringent energy-saving building regulations in the 1970's (e.g., 75% of the Austrian building stock (Bundesanstalt Statistik Österreich 2017)), building retrofit interventions are of particular significance. Inhabitant behavioral aspects and operation parameters are on the other hand determining factors in the conception of energy saving strategies for buildings with higher thermal quality. In this regard encouraging sustainable building operation routines either via informational campaigns or through energy pricing strategies can be attempted.

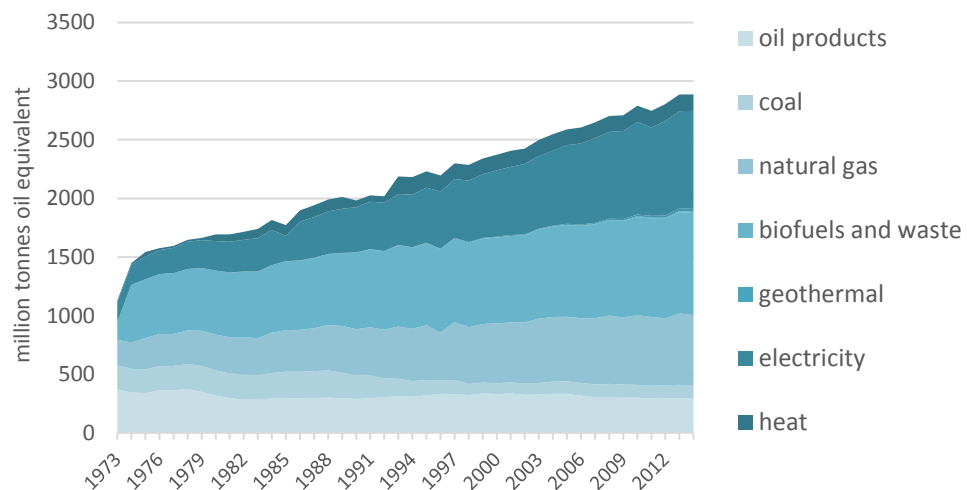
Indirect approaches such as influencing the microclimatic conditions and mitigation of the Urban Heat Island effect (e.g., through introduction of vegetation, shading elements, and waterbodies) can positively affect energy consumption patterns. Moreover, given the significance of urban morphology (e.g., street orientation and mutual shading) for energy use in buildings (Ratti et al. 2005), the design and planning of new urban structures can benefit from efficiency-oriented morphological considerations. Some of these strategies can be applied to individual buildings but others depend on an integrative consideration of buildings within their urban context. In other words, tackling energy efficiency issues at a scale beyond the scope of individual buildings (neighborhood or urban scale), presents us with solutions that are unattainable when limiting efforts to single buildings.

In line with this realization the Smart City initiative depicts urban areas not as agglomerations of independent objects, but as intertwined and connected networks of things, information, and processes. One of the key elements in the Smart City discourse is the “utilization of networked infrastructures to improve economic and political efficiency and enable social, cultural and urban development”(Hollands 2008). This initiative focuses on the utilization of Information and Communication Technologies towards enhancement and amelioration of the urban ecosystem with regard to energy efficiency among other objectives.



New supply side energy management paradigms such as Distributed Generation (DG), smart grids, and net-independent neighborhoods have emerged from this integrative perspective. In an article declaring DG as the “future power paradigm”, Manfren et al. (2011) count the “direct customer’s involvement in energy demand and peak power reduction programs” among the manifold strengths of a DG scheme. Given the increasing reliance of the building sector on electricity (see Figure 1) and the upsurge in affordable and effective small scale solar and wind energy conversion technologies, the aforementioned supply-side energy management concepts provide valuable frameworks for the exploitation of renewable energy potential. On the other hand, district heating and cooling systems may present opportunities to extend the existing urban infrastructures for more energy efficient generation and low-cost distribution of energy. However, spatially and temporally resolved energy supply and demand assessments and projections are required to evaluate the economic and technical feasibility, as well as energy and environmental benefits of various options.

The severity and urgency of the current situation and the substantial financial resources, time and effort required for the deployment of such strategies, leave no room for trial and error in urban development and regeneration activities. Determining the appropriate course of action among the wide spectrum of possible options, in view of the available resources and imminent changes, requires not only reliable and high-resolution information on the buildings' energy behavior, but also methods to predict and investigate the energy implications of various building and urban level change and intervention scenarios.



*Figure 1 Global building-related energy consumption (International Energy Agency 2014)*

Large-scale monitoring activities and smart metering are the most reliable means of building energy information retrieval. Nevertheless, the high cost of the required equipment, as well as data privacy regulations (European Parliament & European Commission 1995) hinder the pervasive realization of such schemes in the near future. Furthermore, monitored information alone provides limited prediction possibilities in view of new and unprecedented circumstances.

Wetter and Van Treeck (2017) emphasize that modeling environments are required to quantify the interactions among buildings and the grid, as well as restrictions caused by the existing urban topologies and infrastructure. Integrative urban-level decision support environments, which allow for the computational investigation and comparative analysis of the implications of various energy and emission management plans, can help insure the effectiveness of the envisaged strategies and an efficient allocation of the available resources. The computational method for the approximation of urban building stock energy behavior or the “Urban Building Energy Model” (UBEM) is the core component of such environments, determining the scope of their utility. Over the past years, efforts towards development of energy and emission models of the urban building stock for the assessment of various urban change and intervention scenarios, and their consequences have been steadily increasing. These models vary substantially in view of the general approach, scenario modeling capabilities, disaggregation level, required input parameters, and temporal and spatial resolution of the results.

In order to inform urban level strategic planning processes, the incorporated Urban Building Energy Model (UBEM) must be capable of investigating the following aspects of the urban environments and their impact on energy consumption:

- Physical and technological properties of buildings
- External boundary conditions including morphological and climatic aspects of the surroundings
- Internal boundary conditions pertaining to building operation, tendencies and actions of inhabitants

These various aspects of urban energy computing have been separately treated through former research and development efforts. Various models including global circulation models (GCMs), regional weather forecasting models, and computational fluid dynamics (CFD) models have been widely applied to generate time-domain urban microclimatic information (Wilby & Wigley 1997; Rizwan et al. 2008; Mirzaei & Haghighat 2010). Elaborate methods have been developed for the representation of occupants’ presence and actions for building performance assessment purposes including schedule-based, rule-

based, stochastic (Yan et al. 2015), and more recently agent-based occupancy models (e.g., Langevin et al. 2015). Sophisticated Building Performance Simulation (BPS) tools have been developed to allow for detailed investigations of the energy behavior of single buildings (e.g., NREL 2017).

However, an environment, which represents the mutual interactions of various influential phenomena and allows for the comparative analysis of the many-fold implications of changes and interventions on these phenomena is lacking. This is mainly attributable to the fact that detailed Building Performance Simulation (BPS) methods, which can provide the necessary versatility and flexibility in the representation of various characteristics of buildings, their boundary conditions and occupants with the appropriate resolution, are not scalable due to their high informational and computational requirements. In a detailed analysis of urban energy system models, including models pertaining to the urban building stock, Keirstead et al. (2012) point out the technical obstacles related to the complexity of the models and data uncertainty, as well as audience-oriented issues concerning model integration and policy relevance as the most prominent challenges in the field of building stock energy modeling.

Seeking to achieve a level of technical feasibility in urban-scale energy modeling, previous efforts have predominantly relied on various domain simplifications. These include, for instance, simplification of the geometry and zonal complexity of modeled buildings, use of reduced order models to represent heat transfer phenomena, and a considerable reduction of the temporal resolution of the modeling results. As a consequence, certain important queries cannot be accommodated with appropriate levels of resolution. Specifically, the temporal dynamics of load patterns and their dependency on transient phenomena (e.g., weather conditions, inhabitants' presence and actions) cannot be realistically captured. The identification of potential sites through analysis of existing customers' distribution, energy demands and load patterns is, however, considered the preliminary phase towards efficient adoption of DG in urban areas. Investigation of the energy implications of behavior change scenarios, caused by demographic changes (e.g., ageing society, elevated birth rate, etc.), or new social developments (e.g., changes in official weekly work hours, increase in part-time employment), as such, can enhance the utility of the urban energy predictive models. Moreover, in the face of the human-induced changes in climate patterns, new urban developments and existing structures can benefit from the predictive capabilities of UBEMs, which are capable of accounting for the implications of projected changes for the urban energy use patterns.

## 1.2. Research Objectives

The present research effort addresses the challenge of finding a reasonable trade-off between the modeling capabilities, the resolution of the results, the required input information, and the computation intensity in urban energy computing. The aim is to develop a computational routine for the estimation of urban building stock's energy demand, able to incorporate detailed information on physical building attributes, as well as internal and external boundary conditions, and produce highly resolved results. The current method is intended as the core energy computation component of an integrative decision support environment for the comparative evaluation of urban change and intervention scenarios pertaining to:

- Physical aspects: Thermal retrofit, densification, etc.
- Technological advancements: Installation of efficient heating systems, etc.
- Climate: Urban Heat Island studies, global warming, etc.
- Inhabitant behavior: Developments induced by demographic changes, lifestyle changes, etc.

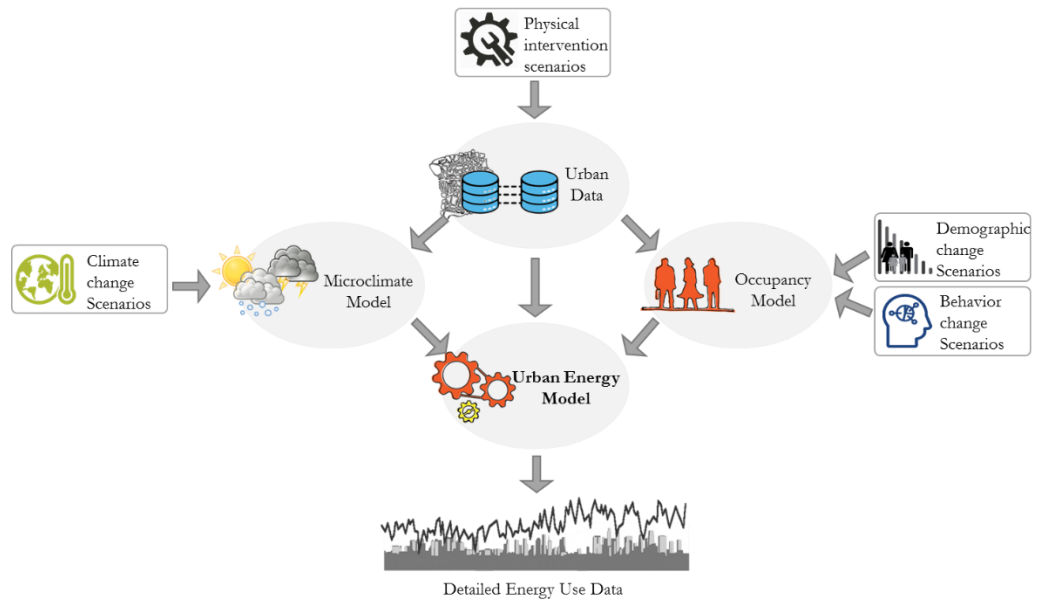
Figure 2 schematically illustrates the envisaged decision support environment. In this modular framework, the UBE is informed by available urban data, as well as detailed descriptions of internal and external boundary conditions, issued by the dedicated occupancy and microclimate models. Note that the development of these specialized modules is not the intention of the present effort. Rather, the focus is on the conception of a computational framework, which supports the incorporation of detailed and highly resolved information on occupant behavior and microclimate conditions. A computational engine that supports detailed representations of internal and external processes, and can generate spatially resolved results, allows for the reproduction of the geographical as well as temporal patterns of energy use.

For this purpose, the present contribution explores the potential of employing already established detailed numeric BPS tools, towards estimation of the energy performance of large assemblies of buildings. In this regard, various challenges must be addressed, mainly pertaining to the following aspects:

- The availability and accessibility of building information required for BPS
- Extensive time and expertise required for the development of BPS models

To address the scarcity of detailed building information required for the deployment of BPS tools, domain reduction routines based on available large-scale data, including Geographic Information Systems (GIS) data, building standards and statistical information

are considered. These methods limit the requirement for high resolution building information to a select number of instances, representing the entire computation domain. In this regard, the organization and incorporation of available information towards efficient representation of an urban domain is no trivial task. The appropriate selection of the representative instances also presents various challenges including the identification of the relevant reduction criteria, as well as a reductive method that can adapt to the dynamic nature of the urban environments. The development of a systematic and automated sampling procedure relying on major energy-relevant building characteristics is intended in the current effort. The intended reductive procedure should be flexible enough to allow for the adjustment of the size of the reduced domain in accordance with the available computational and informational resources.



*Figure 2 Conceptual framework of an integrative urban decision support environment*

A consequence of any reductive procedure, regardless of the efficiency of the method is an inevitable loss of diversity. The present work seeks to explore the possibility of recovering part of the lost diversity, through data-driven readjustment of various model variables. The evaluation of the overall significance of this diversification is also intended.

To establish clear boundaries for the research effort, detailed simulation of the hourly space heating demand (ideal demand) of the urban building stock has been considered. Space

heating is the most energy-intensive activity in both residential and commercial buildings (Figure 3 and Figure 4). Consideration of heating systems is essential for the estimation of the overall energy demand of buildings. However, due to the scarcity of the information with regard to the geographical distribution of various system types, the current effort will not address this matter.

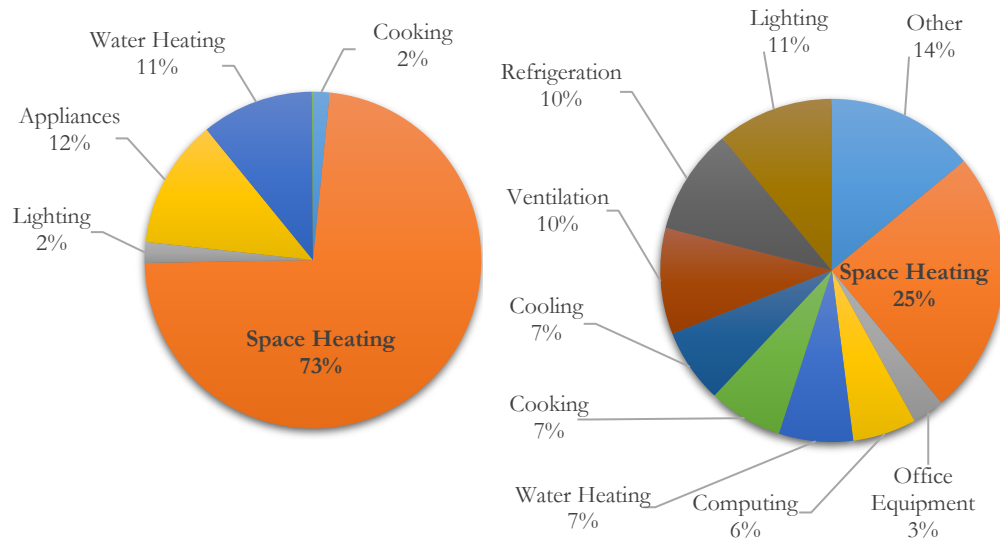


Figure 3 (Left) Residential building energy consumption by end use in Austria (based on data from Austrian Energy Agency 2009)

Figure 4 (Right) Commercial building energy consumption by end use in the United States (based on data from U.S. Energy Information Administration 2016)

For developmental activities, and to demonstrate the capabilities of the developed method, as well as the encountered challenges and the adopted solutions, an urban instance in the city of Vienna is selected. The current implementation will be as such adapted to the Austrian context. The adaptation pertains to the format and the content of the available GIS data, as well as the incorporated standards and statistical information. However, the method should be applicable to other geographical locations with minor modifications to the code, provided that the necessary input data is procured.

### **1.3. Structure of the Work**

A description of the various chapters and the entailed information is presented below:

- **Chapter 2: State of the Art**

This chapter provides an overview of the state of the art in urban energy modeling. It includes an introduction to Urban Building Energy Modeling approaches as well as a structured analysis of some former efforts in view of computational method and modeling domain. Finally, it offers an overview of the building stock classification criteria adopted in former urban energy modeling efforts for the reduction of the computation domain.

- **Chapter 3: General Approach**

This chapter offers a general description of the suggested modeling framework, its prerequisites and components: i) the urban stock representation module ii) the reductive module, and iii) the re-diversification module. To facilitate the communication of the information, an urban instance considered for the developmental activities of the project is described. The particulars of the three model components and the associated results are presented and explained in detail in dedicated chapters (Chapters 4 to 6).

- **Chapter 4: The Building Stock Representation Module**

This chapter describes the steps taken toward the generation of an energy-assessment compliant representation of the urban building stock, from available large-scale data. This step serves as a prerequisite for the informational support of the following steps. The utilized tools and materials for this purpose are listed and the developed method is explained in detail.

- **Chapter 5: The Reductive Module**

In this chapter, the reductive process leading to the identification of the representative buildings in an urban building assembly is described. Toward this end, the adopted building stock classification criteria, and the utilized algorithms and tools are explained. Various classification scenarios are considered, and a method is suggested for the evaluation of the performance of the emerging classification schemas. Lastly, the results of the application of the reductive module to the previously introduced urban instance are presented and discussed.

- **Chapter 6: The Re-diversification Module**

This chapter is dedicated to the description of the re-diversification process, aimed at restoration of part of the urban diversity lost through the reductive procedure. Primarily, an overview of the utilized tools and materials are provided. Then, the protocol followed for the generation of the reference simulation models, as well as the method developed for the generation of the diversified models is explained. Consequently, to demonstrate the impact of the re-diversification module, and the utility of the developed method towards scenario modeling, several illustrative scenarios are suggested and applied. Finally, the results of the re-diversification module as well as the outcome of the applied scenarios (on the formerly introduced urban instance) are presented and discussed in detail.

- **Chapter 7: Conclusion**

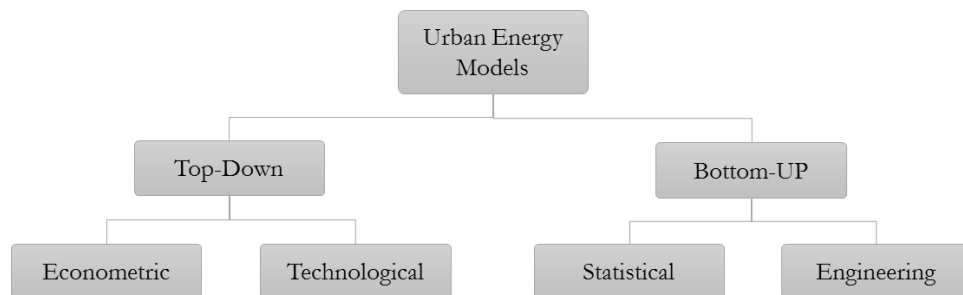
This chapter provides an overview of the developed method, sheds light on the encountered challenges and the particular findings of the research, and suggests issues requiring further investigation, as potential future research horizons.



## State of the Art

### 2.1. Urban Building Energy Modeling Approaches

Over the recent years, a variety of Urban Building Energy Models (UBEM) have been developed. These models vary considerably in terms of overall approach, informational requirements, disaggregation levels, underlying assumptions about buildings and their operation, and the type of results and scenarios they can evaluate (Kavgic et al. 2010). In an extensive review of various modeling approaches aimed at predicting the energy performance of the urban building stock, Swan and Ugursal (2009) have identified two fundamental classes of approaches to urban building stock energy modeling: the top-down and the bottom-up approaches. The terminology is selected to characterize the hierarchal position of data inputs with regard to the building sector as a whole. Figure 5 provides an overview of the different approaches to urban energy modeling.



*Figure 5 An overview of urban energy modeling approaches*

Top-down models work with aggregated national or regional energy consumption data and aim toward disaggregating this total sum based on econometric variables such as income, fuel prices, gross domestic product, or reduced technological aspects such as building type, energy standards, or simplified climatic representations. These approaches require little detail of the actual consumption processes. They treat the building sector as an energy sink and regress or apply factors that affect consumption to determine trends. Top-down approaches are very useful for supply analysis based on long-term projections of energy demand by accounting for historic response (Swan & Ugursal 2009). They can also be used to investigate the implications of a changing economy. However, they falter when encountering discontinuity for instance technological breakthroughs or severe supply shocks. In other words, due to their reliance on historical energy use patterns, top-down UBEMs have no inherent capability to model unprecedented circumstances (Kavgic et al. 2010). Moreover, detailed analysis of the share of various end uses in the overall energy demand cannot be performed through these coarse methods. Therefore, their potential to inform design or intervention decisions are very limited.

In contrast, bottom-up models calculate the energy consumption of individual buildings and then extrapolate these results to represent the energy use of the entire domain (neighborhood/ city/ region/ country). These approaches are either based on statistical methods or building physics principles (engineering techniques).

Bottom-up statistical techniques are used to determine the energy demand contribution of end-uses including behavioral aspects based on data obtained from energy bills and surveys. These techniques including regression models, conditional demand analysis models, and artificial neural networks, rely on historical consumption trends and therefore, require a substantial amount of data on various physical, operational, and contextual properties of buildings as well as the consequent energy demand. The procurement of this information, particularly the demand information, is hampered in many countries by legal barriers. The modeling capabilities offered by these methods are limited by the scope and variety of the data used to develop or train the model. As such, these models are not easily adaptable for various contexts and cannot be relied on to assess the implications of new technologies and conditions, which may not have existed during the accumulation of the underlying dataset.

Engineering (a.k.a. physical) techniques are used to explicitly calculate energy consumption of end-uses based on detailed descriptions of all or a representative set of buildings (Swan & Ugursal 2009). These techniques neglect the relationships between energy use and macroeconomic activity (Kavgic et al. 2010). They rely on building characteristics and end-uses to calculate the energy consumption based on power ratings, use characteristics and/or heat transfer and thermodynamic principles. Consequently, they have the capability of

determining the energy implications of new technologies and discontinuous changes. Moreover, they are able to explicitly address the effect of solar energy gains, which is desirable in urban morphological studies. The selection of the general approach in the development of urban energy computing methods should take the nature of the intended inquiries into account. Swan and Ugursal (2009) emphasize that to evaluate the implications of employment of new technologies, which are more likely to gain public acceptance than taxation or pricing policies, the only option is to use bottom-up engineering modeling techniques. The major drawback of engineering models is their substantial information and expertise requirement, for the generation of reliable representations of buildings, their inhabitants, and their surroundings.

## **2.2. Computation Complexity, Modeling Extent, Resolution**

The modeling capabilities of a bottom-up engineering model and the temporal resolution of the model outcome depend on the underlying building performance computation method. The adopted method determines the ability of the computational routine to represent the various building aspects contributing to the energy behavior of a building. Simplified building performance assessment tools such as standard steady-state or reduced order procedures provide rough impressions of buildings' energetic behavior, but cannot capture the transient patterns of building energy use. The more elaborate dynamic performance simulation methods enable the detailed representation of the physical and technological properties of the buildings as well as their internal and external processes. These methods can generate highly resolved and realistic data on the energy performance of the investigated buildings. However, they are associated with significant modelling effort, extensive input data, and substantial computational power requirements. In other words, the versatility and the resolution of the computation method is strongly correlated with the scope and granularity of its input information. Due to the high informational and computational demand of more elaborate performance assessment tools a fundamental challenge with regard to their urban-scale application involves the identification of a proper balance between the spatial extent of the modeling domain (i.e. the size of the urban segment considered) and the resolution or level of detail of the individually represented entities in the model. The informational and computational requirements of performance simulation increase overwhelmingly as the computation domain broadens. As such, the deployment of performance simulation methods with high-resolution representations of all relevant aspects of buildings is not practical for large assemblies of buildings. Former research efforts have tried to address this challenge in various ways. Reinhart and Cerezo Davila (2016) provide an overview of the application domain, building representation process and computational methods adopted in previous efforts. With regard to the scope of the computations, two general types of bottom up engineering models can be identified: Whole domain models and reductive models.

Whole domain models assess the energy demand of every building within the study domain. To enable such large-scale computations, these models have to rely on a coarse representation of the buildings and their various characteristics, which can be accessed through large scale data sources such as GIS and standards, using automated routines. As such, they typically incorporate reduced order or steady state standard-based computational methods, which are less demanding in terms of input information and representation complexity.

Li et al. (2015) have developed a bottom up engineering model of New York City, using a reduced-order demand computation process. The developed model uses GIS data and statistical information to arrive at the required building representation and produces monthly demand values.

Glawischnig (2016) created a framework for the web-based application of a standard steady-state demand analysis method to the Viennese building stock. The model uses building certification standards, as well as GIS data to provide a representation of buildings compliant with simplified energy assessment routines. The results are in monthly resolution. Although these models provide a useful overview of the spatial distribution of energy demand, they cannot realistically represent the temporal dynamics of load patterns and their dependency on transient phenomena (e.g. weather conditions, inhabitants' presence and actions).

Due to the significance of the temporal patterns of demand and supply potential for the reliability and efficiency of energy networks, high temporal resolution of the results seems to be a fundamental requirement for any modeling environment aimed to inform energy management decisions. In order to facilitate whole domain application of elaborate performance simulation routines, Sansregret and Millette (2009) take a different approach towards building representation. Instead of modeling each building with its authentic characteristics, to represent the Quebecois building stock, they generate synthetic simulation models based on an analysis of a large database of building information. For this purpose, general descriptors such as main building activity, floor area, location, main heating source, and construction year are used to statistically determine probability distributions of other building parameters (e.g., aspect ratio, glazing ratio, thermal properties of the envelope, etc.) from the underlying database. According to these probabilities, simulation compliant generic models of the buildings are automatically generated. Even though this method efficiently represents the urban diversity and building characteristics, due to the archetypical representation of buildings, the mutual influence of microclimate and urban morphology and their energy implications are not efficiently captured by this model. The method does not allow for the consideration of contextual parameters such as adjacency and shading relations. A reliable spatial disaggregation of demand, necessary for proximity-dependent queries (e.g., pertaining to grid optimization and integration of distributed generation plants) cannot be expected.

In a more recent but similar effort, Remmen et al. (2016) have developed a software for the automated generation of reduced order simulation models for large assemblies of buildings. The advantage of this so-called TEASER model over the previously mentioned effort is that it enables the incorporation of the available geo-referenced information, which

allows for a reproduction of the geographical distribution of energy use. It relies, however, also on predefined archetypical building descriptions and statistical data to fill in the missing information in building models. The possibility to consider multiple functionalities (in the case of multi-purpose buildings) has not been demonstrated in urban-level inquiries.

In the framework of the IEA annex 60 project (Wetter and Von Treeck 2017), several tool packages have been developed for urban and district-level building energy assessments by participating research groups. In this project, the focus has been on the development of computational tools, which can accommodate optimization purposes. As such, in the representation of individual buildings various simplifications have been permitted, for instance with regard to zoning schemes and mutual shading effects.

Other researchers have opted for detailed simulations with authentic models of the buildings in the study domain. However, in such efforts the study domain has been limited to small neighborhoods. Baetens and Saelens (2015), for instance, have developed a simulation-supported energy model of a small building assembly to investigate the impact of inhabitant behavior on large-scale energy assessments. The study domain is a synthetic neighborhood composed of buildings formerly identified to represent the Belgian building stock. However, since the computation results were not extrapolated to the entire stock, this effort can be categorized as a whole domain application. In this research, focused on inhabitant behavior, various occupant types (full-time employed, unemployed, minor) for the Belgian context are identified using time-use and household budget survey data. These typical profiles are then used as a basis for the stochastic modeling of occupant presence and activity schedules. The results are used to examine the uncertainties associated with user behavior in neighborhood scale energy assessments. The resulting model efficiently represents diversity of buildings and inhabitants, however, due to the extensive modeling effort required, its applicability to larger building assemblies is limited. Orehounig et al. (2011) apply detailed simulation computations to an entire village in Switzerland and compare the results of this analysis to those derived from a reductive model applied to the same domain.

Reductive models focus on modeling the performance of a smaller domain representative of the entire building agglomeration under study. The resulting energy use patterns are then extrapolated to the larger urban area. Thus, these models can benefit from the analytic and predictive potentials of the more elaborate performance assessment methods through a systematic reduction of the computational domain. This is usually achieved through i) the selection of a sample of buildings, or ii) the development of archetype buildings (synthetic representatives), representing the energy behavior of the urban building stock under investigation. The number of selected samples or developed archetypes depends on the

diversity of energy behavior in the investigated urban area, the computational and informational requirements of the underlying performance assessment method, as well as the available resources. A review of former reductive models revealed that despite a systematic reduction of the modeling domain, most of these efforts still opted for simplified performance computational routines.

Snäkin (2000) developed a non-dynamic bottom up engineering model of the province of North Karelia in Eastern Finland. Through a stock segmentation by building usage, built form, construction/retrofit period, primary heating energy source, and heat distribution type, 4163 building types were identified. The study does not consider heat loads from solar energy and users and focuses on annual demand estimations.

Jones et al. (2001) analyzed data on heated ground floor area, façade, window to wall ratio, and exposed end area to identify 20 typical built forms for the assessment of the energy performance of the building stock of the city of Cardiff (UK). Along with 5 construction periods, this led to the definition of 100 building typologies. The buildings selected to represent these types, were subjected to assessments in a building performance benchmarking tool, UK Standard Assessment Procedure (United Kingdom Department for Business, Energy & Industrial Strategy 2013), which does not represent the temporal distribution of demand or capture the intricate effects of occupant presence and activity on energy demand. This stock energy model is a component of an integrative modeling environment aimed at facilitating urban level policy making, covering various aspects of the urban domain including traffic and industrial processes.

Hens et al. (2001) classified the Belgian building stock according to age, total floor area, built form, primary energy source and heating system type (central heating vs. dispersed heating units), to investigate the effectiveness of several CO<sub>2</sub> emission reduction strategies. The envisaged CO<sub>2</sub> reduction measures included shifting to low emission fuels, installation of heat pumps, conversion to renewable sources and improvements to the energy efficiency of buildings. The identified classes were represented by synthetic archetype buildings, which underwent a steady-state single zone monthly energy demand assessment procedure.

In a study of the energy demand of the Canadian building stock, Parekh (2005) partitioned the building stock into classes of buildings with similar usage, vintage and climate region, generating 56 building types. Other building parameters required for energy assessments were statistically determined in every class. The resulting archetypical buildings were simulated in the HOT2000 energy assessment tool (Natural Resources Canada 2016), used to estimate annual energy consumption of low-rise residential buildings (single-family houses, semi-detached houses, and row houses).

The TABULA project (Intelligent Energy Europe 2012), aiming towards development of residential building typologies for energy assessments across 11 European countries also relies on a building stock classification per climate zone, vintage and dwelling type. Real buildings representing the various characteristics of the buildings in each class are suggested as references, for performance computation purposes. The resulting typology has been used in various reductive urban energy assessment models (e.g., Dascalaki et al. 2011; Ballarini et al. 2014).

Emphasizing the significance of the occupancy related variance in energy demand in view of the increasingly stringent thermal codes, Munoz and Peters (2014) question the efficiency of the reference operational schedules for assessments pertaining to DG schemes. They argue that even though social or behavioral diversity may not play a major role in the current centralized grids, development of decentralized energy generation paradigms requires reliable data on the dynamics of energy demand at a higher spatial resolution. To address this issue, they use the TABULA building typologies for Germany to characterize Hamburg's existing stock. Focusing on the residential buildings, they analyze micro-census data, defining household types and their occurrence likelihood, in each statistical area. These households are allocated to the prototypical buildings within the designated spatial domains, but the resulting configuration is assessed using a simplified standard heat balance model, which is ill suited to incorporate this elaborate representation of inhabitants toward performance assessment.

Fewer former efforts have adopted dynamic performance simulation tools to capture the energy behavior of the representative buildings. In an effort to develop a GIS-based, simulation supported energy model of a small town, Page et al. (2014) propose a typology of buildings based on vintage and usage. The relevance and significance of these criteria are then examined through sensitivity analysis. In addition to construction period and usage, the study points out the importance of the building's urban context, adjacencies and obstructions in the thermal performance of the building, but does not include such factors in the development of building types.

Heiple and Sailor (2008) have developed a simulation-based model applicable to any large US city, which can provide parcel-level hourly consumption predictions. The method employs the capabilities of Geographical Information Systems (GIS). The computation scope is limited to a number of prototypical buildings representing each city.

Huang and Brodrick (2000) used a segmentation scheme by usage, vintage and location, developing a total of 120 commercial and 144 residential prototype/location combinations to characterize the US building stock. Simulation input files for the DOE-2 program (James



J. Hirsch & Associates 2016) were developed for each combination. Results were up-scaled and employed for cogeneration potential studies.

Caputo et al. (2013) defined 56 archetypes based on indicators such as size, number of floors and envelope compactness and construction period to represent Milan's building stock. These archetypes were subjected to detailed energy simulations using Energy Plus BPS software (NREL 2017). Using a GIS platform, a link was established between every building in the study domain and an energy profile pertaining to an archetype building, resulting in a geo-referenced energy consumption map of the city. Similar simulation-supported reductive approaches have been followed by Tuominen et al. (2014) and Orehounig et al. (2011). The latter effort, however, was based on the selection of authentic representative buildings from the study domain.

In a demand model for the residential building sector in the city of Osaka, Japan, Shimoda et al. (2003) incorporate a detailed survey on household demographics and activities to determine 23 household types. The data provides the probability distribution of each living activity such as sleep, meal, work, etc. in 15-minute time intervals for weekdays, Saturdays and Sundays for each family member's category (classified by gender, age, and occupation: employed or not). With 20 dwelling typologies of detached and apartment houses (distinguished by size) and the defined household types, 460 building typologies were defined for which hourly energy consumption was simulated. The results were extrapolated to the entire city. The solid empirical basis allows for a detailed representation of the occupant-dependent aspects including heating, cooling, lighting and appliance use schedules. However, the physical aspects of buildings may have been over-simplified as the only criteria for the segmentation of dwelling types is area.

Table 1 provides an overview of the discussed UEMs. Although the reductive approach facilitates the incorporation of detailed energy investigations, there are two principal challenges associated with this method. The first challenge pertains to the identification of an appropriate set of (authentic or synthetic) representative buildings. The second, concerns the inevitable loss of diversity through the reductive procedure.

Table 1 An overview of the consulted Urban Energy Models

Computation Method	Modeling Extent	Building Representation	Inhabitant Representation	Examples	Comments
simplified	Entire population	Authentic	Simple aggregate representation	(Li et al. 2015; Glawischnig 2016)	<ul style="list-style-type: none"> <li>• Low informational requirements</li> <li>• Low computational cost</li> <li>• Loss of diversity in case of reduced modeling extent</li> <li>• Inefficient representation of contextual parameters in case of reduced modeling extent</li> <li>• Low temporal resolution of results</li> <li>• Limited modeling and representation capabilities (in particular with regard to transient phenomena)</li> </ul>
	Reduced	Synthetic	Simple aggregate representation	(Snäkin 2000; Hens et al. 2001; Jones et al. 2001; Parekh 2005; Dascalaki et al. 2011; Ballarini et al. 2014)	
			Detailed diversity representation	(Munoz H & Peters 2014)	
Dynamic performance simulation	Entire population	Synthetic	Simple operational schedules	(Sansregret & Millette 2009; Remmen et al. 2017)	<ul style="list-style-type: none"> <li>• High informational requirements</li> <li>• High computational cost</li> <li>• High temporal resolution of results</li> <li>• limited application domain due to reliance on a statistical approach towards model generation</li> </ul>
		Authentic	Simple operational schedules	(Orehounig et al. 2011)	
			Stochastic representation	(Baetens & Saelens 2015)	
	Reduced	Authentic	Simple operational schedules	(Orehounig et al. 2011)	<ul style="list-style-type: none"> <li>• High temporal resolution of results</li> <li>• Manageable informational requirements and computational cost</li> <li>• Loss of diversity</li> <li>• Inefficient representation of contextual parameters</li> </ul>
		Synthetic	Simple operational schedules	(Huang & Brodrick 2000; Heiple & Sailor 2008; Caputo et al. 2013; Tuominen et al. 2014; Page et al. 2014)	
			Detailed diversity representation	(Shimoda et al. 2003)	

### **2.3. Domain Reduction Criteria**

Representation of the urban building stock through sample buildings or archetypes is not a new venture. In studies pertaining to urban energy modeling, typically, the existing building stock is subdivided into classes of buildings, which are expected to display similar energy behavior. Each class is then represented through a number of sample buildings or synthetic archetypes, which reflect the main energy-relevant characteristics of the buildings in their associated classes. Performance computations are conducted on this reduced domain and the results are extrapolated to the classes. A review of some contemporary reductive energy assessment methods revealed a frequent lack of explicitly stated arguments, evidence or reasoning in support of the selection of the set of building attributes intended to characterize the energy-relevant features of the building population, a.k.a. the classification criteria. Table 2 offers an overview of some reductive efforts and the various building characteristics used for stock classification or definition of building "types". Although a variety of factors have been considered in these approaches, not all energy-relevant aspects of a building have been included.

Geometric aspects are often expressed in terms of built form (detached, semi-detached, row house, etc.) usually in combination with an indicator of the building's size (volume, floor area, number of stories), (e.g., Farahbakhsh et al. 1998; Snäkin 2000; Ribas Portella 2012). In very few cases (e.g., Jones et al. 2001), more detailed descriptions of the envelope such as compactness, window area, window to wall ratio, or area of façades are included.

Construction period appears in almost all cases as a major indicator of the thermal quality of the buildings' enclosure elements. Some literature, however, state the necessity of involving the U-values of various building components in the classification process (Parekh 2005). Despite having rather similar construction details, buildings of the same construction period may display very different thermal behavior, due to the fact that the sensitivity of transmission losses to the thermal quality of various components is dependent on the share of the respective components in the overall thermal envelope, as well as their boundary conditions. Aksoezen et al. (2015) have questioned the explanatory power of a classification based on construction age and building type (in the context of the Swiss residential stock). They state that based on previous research performed by the Swiss Federal Office of Energy (Dettli & Bade 2007), particularly in the case of older buildings (constructed prior to 1921), such a classification is not suitable for the typological prediction of the performance of an already existing building due to the large intra-class variation.

The urban context's implications for the solar gains through mutual shading are frequently ignored and the effect of surrounding buildings in the reduction of heat emitting envelope area is only sporadically considered. Although built form may give an indication of the exposure of the building to outside air (e.g. detached houses versus row houses), given the significance of adjacency relations (unconditioned spaces, conditioned spaces or outside), a more detailed representation may be required. In few instances, shape of the roof and type of the lowermost building floor are suggested to be involved in the classification, providing a better understanding of the boundary conditions of the horizontal enclosures (Parekh 2005).

In research efforts with wider geographic scopes (national), climate conditions, depicted as climate zones or location, are also considered among classification criteria (e.g., Heiple & Sailor 2008; Deru et al. 2011; Benejam 2011). Additionally, Theodoridou et al. (2011) include the density of the urban area (high vs. low) in their classification, which may be a rough indicator of mutual shading, adjacency relations, and the urban heat island effect on microclimate.

Of course, the selection of the classification criteria must correspond with the characteristics of the urban context. More specifically, in the definition of representative or typical buildings, only the energy-relevant building characteristics, which contribute to the diversity of demand profiles are to be considered. If the variance in a certain building characteristic is insignificant across the urban area, this characteristic, although influential in the overall energy consumption of buildings, is not important in the classification process. In a uniform urban structure where buildings are similarly oriented and are subject to equal amounts of mutual shading, the consideration of this building parameter does not affect the classification process and may instead result in unnecessary computation load. But in the context of the dense and varied morphology of historical European cities, where building orientation, parcel size, street canyon aspect ratio, and adjacency relations are very diverse, contextual parameters are considered influential in characterizing the demand diversity of buildings.

Building's operational characteristics such as occupants' presence and activity, operation schedules, and temperature set points are usually depicted through building usage in former attempts. In some instances, number of occupants or internal temperatures have been considered in addition to usage to represent building operation (e.g., Heiple & Sailor 2008; Shimoda et al. 2003). In fact, in the case of single purpose buildings usage can be a fair indicator of the operational parameters of the building. Nevertheless, if the unit of observation is a single building, in the case of multi-purpose buildings a more elaborate method is required to represent the overall operational characteristics of the building.

Table 2 Building stock classification criteria adopted by previous reductive efforts (part 1/2)

Source	Scale	scope	construction/ renovation year	Geometry										Therm. Quality	Systems	Operational Parameters			Boundary Conditions					
				Built Form	Footprint Shape	Surface to volume ratio	Size/ Volume	Floor area	Exposed end area	Area of facade	Number of levels	Window to wall ratio	Roof shape	Type of lowest floor	U-values of components/ Insulation level	HVAC	Fuel type	Use	Internal temperatures	Occupancy/ schedule	Location	Climate zone	Density of urban area	
(Farahbakhsh et al. 1998)	National	Residential	✓	✓													✓					✓		
(Huang & Brodrick 2000)	National	Various	✓				✓											✓					✓	
Snäkin (2000)	National	Various	✓	✓													✓	✓	✓					
Jones et al. (2001)	City	Various	✓					✓	✓	✓		✓						✓						
Hens et al. (2001)	National	Residential	✓	✓				✓									✓	✓						
Shimoda et al. (2003)	City	Residential		✓			✓													✓				
Parekh (2005)	City	Residential	✓	✓	✓						✓		✓	✓	✓	✓	✓	✓		✓		✓		
Heiple & Sailor (2008)	National	Various	✓					✓									✓		✓		✓		✓	
Sansregret & Millette (2009)	State	Various	✓			✓		✓			✓						✓		✓					
Firth & Lomas (2009)	National	Residential	✓	✓																				
Amtmann (2010)	National	Residential	✓	✓			✓																✓	
Girardin et al. (2010)	City	Various	✓																✓					

Continued

Table 2 Building stock classification criteria adopted by previous reductive efforts (part 2/2)

Source	Scale	scope	construction/ renovation year	Geometry												Therm. Quality	Systems	Operational Parameters		Boundary Conditions		
				Built Form	Footprint Shape	Surface to volume ratio	Size/ Volume	Floor area	Exposed end area	Area of facade	Number of levels	Window to wall ratio	Roof shape	Type of lowest floor	U-values of components/ Insulation level			HVAC	Fuel type	Use	Internal temperatures	Occupancy/ schedule
Benejam (2011)	National	Various	✓	✓														✓			✓	
Dascalaki et al. (2011)	National	Residential	✓				✓														✓	
Deru et al. (2011)	National	Various	✓				✓											✓			✓	
Orchounig et al. (2011)	City	Various	✓																	✓		
Theodoridou et al. (2011a)	National	Residential/Mixed	✓	✓							✓			✓				✓			✓	✓
Dall'O' et al. (2012)	National	Residential	✓																		✓	
Ribas Portella (2012)	National	Various	✓	✓													✓	✓			✓	
Howard et al. (2012)	City	Various						✓										✓		✓		
Caputo et al. (2013)	City	Various	✓	✓			✓											✓			✓	
Page et al. (2014)	City	Various	✓															✓				
Tuominen et al. (2014)	City	Various	✓	✓														✓				
Fonseca & Schlueter (2015)	City	Various	✓															✓				

## General Approach

### 3.1. Suggested Framework

As previously discussed, the objective of the present work is to develop a computational frame-work for an integrative urban decision support environment, which enables the comparative analysis of the energy implications of various urban change and intervention scenarios. Following an extended study of urban energy modeling approaches, the bottom-up engineering approach was considered as the most suitable for the purpose. However, to be able to assess a wide range of scenarios pertaining to different aspects of the urban eco system, including inhabitants and microclimate, this computational framework must be inherently capable of processing relevant information with an appropriate resolution. This led to the selection of detailed dynamic performance simulation tools as the building-level computational engine of the present model. These tools, however, typically demand large amounts of input data and extensive modeling expertise. Former simulation-supported efforts frequently alleviated this issue either through reduction of the modeling domain (via selection of a representative sample of buildings or generation of representative archetypes), or through reduction of modeling effort (via employment of statistical analyses towards automated generation of simulation models).

The framework suggested in the present efforts combines the two techniques to propose a semi-automated routine towards urban stock energy modeling with respect to heating energy demand. Towards this end, a two-fold approach has been envisaged. The first development, the reductive process, is aimed at the automated selection of an appropriate sample of buildings to represent the urban domain under study. This allows for the adjustment of the extent of the modeling domain to the available resources with regard to information, time, and experienced staff. For this purpose, following the example of former

experiences, a set of criteria is defined for the classification of the urban building stock. However, the present project seeks to enhance former efforts through the introduction of a new set of criteria, incorporating contextual, geometric, operational and sematic properties of buildings in the classification process. Moreover, the potential of state of the art data-mining techniques, namely Multivariate Cluster Analysis (MCA) methods, towards classification of the building stock has been explored. Following the classification, a sample of buildings representative of the energy diversity of the urban stock are selected.

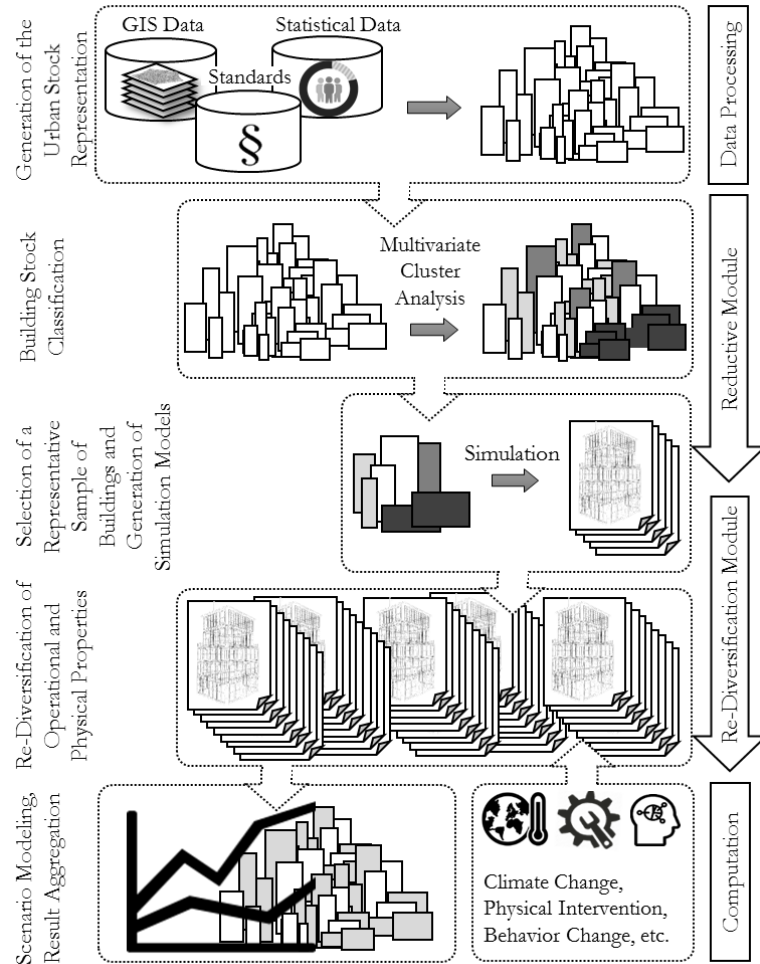
For the extraction of the values of the defined classification criteria, the reductive procedure relies on an energy relevant representation of the urban building stock. As a prerequisite to the reductive procedure, such a representation is generated in an automated process using the available large-scale data such as GIS, statistical data and building standards. This representation serves as a basis for the computational or logical determination of the values of the classification criteria for each individual building within the study domain. The selected building sample is subsequently modeled within a dynamic performance simulation tool, using detailed information. The generation of the simulation models is performed manually. Various change and intervention scenarios can be defined within the scope of the modeling capabilities of the employed simulation software and applied to these models.

The second development, concerns addressing the inevitable loss of diversity due to sampling, through an automated re-diversification of the models prior to aggregation. This step, involves the data-supported automated readjustment of the representative simulation models, to generate new versions, more representative of the energy characteristics of each building within the study domain. As such, it is essentially similar to the automated modeling approach adopted in some former efforts. With the difference that the model is not created from scratch. Rather it inherits the geometric configuration of a representative building. Towards this end, parametric modeling of non-geometric building properties such as thermal quality of components, as well as stochastic representation of occupant presence and actions have been considered. Figure 6 schematically presents the overall structure of the developed framework. Since this framework reduces the computation domain in a first step and enhances it through the re-diversification process, the term “hourglass model” has been adopted by the authors to characterize the method.

The generation of the urban stock representation, as well as the reductive process are facilitated through the development of a plug-in for the open-source GIS platform QGIS (The Open Source Geospatial Foundation 2017), version 2.9.0 Master. The plug-in is developed in the Python Programming Language (Python Software Foundation 2017), version 2.7, and benefits from various packages of the R Project for Statistical Computing (The R Foundation 2017), version 3.2.1, for the classification process. The re-



diversification step is achieved through a Python code that acquires the representative simulation models and relevant building information as input and generates readjusted simulation models associated with each building on the study domain. The presented method is in principle applicable to various geographic contexts. However, the current implementation is tailored to the Austrian context in view of the incorporated data, and its particular format. The following chapters provide detailed descriptions of the urban stock representation generation, the reductive, and the re-diversification processes and shed light to the various encountered challenges and the adopted solutions.



*Figure 6 Overall structure of the proposed urban energy computing framework*

### **3.2. Selected Urban Instance**

As a testing ground for the developmental activities, and to demonstrate the capabilities of the developed method, as well as the encountered challenges and the adopted solutions an urban instance was selected.

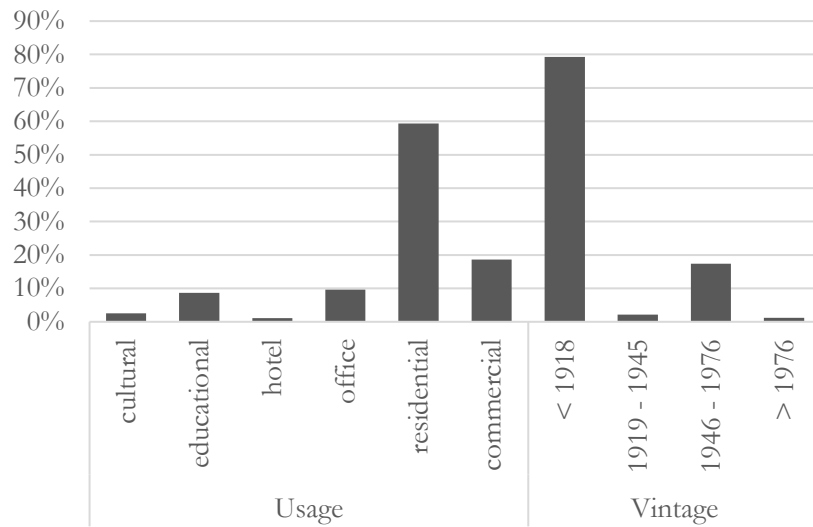
The selected testing ground involves a neighborhood in the center of the city of Vienna, Austria, covering an area of about 1.3 square kilometers. Figure 7 illustrates a bird's eye view of the selected neighborhood, including parts of the 1<sup>st</sup>, 4<sup>th</sup>, and 6<sup>th</sup> districts of the city. The varied orientation and width of the streets, as well as the presence of parks, courtyards, and plazas render the selected neighborhood morphologically diverse. As such, the buildings within the study domain are subject to different external boundary conditions in terms of solar exposure and adjacency relations.



*Figure 7 The bird's eye view of the neighborhood selected as a case study, from Google Maps (Google 2017)*

The selected neighborhood, includes some 750 buildings of various construction periods and usages. It well represents the historical building stock of the Austrian capital, however, new Viennese buildings (constructed after 1945) are underrepresented due to their low count in the central districts. A less central location may have better captured the age diversity of the stock, however, due to their architectural and historical quality, the buildings in the central districts have been better documented in the official GIS data. The area also includes a wide variety of most common building usages, including residential, office, educational and commercial buildings. The share of educational buildings in the area may be relatively high due to the presence of a university complex in the neighborhood. Figure 8 displays the distribution of the buildings by period of construction and primary usage. Note that multi-purpose buildings, especially dominated by residential usage, are very common in the area, especially along major streets.

Buildings with uncommon usages were excluded from the analysis. These included kiosks in a permanent market place, underground station entrances, and a church.



*Figure 8 Distribution of buildings in the study area by usage and age*

# The Urban Building Stock Representation Module

## 4.1. Introductory Comments

Performance estimation procedures relying on buildings' heat balance, regardless of the resolution and precision of the adopted computational method, require descriptions of physical, contextual, and operational characteristics of buildings. As such, generation of such a representation, providing information on the geometric and semantic properties, adjacency relations, and boundary conditions of the envelope components, as well as operational specifications of the building is a prerequisite of urban energy modeling efforts. The resolution and the accuracy of this representation however, depends on the granularity of the available and accessible data on the urban domain. The first developmental step of the current project consists of the superposition of various urban data sources to generate such an energy assessment compliant model of the urban domain. For this purpose, primarily the various data sources, tools, and developments, which could contribute to the development of a pertinent urban building stock representation, were identified. Since the current development is tailored to the Austrian, and in particular the Viennese context, the acquired information pertains to this geographical domain. However, the adopted methods towards model generation, can be applied to any other geographical context, provided that the required information can be obtained. Some readjustments as to the format of the available data may however be required.

## 4.2. Tools and Materials

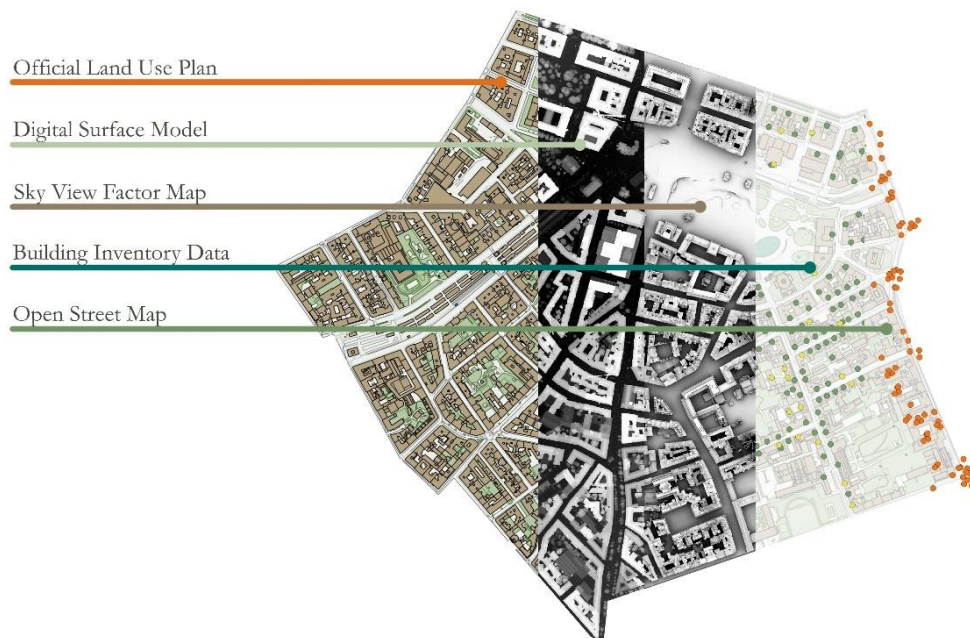
Given the strong reliance on GIS data as the principle source of spatially resolved urban-scale building information, the selection of a GIS platform as the main data visualization and processing environment seemed appropriate. QGIS (The Open Source Geospatial Foundation 2017) is a user friendly Open Source Geographic Information System certified under the GNU General Public License. It is capable of visualizing and manipulating various GIS data formats. QGIS is developed in and therefore compatible with the general-purpose high-level programming language Python (Python Software Foundation 2017). Its utility can, therefore, be enhanced through the employment of various plug-ins developed by the QGIS users' community in Python language. In the current project, the DEMTools plug-in (Hammerberg 2016) was utilized to computationally generate Sky View Factor maps. As mentioned before, the generation of an energy relevant representation of the urban building stock as well as the reductive procedure leading to the identification of a representing sample of buildings, is carried out through a dedicated plug-in, developed by the author for the purpose of the present project.

The available data on the Viennese building stock includes official and crowd-sourced Geographic Information Systems (GIS) data, as well as statistical information and building performance assessment standards:

- **Official GIS Data:** In the framework of Vienna Open Government initiative, basic GIS data of the city of Vienna including land-use plans and digital surface and elevation models of the city have been made accessible to public (Magistrat der Stadt Wien 2017b). Geographical building data inventories are also attainable for research purposes. Official Vienna GIS data includes geometric information on the buildings' foot prints, as well as properties such as height, primary usage, construction period, number of floors, etc.
- **Open Street Map** (Verein OpenStreet Map Austria 2017): Open Street Map (OSM) is an open source of GIS data, created and maintained by a community of freelance mappers. It includes data about roads, stations, buildings and their usages, and can be used for any purpose as long as its contributors are credited. In the present research, OSM data was used to refine the usage profiles of buildings since it includes a more detailed description of various usages present in a building.
- **Sky View Factor Map:** Generated using the DEMTools plug-in (Hammerberg 2016), this geo-referenced raster data layer includes the value of the Sky View Factor on the ground level of the urban area under study.

Various GIS data incorporated in the present development are visualized in Figure 9.

- **Austrian Standard B-8110-6** (Austrian Standards Institute 2014): This standard, which establishes a method for heating demand and cooling demand benchmarking, provides average values for physical parameters (e.g., the ratio of transparent elements to walls, or net to gross floor area) in the absence of more accurate information.
- **Austrian Standard B-8110-5** (Austrian Standards Institute 2011): This standard offers operational profiles for various building usages. These profiles, aimed at informing norm-based performance assessment procedures, provide default values for parameters such as area-related internal gains, air-change rate, temperature set points, and operation schedules. A reference monthly weather data for Austria is also provided in the same standard, which includes orientation-related solar radiation values.
- **Austrian Standard H-5059** (Austrian Standards Institute 2010): Complementary to the previous document, this standard provides default values for the annual area-related lighting energy demand of non-residential buildings.
- **OIB-RL 6 guidelines** (Österreichisches Institut für Bautechnik 2015): include information on the thermal quality of various building components based on construction period for the energy performance assessment of historical buildings.



*Figure 9 GIS data incorporated in the present study visualized in a GIS platform.*

Prior to the integration of various data sources, the data was checked for inconsistencies or deficiencies. The GIS data of the city of Vienna treats buildings as legal rather than architectural entities. As such, buildings pertaining to the same parcel may be identified with a single reference number, or several inventory points may be associated with a single building with several entrances (addresses). The accumulated data was as such pre-processed to ensure that each building (the unit of observation in the study) is associated with a single inventory record and given a unique identifier. Multiple OSM data records can however be associated with one building, demonstrating multiple usages.

Unfortunately, the official GIS data does not accurately represent the actual state of the buildings. This is due to the dynamic nature of the urban building stock and the effort and time associated with the accumulation and organization of the required information, partly through on-site surveys. Some of the inconsistencies encountered pertained to modified building usages. Others concerned physical interventions such as construction of rooftop extensions, very common in the Viennese contemporary architectural tradition. The GIS data available for the study also contains no information on previous refurbishment activities. Such shortcomings will of course affect the quality of the generated building stock representation and consequently the model predictions. The present research effort is not concerned with addressing such data-related issues. As such, buildings are assumed to have maintained their original conditions in terms of geometry and construction components (based on the GIS data). However, an alternative data sources on building usage, OSM, which is more regularly updated by the user community is incorporated to enrich and refine the stock model. Should more information become available in the official data sources, the developed method is capable of incorporating it into the representation generation process. Table 3 offers an overview of the available data sources and their contents.

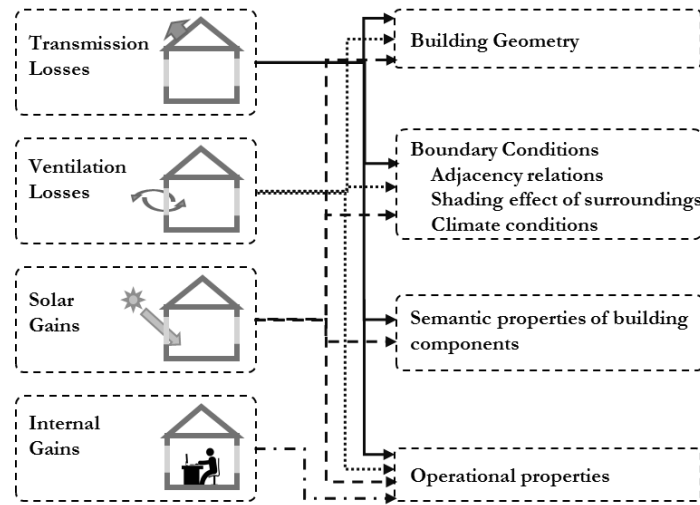
Table 3 Summary of the available data sources and the contained information

	Source	Data type	Contained data
GIS DATA	ViennaGIS	Land Use Plan (Vector layer)	<ul style="list-style-type: none"> <li>• Building Footprint Polygons</li> <li>• Construction type (main/annex)</li> <li>• Relative eaves height</li> <li>• Elevation from ground</li> </ul>
		Digital Elevation Model (Raster layer)	<ul style="list-style-type: none"> <li>• Relative height of every point</li> </ul>
		Building Inventory (Vector layer)	<ul style="list-style-type: none"> <li>• Building construction period or year</li> <li>• Main building usage</li> <li>• Number of Floors</li> </ul>
	OpenStreetMap	Land Use Plan (Vector layer)	<ul style="list-style-type: none"> <li>• Building Footprint Outline</li> <li>• Building usage</li> </ul>
	DEMTTools	Sky View Factor Map (Raster layer)	<ul style="list-style-type: none"> <li>• Sky View Factor of every point at ground level</li> </ul>
STANDARDS	Austrian Standard B-8110-5	Thermal insulation in building construction: Model of climate and user profiles	<ul style="list-style-type: none"> <li>• Usage-based internal gains</li> <li>• Usage-based infiltration rate</li> <li>• Usage-based use hours</li> <li>• Reference weather data</li> </ul>
	Austrian Standard B-8110-6	Thermal insulation in building construction: Principles and verification methods: Heating demand and cooling demand	<ul style="list-style-type: none"> <li>• Average window to wall ratio</li> <li>• Average frame to window ratio</li> <li>• Average net to gross floor area ratio</li> </ul>
	Austrian Standard H-5059	Energy performance of buildings: Energy use for lighting	<ul style="list-style-type: none"> <li>• Annual lighting energy demand of non-residential buildings</li> </ul>
	OIB-RL 6	Guidelines: Energy-technical behavior of buildings	<ul style="list-style-type: none"> <li>• Age-based component U-values</li> <li>• Age-based window solar transmittance</li> </ul>



### 4.3. Building Stock Representation Routine

The intended building stock representation is supposed to cover various energy-relevant aspects of the buildings. To determine the scope of the required information, the various terms in the heat balance of a building were considered and data required for the quantification of each term was identified. Figure 10 displays the building properties required to assess various factors contributing to a building's heating demand.



*Figure 10 Building information required for the quantification of the various heat transfer processes occurring in a building*

Based on the required information and the available data sources a data representation schema was developed for the representation of the individual buildings. This schema was the basis of the object-oriented framework developed for the extraction of the necessary information from the GIS sources as well as the mentioned standards. This data structure is illustrated in Figure 11.

The data extraction and organization process is carried out through a QGIS plug-in. Before the plug-in is run, the necessary GIS data layers must be imported in the GIS platform and arranged in terms of reference coordinate system and projection, such that different information pertaining to the same geographic location are superimposed.

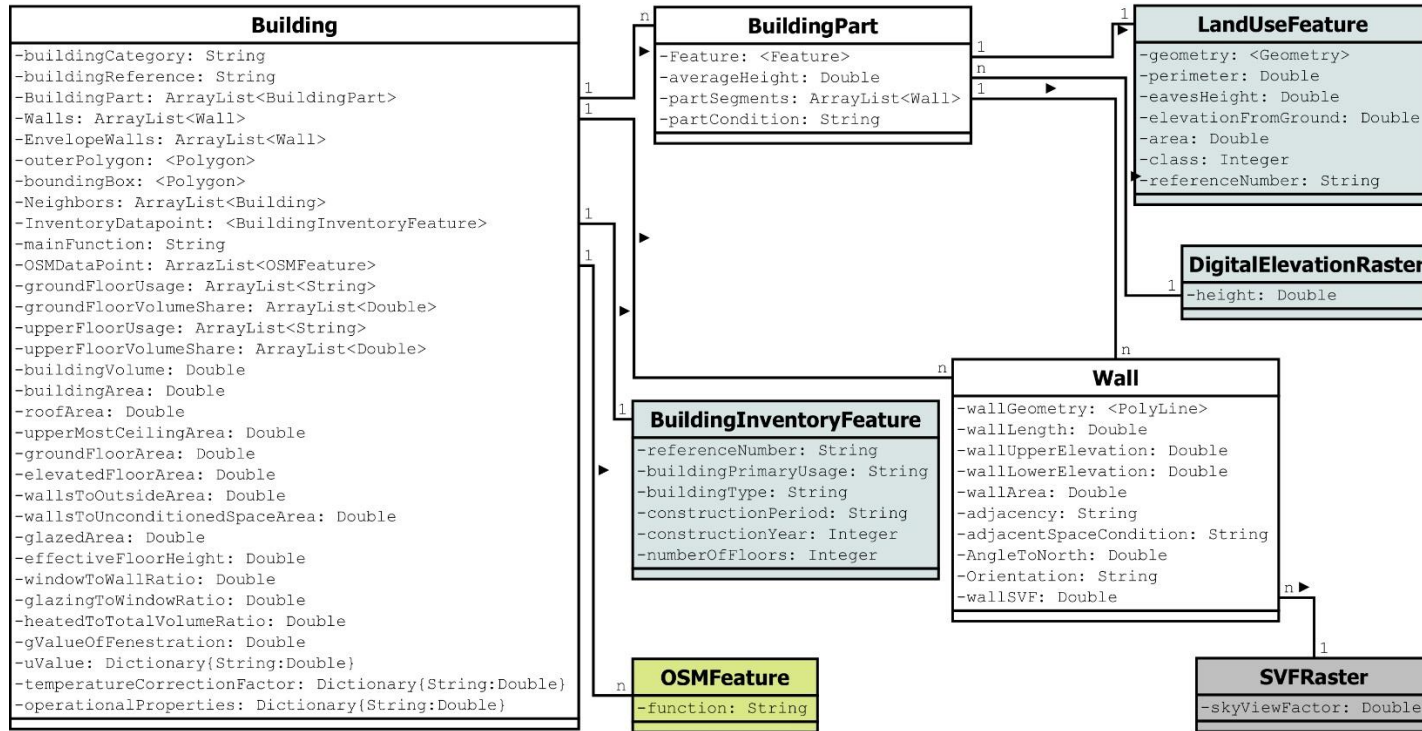
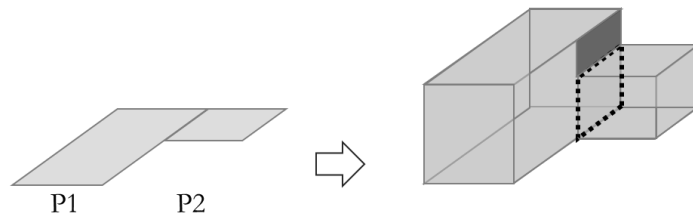


Figure 11 The building data representation schema developed for the structuring of building information from various sources, towards generation of necessary descriptions for all buildings within the study domain

### Envelope Geometry and External Boundary Conditions

Beginning with an analysis of the ViennaGIS Land Use plan, the developed plug-in identifies all unique reference codes, associated with features depicting a building, and creates corresponding “Building” objects. For each “Building” Land Use features (polygons with additional attributes) constituting the building’s footprint are identified by the reference number as “Building Parts”. Utilizing the foot print geometries and eaves height information pertaining to “Building Parts”, vertical building enclosures or “Wall” objects are generated. Since the footprint geometry of the building may be composed of several “Parts” each associated with different height information, geometric reasoning is required to determine which edges define the boundaries enclosing the building volume. For this purpose, edges of all “Building Parts” are compared pair-wise to determine collisions. The height information associated with colliding edges are compared. In case of a height difference, a “Wall” object is constructed with the geometry of the mutual edge and an elevation equal to the difference between the heights of the touching polygons or “Building Parts” (Figure 12). All “Walls” are assumed to be exposed to outside air at this point.



*Figure 12 Identification of vertical elements enclosing the building envelope: polygons 1 is associated with a higher eaves height than polygon 2. Once the polygons are extruded by the given heights to form the volume of the building, the resulting surfaces are not all considered parts of the building envelope. In this case, the shared wall marked in dotted line is excluded from the bounding envelope, whereas the wall element marked in darker color is included.*

The lowermost enclosures are considered to be floor elements adjacent to unheated basements (in line with the Viennese architectural tradition), unless according to the height information of the representing polygon, the building part is elevated from the ground (e.g. the case of a protruding built volume). In the latter case, the floor element is adjacent to outside air (Figure 13).

A rule-based logic is implemented to infer the roof type and condition of the attic according to the difference between the eaves height and the average height of the building. The average height of a “Building Part” (represented by a polygon) is computed using the digital

elevation model raster data. For this purpose, the heights associated with every pixel on this raster image, which falls within the bounds of a “Building Part”, are averaged. If the difference between this value and the eaves height provided by the Land Use data is below 1 meter, the roof is assumed to be flat and adjacent to outside air. If it lies between 1 and 3 meters, the roof is assumed to be sloped, but the attic space is considered unheated, as such the uppermost enclosure of the thermal envelope is the upper-most ceiling, adjacent to an unconditioned attic space. In case of a difference of above 3 meters, the attic space counts as part of the thermal envelope of the building (Figure 14). In this case, the area of the roof and the volume of the attic space are approximated based on the area of the footprint polygon and the height difference.

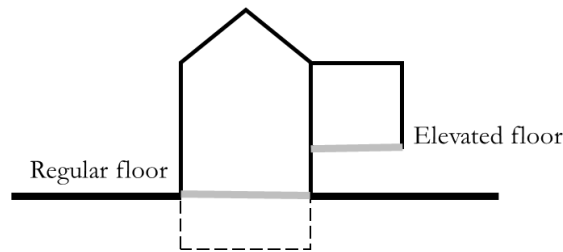


Figure 13 Floor elements are assumed to be adjacent to an unheated basement unless their constituting “Building Part” is associated with an elevation above ground level.

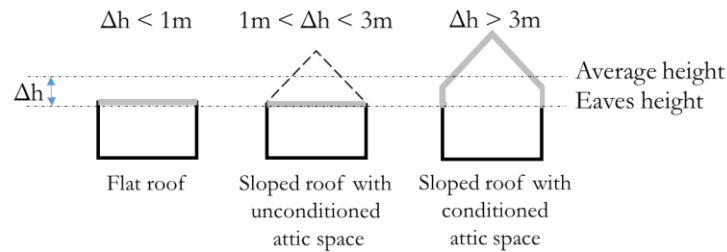
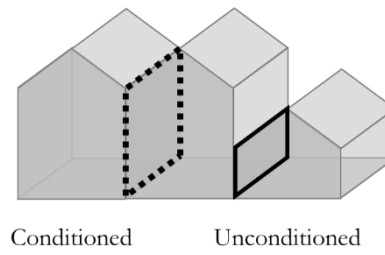


Figure 14 Rule-based identification of roof type and attic space condition

Once the outer bounds of the building are determined, the pertaining information including the definitions of “Wall” objects are stored under the “Building” object. Then, for each building the bounding box of the building footprint is defined. Using the bounding box geometries in the first step, and the footprint geometry in a second step, a neighbor search is performed for each building. The “Walls” of neighbor buildings (wall baseline geometries and lower and upper elevations) are then cross-examined for adjacency relations, resulting in each wall element’s association with an adjacency status: outside, or another space. The

orientation of “Walls” is computationally derived. Area and orientation of the external wall elements can be used along with standard values for window to wall ratio and glazing to window ratio to determine the area and orientation of the transparent building elements.

In the ViennaGIS data, features representing buildings may be parts of a main or a utility (annex) building. This is expressed through the building type property. In the present implementation, utility buildings (such as separate garages or storage cabins) are considered as unconditioned spaces. In the description of “Wall” elements, which are adjacent to another space, the thermal condition of the adjacent space is also stored (Figure 15). The adjacency relation information is required, for instance in the calculation of thermal transmittance for the determination of surface resistances of various elements.



*Figure 15 Determining the external boundary conditions of wall elements: Considering the building in the middle, the wall marked in dotted line is an adiabatic wall (adjacent to a main building), whereas the one marked in continuous line is a wall adjacent to an unconditioned space (utility building). All other wall elements are adjacent to outside air.*

The exact computation of the solar radiation incident on building enclosures is theoretically possible, based on the information contained in the GIS layers. However, for urban scale inquiries, the required computational effort and time may be unreasonable. As such, the shading effect of the surrounding buildings is approximated by the Sky View Factor (SVF) on a point on the ground in the vicinity of the building’s walls. Although SVF is only a measure of the visibility of the sky hemisphere from a given point, a combined use of SVF and orientation can provide a good estimation of incident solar radiation with an accuracy of up to 90% (Robinson 2006). Accordingly, if a larger portion of the sky is visible from a point close to the wall, that wall will receive less shading from its surrounding elements. If the footprint of the wall is longer than 5 m, three points along the wall are selected, and the corresponding SVF values are averaged to better represent the shading conditions.

### **Semantic Properties**

Each “Building” is associated with a feature on the building inventory data layer through the reference number. This feature contains various information fields including the construction year, or in the absence of that, the construction period of the building. Each construction period is associated with a set of building component thermal transmittance (U-value) derived from the standards for the energy assessment of historical buildings in Austria (Österreichisches Institut für Bautechnik 2015). The correct set of component thermal transmittance values is selected and stored in the building object accordingly. If no information on the age of the building is available (incomplete or missing inventory data), the building is assumed to have been built before 1900, since the majority of the buildings in the city are from that period. In cases where the period of construction of the building overlaps with two different historical periods mentioned in the standard (due to dissimilar categorization schemas), the older period is considered for the determination of component thermal properties.

### **Operational Parameters**

The building inventory data layer also provides information on the primary function of buildings (residential, educational, office, etc.). The information contained in the OSM data is integrated to refine and enrich the usage descriptions of buildings. For this purpose, a rule-based method is developed to associate different usages with various portions of the total volume of the building. For this purpose, the OSM building usages are categorized into four classes based on the typical location and spread of these usages in a building:

- Functionalities such as school, embassy, hotel, and pension are typically associated with the entire building. As such, in case the OSM data indicates the presence of one of these usages in a building, the primary usage of the building extracted from ViennaGIS data is overwritten. OSM data in such cases is relied on since crowd-sourced data repositories are expected to be updated more frequently.
- Functionalities such as banks, post offices, supermarkets, restaurants, coffee houses and shops typically occupy the ground floors of buildings. As such, the ground floor volume is divided among such usages.
- Kindergartens, sport facilities, and religious assembly halls are assumed to occupy an entire floor of the building.
- To other usages such as offices a portion of the volume of a floor is allocated. Upper limits are considered to prevent unreasonably large volumes associated with such functions in larger buildings.

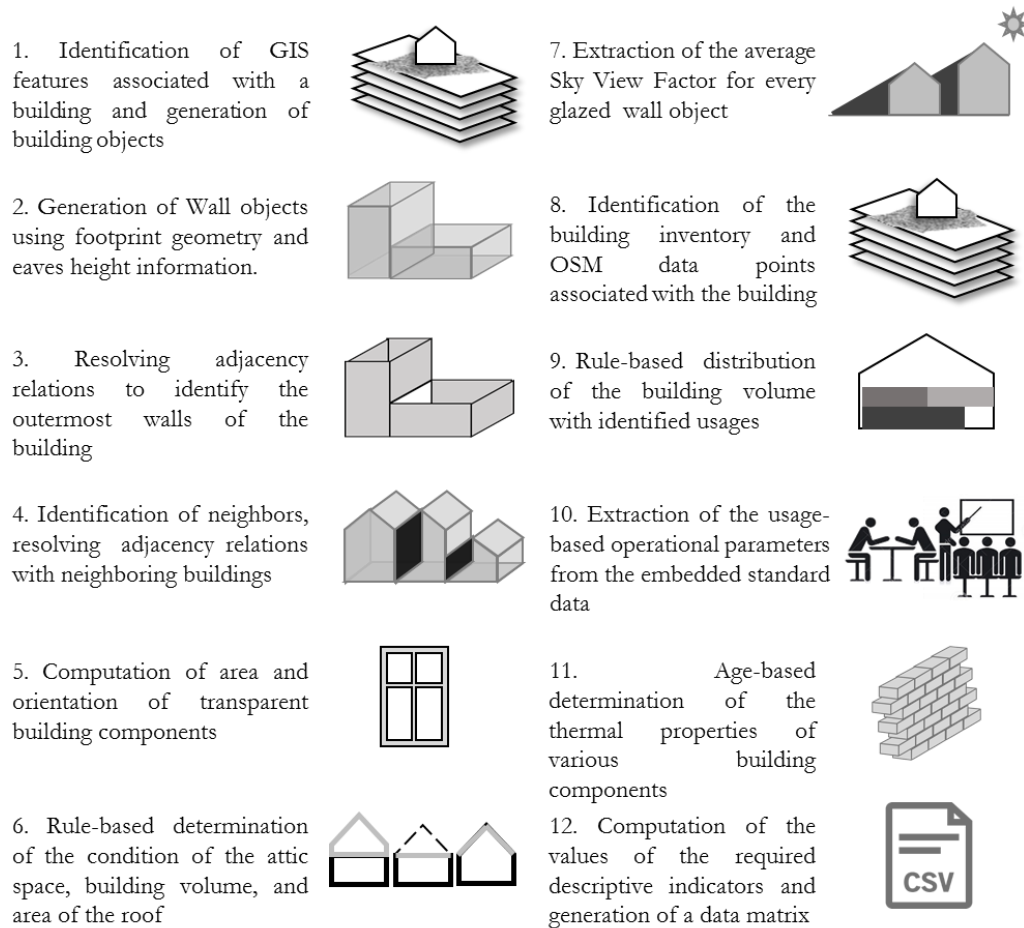
For the computation of the volume associated with every usage, following the above logic, the volume of every building story is required. For this purpose, the volume of the building is divided by the number of floors in a building. The number of floors is given by the building inventory data. This information is however not entirely reliable in case of larger buildings, where the number of floors in various parts of the building may vary. The number of floors can be used to derive the floor height as an indicator of the ratio of volume to useful floor area.

The primary and additional usages of the building, as well as their associated shares of the total building volume are stored in the “Building” object. Sets of operational parameters for different usages including operation schedules, hourly air-change rate, and internal gains are extracted from relevant standards (Austrian Standards Institute 2010, 2011).

Currently the urban stock representation is generated on the fly. The information required for the reductive module, which will be discussed in detail in the following section are automatically extracted from the generated representation and stored in a Comma

Separated Values file. But the developed logic can be implemented to create multi-purpose energy-relevant urban stock data repositories. Recent efforts towards enhancing the CityGML standard urban data model (Open Geospatial Consortium 2017), for energy assessment compliance facilitate the generation of such repositories (Benner et al. 2016; Nouvel et al. 2015). Due to its ontological congruence with energy assessment compliant building models, the developed representation can be adapted to similar energy-based representation schemas with minor modifications.

Figure 16 recaptures and summarizes the necessary steps towards generation of the required urban building stock representation.



*Figure 16 Summary of the steps required for the generation of an energy-assessment-compliant representation of the urban building stock*



# The Reductive Module

## 5.1. Introductory Comments

A majority of former efforts have relied on construction period, usage and built form to classify the urban building stock. For this purpose, typically, bins or threshold values are defined for the considered classification criteria. The data space is then consecutively partitioned on the defined thresholds. The values of the classification variables in the resulting groups will therefore always remain within the limits of a certain predefined bin. A downside of this method is that as the number of the considered criteria and the associated bins grows, the resulting groups become smaller and more numerous. To avoid this, a high level of aggregation in the input parameters is required, which is probably the logic behind the selection of rough indicators such as building usage or age instead of more finely defined characteristics.

Rather than relying on such vague descriptors, in the present research effort a set of indicators are defined to more implicitly capture the various energy-relevant aspects of the buildings, including physical, operational and contextual characteristics. To enable the use of numerous classification criteria, while maintaining control over the number of emerging categories, a different classification method widely incorporated in datamining has been adopted. This allows for the development of a more generic approach towards stock classification in view of the dynamic nature of cities.

## **5.2. Classification Criteria**

Classification criteria have to be defined in such a way as to incorporate the essential energy-relevant building characteristics reflected in the developed representation. These include building geometry and adjacency relations, solar exposure, thermal quality of the envelope, and operational parameters. For this purpose, various properties of the buildings and their components are combined to create meaningful descriptive indicators of various aspects of the building. Obviously, there is no unique or best set of parameters that can cover various energy-relevant building characteristics. Major aspects can be expressed with different levels of accuracy and through different sets of indicators. For instance, different methods can be envisaged to represent operational parameters such as internal gains, ventilation rate, and occupancy schedules. The indicators adopted or developed in the present project to express various building characteristics are described below.

### Geometry and Adjacency Relations

In similar efforts, typically the size and built form of a building, in many instances as categorical variables, are adopted to represent geometry. To facilitate the utilization of datamining methods, however, in the present effort categorical indicators were avoided in favor of numerical indicators. As such, net building volume (Equation 1) and thermally effective envelope area (Equation 2) were selected to represent the size of the building. The latter variable factors in the adjacency relations through a standard-given temperature correction factor. This factor, assuming a value between 0 and 1 is defined in the standard for various building components, depending on their adjacency relations to account for the differences in the contribution of various components to transmission heat losses due to the variance in the boundary conditions. For instance, for exposed components of the envelope this variable assumes a value of 1, whereas components adjacent to unheated spaces (such as unconditioned attic or basement spaces), or earth (such as walls or floors of a conditioned basement) are associated with lower temperature correction factors, to compensate for the higher temperatures of earth and unconditioned spaces compared to outside air in the heating season. As such, temperature correction factor is a good indicator of how different adjacency relations can alter the significance of a building element in contributing to transmission losses.

$$V_n = V \times f_n \quad (1)$$

, where  $V_n$  is the net volume of the building [ $m^3$ ],  $V$  is the gross volume of the building [ $m^3$ ] as computed by in the previous step based on the available GIS data, and  $f_n$  is the ratio of net to gross volume suggested by the standard (Austrian Standards Institute 2014).

$$A_e = \sum(A_i \times f_{t,i}) \quad (2)$$

, where  $A_e$  is the thermally effective envelope area [ $m^2$ ],  $A_i$  is the area of each building enclosure [ $m^2$ ] as computed in the previous step based on the available GIS data, and  $f_{t,i}$  is the temperature correction factor suggested by the standard (Austrian Standards Institute 2014) for the heating season, depending on the adjacency relation of the enclosure in question (see Table 4).

Table 4 Temperature correction factors for various building enclosure elements based on their boundary conditions as given in the Austrian Standard B 8110-6 (Austrian Standards Institute 2014)

Building element	External boundary condition	$f_{t,i}$
Wall	Outside air	1
	Unconditioned space	0.7
Floor	Outside air	1
	Unconditioned basement	0.7
Ceiling/Roof	Outside air	1
	Unconditioned attic	0.9

The built form of a building can be expressed through the ratio of the building's volume to envelope area. Usually the total area of the envelope is incorporated. However, given the significance of adjacency relations, in the current implementation, the thermally effective envelope area has been used to compute the thermal compactness (Equation 3).

The buildings constructed in different historical periods, for various purposes, and different social and economic contexts feature different proportions. For instance, more representative historical buildings have much higher floor to ceiling heights than those constructed after the Second World War. Due to such differences in room height, buildings with similar volumes may be associated with different useful floor areas. Since the useful floor area has strong implications for the estimation of parameters such as internal gains, the effective floor height, defined as the ratio of the building's net volume to the total floor area (Equation 4) is included as an additional geometric indicator in the current project.

$$C_t = \frac{V_n}{A_e} \quad (3)$$

, where  $C_t$  is the thermal compactness of the building [m].

$$h_e = \frac{V_n}{A_f \times n_f} \quad (4)$$

, where  $h_e$  is the effective floor height [m],  $A_f$  is the area of the building's footprint [m<sup>2</sup>] extracted from the GIS data in the previous step, and  $n_f$  is the number of floors associated with the building in the building inventory data. As mentioned before the number of floors may not always be stated accurately in the databases, since some building parts may be higher than others, whereas only one number is associated with every building. In cases where the resulting effective floor height falls outside of a plausible range (under 2.7 m or above 5), the value is replaced by the average effective floor height across the data set.

## Solar Exposure

Solar exposure has often been ignored in former classification schemas. In very few instances, parameters such as exposed envelope area or window to wall ratio have been involved. However, the impact of orientation and mutual shading effect of buildings on solar gains have not been fully considered. In the present research, the solar exposure of the building has been expressed through the effective glazing ratio. This variable has been defined to include information such as climate zone, orientation, and mutual shading (Equation 5). Climate zone and orientation are captured through an orientation factor. Orientation factor for various orientations is defined as the ratio of the average available radiation in that orientation (derived from standard monthly climate data) to the average available radiation toward south (Equation 6). Mutual shading is represented through SVF.

$$GR_e = \frac{WWR \times GWR \times g \times \sum (A_{ow,i} \times f_{o,i} \times SVF_i)}{\sum A_{ow,i}} \quad (5)$$

, where  $GR_e$  is the effective glazing ratio,  $WWR$  and  $GWR$  are the window to wall and glazing to window ratios respectively (Austrian Standards Institute 2014),  $g$  is the solar energy transmission of window glazing (derived from Österreichisches Institut für Bautechnik 2015),  $A_{ow,i}$ ,  $f_{o,i}$ , and  $SVF_i$  are the area, orientation factor, and Sky View Factor pertaining to every wall element exposed to outside air. Wall elements, their orientation and the pertaining SVF values have been derived in the previous step.

$$f_o = \frac{\sum_{j=1}^{12} (r_{o,j})}{\sum_{j=1}^{12} (r_{s,j})} \quad (6)$$

, where  $f_o$  is the orientation correction factor for a certain orientation,  $r_{o,j}$  and  $r_{s,j}$  are the radiation values for the considered orientation and the south orientation for each month  $[kWh \cdot m^{-2}]$ , provided in the standard monthly weather data (Austrian Standards Institute 2011). Sixteen orientations (four cardinal directions and three intervals between each pair) have been considered in the standard climate data. As such, in the definition of wall objects in the previous step, each wall object

has been associated with one of these orientations according to the azimuth angle of the wall element's foot print.

### Thermal Quality of the Envelope

Rather than relying on period of construction as the main indicator of the envelope's thermal quality, in the present effort, an effective average envelope thermal transmittance (U-value) has been adopted. This variable not only involves the U-values of all envelope components, but also weights them according to the share of the said components in the envelope and their significance for transmission losses based on their external boundary conditions (Equation 7). As explained before the thermal transmittance of various building enclosures is determined in the previous phase of the project according to the construction period of the building. But the introduction of area and adjacency-related weights renders this variable more effective in expressing the building's overall quality. The thermal quality of the envelope can also be expressed with a lower level of aggregation through effective wall, ceiling, and floor U-values (Equation 8-Equation 10).

$$U_e = \frac{\sum(U_i \times A_i \times f_{t,i})}{A_e} \quad (7)$$

, where  $U_e$  is the effective envelope U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ],  $U_i$ ,  $A_i$ , and  $f_{t,i}$  are the U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], area [ $m^2$ ], and temperature correction factor associated with every building enclosure.

$$U_{w,e} = \frac{\sum(U_{w,i} \times A_{w,i} \times f_{t,i})}{A_e} \quad (8)$$

, where  $U_{w,e}$  is the average effective wall U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], and  $U_{w,i}$ ,  $A_{w,i}$ , and  $f_{t,i}$  are the U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], area [ $m^2$ ], and temperature correction factor of different wall elements respectively.

$$U_{c,e} = \frac{\sum(U_{c,i} \times A_{c,i} \times f_{t,i})}{A_e} \quad (9)$$

, where  $U_{c,e}$  is the average effective ceiling/roof U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], and  $U_{c,i}$ ,  $A_{c,i}$ , and  $f_{t,i}$  are the U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], area [ $m^2$ ], and temperature correction factor of different ceiling/roof elements respectively.

$$U_{f,e} = \frac{\sum(U_{f,i} \times A_{f,i} \times f_{t,i})}{A_e} \quad (10)$$

, where  $U_{f,e}$  is the average effective floor U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], and  $U_{f,i}$ ,  $A_{f,i}$ , and  $f_{t,i}$  are the U-value [ $W \cdot m^{-2} \cdot K^{-1}$ ], area [ $m^2$ ], and temperature correction factor of different floor elements respectively.



### Operational Parameters

Although usage is a good categorical indicator of a building's operational parameters in single use buildings, its adoption to express the operational characteristics of multi-usage buildings can be very limiting. As such, various indicators were considered to describe the operational characteristics of buildings including the temporal use patterns, internal gains, and ventilation rates. Naturally, concentrating the buildings' use schedules into a few numeric variables will reduce the resolution of the representation. It is, however, necessary for the purpose of the classification. Toward this end, in order to capture the building's temporal use pattern, four indicators were defined. The fraction of annual hours that the building is in use (Equation 11) is used as a measure of the frequency of usage. To distinguish buildings with similar number of use hours but different use duration in daytime and nighttime, three indicators were envisaged: daytime use intensity, which captures the ratio of daytime use hours to overall use hours (Equation 12), or alternatively, the annual daytime and nighttime use ratios (Equation 13, Equation 14). For multi-purpose buildings, all operational parameters are expressed as average values for the various usages present in the building, weighted according to the share of the overall volume associated with every usage. This information is stored in the building representation created in the previous step. Usage-derived operational parameters are extracted from the Austrian standard B 8110-5 (Austrian Standards Institute 2011) and Austrian standard H 5059 (Austrian Standards Institute 2010). For an overview of these parameters see Table 5.

$$O_u = \frac{\sum(t_{use,a,i} \times f_{v,i})}{t_a} \quad (11)$$

, where  $O_u$  is the fraction of the annual hours that the building is in use,  $t_{use,a,i}$  is the number of annual use hours for a certain usage [ $h \cdot a^{-1}$ ],  $f_{v,i}$  is the fraction of the volume associated with that usage, and  $t_a$  is the total number of hours in a year [ $h \cdot a^{-1}$ ].

$$O_{d/u} = \frac{\sum(t_{day,a,i} \times f_{v,i})}{\sum((t_{day,a,i} + t_{night,a,i}) \times f_{v,i})} \quad (12)$$

, where  $O_{d/u}$  is the daytime use intensity,  $t_{day,a,i}$  and  $t_{night,a,i}$  are the annual operation durations in daytime and nighttime for a certain usage [ $h \cdot a^{-1}$ ].

$$O_d = \frac{\sum(t_{day,a,i} \times f_{v,i})}{t_a} \quad (13)$$

, where  $O_d$  is the annual daytime use.

$$O_n = \frac{\sum(t_{night,a,i} \times f_{v,i})}{t_a} \quad (14)$$

, where  $O_n$  is the annual nighttime use.

Target internal temperature is also an essential factor in differentiating the operational characteristics of various space usages. However, the Austrian standards incorporated suggest an indoor temperature of 20 degrees for all usages involved in the present effort. Accordingly, the only building functions with different target indoor conditions are indoor swimming pools, senior care centers and hospitals which are absent in the adopted case study. To be able to generalize the developed method to other geographical locations, target indoor conditions need to be considered for building classification.

To express the internal gains, average area-related rate of internal gains (Equation 15), and daily internal gains (Equation 17) were considered. The former variable ignores the differences in the number of operation hours of various usages in the building, whereas the second variable, involving the operation time, provides a better approximation of the building's internal gains.

$$q_{i,h} = \sum((q_{i,h,i} + q_{i,l,i}) \times f_{v,i}) \quad (15)$$

, where  $q_{i,h}$  is the average area-related internal gains rate [ $W.m^{-2}$ ],  $q_{i,h,i}$  is the area-related rate of internal gains from equipment, and occupants during the heating season [ $W.m^{-2}$ ],  $q_{i,l,i}$  is the area-related rate of internal gains from lighting during the heating season [ $W.m^{-2}$ ] (Equation 16).

$$q_{i,l,i} = \frac{0.5 \times 1000 \times LED_i}{t_{use,a,i}} \quad (16)$$

, where  $LED_i$  is benchmark value for the lighting energy demand associated with a usage [ $kWh.m^{-2}.a^{-1}$ ], provided by the Austrian standard H 5059 (Austrian Standards Institute 2010).

$$Ig_d = \sum((q_{i,h,i} + q_{i,l,i}) \times f_{v,i} \times t_{use,d,i}) \quad (17)$$

, where  $Ig_d$  is the daily area-related internal gains [ $Wh.m^{-2}.d^{-1}$ ], and  $t_{use,d,i}$  is the daily operation time for every usage present in the building [ $h.d^{-1}$ ].

Thermal mass of buildings impacts the effectiveness of heat gains and consequently the heating demand. However, in the case of central European urban settings variance in thermal mass does not significantly contribute to the energy diversity of the building stock. In brick buildings, the effectiveness of heat gains is reduced by only 5% in lightweight constructions as compared to very massive ones (Staniszewski & Gierga 2016). However, the reduction may be more significant, when wooden constructions (e.g., in rural areas) are involved. As such, in the current implementation on an urban context thermal mass has not been involved in the classification and sampling efforts. This, by no means undermines the significance of the thermal mass in the estimation of the temporal dynamics of heating demand. It only means that the diversity in the heating demand of various buildings is not significantly affected by the differences in the thermal mass of these buildings. Had other performance indicators (such as summer overheating) been taken into consideration, this variable might have become more relevant in the classification process.

Similar to the internal gains, the average hourly air-change rate (Equation 18) and the daily air-change rate (Equation 19) were adopted to reflect the ventilation behavior of the buildings. Due to the lack of more detailed information, all buildings are assumed to be naturally ventilated. This assumption agrees with the common practice in most historical buildings.

$$n_v = \sum(n_{v,i} \times f_{v,i}) \quad (18)$$

, where  $n_v$  is the hourly ventilation rate [ $h^{-1}$ ], and  $n_{v,i}$  is the hourly ventilation rate associated with every usage [ $h^{-1}$ ].

$$Ac_d = \sum(n_{v,i} \times f_{v,i} \times t_{use,d,i}) \quad (19)$$

, where  $Ac_d$  is the daily air change rate [ $d^{-1}$ ].

Table 5 An overview of the default values for operational parameters for the usages present in the current project (Source : Austrian Standards Institute 2011; Austrian Standards Institute 2010)

Building usage	$t_{day,a,i}$ [h. a <sup>-1</sup> ]	$t_{night,a,i}$ [h. a <sup>-1</sup> ]	$t_{use,d,i}$ [h. d <sup>-1</sup> ]	$q_{i,h,i}$ [W. m <sup>-2</sup> ]	$LED_i$ [kWh. m <sup>-2</sup> . a <sup>-1</sup> ]	$n_{v,i}$ [h <sup>-1</sup> ]
Residential	5020	3740	24	3.75	-	0.4
Office	2970	258	12	3.75	32.2	1.2
Kindergarten/school	2860	368	12	3.75	24.8	1.2
University	2930	298	12	7.5	24.8	1.8
Hotel	1550	2830	12	7.5	65.1	1.2
Gastronomy	3130	1250	12	7.5	27.1	2
Assembly	1295	1260	7	7.5	27.1	1.8
Sports facility	3690	690	12	7.5	37.9	3
Retail	2970	834	12	3.75	70.6	1.8

Since some of the above-introduced descriptive indicators are just different expressions of the same building characteristic, the combined use of all indicators will result in an unintentional weighting of these aspects of buildings. Weighting of different building characteristics in the classification process may in fact be desired, due to the notion that all building parameters may not contribute to a building's heating demand by equal measures. However, such a weighting schema should be based on an in-depth analysis of the sensitivity of the urban building stock energy demand to different building properties. In the absence of such information, in the current research, no weighting schema has been considered. As such, different combinations of the above-mentioned indicators have been explored toward their potential for an efficient classification of the building stock. These sets have been composed such that all essential aspects of a building are covered, and a certain characteristic is not covered by multiple indicators. The classification scenarios will be introduced in a dedicated section. Table 5 provides an overview of the various indicators that have been considered for building stock classification in the scope of the current

project, along with their computation method and the associated input parameters. Note that these indicators are by no means claimed to be the only aggregate indicators conceivable.

Table 6 An overview of the various indicators that have been considered for building stock classification with their computation method and the associated input parameters (part 1/2)

	Abbr.	Variable	Description	Formula	Parameters	
Geometry	$V_n$	Net Volume [m <sup>3</sup> ]	An indicator of the size of the building	$V_n = V \times f_n$	$V$ $f_n$	Gross volume of the building [m <sup>3</sup> ] Net to gross volume ratio
	$A_e$	Thermally effective envelope area [m <sup>2</sup> ]	Area of the heat emitting envelope corrected for adjacency relations	$A_e = \sum (A_i \times f_{t,i})$	$A_i$ $f_{t,i}$	Area of an envelope element [m <sup>2</sup> ] Temperature correction factor
	$C_t$	Thermal compactness [m]	Ratio of the net building volume to the thermally effective envelope area	$C_t = \frac{V_n}{A_e}$		
	$h_e$	Effective floor height [m]	Ratio of the building volume to the floor area	$h_e = \frac{V_n}{A_f \times n_f}$	$A_f$ $n_f$	Total footprint area [m <sup>2</sup> ] Number of floors
Solar gains	$GR_e$	Effective glazing ratio	Average glazing to wall ratio weighted by orientation and corrected for the shading effect of the surroundings. Weights associated with orientations were based on reference climate data	$GR_e = \frac{WWR \times GWR \times g \times \sum (A_{ow,i} \times f_{o,i} \times SVF_i)}{\sum A_{ow,i}}$	$WWR$ $GWR$ $g$ $A_{ow,i}$ $f_{o,i}$ $SVF_i$	Window to wall ratio Glass to window ratio Solar factor of glazing Area of external wall element [m <sup>2</sup> ] Orientation correction factor Corresponding Sky View Factor
Thermal Quality	$U_e$	Effective average envelope u-value [W.m <sup>-2</sup> .K <sup>-1</sup> ]	Average u-value of the envelope corrected for adjacency relations and weighted by the corresponding areas	$U_e = \frac{\sum (U_i \times A_i \times f_{t,i})}{A_e}$	$U_i$ $A_i$ $f_{t,i}$	U-value of building element [W.m <sup>-2</sup> .K <sup>-1</sup> ] Area of building element [m <sup>2</sup> ] Temperature correction factor
	$U_{w,e}$	Effective Wall U-value [W.m <sup>-2</sup> .K <sup>-1</sup> ]	Average u-value of wall elements corrected for adjacency relations and weighted by the corresponding areas	$U_{w,e} = \frac{\sum (U_{w,i} \times A_{w,i} \times f_{t,i})}{A_e}$	$U_{w,i}$ $A_{w,i}$	U-value of wall element [W.m <sup>-2</sup> .K <sup>-1</sup> ] Area of wall element [m <sup>2</sup> ]

Continued

Table 6 An overview of the various indicators that have been considered for building stock classification with their computation method and the associated input parameters (part 2/2)

	Abbr.	Variable	Description	Formula	Parameters	
Thermal Quality	$U_{c,e}$	Effective Roof/Ceiling U-value [W.m <sup>-2</sup> .K <sup>-1</sup> ]	Average u-value of roof/ceiling elements corrected for adjacency relations and weighted by the corresponding areas	$U_{c,e} = \frac{\sum(U_{c,i} \times A_{c,i} \times f_{t,i})}{A_e}$	$U_{c,i}$	U-value of ceiling/roof element [W.m <sup>-2</sup> .K <sup>-1</sup> ]
	$U_{f,e}$	Effective Floor u-value [W.m <sup>-2</sup> .K <sup>-1</sup> ]	Average u-value of floor elements corrected for adjacency relations and weighted by the corresponding areas	$U_{f,e} = \frac{\sum(U_{f,i} \times A_{f,i} \times f_{t,i})}{A_e}$	$A_{c,i}$	Area of ceiling/roof element [m <sup>2</sup> ]
Operational parameters					$U_{f,i}$	U-value of floor element [W.m <sup>-2</sup> .K <sup>-1</sup> ]
					$A_{f,i}$	Area of floor element [m <sup>2</sup> ]
	$O_u$	Annual use fraction	Fraction of time the building is used annually	$O_u = \frac{\sum(t_{use,a,i} \times f_{v,i})}{t_a}$	$t_{use,a,i}$	Annual operation hours [h.a <sup>-1</sup> ]
					$f_{v,i}$	Associated volume fraction
					$t_a$	Total hours in a year [h.a <sup>-1</sup> ]
	$O_{d/u}$	Daytime use intensity	Ratio of the daytime use hours to total use hours	$O_{d/u} = \frac{\sum(t_{day,a,i} \times f_{v,i})}{\sum((t_{day,a,i} + t_{night,a,i}) \times f_{v,i})}$	$t_{day,a,i}$	Annual daytime use hours [h.a <sup>-1</sup> ]
					$t_{night,a,i}$	Annual nighttime use hours [h.a <sup>-1</sup> ]
	$O_d$	Annual daytime use fraction	Fraction of time the building is used during daytime, annually	$O_d = \frac{\sum(t_{day,a,i} \times f_{v,i})}{t_a}$		
	$O_n$	Annual nighttime use fraction	Fraction of time the building is used during nighttime, annually	$O_n = \frac{\sum(t_{night,a,i} \times f_{v,i})}{t_a}$		
	$q_{i,h}$	Weighted average area related internal gains rate [W.m <sup>-2</sup> ]	Average rate of internal gains per unit of floor area weighted by the share of every usage of the total area	$q_{i,h} = \sum((q_{i,h,i} + q_{i,l,i}) \times f_{v,i})$	$q_{i,h,i}$	Usage-based internal gains rate during the heating season [W.m <sup>-2</sup> ]
					$q_{i,l,i}$	Usage-based internal gains rate from lighting [W.m <sup>-2</sup> ]
	$Ig_d$	Daily area related internal gains [Wh.m <sup>-2</sup> .d <sup>-1</sup> ]	Daily internal heat gains per unit of area during the heating season	$Ig_d = \sum((q_{i,h,i} + q_{i,l,i}) \times f_{v,i} \times t_{use,d,i})$	$t_{use,d,i}$	Daily use hours [h.d <sup>-1</sup> ]
	$n_v$	Weighted average hourly air-change rate [h <sup>-1</sup> ]	Average air-change rate weighted by the share of every usage of the total volume	$n_v = \sum(n_{v,i} \times f_{v,i})$	$n_{v,i}$	Usage-based hourly air-change rate [h <sup>-1</sup> ]
	$Ac_d$	Daily air-change rate [d <sup>-1</sup> ]	Daily air-change rate	$Ac_d = \sum(n_{v,i} \times f_{v,i} \times t_{use,d,i})$		

### **5.3. Classification Algorithms and Tools**

In order to incorporate a larger number of classification criteria, while maintaining control over the number of the resulting groups, the present effort proposes the employment of Multivariate Cluster Analysis (Hair et al. 2010) for stock classification. Multivariate Cluster Analysis (MCA) is a fundamental technique in exploratory data-mining. It is concerned with dividing a multivariate data space into natural clusters or homogeneous groups of objects. MCA is used with an assumption that the object come from a number of distinct populations, but that there is no a priori definition of these populations (The Pennsylvania State University 2017). It has been widely employed in various fields of science including medical studies and market research. Its potential towards building stock classification however, has not been sufficiently explored.

Various techniques and algorithms have been defined to date, for identification of clusters in a data set. Most of these methods can be classified under one of the two main families of MCA methods: i) Hierarchical approaches, and ii) Partitional or relocation approaches. The main distinction between these families is whether or not the identified clusters are nested. Hierarchical clustering methods, which proceed by stages, produce a nested sequence of partitions organized as a tree, through a divisive or an agglomerative procedure. Partitional or relocation clustering methods iteratively move objects from one cluster to another, starting from an initial partitioning schema, such that an objective function is minimized (Wilson et al. 2002; Fraley & Raftery 1998).

In the current project, a classic and well-known algorithm from each family has been investigated for its potential toward efficient building stock classification. Considered algorithms are the hierarchical agglomerative clustering method (Hair et al. 2010) and the k-means method from the partitional family (MacQueen 1967). Later a third method, Model-based clustering (Fraley & Raftery 1998, 2002), was included due to its different approach to determining the number of emerging clusters. This approach incorporates both hierarchical and relocation techniques for a clustering based on probability distributions.

As briefly mentioned before, the generation of the energy relevant building stock representation and the classification and sampling process are both facilitated through a GIS plug-in, developed for the purpose. The Python-based plug-in employs several packages from R statistical software (The R Foundation 2017) to perform the cluster analysis and select the most appropriate partitioning. R is a free software environment specialized in statistical computing and graphics. A large community of statisticians contribute to the R environment through development of task-oriented packages. The rpy2



package (Gautier & rpy2 contributors 2014) for python makes it possible to initiate the R engine and call its various functions and packages within a Python code.

The values of the selected classification criteria are computed for each building and stored as a Comma Separated Values (CSV) file under the address provided by the user in the Python code of the plug-in. Due to the dependence of some of the adopted algorithms on the Euclidean distance between data points, the outcome of these methods is highly sensitive to the magnitude of the values representing the clustering criteria. For instance, parameters such as net volume, which are expressed in larger numerical values, can dominate the clustering process, overshadowing other characteristics such as effective U-value, which assume much smaller numerical values. In order to prevent such unintended skewing of the results, the data set is subjected to standardization prior to the cluster analysis (Equation 20).

$$z_{d,i} = \frac{x_{d,i} - \overline{x_d}}{\sigma_d} \quad (20)$$

, where  $x_{d,i}$  is the raw value of the  $d$  dimension of the  $i$  observation (in the unit of  $d$ ),  $z_{d,i}$  is the standard score of  $x_{d,i}$  (unit-less),  $\overline{x_d}$  and  $\sigma_d$  are the mean and standard deviation of the values of the  $d$  dimension (in the unit of  $d$ ).

A new CSV file is thus generated, which includes the standardized values of the classification criteria. This file is called as an input to the R packages responsible for the cluster analysis and partitioning of the data.

Note that a potential advantage of MCA methods is the possibility to apply weighting schemas to the input variables in order to strengthen the influence of one or more criteria in the clustering scheme. This would be useful in studies focused on a certain feature of the buildings (e.g. solar gains or behavioral aspects). In such cases, fine tuning the MCA process would result in a classification more sensitive to variables, which affect those particular building aspects (e.g. Effective glazing ratio or operational parameters). The current research is, however, not concerned with exploring the potential of weighting schemas for more efficient clustering, due to a lack of supporting information for such a process.

The thorough description of the algorithms, as well as an introduction to the statistical analysis tools adopted for the purpose of the cluster analysis and partitioning follows.

### K-Means Clustering

In this method, the desired number of clusters,  $k$ , is a prerequisite of the clustering process. To initiate the process,  $k$  arbitrary (or pre-determined) data points are selected as the initial seeds of the intended clusters. The data space is then partitioned into  $k$  clusters around these seeds, such that the data points in each cluster are closer to the seed of their own cluster than to the seeds of any other clusters. Then, the centroids of the generated clusters are computed (Equation 21), and used as seeds for the next iteration. This process continues until a convergence is reached, in other words, the data points stop shifting clusters. This process can be expressed as follows:

```

Initiate the seeds of the clusters (through random selection)
Attribute each data point to the closest cluster
Until assignments no longer change:
    Recalculate the centroids of clusters as seeds
    Re-distribute points among clusters based on the new seeds
    
```

$$\vec{m}_A = \frac{1}{n_A} \sum_{x_i \in A} \vec{x}_i \quad (21)$$

, where  $A$  is a cluster,  $\vec{x}_i$  is a multi-dimensional vector belonging to the cluster  $A$ ,  $\vec{m}_A$  is the centroid of the cluster  $A$ , and  $n_A$  is the number of data points within this cluster.

Various metrics can be adopted to measure the distances between data points and cluster centroids in the k-means method. In the present effort, the squared Euclidean distance (Equation 22), has been selected following the most common algorithm for this clustering method, known as the Lloyd's algorithm (Lloyd 1982).

$$dist(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|^2 = \sum_{i=1}^d (x_i - y_i)^2 \quad (22)$$

, where  $dist(\vec{x}, \vec{y})$  is the Euclidean distance between the vectors  $\vec{x}, \vec{y}$ , and  $x_i, y_i$  are the  $i^{th}$  dimensions of the afore-mentioned vectors.

As such, the entire process can essentially be expressed as an optimization problem, in which the objective function to be minimized is the sum of within cluster sums of squared

error. If the data space is the partitioned into a set of  $k$  clusters  $S = \{S_i | 1 \leq i \leq k\}$ , the objective function is as follows (Equation 23):

$$\arg \min_S \sum_{i=1}^k \sum_{\vec{x}_j \in S_i} \|\vec{x}_j - \vec{m}_{S_i}\|^2 = \arg \min_S \sum_{i=1}^k \sum_{\vec{x}_j \in S_i} \sum_{l=1}^d (x_{j_l} - m_{S_{i,l}})^2 \quad (23)$$

, where  $k$  is the number of clusters,  $S_i$  is an arbitrary cluster,  $\vec{x}_j$  is a member of a cluster,  $x_{j_l}$  the dimension  $l$  of the vector  $\vec{x}_j$ ,  $\vec{m}_{S_i}$  is the centroid of the cluster  $S_i$ ,  $m_{S_{i,l}}$  the dimension  $l$  of the centroid, and  $d$  the number of dimensions of the data space.

Figure 17 depicts the application of this method to a sample of two-dimensional points.

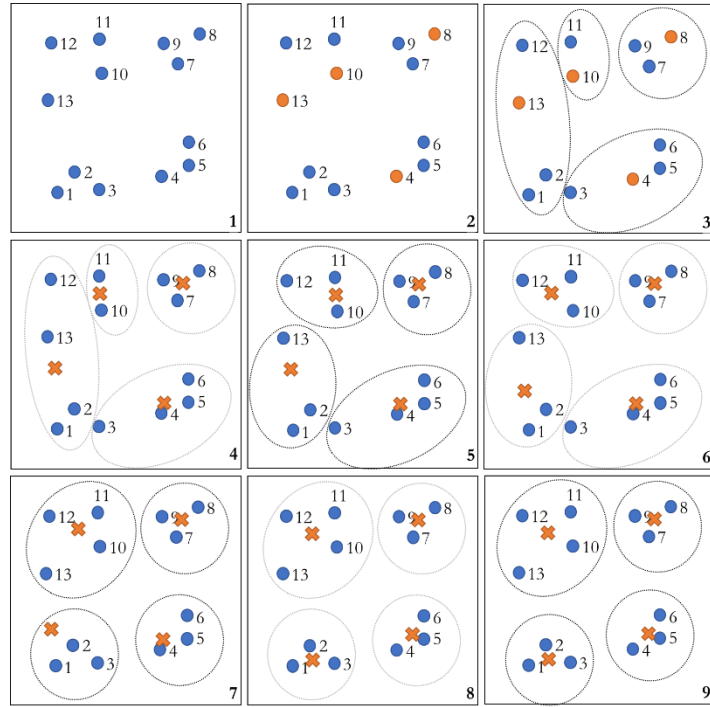


Figure 17 The steps of  $k$ -mean clustering method depicted with a sample of two-dimensional data points. In this example, the desired number of clusters is four, and the initial seeds are selected at random.

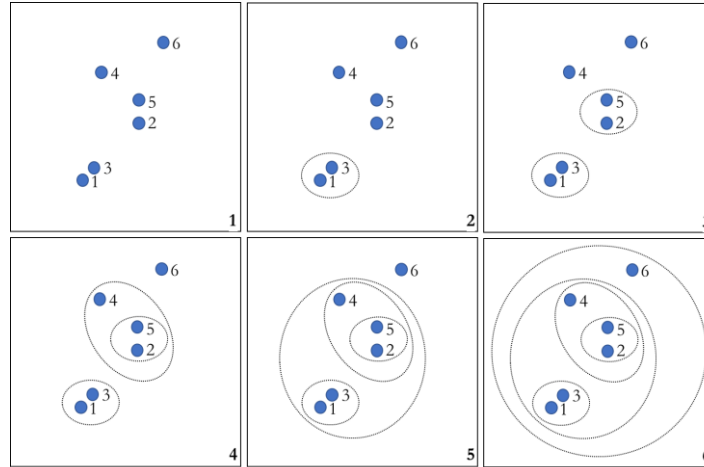
### Hierarchical Agglomerative Clustering

In this method, customarily the process starts with each data point constituting a cluster with a single member (although it is possible to initiate the process from a coarser partition). The closest pair of clusters are consecutively merged until only one cluster remains. The key operation of this clustering technique is the computation of the proximity (distance) between two clusters. The algorithm can be expressed as follows:

```

Start with every instance as a cluster
Compute the proximity matrix for the entire dataset
Until only one cluster remains:
    Merge the closest pair of clusters
    Update the proximity matrix for the new clustering schema
    
```

Figure 18 illustrates the steps of this method, applied to a small sample of two-dimensional objects (points).



*Figure 18 Consecutive steps of hierarchical agglomerative clustering depicted on a sample of six points on the two-dimensional plane.*

In the current research the adopted proximity measure, which determines the pair of clusters to merge at every step is the Ward's method (Ward 1963). The Ward's method, assumes that each cluster is represented by its centroid (Equation 21): a virtual data point with mean values across the cluster for every dimension. It measures the proximity between two clusters in terms of the increase in sum of squared errors (SSE) resulting from the merging of the two clusters. In other words, for any arbitrary pair of clusters the sum of

SSEs (the sum of squared Euclidean distances between each member of the cluster and the cluster centroid) is smaller than the SSE, when the two clusters are merged. The Ward's method measures the difference in the sum of the SSEs of the initial two clusters and the SSE of the emerging cluster (resulting from the merging of the two) as the inconsistency resulting from the merging of the two clusters. It then opts for the merge that minimizes the inevitable increase in inconsistency (Equation 24).

$$\begin{aligned}\Delta(A, B) &= \sum_{x_i \in A \cup B} \|\vec{x}_i - \vec{m}_{A \cup B}\|^2 - \sum_{x_j \in A} \|\vec{x}_j - \vec{m}_A\|^2 - \sum_{x_k \in B} \|\vec{x}_k - \vec{m}_B\|^2 \\ &= \frac{n_A \times n_B}{n_A + n_B} \|\vec{m}_A - \vec{m}_B\|^2\end{aligned}\quad (24)$$

, where  $A, B$  are two clusters,  $\Delta$  is the *merging cost* or the distance between two clusters,  $\vec{x}_i$  is a multi-dimensional vector belonging to a cluster,  $\vec{m}$  is the centroid of a cluster, and  $n$  is the number of data points within a cluster.

The linkages in hierarchical agglomerative clustering can be graphically visualized through a dendrogram. Figure 19 illustrates the dendrogram corresponding to the example provided in Figure 18. The merging steps can be traced back to arrive at a partitioning of the data space, which includes the desired number of clusters. This can be done simply by cutting through the dendrogram at a certain height (Figure 20). As such, the acceptable range for the number of clusters should be known before a decision can be made as to the height at which the dendrogram can be cut. However, deciding which cut provides a more efficient clustering is not a trivial problem.

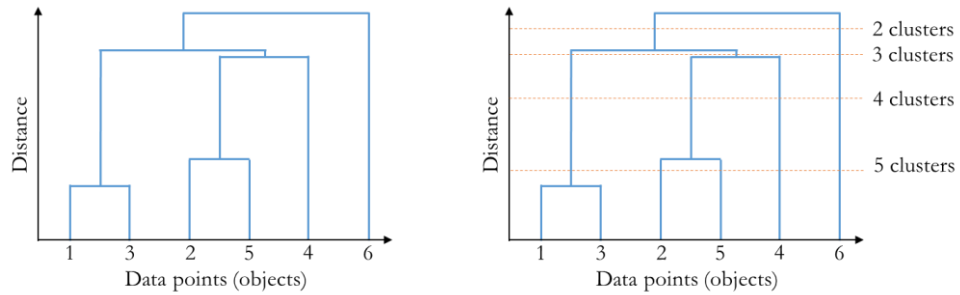


Figure 19 (Left) Dendrogram pertaining to the example illustrated in Figure 18.

Figure 20 (Right) Possible partitioning schemas resulting in the identification of 2, 3, 4, and 5 clusters.

Determining the optimal number of clusters is not a trivial problem. Assuming that each cluster is to be represented through at least one building, the upper limit for the number of the clusters depends on the computational resources and time available for the performance assessments. A lower limit can also be set to ensure a minimum of diversity in the resulting sample. But the hierarchical agglomerative and k-means methods can generate as many clustering schemes as the numbers within the given range. A method is required to identify the optimal number of clusters from the defined range, in view of the quality of the emerging clustering scheme.

In order to quantify the quality of the clustering schemes various approaches have been identified. *Internal* cluster validity approaches relying on “relative criteria” aim at finding the best clustering scheme that a clustering algorithm can generate under certain assumptions about algorithm parameters (Halkidi et al. 2002). The term internal, used to characterize such methods refers to the independence of these methods to outside information on the expected characteristics or labeling of the classes. This is in contrast to external methods, which rely on a prior knowledge on the natural clusters existing in the dataset. The internal approach is deemed appropriate to select the optimal clustering scheme emerging from each clustering algorithm when fed with different values from a predefined range for the number of clusters. For this purpose, the clustering algorithm is run for all values within the given range and the resulting clustering schemes are compared. To enable this comparison, various performance indicators have been defined and introduced by statisticians. These indicators seek to represent the intra-cluster similarity and inter-cluster dissimilarity of each partitioning scheme in terms of a single number. The Dunn’s index (Dunn 1974), for instance, is composed of the largest cluster diameter and the shortest distance between pairs of clusters in a scheme. It identifies the optimal partitioning as the one that minimizes the former and maximizes the latter. The SD index (Halkidi et al. 2000) relies on the average scattering within clusters and the total separation between clusters, identifying a “good” clustering scheme as the one that minimizes the index, thereby insuring well separated and compact clusters. Various other clustering quality assessment methods have been introduced in literature (Rousseeuw 1987; Davies & Bouldin 1979; Tibshirani et al. 2001; Halkidi et al. 2002; Maulik & Bandyopadhyay 2002; Jung et al. 2003; Liu et al. 2010; Charrad et al. 2014). To facilitate the selection of the optimal clustering scheme, Charrad et al. (2014) have developed an algorithm that computes the value of over 20 clustering quality indicators, across a given range of cluster numbers, and yields the clustering scheme suggested by the majority vote. This algorithm, developed as a package, *NbClust*, for the R software (Charrad et al. 2015), is capable of performing hierarchical agglomerative clustering with Ward’s method, as well as k-means clustering with squared Euclidean distance, intended in this study.

### **Model-based Clustering**

In model-based clustering (Fraley & Raftery 1998, 2002), probability models are used as a basis for cluster analysis. In this approach, the data is assumed to come from a mixture of probability distributions, with each distribution representing a cluster. The components of this mixture distribution (the underlying density functions) are assumed to belong to a family of parametric density functions. As such, the clustering problem is recast as a model choice problem, the objective being the identification of the optimal number of model components, as well as the parameters defining those components. For this purpose, a two-fold method composed of hierarchical and relocation steps is utilized.

The maximum number of model components,  $M$ , (or the maximum acceptable number of clusters) as well as the parametrization schema of the underlying components (the parametric family of density functions to which the components belong) is selected. The algorithm used in the present research is based on multivariate normal distributions. Several parametrizations of this family with various levels of complexity are considered, such as spherical or ellipsoidal parametrizations of equal or varying size, shape, and orientation (Fraley & Raftery 2002).

To initiate the process, a hierarchical agglomerative clustering of the data is performed, and all partitioning schemas leading to 2 to  $M$  clusters are identified. This step is essentially similar to the hierarchical agglomerative clustering explained in the previous section, however, the criterion for merging groups in this case is the Maximum Likelihood (ML). In intuitive language, Maximum Likelihood Estimation (Wilks 1938), is a method for estimating the values of the parameters of a statistical model based on some observations, such that the likelihood of generating those observations with the parametric model is maximized. In other words, it is a method of fitting a statistical model to the available set of observations. This criterion can be used for hierarchical agglomerative clustering, to ensure that at every step, the selection of the clusters to merge, is based on maximization of the likelihood of the emerging statistical model representing the observations.

The resulting partitioning schemas including between 2 and  $M$  clusters, are then used to initiate a relocation process. This process is similar to the k-means method, but with a different objective function. The relocation process is performed based on Expectation Maximization, EM (Dempster et al. 1977). In this process, an arbitrary assignment of observations to clusters is considered (resulting from the application of the first step). Then the likelihood function of the model parameters representing the data set is derived. Model parameters are recalculated such that this likelihood function is maximized. The observations are relocated among the clusters according to the current definition of the

underlying model components. This iterative process is continued until a convergence is reached. This process is repeated for all numbers of clusters smaller than  $M$  and for all parametrizations of the multivariate normal distribution family. At this point the optimal solution (the model parameters representing the clusters) for every scenario (pairs of number of clusters and parametrization) is achieved. As a final step, one of the resulting partitioning schemas should be selected. For this purpose, the Bayesian Information Criterion (BIC) is adopted (Schwarz 1978). BIC is a measure of how well a finite mixture model performs in predicting the observations. It is based on the maximum likelihood, and includes a penalty for the increase in the number of model parameters or the complexity of the adopted parametrization.

The model-based clustering method requires the maximum number of clusters allowed, and is equipped with logical processes, which yield the optimal clustering schema. This process as used in the present research can be expressed as follows:

```
Determine a maximum number of clusters as  $M$ 

Perform agglomerative hierarchical clustering based on likelihood
maximization, obtain classifications with up to  $M$  groups

For every classification:
    Until convergence is reached:
        Perform Expectation Maximization
        Compute the value of the Bayesian Information Criterion
    Select the model corresponding to the first decisive local maximum BIC
    value
```

To perform model-based clustering, the R package *mClust* (Fraley et al. 2015) has been used.

Once the partitioning scheme (for each method) is selected, representatives are chosen for each cluster. Currently, one representative per cluster is considered. The most typical building, defined as the building closest to the center of each cluster, is selected as the cluster representative. Proximity to center is computed through Euclidean distance. The center of the cluster in the k-means and hierarchical agglomerative methods is defined as a virtual data point with median values (across cluster) for each dimension. In the model-based method, the mean of each multivariate Gaussian distribution depicting a cluster is considered as the cluster center.



#### 5.4. Clustering Scenarios

Different combinations of the classification criteria are tested towards their efficiency in the partitioning of the data space into groups of buildings with similar energy behavior in terms of heating demand. Each combination includes some expression of the principle building features through inclusion of one or more indicators pertaining to that feature. However, one should bear in mind that the over-representation of a feature through inclusion of several associated indicators will skew the classification towards that specific feature. Also, consideration of a higher number of classification criteria leads to higher computational time, and may as well result in a noisier data space, reducing the chance of the identification of major traits and patterns. offers an overview of the various classification criteria sets considered in the various runs.

Table 7 Various sets of descriptive indicators considered as input for cluster analysis

	Geometry				Solar Gains	Thermal Quality				Operational Parameters								Number of parameters
	$V_n$	$A_e$	$C_t$	$h_e$	$GR_e$	$U_e$	$U_{w,e}$	$U_{c,e}$	$U_{f,e}$	$O_u$	$O_{d/u}$	$O_d$	$O_n$	$q_{i,h}$	$Ig_d$	$n_v$	$Ac_d$	
S1			✓	✓	✓	✓					✓				✓		✓	7
S2	✓		✓	✓	✓	✓					✓				✓		✓	8
S3			✓	✓	✓	✓				✓					✓		✓	7
S4	✓		✓	✓	✓	✓				✓					✓		✓	8
S5			✓	✓	✓	✓				✓	✓				✓		✓	8
S6	✓		✓	✓	✓	✓				✓	✓				✓		✓	9
S7			✓	✓	✓	✓						✓	✓	✓		✓		8
S8	✓		✓	✓	✓	✓						✓	✓	✓		✓		9
S9			✓	✓	✓		✓	✓	✓			✓	✓	✓		✓		10
S10	✓		✓	✓	✓		✓	✓	✓			✓	✓	✓		✓		11
S11	✓	✓		✓	✓		✓	✓	✓			✓	✓	✓		✓		12

Effective glazing ratio ( $GR_e$ ) and effective floor height ( $h_e$ ) are included in all scenarios, since there are no other indicators that express these building characteristics. Scenarios 1 to 10 can be seen as five pairs of similar scenarios differing only in the inclusion of net volume ( $V_n$ ) as an input parameter in the second scenario of each pair. The effective envelope area ( $A_e$ ) is only included in the last scenario, which is the only set that excludes thermal compactness ( $C_t$ ).

The main differences among the first four pairs of scenarios is in their depiction of operational parameters. S1 and S2 depict the operational parameters through daytime use intensity ( $O_{d/u}$ ), daily internal gains ( $Ig_d$ ), and daily air-change rate ( $Ac_d$ ) throughout a typical day. S3 and S4 are similar to the previous scenarios, however the use schedules are represented here through the fraction of the year the building is in use ( $O_u$ ). S5 and S6 combine the above scenarios. In S7 and S8 operational characteristics are represented through the annual daytime and nighttime use fractions ( $O_d, O_n$ ), along with average values of area-related internal gains ( $q_{i,h}$ ), and hourly air-change rate ( $n_v$ ).

S9 and S10 are similar to the former pair of scenarios in their expression of operational parameters, but incorporate a more detailed representation of the thermal quality of components. This is achieved through the consideration of wall, floor, and roof effective average U-values ( $U_{w,e}, U_{f,e}, U_{c,e}$ ) instead of the overall effective average envelope U-value ( $U_e$ ).

The combination of the eleven sets of input parameters and the three discussed clustering algorithms results in a matrix of 33 cluster analysis scenarios. All scenarios were performed to achieve 33 partitioning schemes. In each case, the buildings representing each cluster were identified. The emerging classifications were investigated through an external clustering quality evaluation test based on steady-state heating demand, to select the optimal classification routine for the project. For this purpose, the performance of the emerging set of representatives to predict the annual heating demand of the buildings in their associated cluster was evaluated.

### 5.5. Evaluation of Classification and Sampling Scenarios

Data matrices associated with each of the scenarios introduced in Table 4 are automatically generated by the developed plug-in and subjected to the above-mentioned MCA techniques leading to a total of 33 partitioning schemes. In order to select the most efficient clustering procedure, the performance of the resulting classifications has to be compared in view of the ultimate objective of the model: prediction of heating energy demand. The best clustering procedure (combination of criteria and clustering method, as such, is the one leading to the best representation of the energy performance of the buildings. Due to its dependency on external data, regarding the performance of buildings, this evaluation method falls within the class of external clustering evaluation methods. Due to data privacy issues, a collection of empirical data on the actual energy demand of the neighborhood for model evaluation purposes was not possible. A full-fledged simulation of the buildings within the scope of the case study, was also beyond the means of the current project. For the comparative evaluation of the clustering methods a simple steady-state computation of the monthly heating demand of the buildings was performed using the initially generated building representations. Although this data is not sufficient for a thorough evaluation of the performance of the samples in view of the prediction of the temporal dynamics of load patterns, it provides some hints towards the fittest clustering procedure. For the purpose of this analysis, the method suggested by the Austrian standard B 8110-6 for the computation of the monthly energy demand of buildings (Austrian Standards Institute 2014) was adopted. The essential input parameters of this method, which is based on a simple heat balance equation (Equation 25), are the same building characteristics considered for the cluster analysis. Naturally, this data could only be procured from the large scale available sources and as such is limited to the resolution and precision of these sources.

$$Q_h = (Q_T + Q_V) - \eta_h(Q_i + Q_s) \quad (25)$$

, where  $Q_h, Q_T, Q_V, Q_i, Q_s$  are the monthly heating demand, transmission losses, ventilation losses, internal gains and solar gains respectively [ $kWh.M^{-1}$ ],  $\eta_h$  is the monthly utilization factor for heat gains.

This computation was performed through a dedicated code developed in Python, which utilizes the urban representation generated in the previous step to extract the necessary input parameters. The various steps of this computation and the applied simplifications, necessary in view of informational limitations, are described below.

### Transmission Heat Losses

The utilized standard provides two methods for the computation of the transmission losses. The more elaborate method, requires detailed information on the condition of the unheated adjacent spaces and the nature of the existing thermal bridges. Since this information is unattainable at large scale, the more simpler strategy has been followed (Equation 26):

$$Q_T = \frac{1}{1000} \times L_T \times \Delta\theta \times t \quad (26)$$

, where  $L_T$  is the transmission heat transfer coefficient [ $W.K^{-1}$ ] (Equation 27),  $\Delta\theta$  is difference between internal temperature and average monthly external temperature [ $K$ ], and  $t$  is the duration of the month [ $h.M^{-1}$ ]. The latter two are provided in the standard B 8110-5 (Austrian Standards Institute 2011). Theoretically, the internal temperature is usage-dependent. However, the mentioned standard suggests the same internal temperature for all usages existing in the case study. As such no distinction is necessary among the transmission losses from various conditioned parts of the building.

$$L_T = \sum(U_i \times A_i \times f_{t,i}) + (L_\psi + L_\chi) = (U_e \times A_e) + (L_\psi + L_\chi) \quad (27)$$

, where  $(L_\psi + L_\chi)$  is the sum of the transmission heat coefficient due to linear and punctual thermal bridges [ $W.K^{-1}$ ]. The simpler method provided by the standard has been adopted in this case (Equation 28). Note that the first term in the calculation can be replaced by the effective envelope U-value multiplied by the effective envelope area (See Table 6).

$$(L_\psi + L_\chi) = 0.2 \times \left[ 0.75 - \frac{(U_e \times A_e)}{\sum A_i} \right] \times (U_e \times A_e) \geq 0.1 \times (U_e \times A_e) \quad (28)$$

, where the term  $\sum(U_i \times A_i \times f_{t,i})$  has been replaced by  $(U_e \times A_e)$  as in the previous equation.

### Ventilation Heat Losses

Since no information was available as to the existence and operation mode of mechanical ventilation devices, all building within the study area are assumed to be naturally ventilated (Equation 29). The ventilation losses depend on the operational characteristics of the spaces. Technically, the standard method has been envisaged for a single-usage building (or portion of a building). Therefore, the contribution of each building usage to the overall ventilation losses has to be considered if the same method is to be used for the estimation of the overall ventilation losses of a multi-use building.

$$Q_V = \frac{1}{1000} \times L_V \times \Delta\theta \times t \quad (29)$$

, where  $L_V$  is the ventilation heat transfer coefficient [ $W \cdot K^{-1}$ ] (Equation 30).

$$L_V = c_{vp,L} \times P_L \times v_V \quad (30)$$

, where  $c_{vp,L} \times P_L$  is the volumetric heat capacity of air =  $0.34 [Wh \cdot m^{-3} \cdot K^{-1}]$ , and  $v_V$  is the airflow [ $m^3 \cdot h^{-1}$ ] (Equation 31).

$$v_V = \sum \left( \frac{n_{v,i} \cdot t_{use,m,i}}{t} \times f_{v,i} \right) \times V_V \quad (31)$$

, where  $n_{v,i}$  and  $f_{v,i}$  are the energetically relevant air-change rate of a usage in the building [ $h^{-1}$ ], and its corresponding share of the overall building volume,  $t_{use,m,i}$  is the monthly use hours [ $h \cdot M^{-1}$ ], and  $V_V$  is the effective volume of the building [ $m^3$ ] (Equation 32). This equation has been modified to incorporate the contribution of all functions existing in the building to the building's airflow rate.

$$V_V = 0.8 \times n_f \times A_f \times 2.6 \quad (32)$$

, where the term  $n_f \cdot A_f$ , number of floors multiplied by the area of the building footprint [ $m^2$ ], has replaced the gross building floor area in the original equation.

### **Internal Heat Gains**

Similar to the ventilation heat losses, in the case of the internal gains, also dependent on the operational characteristics of spaces, various usages have to be factored in the computations. As such, the standard calculations have been modified to involve the contribution of various usages. Reference values for internal gains are commonly expressed per unit of floor area. Note that in the distribution of the volume of the buildings among usages, the potential differences in the heights of the different floors has been ignored due to a lack of data in this regard. As such one may assume that the same ratios apply for the association of the floor area to various usages (Equation 33).

$$Q_i = \frac{1}{1000} \times q_{i,h} \times 0.8 \times n_f \times A_f \times t \quad (33)$$

, where  $q_{i,h}$  is the average area-related internal gains rate weighted by the share of every building usage from the overall volume/area [ $W \cdot m^{-2}$ ] (See Table 6). This value includes internal gains from occupants, equipment, as well as lighting devices.

### **Solar Heat Gains**

According to the standard, the solar heat gains in the heating season are computed based on the area and orientation of the transparent building enclosures, according to the standard climate data. The method also includes an estimation of the shading received by the building. This is done through consideration of a shading factor, which incorporates estimations of the average shading received from various obstructions including other buildings, trees, and shading elements such as overhangs and fins. The standard provides default values for the shading received by single family houses and row houses but no simplified method is suggested for the computation of the shading factor in the dense and diverse fabric of the city.

As such, in the present project, this factor has been approximated by the Sky View Factor, which has been argued to provide a good estimation of the shading received on a building element. As such, the solar gains can be computed according to Equation 34.

$$Q_s = \sum_j (I_{s,j} \times \sum_i A_{trans,i,j}) \quad (34)$$

, where  $I_{s,j}$  is the monthly global irradiance on a surface with orientation  $j$  [ $kWh.m^{-2}.M^{-1}$ ], according to the reference climate data provided in Austrian standard B 8110-5 (Austrian Standards Institute 2011),  $A_{trans,i,j}$  is the effective area of a transparent building component with orientation  $j$  [ $m^2$ ], (Equation 35).

$$A_{trans,i} = WWR \times GWR \times A_{ow,i} \times SVF_i \times g \quad (35)$$

, where  $WWR$  is the default value for window to wall ratio ( $=0.15$ ),  $GWR$  is the default value for glazing to window ratio ( $=0.7$ ) (Austrian Standards Institute 2014),  $g$  is the default value for solar energy transmittance coefficient ( $=0.67$ ) (Österreichisches Institut für Bautechnik 2015), and  $A_{ow,i}$ ,  $SVF_i$  the area and corresponding Sky View Factor of exposed wall elements.



### Utilization Factor for Heat Gains

This utilization factor is defined to include the impact of the building's thermal mass in the effective benefitting from the internal and solar heat gains (Equation 36).

$$\eta_h = \begin{cases} \frac{1-\gamma_h^a}{1+\gamma_h^{a+1}} & \text{if } \gamma_h \neq 1 \\ \frac{a}{1+a} & \text{if } \gamma_h = 1 \end{cases} \quad (36)$$

, where  $\gamma_h$  is heat balance ratio in the heating season (Equation 37), and  $a$  is a numerical parameter for the utilization factor (Equation 38).

$$\gamma_h = \frac{Q_i + Q_s}{Q_T + Q_V} \quad (37)$$

$$a = a_0 + \frac{C}{\tau_0 \times (L_T + L_T)} \quad (38)$$

, where  $C$  is the effective heat capacity of the building [ $Wh.K^{-1}$ ] (Equation 39),  $a_0$  is a dimensionless reference numerical parameter (=1), and  $\tau_0$  is the reference time constant = 16 [h].

$$C = f_{BW} \times V \quad (39)$$

, where  $f_{BW}$  is the volumetric heat capacity of the building [ $Wh.m^{-3}.K^{-1}$ ] according to Table 8.

Table 8 Values considered for the volumetric heat capacity of buildings according to construction period.

$f_{BW}$ [Wh. K <sup>-1</sup> ]	Building type	Associated construction period
20	Mid-weight constructions	Post 1976
30	Massive constructions	1945-1976
60	Very massive constructions	Pre 1945

The monthly values of heating demand are summed to obtain the annual heating demand of each building within the study area. This measure is used as an external objective to evaluate the quality of the various partitioning schemes emerging from the applied clustering scenarios. For this purpose, in each partitioning scheme, the volumetric heating demand of the representative of each cluster was used to predict the heating demand of each building within the cluster (Equation 40).

$$Q_{predicted,i,j} = V_{n,i} \times Q_{v,rep,j} \quad (40)$$

, where  $Q_{predicted,i,j}$  is the predicted annual heating demand of the building  $i$ , belonging to the cluster  $j$  [kWh. a<sup>-1</sup>],  $V_{n,i}$  is the net volume of the building [m<sup>3</sup>], and  $Q_{v,rep,j}$  is the volumetric annual heating demand of the building representing the cluster  $j$  [kWh. m<sup>-3</sup>. a<sup>-1</sup>].

The results are compared to the theoretical annual demand values computed for each building according to the previously explained method. The following metrics were defined to assess the performance of each clustering scheme:

- Relative error in the prediction of the annual heating demand of the entire neighborhood, as a measure of how well the sample represents the aggregate heating demand of the urban area under study (Equation 41).

$$\delta_{Total} = \frac{|\sum Q_{predicted,i} - \sum Q_{theoretical,i}|}{\sum Q_{theoretical,i}} \times 100 \quad (41)$$

, where  $\delta_{Total}$  is the relative neighbourhood-level error [%], and  $Q_{theoretical,i}$  is the theoretical annual heating demand computed for the building  $i$  [kWh. a<sup>-1</sup>].

- Mean relative error of building level annual heating demand prediction, as an indicator of the predictive abilities of the sample with regard to individual buildings (Equation 42).

$$\delta_{Mean} = \frac{1}{n} \times \sum \delta_i \quad (42)$$

, where  $\delta_{Mean}$  is the mean building-level error [%],  $\delta_i$  is the relative prediction error of building  $i$  [%] (Equation 43), and  $n$  is the number of buildings within the study area.

$$\delta_i = \frac{|Q_{predicted,i} - Q_{theoretical,i}|}{Q_{theoretical,i}} \times 100 \quad (42)$$

Fraction of the total volume associated with a relative error of above 20% in the prediction of the annual heating demand, to identify the sampling scheme resulting in the least amount of severe prediction errors at buildings level (Equation 44).

$$f_{>20\%} = \frac{(\sum V_i | \delta_i > 20\% )}{\sum V_i} \times 100 \quad (44)$$

, where  $f_{>20\%}$  is the fraction of the volume associated with a building-level prediction error of above 20%, [%].

## 5.6. Results and Discussion

### Evaluation of Cluster Analysis Scenarios

As explained previously, all steps necessary for the generation of an energy-relevant representation of the urban building stock, extraction of the values of the relevant descriptive indicators, multivariate cluster analysis, and selection of the buildings best representing the emerged classes, are performed through a GIS plug-in developed for the QGIS open platform. The final outcome of the operation of the reductive module plug-in for each run, is a map of the investigated urban area, in which clusters are identified by colors (See Appendix 1), a CSV file containing the values of the adopted descriptive indicators with an additional column associating each building with a cluster, as well as a list of buildings representing the clusters. On an average commercial PC, the application of the reductive plug-in on the current case study requires about 15 minutes. In each case, the volumetric heating demand of the representative buildings was used to compute the annual heating demand of the buildings in the associated clusters. Then, the previously-introduced performance indicators were computed for each partitioning schema.

The performance of the three cluster analysis methods regardless of the input parameters is compared in Figure 21, where the ranges of the performance indicators are shown for each method across all 11 input parameter scenarios. As seen in this figure, the three cluster analysis algorithms perform rather similarly with regard to the prediction of the aggregate demand of the neighborhood.

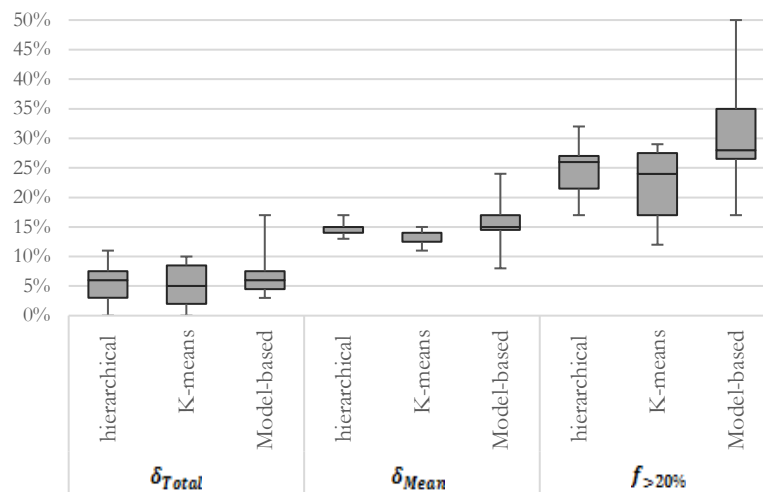


Figure 21 Comparison of the performance of the three clustering algorithms regardless of input parameters.

Except in two cases, the  $\delta_{Total}$  lies below 10 percent, which is an acceptable deviation from the theoretical values. The above 10% deviations pertain to the hierarchical agglomerative and the model-based methods, with a maximum  $\delta_{Total}$  of 11% and 17% respectively. The k-means method displays a slightly lower median deviation in the prediction of the aggregate heating demand of the neighborhood. The k-means clustering method performs consistently better than the two other algorithms in the building level prediction of heating demand. The mean deviation of the building level predictions ( $\delta_{Mean}$ ) of the k-means algorithm from the theoretical values lies below 15%, and less than 25% of the volume of the neighborhood, in average, is associated with a severe building level prediction error ( $f_{>20\%}$ ). The model-based clustering algorithm displays the weakest performance in building level demand prediction, with as much as 50% of the total volume associated with severe errors in the worst case. The hierarchical agglomerative clustering performs rather well with regard to the  $\delta_{Mean}$ . However, it results in a noticeably larger portion of the neighborhood being affected with high prediction errors, compared to the k-means clustering algorithm. As such, regardless of the set of descriptive indicators selected to represent buildings, the k-means method seems to perform best in capturing and representing the diversity of the building stock with regard to annual heating demand.

Figure 22 illustrates the impact of various sets of descriptive indicators on the quality of the resulting partitioning, without considering the clustering method. In this graph, the three performance indicators have been plotted for each scenario (in terms of input variables).

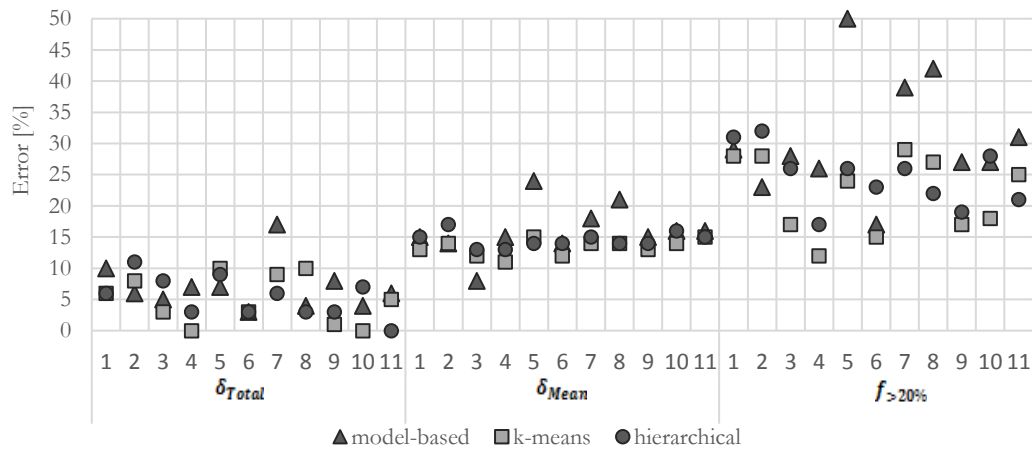


Figure 22 Comparison of the performance of the various sets of performance indicators on the quality of the emerging partitioning and building representation, regardless of the clustering algorithm.

With regard to the prediction of the aggregate heating demand of the neighborhood, with the exception of scenarios 2 and 7, all scenarios result in a deviation of 10% or less from the theoretical values. However, scenarios 3, 4, 6, 9, 10, and 11 perform better than the others, as in these cases, in a majority of cases the deviations lie below 5%. A comparison of the pairs of similar scenarios (See Clustering Scenarios) reveals that the inclusion of volume as a clustering criteria, in almost all cases, results in a better prediction of the aggregate heating demand. It's effect on the prediction of building level demand, however, is not consistently positive. At building level, scenario 3 produces the best result with regard to average deviation from the theoretical values, whereas the best performance with regard to the share of the volume associated with severe errors pertains to scenario 4. Scenarios 1, 4, 6, and 9 are the next best performing scenarios based on  $\delta_{Mean}$ , but scenario 1 results in the misrepresentation of a much larger share of the total volume. An increase in the number of parameters considered for the cluster analysis does not necessarily improve the quality of the classification. This may be due to the extra noise introduced by the inclusion of new, yet insignificant and interrelated variables. As mentioned before, the 33 investigated partitioning schemas, lead to the identification of a range of 6 to 21 clusters. Figure 23 illustrates the relationship between the number of identified clusters (and consequently representative buildings) in the representation of the heating demand of the neighborhood. As apparent from the figure, a higher number of representative buildings does not guarantee a more successful representation of the neighborhood, whereas, partitioning schemas with fewer clusters can perform reasonably well in capturing the energy characteristics of the buildings in the neighborhood.

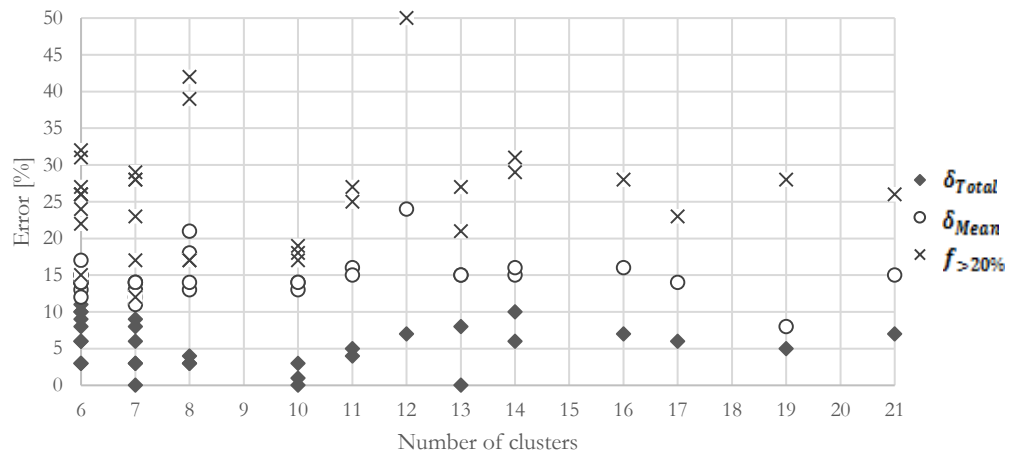


Figure 23 Comparison of the performance of the partitioning schemas in view of the number of the buildings representing the neighborhood in each case.

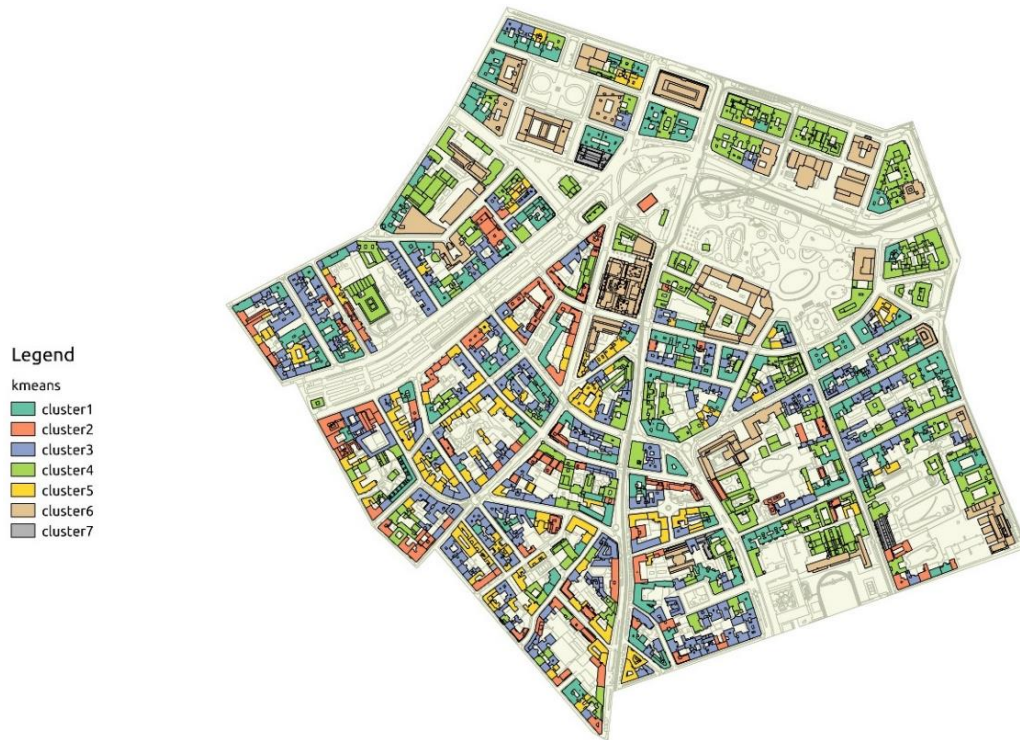
Table 9 displays the values of the performance indicators for all partitioning schemas. All schemas resulting in values below 10%, 15%, and 20% for the  $\delta_{Mean}$ ,  $f_{>20\%}$ , and  $\delta_{Total}$  are highlighted in the table, as the most successful clustering scenarios. Interestingly, none of the best performing schemas are associated with more than 10 classes. Note that the k-means method is responsible for a majority of these partitioning schemas. Scenario 4, with the k-means method, constituting 7 clusters, yields the best results in every aspect.

Table 9 An overview of the values of the performance indicators computed for the 33 partitioning scenarios

Clustering Scenario		Clusters #	$\delta_{Total}$	$\delta_{Mean}$	$f_{>20\%}$
S1	Hierarchical agglomerative	6	6%	15%	31%
	K-means	7	6%	13%	28%
	Model-based	14	10%	15%	29%
S2	Hierarchical agglomerative	6	11%	17%	32%
	K-means	7	8%	14%	28%
	Model-based	17	6%	14%	23%
S3	Hierarchical agglomerative	6	8%	13%	26%
	<b>K-means</b>	<b>7</b>	<b>3%</b>	<b>12%</b>	<b>17%</b>
	Model-based	19	5%	8%	28%
S4	<b>Hierarchical agglomerative</b>	<b>8</b>	<b>3%</b>	<b>13%</b>	<b>17%</b>
	<b>K-means</b>	<b>7</b>	<b>0%</b>	<b>11%</b>	<b>12%</b>
	Model-based	21	7%	15%	26%
S5	Hierarchical agglomerative	6	9%	14%	26%
	K-means	6	10%	15%	24%
	Model-based	12	7%	24%	50%
S6	Hierarchical agglomerative	7	3%	14%	23%
	<b>K-means</b>	<b>6</b>	<b>3%</b>	<b>12%</b>	<b>15%</b>
	<b>Model-based</b>	<b>8</b>	<b>3%</b>	<b>14%</b>	<b>17%</b>
S7	Hierarchical agglomerative	6	6%	15%	26%
	K-means	7	9%	14%	29%
	Model-based	8	17%	18%	39%
S8	Hierarchical agglomerative	6	3%	14%	22%
	K-means	6	10%	14%	27%
	Model-based	8	4%	21%	42%
S9	<b>Hierarchical agglomerative</b>	<b>10</b>	<b>3%</b>	<b>14%</b>	<b>19%</b>
	<b>K-means</b>	<b>10</b>	<b>1%</b>	<b>13%</b>	<b>17%</b>
	Model-based	13	8%	15%	27%
S10	Hierarchical agglomerative	16	7%	16%	28%
	<b>K-means</b>	<b>10</b>	<b>0%</b>	<b>14%</b>	<b>18%</b>
	Model-based	11	4%	16%	27%
S11	Hierarchical agglomerative	13	0%	15%	21%
	K-means	11	5%	15%	25%
	Model-based	14	6%	16%	31%

### **Best Performing Clustering Scenario and Associated Representative Buildings**

The clusters identified by the most successful classification schema (S4-Kmeans) are represented in Figure 24. This map is automatically generated by the plug-in at the end of its operation. Interestingly, most buildings with similar footprint geometry and size are clustered together.



*Figure 24 The visualized results of the reductive procedure for the S4-Kmeans clustering scenario.*

An overview of the mean values of the descriptive indicators adopted in scenario 4, across the emerging seven clusters is provided in Table 10.



Table 10 Mean values of the employed descriptive indicators across clusters S4-Kmeans clustering scenario.

Cluster	$\bar{V}_n$	$\bar{h}_e$	$\bar{C}_t$	$\overline{GR}_e$	$\bar{U}_e$	$\bar{O}_u$	$\overline{I}g_d$	$\overline{Ac}_d$	$\bar{Q}_v$	$Q_{v,rep,j}$
1	13249	5.03	3.52	0.034	1.373	0.95	0.09	10.68	43.93	44.03
2	8048	3.82	3.04	0.054	1.319	0.94	0.09	10.6	43.80	44.68
3	8188	4.77	2.89	0.031	1.407	0.97	0.09	10.18	52.92	45.53
4	11186	4.63	3.13	0.039	1.336	0.4	0.05	18.54	30.12	31.76
5	5616	3.81	2.71	0.031	1.326	0.97	0.09	10.15	53.53	52.76
6	51430	4.63	4.19	0.049	1.334	0.41	0.07	19	25.15	20.60
7	14590	3.89	2.87	0.069	0.726	1	0.09	9.6	23.73	21.87

The range of volumetric heating demand of the buildings in every cluster as well as that of the representing building is presented in Figure 25. As seen in this figure, the energy demand of clusters 3 and 6 is underestimated by the representative building. The representatives of other clusters, however, provide acceptable representativeness.

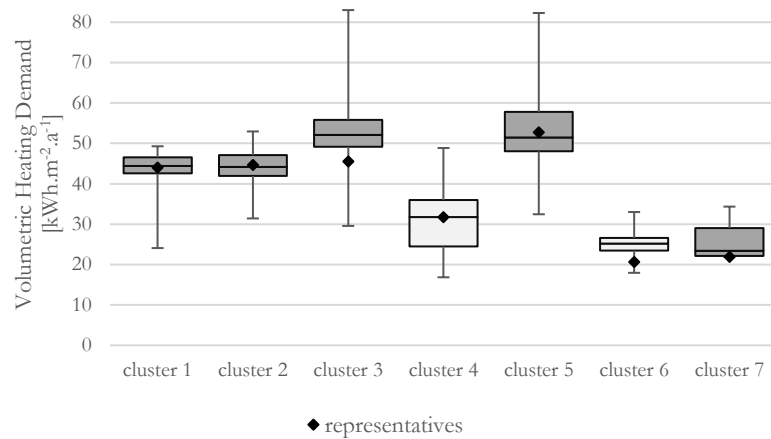


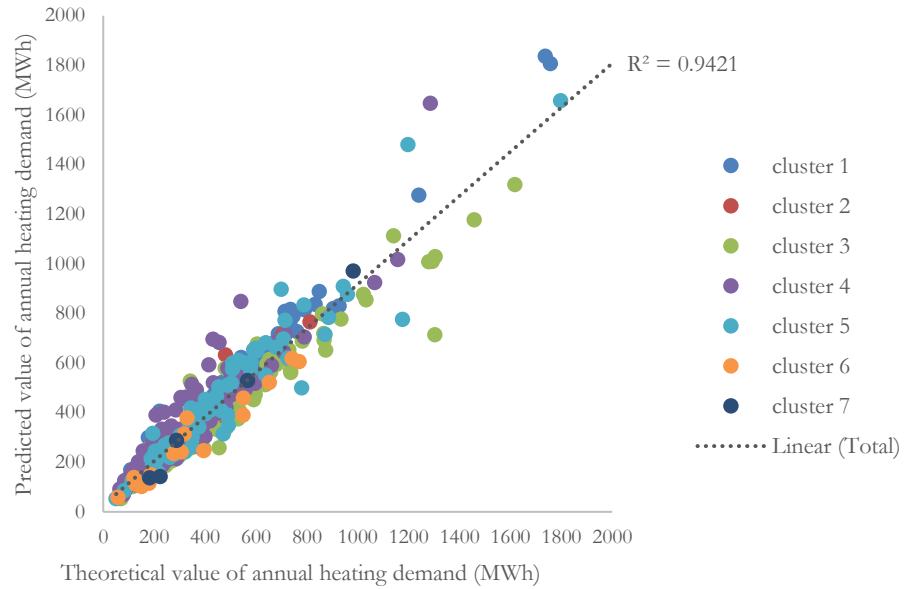
Figure 25 Volumetric heating demand of buildings in each cluster and the cluster representative. The lighter colored columns pertain to non-residential buildings

Table 11 provides an overview of the distribution of buildings by age, construction period, and volume in each cluster, as well as a brief introduction to the representative buildings. A more detailed description of representative buildings can be found in Appendix 2.

Cluster 1 is characterized by prominently sized buildings of mainly residential use with high ceilings, and an L-shaped or U-shaped foot print resulting in high exposure to outside air (lower compactness values). Cluster 2, also a mainly residential cluster, includes buildings

of medium size with relatively low compactness, also much less affected by mutual shading, due to their placement along wider streets or their adjacency to large courtyards. Due to their narrow, linear forms and high exposure to the sun, these buildings receive higher solar gains. Most buildings constructed after 1945 are clustered in this group, although buildings of other construction periods are also present. Clusters 3 and 5 feature a collection of smaller residential buildings. The major distinction between the two classes is in volume and average floor height. Cluster 3 features larger buildings with lower ratio of useful floor area to volume (higher ceilings).

Clusters 4 and 6 are mixes of educational, cultural, commercial and office buildings. Buildings in cluster 6 receive less shading and are considerably less compact than those in cluster 4. Cluster 4 is composed of mid-sized buildings, whereas cluster 6 includes the largest non-residential buildings in the area. Cluster 7 represents all buildings constructed after 1976, whose thermal performance is significantly superior to that of all other classes of building. Among the prominently residential clusters (1, 2, 3, 5, and 7), clusters 1 and 2 have a higher share of mixed use or non-residential buildings, whereas cluster 7 is strictly residential. All clusters, with the exception of cluster 7, are composed of buildings of various construction period. Figure 26 visualizes the accord between the building level predictions of the annual heating demand with the expected theoretical values.



*Figure 26 Theoretical against predicted values of annual heating demand based on the representation resulting from S4-Kmeans scenario.*

Table 11 An overview of the age and usage composition of the generated classes and the associated representative buildings (Part 1/2).

	# by construction period	# by primary usage	# by volume	Representative building	
<b>Cluster 1</b>				<b>Mühlgasse 7</b>  1914 Residential + Gastronomy	
<b>Cluster 2</b>				<b>Rechte Wienzeile 25</b>  1945-1976 Residential	
<b>Cluster 3</b>				<b>Preßgasse 17</b>  1846 Residential + Gastronomy	
<b>Cluster 4</b>				<b>Schwartzenberg-platz 16</b>  1868 Office	

Table 11 An overview of the age and usage composition of the generated classes and the associated representative buildings (Part 2/2).

<b>Cluster 5</b>				<b>Große Neugasse 16</b>  1848-1918 Residential	
<b>Cluster 6</b>				<b>Schillerplatz 4</b>  1872 Office	
<b>Cluster 7</b>				<b>Mattiellistraße 3</b>  2002 Residential	

### Efficiency of the Adopted Classification Compared to Conventional Approaches

An initial goal of the clustering-based reductive procedure was to achieve a more efficient method for the classification of the existing building stock, compared to the commonly adopted age-based classification. In order to demonstrate the insufficiency of a building's construction period to represent the overall thermal quality of a building, the effective envelope U-value of the buildings, as well as the computed heating demand are plotted against the construction year in Figure 27 and Figure 28 respectively. These graphs are generated based on 504 buildings in the dataset, for which the exact year of construction (and not merely the construction period) was known. This sample included no residential buildings constructed between 1945 and 1976, or non-residential buildings constructed after 1976. As seen in the graphs, in the case of buildings constructed prior to 1970's, no meaningful correlation can be found between the overall effective thermal quality or performance of the envelope and construction period. This suggests the inadequacy of construction period as an indicator of the thermal quality of the building. Newer buildings (post 1976), however, perform significantly better due to the drastic changes in the Viennese construction codes in 1976.

One may argue, however, that a combined use of construction period and usage can result in a reasonable classification. To compare the quality of the clustering-based classification schema, with the age/usage based method, the residential buildings in the study area were classified according to the construction period, based on the thresholds provided in the Austrian guidelines for the computation of the energy demand of historical buildings (Österreichisches Institut für Bautechnik 2015). Due to the small number of non-residential buildings in each usage and age category, this analysis was only performed for the residential buildings.

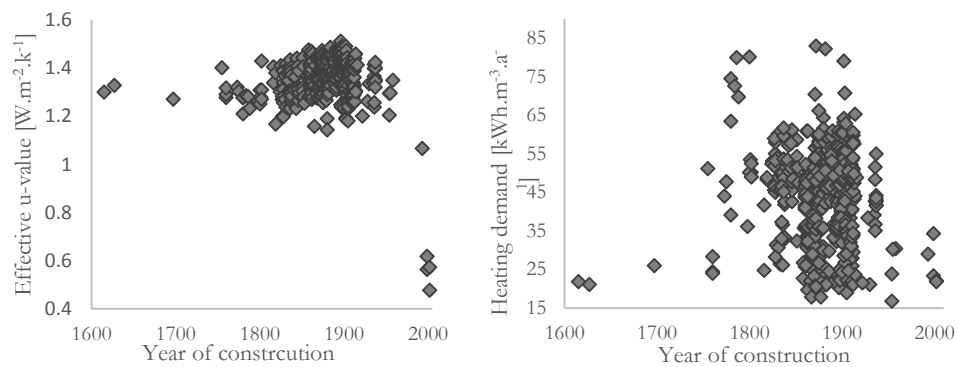


Figure 27 (left) Effective envelope U-value by year of construction.

Figure 28 (Right) Volumetric heating demand by year of construction.

The volumetric heating demand of the resulting classes are illustrated in Figure 29, alongside those of the primarily residential classes emerging from the developed reductive method. The comparison reveals a clear (though not optimal) improvement in the distinction among classes in the developed method. Note that the proposed classification method dedicates a separate cluster (cluster 7) to the buildings constructed after 1976, the only age class which has a significantly different performance compared to the other age groups. Even in this case, the clustering algorithm relocates the only building of this period, which performs worse than the others to a different class to provide a more homogeneous group. Clusters 1 and 2 are composed of buildings with similar volumetric annual heating demand, but these classes perform visibly better than classes 3 and 5. As such, these four clusters feature two distinct heating demand tendencies, whereas the first three age-based clusters are largely overlapping and not very efficient in terms of distinguishing better performing from worse performing buildings. Therefore, the developed method represents the diversity in building performance more reliably, even though each cluster is composed of buildings of various construction periods.

The performed analysis is of course strictly focused on the annual heating demand of buildings, which is not the optimal metric for the evaluation of a building classification schema, aimed at the representation of the temporal patterns of energy use in urban areas. It can however be relied on to offer some insight as to the comparative quality of the developed classification schemas.

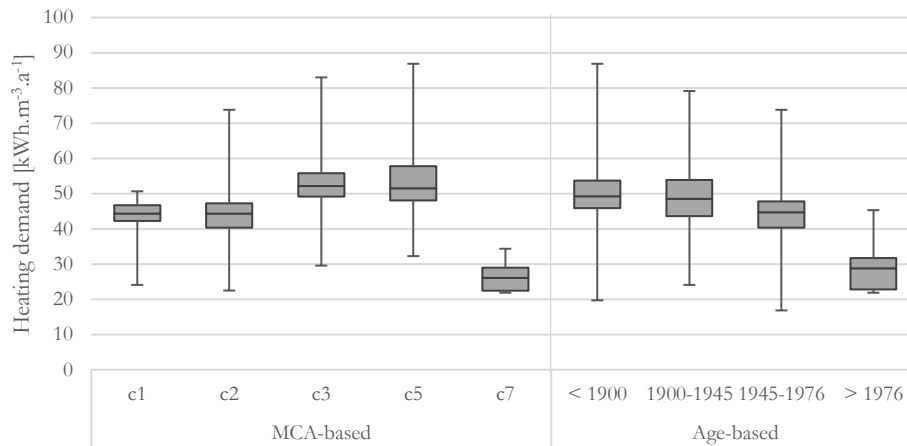


Figure 29 Ranges of volumetric heating demand values across classes emerging from the multivariate cluster analysis procedure adopted in the current effort.

# The Re-Diversification Module

## 6.1. Introductory Comments

Previous studies have shown that the acquisition of pertinent building information and generation of the geometric model of a building are the most time and effort intensive activities in building performance simulation (Mahdavi & El-Bellahy 2005). The developed reductive module, reduces this effort by way of limiting the modeling scope to a manageable number of buildings. As suggested by the simple evaluation performed in the previous phase, buildings selected through the reductive process provide a fair overall representation of the energy demand of the neighborhood. However, much of the diversity of the building stock has been lost in this representation. Loss of diversity is a natural consequence of a reductive process. Moreover, the use of reference schedules and default values for the representation of the operational parameters of the individually assessed representative buildings contributes further to the loss of diversity. The re-diversification module has been developed to reintroduce part of the lost diversity back to the computational model, and to obtain more realistic representations of the spatial and temporal distribution of demand.

The core concept of the re-diversification module is to use the representative buildings as a basis for the automated generation of a diverse set of models that better reflect the various characteristics of the building stock. Assuming that the identified sample of buildings represents the geometric features of the stock with acceptable fidelity, the re-diversification module attempts to readjust some of the non-geometric parameters of the reference simulation models, such that they emulate the characteristics of the represented buildings more closely.

## **6.2. Tools and Material**

Once the representative buildings are selected by the reductive module, reference simulation models are developed based on the detailed plans of these building, procured from the Viennese Building Police (Magistrat der Stadt Wien 2017a). The simulation program Energy Plus (National Renewable Energy Laboratory 2017) was used for the development of the simulation models of sample buildings. Energy Plus is an openly accessible whole building energy simulation program, capable of performing high-resolution building energy assessments, and detailed representation of the operational as well as physical properties of buildings. All required information for Energy Plus simulations is structured and stored in a text-based object-oriented Input Data File (IDF). The modeling process was facilitated by the open studio plug-in (NREL et al. 2017), which utilizes the SketchUp three dimensional modeling environment (Trimble Inc. 2017), to provide a graphical user interface for the development of IDF building models for simulation purposes. The re-diversification module, is developed in Python programming language and requires the IDF's of the above-mentioned reference models as input. The Eppy package for Python (Philip 2013) offers the possibility to browse and modify IDF objects through Python codes. This package was utilized to automate the diversification process.

For various usages present in reference buildings, occupancy, HVAC, lighting, and equipment schedules provided by the Standard 90.1 of the American Society of Heating, Refrigerating, and Air-conditioning Engineers (ASHRAE 2013) were adopted as reference. These reference schedules were readjusted to meet the requirements of Austrian Standards (ÖNORM 2014). For the definition of the layered composition of building components, base case assumptions of the Austrian Handbook for Building Thermal Retrofit (Schöberl et al. 2012) were used. These templates were adjusted such that the national guidelines for performance assessment of historical buildings (OIB 2015) with regard to thermal quality of components were met. Other adopted sources of information for this purpose included the Baubook (Baubook GmbH 2017) and MASEA (Fraunhofer 2017) web-based building product data bases.

Some descriptive indicators extracted in the previous step (See Table 6), were used to inform the diversification process. These include: the effective U-values of wall, roof/ceiling, and floor elements. Additionally, and for the same purpose, annual area-related internal gains and average hourly air-change rate values were extracted from the urban stock representation developed in the first phase of the project (Equation 45 and Equation 46).



$$Ig_{a,Building} = \sum((q_{i,h,i} + q_{i,l,i}) \times f_{v,i} \times t_{use,a,i}) \quad (45)$$

, where  $Ig_{a,Building}$  is the area-related annual internal gains of each building from all sources [ $Wh.m^{-2}.a^{-1}$ ],  $t_{use,a,i}$  is the number of annual hours the building is used [ $h.a^{-1}$ ],  $q_{i,h,i}$  and  $q_{i,l,i}$  are the standard internal gains rate from equipment, people and lighting during the heating season for each usage [ $W.m^{-2}$ ], and  $f_{v,i}$  is the fraction of the volume associated with each usage.

$$\overline{n_{v,Building}} = \frac{\sum(n_{v,i} \times f_{v,i} \times t_{use,a,i})}{t_a} \quad (46)$$

, where  $\overline{n_{v,Building}}$  is the weighted average air change rate of the building [ $h^{-1}$ ],  $n_{v,i}$  is the standard provided air-change rate of every usage [ $h^{-1}$ ], and  $t_a$  is the total annual hours [ $h.a^{-1}$ ].

The computed values, along with the effective U-values of the components are stored in a CSV file and used in the re/diversification process.

### **6.3. Development of Simulation Models**

#### **Geometry and Zoning**

Buildings were modeled based on the detailed drawings provided by the Building Police. However, to maintain representation consistency, operational and material properties were defined according to the information available in the previous phase. As such, the simulation models were high-resolution versions of the building representation derived from large-scale data sources. The inconsistencies between the large-scale data sources and the real situation were ignored. Such inconsistencies pertain mainly to recent thermal retrofit activities, or changes in the usage of spaces. But the incorporation of these changes would damage the representativeness of the sample.

In the definition of the bounds of the buildings' thermal envelope, the common conventions in performance simulation were followed. This means that the inner surfaces of the external walls, lowermost and uppermost enclosures were considered as the outer boundaries. To model all relevant elements within these boundaries, virtual surfaces parallel and equidistant to the facets of the elements were considered. By "relevant elements" those constituting the boundaries of distinct thermal zones are intended. Figure 30 provides a schematic illustration of this convention. Walls adjacent to other conditioned buildings were modeled as adiabatic elements.

Building models were composed of multiple zones. To avoid over-complication of the models, on each building floor, all thermally connected spaces of similar usage were considered as one thermal zone. This rule applied to conditioned, as well as unconditioned spaces. As such, all internal floors were modeled according to the previously described convention, whereas only the internal walls constituting the boundaries of a zone were included in the geometric model. Figure 31 provides an illustrative example of the adopted zoning convention. Representative buildings included between 11 zones in the simplest building (Preßgasse 17) and 76 zones in the most complex building (Mühlgasse 7).

Partition walls within thermal zones, which contribute significantly to the thermal mass of the building, were not represented geometrically, but were included in the thermal mass computations. Internal apertures such as doors and windows were not modeled. Whereas external windows were represented geometrically according to available drawings.

Balconies and protruding façade elements permanently contributing to shading were geometrically represented as shading elements; so were all surrounding buildings obstructing solar radiation. Trees were not included in the models. Also, shading components such as blinds and curtains were not considered due to lack of information on their presence, and operation routines.

### Semantic Properties of Building Components

To represent the thermal quality of the building components, for historical buildings, typical layered compositions of various building elements, based on the period of construction were extracted from the Austrian Handbook for Building Thermal Retrofit (Schöberl et al. 2012). These constructions provided the material and thickness of every layer, but included no information on other material properties required for simulation computations (e.g., thermal conductivity, density, etc.). Additional information on the thermally relevant properties of the suggested materials were extracted from MASEA (Fraunhofer 2017) and Baubook (Baubook GmbH 2017) web-based data bases.

In cases, where a wide range for the thermal conductivity of a certain material was conceivable (e.g., bricks and plasters), the values were selected such that the overall thermal transmittance (U-value) of the construction matched the values suggested by the standards for the computation of the heating demand of historical buildings (Österreichisches Institut für Bautechnik 2015). In case of newer buildings (Mattiellistraße 3 and Rechte Wienzeile 25) the constructions provided in the building documentation were adopted. But the thermal conductivity values were readjusted in accordance with the values provided in the standards, similar to the case of historical buildings. The simplest possible representation, based on U-value was adopted for the modeling of transparent building elements.

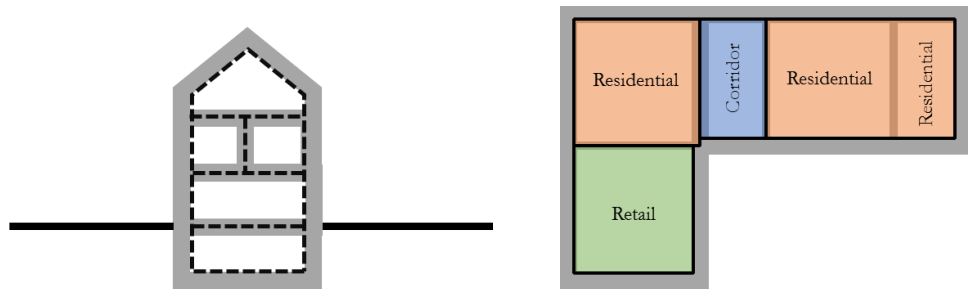


Figure 30 (Left) Modeling of space enclosures in the simulation models.

Figure 31 (Right) An example of the adopted zoning convention based on space usage and connectivity, applied to an arbitrary building floor: Four zones are identified in this floor plan.

### **Operational Parameters**

Operational parameters pertaining to inhabitants' presence, use of appliances and lighting devices, and natural ventilation were represented through schedules and reference values. In this representation, the reference value provides the maximum possible numerical value of a commodity in a time step, and the schedule expresses the fraction of that value associated with every time step. HVAC operation was expressed in terms of a schedule with Boolean values (0 and 1). To represent the temporal patterns of the above mentioned aspects, schedules suggested by ASHRAE (2013) were used. For non-residential buildings, different schedules are available for weekdays and weekends. The standard schedules were however modified, such that the daily operation hours matched the values suggested in the Austrian standards (Austrian Standards Institute 2011) for the pertinent building usage.

For people's activities, a schedule was developed based on the expected activities throughout an average day. Given the dependence of natural ventilation on inhabitants' presence in the building, and in the absence of more detailed information, the occupancy schedule was adopted for ventilation as well. Unconditioned spaces such as corridors, basements, and uninhabited attic spaces were associated with higher permanent infiltration rates, and no elements contributing to internal gains.

As emphasized before, the simulation models are supposed to be high-resolution versions of the representations generated based on large scale data. As such, in the definition of reference values for the above-mentioned parameters, the internal gain values suggested by the Austrian standards (Austrian Standards Institute 2011) and employed in the previous phases must be considered. In other words, the mean rate of internal gains from people, equipment, and lights must match the values provided in the mentioned standard.

For every time-step, Energy Plus computes the values of lighting and equipment gains by multiplying the applicable usage rate (given by the schedule) by the pertinent reference value. In the case of occupants, the hourly gains depend not only on hourly occupancy presence rates, but also on the metabolic rate or activity level assumed for the time step. For the determination of the reference values such that the overall gains agree with standard-based average rates, area-related annual internal gains value was computed for each usage (Equation 47). For non-residential buildings, gains from lighting were computed separately according to (Equation 48).

$$Iga = t_{use,a} \times q_{i,h} \quad (47)$$

, where  $Ig_a$  is the area-related annual internal gains for a certain usage [ $Wh.m^{-2}.a^{-1}$ ],  $t_{use,a}$  is the number of annual operation hours associated with the usage [ $h.a^{-1}$ ], and  $q_{i,h}$  is the standard internal gains rate for the usage in question [ $W.m^{-2}$ ].

$$Ig_{a,L} = t_{use,a} \times q_{i,l} \text{ (Non-residential)} \quad (48)$$

, where  $Ig_{a,L}$  is the area-related annual internal gains from lighting for a certain usage [ $Wh.m^{-2}.a^{-1}$ ], and  $q_{i,l}$  is the standard internal gains rate from lighting for the usage in question [ $W.m^{-2}$ ].

The reference values for the lighting, and equipment use, and the number of occupants had to be selected, such that the sum of the internal gains computed by the simulation matched the above values. For this purpose, first the above aggregate values had to be disaggregated among the three components of internal gains (Except in the case of lighting gains in non-residential usages). This disaggregation was based on statistical information on the contribution of each component to internal gains. Internal gains from each sector can be calculated based on Equation 49, Equation 50, Equation 51.

$$Ig_{a,L} = Ig_a \times f_L \text{ (Residential)} \quad (49)$$

$$Ig_{a,E} = Ig_a \times f_E \quad (50)$$

$$Ig_{a,P} = Ig_a \times f_P \quad (51)$$

, where  $Ig_{a,E}$  and  $Ig_{a,P}$  are the annual area-related internal gains from equipment and people [ $Wh.m^{-2}.a^{-1}$ ], and  $f_L$ ,  $f_E$ , and  $f_P$  are the shares of internal gains from lighting, equipment, and people respectively. For residential spaces, 58%, 19%, and 23% were assumed for gains pertaining to equipment, lighting and people respectively (Kemna & Acedo 2014). For gastronomy 70% and 30% were considered for gains from people and equipment, and for office 30% and 70 %.

The reference values for lighting and equipment power were computed according to Equation 52.

$$P_{L/E} = \frac{Ig_{a,L/E}}{\sum HR_{L/E}} \quad (52)$$

, where  $P_{L/E}$  is the reference area-related lighting/equipment power [ $W.m^{-2}$ ],  $\sum HR_{L/E}$  is the aggregated annual full load hours of lighting/equipment use calculated based on hourly rates provided by the schedule [ $h.a^{-1}$ ].

In the case of occupants, the hourly gains depend not only on hourly occupancy presence rates, but also on the metabolic rate or activity level assumed for the time step. In this case, the following equation was applied (Equation 53):

$$N_p = \frac{Ig_{a,p}}{\sum (HR_p \times H_{mr})} \quad (53)$$

, where  $N_p$  is the area-related number of people [ $person.m^{-2}$ ],  $HR_p$  is the hourly presence rate of inhabitants provided by the schedule [ $h.a^{-1}$ ], and  $H_{mr}$  is the hourly activity level or metabolic rate of the inhabitants provided by the schedule [ $W.person^{-1}$ ].

The same logic was applied to acquire the reference air change rates, such that the annual average matches the expected value. For this purpose, the average hourly air change rate for every usage was calculated based on the standard provided values (Equation 54). Then, the reference air change rate was determined such that this average value would be attained (Equation 55). Note that natural ventilation follows the same schedule as occupancy presence rate.

$$\overline{n_v} = \frac{n_v \times t_{use,a}}{t_a} \quad (54)$$

, where  $\overline{n_v}$  is the average hourly air change rate [ $h^{-1}$ ],  $n_v$  is the standard provided value for air change rate during use hours [ $h^{-1}$ ], and  $t_a$  is the number of hours in a year [ $h.a^{-1}$ ].

$$n_{v,0} = \frac{\overline{n_v} \times t_a}{\Sigma(HR_v)} \quad (55)$$

, where  $n_{v,0}$  is the reference value for air change rate [ $h^{-1}$ ], and  $\Sigma(HR_v)$  is the total number of full load operation hours of ventilation in a year [ $h \cdot a^{-1}$ ].

Once the reference simulation models were developed, they were used as a basis to develop diversified models to represent buildings in the study area.

#### **6.4. Diversification Parameters**

The re-diversification module readjusts the non-geometric parameters of the reference simulation models to recapture part of the lost diversity and generate simulation models that better reflect the physical and operational characteristics of the represented buildings. The building parameters currently subjected to diversification are the thermal properties of the main building components and operational parameters. More specifically, the following model properties are diversified and readjusted:

- Schedules pertaining to the presence and actions of inhabitants, use of lighting and equipment.
- Reference area-related values for the number of inhabitants, equipment and lighting power.
- Reference values for air change rate.
- Thermal performance of the main components of the buildings' thermal envelopes: uppermost and lowermost enclosures, external walls.

In the current effort, fairly constant physical contextual parameters such as adjacency relations and mutual shading have been considered in the clustering process. The contribution of other urban features such as traffic or trees to microclimate variations can be captured via the re-diversification process. This can be achieved through readjustment of standard climate data used in the simulation or operational parameters, based on location dependent variables via a dedicated microclimate model. Similarly, a detailed consideration of inhabitant types, their behavioral traits and the consequent energy implications can be included. This can be facilitated through the coupling of the re-diversification module with an advanced data-informed occupancy model. Such expansions have been foreseen in the initial vision of the envisaged computational environment. However, the present effort focuses only on the diversification of those parameters, for which some source of background information concerning pertinent distribution patterns was available. The diversification process is guided by the information contained in the initially generated building stock representation. The current implementation of the re-diversification of occupant-dependent variables, can be seen as a place holder for more intricate considerations of inhabitant diversity in future attempts.



## 6.5. Diversification Method

### Computational Logic

For the purpose of re-diversification, for every building in the neighborhood, a duplicate of the reference simulation model pertaining to the relevant representative building is created. In the new model, the values of the above-mentioned parameters are modified according to the information available on the target buildings, or based on background information on the distribution of the parameter subject to diversification. This procedure results in the generation of a unique simulation file associated with a unique set of schedules for every building in the study domain. These models share all geometric properties of the reference models, but are expected to emulate the thermal and operational properties of the target buildings more closely.

The resulting simulation files are batch processed from the Energy Plus launcher. Hourly volumetric simulation results for energy loads of these diversified models are used along with the net volume of the target buildings to arrive at the hourly demand of these buildings according to Equation 56.

$$Q_{i,h} = \frac{Q_{Sim,i,h}}{V_{reference,i}} \times V_{n,i} \quad (56)$$

, where  $Q_{i,h}$  is the heating demand of an arbitrary building  $i$  in timestep  $h$  [kWh],  $Q_{Sim,i}$  is the heating demand of the diversified simulation model pertaining to building  $i$  in the same timestep [kWh],  $V_{reference,i}$  and  $V_{n,i}$  are the net volumes of the reference building (based on which the simulation model is generated) and building  $i$  [ $m^3$ ] respectively.

A thorough explanation of the methods adopted to diversify each model parameter follows.

### **Representation of Inhabitant Behavior Diversity**

The reference schedules represent the temporal distribution of internal loads in aggregate terms. However, use of average profiles for demand assessments will result in identical peak load hours and unrealistically monotonous profiles across building classes. In order to stochastically represent occupancy-related factors, for each building based on the reference schedules for various days of the week, a set of randomized schedules were created and stored as CSV files compatible with Energy Plus requirements. Each schedule file has 8760 rows (for every hour of the year), and 5 columns corresponding to occupants' presence, lighting use, equipment use, HVAC operation and activity level (metabolic rate). The first three variables assume real values in the range of 0 to 1, and the fourth is a Boolean variable (0 or 1). The metabolic rate column includes values between 70 and 180 [ $W \cdot person^{-1}$ ]. HVAC schedules are not diversified.

As apparent in Table 11, the representative buildings are associated with three usages: residential, office, and gastronomy. In case of residential and gastronomy usages, schedules for occupancy, activity level, lighting and equipment use were diversified separately. For this purpose, for every hour of the year and every variable, the value suggested by the reference schedule was considered as the mean value of the time step, and a default coefficient of variance (CV) was used to derive the standard deviation of the normal probability distribution representing the variable in question (Equation 57). A value was then generated based on this distribution and assigned to the time step using NumPy an existing Python package (NumPy developers 2017).

The identification of the specifically appropriate CV value is an open research question. Former studies on an office space with eight employees, suggests that for certain applications (e.g., the stochastic generation of presence patterns), CV might display a distinct value range (Mahdavi & Tahmasebi 2015). For instance, in the observed case study, the work duration is associated with a CV of 0.2-0.3 for various subjects, whereas the CV of arrival and departure times reveals a smaller value. As such, a CV value of 0.2 was deployed in the present study. However, as mentioned before, further investigation is required to arrive at CV values, which better describe the distribution of various inhabitant-related aspects of building operation.

$$\sigma_{t,j} = CV_j \times \mu_{t,j} \quad (57)$$

, where  $\sigma_{i,j}$  is the standard deviation of the variable  $j$  in time step  $t$ ,  $CV_j$  is the coefficient of variance associated with variable  $j$  (in the present case for all variables this value equals 0.2), and  $\mu_{t,j}$  is the mean value of the variable  $j$  in time step  $t$  which equals the value provided for the time step in the standard schedules.

Rules were set to ensure that the selected value remained within the acceptable range. These included upper and lower threshold values for various variables: a range of 0 to 1 for inhabitant presence, appliance and lighting use values, and minimum values of 70, and 90 for activity levels (metabolic rates) in residential and non-residential buildings. As such, the generated schedules maintain the overall tendencies of the reference schedules, while featuring unique characteristics, which better emulate the diverse nature of occupant behavior.

For residential and gastronomy usages, all schedules were diversified independently. In the case of office spaces, where equipment and lighting use are very strongly correlated to inhabitant presence, the rate of lighting and equipment use in every time step of the reference schedules was expressed as a function of the occupancy rate (Equation 58 and Equation 59). In this case, occupancy schedule was subjected to diversification. Equipment and lighting rates were computed based on the occupants' presence rate at every time step. Diversified activity level schedules were also generated based on the explained method. Natural ventilation was assumed to follow the occupancy schedule.

$$L_t = O_t \times f_{L,t} + L_{0,t} \quad (58)$$

, where  $L_t$  is the lighting use rate in time step  $t$ ,  $f_{L,t}$  is a factor derived from standard schedules for time step  $t$ , and  $L_{0,t}$  is the minimum lighting use rate of the timestep regardless of occupant presence (also defined based on standards schedules).

$$E_t = O_t \times f_{E,t} + E_{0,t} \quad (59)$$

, where  $E_t$  is the equipment use rate in time step  $t$ ,  $f_{E,t}$  is a factor derived from standard schedules for time step  $t$ , and  $E_{0,t}$  is the minimum equipment use rate of the timestep regardless of occupant presence.

The diversification code links each IDF file (Simulation model) associated with a building to a unique set of diversified schedules according to the functions present in the model. The implementation of more elaborate occupancy models (e.g., Page et al. 2008) was not considered due to the excessively high granularity of the resulting profiles for urban level assessments. However, for certain applications, where an in-depth study of behavioral aspects of energy use is intended, the use of such models can be envisaged. The employment of Energy Plus for the simulation assessments certainly allows for the incorporation of more detailed developments with regard to occupant behavior modeling. Yet the usefulness of the adopted method, and its implications for computational load should be justified.

### Readjustment of Reference Values Pertaining to Internal Gains and Ventilation

Once the schedules are generated, the reference values for equipment and lighting power, number of occupants, and air change rate need to be diversified. This process is informed by a comparison between the annual area-related internal gains ( $Ig_{a,Building}$ ) and average hourly air change rate ( $\overline{n_{v,Building}}$ ) of the target buildings in each cluster with the values associated with the reference building representing them. These values were extracted from the developed urban building stock representation, for the purpose of the re-diversification process (See 906.2). For each target building, if the internal gains or air change rate vary from those of the reference building, in the simulation file the relevant values are readjusted in proportion to the observed difference.

However, a straightforward readjustment of reference values for equipment and lighting power, number of occupants and ventilation rate would not be accurate, since each simulation model is associated with a unique set of schedules. Although the schedules pertaining to each usage are relatively similar in terms of overall tendencies, the annual of full load operation hours may vary slightly from one schedule to another. As such, to determine the reference values, the following steps are taken:

- The relation between the annual internal gains of the target building and the reference building is expressed as a fraction (Equation 60).
- For the target building, the annual internal gains through lighting ( $Ig_{a,L}$ ), equipment use ( $Ig_{a,E}$ ), and people ( $Ig_{a,P}$ ) are calculated for each usage based on the default values of these variables calculated for the generation of the reference models and the multiplier derived in the previous step (Equation 61).
- Using the new values, and the schedules associated with the target building, the reference values for lighting and equipment power, and number of occupants are computed according to Equation 52 and Equation 53.

$$f_{Ig} = \frac{Ig_{a,Building,Target}}{Ig_{a,Building,Reference}} \quad (60)$$

, where  $f_{Ig}$  is the multiplier for the readjustment of internal gains, and  $Ig_{a,Building,Target}$  and  $Ig_{a,Building,Reference}$  are the area-related annual internal gains of the target building and the reference building representing it respectively [ $Wh.m^{-2}.a^{-1}$ ].

$$Ig_{a,L/E/P,Target} = Ig_{a,L/E/P,Reference} \times f_{Ig} \quad (61)$$

, where  $Ig_{a,L/E/P,Target}$  and  $Ig_{a,L/E/P,Reference}$  are the area-related annual internal gains from lighting/equipment/people of the target and reference buildings respectively [ $Wh.m^{-2}.a^{-1}$ ].

Similarly, the multiplier for reference values for air change rate is computed according to Equation 62.

$$f_{\overline{n_v}} = \frac{\overline{n_{v,Building,Target}}}{\overline{n_{v,Building,Reference}}} \quad (62)$$

, where  $f_{\overline{n_v}}$  is the multiplier for the readjustment of average air change rate, and  $\overline{n_{v,Building,Target}}$  and  $\overline{n_{v,Building,Reference}}$  are the average hourly air change rates of the target building and the reference building representing it respectively [ $h^{-1}$ ].

Equation 55 is used to compute the reference value for air change rate for each usage based on the schedule associated with the target building. The resulting value is multiplied by the above identifier.

### **Readjustment of thermal properties**

The diversification of the thermal properties of the envelope relies on the diversity of effective element U-values. The effective element U-value is not only a measure of the thermal quality of the element's construction, but also a function of the significance of the said construction in the overall thermal performance of the building. This significance is determined by the share of the elements associated with a particular construction in the total thermally effective area of the envelope (corrected for adjacencies). The diversification process modifies the pertinent constructions in such a way as to replicate the effective construction U-values computed for every principle building element in the previous phase. Since the geometry of the diversified target model is identical to that of the reference model, any deviations from the effective U-values of the reference building must be accounted for by manipulating the U-values of the constructions in the new model. For instance, if the walls of a building have a more significant share in the building's effective envelope than is the case for the reference building, and both buildings have the same effective wall U-value, the wall element of the diversified model will assume a higher U-value, such that the effective U-value of the wall remains the same.

As mentioned before, for every reference simulation file, building component constructions were defined according to the common practice of the period of construction of the reference building. In every relevant construction, only the most thermally effective layer was subjected to modifications. (e.g., the massive masonry layer, or the ceiling timber in the case of historical buildings, and the insulation layer in case of newer buildings). Since a modification of the thermal mass of the building was not intended, only the thermal conductivity of the layer was changed to reach the target construction U-value (Equation 63).

Rules have been applied to prevent the layer from assuming unreasonably large or negative values. This may happen in rare cases where the effective U-value of the components vary significantly from those of the reference building.

The re-diversification process, for each building computes the value of the diversified variables and modifies the associated IDF model accordingly. As a result, each building is represented by a unique simulation model which mimics the geometry of the pertinent reference building but is expected to represent the operational and thermal properties of the target building more closely.

$$U_{Target} - U_{Reference} = \frac{1}{R_{Target}} - \frac{1}{R_{Reference}} \quad (63)$$

$$\Rightarrow R_{Reference} - R_{Target} = \frac{U_{Target} - U_{Reference}}{U_{Target} \times U_{Reference}}$$

$$\Rightarrow \frac{1}{\lambda_{Reference}} - \frac{1}{\lambda_{Target}} = \frac{U_{Target} - U_{Reference}}{U_{Target} \times U_{Reference} \times d}$$

$$\Rightarrow \lambda_{Target} = \frac{1}{\frac{1}{\lambda_{Reference}} - \frac{U_{Target} - U_{Reference}}{U_{Target} \times U_{Reference} \times d}}$$

, where  $U_{Target}$  and  $U_{Reference}$  are the effective U-values of a certain component (uppermost/lowermost enclosures/external walls) of the target and reference buildings [ $W.m^{-2}.K^{-1}$ ],  $R_{Target}$  and  $R_{Reference}$  are the thermal resistance of the said element in target and reference buildings [ $K.m^2.W^{-1}$ ],  $d$  is the thickness of the layer subjected to diversification [ $m$ ], and  $\lambda_{Target}$  and  $\lambda_{Reference}$  are the thermal conductivities of the layer in the target construction and the reference construction [ $W.m^{-1}.K^{-1}$ ].

The generation of diversified simulation files for all buildings in the case study, on an average PC, requires less than one hour. The simulation process itself, for the 744 buildings within the study area, takes several hours. The computation load association with the simulations can be managed through deployment of distributed computing schemes.



## 6.6. Illustrative Scenarios

Two sets of illustrative scenarios have been developed to demonstrate the impact of the adopted diversification on model predictions. The first set of scenarios aims at comparing the model predictions with various levels of diversification to determine whether or not the integrity of the model and its representativeness is maintained through the re-diversification process.

The second set is intended to display the utility and implications of the re-diversification process in comparative analysis of hypothetical simple behavior change scenarios. The envisaged scenarios in this regard, are purely illustrative and do not claim to represent the complexity of real behavioral traits. Rather, they are developed to demonstrate the different response of the diversified and non-diversified models to potential behavioral changes.

### Diversification Scenarios

To assess the impact of the diversification process on the model predictions, the predictions of the non-diversified model were compared to the predictions resulting from models with two levels of diversification. The non-diversified model (NDM) is based on the reference simulation files with standard assumptions for thermal quality of components, schedules, area-related internal gains, and air change rate (As presented in 6.3). In this case, the volumetric hourly demand values of the reference simulation models are multiplied by the volume of the represented buildings to yield the hourly heating demand of these buildings. In the second model (DMS), as a first diversification step, only the schedules are diversified. The third model (DMA) involves all the diversification steps described in the previous chapter (Table 12). In each case, simulation results are used to compute the heating demand of the entire neighborhood according to Equation 56.

Table 12 Overview of the diversification scenarios

Abbr.	Schedules	Thermal properties	Internal gains	Number of simulations
<b>NDM</b>	Not diversified	Not diversified	Not diversified	7
<b>DMS</b>	Diversified	Not diversified	Not diversified	744
<b>DMA</b>	Diversified	Diversified	Diversified	744

### Behavior Change Scenarios

To demonstrate the advantages of the developed computational model for the investigation of various change and intervention scenarios, three simple illustrative scenarios pertaining to changes in the operational parameters of buildings (occupant behavior) were designed. The first scenario follows the standard assumptions for internal temperature and HVAC availability hours. The second scenario assumes a set-back heating set point for the vacant hours in non-residential spaces, which is closer to the actual building operation tendencies. The third scenario, emulating the behavior of a more energy-aware population, maintains the set-back threshold, and modifies the internal heating set point temperatures in proportion to the occupancy rate of the building in every time step. This scenario is based on the simple assumption that if the occupancy rate is lower, fewer spaces are heated, thereby reducing the average internal temperature of the building. These scenarios were simulated with the NDM and DMA models. Note that the developed behavior change scenarios are intended as illustrative examples to demonstrate the modeling possibilities of the developed framework and are not presumed to realistically capture the behavior of inhabitants. Table 13 describes the rules defining the HVAC operation in each scenario.

Table 13 Overview of the base case and scenario assumptions for HVAC operation

		Residential		Non-Residential	
S0	Set point assumptions [°C]	20		20	
	HVAC Availability	24 hours a day		14 hours on weekdays	
S1	Set point assumptions [°C]	20		20 during work hours 14 other times	
	HVAC Availability	24 hours a day		24 hours a day	
S2	Set point assumptions [°C]	16	Night hours	14	Not working hours
		16	Occupancy rate <25%	16	occupancy rate <25%
		20	Occupancy rate > 55%	20	Occupancy rate > 75%
		Interpolate	Other times	Interpolate	Other times
	HVAC Availability	24 hours a day		24 hours a day	

## 6.7. Results and discussion

### Diversified Schedules

Figure 32 displays examples of hourly metabolic rates (activity level values) of a typical day for each usage. As apparent in the graph, the values remain within a plausible range, yet display variations.

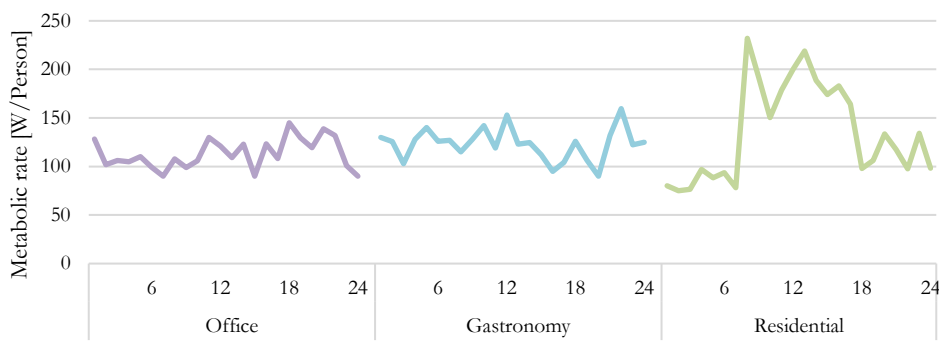
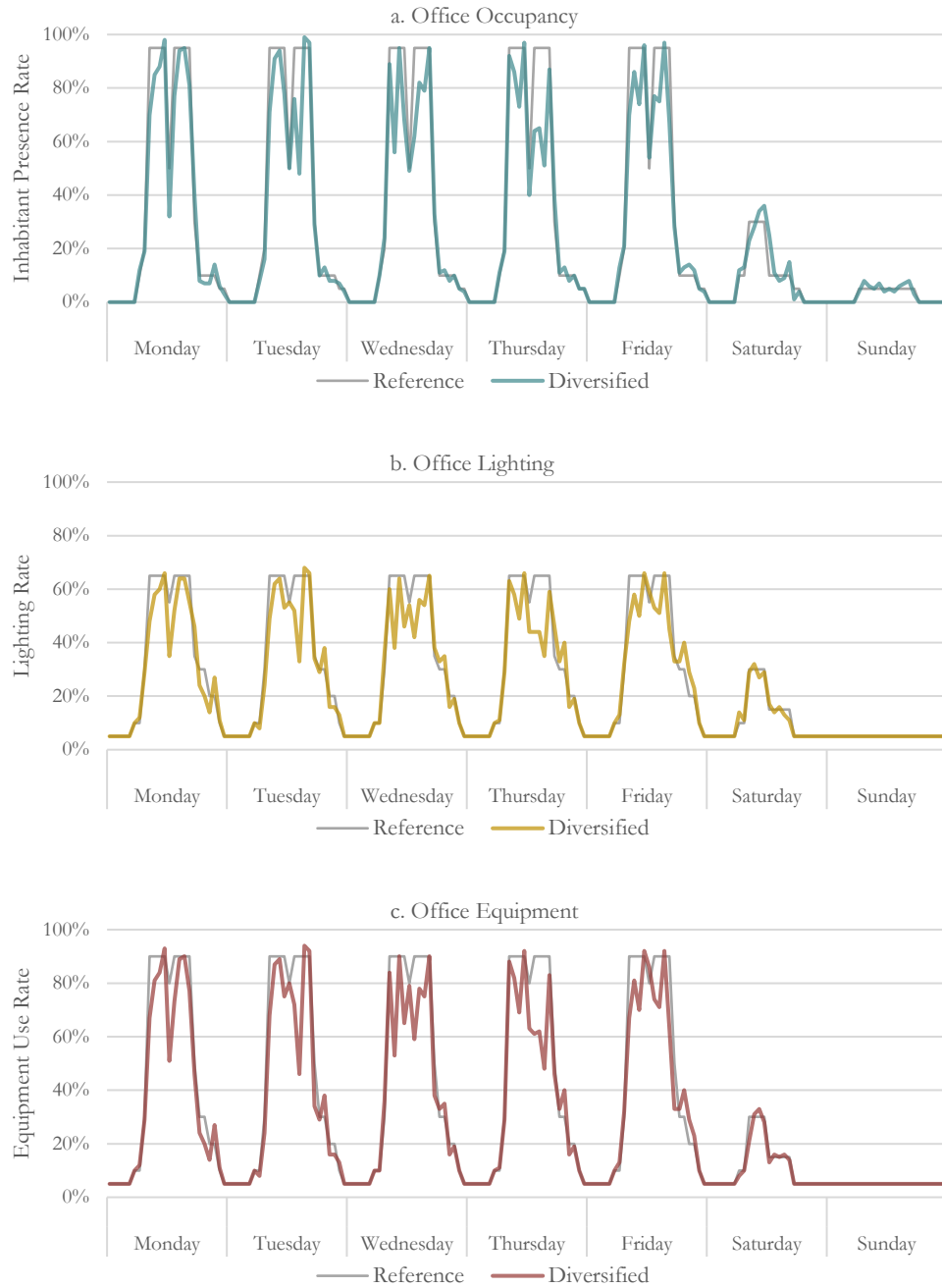


Figure 32 An instance of hourly metabolic rate values generated for the three usages on a typical day.

Figure 33, Figure 34, and Figure 35 illustrate the standard schedules for occupancy, lighting, and equipment use as well as instances of a week of diversified operational data generated. Note that the diversified schedules closely follow the tendencies of the reference schedules, yet represent the stochastic nature of occupant behavior. The presented examples display a single week of data generated for three instances of office, gastronomy, and residential spaces. The generated schedule files include 8760 rows of stochastically generated data for each commodity, and are produced individually for each building. Note that the usages present in the simulation file produced to represent each target building are the same as those incorporated in the relevant reference model. As such, the general temporal distribution of loads in the target models are not much different from the reference model. The variations in load distributions due to diversification of schedules maintain the general patterns but add some noise to better represent the stochastic nature of inhabitant-dependent building characteristics.



*Figure 33 Reference schedules versus diversified schedules generated for a building with office usage representing a. Occupancy, b. Lighting, and c. Equipment.*

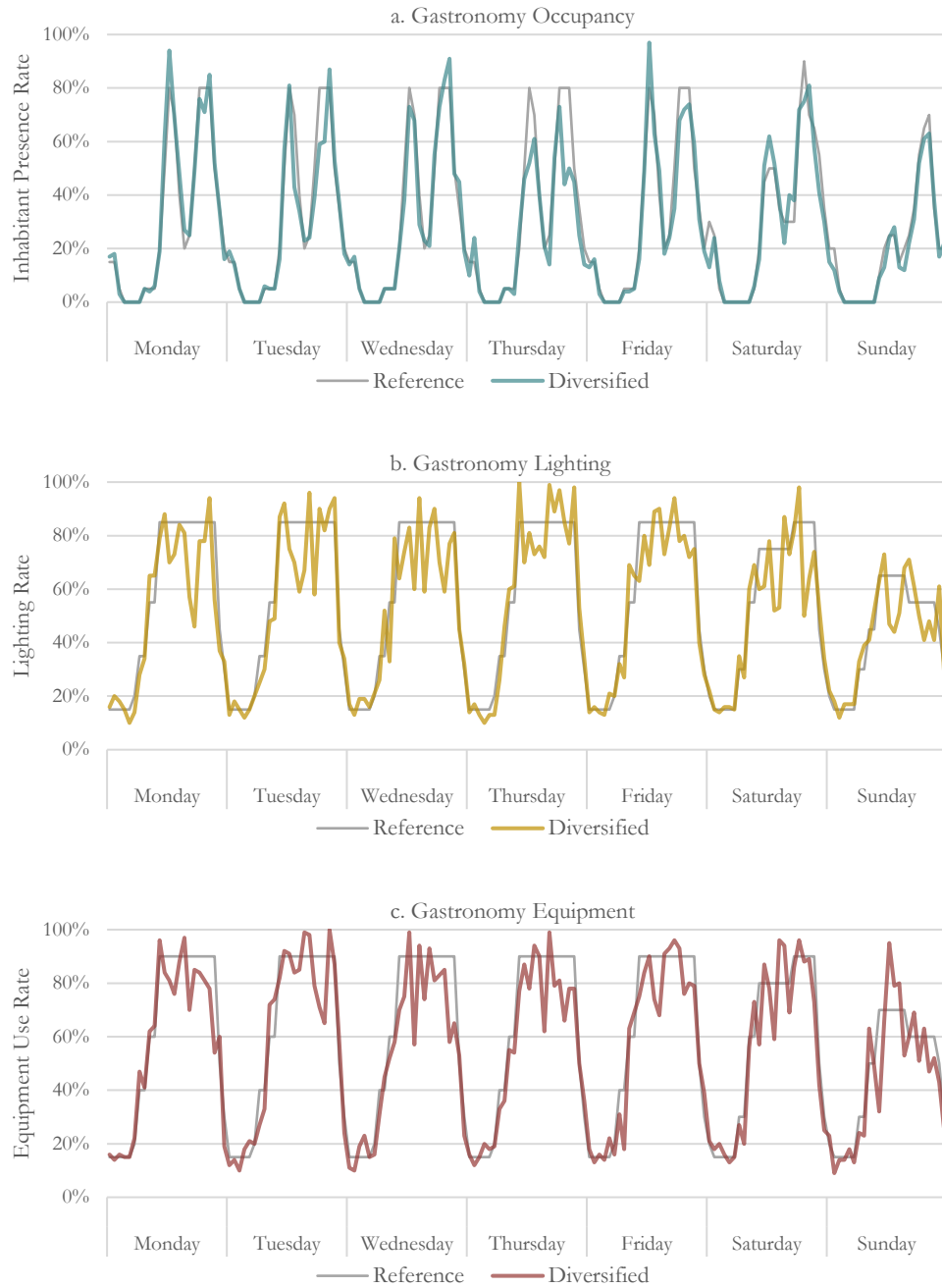
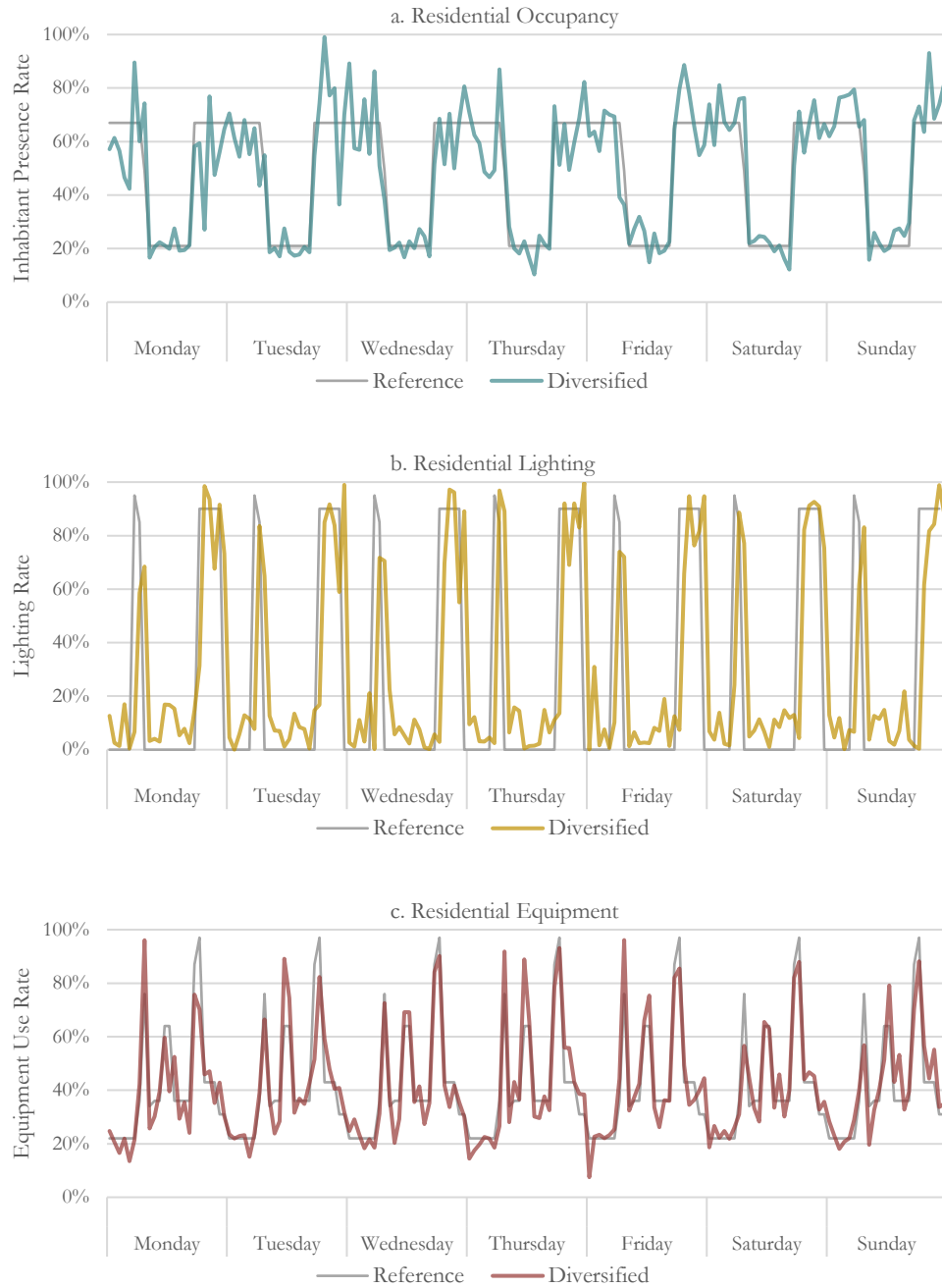


Figure 34 Reference schedules versus diversified schedules generated for a building with gastronomy usage representing a. Occupancy, b. Lighting, and c. Equipment.



*Figure 35 Reference schedules versus diversified schedules generated for a building with residential usage representing a. Occupancy, b. Lighting, and c. Equipment.*

### Results of Diversification Scenarios

Table 14 summarizes the results of the modeled cases. As seen in the table, both levels of diversification result in minor changes in the model outcome in terms of the aggregate energy demand of the neighborhood, as well as annual peaks. The diversification of the schedules results in small modifications in the annual peak load (+1%) and the aggregated annual demand of the neighborhood (-1%). The cumulative effect of the diversified schedules resembles closely, that of the impact of the average schedule, which was to be expected. The additional readjustment of the buildings' thermal properties, internal loads and ventilation rates causes slightly more significant changes in the overall predictions of the model (-3.4%). However, the deviation at the aggregated level is not so substantial as to compromise the representativeness of the model.

*Table 14 Summary of the results of modeled diversification scenarios. The non-diversified model is considered as the reference model in the calculation of the relative values.*

Models	Maximum hourly load [MWh]	Relative deviation from reference scenario [%]	mean hourly demand [MWh]	Standard deviation [MWh]	Total annual space heating load [MWh]	Relative deviation from reference scenario [%]
<b>NDM</b>	153.13	0	22.64	26.83	198354.47	0
<b>DMS</b>	154.84	1.11	22.4	26.34	196406.17	-0.98
<b>DMA</b>	151.39	-1.13	21.88	25.97	191659.24	-3.38

Nonetheless, the impact of the diversification process is magnified when the scale of observation reduces. A comparison of the aggregated annual demand of various clusters computed with the three models can be seen in Table 15. At the level of the clusters, the deviations from the NDM are more substantial in the case of DMA. The non-residential building clusters are more sensitive to changes in schedules.

*Table 15 Deviations in the annual demand of clusters as predicted by diversified models from the non-diversified model predictions.*

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
NDM	Aggregated annual heating demand [MWh]	45584.23	22397.46	38435.705	41929.35	16690.24	32003.70	1313.78
DMS		45489.35	22343.98	38334.817	41001.84	16663.19	31264.58	1308.41
DMA		43109.15	25443.04	32301.631	37494.41	17984.79	33739.19	1587.04
DMS	Relative deviation from NDM [%]	0	0	0	-2	0	-2	0
DMA		-5	14	-16	-11	8	5	21

The percentage deviation of the volumetric annual heating demand of the buildings acquired from the DMA, from the reference (non-diversified) values are shown in Figure 36. For the majority of the buildings, the DMA volumetric demand remains within 20% of the reference value. The values predicted by DMS do not vary significantly from the reference values.

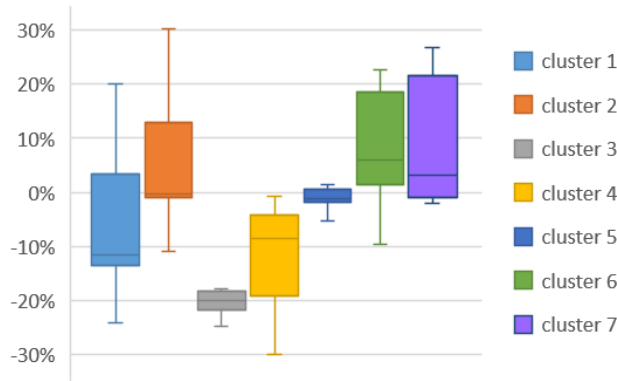


Figure 36 Percentage deviation of the volumetric annual heating demand of buildings in each cluster, computed by the DMA from that of the relevant reference model (NDM).

If the observation scale is further reduced to a single time step, both DMS and DMA result in noticeable deviations from the non-diversified hourly predictions. displays the diversity caused by the two diversification scenarios in the hourly heating demand values of the buildings in a single time step (11 am, January 8th). In this graph, the percentage deviation from the hourly demand value predicted by the non-diversified model has been shown. The two diversification scenarios, DMS and DMA, result in 2% and 8% average deviation from the hourly predictions of the non-diversified model respectively.

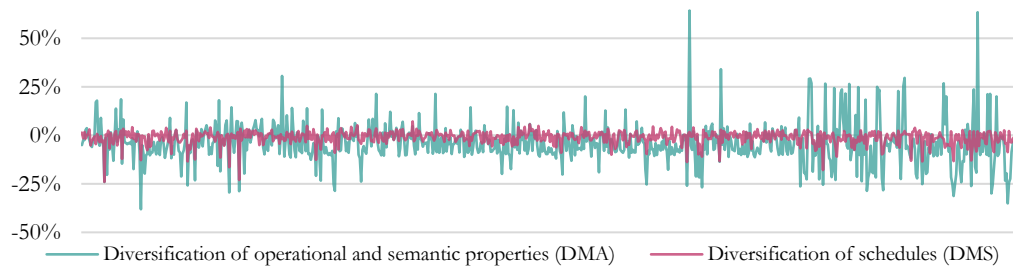


Figure 37 Relative deviation of hourly demand results of all buildings as predicted by the DMA and DMS from NDS predictions for a single time step. The x-axis represents the 744 buildings in the study area.



These variations, although unnoticeable at aggregate scale, can have significant implications for instance for the design and deployment of small scale distributed generation schemes. Although the variance resulting from the re-diversification process produces seemingly more realistic results, the performance of this model towards accurate representation of urban energy demand requires further validation. The ultimate reliability of the predictions with regard to heating demand, can only be assessed and validated based on actual demand data, currently unavailable.

### Results of Illustrative Behavior Change Scenarios

The three sets of assumptions about inhabitant behavior have been applied to the non-diversified model as well as the fully diversified model. The results of this investigation are summarized in Table 16. All scenarios have been compared to the base case assumptions with the non-diversified model (as the reference scenario). As mentioned before, at the aggregated level, peak, mean, and total heating demand simulated for the base case assumptions (S0) change little due to the inclusion of re-diversification in the modeling procedure. The application of the first behavior change scenario does not result in the divergence of the tendencies of non-diversified and the diversified model. The consideration of a setback value for the operation of non-residential spaces leads to a minor increase (2%) in the overall energy demand of the neighborhood based on both models. This is to be expected as the modifications applied in this scenario are somewhat independent of the occupancy-related aspects (they apply only to non-residential spaces in non-occupied hours). However, the implications of this scenario for the annual peak load are much more significant (16-17% depending on the computational model). Lower peak loads are advantageous for grid stability particularly in distributed generation systems.

Table 16 Results of the behavior change scenarios as simulated by the non-diversified and diversified computational models

Scenarios		Annual Peak load [MWh]	Relative deviation from NDM-S0 [%]	Total annual space heating load [MWh]	Relative deviation from NDM-S0 [%]
NDM	S0	153.1	0	198.35	0
	S1	128.2	-16.3	200.70	1.2
	S2	122.6	-19.9	169.14	-14.7
DMA	S0	151.4	-1.1	191.66	-3.4
	S1	124.5	-18.7	195.22	-1.6
	S2	111.7	-27.0	170.30	-14.1

The differences between the two models become more visible in case of the second scenario, which is based on maintaining the set-back values in non-residential buildings during non-occupied hours, occupancy sensitive control of heating set points during daytime, and lower night-time indoor thermostat settings in residential buildings. The comparison of the second scenario's predictions of both models (NDM-S2, DMA-S2) with the respective base case predictions of the same models (NDM-S0, DM-2-S0) shows that in the non-diversified model the application of the occupant-sensitive HVAC control displays a more significant decrease in demand than in the diversified model (14.7%

compared to 10.7%). As such, the non-diversified model may over-estimate the energy saving impact of similar scenarios. Comparing the S2 scenario to S1, reveals an even larger improvement in the case of the non-diversified model (-15.9% versus -12.5%). However, the peak load predicted by the non-diversified model, when subjected to S2, does not display much improvement compared to the case of S1. The non-diversified model predicts an additional 3.6% improvement in peak load when applying S2, whereas the improvement predicted by the diversified model mounts up to 8.2%. This seems to be inconsistent with the higher overall demand reduction predicted by the former model. As such, the non-diversified model appears to overestimate annual demand reduction due to occupant behavior change, while failing to realistically predict the impact of these improvements on the peak loads. Figure 38 displays the range of hourly loads (of the entire domain) for various scenarios applied to the diversified model in terms of boxplots. In Figure 39, the cumulative frequency graph for hourly demand of the neighborhood in all scenarios has been plotted. Note that even though S1 and S2 result in decreases in the magnitude of the peak load, they increase the frequency of hours with higher demands. Such information can support an efficient deployment of supply side energy management strategies.

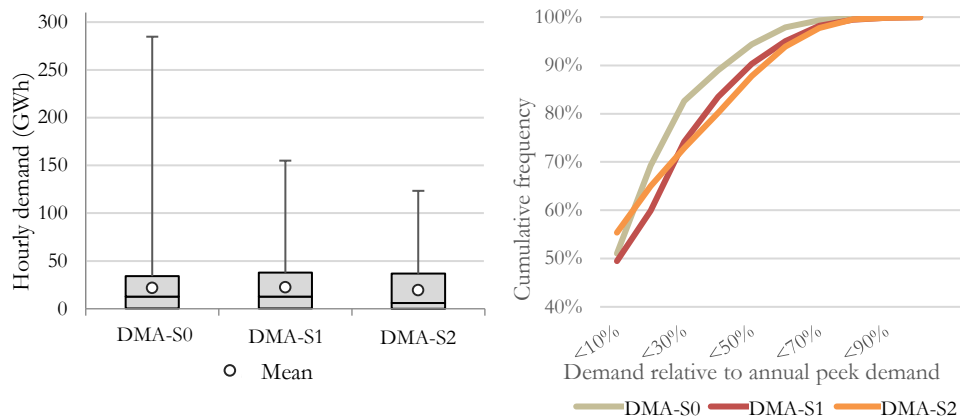


Figure 38 (Left) Comparative analysis of the hourly demand predicted in various scenarios by DMA

Figure 39 (Right) Cumulative relative frequency of the hourly heating load in relation to annual peak load

The current study cannot provide solid evidence as to whether or not the developed diversified model is successful in realistic capturing of the actual response of urban neighborhoods to inhabitant behavior change. However, the existing urban diversity, and the significant differences between the response of the diversified and non-diversified models to various assumptions with regard to building operation, suggest that essential information can be lost or misrepresented in common reductive urban stock models.

# Conclusion

## 7.1. Summary of Contributions

### **Large-scale implementation of detailed building performance simulation**

Addressing the requirement for more versatile urban energy computing methods, the *Hourglass* framework enables the utilization of full-fledged building performance simulation tools for high-resolution assessment of the performance of large building assemblies. These tools can model the intricate interrelations between buildings, their surroundings, and their internal processes, yet their large-scale application is hampered by limitations in informational and computational resources. The suggested approach relies on three fundamental developments to eliminate the practical obstacles toward implementation of simulation tools at urban level: a. the urban building stock representation module, b. the reductive module, and c. the re-diversification module. An urban instance in the city of Vienna, Austria, has been selected as a laboratory for the developmental and evaluative activities of the current project. As such, the current implementation is tailored to the Austrian context but can be readjusted for other geographical contexts with minor modifications.

**Automated generation of a geo-referenced, versatile, and energy-relevant urban building stock representation using available official and crowd-sourced data**

The first component, incorporates the data available at large scale to generate an energy relevant representation of the urban building stock. For this purpose, official and crowd-sourced web-based GIS data, relevant standards and guidelines, and statistical information are utilized. The developed routine, integrated within a plug-in for an open source GIS platform, superimposes the available data to extract information on buildings' geometric, thermal, contextual, and operational properties. The adopted method points out the shortcomings of the official GIS data in providing useful descriptions of buildings. The lack of an appropriate task-oriented ontology for the accumulation and organization of such data has resulted in inconsistencies in the available information, pertaining to physical and operative aspects. Moreover, due to the time, effort, and resources required for the accumulation of the official GIS information, and given the dynamic nature of the urban environments, the available information is outdated in many instances. Unofficial web-based sources on the other hand, which are fuelled by volunteer communities, if used with care can provide insight into the current state of the urban fabric. The present effort is a case in point of how crowd-sourced web-based data repositories can be utilized to evaluate, enrich, and update the available official data sources.

The generated representation does not only provide the basis for the following steps of the current project, but can also cater for other applications, which require three-dimensional urban stock models (e.g., solar or wind energy availability analyses). Since the entire process relies on geo-referenced information, the model can be further enhanced through incorporation of other information layers such as energy and mobility networks, renewable energy sources, etc. Currently, a building data representation schema created for the purpose of this project is being used. However, the recent advancements in the development of standard energy-compliant urban data representation schemas (Nouvel et al. 2015; Benner et al. 2016) facilitate the integration of the developed routine for the generation of multi-purpose city models.

**Employment of cluster analysis techniques for the identification of authentic representative buildings from an urban neighborhood**

The second component of the project, employs well-known data mining methods towards an energy-based classification and sampling of buildings. In lieu of the conventional usage and age based classification schemes, to more systematically capture the dynamic nature of the urban building stock and its transformations through retrofit and densification, as well as operative changes, an original set of energetically relevant indicators was assembled for stock segmentation. The building stock representation provided by the first module was used to extract the values of these descriptive indicators and the matrix of building information expressed in terms of these indicators was then subjected to Multivariate Cluster Analysis (MCA). The resulting classification was visualized in the adopted GIS environment. Average buildings were selected from each class as the class representative.

Contrary to the conventional classification efforts, where an increase in the number of classification criteria results in an increase in the number of classes, MCA-based classification enables the involvement of various parameters, while maintaining control over the number of resulting classes. This facilitates the use of explicit energy-influential characteristics such as average envelope U-value, as opposed to vague signals such as building age. In view of the dynamic nature of the urban building stock and its transformations through retrofit and densification, as well as operative changes, this may provide a more generic stock segmentation and sampling possibility. The cluster analysis algorithms allow for flexibility in the definition of the acceptable range for the number of emerging clusters. As such, the number of clusters can be adapted to the available computational and informational resources of the users.

Ideally, the identification of the most suitable set of descriptive indicators for classification and the potential weighting of these factors, should be based on detailed analysis of the sensitivity of the urban energy use patterns to various building characteristics. Since this analysis requires currently unavailable highly resolved information on the actual energy demand of the urban building stock, a different method was adopted to support the selection of the criteria for the classification process. Various scenarios in terms of classification criteria and clustering algorithms were tested for their performance toward efficient prediction of the aggregated and disaggregated values of heating demand. Preliminary tests carried out on the selected urban instance involving 744 buildings, based on simplified normative procedures suggest that the adopted classification and sampling schema can reliably represent the aggregate annual energy performance of an urban neighborhood. The best performing clustering schema results in the partitioning of the studied building agglomeration into 7 distinct building clusters. Each cluster is represented

through one building, resulting in a sample size of less than 1% of the entire domain. Despite the small sample size, this method leads to relatively small errors in the prediction of the aggregated neighborhood performance, as well as low average error in the building level predictions. A comparison between the conventional “construction period and usage” oriented classification and the suggested method (performed for residential buildings) reveals the superiority of the developed method in view of the representation of heating demand diversity. Although this evaluation has provided some indication of the efficiency of the selected method for building classification, due to the low temporal resolution of the incorporated information, further investigation with highly resolved data is necessary to readjust, enhance, and validate the model.

The selection of authentic buildings to represent the urban neighborhood, on the one hand enables the detailed modeling of the various aspects of these buildings based on actual data. On the other hand, use of authentic buildings as opposed to synthetic archetypes, facilitates the future utilization of punctual monitoring data for the calibration of large-scale energy models. Such methods can significantly improve the quality of model predictions, while minimizing the investment and effort required for data acquisition through monitoring activities.

**Consideration of the impact of urban diversity on the model predictions**

The third component of the environment utilizes stochastic methods and the generated stock representation to recover part of the diversity loss, inherent to reductive methods. Currently, several non-geometric building characteristics are subjected to diversification. These diversified aspects pertain to the magnitude and temporal distribution of internal gains and natural ventilation, as well as thermal properties of the building envelope. The information contained in the representation developed in the first method is used to create permutations of the simulation models of the selected representative buildings, such that the resulting models better express various characteristics of the target buildings.

The impact of the diversification process on the model predictions have been demonstrated with simple illustrative examples pertaining to behavior change scenarios. In the diversification of the reference schedules, due to the nature of the stochastic method adopted, in aggregate terms, the tendencies of the original standard schedules are maintained. However, the diversification results in a more realistic representation of the temporal as well as spatial distribution of energy demand. The data-oriented diversification of internal gains and thermal properties leads to more significant changes in the overall demand predictions of the model. However, it is expected to improve the performance of the model for comparative analysis of the energy impact of various scenarios due to the better representation of urban stock diversity.

A model validation process is essential to prove that the re-diversification process improves the precision and accuracy of the sample-based energy model's predictions. However, a comparison between the predictions of the diversified and non-diversified models when subjected to behavior change scenarios demonstrates the significance of accounting for operative diversity in such studies. The applied tests suggest that due to its unrealistic representation of the occupants' presence and behavior, the non-diversified model appears to overestimate the urban-level consequences of occupancy-driven changes in the settings of system controls. This could lead to major shifts between the expected improvements and delivered outcomes in the planning of urban-level demand side energy strategies. On the other hand, the predicted impact of the envisaged changes on the peak load was not consistent with the projected changes in the overall demand. Such miscalculations may falsify, amongst other things, the design and deployment process of renewable distributed energy generation schemes. As such, the presented illustrative examples suggest that the consideration of urban diversity, often ignored in reductive energy models, is of crucial significance in the assessment of the impact of various change and intervention scenarios.



### **Catering for the computational requirements of an integrative urban decision support environment**

The proposed *Hourglass* approach is envisaged as the core computational component of an integrative urban decision support environment, which is intended to cater for urban scale energy inquiries in view of various change and intervention scenarios. This environment is targeted at the assessment and comparative analysis of the energy and emission implications of a wide range of scenarios pertaining to the following aspects:

- Physical interventions: Thermal retrofit, densification, etc.
- Technological and infrastructural interventions: Integration of distributed generation systems, efficient heating systems, etc.
- Climatic changes: Urban Heat Island Studies, etc.
- Occupant behavior changes: Induced by demographic changes, lifestyle changes, etc.

Figure 40 schematically illustrates how the current development can be incorporated within such an environment. The generated routine represents buildings and their internal and external boundary conditions with the level of detail suited to intricate simulation assessments. As such, it has the appropriate representation resolution to support the influx of data from dedicated modules of inhabitant behaviour and microclimate for better predictive performance. The computational framework is inherently capable of incorporating detailed representations of building systems. Even though the current implementation does not consider this commodity, the method can be extended to involve system types in the clustering and/or the re-diversification process.

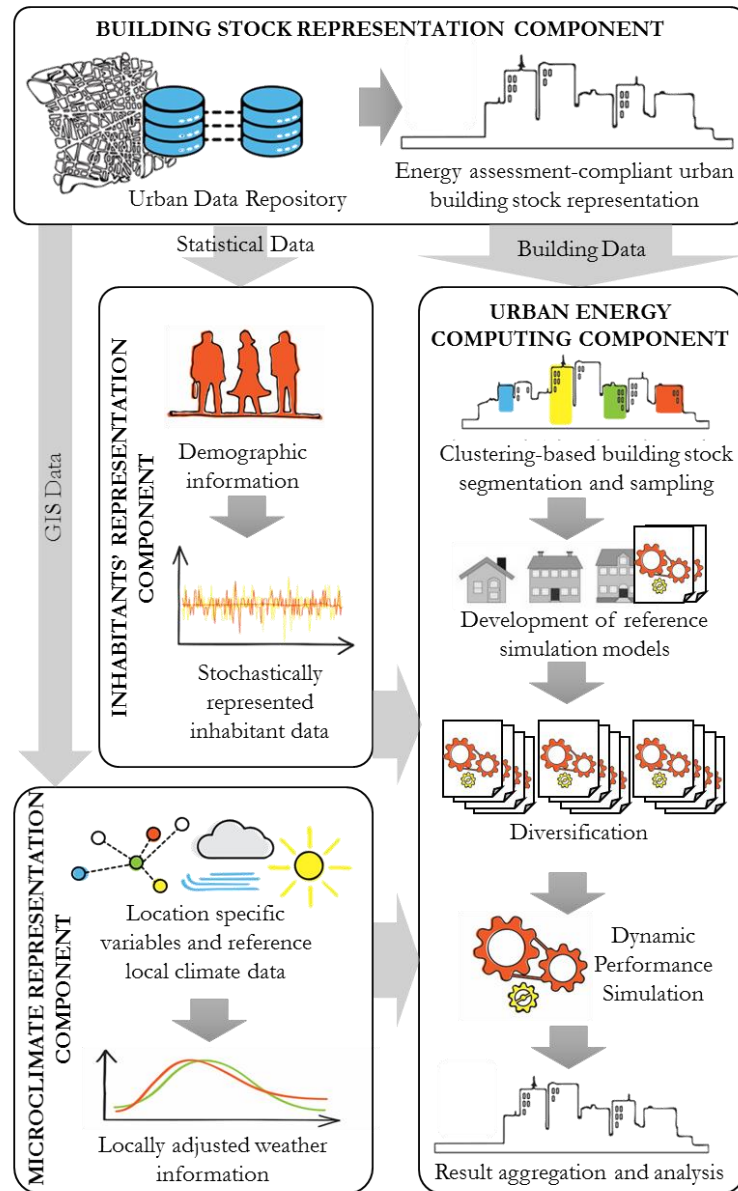


Figure 40 Schematic depiction of the incorporation of the developed Hourglass model in an integrative urban decision support environment.

## **7.2. Future Research Horizons**

The reported effort paves the way toward generation of generic urban stock energy models, which do not rely on expert or location dependent knowledge on building typologies. It enables the use of detailed simulation computations for urban-level energy inquiries, while attempting to represent the urban diversity. Despite the efforts to assess the performance of the developed model using low-resolution estimated demand data, a thorough validation of the method via highly resolved energy demand information remains a necessary step to establish its merit and usability. If the required data for such an analytical step becomes available, this data can be utilized to take a step backwards and re-evaluate and improve the set of descriptive indicators adopted for the classification process, as well as reconsider the benefits of various MCA algorithms for the efficient partitioning of the urban building stock.

Moreover, following in-depth analyses, a suitable weighting schema can be determined for the classification criteria. In a broader approach, further research on various urban contexts with diverse spatial and morphological compositions can help better understand the relationship between various city and building characteristics and their implications for energy demand. Such studies can support and improve the development of advanced tools to inform urban developmental activities and strategic planning.

The consideration of urban diversity is shown to be of marked significance in the planning of urban intervention strategies and the prediction of the impact of various changes on the urban energy demand. In the current project, some non-geometric building features were subjected to diversification. However, accounting for the diversity of some geometric aspects such as window to wall ratio can be done with acceptable effort. Future research intentions in this regard include exploring the potential of diversifying other building characteristics such as solar exposure. Also, the underlying assumptions (constant CV values) for the stochastic generation of schedules require further refinement and evaluation through empirical studies, to insure realistic representation of operational parameters. The diversification process can be a potential instrument for the calibration of urban energy models. Another interesting research endeavor would be to explore the utility of the diversification process for calibration of reductive urban energy demand models based on detailed monitored data from a subset of buildings.

Ultimately, further research is required to insure the incorporation of other useful sources of information in such energy computing schemes. Toward this end, integration of available Building Information Models, simulation results from on-site predictive control systems, and monitored demand data can be considered for real-time model evaluation and

calibration. For instance, utilization of monitored data pertaining to the representative buildings can help calibrate the model for more precision. In a more general approach, available and growing monitored data can be used to refine estimations of neighborhood level energy behavior. In this regard, the potential of statistical methods (e.g., Bayesian statistics), which facilitate the correction and readjustment of predictions as more data becomes available, can be explored. Such attempts may lead to the development of a computational environment, which is capable of aligning all available data streams to deliver consistent and reliable projections of urban energy use patterns.

## References

### Project Related Publications

Notice: Major parts of this contribution have been formerly included in the following publications.

Ghiassi, N., Mahdavi, A. 2017. Re-diversification of Predictions of a Reductive Urban Energy Modeling Method. *Proceedings of IBPSA 2017*. San Francisco, California.

Mahdavi, A., Ghiassi, N., Vuckovic, M., Taheri, M., Tahmasebi, F. 2017. High-resolution representations of internal and external boundary conditions in urban energy modelling. *Proceedings of IBPSA 2017*. San Francisco, California.

Ghiassi, N., Mahdavi, A. 2017. Reductive bottom-up urban energy computing supported by Multivariate Cluster Analysis. *Energy and Buildings*. Accepted on 2.3.2017. DOI: 10.1016/j.enbuild.2017.03.004

Ghiassi, N., Tahmasebi, F., Mahdavi, A. 2017. Harnessing buildings' operational diversity in a computational framework for high-resolution urban energy modeling. *Building Simulation*. Accepted on 8.2.2017. DOI: 10.1007/s12273-017-0356-1

Mahdavi, A., Ghiassi, N. 2017. Urban energy computing: an hourglass model. *Building Simulation Applications Proceedings*. M Baratieri, V. Corrado, A. Gasparella, F. Patuzzi (ed.); publisher of the Free University of Bolzano.

Ghiassi, N., Mahdavi, A. 2016. Utilization of GIS Data for Urban-Scale Energy Inquiries: A Sampling Approach. *Proceedings of ECPPM 2016*. Limassol, Cyprus.

- Ghiassi, N., Mahdavi, A. 2016. A GIS-Based Framework for Semi-Automated Urban-Scale Energy Simulation. *Proceedings of CESB 2016*. Prague, Czech Republic.
- Ghiassi, N., Mahdavi, A. 2016. Urban Energy Modeling Using Multivariate Cluster Analysis. *Proceedings of BAUSIM 2016*. Dresden, Germany.
- Mahdavi, A., Glawischnig, S., Ghiassi, N. 2016. Urban Energy Computing: A Multi-Layered Approach. *CEUR Workshop-Proceedings*, 1559 (2016), Vienna, Austria.
- Ghiassi, N., Hammerberg, K., Taheri, M., Pont, U., Sunanta, O., Mahdavi, A. 2015. An Enhanced Sampling-Based Approach to Urban Energy Modeling. *Proceedings of IBPSA 2015*. Hyderabad, India.
- Ghiassi, N., Glawischnig, S., Pont, U., Mahdavi, A. 2014. Toward a Data Driven Performance Guided Urban Decision Support Environment. *Information and Communication Technology*. Ir. Linawati, M.S. Mahendra, E.J. Neuhold, A. Tjoa, I. You (Eds.). Bali, Indonesia.

## **Bibliography**

- Aksoezen, M., Daniel, M., Hassler, U., and Kohler, N. 2015. Building age as an indicator for energy consumption, *Energy and Buildings*, vol. 87.
- Amtmann, M. 2010. TABULA: Reference buildings- The Austrian building typology, Austrian Energy Agency. Austria.
- ASHRAE. 2013. Standard 90.1: Energy Standard for Buildings Except Low-Rise Residential Buildings. U.S. Department of Energy, pp. 57–66. US.
- Austrian Energy Agency. 2009. Energy efficiency policies and measures in Austria: monitoring of energy efficiency in EU-27. Austria.
- Austrian Standards Institute. 2011. ÖNORM B 8110-5: Wärmeschutz im Hochbau, Teil 5: Klimamodell und Nutzungsprofile Thermal. Austria.
- Austrian Standards Institute. 2014. ÖNORM B 8110-6: Wärmeschutz im Hochbau, Teil 6: Grundlagen und Nachweisverfahren-Heizwärmebedarf und Kühlbedarf.
- Austrian Standards Institute. 2010. ÖNORM H 5059: Gesamtenergieeffizienz von Gebäuden, Beleuchtungsenergiebedarf. Austria.
- Baetens, R., and Saelens, D. 2015. Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. *Journal of Building Performance Simulation*, vol. 1493, no. September, pp. 1–17.
- Ballarini, I., Corgnati, S. P., and Corrado, V. 2014. Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project. *Energy Policy*, vol. 68, pp. 273–284.
- Baubook GmbH. 2017. Baubook rechner für bauteile. [Online]. Available: <https://www.baubook.info/zentrale/?SW=5>. [Accessed: 02-May-2017].
- Benejam, G. M. 2011. Bottom-up characterisation of the Spanish building stock – Archetype buildings and energy demand. Chalmers University of Technology.
- Benner, J., Geiger, A., and Häfele, K. 2016. Virtual 3D City Model Support for Energy Demand Simulations on City Level – The CityGML Energy Extension. In *REAL CORP 2016*, vol. 2, no. June, pp. 777–786.
- Bundesanstalt Statistik Österreich. Statistik Austria. 2017. [Online]. Available: [https://www.statistik.at/web\\_de/statistiken/index.html](https://www.statistik.at/web_de/statistiken/index.html). [Accessed: 27-Feb-2017].

- Caputo, P., Costa, G., and Ferrari, S. 2013. A supporting method for defining energy strategies in the building sector at urban scale. *Energy Policy*, vol. 55, no. Apr. pp. 261–270.
- Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, vol. 61, no. 6.
- Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. 2015. Package ‘NbClust’. [Online]. Available: <https://sites.google.com/site/malikacharrad/research/nbclust-package>. [Accessed: 19-Apr-2017].
- Citiscopes. 2015. COP 21. [Online]. Available: <http://citiscopes.org/topics/new-urban-agenda>. [Accessed: 01-Jan-2017].
- Dall’O’, G., Galante, A., and Torri, M. 2012. A methodology for the energy performance classification of residential building stock on an urban scale. *Energy and Buildings*, vol. 48, no. May. pp. 211–219.
- Dascalaki, E. G., Droutsa, K. G., Balaras, C. A., and Kontoyiannidis, S. 2011. Building typologies as a tool for assessing the energy performance of residential buildings – A case study for the Hellenic building stock. *Energy and Buildings*, vol. 43, no. 12 (Dec) pp. 3400–3409.
- Davies, D. L. and Bouldin, D. W. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 224–227.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, vol. 39, no. 1, pp. 1–38.
- Deru, M., Field, K., Studer, D., Benne, K., Griffith, B., Torcellini, P., Liu, B., Halverson, M., Winiarski, D., Rosenberg, M., Yazdanian, M., Huang, J., and Crawley, D. 2011. U.S. Department of Energy commercial reference building models of the national building stock. U.S. Department of Energy.
- Dettli, R., and Bade, S. 2007. Vorstudie zur Erhebung von Energiekennzahlen von Wohnbauten no. Bundesamt für Energie BFE. Switzerland.
- Dunn, J. C. 1974. Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, vol. 4, no. 1 (Jan), pp. 95–104.



- European Parliament and European Commission. 1995. Directive 1995/46/EC on protection of individuals with regard to the processing of personal data on the free movement of such data. *Official Journal of the European Communities*, no. L281/31. pp. 31–39.
- Farahbakhsh, H., Ugursal, V. I., and Fung, A. S. 1998. A residential end-use energy consumption model for Canada. *International Journal of Energy Research*, vol. 22, no. 13, pp. 1133–1143.
- Firth, S. K., and Lomas, K. J. 2009. Investigating Co2 Emission Reductions in Existing Urban Housing Using a Community Domestic Energy Model. *Building Simulation*, pp. 2098–2105.
- Fonseca, J. A., and Schlueter, A. 2015. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. *Applied Energy*, vol. 142, pp. 247–265.
- Fraley, C., and Raftery, A. E. 1998. How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis. *The Computer Journal*, vol. 41, no. 8, pp. 578–588.
- Fraley, C., and Raftery, A. E. 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631.
- Fraley, C., Raftery, A. E., Scrucca, L., Murphy, T. B., and Fop, M. 2015. Package ‘mclust’. [Online]. Available: <http://www.stat.washington.edu/mclust/>. [Accessed: 19-Apr-2017].
- Fraunhofer. 2017. MASEA geprüfte Datenbank. [Online]. Available: <http://www.masea-ensan.de/>. [Accessed: 02-May-2017].
- Gautier, L., and rpy2 contributors. 2014. rpy2 package. [Online]. Available: <http://rpy.sourceforge.net/rpy2/doc-dev/html/index.html>. [Accessed: 18-Apr-2017].
- Girardin, L., Marechal, F., Dubuis, M., Calame-Darbellay, N., and Favrat, D. 2010. EnerGis: A geographical information based system for the evaluation of integrated energy conversion systems in urban areas. *Energy*, vol. 35, no. 2 (Feb), pp. 830–840.
- Glawischnig, S. 2016. An Urban Monitoring System for Large-Scale Building Energy Assessment. *Dissertation*. TU Vienna. Austria.
- Google. 2017. Google Maps. [Online]. Available: 12.03.2017.
- Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. 2010. *Multivariate Data Analysis: A Global Perspective*. Pearson Education.

- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. 2002. Clustering validity checking methods. *ACM SIGMOD Rec.*, vol. 31, no. 3 (Sep), p. 19.
- Halkidi, M., Vazirgiannis, M., and Batistakis, Y. 2000. Quality scheme assessment in the clustering process. *Principles of Data Mining and Knowledge Discovery*. pp. 265–276.
- Hammerberg, K. 2016. DEMTools Plug-in for QGIS. [Online]. Available: <https://plugins.qgis.org/plugins/DEMTools/>. [Accessed: 13-Mar-2017].
- Heiple, S. and Sailor, D. J. 2008. Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles. *Energy and Buildings*, vol. 40, no. 8 (Jan), pp. 1426–1436.
- Hens, H., Verbeeck, G., and Verdonck, B. 2001. Impact of energy efficiency measures on the CO<sub>2</sub> emissions in the residential sector, a large scale analysis. *Energy and Buildings*, vol. 33, pp. 3–9.
- Hollands, R. G. 2008. Will the real smart city please stand up? *City*, vol. 12, no. 3, pp. 303–320.
- Howard, B., Parshall, L., Thompson, J., Hammer, S., Dickinson, J., and Modi, V. 2012. Spatial distribution of urban building energy consumption by end use. *Energy and Buildings* vol. 45, no. Feb, pp. 141–151.
- Huang, Y. J., and Brodrick, J. 2000. A Bottom-Up Engineering Estimate of the Aggregate Heating and Cooling Loads of the Entire US Building Stock Prototypical Residential Buildings. In *2000 ACEEE Summer Study on Energy Efficiency in Buildings*, pp. 135–148.
- Intelligent Energy Europe. 2012. Project TABULA. [Online]. Available: <http://episcopes.eu/iee-project/tabula/>. [Accessed: 03-Mar-2017].
- International Energy Agency. 2014. Final consumption: World. [Online]. Available: [https://www.iea.org/Sankey/#?c=Austria&s=Final consumption](https://www.iea.org/Sankey/#?c=Austria&s=Final%20consumption). [Accessed: 12-Oct-2016].
- James J. Hirsch & Associates. 2016. DOE2. [Online]. Available: <http://doe2.com/>. [Accessed: 03-Mar-2017].
- Jones, P., Lannon, S., and Williams, J. 2001. Modelling building energy use at urban scale. *Seventh International IBPSA Conference*, pp. 175–180. Rio de Janeiro, Brazil.
- Jung, Y., Park, H., Du, D. Z., and Drake, B. L. 2003. A decision criterion for the optimal number of clusters in hierarchical clustering. *Journal of Global Optimization*, vol. 25, no. 1, pp. 91–111.

- Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., and Djurovic-Petrovic, M. 2010. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, vol. 45, no. 7 (Jul), pp. 1683–1697.
- Keirstead, J., Jennings, M., and Sivakumar, A. 2012. A review of urban energy system models: Approaches, challenges and opportunities. *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6 (Aug), pp. 3847–3866.
- Kemna, R., and Acedo, J. 2014. Average EU building heat load for HVAC equipment. *DG ENER C.3*. Van Holsteijn en Kemna B.V. The Netherlands.
- Langevin, J., Wen, J., and Gurian, P. L. 2015. Simulating the human-building interaction: Development and validation of an agent-based model of office occupant behaviors. *Building and Environment*, vol. 88, pp. 27–45.
- Levine, M., Uerge-Vorsatz, D., Blok, K., Geng, L., Harvey, D., Lang, S., Levenmore, G., Mongameli Mehlwana, A., Mirasgedis, S., Novikova, A., Rilling, J., and Yoshino, H. 2007. Residential and commercial buildings. In *Climate Change 2007: Working Group III: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Metz, B., Davidson, O. R., Bosch, P., Dave, R., and Meyer, L. A., Eds. Cambridge University Press.
- Li, Q., Quan, J. S., Augenbroe, G., Yang, P.-J., and Brown, J. 2015. Building Energy Modelling at Urban Scale: Integration of Reduced Order Energy Model with Geographical Information. *14<sup>th</sup> IBPSA International Conference*, p. 190. Hyderabad, India.
- Liu, Y., Li, Z., Xiong, H., Gao, X., and Wu, J. 2010. Understanding of internal clustering validation measures. *IEEE International Conference on Data Mining*, pp. 911–916,
- Lloyd, S. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137.
- MacQueen, J. B. 1967. Kmeans Some Methods for classification and Analysis of Multivariate Observations. *5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 233, pp. 281–297.
- Magistrat der Stadt Wien. 2017a. Baupolizei (MA 37). [Online]. Available: <https://www.wien.gv.at/wohnen/baupolizei/>. [Accessed: 01-May-2017].
- Magistrat der Stadt Wien. 2017b. Vienna GIS. [Online]. Available: <https://www.wien.gv.at/viennagis/>. [Accessed: 13-Mar-2017].

- Mahdavi, A., and El-Bellahy, S. 2005. Effort and effectiveness considerations in computational design evaluation: A case study. *Building and Environment*, vol. 40, no. 12, pp. 1651–1664.
- Mahdavi, A., and Tahmasebi, F. 2015. The Inter-Individual Variance of the Defining Markers of Occupancy Patterns in Office Buildings: A Case Study. *14<sup>th</sup> IBPSA International Conference*, pp. 2243–2247. Hyderabad, India.
- Manfredi, M., Caputo, P., and Costa, G. 2011. Paradigm shift in urban energy systems through distributed generation: Methods and models. *Applied Energy*, vol. 88, no. 4, pp. 1032–1048.
- Maulik, U., and Bandyopadhyay, S. 2002. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1650–1654.
- Mirzaei, P. A., and Haghighat, F. 2010. Approaches to study Urban Heat Island - Abilities and limitations. *Building and Environment*, vol. 45, no. 10, pp. 2192–2201.
- Munoz, E. M., and Peters, I. 2014. Constructing an Urban Microsimulation Model to Assess the Influence of Demographics on Heat Consumption. *International Journal of Microsimulation*, vol. 7, pp. 127–157.
- Natural Resources Canada. 2016. HOT2000 Software Suite. [Online]. Available: <http://www.nrcan.gc.ca/energy/efficiency/housing/home-improvements/17725>. [Accessed: 03-Mar-2017].
- Nouvel, R., Kaden, R., Bahu, J., Kaempf, J., Cipriano, P., Lauster, M., Benner, J., Munoz, E., Tournaire, O., and Casper, E. 2016. Genesis of the CityGML Energy ADE. In *Cisbat2015*, pp. 931–936. Lausanne, Switzerland.
- NREL. 2017. EnergyPlus. EnergyPlus Energy Simulation Software National Renewable Energy Laboratories. U.S. Department of Energy. [Online]. Available: <https://energyplus.net/>. [Accessed: 01-Mar-2017].
- NREL, ANL, LBNL, ORNL, and PNNL. 2017. OpenStudio. [Online]. Available: <https://www.openstudio.net/>. [Accessed: 01-May-2017].
- NumPy developers. 2017. NumPy. [Online]. Available: <http://www.numpy.org/>. [Accessed: 04-May-2017].
- Open Geospatial Consortium. 2017. CityGML. [Online]. Available: <http://www.opengeospatial.org/standards/citygml>. [Accessed: 15-Mar-2017].

- Orehounig, K., Mavromatidis, G., Evins, R., Dorer, V., and Carmeliet, J. 2014. Predicting Energy Consumption of a Neighborhood Using Building Performance Simulations. *Building Simulation and Optimization Conference*, London, England.
- Österreichisches Institut für Bautechnik. 2015. Leitfaden OIB-Richtlinie 6: Energietechnisches Verhalten von Gebäuden. Vienna, Austria.
- Page, J., Dervey, S., and Morand, G. 2014. Aggregating building energy demand simulation to support urban energy design. *PLEA 2014*, pp. 1–8.
- Page, J., Robinson, D., Morel, N., and Scartezzini, J. L. 2008. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings*, vol. 40, no. 2, pp. 83–98.
- Parekh, A. 2005. Development of Archetypes of Building Characteristics Libraries for Simplified Energy Use Evaluation of Houses. *Ninth International IBPSA Conference*, pp. 921–928. Montreal, Canada.
- Philip, S. 2013. Eppy Package. [Online]. Available: <http://pythonhosted.org/eppy/>. [Accessed: 04-May-2017].
- Python Software Foundation. 2017. Python. [Online]. Available: <https://www.python.org/>. [Accessed: 12-Mar-2017].
- Ratti, C., Baker, N., and Steemers, K. 2005. Energy consumption and urban texture. *Energy and Buildings*, vol. 37, no. 7 (Jul), pp. 762–776.
- Reinhart, C. F., and Cerezo Davila, C. 2016. Urban building energy modeling - A review of a nascent field. *Building and Environment*, vol. 97, pp. 1-15. Taylor and Francis.
- Remmen, P., Lauster, M., Mans, M., Fuchs, M., Osterhage, T., and Mueller, D. 2017. TEASER: an open tool for urban energy modelling of building stocks. *Journal of Building Performance Simulation*, vol. 1493 (Oct), pp. 762–776.
- Ribas Portella, J. M. 2012. Bottom-up description of the French building stock, including archetype buildings and energy demand. *Thesis*. Chalmers university of technology. Gothenburg, Sweden.
- Rizwan, A. M., Dennis, L. Y. C., and Liu, C. 2008. A review on the generation, determination and mitigation of Urban Heat Island. *Journal of Environmental Sciences*, vol. 20, no. 1, pp. 120–128.
- Robinson, D. 2006. Urban morphology and indicators of radiation availability. *Solar Energy*, vol. 80, no. 12 (Dec), pp. 1643–1648.

- Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65.
- Sansregret, S., and Millette, J. 2009. Development of a Functionality Generating Simulations of Commercial and Institutional Buildings Having Representative Characteristics of a Real Estate Stock in Québec (Canada). *11th IBPSA Conference*, pp. 1437–1443. Glasgow, Scotland.
- Schöberl, H., Hofer, R., and Lang, C. 2012. *Handbuch thermische Gebäudesanierung: Optimale Ausführungsvarianten*. Landesinnung Bau Niederösterreich. Austria.
- Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464.
- Shimoda, Y., Fujii, T., Morikawa, T., and Mizuno, M. 2003. Development of Residential Energy End-Use Simulation Model at City Scale. *Eighth International IBPSA Conference*, pp. 1201–1208. Eindhoven, The Netherlands.
- Snäkin, J. 2000. An engineering model for heating energy and emission assessment: The case of North Karelia, Finland. *Applied Energy*, vol. 67.
- Staniszewski, A., and Gierga, M. 2016. *Energie-einsparverordnung: Leitfaden für Wohngebäude*. Bonn, Germany.
- Swan, L. G. and Ugursal, V. I. 2009. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8 (Oct), pp. 1819–1835.
- The Open Source Geospatial Foundation. 2017. QGIS. [Online]. Available: <http://www.qgis.org/en/site/about/index.html>. [Accessed: 12-Mar-2017].
- The Pennsylvania State University. 2017. Applied Multivariate Statistical Analysis: Cluster Analysis. (*Online course*) [Online]. Available: <https://onlinecourses.science.psu.edu/stat505/node/138>. [Accessed: 12-Apr-2017].
- The R Foundation. 2017. The R Project for Statistical Computing. [Online]. Available: <https://www.r-project.org/>. [Accessed: 12-Mar-2017].
- The World Bank. 2016. World Bank Open Data. [Online]. Available: <http://data.worldbank.org/>. [Accessed: 12-Oct-2016].
- Theodoridou, I., Papadopoulos, A. M., and Hegger, M. 2011a. A typological classification of the Greek residential building stock. *Energy and Buildings*, vol. 43, no. 10, pp. 2779–2787.

- Theodoridou, I., Papadopoulos, A. M., and Hegger, M. 2011b. Statistical analysis of the Greek residential building stock. *Energy and Buildings*, vol. 43, no. 9, pp. 2422–2428.
- Tibshirani, R., Walther, G., and Hastie, T. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, pp. 411–423.
- Trimble Inc. 2017. SketchUp. [Online]. Available: <https://www.sketchup.com/>. [Accessed: 01-May-2017].
- Tuominen, P., Holopainen, R., Eskola, L., Jokisalo, J., and Airaksinen, M. 2014. Calculation method and tool for assessing energy consumption in the building stock, *Building and Environment*, vol. 75, pp. 153–160.
- U.S. Energy Information Administration. 2016. Commercial Energy Consumption Survey. [Online]. Available: <https://www.eia.gov/consumption/commercial/reports/2012/energyusage/>. [Accessed: 01-Mar-2017].
- UNEP. 2009. *Buildings and Climate Change: Summary for Decision Makers*. pp. 1–62.
- United Kingdom Department for Business Energy & Industrial Strategy. 2013. Standard Assessment Procedure. [Online]. Available: <https://www.gov.uk/guidance/standard-assessment-procedure>. [Accessed: 02-Mar-2017].
- Verein OpenStreet Map Austria. 2017. OpenStreetMap Austria. [Online]. Available: <https://openstreetmap.at/>. [Accessed: 13-Mar-2017].
- Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, vol. 58, no. 301.
- Wetter, M., and Van Treeck, C. 2017. *New generation computational tools for building and community energy systems (Annex 60 final report)*. International Energy Agency, pp. 363–382.
- Wilby, R. L., and Wigley, T. M. L. 1997. Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, vol. 21, no. 4, pp. 530–548.
- Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62.
- Wilson, H. G., Boots, B., and Millward, A. 2002. A comparison of hierarchical and partitional clustering techniques for multispectral image classification. *IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, no. 519, pp. 1624–1626.

Yan, D., O'Brien, W., Hong, T., Feng, X., Burak Gunay, H., Tahmasebi, F., and Mahdavi, A. 2015. Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, vol. 107, pp. 264–278.



## List of Abbreviations

### General

---

DG	Distributed Generation
UBEM	Urban Building Energy Model
GCM	Global Circulation Model
CFD	Computational Fluid Dynamics
BPS	Building Performance Simulation
GIS	Geographic Information Systems
SVF	Sky View Factor
OSM	Open Street Map
MCA	Multivariate Cluster Analysis
EM	Expectation Maximization
BIC	Bayesian Information Criterion
CSV	Comma Separated Values
CV	coefficient of variance
IDF	Input Data File
HVAC	Heating Ventilation and Air-Conditioning

---

### Project Related

---

NDM	Non-diversified model
DMS	Model with diversified schedules
DMA	Model with diversification of all considered parameters

---

## List of Symbols and Units

### The Reductive Module

Symbol	Description	Units
$V_n$	net volume of the building	$m^3$
$V$	Urban air temperature	$m^3$
$f_n$	ratio of net to gross volume	-
$A_e$	thermally effective envelope area	$m^2$
$A_i$	area of a building enclosure	$m^2$
$f_{t,i}$	temperature correction factor	-
$C_t$	thermal compactness of the building	$m$
$h_e$	effective floor height	$m$
$A_f$	area of the building's footprint	$m^2$
$n_f$	number of floors	-
$GR_e$	effective glazing ratio	-
$WWR$	window to wall ratio	-
$GWR$	glazing to window ratio	-
$g$	solar energy transmission of glazing	-
$A_{ow,i}$	Area of an external wall	$m^2$
$f_{o,i}$	orientation factor of a wall	-
$SVF_i$	Sky View Factor of a wall	-
$r_{o,j}$	Monthly radiation value for an orientation	$kWh \cdot m^{-2}$

Symbol	Description	Units
$r_{s,j}$	Monthly radiation value for the south orientation	$kWh \cdot m^{-2}$
$U_e$	Average effective envelope U-value	$W \cdot m^{-2} \cdot K^{-1}$
$U_i$	U-value of an enclosure	$W \cdot m^{-2} \cdot K^{-1}$
$U_{w,e}$	average effective wall U-value	$W \cdot m^{-2} \cdot K^{-1}$
$U_{w,i}$	U-value of a wall element	$W \cdot m^{-2} \cdot K^{-1}$
$A_{w,i}$	area of a wall element	$m^2$
$U_{c,e}$	average effective ceiling/roof U-value	$W \cdot m^{-2} \cdot K^{-1}$
$U_{c,i}$	U-value of a ceiling/roof element	$W \cdot m^{-2} \cdot K^{-1}$
$A_{c,i}$	area of a ceiling/roof element	$m^2$
$U_{f,e}$	average effective floor U-value	$W \cdot m^{-2} \cdot K^{-1}$
$U_{f,i}$	U-value of a floor element	$W \cdot m^{-2} \cdot K^{-1}$
$A_{f,i}$	area of a floor element	$m^2$
$O_u$	Annual use hours fraction	-
$t_{use,a,i}$	number of annual use hours for a usage	$h. a^{-1}$
$f_{v,i}$	fraction of the volume associated with a usage	-
$t_a$	total number of hours in a year	$h. a^{-1}$
$O_{d/u}$	daytime use intensity	-
$t_{day,a,i}$	annual operation hours in daytime	$h. a^{-1}$
$t_{night,a,i}$	annual operation hours in nighttime	$h. a^{-1}$

Symbol	Description	Units
$O_d$	annual daytime use fraction	-
$O_n$	annual nighttime use fraction	-
$q_{i,h}$	average area-related internal gains rate	$W.m^{-2}$
$q_{i,h,i}$	area-related rate of internal gains from equipment, and occupants in heating season	$W.m^{-2}$
$q_{i,l,i}$	area-related rate of internal gains from lighting during the heating season	$W.m^{-2}$
$LED_i$	benchmark value for the lighting energy demand	$kWh.m^{-2}.a^{-1}$
$Ig_d$	daily area-related internal gains	$Wh.m^{-2}.d^{-1}$
$t_{use,d,i}$	daily operation time for a usage	$h.d^{-1}$
$n_v$	hourly ventilation rate	$h^{-1}$
$n_{v,i}$	ventilation rate for a usage	$h^{-1}$
$Ac_d$	Daily air-change rate	$d^{-1}$
$Q_h$	monthly heating demand	$kWh.M^{-1}$
$Q_T$	Monthly transmission losses	$kWh.M^{-1}$
$Q_V$	Monthly ventilation losses	$kWh.M^{-1}$
$Q_i$	Monthly internal gains	$kWh.M^{-1}$
$Q_s$	Monthly solar gains	$kWh.M^{-1}$
$\eta_h$	monthly utilization factor for heat gains	-
$L_T$	transmission heat transfer coefficient	$W.K^{-1}$

Symbol	Description	Units
$\Delta\theta$	difference between internal temperature and average monthly external temperature	$K$
$t$	duration of the month	$h.M^{-1}$
$L_\psi$	transmission heat coefficient due to linear thermal bridges	$W.K^{-1}$
$L_\chi$	transmission heat coefficient due to punctual thermal bridges	$W.K^{-1}$
$L_V$	ventilation heat transfer coefficient	$W.K^{-1}$
$c_{Vp,L} \times P_L$	volumetric heat capacity of air	$Wh.m^{-3}.K^{-1}$
$v_V$	Hourly airflow	$m^3.h^{-1}$
$t_{use,m,i}$	monthly use hours	$h.M^{-1}$
$V_V$	effective volume of the building	$m^3$
$I_{s,j}$	monthly global irradiance on a surface with orientation $j$	$kWh.m^{-2}.M^{-1}$
$A_{trans,i,j}$	effective area of a transparent building component with orientation $j$	$m^2$
$\gamma_h$	heat balance ratio in the heating season	-
$a$	numerical parameter for the utilization factor	-
$a_0$	dimensionless reference numerical parameter	-
$C$	effective heat capacity of the building	$Wh.K^{-1}$
$\tau_0$	the reference time constant	$h$

Symbol	Description	Units
$f_{BW}$	volumetric heat capacity of the building	$Wh.m^{-3}.K^{-1}$
$Q_{predicted,i,j}$	predicted annual heating demand of the building $i$ , belonging to the cluster $j$	$kWh.a^{-1}$
$V_{n,i}$	net volume of the building $i$	$m^3$
$Q_{v,rep,j}$	volumetric annual heating demand of the building representing the cluster $j$	$kWh.m^{-3}.a^{-1}$
$\delta_{Total}$	relative neighborhood-level error	-
$Q_{theoretical,i}$	theoretical annual heating demand computed for the building $i$	$kWh.a^{-1}$
$\delta_{Mean}$	mean building-level error	-
$\delta_i$	relative prediction error of building $i$	-
$n$	number of buildings within the study area.	-
$f_{>20\%}$	the fraction of the volume associated with a building-level prediction error of above 20%	-

**The Re-Diversification Module**

Symbol	Description	Units
$Ig_{a,Building}$	area-related annual internal gains of each building	$Wh.m^{-2}.a^{-1}$
$\overline{n_{v,Building}}$	weighted average air change rate of the building	$h^{-1}$
$Ig_a$	area-related annual internal gains for a usage	$Wh.m^{-2}.a^{-1}$
$Ig_{a,L}$	area-related annual internal gains from lighting	$Wh.m^{-2}.a^{-1}$
$Ig_{a,E}$	annual area-related internal gains from equipment	$Wh.m^{-2}.a^{-1}$
$Ig_{a,P}$	annual area-related internal gains from people	$Wh.m^{-2}.a^{-1}$
$f_L$	share of internal gains from lighting	-
$f_E$	share of internal gains from equipment	-
$f_P$	share of internal gains from people	-
$P_{L/E}$	Reference area-related lighting or equipment power	$W.m^{-2}$
$HR_{L/E}$	aggregated annual full load hours of lighting or equipment use	$h.a^{-1}$
$N_P$	area-related number of inhabitants	$person.m^{-2}$
$HR_P$	hourly presence rate of inhabitants	$h.a^{-1}$
$H_{mr}$	metabolic rate of inhabitants	$W.person^{-1}$
$\overline{n_v}$	average hourly air change rate	$h^{-1}$

Symbol	Description	Units
$n_{v,0}$	reference value for air change rate	$h^{-1}$
$\sum (HR_V)$	number of full load operation hours of ventilation in a year	$h. a^{-1}$
$Q_{i,h}$	heating demand of an arbitrary building $i$ in timestep $h$	$kWh$
$Q_{Sim,i,h}$	heating demand of the diversified simulation model associated with building $i$ in timestep $h$	$kWh$
$V_{reference,i}$	the net volume of the reference building	$m^3$
$L_t$	lighting use rate in time step $t$	-
$O_t$	Inhabitant presence in time step $t$	-
$f_{L,t}$	multiplier derived from standard schedules for time step $t$ (lighting)	-
$L_{0,t}$	minimum lighting use rate of the timestep regardless of occupant presence	-
$E_t$	Equipment use rate in time step $t$	-
$f_{E,t}$	multiplier derived from standard schedules for time step $t$ (equipment)	-
$E_{0,t}$	minimum equipment use rate of the timestep regardless of occupant presence	-
$f_{Ig}$	multiplier for the readjustment of internal gains	-
$Ig_{a,Building,Target}$	area-related annual internal gains of the target building	$Wh. m^{-2}. a^{-1}$



Symbol	Description	Units
$Ig_{a,Building,Reference}$	area-related annual internal gains of the reference building	$Wh.m^{-2}.a^{-1}$
$Ig_{a,L/E/P,Target}$	area-related internal gains of target building (lighting/equipment/people)	$Wh.m^{-2}.a^{-1}$
$Ig_{a,L/E/P,Reference}$	area-related internal gains of reference building (lighting/equipment/people)	$Wh.m^{-2}.a^{-1}$
$f_{\overline{n_v}}$	multiplier for the readjustment of average air change rate	-
$\overline{n_{v,Building,Target}}$	average hourly air change rates of the target building	$h^{-1}$
$\overline{n_{v,Building,Reference}}$	average hourly air change rates of the reference building	$h^{-1}$
$U_{Target}$	effective U-values of a component of target building	$W.m^{-2}.K^{-1}$
$U_{Reference}$	effective U-values of a component of reference building	$W.m^{-2}.K^{-1}$
$R_{Target}$	thermal resistance of a component in target building	$K.m^2.W^{-1}$
$R_{Reference}$	thermal resistance of a component in reference building	$K.m^2.W^{-1}$
$\lambda_{Target}$	thermal conductivities of a material in the target building	$W.m^{-1}.K^{-1}$
$\lambda_{Reference}$	thermal conductivities of a material in the reference building	$W.m^{-1}.K^{-1}$
$d$	thickness of a material layer in the target/reference buildings	$m$

## List of Tables (Main text)

Table 1 An overview of the consulted urban energy models.....	20
Table 2 Building stock classification criteria adopted by previous reductive efforts .....	23
Table 3 Summary of the available data sources and the contained information.....	34
Table 4 Temperature correction factors for various building enclosure elements based on their boundary conditions as given in the austrian standard b 8110-6 (austrian standards institute 2014).....	46
Table 5 An overview of the default values for operational parameters for the usages present in the current project (source : austrian standards institute 2011; austrian standards institute 2010).....	54
Table 6 An overview of the various indicators that have been considered for building stock classification with their computation method and the associated input parameters .....	56
Table 7 Various sets of descriptive indicators considered as input for cluster analysis .....	67
Table 8 Values considered for the volumetric heat capacity of buildings according to construction period.....	76
Table 9 An overview of the values of the performance indicators computed for the 33 partitioning scenarios.....	81
Table 10 Mean values of the employed descriptive indicators across clusters s4-kmeans clustering scenario.....	83
Table 11 An overview of the age and usage composition of the generated classes and the associated representative buildings .....	85
Table 12 Overview of the diversification scenarios.....	107
Table 13 Overview of the base case and scenario assumptions for hvac operation .....	108
Table 14 Summary of the results of modeled diversification scenarios. The non-diversified model is considered as the reference model in the calculation of the relative values.....	113
Table 15 Deviations in the annual demand of clusters as predicted by diversified models from the non-diversified model predictions.....	113
Table 16 Results of the behavior change scenarios as simulated by the non-diversified and diversified computational models.....	116

## **List of Tables (Appendix 2)**

Table 2.1 Description of the building representing cluster 1 .....	192
Table 2.2 Description of the building representing cluster 2 .....	194
Table 2.3 Description of the building representing cluster 3 .....	196
Table 2.4 Description of the building representing cluster 4 .....	198
Table 2.5 Description of the building representing cluster 5 .....	200
Table 2.6 Description of the building representing cluster 6 .....	202
Table 2.7 Description of the building representing cluster 7 .....	204

## List of Figures (Main text)

Figure 1 Global building-related energy consumption (international energy agency 2014)	3
Figure 2 Conceptual framework of an integrative urban decision support environment ...	7
Figure 3 (Left) Residential building energy consumption by end use in Austria (based on data from Austrian Energy Agency 2009) .....	8
Figure 4 (Right) Commercial building energy consumption by end use in the United States (based on data from U.S. Energy Information Administration 2016) .....	8
Figure 5 An overview of urban energy modeling approaches .....	11
Figure 6 Overall structure of the proposed urban energy computing framework.....	27
Figure 7 The bird's eye view of the neighborhood selected as a case study, from Google Maps (Google 2017) .....	28
Figure 8 Distribution of buildings in the study area by usage and age .....	29
Figure 9 GIS data incorporated in the present study visualized in a GIS platform. ....	32
Figure 10 Building information required for the quantification of the various heat transfer processes occurring in a building.....	35
Figure 11 The building data representation schema developed for the structuring of building information from various sources, towards generation of necessary descriptions for all buildings within the study domain .....	36
Figure 12 Identification of vertical elements enclosing the building envelope: polygons 1 is associated with a higher eaves height than polygon 2. Once the polygons are extruded by the given heights to form the volume of the building, the resulting surfaces are not all considered parts of the building envelope. In this case the shared wall marked in dotted line is excluded from the bounding envelope, whereas the wall element marked in darker color is included. ....	37
Figure 13 Floor elements are assumed to be adjacent to an unheated basement unless their constituting "Building Part" is associated with an elevation above ground level. ....	38
Figure 14 Rule-based identification of roof type and attic space condition .....	38
Figure 15 Determining the external boundary conditions of wall elements: Considering the building in the middle, the wall marked in dotted line is an adiabatic wall (adjacent to a main building), whereas the one marked in continuous line is a wall adjacent to an unconditioned space (utility building). All other wall elements are adjacent to outside air. ....	39
Figure 16 Summary of the steps required for the generation of an energy-assessment-compliant representation of the urban building stock .....	42

Figure 17 The steps of k-mean clustering method depicted with a sample of two-dimensional data points. In this example, the desired number of clusters is four, and the initial seeds are selected at random.....	61
Figure 18 Consecutive steps of hierarchical agglomerative clustering depicted on a sample of six points on the two-dimensional plane.....	62
Figure 19 (Left) Dendrogram pertaining to the example illustrated in Figure 18.....	63
Figure 20 (Right) Possible partitioning schemas resulting in the identification of 2, 3, 4, and 5 clusters.....	63
Figure 21 Comparison of the performance of the three clustering algorithms regardless of input parameters.....	78
Figure 22 Comparison of the performance of the various sets of performance indicators on the quality of the emerging partitioning and building representation, regardless of the clustering algorithm.....	79
Figure 23 Comparison of the performance of the partitioning schemas in view of the number of the buildings representing the neighborhood in each case.....	80
Figure 24 The visualized results of the reductive procedure for the S4-Kmeans clustering scenario.....	82
Figure 25 Volumetric heating demand of buildings in each cluster and the cluster representative. The lighter coloured columns pertain to non-residential buildings ..	83
Figure 26 Theoretical against predicted values of annual heating demand based on the representation resulting from S4-Kmeans scenario.....	84
Figure 27 (left) Effective envelope U-value by year of construction.....	87
Figure 28 (Right) Volumetric heating demand by year of construction.....	87
Figure 29 Ranges of volumetric heating demand values across classes emerging from the multivariate cluster analysis procedure adopted in the current effort.....	88
Figure 30 (Left) Modeling of space enclosures in the simulation models.....	93
Figure 31 (Right) An example of the adopted zoning convention based on space usage and connectivity, applied to an arbitrary building floor: Four zones are identified in this floor plan.....	93
Figure 32 An instance of hourly metabolic rate values generated for the three usages on a typical day.....	109
Figure 33 Reference schedules versus diversified schedules generated for a building with office usage representing a. Occupancy, b. Lighting, and c. Equipment.....	110
Figure 34 Reference schedules versus diversified schedules generated for a building with gastronomy usage representing a. Occupancy, b. Lighting, and c. Equipment.....	111
Figure 35 Reference schedules versus diversified schedules generated for a building with residential usage representing a. Occupancy, b. Lighting, and c. Equipment.....	112

Figure 36 Percentage deviation of the volumetric annual heating demand of buildings in each cluster, computed by the DMA from that of the relevant reference model (NDM). .....	114
Figure 37 Relative deviation of hourly demand results of all buildings as predicted by the DMA and DMS from NDS predictions for a single time step. The x-axis represents the 744 buildings in the study area. ....	114
Figure 38 (Left) Comparative analysis of the hourly demand predicted in various scenarios by DMA.....	117
Figure 39 (Right) Cumulative relative frequency of the hourly heating load in relation to annual peak load .....	117
Figure 40 Schematic depiction of the incorporation of the developed Hourglass model in an integrative urban decision support environment. ....	124

## **List of Figures (Appendix 1)**

Figure 1.1 Visualized results of clustering scenario 1, K-means Method.....	158
Figure 1.2 Visualized results of clustering scenario 1, Hierarchical Method.....	159
Figure 1.3 Visualized results of clustering scenario 1, Model-based Method.....	160
Figure 1.4 Visualized results of clustering scenario 2, K-means Method.....	161
Figure 1.5 Visualized results of clustering scenario 2, Hierarchical Method.....	162
Figure 1.6 Visualized results of clustering scenario 2, Model-based Method.....	163
Figure 1.7 Visualized results of clustering scenario 3, K-means Method.....	164
Figure 1.8 Visualized results of clustering scenario 3, Hierarchical Method.....	165
Figure 1.9 Visualized results of clustering scenario 3, Model-based Method.....	166
Figure 1.10 Visualized results of clustering scenario 4, K-means Method.....	167
Figure 1.11 Visualized results of clustering scenario 4, Hierarchical Method.....	168
Figure 1.12 Visualized results of clustering scenario 4, Model-based Method.....	169
Figure 1.13 Visualized results of clustering scenario 5, K-means Method.....	170
Figure 1.14 Visualized results of clustering scenario 5, Hierarchical Method.....	171
Figure 1.15 Visualized results of clustering scenario 5, Model-based Method.....	172
Figure 1.16 Visualized results of clustering scenario 6, K-means Method.....	173
Figure 1.17 Visualized results of clustering scenario 6, Hierarchical Method.....	174
Figure 1.18 Visualized results of clustering scenario 6, Model-based Method.....	175
Figure 1.19 Visualized results of clustering scenario 7, K-means Method.....	176
Figure 1.20 Visualized results of clustering scenario 7, Hierarchical Method.....	177
Figure 1.21 Visualized results of clustering scenario 7, Model-based Method.....	178
Figure 1.22 Visualized results of clustering scenario 8, K-means Method.....	179
Figure 1.23 Visualized results of clustering scenario 8, Hierarchical Method.....	180
Figure 1.24 Visualized results of clustering scenario 8, Model-based Method.....	181
Figure 1.25 Visualized results of clustering scenario 9, K-means Method.....	182
Figure 1.26 Visualized results of clustering scenario 9, Hierarchical Method.....	183
Figure 1.27 Visualized results of clustering scenario 9, Model-based Method.....	184
Figure 1.28 Visualized results of clustering scenario 10, K-means Method.....	185
Figure 1.29 Visualized results of clustering scenario 10, Hierarchical Method.....	186
Figure 1.30 Visualized results of clustering scenario 10, Model-based Method.....	187
Figure 1.31 Visualized results of clustering scenario 11, K-means Method.....	188
Figure 1.32 Visualized results of clustering scenario 11, Hierarchical Method.....	189
Figure 1.33 Visualized results of clustering scenario 11, Model-based Method.....	190

## List of Figures (Appendix 2)

Figure 2.1 South-West View .....	192
Figure 2.2 North-East View .....	192
Figure 2.3 Typical floor plan, the conditioned area is highlighted .....	193
Figure 2.4 Main elevation .....	193
Figure 2.5 (Left) North-West View.....	194
Figure 2.6 (Right) South-East View .....	194
Figure 2.7 Typical floor plan, the conditioned area is highlighted .....	195
Figure 2.8 Main elevation .....	195
Figure 2.9 (Left) East View.....	196
Figure 2.10 (Right) West View .....	196
Figure 2.11 Typical floor plan, the conditioned area is highlighted .....	197
Figure 2.12 Courtyard elevation-section .....	197
Figure 2.13 (Left) East View .....	198
Figure 2.14 (Right) West View .....	198
Figure 2.15 Typical floor plan, the conditioned area is highlighted.....	199
Figure 2.16 Main elevation .....	199
Figure 2.17 (Left) West View .....	200
Figure 2.18 (Right) East View .....	200
Figure 2.19 (Left) Typical floor plan, the conditioned area is highlighted .....	201
Figure 2.20 (Right) Main elevation .....	201
Figure 2.21 (Left) North view .....	202
Figure 2.22 (Right) South view .....	202
Figure 2.23 Typical floor plan, the conditioned area is highlighted .....	203
Figure 2.24 East elevation .....	203
Figure 2.25 (Left) West view .....	204
Figure 2.26 (Right) East view .....	204
Figure 2.27 Typical floor plan, the conditioned area is highlighted .....	205
Figure 2.24 Main elevation .....	205



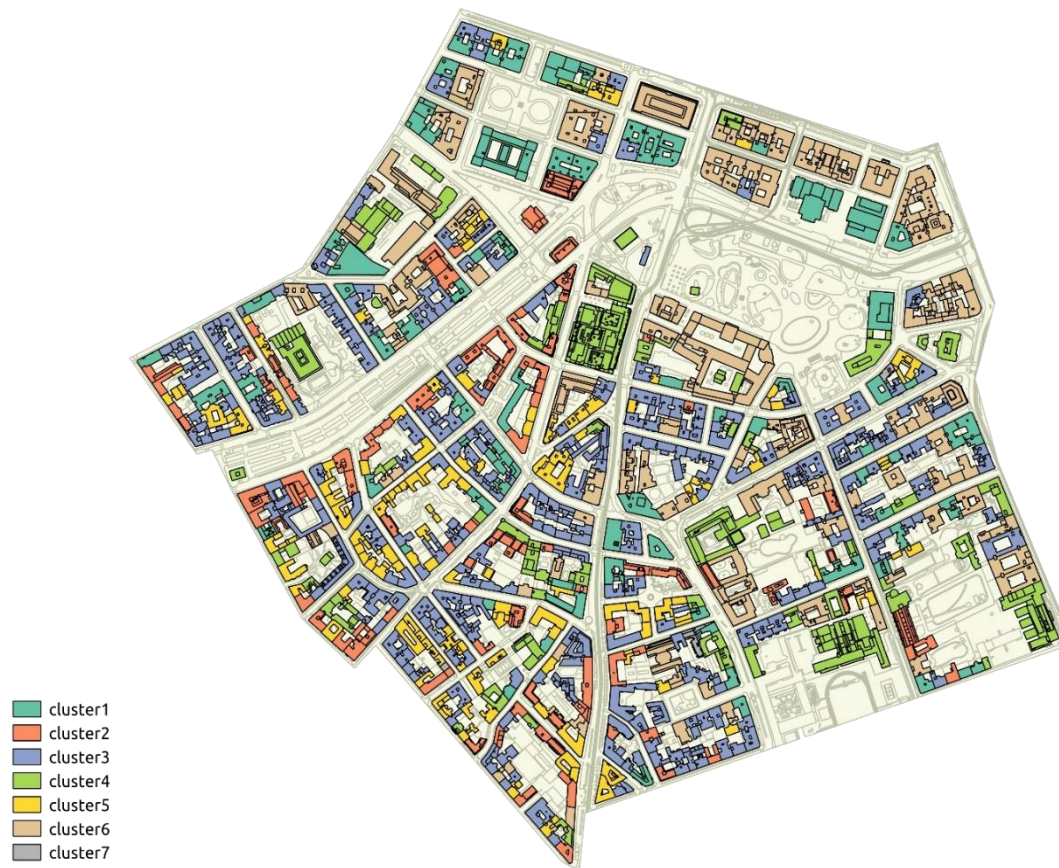
## **List of Equations**

Equation 1 .....	45
Equation 2 .....	45
Equation 3 .....	46
Equation 4 .....	46
Equation 5 .....	48
Equation 6 .....	48
Equation 7 .....	49
Equation 8 .....	49
Equation 9 .....	49
Equation 10 .....	50
Equation 11 .....	51
Equation 12 .....	51
Equation 13 .....	52
Equation 14 .....	52
Equation 15 .....	52
Equation 16 .....	53
Equation 17 .....	53
Equation 18 .....	53
Equation 19 .....	54
Equation 20 .....	59
Equation 21 .....	60
Equation 22 .....	60
Equation 23 .....	61
Equation 24 .....	63
Equation 25 .....	69
Equation 26 .....	70
Equation 27 .....	70
Equation 28 .....	70
Equation 29 .....	71
Equation 30 .....	71
Equation 31 .....	71
Equation 32 .....	71
Equation 33 .....	73
Equation 34 .....	74
Equation 35 .....	74
Equation 36 .....	75

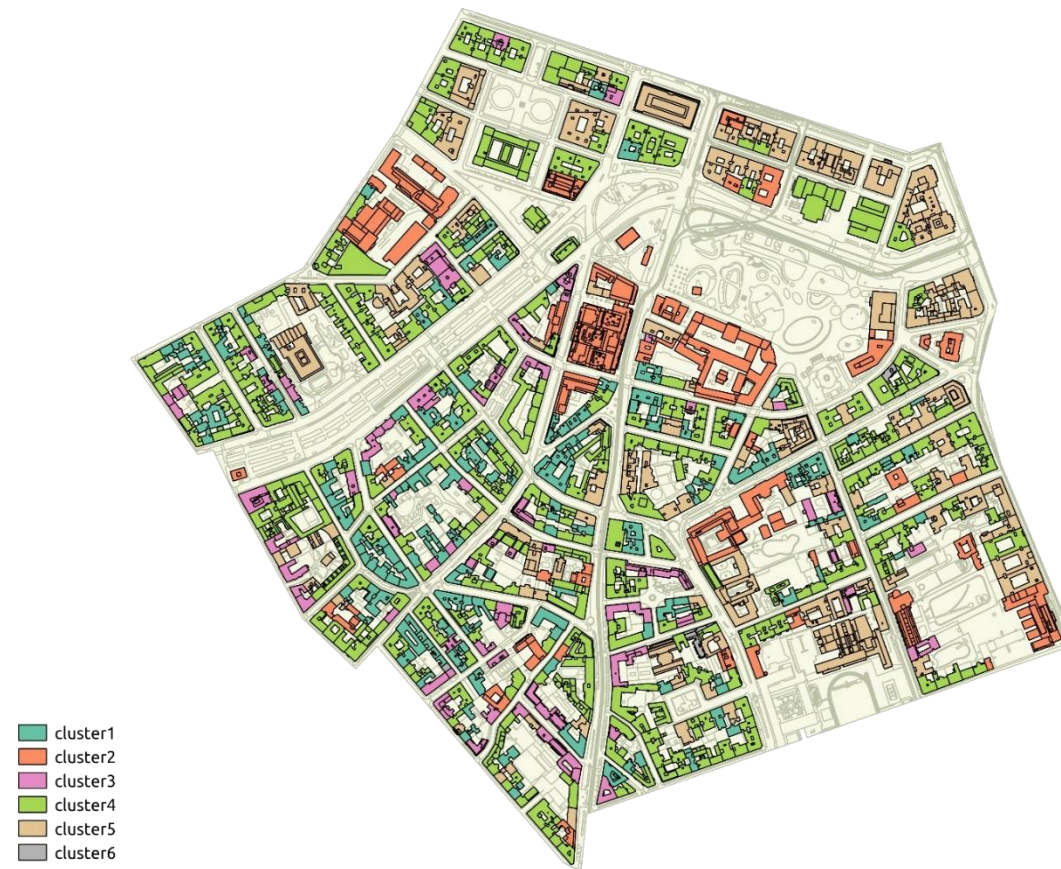
Equation 37 .....	75
Equation 38 .....	75
Equation 39 .....	75
Equation 40 .....	76
Equation 41 .....	76
Equation 42 .....	77
Equation 43 .....	77
Equation 44 .....	77
Equation 45 .....	91
Equation 46 .....	91
Equation 47 .....	94
Equation 48 .....	95
Equation 49 .....	95
Equation 50 .....	95
Equation 51 .....	95
Equation 52 .....	96
Equation 53 .....	96
Equation 54 .....	96
Equation 55 .....	97
Equation 56 .....	99
Equation 57 .....	100
Equation 58 .....	101
Equation 59 .....	101
Equation 60 .....	103
Equation 61 .....	104
Equation 62 .....	104
Equation 63 .....	106

## Appendix 1

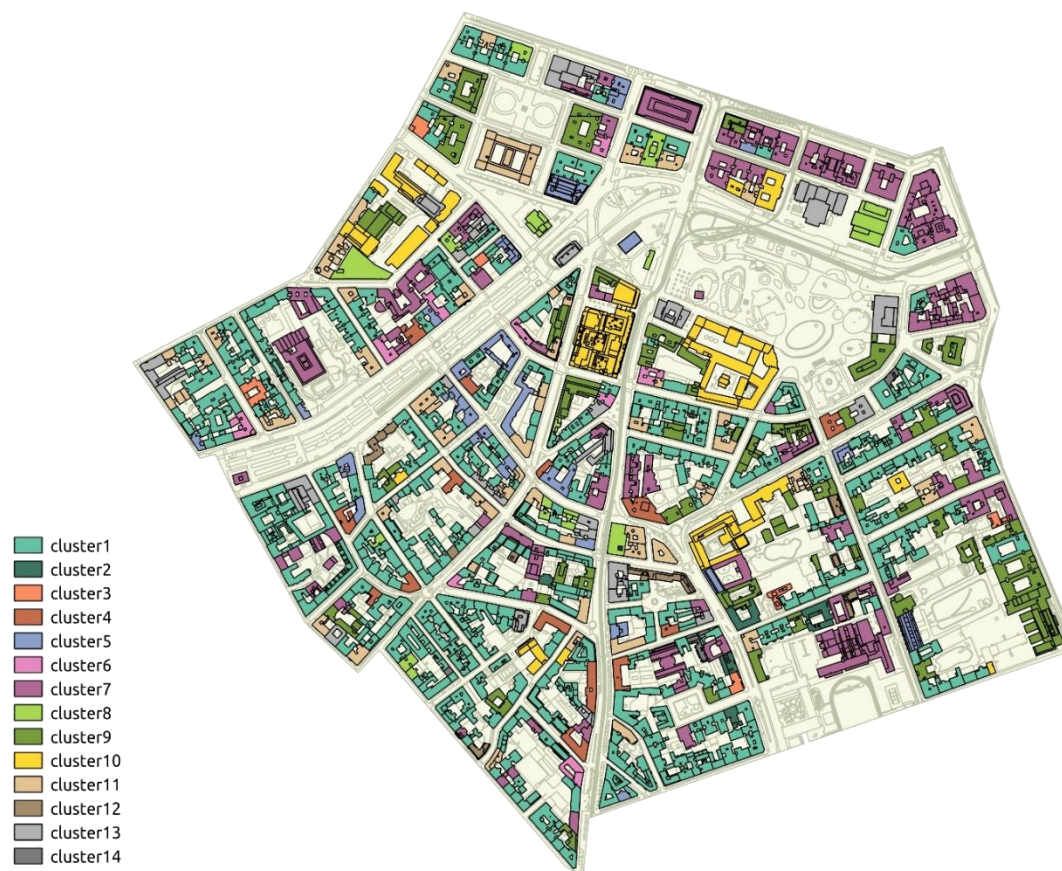
The results of all applied cluster analysis scenarios are presented in this appendix. These scenarios vary in the selection of the descriptive indicators constituting the clustering criteria, as well as the adopted clustering algorithms. The list of the input parameters in each case can be found in Table 7. The results have been visualized by the GIS plug-in developed for the reductive process. In each image, the buildings belonging to the same cluster are marked in the same color.



*Figure 1.1 Visualized results of clustering scenario 1, K-means Method*

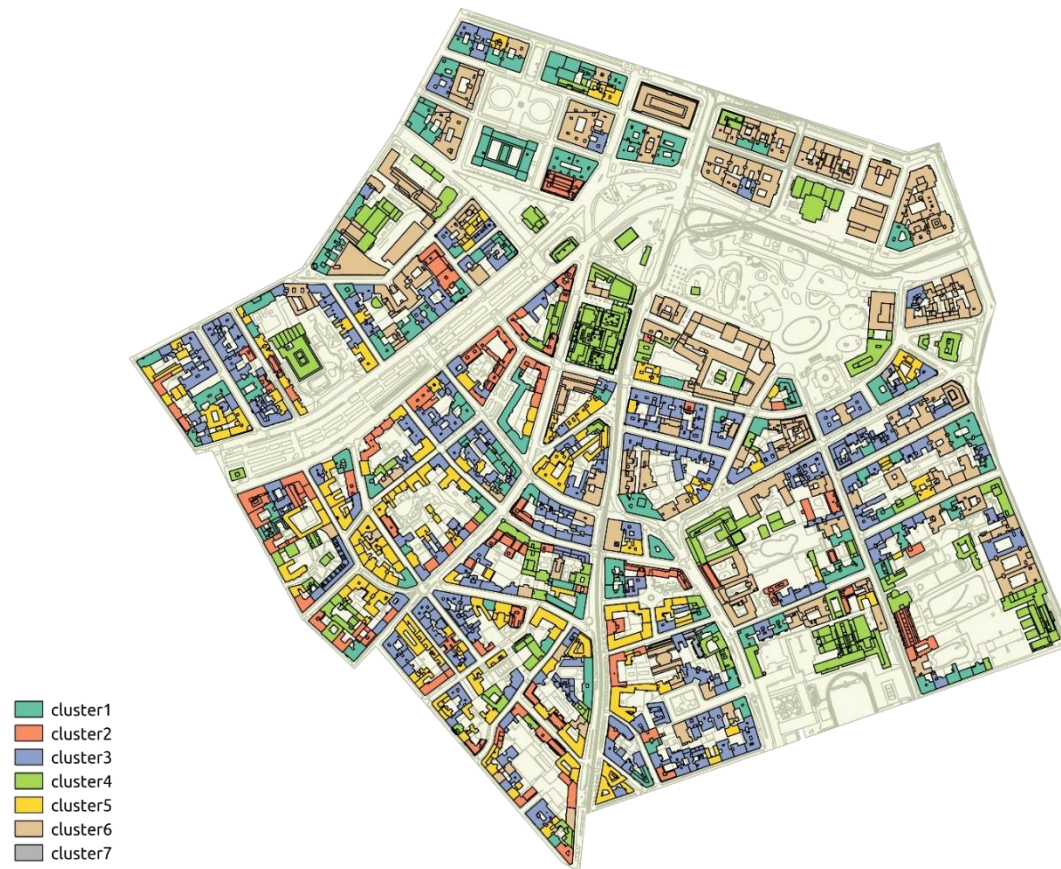


*Figure 1.2 Visualized results of clustering scenario 1, Hierarchical Method*

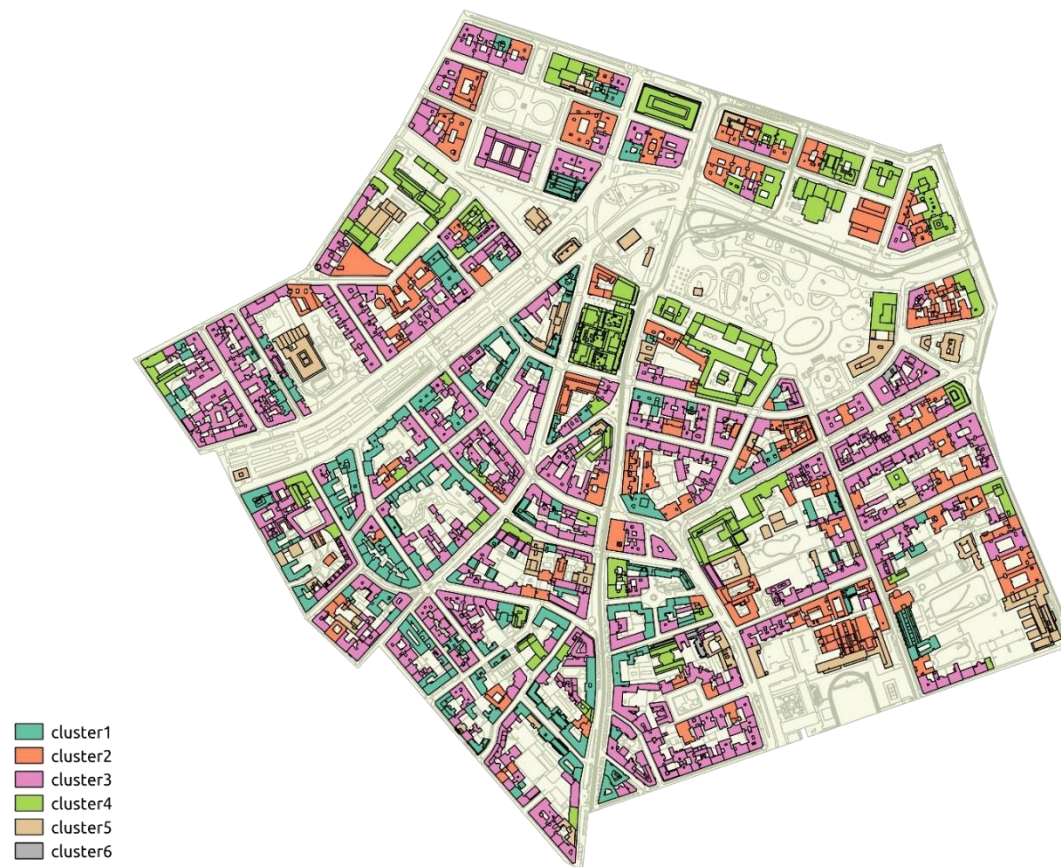


*Figure 1.3 Visualized results of clustering scenario 1, Model-based Method*



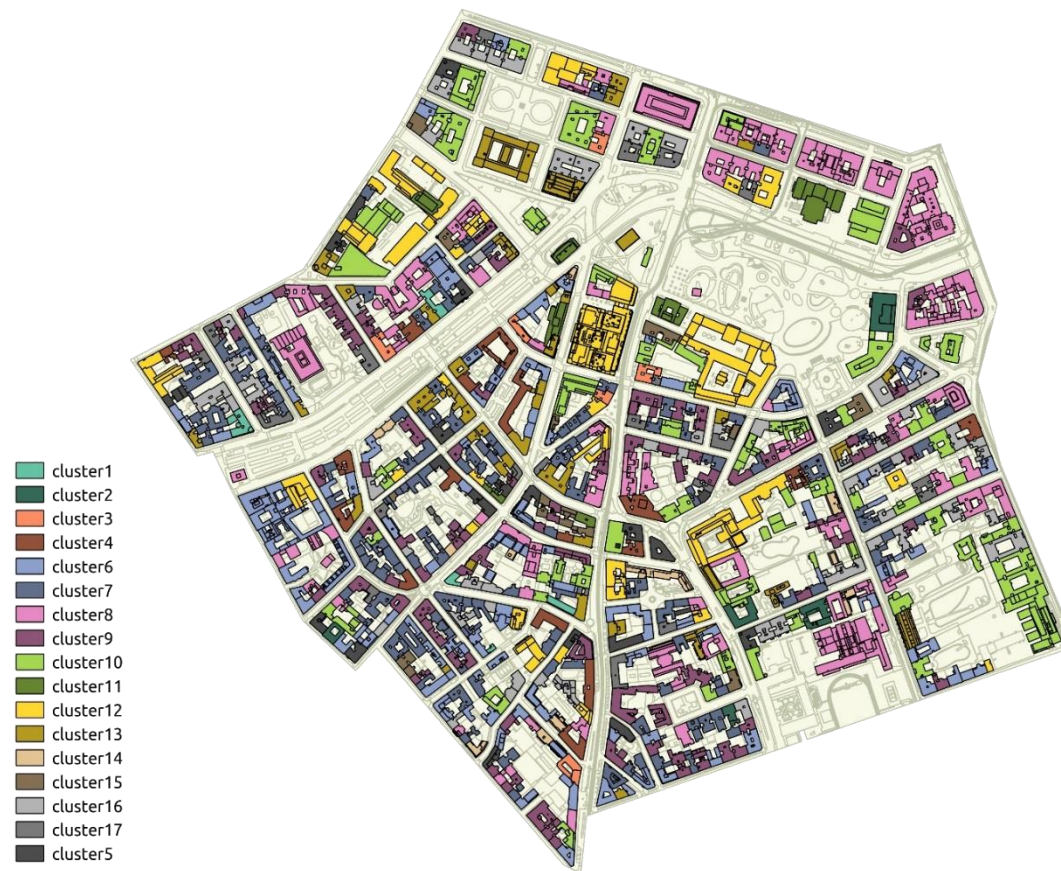


*Figure 1.4 Visualized results of clustering scenario 2, K-means Method*

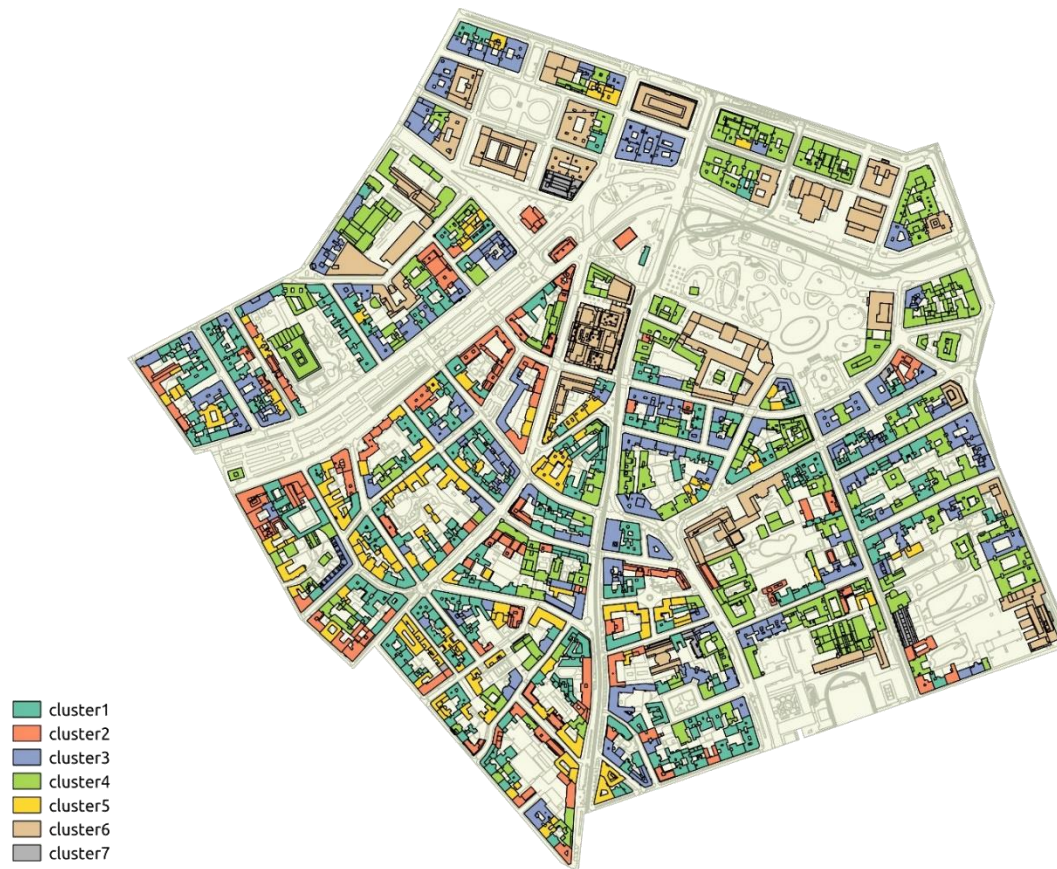


*Figure 1.5 Visualized results of clustering scenario 2, Hierarchical Method*

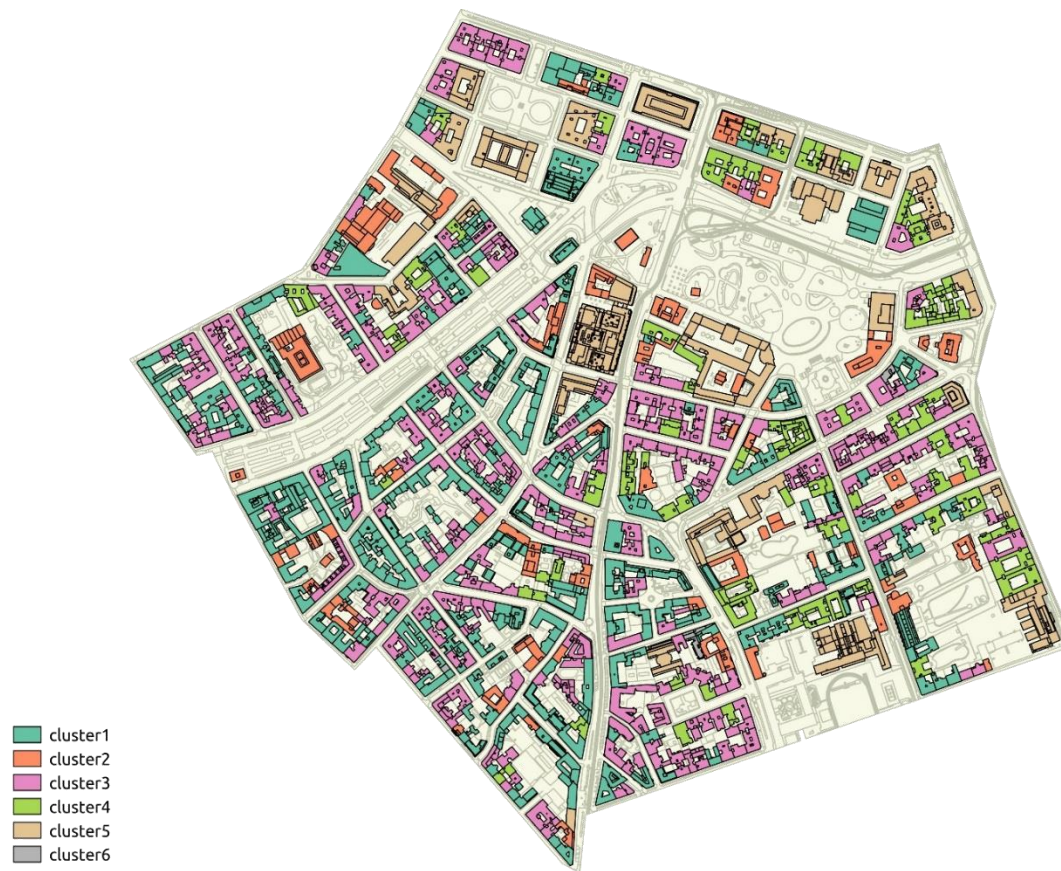




*Figure 1.6 Visualized results of clustering scenario 2, Model-based Method*



*Figure 1.7 Visualized results of clustering scenario 3, K-means Method*

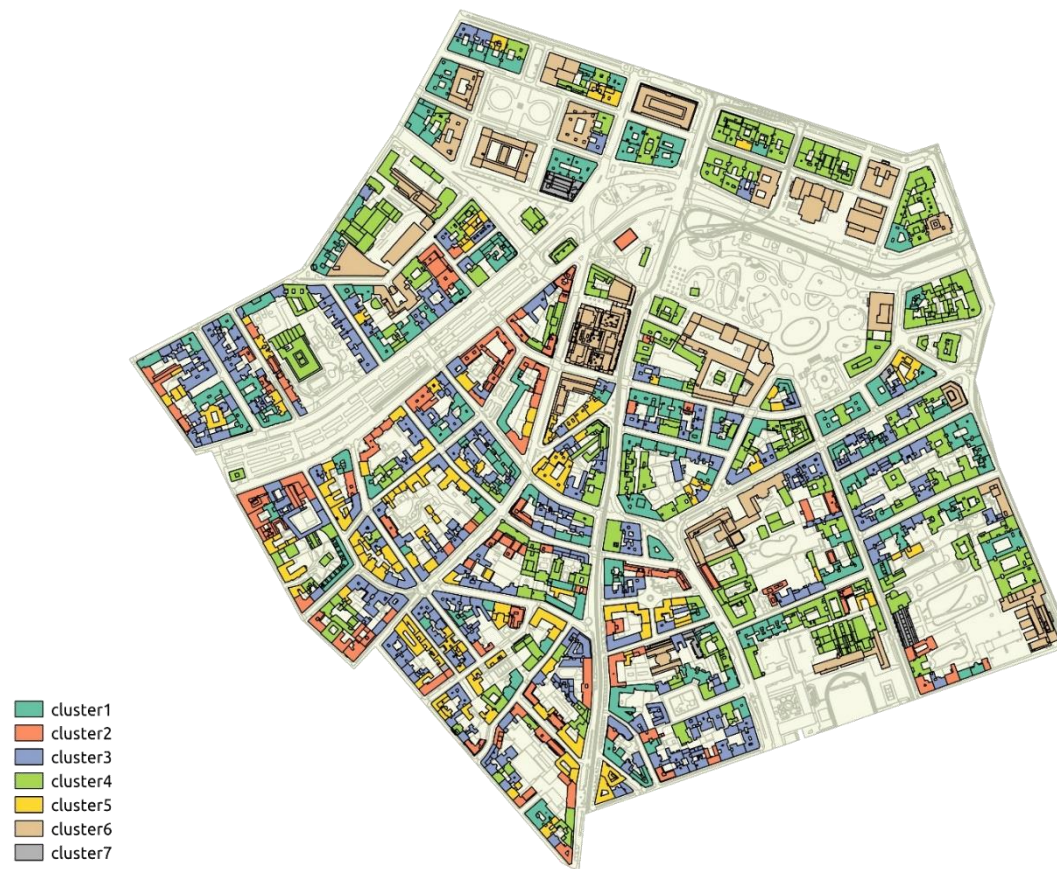


*Figure 1.8 Visualized results of clustering scenario 3, Hierarchical Method*

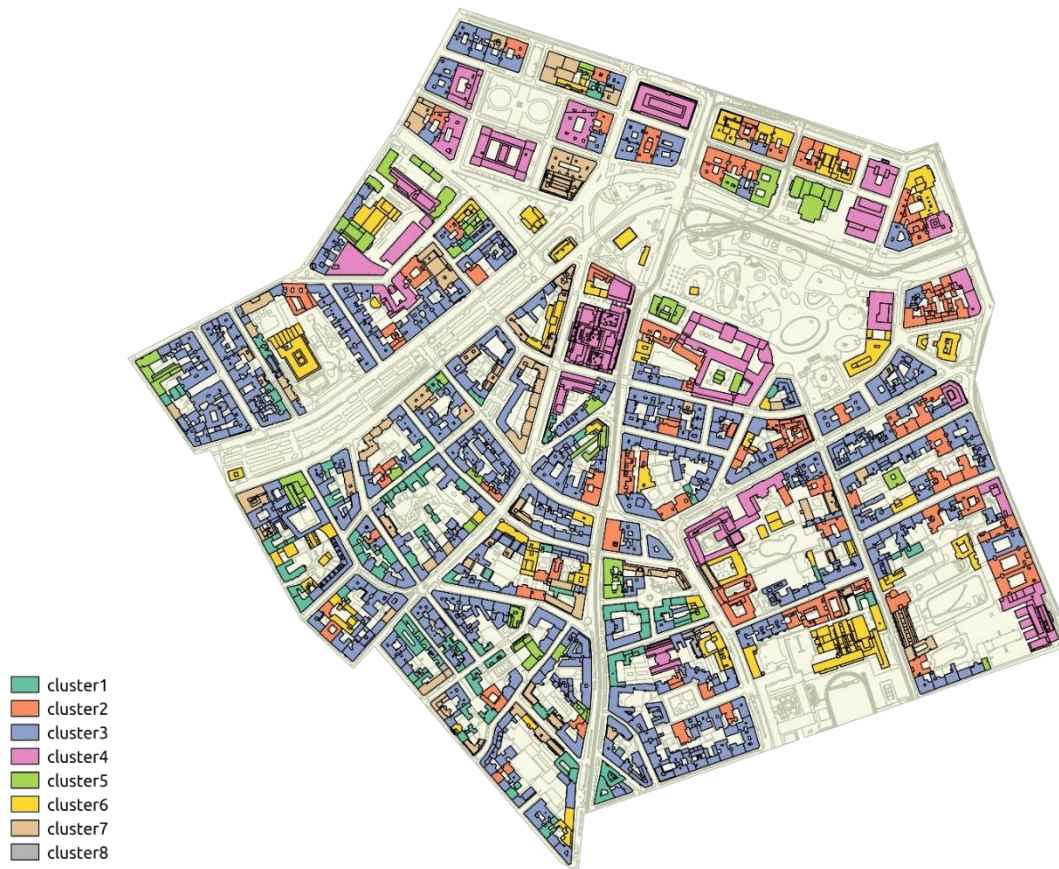




*Figure 1.9 Visualized results of clustering scenario 3, Model-based Method*



*Figure 1.10 Visualized results of clustering scenario 4, K-means Method*

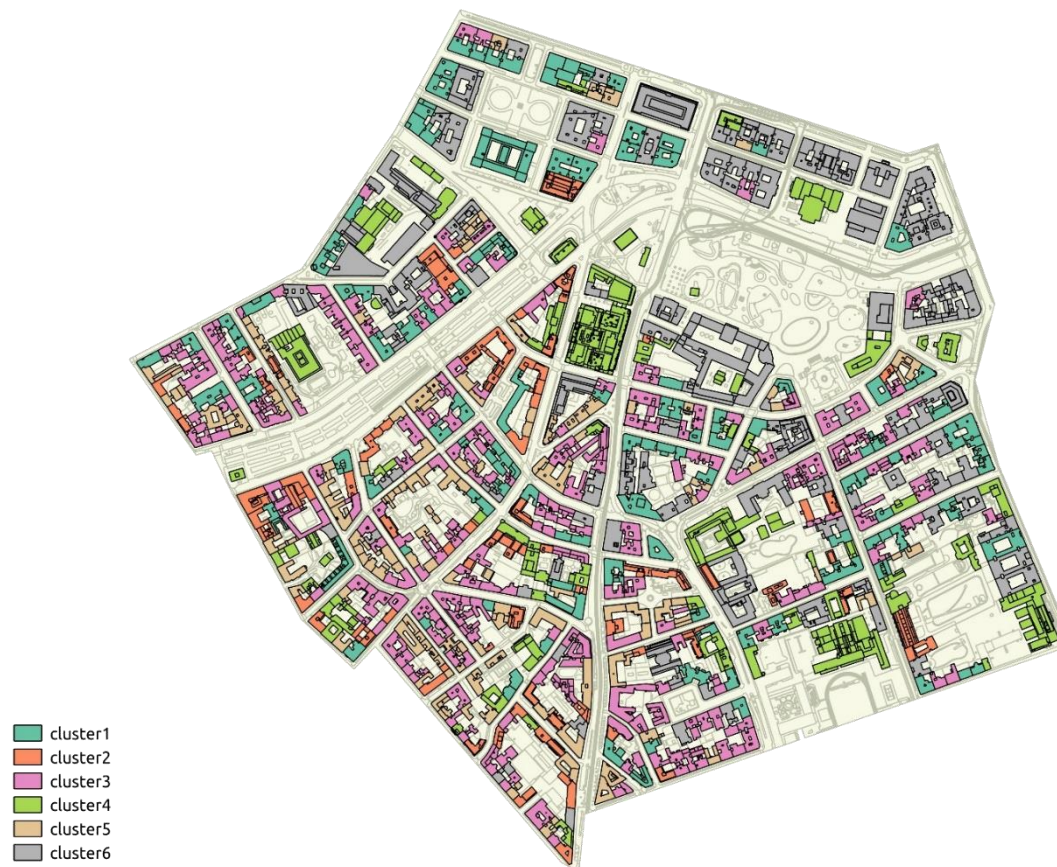


*Figure 1.11 Visualized results of clustering scenario 4, Hierarchical Method*



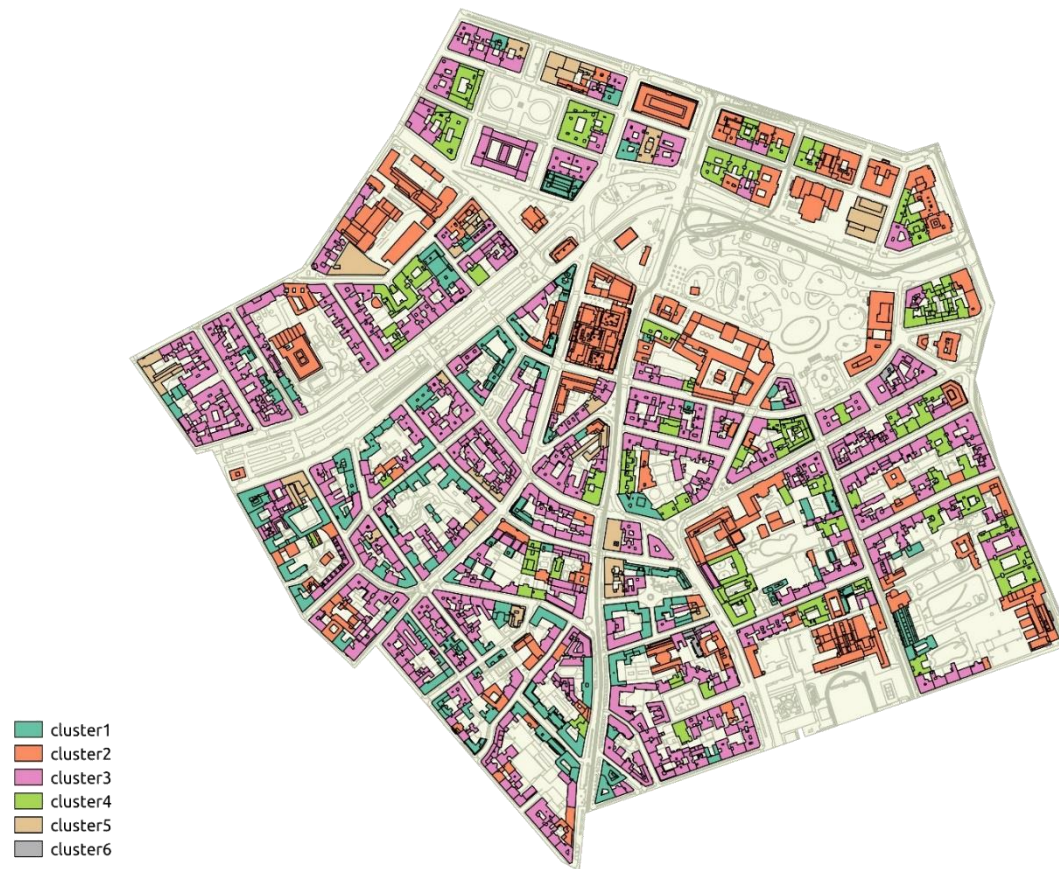


*Figure 1.12 Visualized results of clustering scenario 4, Model-based Method*



*Figure 1.13 Visualized results of clustering scenario 5, K-means Method*

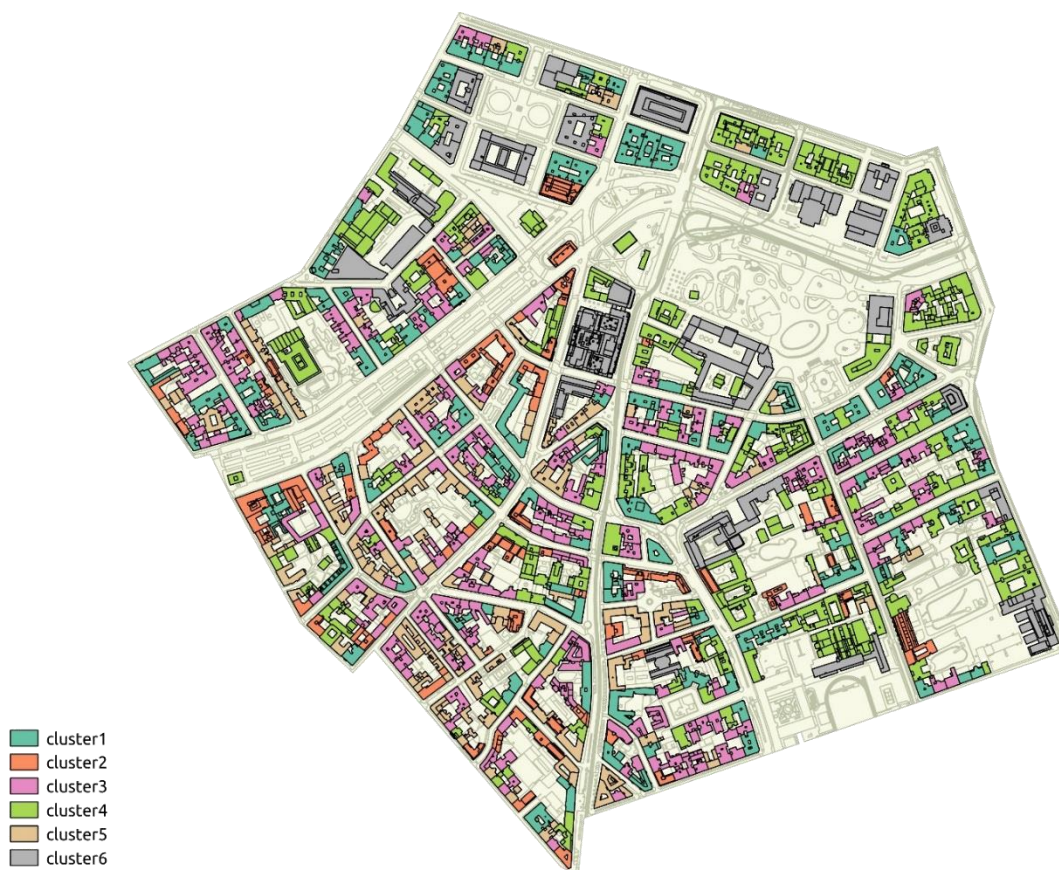




*Figure 1.14 Visualized results of clustering scenario 5, Hierarchical Method*

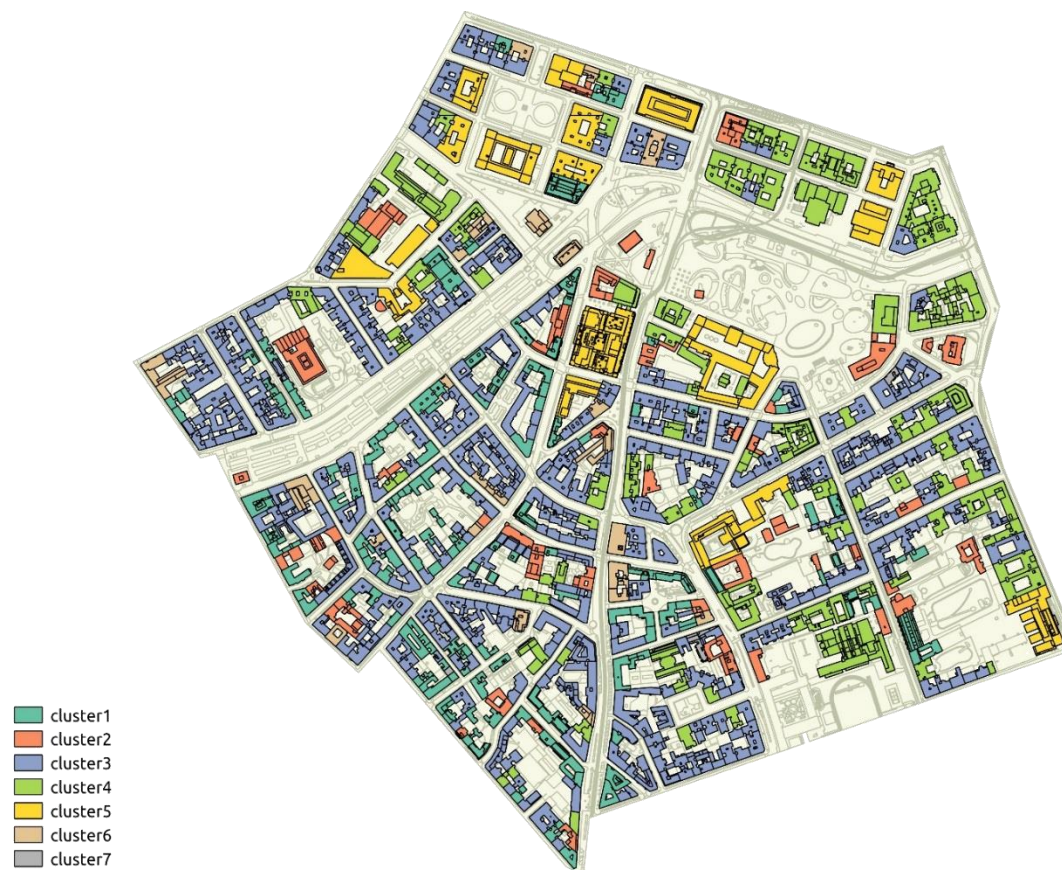


*Figure 1.15 Visualized results of clustering scenario 5, Model-based Method*

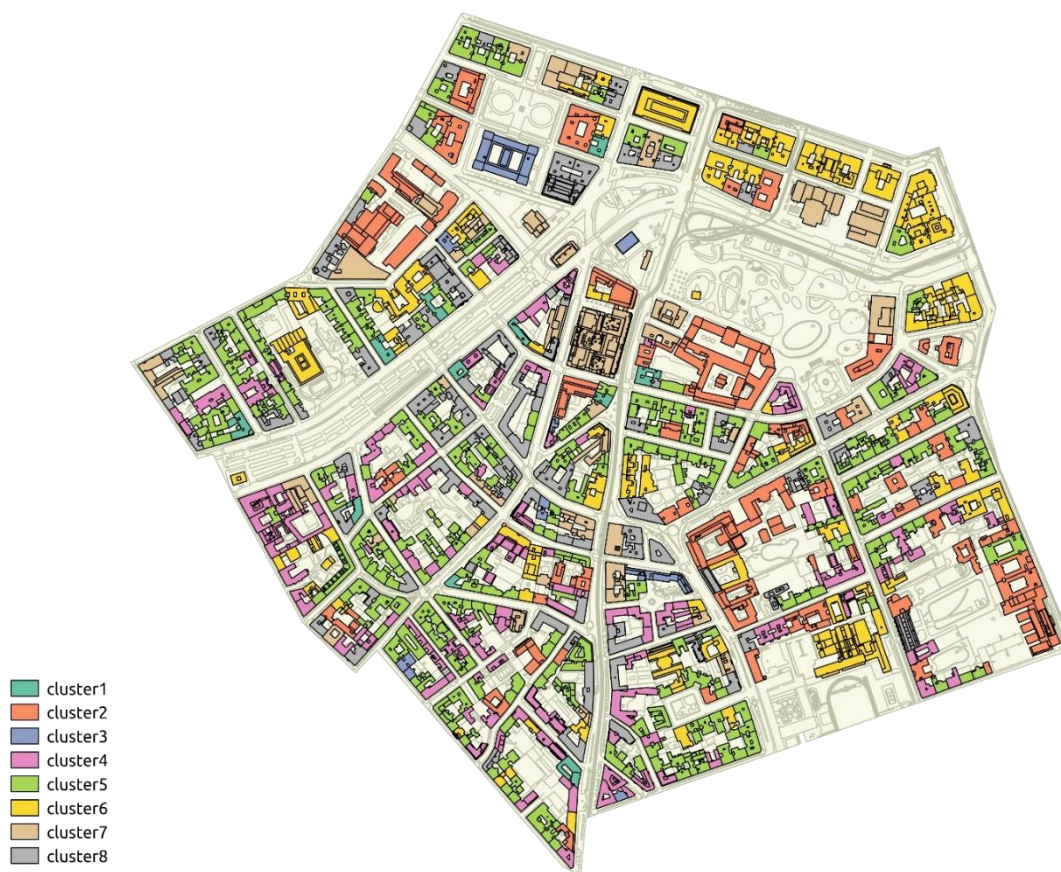


*Figure 1.16 Visualized results of clustering scenario 6, K-means Method*

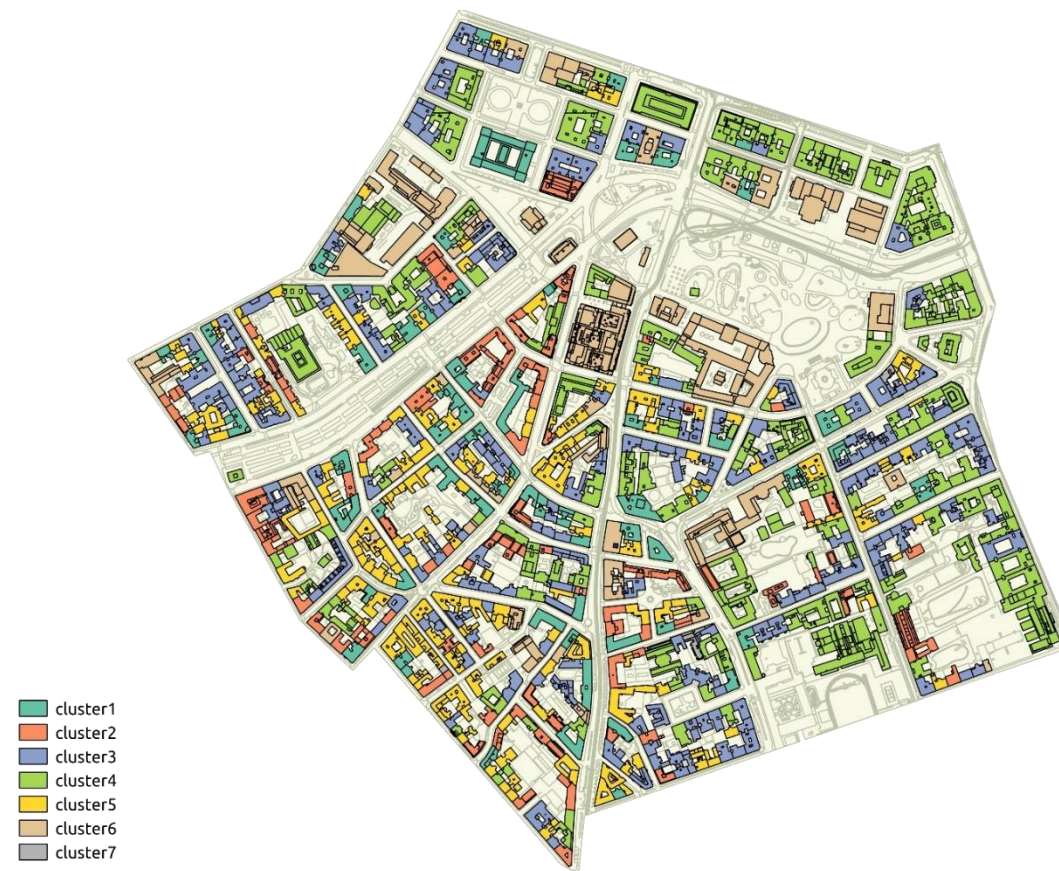




*Figure 1.17 Visualized results of clustering scenario 6, Hierarchical Method*

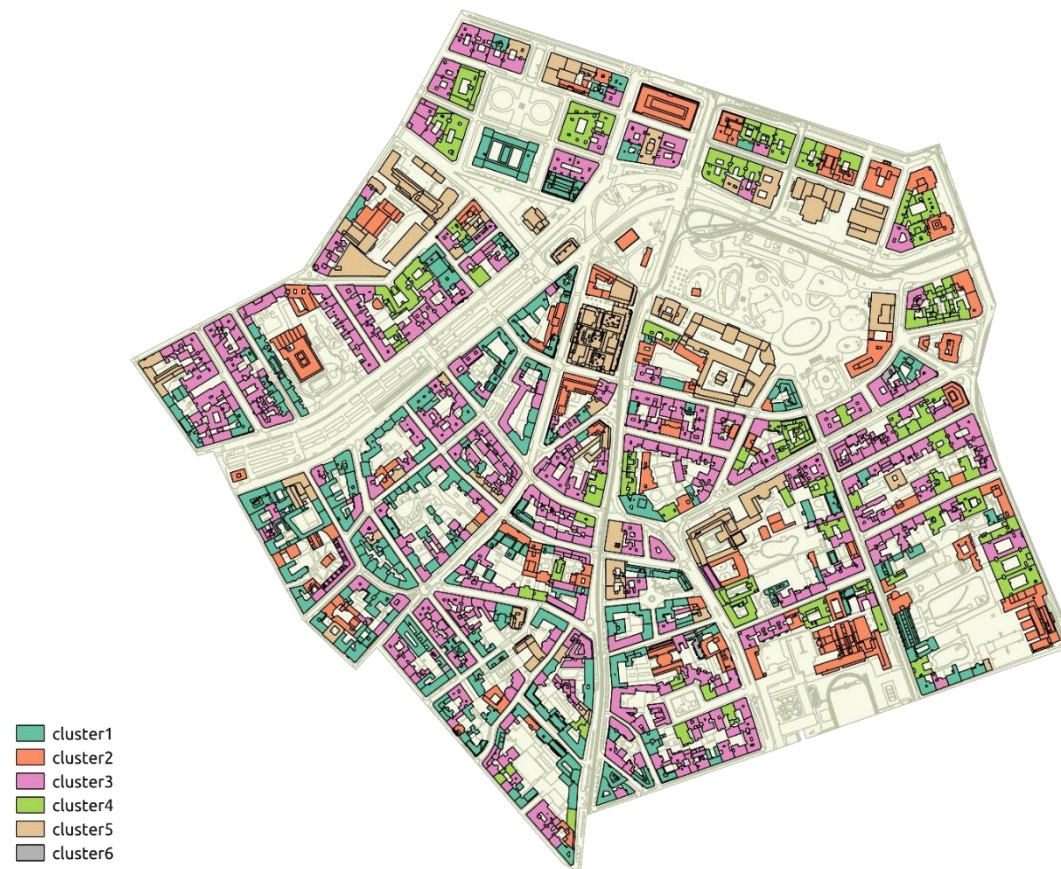


*Figure 1.18 Visualized results of clustering scenario 6, Model-based Method*

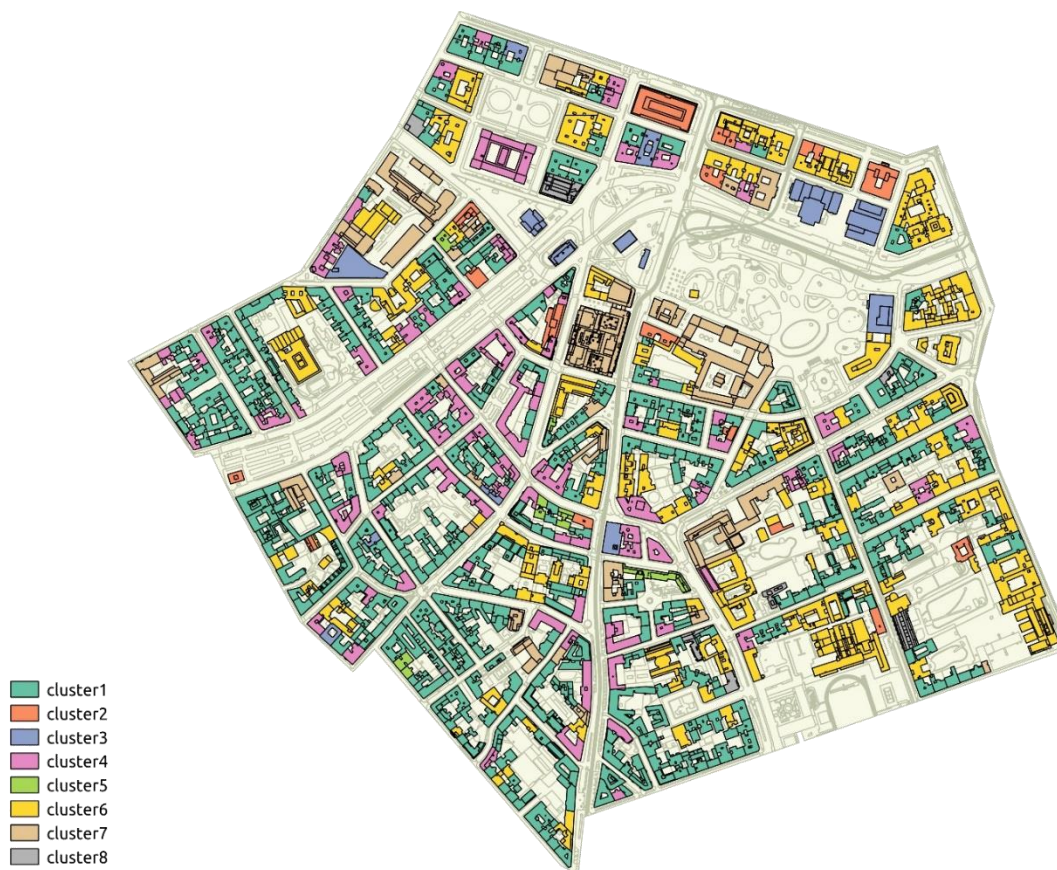


*Figure 1.19 Visualized results of clustering scenario 7, K-means Method*



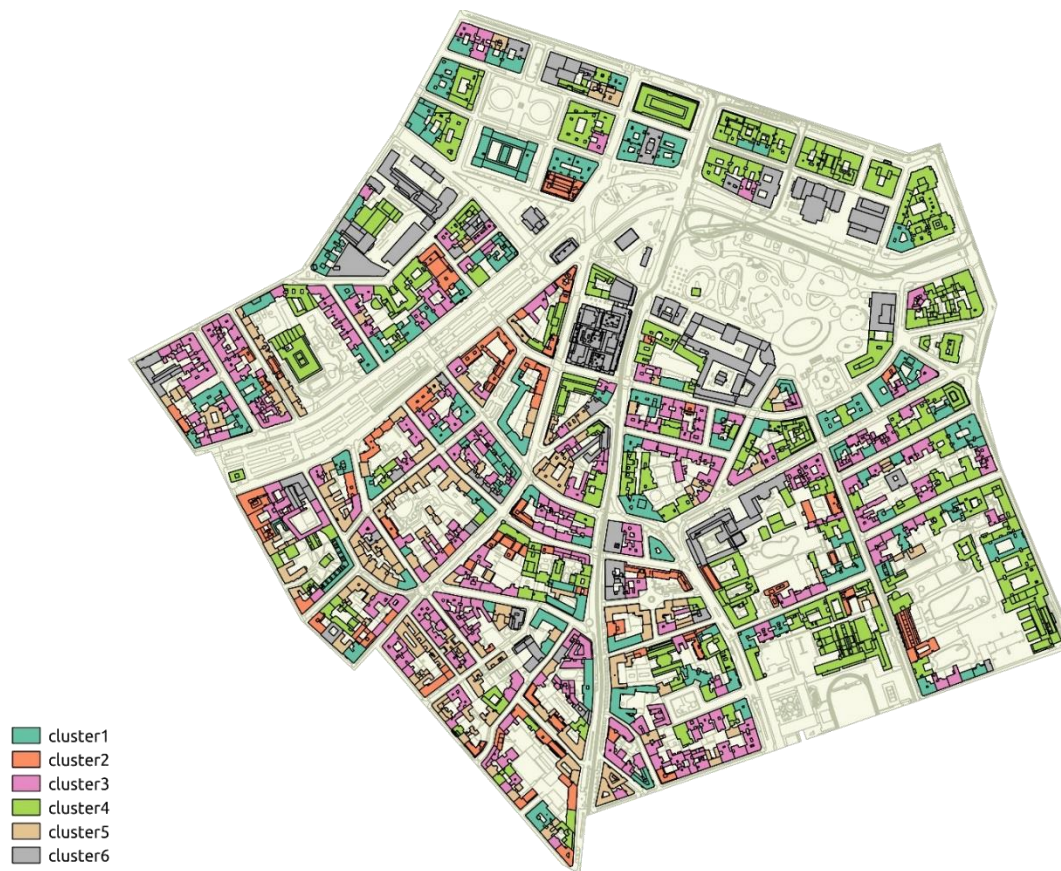


*Figure 1.20 Visualized results of clustering scenario 7, Hierarchical Method*

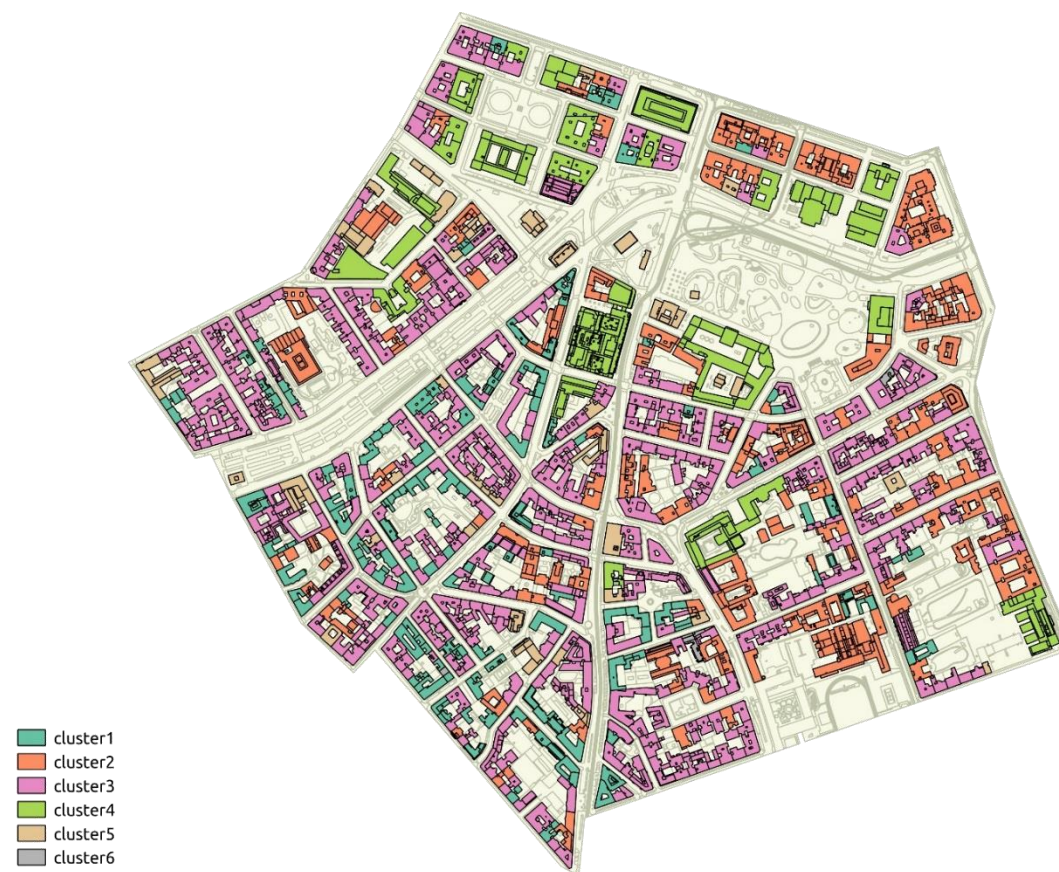


*Figure 1.21 Visualized results of clustering scenario 7, Model-based Method*

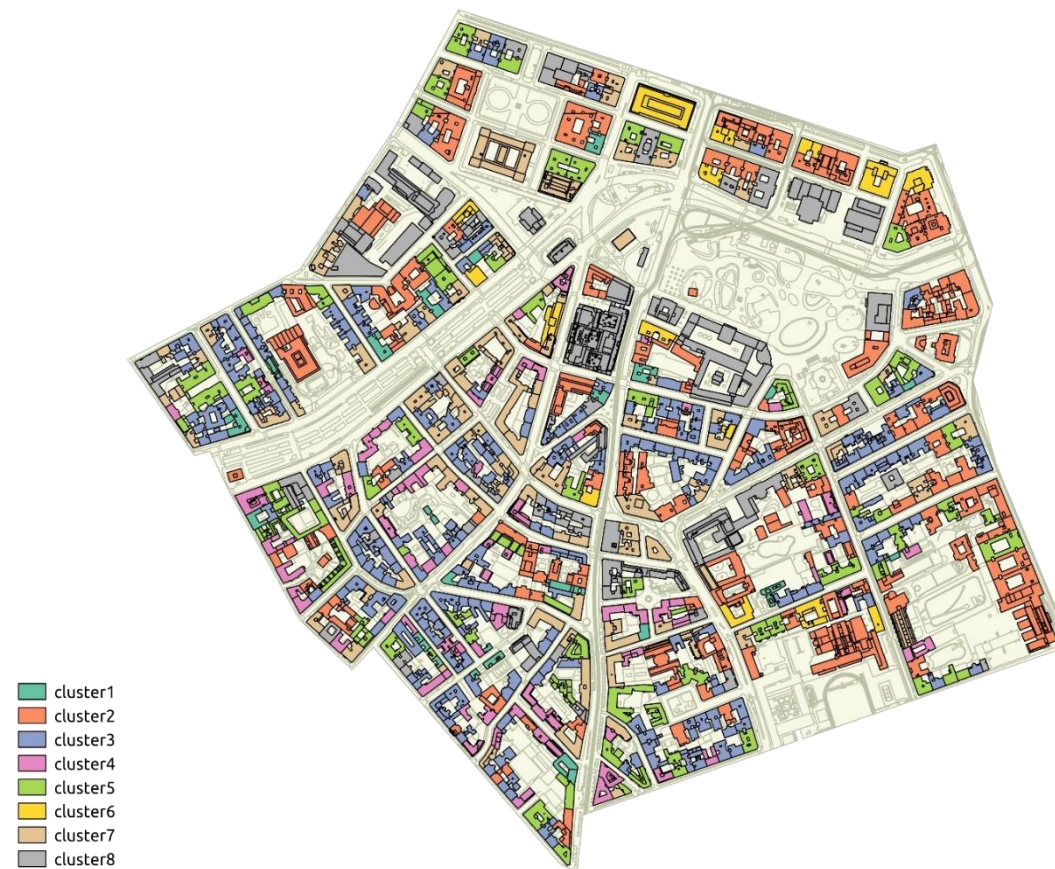




*Figure 1.22 Visualized results of clustering scenario 8, K-means Method*



*Figure 1.23 Visualized results of clustering scenario 8, Hierarchical Method*

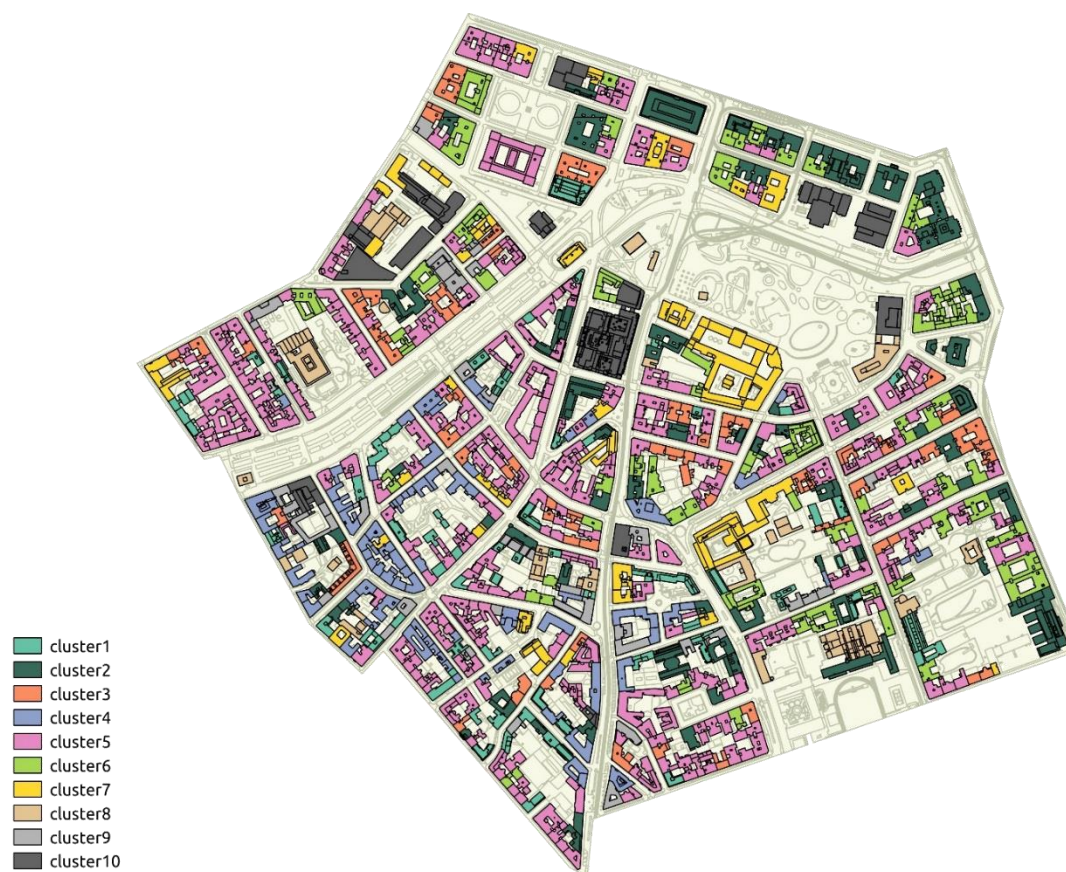


*Figure 1.24 Visualized results of clustering scenario 8, Model-based Method*

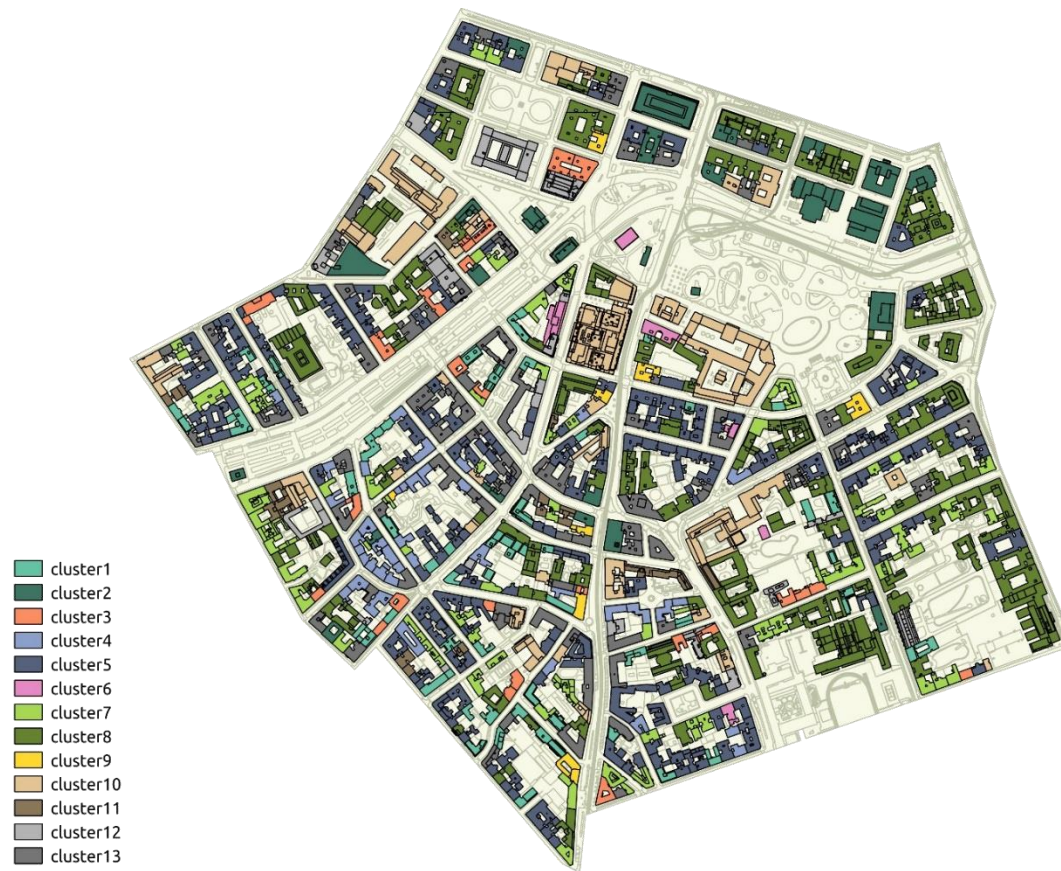




*Figure 1.25 Visualized results of clustering scenario 9, K-means Method*



*Figure 1.26 Visualized results of clustering scenario 9, Hierarchical Method*



*Figure 1.27 Visualized results of clustering scenario 9, Model-based Method*



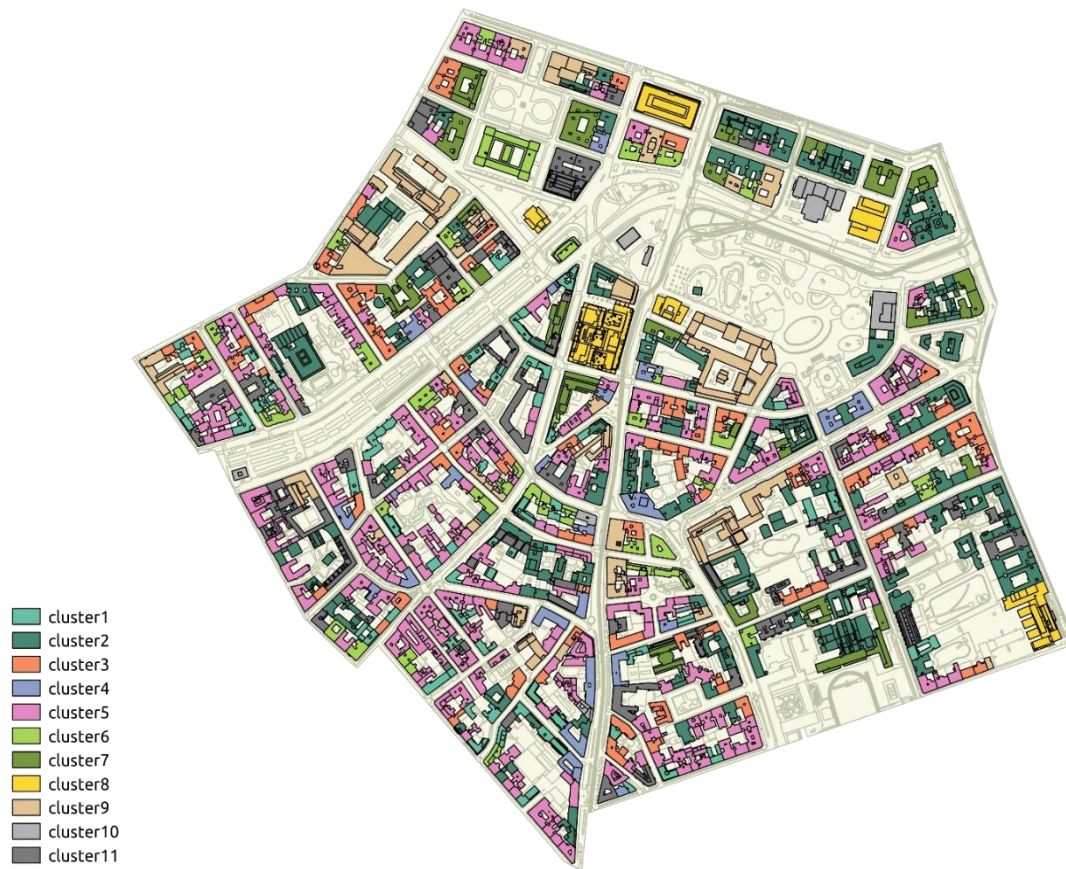


*Figure 1.28 Visualized results of clustering scenario 10, K-means Method*

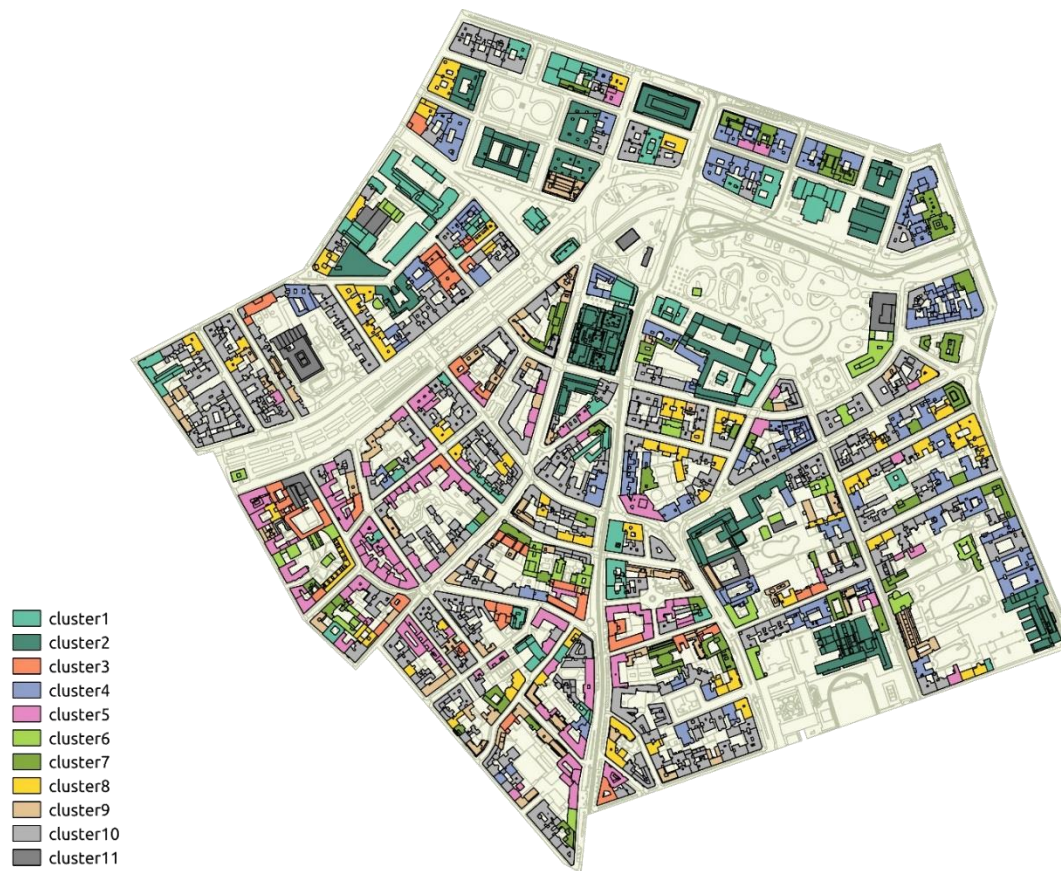


*Figure 1.29 Visualized results of clustering scenario 10, Hierarchical Method*





*Figure 1.30 Visualized results of clustering scenario 10, Model-based Method*



*Figure 1.31 Visualized results of clustering scenario 11, K-means Method*



*Figure 1.32 Visualized results of clustering scenario 11, Hierarchical Method*





*Figure 1.33 Visualized results of clustering scenario 11, Model-based Method*

## Appendix 2

This appendix provides more information on the representative buildings emerging from the most successful clustering schema (Scenario 4, K-means clustering method). Accordingly, for each representative building, a descriptive table, as well as the typical floor plan, and an elevation is provided.

## Cluster 1

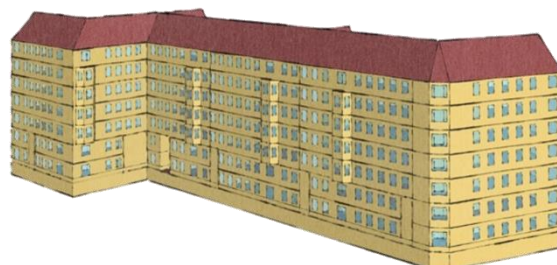


Figure 2.1 South-West View

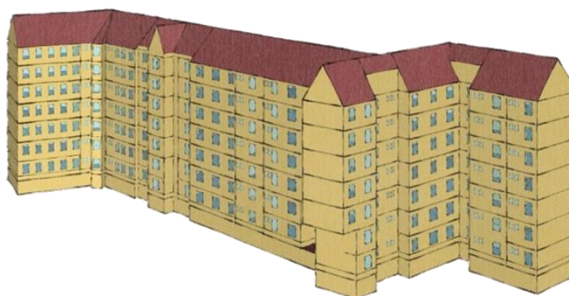
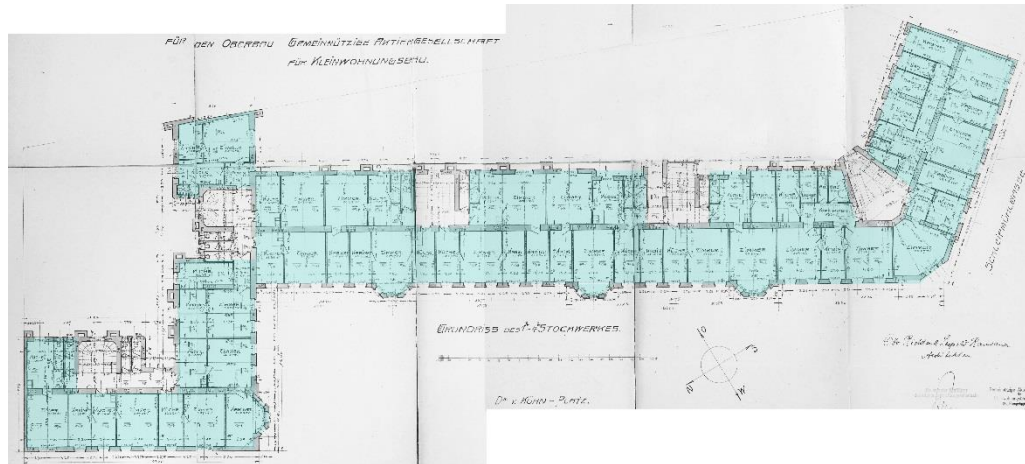


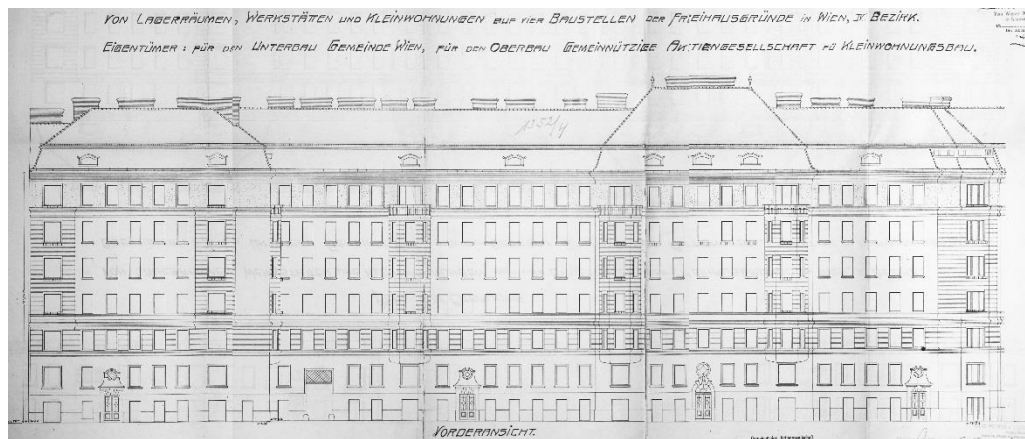
Figure 2.2 North-East View

Table 2.1 Description of the building representing cluster 1

Address	Mühlgasse 7, 1040
Total Area [m <sup>2</sup> ]	10489
Conditioned Area [m <sup>2</sup> ]	6888
Total Volume [m <sup>3</sup> ]	34218
Conditioned Volume [m <sup>3</sup> ]	23217
Year of construction	1914
Number of Zones	76
Primary Usage	Residential
Associated Area [m <sup>2</sup> ]	5987
Other Usage(s)	Gastronomy
Associated Area [m <sup>2</sup> ]	900
Heating Demand per Total Building Area [kWh.m <sup>-2</sup> .a <sup>-1</sup> ]	99.37
Heating Demand per Conditioned Building Area [kWh.m <sup>-2</sup> .a <sup>-1</sup> ]	151.33

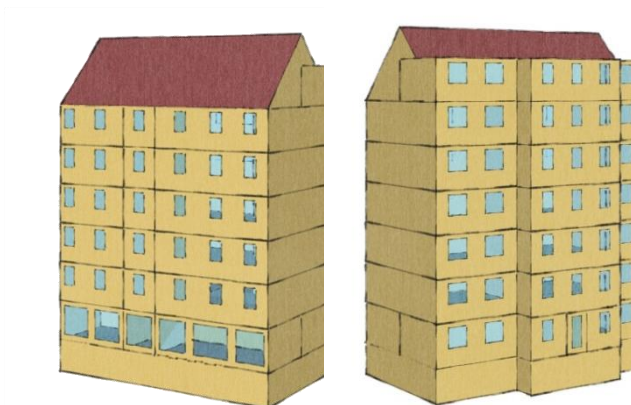


*Figure 2.3 Typical floor plan, the conditioned area is highlighted*



*Figure 2.4 Main elevation*

## Cluster 2



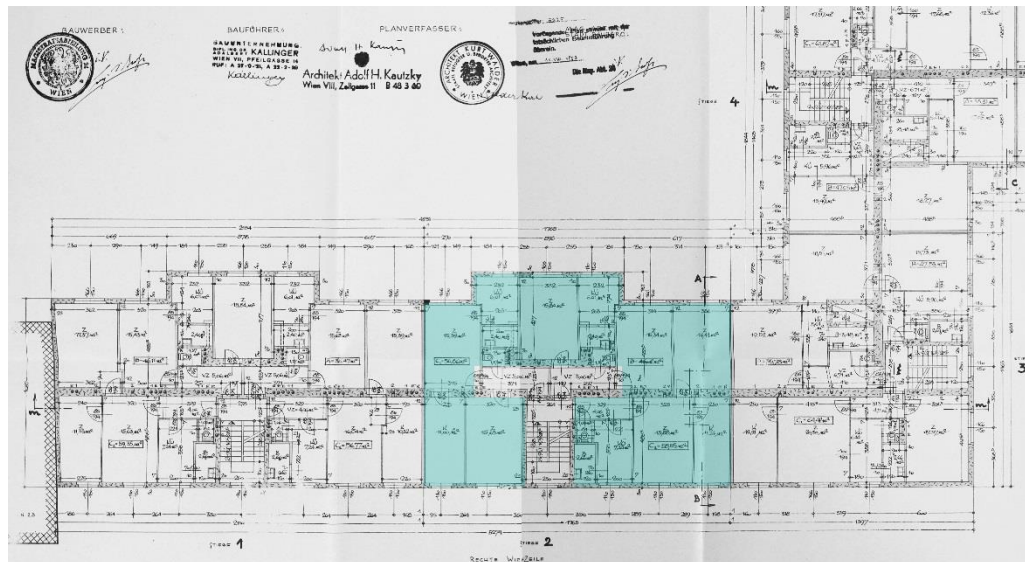
*Figure 2.5 (Left) North-West View*

*Figure 2.6 (Right) South-East View*

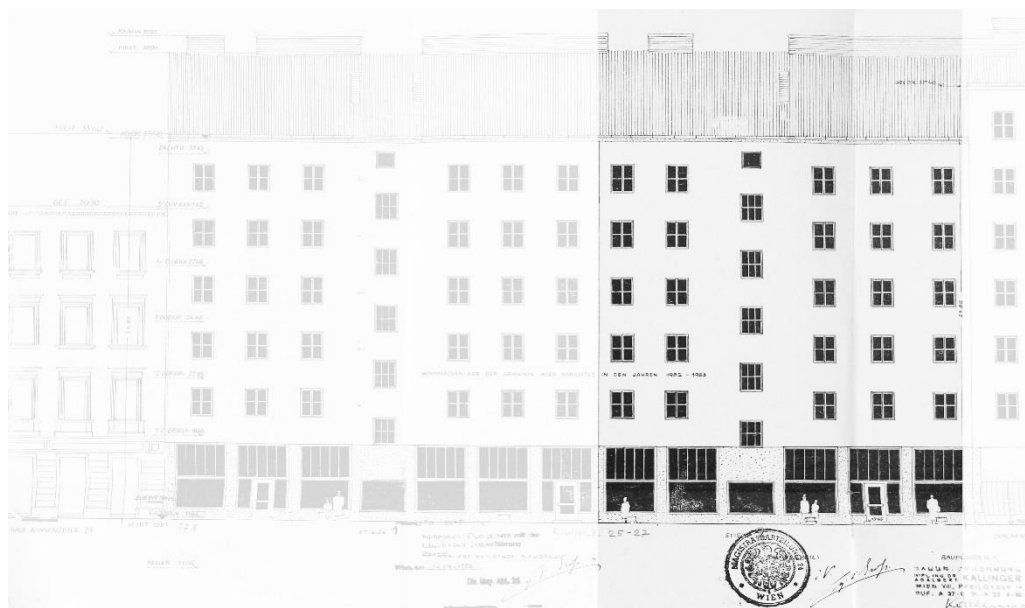
*Table 2.2 Description of the building representing cluster 2*

<b>Address</b>	Rechte Wienzeile 25, 1040
<b>Total Area [m<sup>2</sup>]</b>	1608
<b>Conditioned Area [m<sup>2</sup>]</b>	1180
<b>Total Volume [m<sup>3</sup>]</b>	4595
<b>Conditioned Volume [m<sup>3</sup>]</b>	3541
<b>Year of construction</b>	1945-1976
<b>Number of Zones</b>	18
<b>Primary Usage</b>	Residential
<b>Associated Area [m<sup>2</sup>]</b>	1180
<b>Other Usage(s)</b>	-
<b>Associated Area [m<sup>2</sup>]</b>	-
<b>Heating Demand per Total Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	86.38
<b>Heating Demand per Conditioned Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	117.68



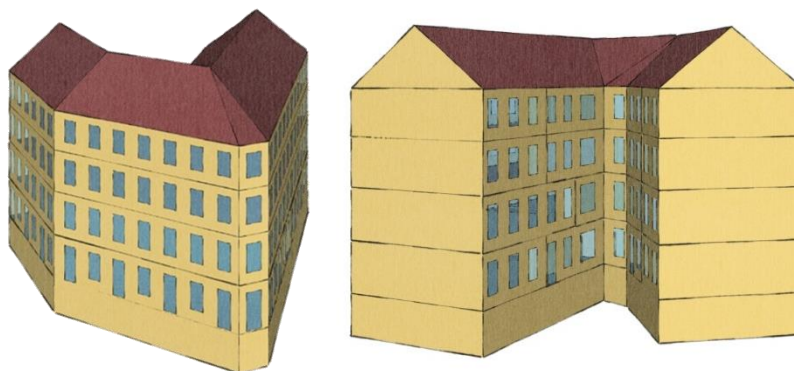


*Figure 2.7 Typical floor plan, the conditioned area is highlighted*



*Figure 2.8 Main elevation*

### Cluster 3



*Figure 2.9 (Left) East View*

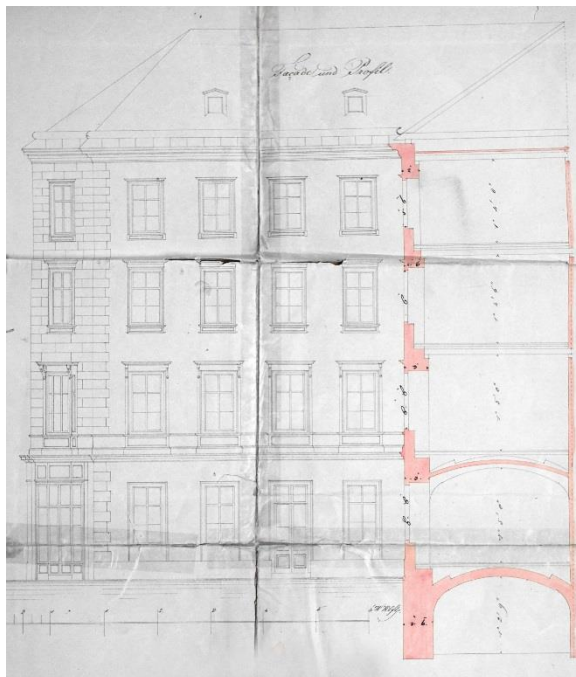
*Figure 2.10 (Right) West View*

*Table 2.3 Description of the building representing cluster 3*

<b>Address</b>	Preßgasse 17, 1040
<b>Total Area [m<sup>2</sup>]</b>	2777
<b>Conditioned Area [m<sup>2</sup>]</b>	1647
<b>Total Volume [m<sup>3</sup>]</b>	8441
<b>Conditioned Volume [m<sup>3</sup>]</b>	5536
<b>Year of construction</b>	1846
<b>Number of Zones</b>	11
<b>Primary Usage</b>	Residential
<b>Associated Area [m<sup>2</sup>]</b>	1252
<b>Other Usage(s)</b>	Gastronomy
<b>Associated Area [m<sup>2</sup>]</b>	395
<b>Heating Demand per Total Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	85.2
<b>Heating Demand per Conditioned Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	143.67

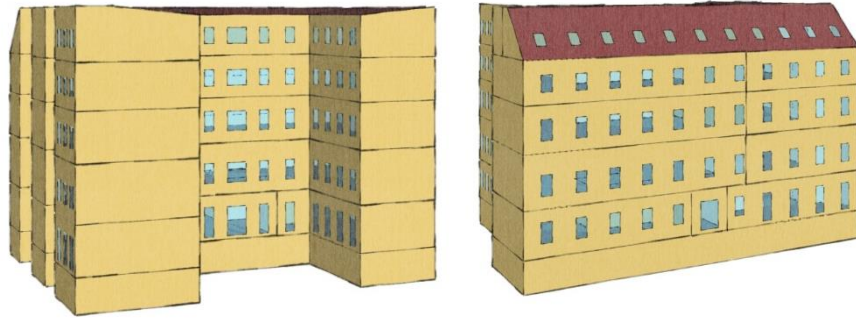


*Figure 2.11 Typical floor plan, the conditioned area is highlighted*



*Figure 2.12 Courtyard elevation-section*

#### Cluster 4

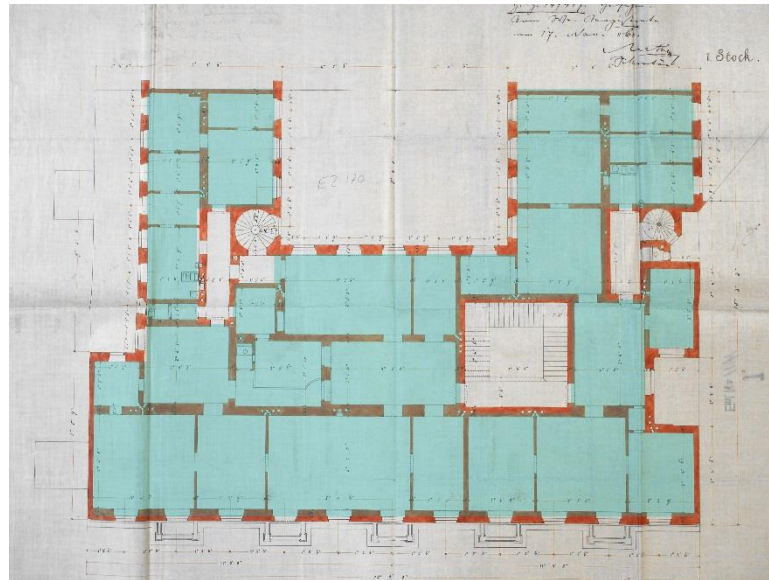


*Figure 2.13 (Left) East View*

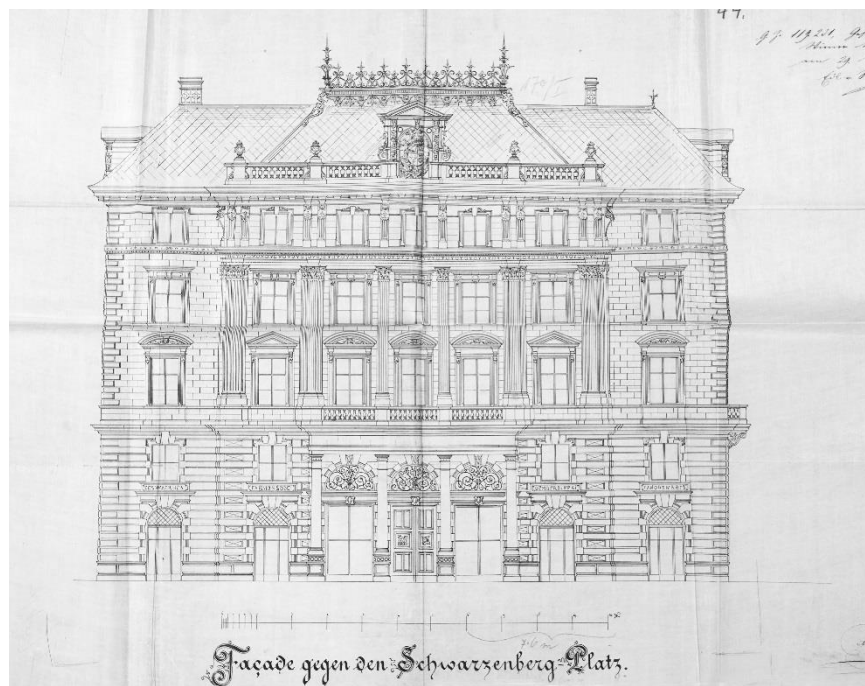
*Figure 2.14 (Right) West View*

*Table 2.4 Description of the building representing cluster 4*

<b>Address</b>	Schwarzenbergplatz 16, 1010
<b>Total Area [m<sup>2</sup>]</b>	5099
<b>Conditioned Area [m<sup>2</sup>]</b>	38512
<b>Total Volume [m<sup>3</sup>]</b>	22508
<b>Conditioned Volume [m<sup>3</sup>]</b>	17863
<b>Year of construction</b>	1868
<b>Number of Zones</b>	17
<b>Primary Usage</b>	Office
<b>Associated Area [m<sup>2</sup>]</b>	38512
<b>Other Usage(s)</b>	-
<b>Associated Area [m<sup>2</sup>]</b>	-
<b>Heating Demand per Total Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	106.18
<b>Heating Demand per Conditioned Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	140.56



*Figure 2.15 Typical floor plan, the conditioned area is highlighted*



*Figure 2.16 Main elevation*



## Cluster 5



*Figure 2.17 (Left) West View*

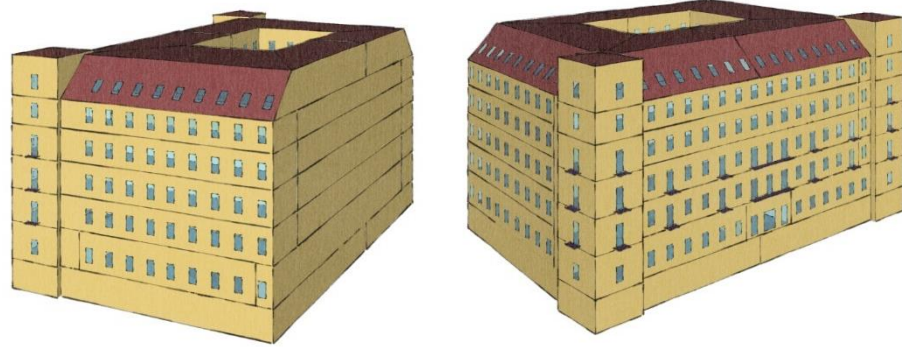
*Figure 2.18 (Right) East View*

*Table 2.5 Description of the building representing cluster 5*

<b>Address</b>	Große Neugasse 16, 1040
<b>Total Area [m<sup>2</sup>]</b>	25512
<b>Conditioned Area [m<sup>2</sup>]</b>	1494
<b>Total Volume [m<sup>3</sup>]</b>	9464
<b>Conditioned Volume [m<sup>3</sup>]</b>	5963
<b>Year of construction</b>	1872
<b>Number of Zones</b>	21
<b>Primary Usage</b>	Residential
<b>Associated Area [m<sup>2</sup>]</b>	1494
<b>Other Usage(s)</b>	-
<b>Associated Area [m<sup>2</sup>]</b>	-
<b>Heating Demand per Total Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	96.33
<b>Heating Demand per Conditioned Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	164.52

Figure 2.20 (Right) Main elevation

## Cluster 6



*Figure 2.21 (Left) North view*

*Figure 2.22 (Right) South view*

*Table 2.6 Description of the building representing cluster 6*

<b>Address</b>	Schillerplatz 4, 1010
<b>Total Area [m<sup>2</sup>]</b>	14502
<b>Conditioned Area [m<sup>2</sup>]</b>	9757
<b>Total Volume [m<sup>3</sup>]</b>	64181
<b>Conditioned Volume [m<sup>3</sup>]</b>	43683
<b>Year of construction</b>	2002
<b>Number of Zones</b>	40
<b>Primary Usage</b>	Office
<b>Associated Area [m<sup>2</sup>]</b>	9757
<b>Other Usage(s)</b>	-
<b>Associated Area [m<sup>2</sup>]</b>	-
<b>Heating Demand per Total Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	85.01
<b>Heating Demand per Conditioned Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	126.35



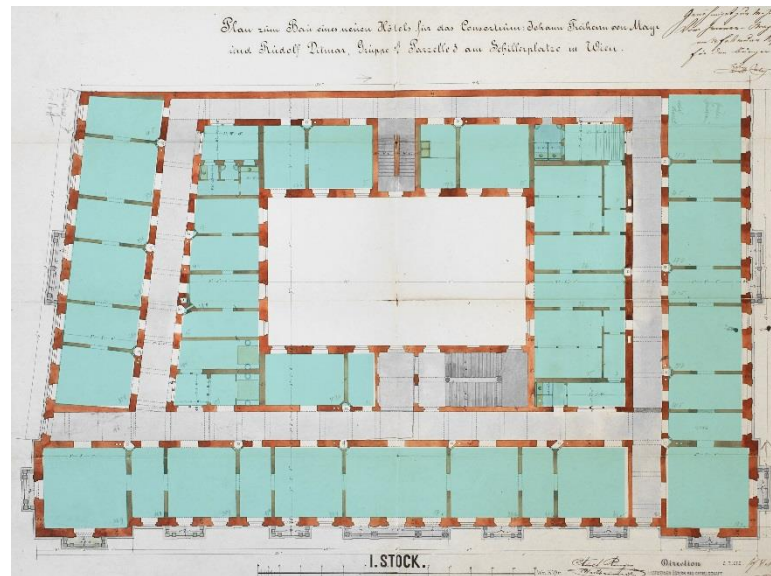


Figure 2.23 Typical floor plan, the conditioned area is highlighted

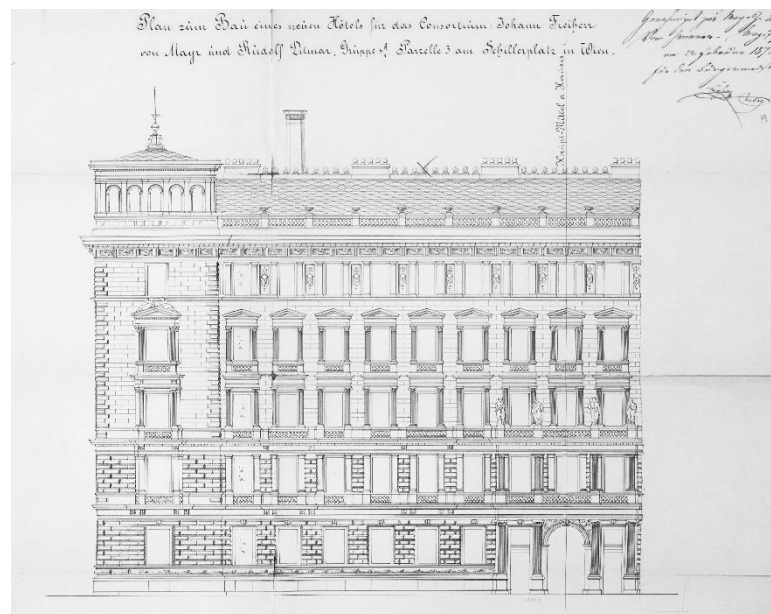
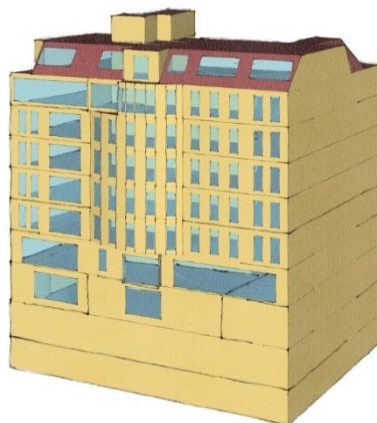
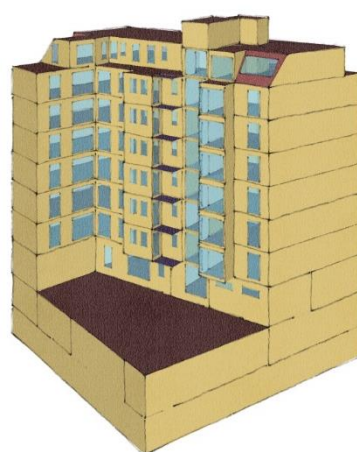


Figure 2.24 East elevation

## Cluster 7



*Figure 2.25 (Left) West view*



*Figure 2.26 (Right) East view*

*Table 2.7 Description of the building representing cluster 7*

<b>Address</b>	Mattiellistraße 3, 1040
<b>Total Area [m<sup>2</sup>]</b>	3667
<b>Conditioned Area [m<sup>2</sup>]</b>	2169
<b>Total Volume [m<sup>3</sup>]</b>	10622
<b>Conditioned Volume [m<sup>3</sup>]</b>	6108
<b>Year of construction</b>	
<b>Number of Zones</b>	22
<b>Primary Usage</b>	Residential
<b>Associated Area [m<sup>2</sup>]</b>	2169
<b>Other Usage(s)</b>	-
<b>Associated Area [m<sup>2</sup>]</b>	-
<b>Heating Demand per Total Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	43.46
<b>Heating Demand per Conditioned Building Area [kWh.m<sup>-2</sup>.a<sup>-1</sup>]</b>	73.49

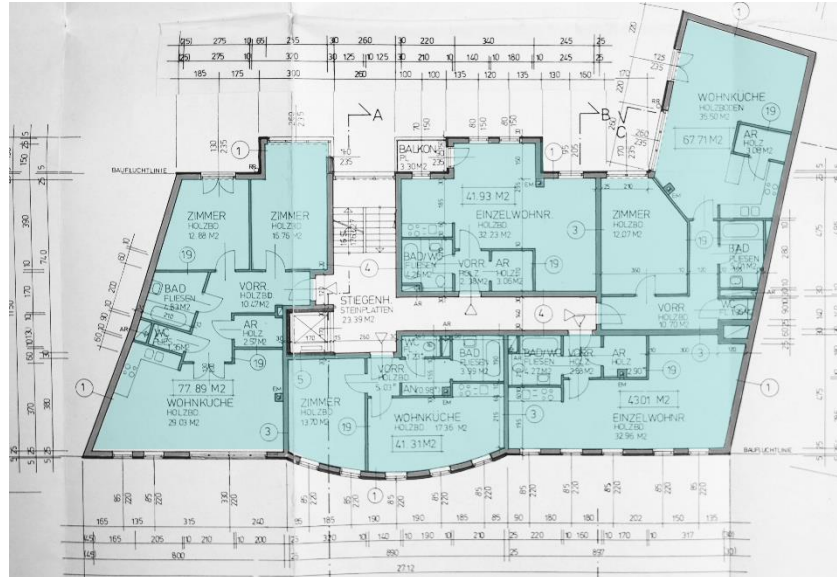


Figure 2.27 Typical floor plan, the conditioned area is highlighted

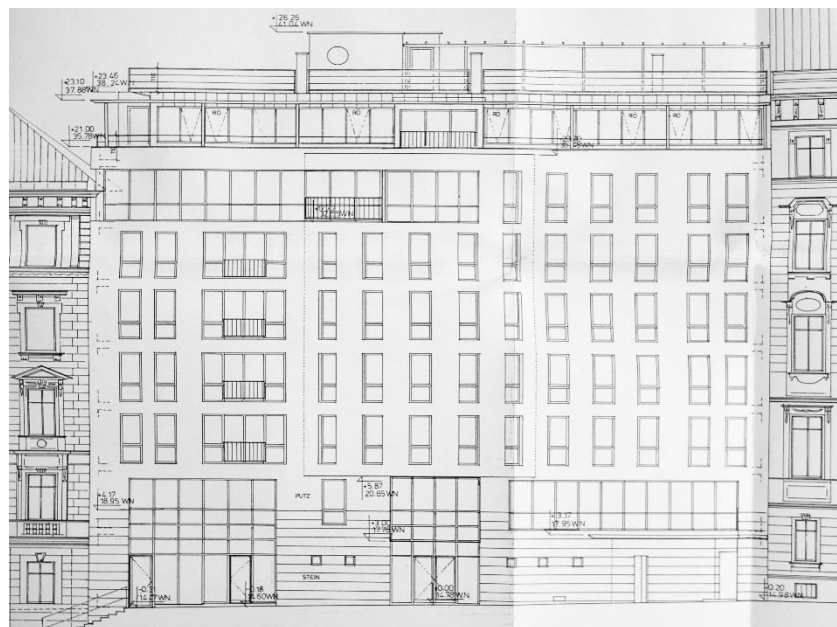


Figure 2.28 Main elevation

## CURRICULUM VITAE

### MSc. DI. Neda Ghiassi

Address: Währinger Straße 51/9 | 1090 Wien  
Telephone: (+43) 680 322 95 68  
E-mail: [neda.ghiassi@gmail.com](mailto:neda.ghiassi@gmail.com)  
Date of Birth: 08.03.1983  
Place of Birth: Teheran, Iran  
Nationality: Iran - Austria



### Educational Background

2013 – 2017	PHD in computer science, Environmental Informatics PHD College, <i>TU Wien</i> , Vienna, Austria Topic: Simulation-Supported Urban Building Stock Energy Demand Modeling
2010 -2013	Master of Building Science and Technology in <i>TU Wien</i> , Vienna, Austria (Graduated with honors) Final Project: Development of a building data model for a performance-guided optimization environment (SEMERGY)
2009	Design Project: Reincarnation of a historical neighborhood in Tehran Research Project: The Influences of the French School of Fine Arts on the Architectural Practice and Education in Iran
2006 - 2009	Master of Architecture in <i>ENSA de Paris-la-Villette</i> , Paris, France
2005	Final Project: Creativity Discovery Centre for Children
2001 - 2005	Bachelor of Architecture in Azad University, Faculty of Art and Architecture, Tehran, Iran
2000	Highschool graduation from Farzanegan High School (NODET), Tehran, Iran

### Professional Background

June 2016 – Aug. 2017	Project assistant, BPI, <i>TU Wien</i> , Vienna, Austria In MOTIVE project: Modellierung, Optimierung, und technische Integration von Vakuumglas-Elementen
May 2015 – May 2016	Project assistant, BPI, <i>TU Wien</i> , Vienna, Austria In BAU-Web project: Facilitation of Web-Based Building Product Data Procurement through Development of a Common Domain Ontology
Nov. 2013 – Feb. 2016	PHD college member in the Environmental Informatics doctoral college, <i>TU Wien</i> , Vienna, Austria
Nov. 2012 – Nov. 2014	Project assistant, BPI, <i>TU Wien</i> , Vienna, Austria In SEMERGY Project: Semantic Web Technologies for Performance-Oriented Building Design Optimization
March 2011 – Sep. 2012	Tutor in the Department of Building Physics and Building Ecology (BPI), <i>TU Vienna</i> , Vienna, Austria
Sep. 2007 – Oct. 2008	Assistant architect in <i>P. Nabon's Bureau d'Architecture</i> , Paris, France
Apr. – July 2006	Research coordinator in the <i>National Heritage Organization</i> , Teheran, Iran
March – May 2005	Internship in <i>A.R. Rouhani Architecture Firm</i> , Tehran, Iran
June – Sep. 2003	Internship in <i>Mojda Consulting Engineers</i> , Tehran, Iran
Jan. 2011 – Aug. 2013	Architect in <i>Abendroth Architekten</i> , Vienna, Austria

## Language skills

Persian	Native speaker
English	Near native
French	Excellent command
German	Very good command
Azeri	Good command
Arabic	Basic communication skills

## Computer Skills

Programming/DBMS	Python, Matlab, PostgreSQL, PostGIS
Data Analysis	R project for statistical computing, Microsoft Excel
GIS	QGIS
Performance Assessment	Energy Plus, EDSL-Tas, Dialux, Odeon, Archiphysik
CAD	AutoCAD, Revit, Sketch-up, 3D Studio Max, Rhinoceros, Grasshopper
DTP/image processing	Adobe InDesign, Adobe Photoshop

## Academic Publications

**N. Ghiassi.** 2013. Development of a Building Data Model for Performance-Based Optimization Environment, Master Thesis, Supervised by: A. Mahdavi, TU Wien. Vienna, Austria.

M. Emamgholivand, **N. Ghiassi.** 2009. La réincarnation d'un quartier du centre de Téhéran, Supervised by: Y.H. Shin. Ecole d'Architecture de Paris-la Villette. Paris, France.

**N. Ghiassi.** 2009. Les influences de l'école des beaux-arts sur l'architecture de l'Iran. Supervised by: C. Moley. Ecole d'Architecture de Paris-la Villette. Paris, France.

**N. Ghiassi.** 2006. Creativity Discovery Centre for Children. Supervised by: H.R. Nasernaseer. Faculty of Art and Architecture of Azad University, Tehran Center Branch. Tehran, Iran.

## Scientific Journal Entries

**N. Ghiassi, A. Mahdavi.** 2017. Reductive bottom-up urban energy computing supported by Multivariate Cluster Analysis. Energy and Buildings. Elsevier. Accepted on 2.3.2017. DOI: 10.1016/j.enbuild.2017.03.004

**N. Ghiassi, F. Tahmasebi, A. Mahdavi.** 2017. Harnessing buildings' operational diversity in a computational framework for high-resolution urban energy modeling. Building Simulation. Springer. Accepted on 8.2.2017. DOI: 10.1007/s12273-017-0356-1

M. Covic, U. Pont, **N. Ghiassi, M. Taheri, R. Bräuer, A. Mahdavi.** An Inquiry into the Current Practice of Building Product Data Handling by Different Stakeholders in Austria. Applied Mechanics and Materials, Vol. 861. DOI: 10.4028/www.scientific.net/AMM.861.547

A. Mahdavi, U. Pont, **N. Ghiassi, F. Shayeganfar, S. Fenz, J. Heurix, A. Anjomshoaa, A. Tjoa.** 2015. Performance-based building design and retrofit optimization : The SEMERY approach. Dosya (Bimonthly Magazine of the Chamber of Architects of Turkey), Digital Opportunities in Architecture, 35, Turkey.

U. Pont, **N. Ghiassi, S. Fenz, J. Heurix, A. Mahdavi.** 2015. SEMERGY: Application of Semantic Web Technologies in Performance-Guided Building Design Optimization. Journal of Information Technology in Construction, Vol. 20. DOI: 10.1201/b17396-38

S. Fenz, J. Heurix, T. Neubauer, A. Tjoa, **N. Ghiassi**, U. Pont, A. Mahdavi. 2014. SEMERGY.net - automatically identifying and optimizing energy-efficient building designs. Computer Science - Research and Development. Vol. 31. Issue 3. Springer. DOI: 10.1007/s00450-014-0285-5

D. Wolosiuk, **N. Ghiassi**, U. Pont, F Shayeganfar, A. Mahdavi, S. Fenz, J. Heurix, A. Anjomshooa, A. Tjoa. 2014. SEMERGY: Performance-Guided Building Design and Refurbishment within a Semantically Augmented Optimization Environment. Advanced Materials Research, Vol. 899. DOI: 10.4028/www.scientific.net/AMR.899.589

## Conference Publications

**N. Ghiassi**, A. Mahdavi. 2017. Re-diversification of Predictions of a Reductive Urban Energy Modeling Method. Proceedings of IBPSA 2017. San Francisco, California.

A. Mahdavi, **N. Ghiassi**, M. Vuckovic, M. Taheri, F. Tahmasebi. 2017. High-resolution representations of internal and external boundary conditions in urban energy modelling. Proceedings of IBPSA 2017. San Francisco, California.

A. Mahdavi, **N. Ghiassi**. 2017. Urban energy computing: an hourglass model. Building Simulation Applications Proceedings. M Baratieri, V. Corrado, A. Gasparella, F. Patuzzi (ed.); publisher of the Free University of Bolzano.

**N. Ghiassi**, A. Mahdavi. 2016. Utilization of GIS Data for Urban-Scale Energy Inquiries: A Sampling Approach. Proceedings of ECPPM 2016. Limassol, Cyprus.

**N. Ghiassi**, M. Taheri, U. Pont, A. Mahdavi. 2016. A Semantic Web Approach to Efficient Building Product Data Procurement. Proceedings of ECPPM 2016. Limassol, Cyprus.

**N. Ghiassi**, A. Mahdavi. 2016. A GIS-Based Framework for Semi-Automated Urban-Scale Energy Simulation. Proceedings of CESB 2016. Prague, Czech Republic.

**N. Ghiassi**, A. Mahdavi. 2016. Urban Energy Modeling Using Multivariate Cluster Analysis. Proceedings of BAUSIM 2016. Dresden, Germany.

A. Mahdavi, S. Glawischnig, **N. Ghiassi**. 2016. Urban Energy Computing: A Multi-Layered Approach. CEUR Workshop-Proceedings, 1559 (2016), Vienna, Austria.

**N. Ghiassi**, K. Hammerberg, M. Taher, U. Pont, O. Sunanta, A. Mahdavi. 2015. An Enhanced Sampling-Based Approach to Urban Energy Modeling. Proceedings of IBPSA 2015. Hyderabad, India.

**N. Ghiassi**, S. Glawischnig, U. Pont, A. Mahdavi. 2014. Toward a Data Driven Performance Guided Urban Decision Support Environment. Information and Communication Technology. Ir. Linawati, M.S. Mahendra, E.J. Neuhold, A. Tjoa, I. You (Eds.). Springer. Bali, Indonesia.

**N. Ghiassi**, U. Pont, A. Mahdavi, J. Heurix, S. Fenz. 2014. Efficient building design model generation and evaluation: The SEMERGY Approach. Proceedings of the 30th PLEA Conference. Ahmedabad, India.

**N. Ghiassi**, F Shayeganfar, U. Pont, A. Mahdavi, J. Heurix, S. Fenz, A. Anjomshooa, A. Tjoa. 2013. A comprehensive building model for performance-guided decision support. Proceedings of the 2nd CESBP. A. Mahdavi, B. Martens (ed.). ÖKK-Editions. Vienna, Austria.

**N. Ghiassi**, F Shayeganfar, U. Pont, A. Mahdavi, S. Fenz, J. Heurix, A. Anjomshooa, T. Neubauer, A. Tjoa. 2012. Improving the usability of energy simulation applications in processing common building performance inquiries. Proceedings of Simulace Budov a Techniky Prostrédi - 7. narodni konference s mezinarodni ucasti. Sikula, J. Hirs (ed.); Ceska Technika - nakladatelstvi CVUT. Brno, Czech Republic.

A. Mahdavi, U. Pont, F Shayeganfar, **N. Ghiassi**, A. Anjomshooa, S. Fenz, J. Heurix, T. Neubauer, A. Tjoa. 2012. Exploring the utility of semantic web technology in building performance simulation. Proceedings of BauSIM 2012.

C. Nytsch-Geusen et al. (ed.); Eigenverlag / wissenschaftliches Komitee der IBPSA Germany-Austria. Berlin, Germany.

A. Mahdavi, U. Pont, F. Shayeganfar, **N. Ghiassi**, A. Anjomshoaa, S. Fenz, J. Heurix, T. Neubauer, A. Tjoa. 2012. SEMERGY: Semantic web technology support for comprehensive building design assessment. Proceedings of ECPPM2012: eWork and eBusiness in Architecture, Engineering and Construction Reykjavík, Iceland. G. Gudnason, R. Scherer (Ed.); Taylor & Francis.

### Scientific Reports and Project Proposals

U. Pont, P. Schober, **N. Ghiassi**, K. Hauer, A. Mahdavi, M. Schuss, M. Wieser, D. Wolosiuk, A. Rubick. 2017. MOTIVE - Modellierung, Optimierung, und technische Integration von Vakuumglas-Elementen: Sondierung über die Detaillierung von Vakuumgläsern in neuen Holz(Alu)Fenster-Konstruktionen: Detaillierung, Bau und Simulation. Report for FFG - Programm Stadt der Zukunft. Research Proposal. **(Approved for funding)**

A. Mahdavi, U. Pont, **N. Ghiassi**, M. Vuckovic, P. Pfaffenbichler, T. Brezina, V. Bühlmann. 2017. MOMBASA - Mobility, Microclimate, Buildings, and Social Aspects of Urban Densification. Report for WWTF - Vienna Science and Technology Fund. Research Proposal.

A. Mahdavi, **N. Ghiassi**, M. Vuckovic, M. Taheri, F. Tahmasebi, S. Fenz, K. Kiesel, S. Glawischnig, U. Pont. 2016. Forschungsproposal EMULATE - Energy Modeling for Urban Level Assessment Tools and Environments / Research Studios Austria, 5. Report for FFG - Programm Research Studios Austria. Research Proposal.

U. Pont, F. Tahmasebi, **N. Ghiassi**, A. Mahdavi. 2016 MAIN\_STREAM - Menschenbezogene Aspekte in städtischer und regionaler Energie Anwendungs-Modellierung. Report for FFG - Programm Smart Cities Demo. Research Proposal.

**N. Ghiassi**, F. Tahmasebi, U. Pont, A. Mahdavi. 2016. URECA - Urban Energy Computing Approach. Report for FFG - Programm Stadt der Zukunft. Research Proposal.

U. Pont, B. Sommer, **N. Ghiassi**, A. Mahdavi. 2015. EDEN - Entwicklung einer strukturierten und fehlerminimierten Datenaufbereitung und Dokumentation für Energieausweise. Report for FFG - Stadt-der\_Zukunft\_2nd Call. Research Proposal. **(Approved for funding)**

Pirstinger, U. Pont, **N. Ghiassi**, A. Mahdavi. 2015. EPIKUR - Energieeffizienz-Potential intelligenter Kernverdichtung des urbanen Raums. Report for FFG - Stadt-der\_Zukunft\_2nd Call. Research Proposal. **(Approved for funding)**

U. Pont, **N. Ghiassi**, S. Glawischnig, A. Mahdavi. 2014. BAU\_WEB: Ontologien für automatisierte Erschließung web-basierter Bauproduktaten für Industrie, Planer und Klienten. Report for FFG - Produktion der Zukunft Initiative. Research Proposal. **(Approved for funding)**

S. Fenz, J. Heurix, T. Neubauer, A. Anjomshoaa, F. Shayeganfar, **N. Ghiassi**, M. Taheri, K. Hammerberg, D. Wolosiuk, C. Sustr, U. Pont, A. Mahdavi, A. Tjoa. 2014. SEMERGY - Identifying optimal building refurbishment strategies. Report for FFG. Intermediate Review.