URBAN BUILDING ENERGY MODELING

USING 3D CITY MODEL AND MINIMIZING UNCERTAINTY THROUGH BAYESIAN INFERENCE

MSc thesis in Geomatics for the Built Environment

Cheng-Kai Wang 2018

TU Delft

URBAN BUILDING ENERGY MODELING USING 3D CITY MODEL AND MINIMIZING UNCERTAINTY THROUGH BAYESIAN INFERENCE - A CASE STUDY FOCUSES ON AMSTERDAM RESIDENTIAL HEATING DEMAND SIMULATION

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Geomatics for the Built Environment

by

Cheng-Kai Wang

October 2018

The work in this MSc thesis is co-supervised by:

3D geoinformation group
Department of Urbanism
Faculty of Architecture & the Built Environment
Delft University of Technology

Intelligent electrical power grids group
Department of Electrical Sustainable Energy
Faculty of EEMCS
Delft University of Technology

Building and urban data science group
Department of Building
School of Design and Environment
National University of Singapore

Supervisors:	Prof.dr. Jantien Stoter
	Dr. Simon Tindemans
	Dr. Giorgio Agugiaro
Ext. supervisor:	Dr. Clayton Miller
Co-reader:	Daniela Maiullari MSc.
Board of Examiner:	Dr. Martijn Stellingwerff

ABSTRACT

To cut down immense greenhouse gases emission and energy consumption in the rapidly urbanizing world, a holistic understanding and rethinking of our dynamic urban energy system are inevitable. Performing bottom-up building energy modelings at urban scale based on Geographical Information System (GIS) and semantic 3D city models could be a promising option to provide quantitative and integrated energy solutions.

Nevertheless, input uncertainties either caused by limited data accessibility in most cities or parameters with stochastic variability (e.g. house occupancy profile) become one of the biggest obstacles to produce reliable and acceptable building energy modeling results. This study aims to address the heating demand simulation performance gap caused by input uncertainties. In this case study based on Amsterdam residential building stock, parameter importance ranking of the 14 simulation inputs are first derived according to the sensitivity analysis. The selected key uncertain parameters are then modeled a in probabilistic distribution way at postcode 6 level (approximately or slightly more than 10 buildings). Model calibration is based on the Bayesian approach and given six years (2010-2015) of gas consumption data to infer parameter posterior distributions. After the training phase, the calibrated annual heating demand simulation results of the validation years show significant improvement in modeling accuracy. Comparing the baseline and the calibrated simulation results, the averaged absolute percentage errors of energy use intensity (EUI) among at least 84 valid postcodes have decreased from 24.96% to 8.31% in 2016 and from 19.93% to 7.70% in 2017 respectively.

The calibrated urban building energy model would be most interested by municipalities, urban planners, and engineering consultancies. It can be used to evaluate long-term energy supply and demand strategies, identify building renovation saving potential, perform large-scale building performance mapping, and carry out retrofit measures assessment.

ACKNOWLEDGEMENTS

Incorporating the building energy simulation field into this Geomatics master thesis was an exciting and thrilling move made in 2017 summer. In one hand, the acquired knowledge and skills could hopefully make a little positive contribution to the urgent energy transition issues, to fulfill my persistent enthusiasm and commitment toward environmental sustainability. On the other hand, exploring a new field can always bring unexpected *uncertainties*. That is why I would like to express my gratitude to my interdisciplinary mentor team to make this research project happen.

First I would like to thank my mentor Jantien. Thank you for your trust and gave me the freedom to explore and develop the idea. Your critical feedback ensured different aspects are more thoroughly discussed and addressed. I am grateful to have Simon closely involved in this project. Simon immediately agreed to supervise the project after an email, even though he was still in the UK at that moment. Your humble personality, insightful and detail oriented comments taught me a great deal. I was also very inspired by how a physicist sees the world. Clayton complements a lot in giving helpful guidance with respect to building energy simulation when I was exploring this field in the beginning. Your comments made me think deeper into how to make the research results have more practical impacts. Unfortunately could not have Giorgio involved in the project earlier, otherwise, the challenging data harmonization part could probably be based on Energy ADE rather than starting from scratch.

A lot of thanks to Danielle, you are the best boss and friend during my time at ETH Zurich. I appreciate a lot for your initial decision, so I had the fortune to participate in the project at the Chair of Information Architecture, enjoy the cozy office, and meet many awesome friends. Most importantly, I enjoyed very much working with you as well as any random chat we had when sitting next to the beautiful window in the office. Fortunately, I can still have your constant support even I have returned to the Netherlands.

Sincere and special thanks to Marcel. The opportunity gave by you in March 2015 is probably the starting point of this incredible journey, so I had an opportunity to explore urban science together with many talented colleagues at Future Cities Laboratory in Singapore. Definitely a life-changing experience for me. Thank you.

I am grateful to have infinite supports from my family, without this, I could not have a chance to start this journey till nowadays. To my best friend, Yi-Jou, your constant support, being patient, and listening mean a lot to me, so I never feel alone during this sometimes stressful adventure. Also, thanks to many Taiwanese and Geomatics friends, hanging out together with you are the best moments during these two years.

Finally, "We are the first generation to feel the impact of climate change, and the last generation that can do something about it" - Barack Obama. Thanks to people dedicated to sustainable development...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AHN3 Actueel Hoogtebestand Nederland . . . 38

ASHRAE American Society of Heating, Refrigerating and Air-Conditioning Engineers . . . 35

BAG Basisregistratie Adressen en Gebouwen . . . 38

BES Building Energy Simulation . . . 2

BESTEST Building Energy Simulation Test . . . 17

CBS Centraal Bureau voor de Statistiek . . . 44

CFD Computational Fluid Dynamics . . . 15

CIM Canopy Interface Model . . . 16

EPBD Energy Performance of Buildings Directive . . . 41

EED Energy Efficiency Directive . . . 41

EUI Energy Use Intensity . . . 4

ESMs Energy Saving Measures

GIS Geographical Information System . . . 1

HVAC Heating, ventilation and air conditioning . . . 44

IEA International Energy Agency . . . 17

KNMI Koninklijk Nederlands Meteorologisch Instituut . . . 38

LOD Level of Detail . . . 6

SHAERE Sociale Huursector Audit en Evaluatie van Resultaten Energiebesparing . . . 41

UBEM Urban Building Energy Modeling . . . 1

UHI Urban Heat Island . . . 2

WFS Web Feature Service . . . 38

1 INTRODUCTION

Urban population is rapidly increasing. In 2014, 54% population lived in cities, and it is expected to climb to 66% by 2050 [10]. This fact is accompanied by increased energy demand per capita by 32% in the last 40 years [11]. Undeniably, urban energy systems worldwide face a tremendous challenge to support increasing energy demand in the built environment while achieving decarbonization target. To cope with such challenge, Urban Building Energy Modeling (UBEM) [12, 13] has been developed in many cities and aims at characterizing building stock energy consumption or predicting energy demand given hypothetical scenarios. Among many applications, urban building energy model has an important role to assist decision making and perform scenario analysis, for instance, evaluating urban building energy performance, identifying cost-effective building retrofit potential, balancing energy demand and supply, and supporting energy efficient building design and district planning [14, 15, 16].

Two urban building energy modeling approaches can be generally distinguished: top-down and bottom-up methods according to Swan and Ugursal [1]. Top-down approach models long-term total energy demand of the building stock often in country scale based on macroeconomic and socioeconomic parameters. It has a relatively coarse spatial-temporal resolution of energy consumption of the building stock. On the other hand, the bottom-up approach is more capable of developing a high spatial-temporal granularity energy modeling for a city. This feature enables many more applications, for instance, detailed end-use consumption calculations, assessing the consequence of introducing new technologies, and evaluating district-wide energy demand and supply strategies. Due to these advantages, it is the adopted approach for this study.

Bottom-up approach can be further separated into statistical method and engineering method [1]. Statistical method relies on historical consumption data, and building stock characteristics data, which are often derived from Geographical Information System (GIS) layers (e.g. building construction year, building surface area) or census statistics (e.g. population, income level), in order to build a mathematical model to predict urban building energy consumption [14, 17, 18]. On the other hand, engineering method often requires high level of detail data, from geospatial to building level data such as 3D building geometries, building thermal envelope properties, window to wall ratio, occupant number and occupancy profile, HVAC system (Heating, Ventilation, and Air Conditioning), etc. as inputs for a physical model or a simulation engine to estimate dynamic building energy consumption.

Although bottom-up engineering method is a more versatile tool to assess dynamic energy consumption, high computational cost of applying a building energy simulation engine at an urban scale is a significant challenge. Another obvious obstacle is how to properly deal with inherent simulation

uncertainties, particularly, input uncertainties alone could be significant factors to cause simulation performance gap [13, 19, 20].

Archetype classification is one of the solutions to assist input uncertainty issues. As urban building energy modeling usually covers hundreds or thousands of buildings, it is expensive and not yet possible to collect detailed individual building data according to the current registration system. A more common approach is to classify the building stock into different groups (archetypes) with each unique archetype sharing similar characteristics such as construction period, main building function, etc. [13, 21]. Modelers can then derive building thermal envelope properties from construction database such as European Building Stock Observatory[1], technical reports, or local experts, based on the corresponding construction period, or estimate occupancy schedule based on the main building function.

Input uncertainties can be further separated into *subjective* type and *stochastic* type uncertainty. In the first case, deterministic parameter value does exist but unknown to the modeler because of incomplete information. For instance, thermal transmittance U-values of wall, roof, floor, etc. On the other hands, some simulation inputs are inherently uncertain and fluctuating. For instance, it makes little sense to define occupancy schedule, thermostat setting of the building, etc., deterministically. Cerezo et al. [21] suggest that an additional level of detail can be introduced by assigning the key uncertain parameters in a probabilistic way. Depending on the modeling scale and data resolution, this can be based on a single building or to a group of buildings of the same archetype. The Bayesian inference approach and the measurement data can be applied to compute the parameter posterior distributions. The subjective type uncertain parameter could potentially be reduced to a deterministic value if sufficient data is given; while the distribution of the stochastic type uncertain parameter could be effectively refined to describe the underlying random process.

Nevertheless, one should be aware that the calibration process is an over-specified and under-determined problem. In most cases, there are relatively few measured energy data available, but way more model inputs. This could lead to over-fitting issues [22, 23]. To overcome this problem, sensitivity and uncertainty analysis [3, 24] plays an important role to identify the most influential variables affecting the simulation results. The modeler can thus prioritize data collection procedure or make more sensible assumptions for the prior probability distributions of these key uncertain inputs.

As Saltelli et al. [24] point out, "A model, as a human representation of a given problem, represents essential simplifications and simulation constraints". An additional aspect to be considered in energy modeling is thus the model adequacy in terms of the modeling scale and purpose. Building Energy Simulation (BES) has been widely used to model building performance for decades, the popular tools include EnergyPlus, TRNSYS, Modelica, etc. Some of the sophisticated tools are capable of simulating building energy performance in details, but often require diverse and detailed inputs and have high computational cost. Additionally, complicated urban microclimate, e.g. the Urban Heat Island (UHI), has a direct impact on building

1 European Building Stock Observatory: https://ec.europa.eu/energy/en/data-analysis/building-stock-observatory

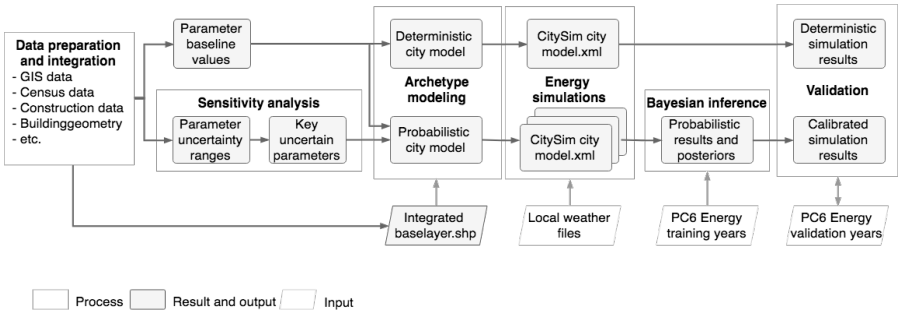


Figure 1.1: The generalized workflow followed by this research. Only key processes and results are shown. Complete breakdown of this workflow is discussed in Chapter 3

energy consumption [15, 25, 26, 27], while these aspects are often simplified or ignored at BES scale.

Among numerous simulation tools, CitySim is selected because of the following characteristics. CitySim models the dynamic irradiation on the exterior building surfaces to consider the effects of inter-building obstruction. Meanwhile, a resistor-capacitor (R-C) thermal model is implemented in CitySim to calculate the thermal exchange between the outdoor and indoor environment [28]. As a consequence, CitySim can simultaneously consider important geometric features at an urban scale including the building size, shape, orientation and density in response to local weather data and at the appropriate level of detail. The simplified building thermal model not only reduces computational cost but also eases the burden of managing detailed building level data, which is often the biggest obstacle at an urban scale simulation.

Figure 3.2 shows the generalized workflow followed by this research, which summarizes the key processes and (intermediate) outputs for developing a calibrated urban building energy model. This workflow can be further broken down into several research questions and steps as discussed below.

1.1 RESEARCH QUESTIONS

The main research question of the project is defined as follow:

- *How to realize Urban Building Energy Modeling (UBEM) using 3D city model and minimizing parameter and simulation uncertainties based on the Bayesian approach?*

In order to answer this main question, these sub-questions need to be answered subsequently.

- *What open-source GIS layers, statistics data, and technical datasets related to Amsterdam are available for the UBEM development? and how to integrate these datasets in a sensible way to build up a simulation ready 3D city model?*
- *What is the added value and appropriateness of adopting CitySim as a simulation engine in this research scope?*

- *Given a number of simulation parameters with the associated uncertainty ranges, which ones are the key parameters affecting annual heating Energy Use Intensity (EUI) calculation and which ones are minimal and even ignorable, according to the sensitivity analysis?*
- *How to infer building dwelling type in order to assist deterministic archetype classification?*
- *Which key uncertain parameters should be applied to probabilistic archetype modeling?*
- *How could the Bayes' theorem be applied to infer posterior probability distributions of the uncertain building parameters and how effective is it to improve simulation accuracy?*

1.2 THESIS OUTLINE

The following will be discussed in this document.

- Chapter 2 discusses the theoretical background related to the Urban Building Energy Modeling (UBEM) development. This includes motivations, existing modeling approaches, considerations in terms of modeling domain and scale, introducing the selected simulation engine: CitySim and its characteristics, the role of the semantic 3D city model, introducing the uncertainty nature in energy simulation, how to tackle uncertainty through available sensitivity analysis methods.
- Chapter 3 introduces the adopted methodology of this research. The essential steps such as data preparation and uncertainty quantification, sensitivity analysis of the simulation engine, probabilistic archetype modeling, Bayesian calibration framework and validation are introduced sequentially.
- Chapter 4 is the first implementation section of the research. Data requirements, sources and data quality will be discussed. Deterministic (baseline) values and uncertainty ranges of the inputs are summarized. The data selection and cleaning rules and how to integrate heterogeneous datasets together and transform into a simulation ready 3D city model will be discussed.
- Chapter 5 takes a closer look at CitySim simulation characteristics and discusses its pros and cons. The sensitivity analysis result of CitySim will be discussed.
- Chapter 6: The Bayesian inference and model calibration framework is introduced. The implementation and the validation results are presented.
- Chapter 7 discusses and concludes the research results, problems encountered in this research, the implemented solutions and the possible future work.

2 THEORETICAL BACKGROUND AND RELATED WORK

2.1 OVERVIEW OF URBAN BUILDING ENERGY MODELING

The building sector plays an indispensable role in achieving a low-carbon future as it accounts for more than one-third of total final energy use and CO_2 emissions [29, 30]. Meanwhile, fulfilling increasing energy demand and achieving decarbonization target in the rapidly urbanizing world is a tremendous challenge [13]. Nevertheless, the building sector has tremendous potential to reduce energy consumption and greenhouse gases emission by improving building energy efficiencies such as enforcing appropriate retrofit measures and net-zero emission buildings and urban districts planning [29], where energy models play an important role.

Frayssinet et al. [27] breaks down urban energy challenges into three aspects and these facts illustrate why urban energy modeling remains an active field for the last 30 years [12].

- Urban population is rapidly increasing, with 54% population lived in cities in 2014 and is expected to climb to 66% by 2050 [10]. This fact is accompanied by increased energy demand per capita by 32% in the last 40 years [11]. Sustainable urban development and holistic energy policies to balance the increased resource demands is thus a crucial issue.
- Renewable energy sources are changing the landscape of energy market rapidly [31]. Because of the decentralized and intermittent characteristics of renewable energy sources, a comprehensive understanding of urban energy systems is crucial to bridge the gap between energy demand and production.
- Urban heat stress and increased cooling demands caused by urban heat island (UHI) effects may be further intensified and become more frequent in the context of climate change [32] and lead to public health problems.

Allegrini et al. [15] complement to the second point by pointing out energy hubs, networks, and multi-energy grids are promising technologies to achieve greater penetration of renewable generation, as balancing demand and supply between different building types can improve load matching and resource utilization. However, the temporal mismatch between energy demands and availability of energy sources is still a significant barrier.

Since buildings and interconnected infrastructures and services typically have a course of decades life cycle, data-driven planning, parametric modeling, and energy optimization design at building or urban level are of increasing interest to achieve sustainable urban development and to avoid

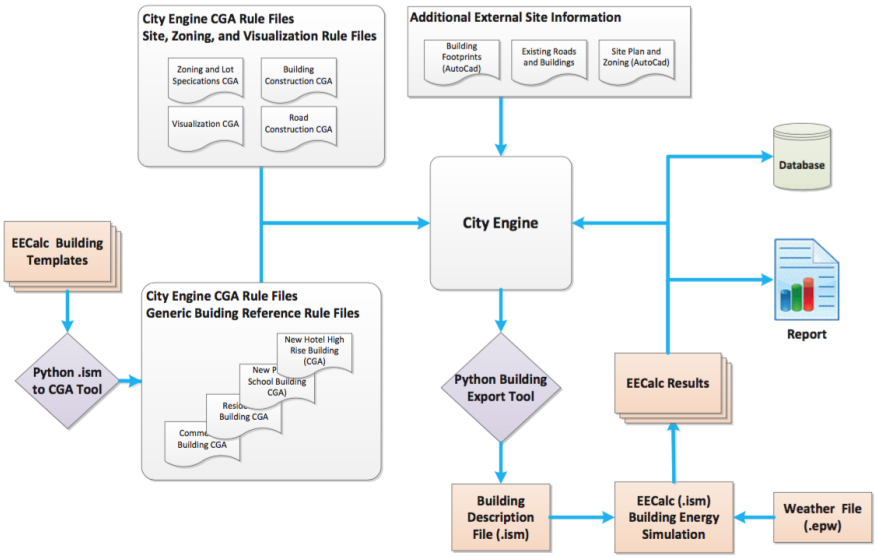


Figure 2.1: Data flow and platform architecture of the LakeSim project, which has building energy simulation model integrated into the application to assist parametric design assessment [16].

unsatisfactory energy performances [16, 33]. These kinds of design optimization or decision support applications are particularly important to the rapidly urbanizing regions or where refurbishment measures are urgently needed. This leads to emerging spatial decision support environments such as TRANSFORM[1], CityBES[2], SusCity[3], MUSIC-iGUESS[4] and LakeSim[5] where energy models are important components of the platforms (see Figure 2.1 for example). However, urban building energy modeling remains challenging for multiple reasons, for instance, high Level of Detail (LOD) data is rarely accessible, inherent simulation uncertainties, computational cost, dynamic urban microclimate, stochastic occupant behavior, etc., all add complexity to the urban energy modeling.

Urban energy modeling has broad application fields. Increased accessibility and improved quality of spatial, non-spatial data as well as advancement in geometric and geo-data modeling capabilities become indispensable components to enable urban-level energy analysis [15]. The scope of this research will be a case study based in Amsterdam, the Netherlands, focusing on residential building heating demand simulation and dealing with input uncertainties. Nevertheless, the review below will start from a broad introduction of urban building energy modeling approaches and subsequently dive into the discussions with respect to modeling scales, options of simulation engines, and the chosen one for urban scale modeling in this research and its challenges.

1 TRANSFORM: http://urbantransform.eu/decisionsupportenvironment/
2 CityBES: https://citybes.lbl.gov/
3 SusCity: http://groups.ist.utl.pt/suscity-project/home/
4 MUSIC-iGUESS: https://www.list.lu/en/research/project/music/
5 LakeSim: https://www.ci.uchicago.edu/blog/lakesim-designing-future-cities

Table 2.1: Three major residential energy modeling approaches summarized by Swan and Ugursal [1]

Approach	• Advantage	• Disadvantage
Top-down	• Long term forecasting in the absence of any discontinuity • Inclusion of macroeconomic and socioeconomic effects • Simple input information • Encompasses trends	• Reliance on historical consumption information • No explicit representation of end-uses • Coarse analysis
Bottom-up statistical	• Determination of typical end-use energy contribution • Encompasses occupant behaviour • Inclusion of macroeconomic and socioeconomic effects • Uses billing data and simple survey information	• Reliance on historical consumption information • Multicollinearity • Large survey sample to exploit variety
Bottom-up engineering	• Determination of each end-use energy consumption by type, rating, etc • "Ground-up" energy estimation • Model new technologies • Determination of end-use qualities based on simulation	• Assumption of occupant behaviour and unspecified end-uses • Detailed input information • Computationally intensive

2.2 MODELING APPROACHES: TOP-DOWN AND BOTTOM-UP MODEL

Many studies have investigated how different modeling approaches can be used to predict building and urban scale energy consumption. Swan and Ugursal [1] have categorized two main approaches to address urban energy modeling issues: top-down and bottom-up model, where bottom-up can be further distinguished by statistical and engineering method. Table 2.1 summarizes the basic characteristics of these modeling approaches.

2.2.1 Top-down model

Top-down approach models total energy consumption of building stock, often at the country scale, and based on macroeconomic and socioeconomic parameters such as energy price, income, and population density. Since detailed technological components of the city are not considered explicitly, the top-down model is not able to represent energy consumption of buildings at high spatial-temporal resolution [1, 27, 34].

2.2.2 Bottom-up model

Bottom-up approach is able to generate high temporal and (potentially) spatial energy demand estimations of the building stock and enables more detailed end-use consumption calculations [1]. This can be further distinguished by statistical method and engineering method, namely, data-driven and law-driven model according to Coakley et al. [22]

2.2.3 Bottom-up statistical method

At building scale, bottom-up statistical modeling uses recorded data as inputs of the mathematical models to predict energy consumption. These models can be regression fitting, artificial neural network (ANN), random forest, support vector machine (SVM), etc. [22, 35, 36, 37, 38]. Many studies have indicated that statistical method results in better prediction accuracy than engineering method [35, 37]. However, the downside is that it relies on site-specific recorded data and consequently more difficult to generalize when unobserved conditions happen. Unlike engineering method, statistical method is much limited in predicting the impact of new technologies and energy saving potential through retrofit measures [14]. Additionally, the complex interaction between input and output data is often sophisticated and difficult to interpret and thus has less transparency [37].

When statistical method is applied at urban scale, building characteristic such as construction year, envelope properties, building main function, building shape, etc., become commonly selected predictors to study its correlation with energy consumption. Tso and Yau [39] studied three methods of predicting electricity consumption: regression, decision tree, and neural network. The results show that decision tree and neural network have better predicting performance but the difference among three methods are minimal. This indicates that regression is a valid model for predicting energy consumption and with the benefits of more interpretable parameters introduced in the analysis.

Many studies have utilized multiple linear regression model and together with GIS as a data collection, management, analysis and visualization platform to estimate energy consumption at an urban scale [14, 17, 18, 40]. The general form of multiple linear regression can be written as follow:

$$y = I + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \epsilon \tag{2.1}$$

Where y is the dependent variable (energy consumption); I the general model intercept; β_i the regression coefficient ($i = 1, 2, ..., p$); x_i the predictors ($i = 1, 2, ..., p$); ϵ the error term

Howard et al. [17] analyzed the spatial distribution of end-use energy intensities for domestic hot water, space heating and cooling, base electricity in New York City. Building function and building floor area are predictors of the robust multiple linear regression model in this study. The analysis is performed in the aggregate level of 170 zip-code data points, individual building on the same tax lot is not distinguishable. Visualization of the result can be seen in Figure 2.2

Torabi Moghadam et al. [18] developed a bottom-up statistical model via robust multiple linear regression analysis and identified period of construction, heated volumes, type of ground floor, occupation factor, air temperature, type of roof, and the installed power are most related to space heating demands of the residential building stock of the city Settimo Torinese, Italy.

Mastrucci et al. [14] used statistical method to breakdown aggregated level energy consumption to end-use by multiple linear regression. To identify significant independent variables of the regression model, the authors

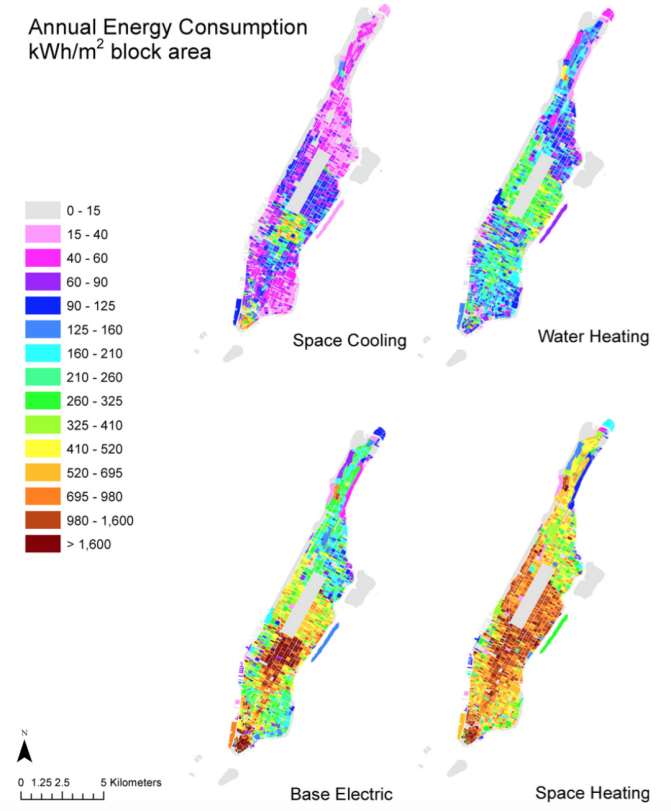


Figure 2.2: Spatial distribution of end-use energy intensity in New York City estimated by multiple linear regression model. Retrieved from Howard et al. [17]

carried out a two steps procedure: first identifying main influential variables on energy consumption of dwellings based on the current state of the art; secondly performing stepwise regression analysis to select the statistically significant variables. Dwelling type, construction period, floor area are the main predictors of gas consumption.

Braulio-Gonzalo et al. [40] took year of construction, and additional features of surrounding area such as building shape coefficient (represents surface to volume ratio, S/V), solar orientation of the main facade, street aspect ratio and urban block pattern as predictors to estimate building energy performance and discomfort hours for the cooling and heating in Castellón de la Plana, Spain. The statistical analysis is based on approximate Bayesian inference using Integrated Nested Laplace Approximation (INLA) and identified building shape coefficient and construction year are the most influential parameters while solar orientation of the main facade of the building is least significant.

Feature and model selection are important steps especially when dealing with high-dimensional and multi-domain building energy data. Identifying key independent predictors with minimal interactions among them is crucial to give better model performance, stability, interpretation and to reduce over-fitting and multicollinearity (predictors are correlated) issues [1, 41]. This is part of the scope of sensitivity and uncertainty analysis, which will

Table 2.2: Key variables identified by different statistical methods in case studies

Method	Purpose	Key variables
Robust MLR with M-estimators	End-use energy use intensities of domestic hot water, space heating and cooling, base electricity in New York City [17]	building function, building floor area
Robust MLR with AIC	Heating demands of the residential building stock of the city Settimo Torinese, Italy [18]	construction period, heated volumes, type of ground floor, occupation factor, air temperature, type of roof, installed power
Stepwise MLR	End-use energy consumption in residential building stock, Rotterdam, the Netherlands [14]	dwelling type, period of construction, floor area
Bayesian inference INLA	Building energy performance and discomfort hours for the cooling and heating in Castellón de la Plana, Spain [40]	year of construction, shape coefficient
Stepwise MLR with SRC	Heating energy consumption in historical dwellings in Bath, UK [42]	number of open flues

MLR: Multiple Linear Regression; AIC: Akaike Information Criterion; INLA: Integrated Nested Laplace Approximation; SRC: Standardized Regression Coefficients

be introduced in section 2.6.

Other method such as principal component analysis (PCA) is studied by [43] and is used to filter and select key predictors of consumption for district heating, building electricity use, cold water consumption and total heat loss in Stockholm. Table 2.2 summarizes the identified key influential variables by different statistical methods in different building stocks.

2.2.4 Archetype modeling

Before diving into engineering method, it is worth to introduce archetype characterization. Urban building energy modeling usually deals with more than hundreds or thousands of buildings, it is expensive and not realistic to collect detailed building data for each of them according to the current registration system. Consequently, a common approach is to classify the building stock into different groups (archetypes) with each archetype sharing similar characteristics such as construction period, main building function, etc.

There is no consensus about general approach for archetype modeling because fundamental interactions may differ by geographic context and consequently often rely on generic assumptions and literature [13]. Reinhart and Cerezo Davila [13] review 17 works, the number of archetypes can range from 5 to 3168. European TABULA project[6] generates country-wide archetypes for 13 European nations based on building dwelling type (detached, terrace, etc.) and construction period [6]. Monteiro et al. [44] classified partial residential building stocks of Lisbon into 18 archetypes based on construction period, size class (single-family, multi-family), roof type, and neighboring.

6 TABULA: http://webtool.building-typology.eu/#bm

Cerezo et al. [21] proposed three approaches in the case study to classify the building stock:

- Available literature: characterizing building properties such as construction and occupancy details deterministically based only on available literature, e.g. technical reports, engineering database, or scientific papers.

- Local expertise: acquiring a deeper knowledge from local construction and engineering practices and sometimes applying the information derived from the representative buildings for archetype classification.

- Probabilistic parameters: while the local expertise could significantly reduce the uncertainties of specific parameters, it makes little sense to define occupancy profile, lighting loads, thermostat setting, etc., deterministically since these variables could either have *subjective uncertainty* or *stochastic variability*. Consequently, the authors introduce an additional level of detail by assigning probabilistic distributions to key uncertain parameters. Depending on the modeling scale and data granularity, this can be based on a single building or a group of buildings of the same archetype. The Bayesian inference approach and the measurement data can be applied to compute the parameter posterior distributions. The subjective type uncertain parameter could potentially be reduced to a deterministic value if sufficient data is given; while the distribution of the stochastic type uncertain parameter could be effectively refined to describe the underlying random process.

The advantage of estimating building characteristics by probabilistic distribution also help modelers to examine energy simulation result in probabilistic ranges rather than a deterministic value [21].

2.2.5 Bottom-up engineering method

Engineering method requires detailed building characteristics data as model inputs, for instance, occupancy profile, thermostat setting, air infiltration rate, and so on (detailed parameters see Figure 3.3), and it simulates energy demand based on the science of building physics. The capability to generalize and predict system behavior given previously unobserved conditions is the biggest advantage of this modeling approach, and thus it is often used to assess and quantify the impacts of retrofit measures, future climate scenarios, new technologies or to assist policy making [1, 22, 27, 34]. These advantages make it the preferred approach in the scope of this research.

Nevertheless, this method comes with the cost that it is computationally more intensive than statistical method. Depending on the modeling purposes and how sophisticated the underlying calculations or simulation engines are, the required inputs can vary dramatically and might up to hundreds parameters or even more [1, 22, 41]. Accessibility, quality, resolution and uncertainty of the model inputs are common challenges of this approach [12].

The next section will go further into bottom-up engineering method and introduce the modeling scale aspect and general features of specific simulation models and engines.

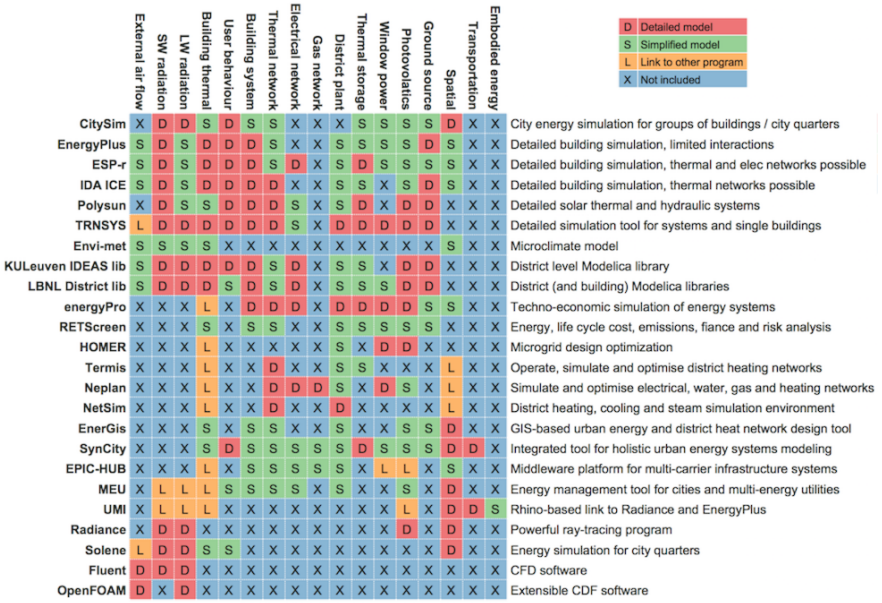


Figure 2.3: Allegrini et al. [15] assess 20 energy simulation tools based on 17 modeling capabilities and its level of details.

2.3 SCOPE AND SCALE OF MODELING

2.3.1 Scope of modeling

Urban energy modeling is a broad topic and consists of numerous problem domains. Allegrini et al. [15] identify and classify district-scale energy system into three key parts, district energy systems (such as heat networks, multi-energy systems and low temperature networks), renewable energy generation (for instance, solar, bioenergy, wind and the related topic of seasonal storage), and the urban microclimate which affects building energy demands. Although each of them addresses different purposes and are modeled with different tools, the authors emphasize the needs to consider district-level interaction in energy systems and assess the theory of modeling approaches and capabilities of 20 existing simulation engines which can be seen in Figure 2.3.

It is clear that one should choose a suitable simulation engine depending on the modeling purpose and scale. In some cases, especially when the study covers broad spatial or temporal scales, a single simulation engine might not be able to give reasonable result to the research problem. Coupling different engines to benefit from the strength of each tool is also a common modeling approach [26, 27, 45]. Additionally, a compromise between inputs LOD and computational cost is always an issue to be considered.

The following paragraphs specifically address scale aspect and requirements for UBEM.

2.3.2 Simulation engine options and considerations

Many models and simulation engines have been developed and mainly used in the community for the whole building energy simulation. Simulation at this scale can generally be called as Building Energy Simulation, BES, and it is unnecessary to precisely consider the local urban effects [27]. The missing or simplified aspects at BES level will be further discussed in section 2.3.3.

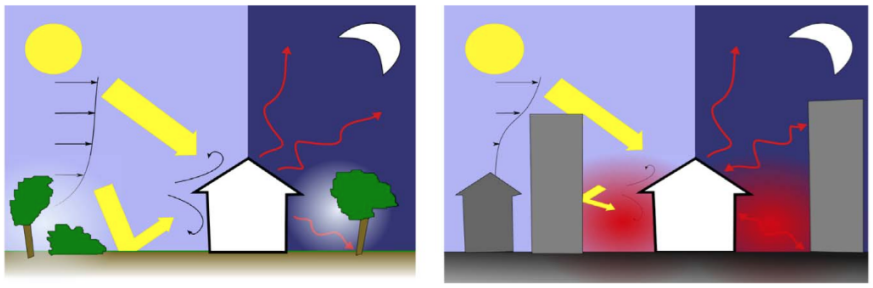
Heo et al. [46], Booth et al. [47] carried out building energy retrofit analysis under uncertainty, and applying Bayesian calibration on normative models (CEN/ISO), which are light-weight, quasi-steady state formulations of heat balance equations and thus more efficient to model large sets of buildings. Buffat et al. [48] studied building stock modeling with large-scale GIS data and modeled with the SIA 380/1 Norm, which is a building thermal model also established on the ISO model and uses a monthly steady-state method to estimate and model the heat balance of a building. The arguments for choosing simplified ISO models are mostly due to compromise between level of details and computational effort [46, 47, 48]

Except ISO models, many researchers have carried out district to urban level building energy modeling with more advanced dynamic simulation engines such as EnergyPlus, and TRNSYS. In the review work of [15], these are classified as holistic simulation tools as they have capabilities to span many areas of interests, see the matrix presented in Figure 2.3. EnergyPlus, developed by the U.S. Department of Energy (DOE) [49], has gained popularity in the research and simulation community for decades. It primarily simulates building level energy costs given hourly local weather information, a building geometry, HVAC description, utility rate structure, and etc. [15, 22, 49]. Davila et al. [50], Sokol et al. [51] used EnergyPlus to carry out urban residential building energy modeling and calibrated by Bayesian model.

TRNSYS is a transient system simulation program with a modular structure, which implements component-based simulation approach and allowing the users to implement new components and mathematical models for various purposes. The tool is intentionally designed for detailed energy system simulations rather than modeling general energy flows at district or city scale [15, 22].

2.3.3 Urban scale energy modeling considering microclimate

The simulation models and tools discussed in the previous section are mainly designed for BES and with limited capabilities to take interactions between adjacent buildings into account, or often only shadowing is modelled [15]. This is no longer sufficient in district or urban scale modeling as microclimate has significant influence on building energy demands. For instance, air temperature is higher in urban areas due to the Urban Heat Island (UHI) effect and wind speeds are lower due to sheltering effect [25]. These local interactions motivates that microclimate should be considered when performing urban scale energy simulation. Coupled modeling with different simulation engines can also be executed rather than the microclimate being



(a) Stand-alone building (b) Urban building

Figure 2.4: Modification of the energy balance of an urban building compared to a stand-along building. Retrieved from Frayssinet et al. [27]

a predetermined boundary condition [15, 26, 27, 52]

Frayssinet et al. [27] illustrated three spatial scales of energy modeling (Figure 2.5) and summarized four main effects to be accounted for at urban scale modeling. The general difference is illustrated in Figure 2.4:

- Obstructions and sheltering effects, which has direct impacts on solar gains and the radiative cooling to the sky
- Surrounding surfaces account for the reflected solar radiations and emitted longwave radiations from the ground surface.
- Urban morphology will affect urban airflows and ventilation, which influence convective heat exchanges.
- The general Urban Heat Island (UHI) effect means city center has a higher temperature than suburban areas because of above mentioned reasons, extra the anthropogenic heat sources as well as less evaporative cooling due to lack of vegetation. These facts account for reduced heating demands and increased cooling demands in the city center.

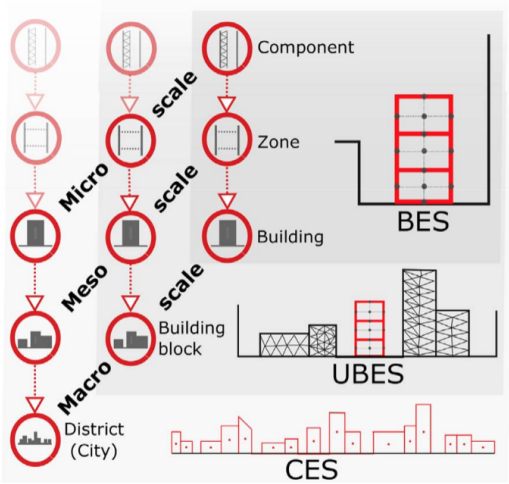


Figure 2.5: Frayssinet et al. [27] classified energy simulation into three spatial scales.

Due to multi spatial-temporal scale modeling characteristics (Figure 2.6) and high computational cost in general, the authors indicates that there is

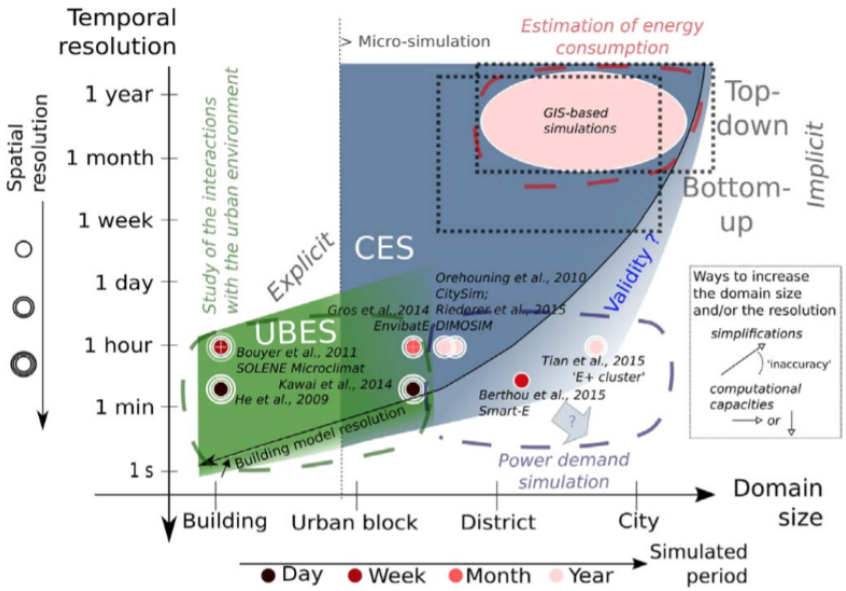


Figure 2.6: Multi-scale and multi-domain nature of energy simulation. Retrieved from Frayssinet et al. [27]

still no entirely validated simulation tool able to model accurately and explicitly the energy demand of urban buildings at the city scale [27]

Dorer et al. [26] studied the influence of urban microclimate on the building energy demands in street canyons with different aspect ratios (building height/street canyon width). The authors used BES tool: TRNSYS and adapted with other models such as Computational Fluid Dynamics (CFD) to account for (1) the radiation exchange between buildings; (2) the convective heat transfer adapted to the local flow field and (3) the UHI effect. The influence from urban microclimate is obvious (Figure 2.7) and the authors explained that due to multiple reflections and more solar and thermal radiation is entrapped, wider street canyon require higher cooling demands, while lower cooling demands in narrow street canyons is expected as less solar radiation entering the street canyon.

Above discussion justifies that taking environmental interactions into account at urban scale simulation is needed. However, a certain level of simplification of the models to reduce computational cost should also be considered. This leads to the discussion of why choosing CitySimin in this case study. The general CitySim characteristics are introduced below.

2.4 OVERVIEW OF CITYSIM FOR URBAN SCALE ENERGY MODELING

Compromise between level of details and computational cost is always an issue to be concerned. Although sophisticated simulation tool such as TRNSYS has components to simulate both radiative and convective model, it requires high level of detail inputs, modelers experience, and high compu-

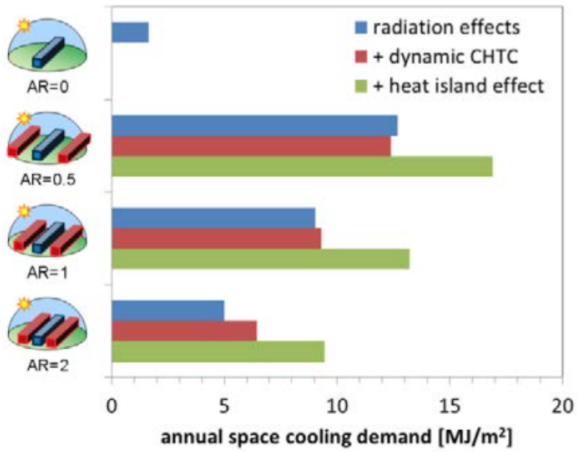


Figure 2.7: The experimental result modeled in TRNSYS and CFD on annual cooling demand estimations of a stand-along building and street canyons with different aspect ratios. Retrieved from Dorer et al. [26]

tational cost and thus is not the first choice for an urban scale simulation.

On the other hand, CitySim developed at the Swiss Federal Institute of Technology Lausanne (EPFL) by [28] is a holistic simulation tool to model resource flows in urban environment and specifically focusing on inter-building interactions such as shadowing, light inter-reflection, and infrared exchanges [28, 53]. It consists of a simple Resistor-Capacitor (R-C) network, analogy with the electrical circuit, to perform energy simulation. The short-wave and longwave radiation model (simply denoted as radiation model in this document) accounts for solar gains on facades and roofs. Beside that, it includes simplified occupant behavior with deterministic or stochastic modeling options, and plant/equipment models [15, 28, 45].

CitySim requires 3D building model and ground surface to model environmental interactions between buildings. Building geometry and construction details such as facade U-value (thermal transmittance coefficient, W/m^2K), window to wall ratio (glazing ratio) and etc. of each surface; operation details such as number of occupants and occupancy profile of the building can be specified in the specific CitySim XML schema or CityGML schema for simulation.

Like most simulation engines listed in matrix 2.3, CitySim does not include convective model to account for external air flow and ventilation; weather file is usually collected from the historical hourly observation data of a nearby meteorological station or from Meteonorm[7]. These facts indicate that the influence of the UHI effect is simplified to some extent. To evaluate microclimate in greater detail, Mauree et al. [54] tried to couple CitySim with Canopy Interface Model (CIM), where CIM provides high-resolution vertical meteorological profiles to CitySim while CitySim is capable of representing surface and building thermal performance.

To validate whether the simplified thermal model of CitySim can still produce valid result, Walter and Kämpf [53] conducted a comparative testing

7 Meteonorm: http://www.meteonorm.com/

approach within the frame of the Building Energy Simulation Test (BESTEST) method, which is developed by the US National Renewable Energy Laboratory, the International Energy Agency (IEA) and the US Department of Energy [55]. The IEA BESTEST suite used in the study is composed of reference kernel of ESP (UK), BLAST 3.0 Level 215 (USA), DOE2.1E-W54 (USA), SERIRES/SUNCODE (USA), SERIRES-1.2 (USA), S3PAS (Spain), TRNSYS (USA) and TASE (Finland). The comparison is carried out in sequence of different subsequent changes of the base case such as building mass, windows orientation and shadings. Annual heating and cooling loads and annual peak heating and cooling loads considering ideal heating and cooling control system are compared with validity ranges generated from the reference programs. The authors conclude CitySim results are consistent with those of more sophisticated program except two cases, peak cooling requirement (case 610) and annual heating requirement (case 960) are outside the range by approximately 0.14% and 0.54% (Figure 2.8).

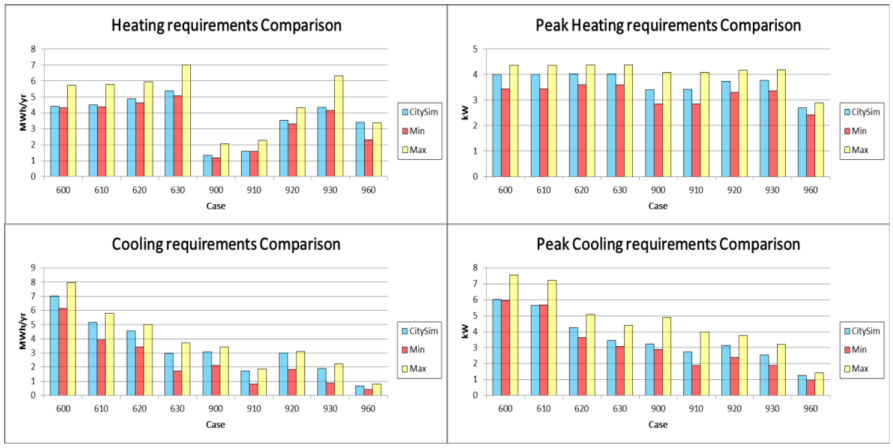


Figure 2.8: Comparative testing results of CitySim compared with reference kernels based on BESTEST for the annual heating, cooling loads, and peak heating and cooling requirements. Retrieved from Walter and Kämpf [53]

Meanwhile, the experimental verification with the monitored data for the EPFL campus building shows that CitySim simulated result overestimates heating demand by 5.1% in year 2012 and 0.6% in year 2013, but are within acceptable range. The authors explained the deviation could be caused by the simplified occupancy behaviour, uncertainties associated with temperature set-point, air tightness of the envelope, and the averaged weather file obtained from Meteonorm [53].

2.5 BOTTOM-UP ENERGY SIMULATION ENABLED BY SEMANTIC CITY MODEL

Regardless of the modeling scale and purpose, bottom-up simulation generally requires way more inputs than top-down approach or bottom-up statistical method. Collecting, cleaning and integrating such detailed information from heterogeneous data sources is a rather complex and resource-demanding task, let alone aggregating, disaggregating or updating existing datasets. To deal with this challenge, semantic data (city) model is al-

most indispensable for most UBEM developments. Two previous bottom-up UBEM projects, TRANSFORM[8] and MUSIC-iGUESS[9] [10] with focus on Amsterdam and Rotterdam respectively, have developed respective data model and workflow (script) to manage and integrate large amount of spatial/non-spatial data based on the relational database management system (both use PostgreSQL[11]) [56, 57]. Monteiro et al. [44] developed an urban building database (UBD) to run an urban building energy modeling in Lisbon, Portugal. While most of these projects, to name just a few, have achieved the individual study purpose, the developed semantic data (city) models are application specific and require additional efforts to be reused by other simulation tools or domains.

An international standardized data model such as CityGML based on the Open Geospatial Consortium (OGC) standard [58] is an alternative option and has potential to increase data interoperability and to facility data exchange for multi-scale and multi-domain simulations. CityGML is based on the Geography Markup Language (GML) to represent and exchange virtual 3D city models. It can define 3D geometry, semantics, ontologies and appearance of most relevant topographic objects of different spatial scales on different levels of detail (LODs, see Figure 2.9 left) [58]. Beside the existing CityGML thematic modules (bridge, building, city furniture, and so on, see Figure 2.9), it is possible to extend the new classes and attributes by the Application Domain Extension (ADEs) such as Energy ADE [59], Utility Network ADE [60], etc, where the Energy ADE is highly relevant to the urban building energy modeling purpose.

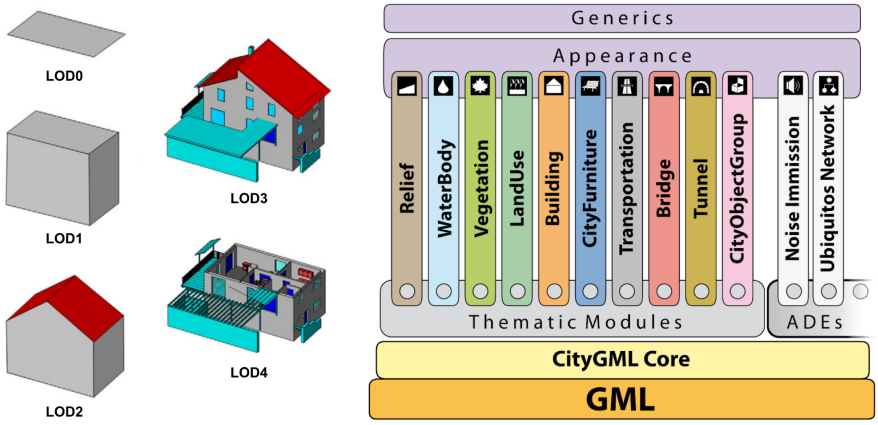


Figure 2.9: Left: Five level of detail representations in CityGML. Right: General CityGML architecture, thematic modules and application domain extensions (ADEs). Retrieved from [58, 61]

Since its standardization in 2008, there are many urban energy modeling related researches and projects based on CityGML. Few examples include, Hong et al. [62] adopt CityGML and Energy ADE as one main component within the platform architecture to develop CityBES, a web-based platform to support city-scale building energy efficiency. Nouvel et al. [63] investigate

8 TRANSFORM: http://urbantransform.eu/decisionsupportenvironment/
9 MUSIC: https://www.list.lu/en/research/project/music/
10 iGUESS: https://github.com/ERIN-LIST/iguess
11 PostgreSQL: https://www.postgresql.org/

the influence of data quality on annual urban heating demand modeling based on CityGML LOD1 and LOD2 3D city models and energy simulation tool SimStadt [64]. The authors classify the simulation inputs into three data categories depending on relevance, where the *Must-have* category includes building construction year, building function, refurbishment information, residence type (main, secondary, vacant); LOD2 city model belongs to *Nice-to-have* category, the third relevance category. Case studies in partial districts of two German cities (Ludwigsburg, Karlsruhe) and Dutch city (Rotterdam) have also gone through energy modelings by using CityGML 3D city model [19]. More CityGML and Energy ADE use cases can be found in [59].

Despite the above mentioned benefits of using the standardized data model like CityGML and Energy ADE, which is also a valid format to CitySim, it is not yet implemented in this work merely because the author had founded the way and developed the relevant scripts to integrate datasets and generate (overwrite) CitySim XML during the course of exploring CitySim, see Figure 4.6, and Figure 4.10.

2.6 UNCERTAINTY IN URBAN BUILDING ENERGY MODELING

As Saltelli et al. [24] point out, "A model, being a human representation of a given problem,...represents essential simplifications and simulation constraints, and might sometimes be erroneous or a poor representation of reality". Awareness of this fact brings to the discussion of uncertainty and sensitivity analysis for (urban) building energy simulation process. Uncertainty and sensitivity analysis can not only help to overcome building knowledge gap, identifying and ranking the uncertainty sources, might also assist calibration process to obtain a better probabilistic description [65] e.g. helping probabilistic archetype modeling, which is briefly discussed in section 2.2.4.

The relationship between the observations z_i, the true process $\zeta(\cdot)$, and the computer model process $\eta(\cdot,\cdot)$ can be described in Equation 2.2, according to Kennedy and O'Hagan [23].

$$z_i = \zeta(\boldsymbol{X_i}) + e_i = \rho\eta(\boldsymbol{X_i}, \boldsymbol{\theta}) + \delta(\boldsymbol{X_i}) + e_i \tag{2.2}$$

Where e_i is the observation error for the ith observation; ρ is an unknown model parameter; $\delta(\cdot)$ is a model inadequacy function that is independent of the code output $\eta(\cdot,\cdot)$; $\boldsymbol{X_i}$ is vector of known inputs while $\boldsymbol{\theta}$ is vector of uncertain inputs. This formula can be translated and corresponds to the source of uncertainty listed in Table 2.3

In the scope of this study, building characteristics and operational uncertainty (simply denoted as parameter or input uncertainty in this document) will be the main focus for uncertainty and sensitivity analysis, while observation error is assumed minimized (data selection process see section 4.1.6). Weather observation data is collected from meteorological station and thus has little control over it in the current research scope. CitySim model adequacy is discussed in the previous section and treated as an adequate black box.

Table 2.3: Source of uncertainty in building energy models. From Heo [2]

Category	Factors
Scenario uncertainty	Outdoor weather conditions Building usage/occupancy schedule
Building physical/operational uncertainty	Building envelope properties Encompasses occupant behaviour HVAC systems Operation and control settings
Model inadequacy	Modeling assumptions Simplification in the model algorithm Ignored phenomena in the algorithm
Observation error	Metered data accuracy

2.7 OVERVIEW OF SENSITIVITY AND UNCERTAINTY ANALYSIS

When dealing with complex system like building energy simulation model or statistical method with multiple inputs, sensitivity and uncertainty analysis is a powerful tool to answer the question like how can we define the importance of each variable unambiguously? which variables are ignorable when limited time and computational resources become constraints? and how to quantify uncertainty ranges for the variables? [3, 24, 66]. Specifically, sensitivity analysis (SA) is the study of relative importance of different inputs on the model output, while uncertainty analysis (UA) focuses on quantifying the uncertainty in model output. The general framework of performing sensitivity analysis is illustrated in Figure 2.10

Sensitivity analysis can be further distinguished by: local sensitivity analysis and global sensitivity analysis. Local sensitivity emphasizes on the effects of uncertain inputs around a point (or a base case). Global sensitivity studies the influences of uncertain inputs over the entire input space, thus it is more a reliable method but also computationally more expensive [3, 24].

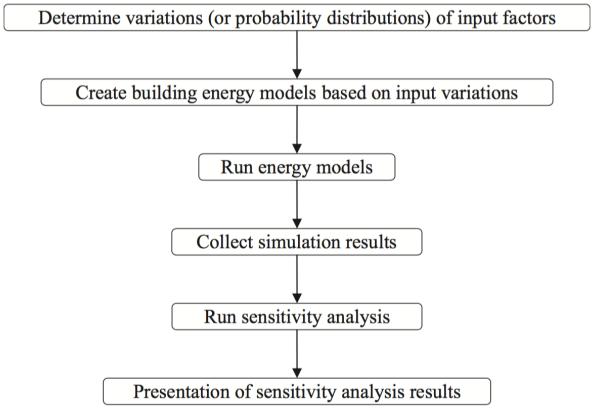


Figure 2.10: Typical workflow of sensitivity analysis in building performance analysis. Retrieved from Wei [3]

Table 2.4: Global sensitivity methods and characteristics. Adapted from Wei [3]

Method	Characteristics	Subtype
Regression	• Fast and easy to interpret • SRC and t-value are only suitable for linear models and can not be used in the presence of correlated factors • SRRC and PCC can be used for non-linear but monotonic functions • Applicable to observational study	SRC SRRC PCC t-value step-wise adjust R square AIC
Screening	• Qualitative measures to rank factors, not suitable for uncertainty analysis • Model free approach, suitable for large number of inputs and computationally intensive models • No self-verification	**Morris**
Variance	• Model free appraoch, suitable for complex non-linear and non-additive models • Quantify all the variance of the output and consider interaction effects among variables • Highest computational cost among all global methods • FAST is not suitable for discrete distribution and only consider non-linear effects, but not interaction effects	Sobol FAST
Meta-model	• Suitable for complex and computationally intensive models • Quantify output variance of different inputs • The accuracy dependent on the applied meta-model • Applicable to observational study	MARS ACOSSO SVM GP TGP

SRC: Standardized Regression Coefficients; SRRC: Standardized Rank Regression Coefficient; PCC: Partial Correlation Coefficients; AIC: Akaike Information Criterion; FAST: Fourier Amplitude Sensitivity Test; MARS: Multivariate Adaptive Regression Splines; ACOSSO: Adaptive Component Selection and Smoothing Operator; SVM: Support Vector Machine; GP: Gaussian Process; TGP: Treed Gaussian Process

Wei [3] conducted a comprehensive overview regarding different sensitivity techniques applied in building performance analysis and observational study. The introduction below presents basic characteristics and limitations of these methods based on the works done by [3, 24, 66, 67], and summarized in Table 2.4 (only global methods are presented).

2.7.1 Local sensitivity analysis

Local sensitivity analysis, also named as differential sensitivity analysis, belongs to the class of the one-factor-at-a-time methods (OAT). The usual approach is to study how output is changed by changing one factor while all other factors are fixed. Comparatively straightforward approach gives this method advantage regarding computational cost while the drawbacks include: it only explores a reduced space of the input factor around a base case; the interactions are not considered; no self-verification in this method, which means the analyst does not know how much of the total variances of outputs have been taken into account in the analysis [3].

2.7.2 Global sensitivity analysis

Global sensitivity analysis examines the influence of uncertain parameters over the whole parameter range. It includes regression methods, screening-

based methods, variance-based and meta-modeling approaches.

Be noted that the screening-based method (Morris method) is the only adopted sensitivity analysis method in this study. The other methods are introduced because of their relevance to the building energy performance study, reading comprehension shall not be affected even without the prior knowledge of these methods. The overview paragraphs below are mainly summarized from [3, 24, 66, 67].

Regression method

Regression method is fast to compute and easy to interpret, and thus widely used in building energy analysis. Regression and variance based method are based on the decomposition of the Equation 2.3 [68]

$$Y(\boldsymbol{X}) = f_0 + \sum_{i=1}^{k} f_i(X_i) + \sum_{j>i}^{k} f_{ij}(X_i, X_j) + f_{12\ldots k}(X_1, \ldots X_k) \tag{2.3}$$

The model response $Y(\boldsymbol{X})$ is a vector of one model output. $\boldsymbol{X}$ is a $N \times k$ matrix of model inputs X with N samples of k input parameters defined with the parameter space by the lower and upper bounds for each parameter, X_{min}, X_{max} respectively. In order to estimate sensitivity indices based on regression analysis, the model response $Y(\boldsymbol{X})$ can be approximated by a linear multidimensional model $F(\boldsymbol{X})$ with a regression coefficient f_i for each input parameter X_i as shown below, and assuming the decomposed individual input factors are independent:

$$F(\boldsymbol{X}) = f_0 + \sum_{i=1}^{k} f_i(X_i) \tag{2.4}$$

Estimated regression coefficients become comparable when they are standardized using the variance of the model response $V(Y(\boldsymbol{X}))$ and the variance of the corresponding input $V(X_i)$, which has the following form:

$$SRC_i = f_i \frac{V(X_i)}{V(Y(\boldsymbol{X}))} \tag{2.5}$$

The absolute value of SRC_i indicates the importance of variable on model output, while the sign indicates the positive or negative correlation between input and output. One should be noted that SRCs are only applicable when the model response $Y(\boldsymbol{X})$ can be sufficiently represented by the approximated linear regression model, namely, it is not suitable when the building model is highly non-linear [68]. To understand how well the approximated linear model fits the (possibly non-linear) building model, coefficient of determination R^2 is used, which means how much of the building energy model variance $V(Y(\boldsymbol{X}))$ can be explained by the variance of the approximated linear model $V(F(\boldsymbol{X}))$:

$$R^2 = \frac{V(F(\boldsymbol{X}))}{V(Y(\boldsymbol{X}))} \tag{2.6}$$

Low R^2 might indicate relevant variables (predictors) are not included in SA, or due to non-linearity or interaction effects that are not captured by the regression analysis. A threshold of $R^2 = 0.7$ is defined by Saltelli et al. [69] for the acceptance of the approximated regression model and the resulting

SRCs.

Another commonly used technique in regression method includes: forward (backward) stepwise regression with selection criteria by means of SRC, t-values, adjusted R^2, and Akaike Information Criterion (AIC).

A example case of using regression method for SA would be applying Monte Carlo or Latin hypercube sampling method to sample input space and to perform multiple simulation runs. The simulation inputs and outputs are then collected in order to run a sensitivity analysis.

Screening-based method

The Morris method is one of the most commonly used approaches due to its efficiency in combination with factorial sampling strategy. The computation cost of Morris method depends on number of parameters k, number of trajectory t and leads to a total number of observations: $t * (k + 1)$. Morris method gives two indices: μ_i^* is sensitivity index used to estimate the elementary (main) effect of a variable on the model output and to rank the importance of model variables. The second index δ_i can be interpreted as a measure for non-linearity and parameter interactions. The illustration of variable ranking by Morris method can be seen in Figure 2.11 and Figure 5.6. Morris method will be more thoroughly discussed in section 3.4

$$EE_i = \frac{y(x_i, ..., x_{i-1}, x_i + \Delta, ..., x_k) - y(x_i, ..., x_k)}{\Delta} \tag{2.7}$$

$$\mu_i^* = \frac{1}{r} \sum_{t=1}^{r} |EE_{i,t}| \tag{2.8}$$

$$\sigma_i = \sqrt{\frac{1}{r-1} \sum_{t=1}^{r} |EE_{i,t} - \mu_i^*|^2} \tag{2.9}$$

The drawbacks of Morris method include: it provides qualitative measures by ranking input variables but does not quantify the effects of different variables on outputs. Namely it does not allow self-verification, because the modeler does not know how much of the total variances of outputs have been taken into account [3].

Variance-based method

The variance-based method tries to decompose the uncertainty of outputs and apportion to different sources of inputs based on the functional decomposition scheme shown in Equation 2.3. This leads to the ANOVA decomposition of total model variance (Equation 2.10). The notation discussed here follows section 2.7.2, and the detailed derivation of the following formulas can refer to [67]

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + ... + V_{12...k} \tag{2.10}$$

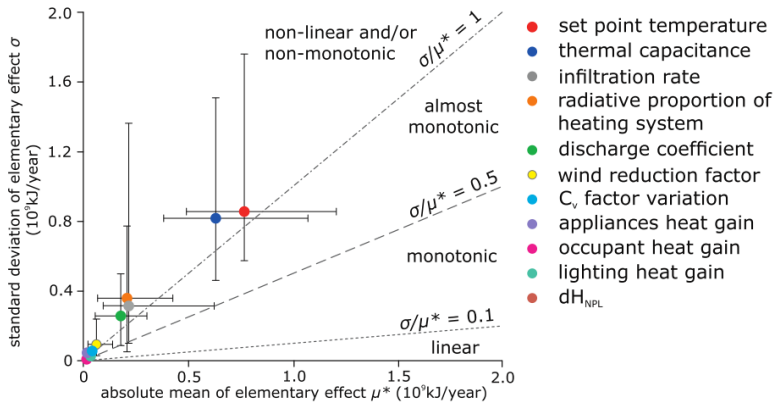


Figure 2.11: Example of using Morris method to rank variable importance. Retrieved from Menberg et al. [66]

where $V_i = V(E(Y|X_i))$ (equivalents to Equation 2.11) is the contribution of each input to the total variance of the output with the inner expectation operator denoting that the mean of Y is taken over all possible values of X_i; $V_{ij} = V(E(Y|X_i, X_j)) - V_i - V_j$ is the joint effect of the parameter pair (X_i, X_j) on the outcome, and accordingly for all other higher-order effects.

Two indices are often used: first order effects and total effect. First order effects consider the main effects for the output variations due to the corresponding input and has the following form:

$$V_{X_i}(E_{\boldsymbol{X}_{\sim i}}(Y|X_i)) \tag{2.11}$$

where $\boldsymbol{X}_{\sim i}$ denotes the $N \times k$ matrix of all inputs but X_i. In other words, it is the expected reduction in variance that would be obtained if X_i could be fixed. The normalized first order sensitivity coefficient is written as:

$$S_i = \frac{V_{X_i}(E_{\boldsymbol{X}_{\sim i}}(Y|X_i))}{V(Y)} \tag{2.12}$$

Total effects represent total contributions to the output variance due to the corresponding input, where first, second, and all higher-order effects caused by interactions are included (Equation 2.13):

$$S_{Ti} = \frac{E_{\boldsymbol{X}_{\sim i}}(V_{X_i}(Y|\boldsymbol{X}_{\sim i}))}{V(Y)} = 1 - \frac{V_{\boldsymbol{X}_{\sim i}}(E_{X_i}(Y|\boldsymbol{X}_{\sim i}))}{V(Y)} = S_i + S_{ij} \ldots + S_{ij \ldots k} \tag{2.13}$$

where $E_{\boldsymbol{X}_{\sim i}}(V_{X_i}(Y|\boldsymbol{X}_{\sim i}))$ is the expected variance that would be left if all inputs but X_i could be fixed; $V_{\boldsymbol{X}_{\sim i}}(E_{X_i}(Y|\boldsymbol{X}_{\sim i}))$ is the expected reduction in variance that would be obtained if all inputs but X_i could be fixed. Equation 2.13 also indicates that the difference of total effects and first order effects can be used to indicate overall higher order effects of an input (Equation 2.14).

$$S_H = S_{Ti} - S_i = S_{ij} \ldots + S_{ij \ldots k} \tag{2.14}$$

Similarly, based on Equation 2.10, for the second order effects S_{ij} for parameter pairs:

$$S_{ij} = \frac{V_{\boldsymbol{X}_{ij}}(E_{\boldsymbol{X}_{\sim ij}}(Y|X_i, X_j)}{V(Y)} - S_i - S_j \tag{2.15}$$

As a model free approach, variance-based method is suitable for complex nonlinear and non-additive models. Additional advantage includes the capability to quantify all the variance of output and also consider the interaction effects among variables. The downside of this approach is its high computational cost. Fourier Amplitude Sensitivity Test (FAST) and Sobol method are two most commonly used numerical approaches to estimate sensitivity indices S_i, S_{Ti}, S_H, S_{ij} while Sobol method requires much more computational cost compared to other global sensitivity analysis. The illustration of Sobol sensitivity result can be seen in Figure 2.12

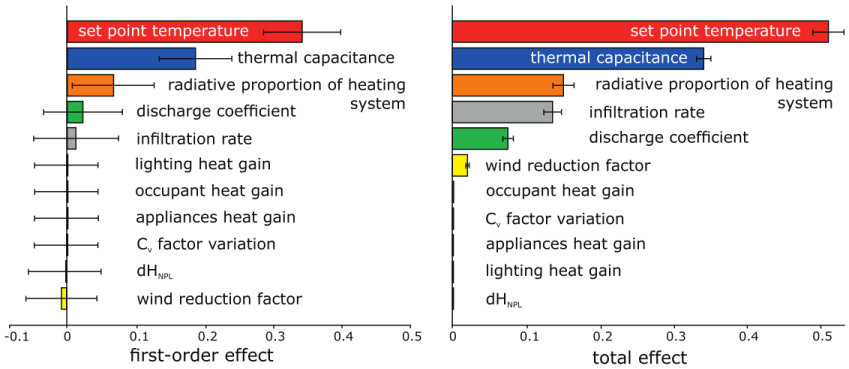


Figure 2.12: Example result of the first order and total effect of the variables conducted by Sobol method. Retrieved from Menberg et al. [66]

Meta-model based method

Meta-model based method is suitable for complex (linear, non-linear) models as it is created using non-parametric regression methods to approximate the objective functions based on statistical or machine learning models, which means there is no predetermined form. Sensitivity analysis measures are calculated by using meta-model based the on variance-based method. Efficient calculation is the biggest advantage of adopting meta-model based method, which is faster than running detailed building energy simulation models. Meanwhile, it also allows variance quantification of the output for different inputs since it uses variance-based method in the second step.

According to the review done by Wei [3], commonly used meta-model techniques include: MARS (Multivariate Adaptive Regression Splines), SVM (Support Vector Machine), ACOSSO (Adaptive COmponent Selection and Smoothing Operator), GP (Gaussian Process) and TGP (Treed Gaussian Process).

Table 2.5 summarizes the different sensitivity analysis methods applied in different case studies as well as the most influential variables, based on the review work of Wei [3]. The table reveals that key influential variables not just varies according to the modeling purposes, but also geographical locations and building (stock) characteristics.

Table 2.5: Global sensitivity analysis techniques applied in different simulation engines and case studies. Summarized from Wei [3]

Method	Purpose	Key variables
SRC	Energy performance of office buildings in four USA cities by EnergyPlus [70]	key variables changed by different climate zone
	Cooling energy for a residential building, Italy [71]	solar shading, window area, window insulation
SRRC	Annual cooling load in low-rise apartment in hot-humid climate by EnergyPlus [72]	natural ventilation, window area, solar heat gain coefficient
	Heating energy use in a mixed-mode office buildings, UK [73]	infiltration rate, lighting gains, equipment gains
t-value	Heating energy and overheating risk in an office building, UK [74]	internal heat gains
Morris	Office building energy use, Denmark [75]	lighting control, ventilation during winter
	Energy rating in a house, Italy [76]	indoor temp., air change rate, number of occupants, metabolic rate, equipment heat gains
FAST	Building thermal performance for typical office buildings, Italy [77]	envelope transparent surface ratio
Sobol	Air temperature for an experimental house, France [78]	heating capacity, infiltration, fibreglass thickness, heat exchanger efficiency, internal heat gains, fibreglass conductivity
	Final energy consumption in office buildings in six european cities [79]	climate zone is the most influential factor
MARS	Influences of climate change on a office building for overheating risk, UK [80]	lighting gains, equipment gains, weather conditions
ACOSSO	Thermal performance of a campus building, UK [81]	lighting gains, solar heat gain coefficients of windows, cooling degree days, equipment heat gains

SRC: Standardized Regression Coefficients; SRRC: Standardized Rank Regression Coefficient; FAST: Fourier Amplitude Sensitivity Test; MARS: Multivariate Adaptive Regression Splines; ACOSSO: Adaptive Component Selection and Smoothing Operator; SVM: Support Vector Machine; GP: Gaussian Process; TGP: Treed Gaussian Process

3 METHODOLOGY

This chapter starts with introducing the selected test area for the UBEM development in this case study. It is then followed by methodology overview, which is broken down into several steps and research questions. These steps include data preparation, parameter uncertainty quantification and data integration, energy model sensitivity analysis, building archetype classification, Bayesian inference and calibration, and validation of the calibrated model.

3.1 TEST AREA

The whole project is studying how to realize Urban Building Energy Model (UBEM) development from 3D city model and with particular focus on residential building heating demand simulation and model calibration. In the first place, a semantic 3D city model of the test area, managed in the form of a programmable CitySim XML file, is required by the simulation engine CitySim to perform dynamic heating demand simulations.

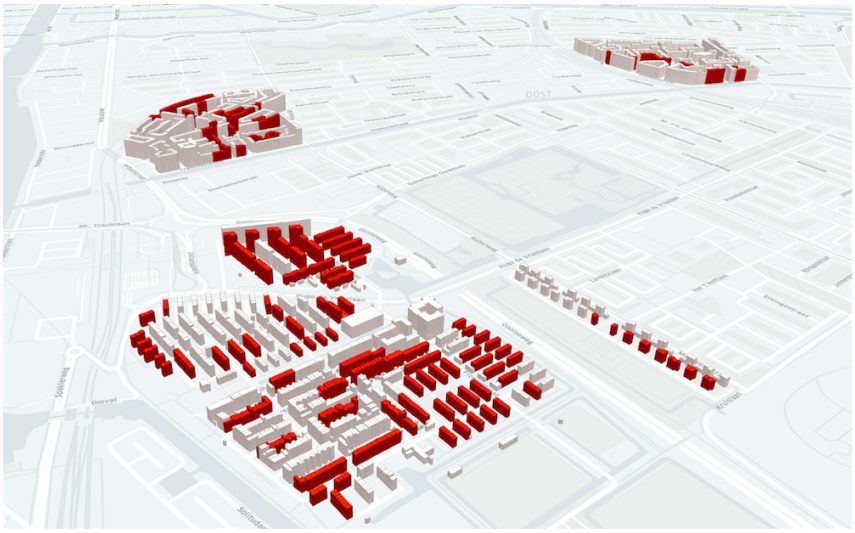


Figure 3.1: 3D city model of the partial residential districts in Amsterdam-Oost. Buildings colored in red have completed data (building metadata, energy, etc.) to perform heating demand simulation and calibration.

In order to reflect adequate building stock diversity with regards to construction periods and dwelling types, three residential building districts in Amsterdam-Oost (Amsterdam-East) are selected, see Figure 3.1. Several years of annual metered gas consumption data is collected and will be used to calibrate the model. Due to the fact that the highest spatial resolution of the annual metered gas consumption data is only accessible at an aggregate

postcode 6 level (approximately or slightly more than 10 buildings), datasets with different spatial resolutions will at some point be integrated and aggregated to the same level of details. In total, 2178 buildings are included in the test area. At least 84 postcodes within the test area fulfill simulation data requirement (selection criteria is explained in Chapter 4), and will be used for model calibration and validation.

3.2 RESEARCH METHODOLOGY

A complete bottom-up Urban Building Energy Modeling (UBEM) is often built from several interdependent tasks, which is illustrated in Figure 3.2. This general framework is concluded from extensive but non-exhaustive literature review and mainly discussed in the works done by [2, 21, 22, 47]. The following paragraphs will go through these steps in depth.

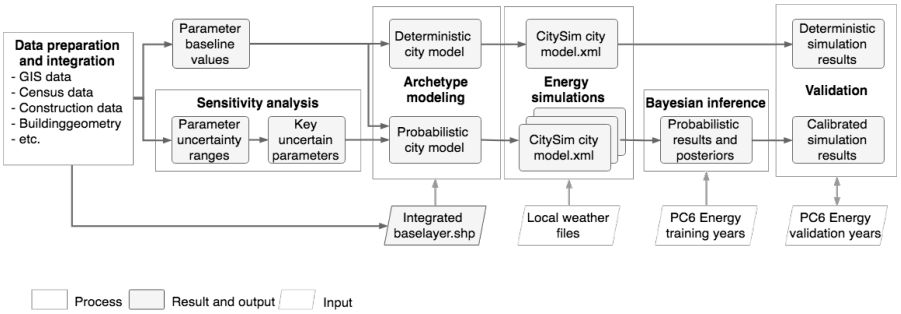


Figure 3.2: The research project is following this generalized framework. Only key processes and results are presented. Complete breakdown of this workflow is discussed in respective sections in this chapter.

3.3 DATA PREPARATION AND UNCERTAINTY QUANTIFICATION

While the required inputs for running an energy model is significantly dependent on the adopted simulation engine, they can in general be grouped into these five data categories, weather, geometry, construction, system, and operation respectively, see Figure 3.3. These data categories also apply to CitySim inputs, the table part in Figure 3.3 summarizes the overwritten CitySim parameters in this case study.

In the final step of data preparation, all heterogeneous datasets are cleaned, harmonized and integrated mostly by GIS operations (in FME platform [1]) and Python scripts. The end results are managed in database software PostgreSQL and some as GIS layers (in shapefile format [82]).

At the stage of data preparation, one may easily face a difficulty of assigning some inputs with deterministic values. This is because those parameters, for instance, room temperature setting, occupancy profile, etc., are highly fluctuating and exist stochastic variability. Another typical case is subjective

1 FME: https://www.safe.com/

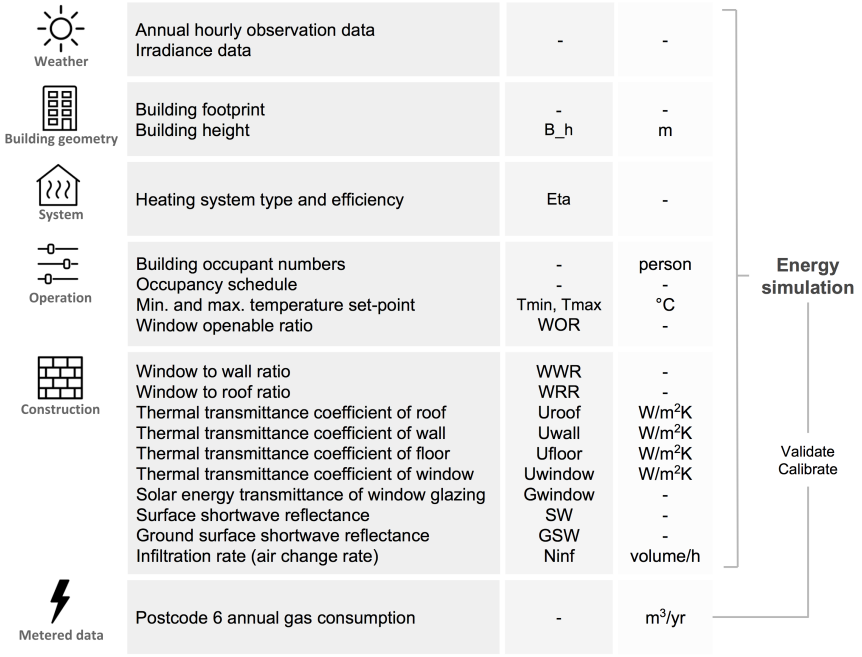


Figure 3.3: General data categories required by UBEM as well as the modified CitySim parameters in this case study. Parameter, symbol, and unit are listed from left to right.

uncertainty, for instance, window to wall ratio (WWR) of a building is a deterministic value; however, due to high data collection cost at an urban scale, this is often given as an averaged value of a certain building type from technical reports (e.g. residential, commercial).

Due to the above mentioned reasons, the main task in this step is not only assigning deterministic values for the simulation inputs, but also defining reasonable uncertainty ranges or probabilistic distributions to the relevant parameters. Chapter 4 will further discuss data preparation in details.

3.4 SENSITIVITY ANALYSIS

Depending on the modeling purposes, energy simulation can sometimes be very sophisticated and requires tens to hundreds of inputs to run [1, 22, 41]. Identifying the key parameters influencing heating demand calculation can be a critical step when available datasets and computational resources are constrained.

The main task in this step is to answer one of the research questions raised in section 1.1. *Given a number of simulation parameters with the associated uncertainty ranges, which ones are the key parameters affecting annual EUI calculation and which ones are minimal and even ignorable, according to the sensitivity analysis?*

Morris method, which is a commonly used statistical method in energy simulation study, will be used to answer this question considering its com-

putational efficiency. Morris method is a global sensitivity method to evaluate the influence of uncertain parameters over the whole parameter range. It applies Monte Carlo based evaluation by running CitySim multiple times (hundred to thousand times) on randomly selected input samples from the entire input space, see Figure 3.4. The modeler defines k parameters with the associated uncertainty ranges. Each parameter is normalized to $[0, 1]$ scale and divided into p levels. The input space Ω is now a k dimensional space with p levels in each dimension, consequently leads to p^k points in the entire input space, see Figure 3.5 for illustration.

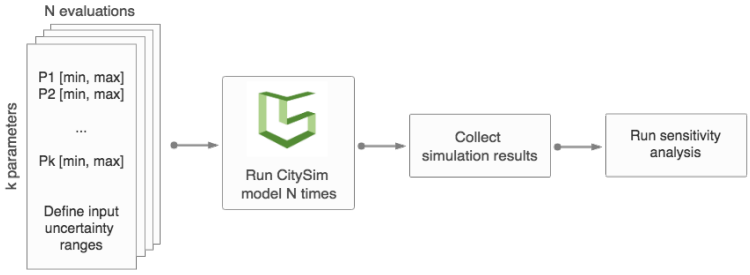


Figure 3.4: Procedure to perform Morris sensitivity analysis

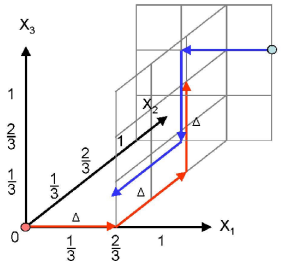


Figure 3.5: The whole input space Ω is divided into p levels in each dimension and leads to p^k points in the entire input space. The blue and red arrows represent two example trajectories. Retrieved from Prempraneerach et al. [83]

During sampling, the method randomly picks up an initial point in the input space and then moves along each dimension by the defined grid jump steps (Δ). Each sampling process forms a trajectory. Depending on parameter k and trajectory number (t), the total number of evaluation N is: $t \times (k + 1)$. Morris method gives two indices: μ_i^* and σ_i. The first measure, μ_i^*, is the sensitivity index to estimate the average elementary effect of a parameter, which has the following form shown in Equation 3.1:

$$\mu_i^* = \frac{1}{r} \sum_{t=1}^{r} |EE_{i,t}| \tag{3.1}$$

where

$$EE_i = \frac{y(x_i, ..., x_{i-1}, x_i + \Delta, ..., x_k) - y(x_i, ..., x_k)}{\Delta} \tag{3.2}$$

The second measure, σ_i, can be interpreted as a measure for non-linearity and parameter interaction effects.

$$\sigma_i = \sqrt{\frac{1}{r-1}\sum_{t=1}^{r}|EE_{i,t} - \mu_i^*|^2} \tag{3.3}$$

In this research, the sensitivity analysis is performed for a simple generic model: a cubic building with dimension of 13.5 meter in width, length and height standing along on the center of the ground surface (192m×176m) to run sensitivity analysis. In total, 14 parameters are evaluated, and the upper and lower bound value of the parameter uncertainty is defined respectively, see Table 4.8 for the complete input setting. Different trajectory numbers (t = 10, 30, 50, 100, 150) are being tested to ensure the stability of the results. More experimental setup and results will be discussed in full length in Chapter 5: CitySim characteristics and sensitivity analysis.

The end result of Morris method provides a qualitative measures to rank parameter importance. This gives the modeler a guidance for building archetype classification, and allows prioritizing calibration parameters.

3.5 BUILDING STOCK ARCHETYPE MODELING

As UBEM is dealing with more than hundreds or thousands of buildings, it is expensive and not yet available to have such comprehensive documentation of the built environment, or it might simply not be freely accessible. As a result, a common approach is to classify the building stock into different groups (archetypes), with each archetype shows similar heating demand pattern, based on the similar building properties such as construction period, main building function, and etc.

According to Dutch national reference home standard [84, 85], the example building is distinguished by construction period and dwelling type (detached house, semi-detached house, terrace house between, terrace house corner, gallery complex, apartment complex). In addition, the EPISCOPE and TABULA project, supported by the Intelligent Energy Europe program of the European Union, also develops residential building typologies for 13 European countries with the classification scheme mainly based on dwelling type (single-family house, terraced house, multi-family house, and apartment block), and construction period [6]. Based on these references, the research classifies the residential building stock of Amsterdam into 18 building archetypes, with six distinguished construction periods and three major dwelling types, see Figure 3.6. While Dutch standard and TABULA have six and four dwelling types respectively, it is aggregated into three main types in this research: single-family house (SFH), terrace house (TH) and multi-family house (MFH)

This deterministic archetype classification scheme now serves as a basis. As suggested by Cerezo et al. [21], assigning probabilistic distribution to key

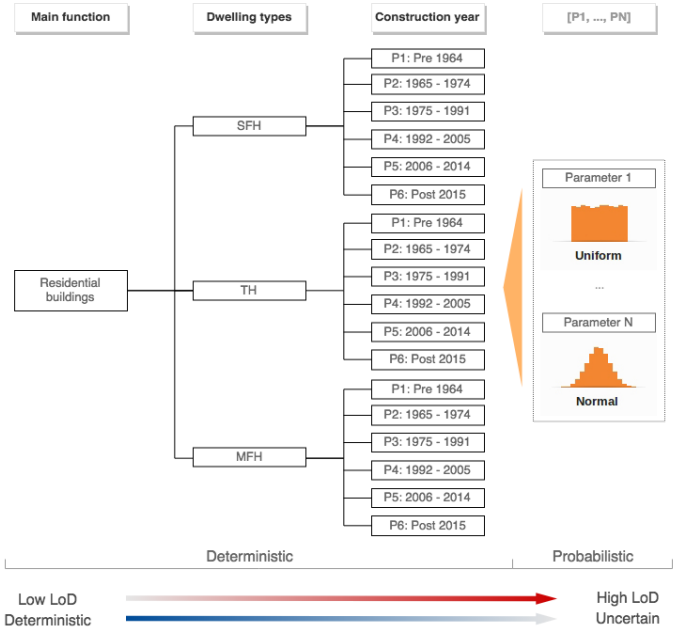


Figure 3.6: The basis archetype classification follows the existing schemes by separating the buildings stock into groups based on dwelling type and construction period. Key uncertain parameters can be modeled in a probabilistic way and add additional level of detail to the deterministic archetype model.

uncertain parameters can introduce an additional level of detail to the deterministically defined archetype. The results from the sensitivity analysis can help the modeler determines which parameters are the ideal candidates to be modeled in a probabilistic way and should go through Bayesian process to infer posterior. Among these selected uncertain parameters, some may be inherently stochastic such as thermostat setting and infiltration rate, while some are just unclear to the modeler. As the current CitySim version does not allow probabilistic inputs, and to simplify the entire model calibration process, it is assumed that the parameter with either kind of uncertainty can be both reduced to a single deterministic value, which is represented by the highest posterior probability. These optimal parameter values inferred from the posterior are building or postcode level specific and will be used to run the calibrated energy simulation.

Besides, the prior probability is now only modeled in a uniform distribution; nevertheless, non-uniform distribution is also possible given adequate prior knowledge of the specific parameter, see Figure 3.6.

As dwelling type information is not available in the existing collected datasets, additional classification task is required, and will be explained in more details in section 4.2.2: Building archetype classification in Chapter 4.

Although the archetype information is assigned to each building, due to the fact that spatial resolution of the gas consumption is at an aggregate postcode 6 level. As a result, the calibration can only be carried out at the same spatial resolution. Building archetype will be aggregated into postcode 6 level based on the dominant building volume share within the same postcode.

3.6 BAYESIAN INFERENCE AND MODEL CALIBRATION

Previous steps and works put essential data together and establish a foundation to run an urban scale heating demand simulation. At this stage, the modeler can carry out a baseline heating demand simulation given a deterministic city model (inputs with only deterministic values). This baseline result, however, is very likely to deviate from the metered data due to limited understanding of some key uncertain parameters. To minimize this gap, Bayes' theorem is adopted to perform model calibration.

Based on the sensitivity analysis (the result is presented in Chapter 5), minimum thermostat setting (*Tmin*) and infiltration rate (*Ninf*) are two of the most uncertain and influential parameters affecting annual heating demand calculation. The calibration framework assumes there exists a postcode specific representative value for each of these parameters, which is approximated by the highest posterior probability of *Tmin* and *Ninf* in the end of the training phase. Previously defined uncertainty ranges of these two parameters are divided into 5 sections, and it is assumed that each parameter prior has a uniform probability distribution. Theoretically, it is possible to select more parameters or dividing input space into finer sections for calibration. Nevertheless, it will significantly increase the number of parameter combinations and will thus require much more computational resources to complete the calibration task. In addition, it is also possible to model the parameter uncertainty in a non-uniform distribution way as long as adequate prior knowledge is given.

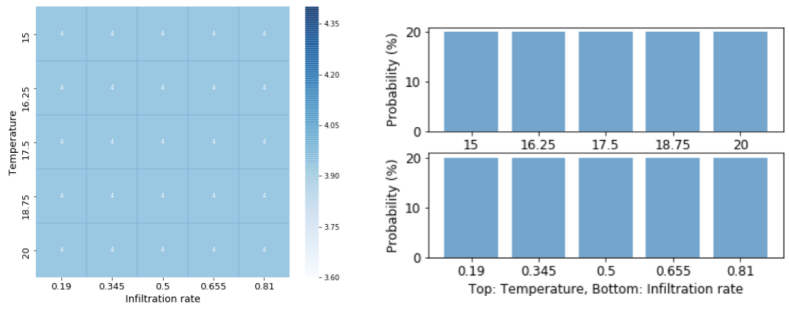


Figure 3.7: Prior probability distribution illustrated in a two dimensional grid. Each input combination has 4% prior probability given the uniform probability distribution assumption.

This setting generates 25 input combinations and the prior probability $P(\boldsymbol{\theta})$ of the joint distribution having equally 4% probability for each input combination (see Figure 3.7), which is denoted as uncertain input vector $\boldsymbol{\theta}$. Meanwhile, the remaining deterministic parameters are grouped into a vector $\boldsymbol{X}$. We can then think of simulation process in the following form:

$$\boldsymbol{y} = G(\boldsymbol{X}, \boldsymbol{\theta}) \tag{3.4}$$

where G is the simulation engine algorithm, and $\boldsymbol{y}$ is the vector contains energy simulation outputs, y_i, and $1 \leqslant i \leqslant 25$. Based on the Bayes' theorem

(Equation 3.5), observed annual gas consumption EUI g_{eui} can be used to infer posterior probability $P(\boldsymbol{\theta}|g_{eui})$.

$$P(\boldsymbol{\theta}|g_{eui}) = \frac{P(g_{eui}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(g_{eui})} \tag{3.5}$$

and

$$P(g_{eui}) = \int_{\boldsymbol{\theta}} P(g_{eui}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta}) d(\boldsymbol{\theta}) \tag{3.6}$$

Since there is no explicit form, it is assumed that the likelihood function $P(g_{eui}|\boldsymbol{\theta})$ can be described by Gaussian normal distribution function.

$$P(g_{eui}|\boldsymbol{\theta}) \approx P(g_{eui}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(g_{eui} - \mu)^2}{2\sigma^2}) \tag{3.7}$$

where g_{eui} is the individual postcode 6 EUI of the training year; μ is the simulated EUI of the corresponding postcode of the same year given the specific input combination $\boldsymbol{\theta}$; σ is estimated from the standard deviation of the measured consumptions of the same building archetype.

In the current implementation, annual metered data from year 2010 to 2015 are used to train the model (infer posterior probability of the uncertain parameters). As mentioned above, the prior probability $P(\boldsymbol{\theta})$ is initialized as uniform distribution. This iterative calibration process uses the posterior probability of the N year as a new prior of the $N+1$ year in the training phase. When the training phase is done, the input combination $\boldsymbol{\theta}$ with the highest posterior probability is selected as a calibrated input to run heating demand simulation of year 2016, and 2017.

3.7 VALIDATION

The whole framework applies annual metered EUI from year 2010 to 2015 to train the model. The annual metered EUI of the year 2016 and 2017 are used to validate the calibrated model. Absolute percentage error (PE) in the following form is used to assess and compare baseline and calibrated results:

$$PE = \left| \frac{EUI_{metered} - EUI_{sim}}{EUI_{metered}} \right| \times 100\% \tag{3.8}$$

Be noted that Energy Use Intensity, EUI, is calculated as the sum of gas consumption in kilowatt-hours per cubic meter (kWh/m^3) instead of common EUI reference (kWh/m^2), which is normalized to the square meter of the floor area. The main reason of applying this unit is considering the fact that accurate calculation of total floor area of the building is much more difficult than volume due to incomplete and missing values sometimes found in the floor area features in the address layer (BAG.verblijfsobject). On the contrary, building volume can be more easily and efficiently calculated based on

the 3D city model. Meanwhile, based on the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) Guideline 14-2002 recommendations, the allowable maximum percentage error of annual calibrated model is set to be 5%.

4 SEMANTIC CITY MODEL FROM HETEROGENEOUS DATASETS

Establishing a 3D semantic city model for Urban Building Energy Modeling involves several data collection, selection, cleaning, and integration processes. Due to large amount of information involved at this stage, it is exclusively discussed in this new chapter. It starts with data preparation and uncertainty quantification, and followed by multi-datasets integration and semantic city model enrichment.

4.1 DATA PREPARATION AND UNCERTAINTY QUANTIFICATION

This section aims at providing an overview of available data sources for the UBEM development applicable to the Netherlands, and more specifically, in Amsterdam. The collected data shall be able to serve as useful information for the relevant projects or applying to other simulation softwares such as City Energy Analyst for future UBEM development when necessary.

General data categories required by UBEM can be seen in Figure 3.3. Data preparation of meteorological data, building geometry, construction data, system, operation, and postcode 6 level gas consumption will be discussed in sequence. In the end of this section, deterministic values and uncertainty ranges of the model inputs are summarized in Table 4.6 to Table 4.8 respectively.

Collecting and carefully selecting heterogeneous datasets from available GIS layers, statistics data, technical reports, and energy data sources is sometimes a time-consuming and tedious process. Nevertheless, it is an indispensable step to define appropriate model inputs, or at least to help the modeler making a reasonable assumption about input uncertainty ranges. The work done in this section shall be able to answer one of the research questions: *What open-source GIS layers, statistics data, and technical datasets related to Amsterdam are available for UBEM development? and how to integrate these datasets in a sensible way to build up a simulation ready 3D city model?*

4.1.1 Meteorological data

Hourly time-series meteorological data of the study area is often required by energy simulation softwares in order to run a dynamic building energy simulation. CitySim takes the following meteorological parameters into account as listed in Table 4.1. Apart from Diffuse horizontal irradiance (G_Dh) and solar normal irradiance (G_Bn), historical observation records measured at the Schiphol station (which is approximately 10km from the city center of Amsterdam) is accessed from the Royal Dutch Meteorological Institute,

Koninklijk Nederlands Meteorologisch Instituut (KNMI)[1] and translated to CitySim compatible format. Meanwhile, because G_Dh and G_Bn records are not included in KNMI records, these values are retrieved from EnergyPlus weather data repository, Amsterdam epw.[2].

Overall, 8 years of meteorological data from year 2010 to 2017 are collected and translated into CitySim readable weather files.

Table 4.1: Meteorological parameters required by CitySim Pro

Abbreviation	Data	Unit	Source
d	Day		
m	Month		
h	Hour		
G_Dh	Diffuse horizontal irradiance	W/m^2	Amsterdam.epw
G_Bn	Beam (solar) normal irradiance	W/m^2	Amsterdam.epw
Ta	Air temperature	°C	KNMI
FF	Wind speed	m/s	KNMI
DD	Wind direction	°	KNMI
RH	Relative humidity	%	KNMI
RR	Precipitation	mm	KNMI
N	Nebulosity	*Octas*	KNMI

4.1.2 CitySim specific building geometry preparation

The building geometry preparation is subjected to the adopted simulation engine, the discussion in this section applies to CitySim specifically

In the Netherlands, the Basic Registration Addresses and Buildings data Basisregistratie Adressen en Gebouwen (BAG), managed by Kadaster[3] contains detailed, up to date, and georeferenced building (BAG.pand) and address (BAG.verblijfsobject) level data of the entire country (attributes contained in each layer can be seen Table A.1 and Table A.2 in Appendix). Attributes like year of construction, building function and building footprint, and etc. are included in BAG and can be freely accessed via Web Feature Service (WFS) maintained by Nationaal Georegister (NGR)[4].

Building geometry is modeled in level of detail 1 (LOD1) block model, the coarsest volumetric representation defined in Open Geospatial Consortium (OGC) CityGML standard [58]. LOD1 building model is usually acquired with extrusion from 2D building footprint with building height estimated from point cloud data [86]. 2D building footprint in the current implementation is a standard GIS file in shapefile format [82]. Point cloud data, Actueel Hoogtebestand Nederland (AHN_3), published by de waterschappen, de provincies en Rijkswaterstaat provides detailed and precise elevation data collected by airborne laser scanning techniques (or LiDAR-Light Detection And Ranging) and has average of eight height measurements per square meter covering the whole Netherlands. These datasets are an open-source and are accessed from PDOK[5].

1 KNMI: https://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi
2 EnergyPlus Weather (.epw): https://energyplus.net/weather/sources
3 Kadaster: https://www.kadaster.nl/bag
4 National Georegister: http://www.nationaalgeoregister.nl
5 PDOK: https://www.pdok.nl/nl/ahn3-downloads

Building reference height is estimated from the median height of the points positioned within the building footprint. This value approximately corresponds to the half of the height of the roof, which is considered to be the most suitable LOD1 height reference in terms of building volume computation [87].

Considering the level of detail of the collected data is mostly in building and postcode 6 level. In order to reduce simulation complexity, each building in the current implementation is modeled as a single thermal zone. This greatly simplifies geometric processing complexity, even though defining multiple thermal zones within a building is applicable in CitySim. To generate a valid 3D city model for thermal simulation in CitySim, the contiguous building wall areas have to be removed to avoid overestimating heating demand; otherwise, CitySim will assume these areas are touching the air and consequently overestimates heat exchanges. While many approaches may exist to perform adjacency detection and removal, this task can be most efficiently and effectively achieved by using the 3D modeling tool: Rhinoceros 3D in combination with its plugins: Grasshopper and Honeybee according to the author's experience.

Rhinoceros 3D is a CAD modeling software allows miscellaneous geometric processing, especially when it is used with its versatile Grasshopper plugin. Meanwhile, Honeybee and Ladybug are open-source, comprehensive environmental design tools built upon Grasshopper environment. These additional tools enable parametric scripting to carry out geometry preparation suitable for energy simulation, e.g. detecting adjacency and removal (Figure 4.2), decomposing surfaces by types, and so on. The procedure here starts with preparing GIS baselayer contains unique building IDs, 2D footprints and reference heights estimated from the point cloud data. The 2D building footprints have gone through topology check to remove polygon overlaps, slivers, merge gaps, and a generalization process to reduce geometry complexity. As CitySim applies detailed radiation model, simplified geometry (e.g. less triangulated faces) can significantly reduce computation time, see Figure 4.3.

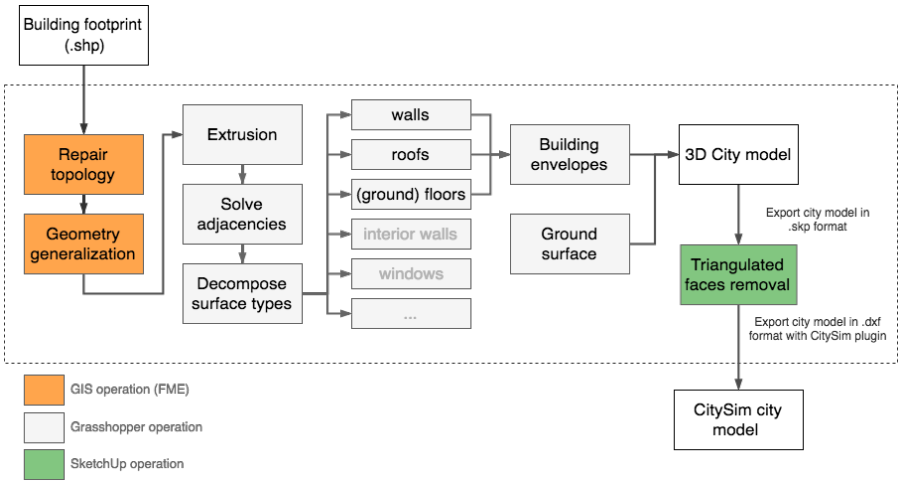


Figure 4.1: The diagram presents the generalized workflow for preparing a CitySim compatible 3D city model. The complete Grasshopper workflow can refer to Figure D.1 and Figure D.2 in Appendix.

LOD1 building extrusion and adjacency detection is followed by baselayer cleaning and generalization. The exported 3D city model has LOD1 buildings standing on the ground surface, where the ground surface area overlapped by the building footprints are removed. The conceptual workflow can be seen in Figure 4.1 and complete Grasshopper workflow is presented Figure D.1 and Figure D.2 in Appendix. After all these processes, the 3D city model now contains ground surface and LOD1 buildings with unique building IDs. CitySim can translate the exported AutoCAD DXF file into XML format, which is the basis for semantic city model enrichment in the later steps and will be discussed in section 4.2.

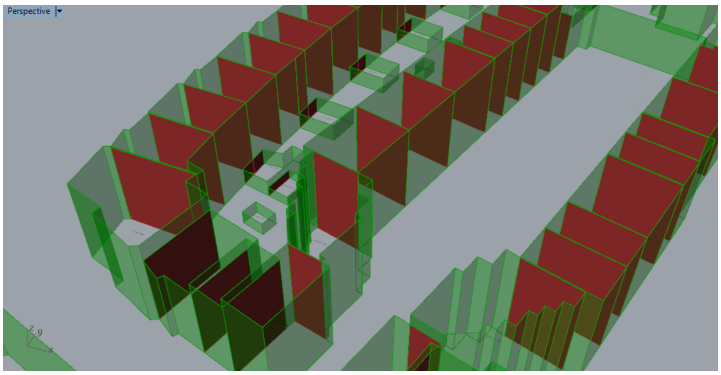


Figure 4.2: Red surfaces indicate overlaps between two neighboring buildings. In order to generate a valid 3D city model for running an energy simulation, these areas are removed by adopting Honeybee components in Grasshopper.

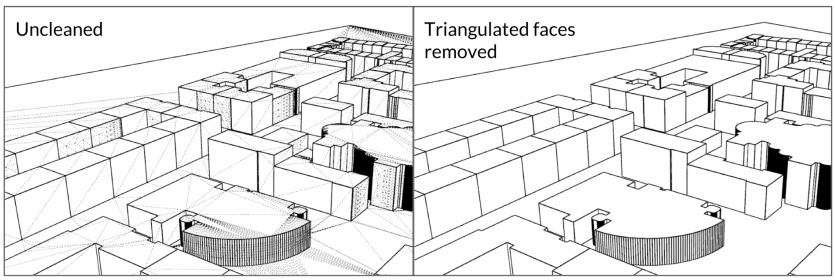


Figure 4.3: Removing triangulated faces in SketchUp to simplify the city model.

4.1.3 Construction data

Construction data refers to thermal transmittance coefficient (U-value) of roof, wall, floor and window (denoted as Uroof, Uwall, Ufloor, Uwindow); solar energy transmittance of window glazing (Gwindow); building infiltration rate (Ninf); window to wall ratio, window to roof ratio (WWR, WRR); (ground) surface shortwave reflectance (GSW, SW).

It is not difficult to imagine that certain building construction parameters, such as envelope U-values, can have significant influence on building energy

performance in moderate climate zone. The relation between specific U-value and heat flow (Φ) is illustrated in the follow form.

$$\Phi = A \times U \times (T_i - To) \tag{4.1}$$

where heat flow is represented by Φ [W]; A [m^2] is exposed surface area; U [W/m^2K] is thermal transmittance of the construction material (U-value), T_i [K] represents indoor temperature and T_o [K] represents outdoor temperature.

While it is impractical to collect deterministic construction values for each building, some construction parameters, especially U-values, are related to and can be inferred from building construction period. These year-dependent data can be found from the sources such as, European Building Stock Observatory[6], EPISCOPE and TABULA project web portal[7], Ecofys report [9], Sociale Huursector Audit en Evaluatie van Resultaten Energiebesparing (SHAERE) database [4], and so on.

Figure 4.4 summarizes and compares U-value statistics for each construction material (roof, wall, floor, and window) in different construction periods. Statistics values from these data sources are summarized in Appendix B.1, B.2, B.3. SHAERE database values are not included here as it requires additional authorization to access the database; however, large amount of records and diverse construction metadata (with 50-60% reporting rate of non-profit buildings in the Netherlands each year since 2010 [4]) might be a potential data source to be used for the UBEM development.

Lower U-value means less heat exchange between surfaces due to better insulation. Clear descending trends can be observed in all U-values of the construction materials, mainly due to technology advancement and energy performance regulation such as Energy Performance of Buildings Directive (EPBD) or Energy Efficiency Directive (EED). However, one may soon find it difficult to determine which data source values should be used as simulation inputs since some values are not consistent in some close construction periods. This could be caused by different samples and sampling approaches used by the data providers. This fact also reflects the inherent uncertainty within the simulation inputs and emphasizes the importance of model calibration as an essential UBEM step.

U-values accessed from EPISCOPE and TABULA project are chosen and used as baseline inputs in this project in the end, as U-values of the represented construction periods are consistent with the building classification scheme introduced in the Dutch national reference home standard [84, 85]. Another advantage of using this data source is because of its scalability. EPISCOPE and TABULA project is supported by the Intelligent Energy Europe program of the European Union with the aim of making the energy refurbishment processes in the European housing sector transparent and effective. In the precedent TABULA project, residential building typologies have been developed for 13 European countries with the classification scheme mainly based on dwelling type (single-family house, terrace house,

6 European Building Stock Observatory: https://ec.europa.eu/energy/en/data-analysis/building-stock-observatory

7 The TABULA WebTool: http://webtool.building-typology.eu/#bm

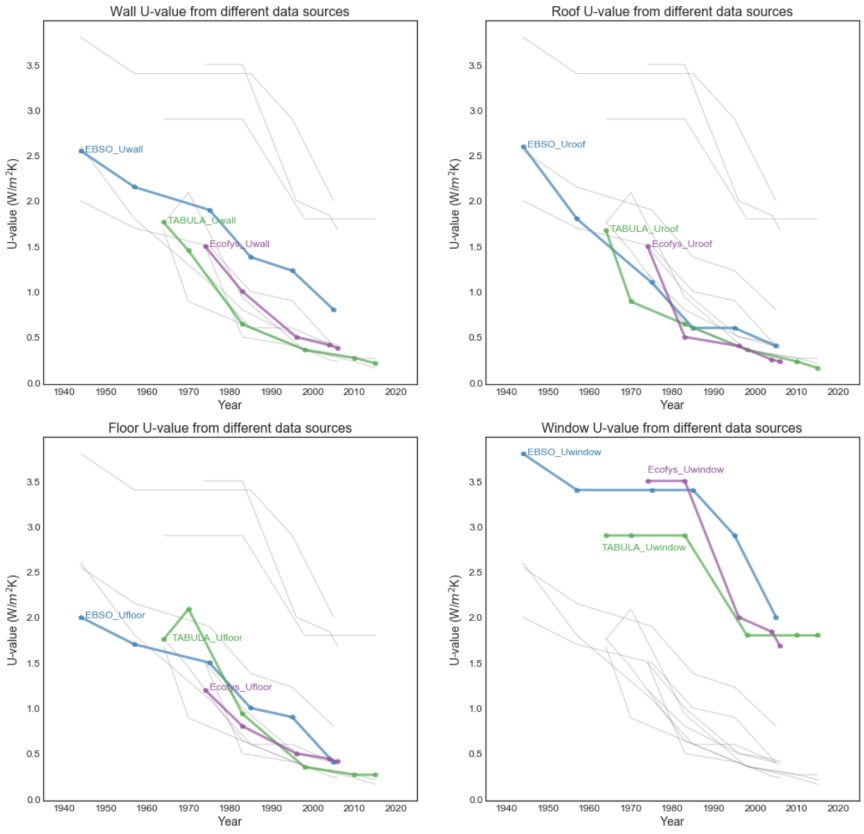


Figure 4.4: Comparison of multiple U-values of the construction materials collected from different sources

multi-family house, and apartment block), and construction period [6]. Each building archetype has an exemplary building with the associated construction details such as U-values of wall, roof, etc., and freely accessible via TABULA web portal[8]. This will undoubtedly speed up data preparation process if the developed methodology in this work is going to apply to the building stock of another European country.

While EPISCOPE and TABULA values are used as baseline inputs, the remained construction U-values are still helpful sources to be used for uncertainty quantification (e.g. defining upper and lower bound of the U-values within the similar construction periods).

At the stage of data collection, it is easily realized that U-value references are generally more accessible and comprehensive than the other construction parameters. This could be due to high data collection cost (e.g. building infiltration rate) [88], or the parameter itself has stochastic variability and thus can not be described deterministically. To cope with such situation, uncertainty assumptions (assumed uniform probabilistic distribution within the defined boundary) are made for these parameters and presented in Table 4.2.

8 The TABULA WebTool: http://webtool.building-typology.eu/#bm

Table 4.2: Defined uncertainty ranges of the construction parameters

Parameter	Baseline value	Probabilistic	Baseline source
Gwindow	0.76	U(0.30, 0.85)	[9]
Ninf	0.60	U(0.19, 0.81)	[89]
WWR	0.18(SFH), 0.24(MFH)	U(0.15, 0.45)	[9]
WRR	0.00	U(0.00, 0.15)	CitySim default
GSW	0.20	U(0.20, 0.50)	CitySim default
SW	0.20	U(0.20, 0.50)	CitySim default

1. SFH: Single family house; MFH: Multi-family house

Except for the self-explanatory parameters, *Ninf* here stands for building air volume change rate per hour in normal conditions (unit in h^{-1}), and taking both air infiltration through the envelope, air flow through the doors, windows, and casual or for ventilation purposes into account [88]. Most studies indicate air change rates are decreased with the construction periods of the buildings, and *Ninf* of older buildings is around 2 to 4 times higher than the more recent buildings [90, 91]. Additionally, cold countries (Estonia, Finland, Canada, Sweden and Norway, with *Ninf* reported around 0.19 to 0.37 h^{-1}) are found to have lower average *Ninf* than more temperate climate countries (Belgium, England, Greece, USA, and Italy, with *Ninf* be around 0.44 to 0.81 h^{-1}) according to [91], which can be used to define uncertainty range of *Ninf* as listed in Table 4.2. All baseline input values and uncertainty ranges of the construction parameters are summarized in Table 4.6, Table 4.7, and Table 4.8.

4.1.4 System data

In the Netherlands, majority of heating demand (80%) is fulfilled by combustion of natural gas through gas boilers or cogeneration plants, this percentage will gradually decrease to 77% by 2020 and to 71% by 2030 according to the existing policy [92]. Table 4.3 provides an insight into the share of household heating system type and efficiency of the non-profit buildings accessed from SHAERE database [4]. It clearly shows that the condensing boiler with efficiency $\geqslant$ 0.95 is the dominant installation in the existing building stock. In addition, to access detailed heating system type and efficiency per building is challenging due to privacy concerns. As a result, in the current implementation, heating system diversity and uncertainty is less addressed and treated as homogeneously distributed in the building stock when compared with other simulation inputs.

Table 4.3: Distribution of heating system types in the Netherlands. Retrieved from Filippidou [4]

Type of heating system	Frequency	Percentage
Condensing boiler ($\eta \geq$0.95)	930127	74%
Improved non-condensing boiler (η=0.80-0.90)	178557	14%
Condensing boiler (η=0.90-0.925)	42026	3%
Gas/oil stove	40548	3%
Conventional boiler ($\eta <$0.80)	29973	2%
Condensing boiler (η=0.925-0.95)	19595	2%
Heat pump	16722	1%
μCHP	2751	0%
Electric stove	484	0%
Total	1260783	100%

4.1.5 Operation data

In the context of this research, operation parameters refer to number of occupants per building, occupancy schedule (profile), minimum set-point temperature of the heating system (Tmin, when indoor temperature is lower than Tmin, heating system starts to operate), and window openable ratio (WOR).

Needless to say, human behavior has significant impacts on instantaneous energy consumption particularly when it is examined with high temporal resolution (e.g. hourly, and daily energy consumption). For instance, the presence of occupants directly influence how the building would be operated (e.g. Heating, ventilation and air conditioning (HVAC)), as well as internal heat gain (use of appliances) [93, 94, 95]. As occupant behavior is one of the biggest uncertainty in energy modeling [47], how to properly model occupancy schedule and human behavior in great details remains a popular yet challenging topic, and it is currently out of the scope of this research. Meanwhile, the UBEM model developed in this case study is validated with and calibrated by annual gas consumption data at aggregated postcode 6 level, which gives us more confidence to assume that the fluctuation in heating demand consumption is mainly affected by building characteristics, while occupancy uncertainty can be approximated by a standardized profile in annual scale, see Figure 4.5.

Operation data preparation first starts with estimating the number of occupant per residential building. Since occupant number of each building is not directly available in the building layer (BAG.pand), it is derived from the census data at postcode 6 level (PC6) published by the Central Bureau of Statistics, Centraal Bureau voor de Statistiek (CBS) [9] of year 2008 to 2010 (the PC6 census data of year 2012-2014 requires additional cost to access), and denoted as OCC_{PC6}. Detailed attributes included in CBS census data can be viewed in Appendix Table A.5. Following the equation below, we can estimate the number of occupants within building i, denoted as $OCC_{building,i}$

$$OCC_{building,i} = OCC_{PC6} \times \frac{A_i}{A_{pc6}} \tag{4.2}$$

where A_i stands for the summed residential surface areas within the building i; A_{pc6} represents the total residential surface areas of the belonged postcode. $OCC_{building,i}$ is rounded to the nearest integer before running simulation.

The set-point temperature of the heating system is another influential yet uncertain input on heating demand calculation. According to the study of Leidelmeijer and Grieken [5], Tigchelaar and Leidelmeijer [96], which is partially based on the WoON survey [10] data in the Netherlands, different temperature setting profiles throughout different time of a day and weekend are observed and summarized in Table 4.4. Although different temperature

9 CBS: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische%20data/gegevens-per-postcode

10 WoOn: https://www.wooninfo.nl/

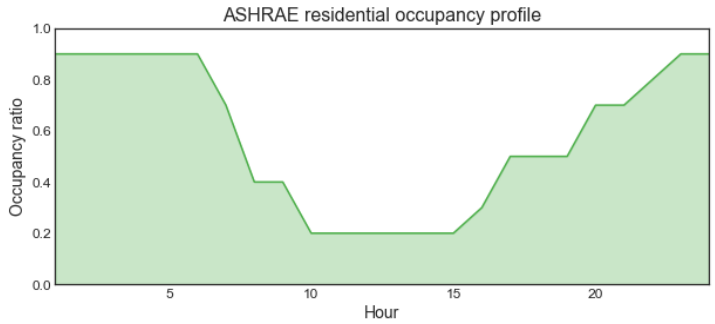


Figure 4.5: Standardized occupancy profile of the residential building provided by ASHRAE Standard 90.1. 2004 is adopted for energy simulation in the project.

setting profiles exist, CitySim only allows a single value (T minimum and T maximum respectively) as set-point inputs per building. As a consequence, the averaged value 18.38 °C is adopted as a baseline input. Nevertheless, this profile gives us a better insight and helps to quantify uncertainty range of the set-point temperature in average Dutch households by taking the minimum and maximum temperature as lower and upper bounds from weighted average (approximately between 15 to 20 °C).

Window openable ratio (WOR) is barely found in any technical report or literature reviewed by the author, baseline value and uncertainty range of this parameter is thus made with an intuitive but rational assumption, see Table 4.6 and Table 4.8. Sensitivity analysis result presented in Chapter 5 confirms that the parameter influence on heating demand calculation is minimal and almost ignorable.

Table 4.4: Different thermostat setting (°C) profiles in Dutch households. Adapted from Leidelmeijer and Grieken [5]

	Morning	Afternoon	Evening	Weekend	Share
profile 1	15.2	17.4	14.1	16.2	4%
profile 2	18.4	18.8	15.6	18.5	16%
profile 3	19.7	20.2	15.2	20.0	35%
profile 4	19.6	20.0	11.6	19.8	8%
profile 5	14.9	20.2	14.7	20.1	11%
profile 6	20.9	21.2	20.4	21.1	5%
profile 7	21.6	22.0	15.5	21.7	20%

4.1.6 Energy consumption data

Metered gas consumption data is not only used to validate simulation results but also applied to model calibration. Because of privacy concern, high spatial-temporal resolution metered data, for instance, sub-hourly energy data per building, is often not readily available. As a consequence, the research focuses on using open-source data and starts with energy datasets released by CBS and Liander (local grid operator and energy provider), which both open annual energy consumption (gas and electricity) data at aggregated postcode 6 level.

Brief analysis on both datasets quickly leads to the final decision of adopting Liander dataset for this project. This is because, CBS dataset has only 2014 consumption data freely accessible through WFS [11] ; while 10 years of energy consumption records starts from 2008 to 2017 is made available via Liander open data portal [12]. The energy consumption records managed in Liander dataset also come with much detailed metadata. For instance, delivery percentage (network supply minus customer self-generation), network status, and so on, which gives user a better control over data quality. Detailed attributes included in CBS and Liander energy data can be seen in Table A.3 and Table A.4 in Appendix.

4.1.7 Inputs summary

Running a bottom-up heating demand simulation requires diverse inputs from different data categories as illustrated in Figure 3.3. The main efforts spent in this section is trying to source deterministic values as baseline inputs and to quantify uncertainty ranges for these parameters based on non-exhaustive literature and technical reports. These numerical values are summarized in the following tables to provide a general overview.

Table 4.5: Data collections applied to the Amsterdam UBEM development

Data category	Dataset	Data period	Remark	Source
Weather	Annual hourly observation	2010-2017	[1]	KNMI
	Irradiance	1982-1999	[2]	Amsterdam.epw
Geometry	Building footprint	up to date	-	BAG(WFS)
	Building height	2014-2019	[3]	AHN3
System	Heating system statistics	2010-2014	[4]	[4]
Operation	PC6 population	2008-2010	[5]	CBS
	Occupancy schedule	-	[6]	ASHRAE
	Temperature set-point	- 2005	-	[5]
	Window openable ratio	-	[7]	-
Construction	See Table 4.6	-	-	-
Energy	PC6 annual gas consumption	2010-2017	[8]	Liander

1. The measurement is made at Schiphol meteorological station, which is approximately 10 km away from Amsterdam center.
2. This data is based on IWEC data in Amsterdam and managed in EPW (EnergyPlus Weather) format. IWEC data is derived from long term observation sometimes up to 18 years (1982-1999 for most stations). Details refer to https://energyplus.net/weather/sources#IWEC
3. AHN3 point cloud data collection period is scheduled to 2019 http://ahn.maps.arcgis.com/apps/Cascade/index.html?appid=75245be5e0384d47856d2b912fc1b7ed
4. Statistical distribution data of the heating system type and efficiency collected from the non-profit building stock database in the Netherlands.
5. Only 2008-2010 is made freely accessible. 2012-2014 data is made available at cost.
6. Standardized residential occupancy profile
7. Barely found reliable data source, rational assumption is made for this parameter
8. Liander energy data is better than the CBS data quantitatively and qualitatively (several years of consumption data and much detailed metadata)

11 2014 Gas and electricity supply: http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/9a3ae78f-9d4d-4cd9-9aec-10c16f8597f4?tab=relations

12 Liander open data: https://www.liander.nl/over-liander/innovatie/open-data

Table 4.6: Deterministic (baseline) values of the simulation inputs

Parameters	Symbol	Unit	Baseline	Source [a]
Building construction parameters				
Window to wall ratio	**WWR**	-	0.21	[9]
Window to roof ratio	**WRR**	-	0.00	-
Thermal transmittance coefficient of roof	**Uroof**	W/m^2K	Table4.7	TABULA
Thermal transmittance coefficient of wall	**Uwall**	W/m^2K	Table4.7	TABULA
Thermal transmittance coefficient of floor	**Ufloor**	W/m^2K	Table4.7	TABULA
Thermal transmittance coefficient of window	**Uwindow**	W/m^2K	Table4.7	TABULA
Solar energy transmittance of window glazing	**Gwindow**	-	0.76	[9]
Surface shortwave reflectance	**SW**	-	0.20	Default
Ground surface shortwave reflectance	**GSW**	-	0.20	Default
Infiltration rate (air change rate)	**Ninf**	$Volume/h$	0.60	[89]
Operation parameters				
Minimum set-point temperature	**Tmin**	°C	18.38	[5]
Window openable ratio	**WOR**	-	0.25	-
System parameter				
Heating system efficiency	**Eta**	-	0.95	[4]

a If no reliable source is being found, a generic assumption is made.

Table 4.7: Baseline U-values in different construction periods [6]

.

Parameters	Pre 1965	1965-1974	1975-1991	1992-2005	2006-2014	Post 2014
Uroof	1.68	0.89	0.64	0.36	0.23	0.16
Uwall	1.76	1.45	0.64	0.36	0.27	0.21
Ufloor	1.75	2.09	0.94	0.35	0.27	0.27
Uwindow	2.90	2.90	2.90	1.80	1.80	1.80

Table 4.8: Defined uncertainty ranges of the simulation inputs

Parameters	Symbol	Unit	Uncertainty	Source [a]
Building construction parameters				
Window to wall ratio	**WWR**	-	U(0.15-0.45)	[9]
Window to roof ratio	**WRR**	-	U(0.00-0.15)	-
Thermal transmittance coefficient of roof	**Uroof**	W/m^2K	U(0.16-2.60)	[6, 8, 9]
Thermal transmittance coefficient of wall	**Uwall**	W/m^2K	U(0.21-2.55)	[6, 8, 9]
Thermal transmittance coefficient of floor	**Ufloor**	W/m^2K	U(0.27-2.09)	[6, 8, 9]
Thermal transmittance coefficient of window	**Uwindow**	W/m^2K	U(1.68-3.80)	[6, 8, 9]
Solar energy transmittance of window glazing	**Gwindow**	-	U(0.30-0.85)	-
Surface shortwave reflectance	**SW**	-	U(0.20-0.50)	-
Ground surface shortwave reflectance	**GSW**	-	U(0.20-0.50)	-
Infiltration rate (air change rate)	**Ninf**	$Volume/h$	U(0.19-0.81)	[91]
Operation parameters				
Minimum set-point temperature	**Tmin**	°C	U(15.0-20.0)	[5]
Window openable ratio	**WOR**	-	U(0.00-0.35)	-
System parameter				
Heating system efficiency	**Eta**	-	U(0.80-0.95)	[4]
Geometry parameter				
Building height uncertainty	**B_h**	-	U(0.90-1.10)	[87]

a If no reliable source is being found, a generic assumption is made.

4.2 MULTI-DATASETS INTEGRATION AND SEMANTIC CITY MODEL ENRICHMENT

This section expatiates the technical details of integrating heterogeneous datasets and explains what selection constraints are applied to the specific datasets.

The entire data cleaning, integration and city model enrichment process involves several steps and is accomplished by combining multiple tools (environments) in use, such as GIS processing in FME software, Geometry processing in Rhinoceros 3D together with Grasshopper plugin, PostgreSQL database and Python programming.

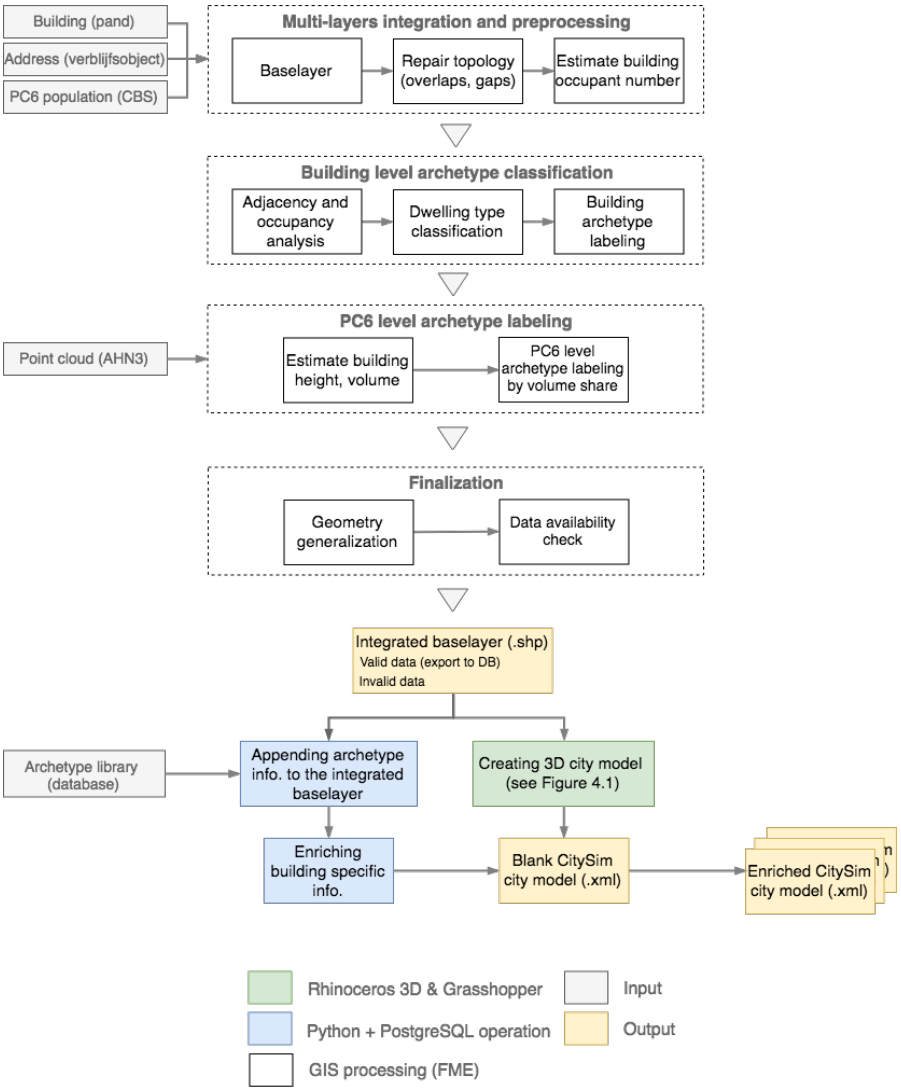


Figure 4.6: Generalized workflow presents some of the key steps in the data integration process. Be noted that the sequence of some steps is adjustable, but this is how the workflow is implemented in the project. The complete FME script used for data integration can refer to Figure C.1 in Appendix.

Figure 4.6 summarizes the key steps from data integration to city model enrichment. Figure 4.7 further illustrates the key attributes, constraints, and relationships of these datasets. GIS processing in FME platform is extensively carried out here and can be broken down into four main stages and explained in the following sequence.

4.2.1 Multi-layers integration and preprocessing

The first step starts with combining building layer (BAG.pand) and address layer (BAG.verblijfsobject) together by the shared attribute, building ID (pandidentificatie), as the join key. The output layer is now called *baselayer* and additional information will be stacked upon this layer subsequently. Although the BAG layers are regularly maintained by Kadaster, topology repairing process is performed to ensure valid building footprint polygon (without overlaps, slivers, gaps, and etc.) for later processes. As occupant

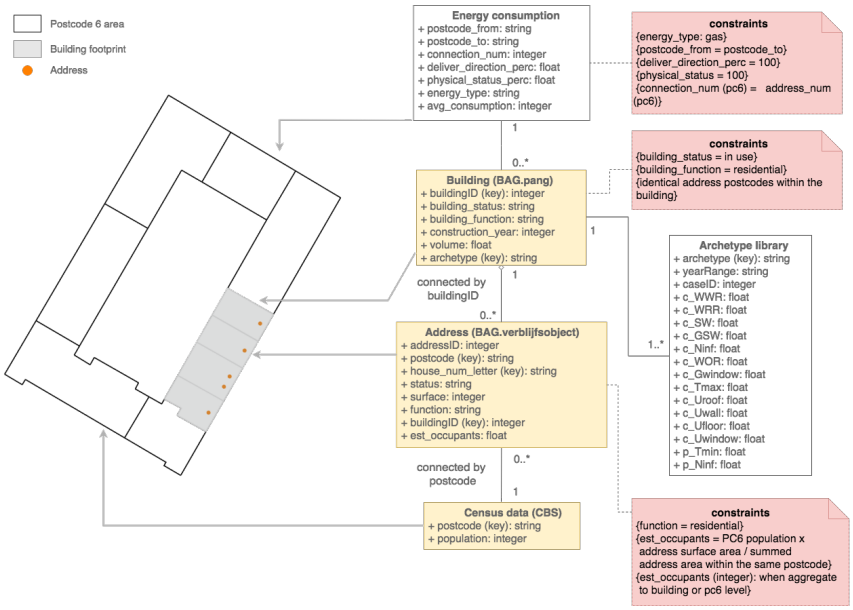


Figure 4.7: UML diagram presents the key attributes, feature constraints, and relationships of these heterogeneous datasets. The spatial resolution of each data is illustrated on the left. The *integrated baselayer* is integrated from the UML classes colored in yellow, and the data filtering rules are colored in red.

number per building is not available as an open-source data, it is estimated from postcode 6 population statistics from CBS dataset, and calculation method is introduced in section 4.1.5.

Be noted that specific constraints applied to the corresponding datasets is illustrated in Figure 4.7. These constraints and data filtering rules, for instance, confining building function to be residential building, energy supply minus energy return is 100%, namely, no self-generated energy, and so on, are applied to ensure heterogeneous datasets are integrated in a sensible way.

4.2.2 Building archetype classification

The goal at this stage is to classify the building stock into three dwelling types: single-family house (SFH), terrace house (TH), and multi-family house (MFH) as well as 6 construction periods as introduced in section 4.2.2. Construction periods can be easily derived from the construction year stored in the baselayer. Meanwhile, GIS operations and topology relation analysis are performed to infer three dwelling types. The applied classification rule (Figure 4.9) can distinguish 6 dwelling types, it is then aggregated into three main dwelling types for the final use.

Dwelling type classification based on geometry features introduced in the literatures [18, 97] has also been experimented. Two geometric indices, shape coefficient C_s (exposed surface to volume ratio, also denoted as S/V

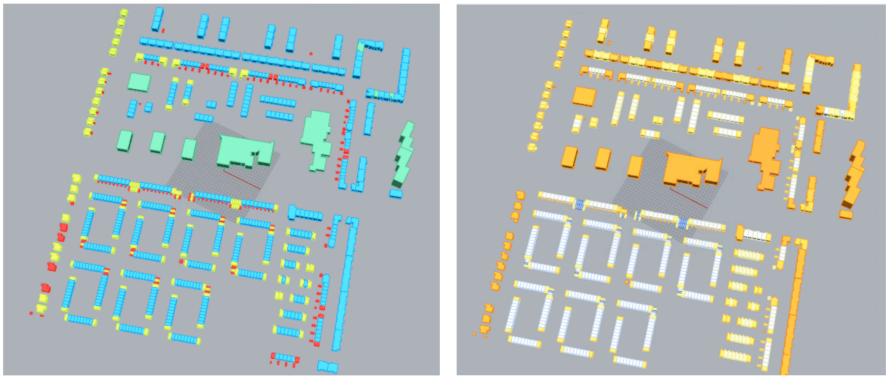


Figure 4.8: Left: three S/V ratios are applied to the city model and try to distinguish four dwelling types, which are presented by different colors. However, misclassification is currently unavoidable by this approach. Right: relative compactness shown in color ramp. From isolated (orange), intermediate (yellow) to compact (blue).

ratio) and relative compactness (RV) of the building with the following forms are applied to the 3D city model.

$$C_s = \frac{S}{V} \tag{4.3}$$

where S represents total surface areas in contact with outside air and V represents building volume. C_s indicates a relation of area per unit volume. This index is used by Torabi Moghadam et al. [18] to classify the building stock of the city Settimo Torinese located in northwestern Italy with the following classification rules: Detached house (DH): $S/V \geqslant 0.8m^{-1}$; Terrace house (TH): $0.6 \leqslant S/V \leqslant 0.8m^{-1}$; Multi-family house (MFH): $0.4 \leqslant S/V \leqslant 0.6m^{-1}$; Apartment block (AB): $S/V \leqslant 0.4m^{-1}$. Relative compactness has the following form:

$$RV = \frac{6V^{2/3}}{S} \tag{4.4}$$

where V represents building volume and S is surface area of the building.

In this experiment, it is found that the previously mentioned classification values can not be directly apply to the building stock of Amsterdam. After some trial and error, it is found that although defining classification boundaries with alternative values can classify building stock into different dwelling types in general, misclassification becomes another issue to be solved with this approach, see Figure 4.8. As a consequence, this approach is not adopted in the current implementation for the dwelling type classification task.

4.2.3 Postcode 6 level archetype labeling

Since the availability of gas consumption data is at postcode 6 level, the inferred building archetype will be aggregated to the corresponding level at this stage. The aggregated archetype is assigned based on the dominant building volume share within the corresponding postcode. To estimate

LOD1 building volume, AHN3 point cloud data is used to calculate reference building height and then calculate building volume. The detail of this calculation is introduced in section 4.1.2.

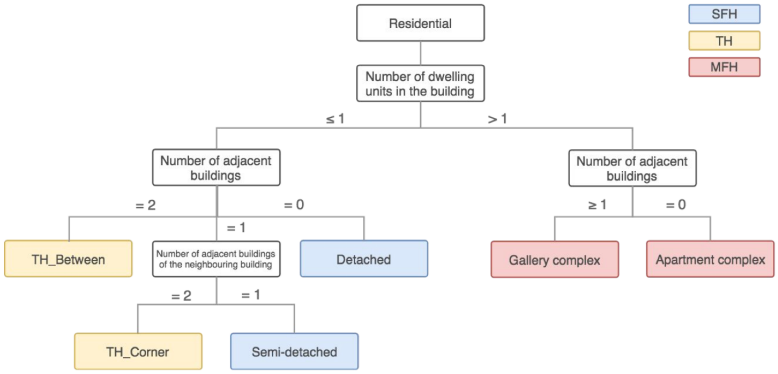


Figure 4.9: Classification rules based on topology relation and the number of address per building is implemented for the dwelling type classification.

4.2.4 Finalization

It is found that simulation time is significantly affected by geometry complexity as a result of the detailed radiation model implemented in CitySim. To speed up computation time, building footprint geometry is simplified and generalized (using Douglas-Peucker algorithm with tolerance set to be 0.5 and the shared boundaries are preserved) through GIS operations in FME at this stage. For instance, reducing the number of nodes in a curved boundary and thus decreases the number of faces of the LOD1 model. Furthermore, data availability check is carried out after going through these steps to examine whether the building has valid and complete data to carry out heating demand simulation or not (constraints applied to the integrated data can be seen in Figure 4.7). Valid data is duplicated, exported and managed in PostgreSQL database. Meanwhile, *integrated baselayer* is exported as a .SHP file for later use.

4.2.5 Enriching CitySim 3D city model

This is the final step of data integration before executing heating demand simulation. At this stage, the intermediate result: *integrated baselayer* will be used to update CitySim 3D city model, which has XML format with default values. On the other hand, archetype library information, which contains archetype specific deterministic and probabilistic values (uncertainty ranges) defined in section 4.1.7, will be appended to valid building data via PostgreSQL operation. As CitySim 3D city model is managed in XML format, it can be easily parsed and overwritten by the developed Python script (see Figure 4.10). The main task done by the Python script is to overwrite the CitySim default values with the defined deterministic or probabilistic inputs for each building. Because two simulation inputs, $Ninf$ and $Tmin$ are defined probabilistically in the current implementation (explanation see Chapter 5), this generates 25 enriched CitySim city model files instead of a

single file. These enriched city models will go through heating demand simulation iteratively and using metered data to carry out model calibration.

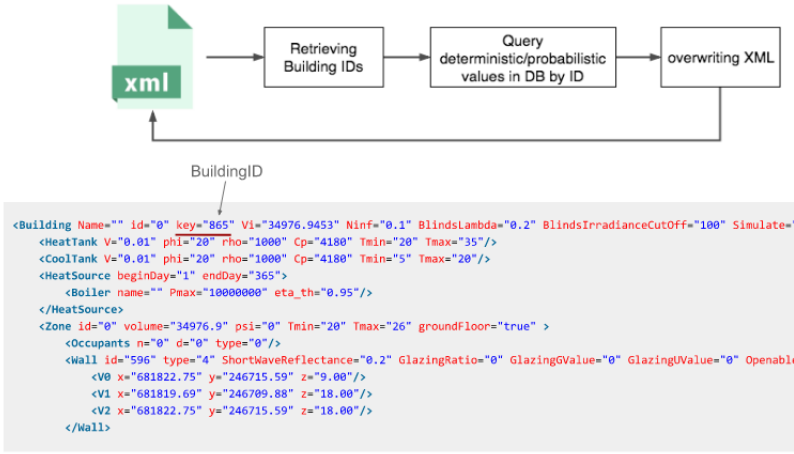


Figure 4.10: The simplified overwriting logic implemented in the Python script. A piece of CitySim XML code is presented

5 CITYSIM CHARACTERISTICS AND SENSITIVITY ANALYSIS

The main goal of this chapter is to further investigate the added-value and appropriateness of adopting CitySim as a simulation engine in this case study. Meanwhile, the sensitivity analysis is performed in order to get the insight of input importance ranking and to be used as a reference for probabilistic archetype classification.

5.1 CITYSIM ENERGY SIMULATION CHARACTERISTICS

A large paragraph from section 2.3 to section 2.4 in Chapter 2 introduces what factors should be considered when choosing a simulation engine for the project. In essence, the requirements for this research are: i) the simulation engine should be able to take energy exchange in the urban environment into account as urban microclimate is found to have significant influences on urban thermal condition and consequently affect energy consumption [15, 25, 27]; ii) computation efficiency; iii) the scale of the simulation matches the level of detail of the available data; iv) easily programmable. According to these requirements, advanced energy simulation engine such as EnergyPlus, TRNSYS, and etc. become an expensive option as they usually require very detailed inputs to perform sophisticated simulation which will require more computational resources and time compared with CitySim. Additionally, these traditional simulation engines are designed mainly for building scale simulation and consequently the environmental interactions between buildings are less addressed.

CitySim becomes the option because it is the one most fulfills these requirements and also the operation is familiar by the author. The experiment below addresses specifically about the first requirement, and trying to understand what is the added-value of adopting CitySim for this UBEM project.

5.1.1 Experiment set-up

The first experiment tries to understand how radiation effect (shadowing, multiple reflections) in the urban environmental causes annual heating demand differences under different building layouts. Rectangular building(s) with dimension of length: 110.5m; width: 13.5m; height: 13.5m are placed in parallel with four different aspect ratios (building height/street canyon width): 0, 0.5, 1, 2, see Figure 5.1. This set-up roughly corresponds to multi-family building size, and is adopted from the work done by [26], where the same building dimension and layouts are applied, but tested with TRNSYS.

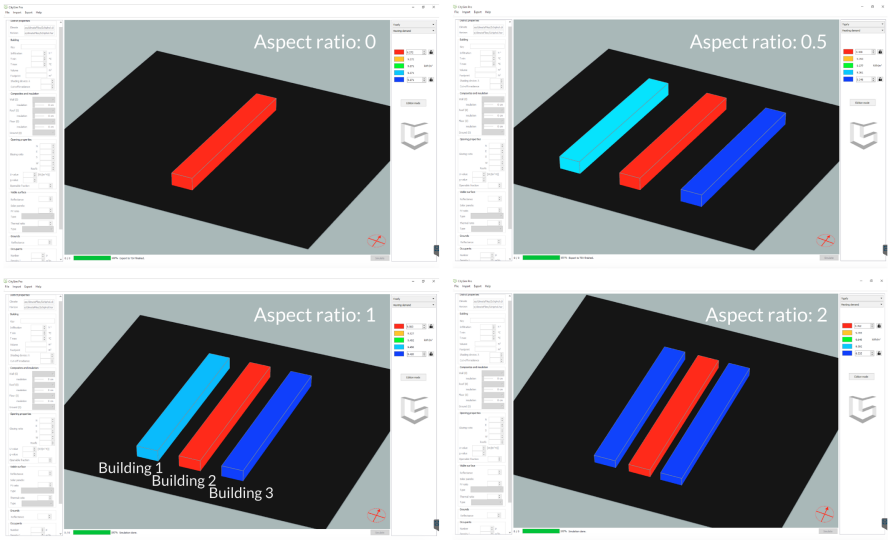


Figure 5.1: Annual heating demand results with four different aspect ratio set-ups. Heating demand from high to low is colored in red to blue.

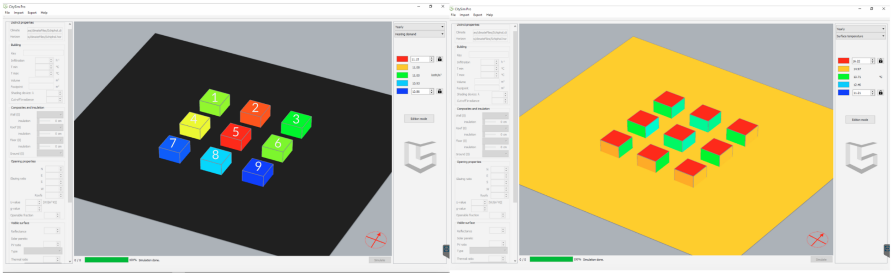


Figure 5.2: Left: Annual heating demand differences caused by building layout. The one in the center has the highest heating demand because it is most shaded by the surrounding buildings. Right: The surface temperature calculation made with the radiation model of CitySim.

The Schiphol weather record is used. The construction, operation, and system parameters of the buildings are all kept as default, so we can observe the effect of building geometry (related to radiation) on annual heating demand calculation.

The goal of the second experiment is similar to the first one, but the building dimension (13.5m in all dimensions) as well as layouts are changed, see Figure 5.2. The S/V ratio (surface to volume ratio) of the rectangular building in the first experiment is approximately: 0.24, while the S/V ratio of the cubic building is: 0.37 in the second experiment.

5.1.2 Results and discussion

The results of these two experiments can be best illustrated in Figure 5.3, 5.4. From both figures, one can easily observe that compact built environment leads to slightly higher heating demand, mainly because of less external radiation received by the building due to geometry shadowing. One may question that this result is a bit counterintuitive, for instance, the Urban Heat Island effect happens in the densely populated urban environment,

where the temperature can be few Celsius higher than the suburban area [25, 27]. The author believes that this result can be explained by the fact that the current CitySim version applies a radiation model to calculate environmental energy exchange; however, this does not include energy exchange caused via convection. To be able to simulate convective energy exchange, it often requires coupling with Computational Fluid Dynamics (CFD) software to run more advanced simulations [54], which is out of the scope of this research.

On the other hand, a positive correlation between heating energy demand and building surface to volume ratio (S/V) can be observed when comparing two results. This observation is not too surprising as heat flow (Φ) is proportional to exposed surface area (A) as shown in Equation 4.1. When exposed surface area per unit volume (S/V) is higher and all the other conditions are fixed, one can expect higher heat flow per unit volume, namely, more heat loss and consequently higher heating demand.

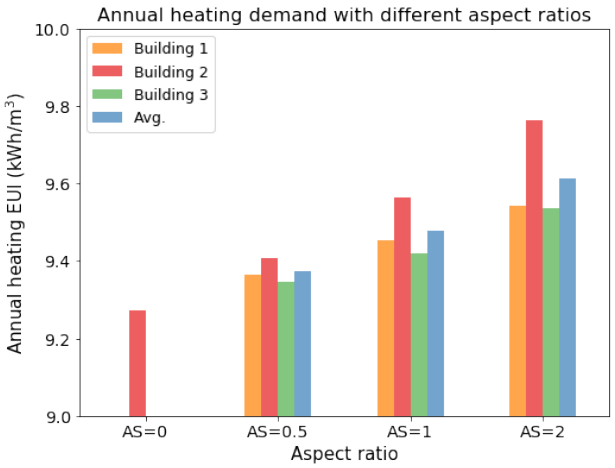


Figure 5.3: The first experiment reveals that high aspect ratio layout (more compact and more shading for each building) leads to higher annual heating demand than low aspect layout.

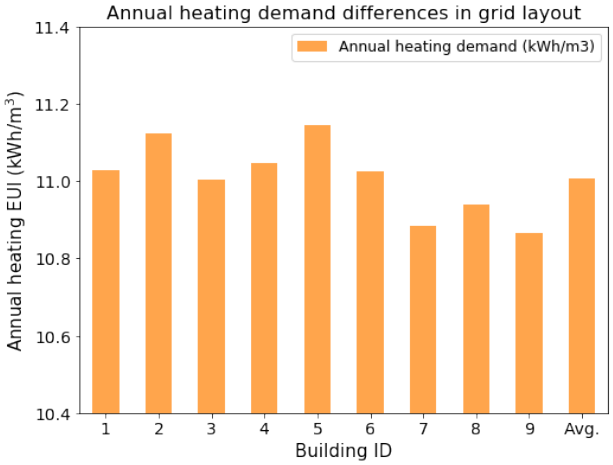


Figure 5.4: Similarly, the shaded buildings have higher heating demand. Furthermore, larger S/V ratio leads to higher heating demand compared with the first experiment. Building numbering refer to Figure 5.1, 5.2

To briefly conclude, two experiments present the simulation characteristics CitySim have when it is applied to the urban environment. Although without a convection model, the environmental radiation model takes building geometry characteristics and shadowing into account. Nevertheless, the normalized annual heating demand differences are rather modest in both cases.

5.2 SENSITIVITY ANALYSIS

Chapter 4 introduces the data requirements for UBEM development of Amsterdam, where the deterministic parameter values and uncertainty ranges are defined and presented. The work of this section is trying to answer one of the research questions: *Given a number of parameters with the defined uncertainty ranges to describe building characteristics, which ones are the key parameters affecting heating demand calculation and which ones are minimal and even ignorable, according to the sensitivity analysis?*

The adopted sensitivity analysis method is Morris method, due to its capability to give parameter importance ranking with less computational resources, compared with other methods such as Sobol method. The theoretical details of sensitivity analysis are extensively discussed in section 2.7: Overview of sensitivity and uncertainty analysis, and the execution details are briefly discussed in section 3.4: Sensitivity analysis of the simulation engine. The sensitivity analysis is carried out by the Python script with external library, SALib[1] [98], which contains commonly used sensitivity analysis methods. The following paragraphs will expand on these backgrounds and start with the experiment set-up.

5.2.1 Sensitivity analysis set-up

A stand-alone cubic building with $13.5m$ in all dimensions positioned in the center of the ground surface is used for the sensitivity analysis. Input uncertainty ranges used in the sensitivity analysis is summarized in Table 4.8. Each parameter is divided into 10 levels (p), which leads to 10^{14} input combinations (Ω). The grid jump size (Δ) is set to be 2. Since Morris method applies Monte Carlo based evaluation by running simulation multiple times with each input sampled from the whole input space (Ω) using trajectory sampling (see Figure 3.5), it is likely to give a more stable result (parameter ranking) given more evaluations. The total number of evaluation (N) is defined as $N = t \times (k + 1)$, namely, $N \propto t$. To understand how many trajectories (t) can roughly give a stable result, sensitivity analysis is carried out 5 times with $t = [10, 30, 50, 100, 150]$. Based on this finding, a stable threshold t is adopted to run the further sensitivity analysis, which should give a more reliable parameter importance ranking result.

5.2.2 Sensitivity analysis results and discussion

The relative parameter importance ranking based on the sensitivity analysis with different number of trajectories (t) is illustrated in Figure 5.5. Param-

1 SALib https://salib.readthedocs.io/en/latest/

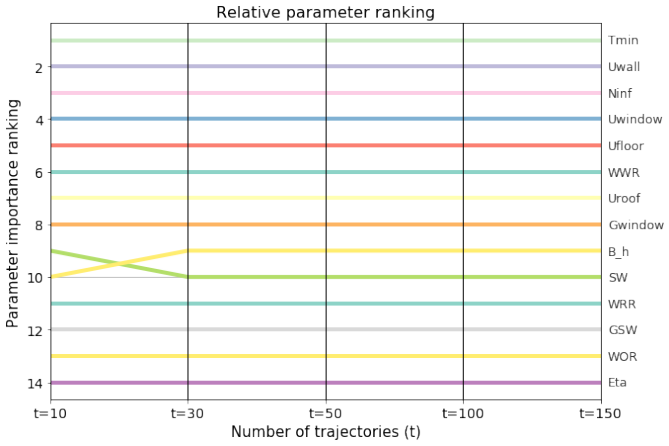


Figure 5.5: Relative parameter importance ranking tested with different number of sampling trajectories (t). A stable result is expected when $t \geqslant 50$.

eter ranking is slightly changed when t is increased from 10 to 30. The ranking remains consistent afterwards, and concludes that the sensitivity analysis results with $t \geqslant 30$ are more stable. It is believed that setting $t = 50$ could give a more reliable result and still under acceptable computation time in this case.

Based on this finding, Figure 5.6 presents the sensitivity analysis result with $t = 50$. Parameter sensitivity estimated from the average elementary effect μ^* as well as parameter interaction effect measured by σ are presented. It is found that under the defined uncertainty ranges, operation parameter: minimum thermostat setting $Tmin$ has the most significant effect on annual heating demand calculation, and followed by construction parameters $Uwall$ and $Ninf$. Furthermore, the result clearly shows that construction parameters, especially U-values of wall, floor, window, and roof have significant to moderate influence on annual heating demand calculation, while building surface shortwave reflectance (SW), ground surface shortwave reflectance (GSW), and window openable ratio (WOR) are insignificant. The ignorable result of SW and GSW might be caused by the experiment setting that there is no surrounding building in this case.

Meanwhile, as building reference height is estimated from AHN3 point cloud data, it is possible that building height (volume) estimation uncertainty exists. The experiment allows the building reference height to vary between 90% to 110% of the estimated height, and according to sensitivity analysis result, one can conclude that building height estimation uncertainty within this range has a comparatively insignificant influence on annual heating demand calculation. Finally, it was surprisingly found that heating system efficiency (Eta) shows no influence in this case. Further examinations revealed that the maximum thermal power of the system is also required by CitySim in order to give sensible results. Nevertheless, detailed system-related information is difficult to collect and thus prior information remains ambiguous, the rest experiments are thus carried out with this sensitivity analysis result and believed it is more than sufficient.

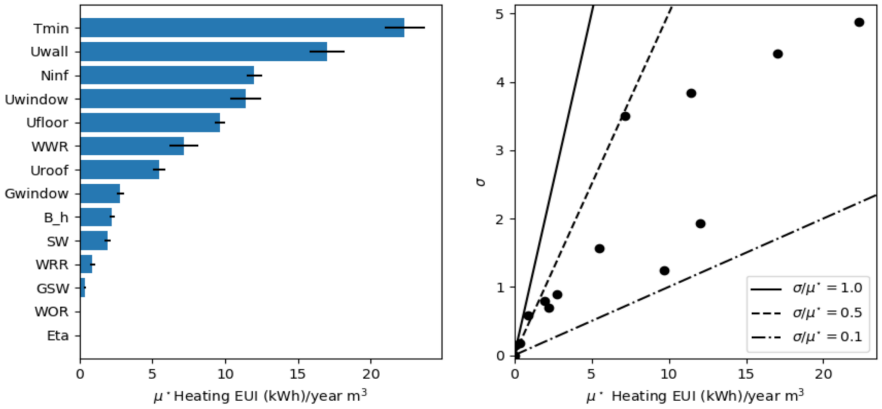


Figure 5.6: Sensitivity analysis result on a simple cubic building with $k = 14$, $p = 10$, $\Delta = 2$, and $t = 50$. Diagram on the left shows parameter sensitivity estimated from the average elementary effect μ^*; diagram on the right shows additional parameter interaction by the measure of σ.

5.2.3 Probabilistic archetype modeling

Based on the sensitivity analysis result, the basis archetype classification, which classifies the building stock into 18 archetypes based on dwelling types and construction periods (refer Figure 3.6), can be further expanded. Two key uncertain parameters: $Tmin$ and $Ninf$, divided into 5 sections respectively and with assumed uniform prior probability distribution, are used to expand the existing archetype, see Figure 5.7. When metered gas consumption data is available, and by using Bayesian inference approach, the prior assumption with respect to these parameters will be updated to posterior. The optimal parameter combination of the postcode with the highest posterior probability is selected and used to run calibrated simulation deterministically.

It is theoretically possible to include more key uncertain parameters in a probabilistic archetype modeling and prepare for the later calibration process. However, this will significantly increase the required number of calibrating simulations (N_c), which can be written as: $N_c = P_s^K$, where P_s here represents the number of levels of the sensitive parameter, and K indicates how many key parameters to be calibrated. In addition, calibration process is an over-specified and under-determined problem. In this project, six years of annual gas consumption data of at least 84 respective postcodes is used for calibration. Applying comparatively few and low-resolution measurements to calibrate the complex model could easily lead to over-fitting issues, this explains why only two key uncertain parameters are currently selected for probabilistic archetype modeling and calibration.

Despite the above archetype model is the final implemented solution, one idea was trying to investigate whether it is possible to group and model U-value parameters (Uroof, Uwall, Ufloor, and Uwindow; these parameters are sensitive and related to construction periods, see Figure 4.4) as an effective construction year, which takes building renovation into account, and expecting Bayesian inference can update building construction year into effective construction year. Nevertheless, based on the fact that renovation

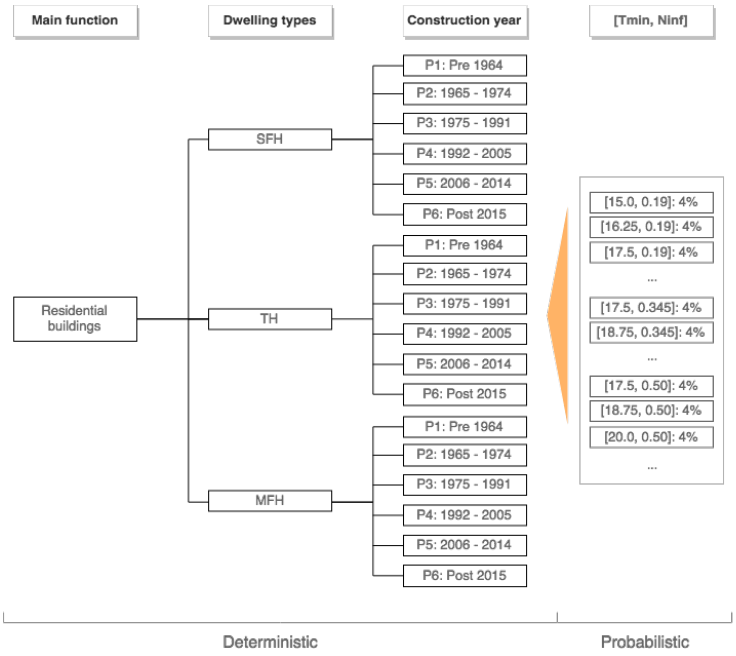


Figure 5.7: *Tmin* and *Ninf* are two key uncertain parameters according to the sensitivity analysis. This information enables the modeler to expand each of the existing deterministic archetype with a probabilistic definition.

mostly takes place on respective building components (Figure 5.8) rather a whole building renovation [4], this idea was not implemented in the end.

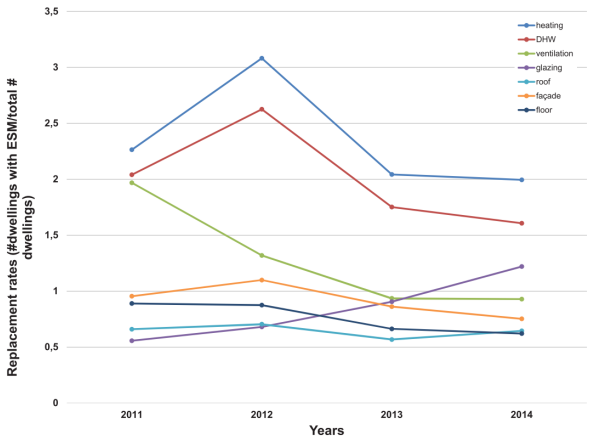


Figure 5.8: Replacement rate based on Energy Saving Measures ESMs according to SHAERE database. From Filippidou [4]

6 BAYESIAN INFERENCE, MODEL CALIBRATION AND VALIDATION

This chapter explains how Bayesian inference is applied to infer posterior probability distribution of the key uncertain parameters and how model calibration is performed given annual gas consumption data. The following paragraph starts with the recap of the Bayesian calibration framework, which is followed by the model validation, and discussion.

6.1 BAYESIAN INFERENCE AND MODEL CALIBRATION

6.1.1 Bayesian inference and model calibration set-up

All of the previous works conclude to two important outputs, deterministic parameters describing building characteristics, which are grouped into a vector $\boldsymbol{X}$, and two key uncertain parameters, $Tmin$ and $Ninf$, denoted as a vector $\boldsymbol{\theta}$. $Tmin$ and $Ninf$ is divided into 5 levels with uniform prior probability distribution respectively and this leads to totally 25 input combinations. This can be interpreted as that $(\boldsymbol{X}, \boldsymbol{\theta})$ together generates 25 models for each valid postcode. We are particularly interested in which input combination, $\boldsymbol{\theta}$, is likely to give the most reasonable simulation result when compared with the metered data g_{eui}, namely, posterior probability $P(\boldsymbol{\theta}|g_{eui})$. Posterior probability $P(\boldsymbol{\theta}|g_{eui})$ can be calculated according to Bayes' theorem (Equation 6.1), where $P(g_{eui}|\boldsymbol{\theta})$ is the likelihood function and evaluated in the form shown in Equation 6.3; $P(\boldsymbol{\theta})$ is prior probability and initialized as a uniform distribution, see Figure 3.7.

$$P(\boldsymbol{\theta}|g_{eui}) = \frac{P(g_{eui}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(g_{eui})} \tag{6.1}$$

and

$$P(g_{eui}) = \int_{\boldsymbol{\theta}} P(g_{eui}|\boldsymbol{\theta}) \times P(\boldsymbol{\theta}) d(\boldsymbol{\theta}) \tag{6.2}$$

Since there is no explicit form, it is assumed that the likelihood function $P(g_{eui}|\boldsymbol{\theta})$ can be described by Gaussian normal distribution function.

$$P(g_{eui}|\boldsymbol{\theta}) \approx P(g_{eui}; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(g_{eui} - \mu)^2}{2\sigma^2}) \tag{6.3}$$

where g_{eui} is the individual postcode 6 EUI (kWh/m^3) of the training year; μ is the simulated EUI of the corresponding postcode of the same year given the specific input combination $\boldsymbol{\theta}$; σ is estimated from the standard deviation

of the measured consumptions of the same building archetype. σ is further scaled down by a factor of 5 because most archetype groups have numerous sample points, which could overestimate the EUI variability of the respective archetype. When the sample number of the specific archetype is not enough, for instance, archetype SFH_1964 has only one sample in this case, a constant value 5 (kWh/m^3) is adopted for σ.

It should be noted that using Gaussian normal distribution to describe the likelihood function is an assumption. This might fall short to accurately describe the real likelihood distribution for the respective input combination. Further investigations in future work are required.

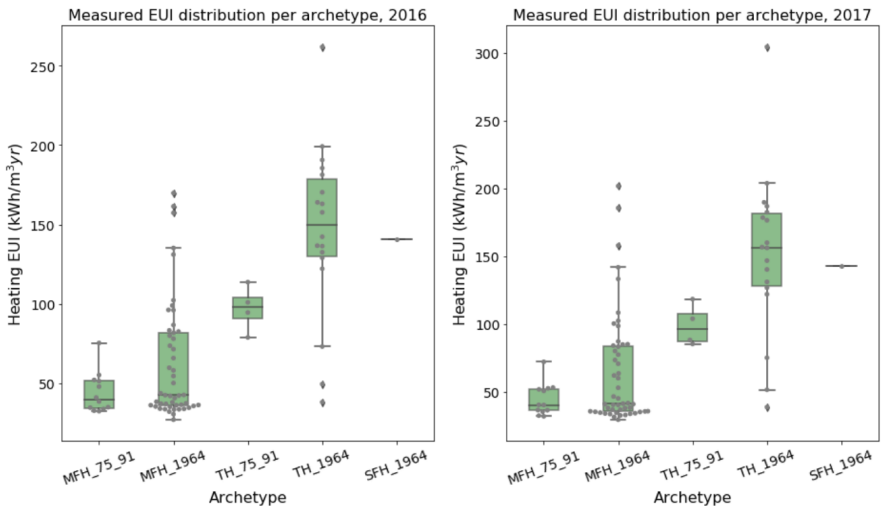


Figure 6.1: Measured postcode level EUI distribution per archetype included in the test site. 84 and 85 samples available in 2016, 2017 respectively.

Annual metered data from year 2010 to 2015 are used to train the model. As mentioned, the prior probability $P(\boldsymbol{\theta})$ is initialized as a uniform distribution. In the training phase, the iterative calibration process uses the posterior probability of the N year as a new prior of the $N+1$ year. If the measurement EUI deviates more than ± 1.5 EUI of the possible simulation results, the posterior of the year will not be updated since all the simulation results generated by the given input combinations can poorly explain this measurement. Namely, such observation deviation could only be explained when other input variabilities are taken into account.

When the training phase is complete, the input combination $\boldsymbol{\theta}$ with the highest posterior probability is selected as a calibrated input to run heating demand simulation deterministically given 2016, and 2017 weather data. The validation step will compare the baseline simulation result as well as the calibrated simulation result to the measurement data, see Figure 6.3.

6.1.2 Results, validation, and discussion

An example Bayesian inference process of the postcode 1094 SH is visualized in Figure 6.2. It clearly shows that after given six years of training data (annual gas consumption data per postcode), the posterior probability of the

few specific input combinations stand out from the rest. It also shows how the marginal posterior probabilities of the parameters $Tmin$ and $Ninf$ are updated in the right bar chart in Figure 6.2. The optimal input combination with the highest posterior probability will be used to run a calibrated heating demand simulation deterministically. Such Bayesian inference process is applied to all valid postcodes (fulfill data requirement) in the test area by the Python script.

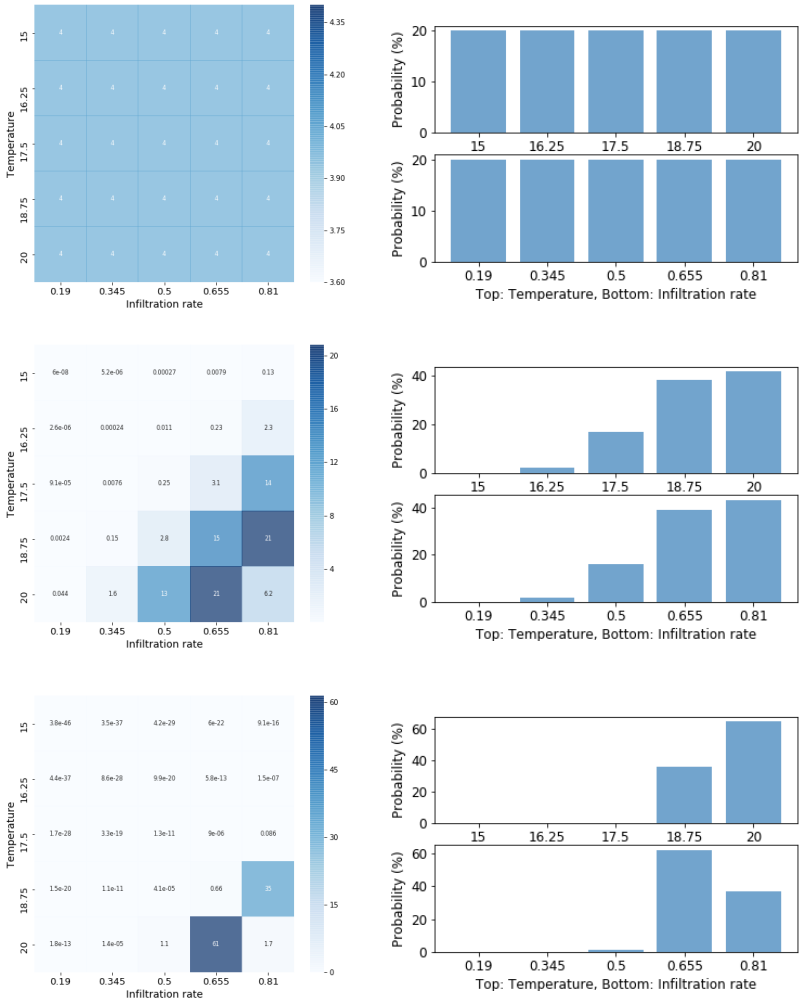


Figure 6.2: Visualizing the Bayesian inference process of the postcode 1094 SH over the course of the training phase. Joint probability distribution is visualized in a 2D grid in left while marginal probability distributions of $Tmin$ and $Ninf$ are presented in right. From top to down: the prior probability, the posterior of 2010, and the posterior of 2015.

Figure 6.3 presents the comparison between EUI of the Liander measurement, the baseline simulation and the calibrated simulation result of approximately 85 postcodes in the test area in year 2016, 2017 (Figure 3.1). The improvement is clearly visible. It is found that the baseline simulation tends to over estimate heating demand in this case, and the calibrated simulation successfully reduces simulation performance gaps. To compare in a quantitative way, absolute percentage error (see Equation 6.4) of the baseline and

the calibrated result of the partial postcodes in the test site are presented in Figure 6.4.

$$PE = \left| \frac{EUI_{metered} - EUI_{sim}}{EUI_{metered}} \right| \times 100\% \tag{6.4}$$

Based on the validation results, the effectiveness of applying Bayesian inference framework to perform model calibration is clear. Although some calibrated postcodes do not completely fulfill ASHRAE standard: allowable maximum 5 percentage error of annual calibrated model (see Figure 6.4), the simulation performance gaps are in general significantly reduced.

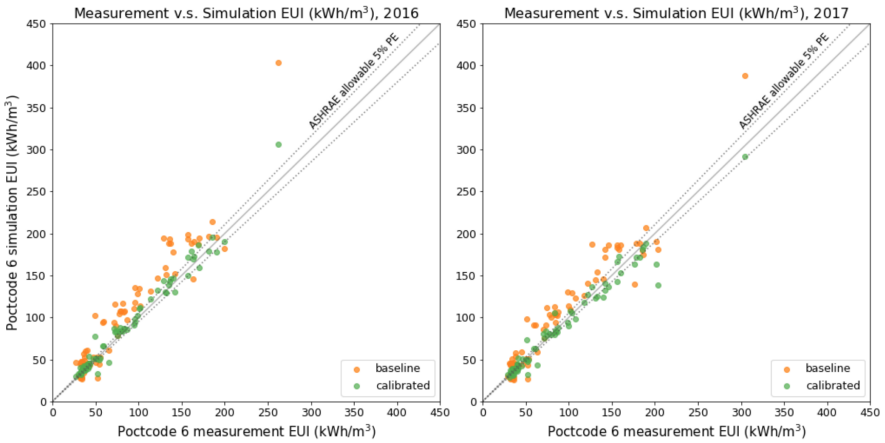


Figure 6.3: Comparison between the 85 postcode 6 measurements and simulation results. The dash lines indicate 5% of allowable percentage error suggested by the ASHRAE standard. In the case of the uncommonly high energy use shown in upper right of the plots, it could be flagged as a priority postcode (buildings) for further examination.

For the purpose of model calibration, the Bayesian calibration framework applied in this case study has picked up an optimal deterministic input combination for each valid postcode from the refined posterior distributions. To understand whether the posterior distributions show statistically significant pattern for the respective archetype (a group of postcodes in this case). The marginal posterior probability distributions of $Tmin$ and $Ninf$ are visualized on each row of the heat maps shown in Figure 6.5, and postcodes with the same archetype are placed together. Visual inspection reveals that it might be able to say that the distributions of $Tmin$ and $Ninf$ of the particular archetypes are further reduced, for instance, $Tmin$ and $Ninf$ of MFH_75_91 and TH_1964. However, due to the limited number of sample of the certain archetypes, it is in general difficult to conclude that the posterior distributions of $Tmin$ and $Ninf$ show statistically significant pattern or cluster for most archetypes at the moment

Nevertheless, when other area of interests have poor quality inputs or no measured data is available for the model calibration, the posterior distributions of the particular archetypes might still assist the modeler to form a more sensible prior assumption. These posterior distributions are also

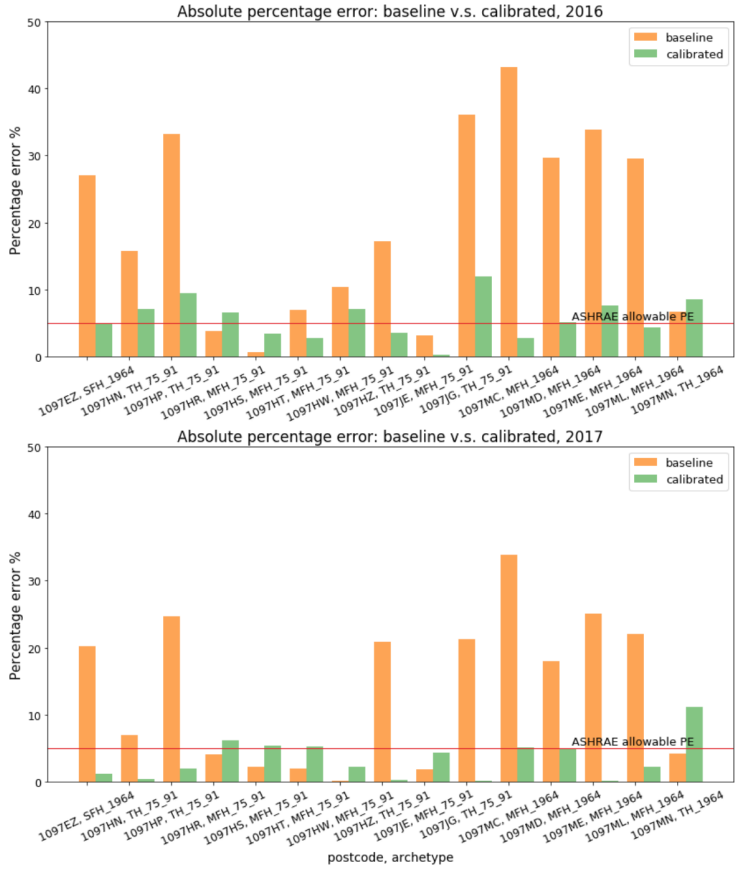


Figure 6.4: Baseline and calibrated simulation results of the partial postcodes presented in absolute percentage error. The red horizontal line indicates the ASHRAE allowable maximum percentage error (5%) of annual calibrated model.

valuable when probabilistic energy modeling is needed. For instance, when using UBEM to predict retrofit measures saving, probabilistic simulation results can serve as confidence intervals around the saving amounts [51].

Computation time is another aspect worth discussion. According to the current implementation, calibrating two parameters requires 25 simulation runs in total, and this process is iteratively conducted 6 times given 6 years of annual gas consumption data. This is 150 simulations in total. Depending on the partitioned city model scale of the test site and geometry complexity (Figure 6.6), small scale simulation with 226 buildings can take 27 hours and the largest scale simulation with 1363 buildings requires almost 10 days to complete the expensive training phase on a personal computer with 4 cores 3.60GHz processor and 16GB RAM. Although CitySim simulation scales well in general, such time constraint could still become an obstacle to developing UBEM into an interactive platform for decision support applications.

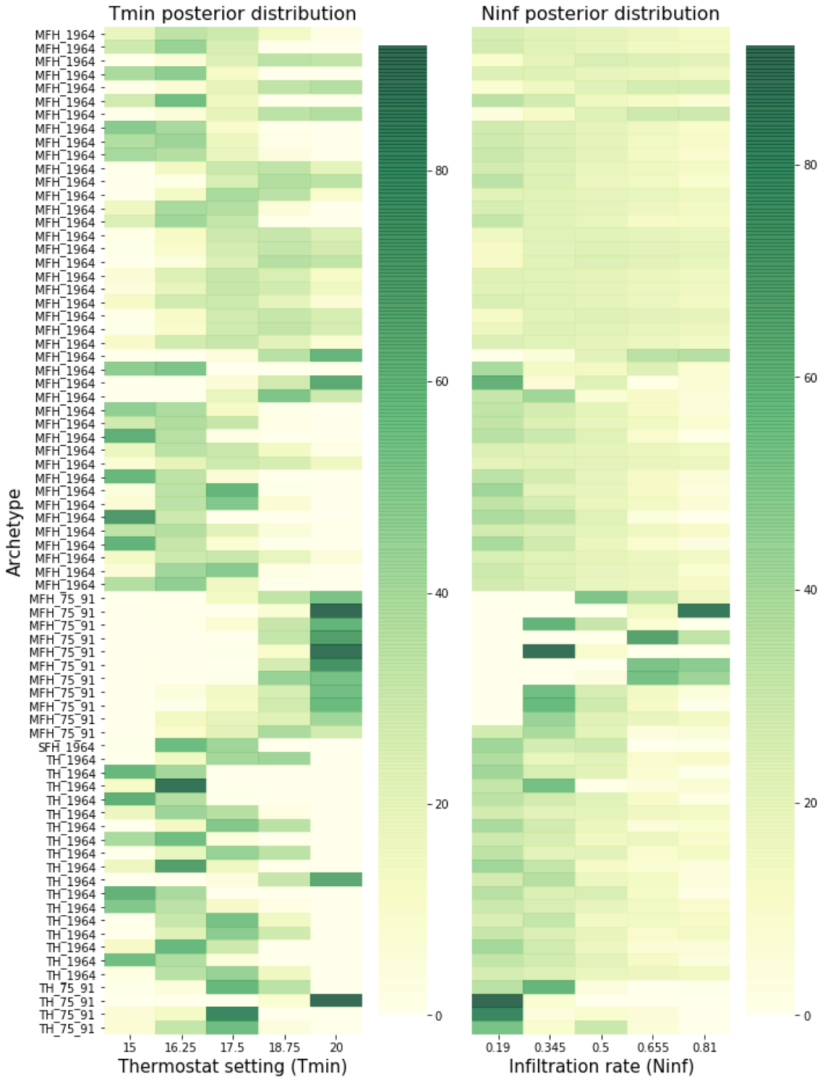


Figure 6.5: Marginal posterior distribution of $Tmin$ and $Ninf$ visualized in heat map. Each row represents one postcode, and the postcodes with same archetype are place together. Color intensity represents probability (%).

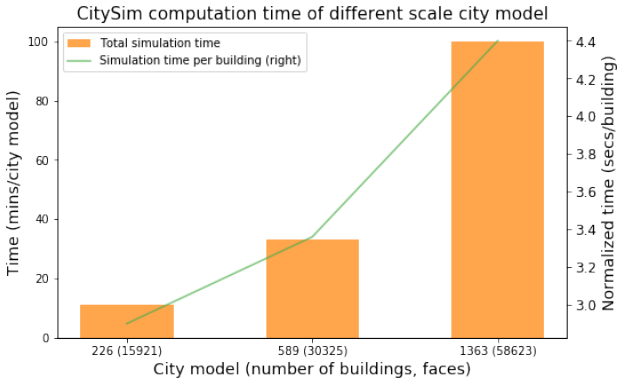


Figure 6.6: The test site is partitioned into three city models, the diagram shows how much CitySim computation time is needed to run one complete simulation for each case.

Figure 6.7: Visualizing the spatial distribution of the optimal posterior values of $Tmin$ and $Ninf$ of the calibrated postcodes.

7 DISCUSSION AND CONCLUSION

7.1 DISCUSSION

7.1.1 General discussion

To cut down immense greenhouse gases emission and energy consumption in the built environment requires a holistic understanding and rethinking of our dynamic urban energy system. Multi-scales and multi-domains energy flow data within the built environment is valuable information, especially when the penetration of decentralized and intermittent sustainable energy sources is constantly increasing. The research focus of the project, annual heating demand simulation, is merely one domain of the broad scope of UBEM. The modeling scale and purpose of UBEM can be very diverse.

The project has developed a calibrated annual heating demand modeling for the partial districts of Amsterdam based on open-source data collections. Comparing the baseline and the calibrated simulation results, the averaged absolute percentage error at postcode 6 level (approximately a group of 10 buildings or slightly more) of the validation years, 2016 and 2017, has improved from 24.96% to 8.31% and 19.93% to 7.70% respectively. Postcodes with absolute percentage error less than 10% has increased from 23.8% to 75% in 2016 (number of calibrated postcodes = 84) and from 32.9% to 78.8% in 2017 (number of calibrated postcodes = 85).

These error ranges are acceptable when the UBEM is served as a guidance to assist urban planning and retrofit measures assessment or to provide decision support [13]. However, simulation accuracy can significantly decreases when analyzed in higher spatial-temporal resolution, for instance 40% deviation is found in single building results according to Nouvel et al. [19], and 4% to 66% deviation in a building basis according to Fonseca and Schlueter [20] (more reported simulation errors of UBEM studies can refer to Reinhart and Cerezo Davila [13]), and these deviations are largely accounted by the stochastic type uncertainties.

How to generate energy flow information of high spatial-temporal granularity becomes an indispensable discussion in order to open more engineering applications based on UBEM. The impact of data availability and granularity on simulation results is obvious, this is not only a call for harmonizing diverse data collections for the sake of high-quality data but also emphasizing the need of releasing available datasets, under the secured condition, in order to fill the data gap in UBEM development.

As for what kind of spatial and temporal data matters to the specific modeling scale and task, sensitivity analysis such as Morris method provides a very efficient and interpretable way for prioritizing calibration targets. Nevertheless, narrowing down the data collection scope can sometimes be a

challenging task already, as the influence of the specific input could be unknown in the early development phase. As a consequence, it might be interesting to start discussing and incorporating the levels of detail (LODs) framework for the energy simulation inputs, based on the existing LODs framework applied in GIS domain, domain sizes and spatial-temporal resolution of the modeling techniques, similar to the scheme presented in Figure 2.6 from Frayssinet et al. [27].

This could possibly start with reviewing and summarizing the existing works, as many studies have tried to identify the key variables on different modeling scales and purposes (Table 2.2 presents some case studies), but no comprehensive review starting with this perspective exists yet. Secondly, sensitivity analysis can be a powerful tool in assisting such task when high quality and high granularity data is not directly available. When interpreting sensitivity analysis results, one should be careful that the result is model specific and the defined uncertainty range is a strong influencer. By performing aforementioned analysis, this could result in a multi-dimensions LODs framework, which can serve as a guideline for future UBEM development.

Energy data of high spatial-temporal resolution is a valuable asset for many engineering applications, but often not readily accessible. If no corresponding high frequency temporal inputs are given (e.g. precise occupancy profile, HVAC schedule, etc.), it might seem challenging to produce reliable energy data of high spatial-temporal resolution via UBEM because of stochastic uncertainty, which may be the biggest limitation of UBEM at the moment. Nevertheless, we have seen Bayesian inference is an effective approach to reduce subjective uncertainty, which in some way, can reduce to a single usable input value and produce acceptable energy data through simulation engine at reasonable temporal scale. Following this framework, it is expected given higher frequency (monthly or perhaps weekly) measured energy data, it is possible to more effectively and accurately reduce multiple parameter uncertainties (subjective type in particular) as the inference process has been tested with various conditions (e.g. seasonal variation).

In a rigorous sense, the parameter posterior distributions should be interpreted on postcode basis in this study. However, if the parameter posteriors of the specific building archetype (not necessary confined to the archetype definition of this study, could also be spatially close buildings, etc.) show statistically significant pattern, the parameter posteriors might be able to apply to the untrained postcodes or buildings to fill the spatial and energy data gap often seen at the urban scale. Additionally, since simulation inputs and energy data have both gone through the calibration process, the confidence level could increase when breaking down aggregated energy data into different end-use consumptions. For instance, if the heating system efficiency and domestic hot water system efficiency are both calibrated (subjective uncertainty in this case), the aggregated heating simulation result of the postcode (building) can be decomposed into respective consumption types, which generates finer energy flow information and makes further analysis and energy saving possible.

Table 7.1: 10 building characteristics are used to calculate definite energy label [7].

Construction year	Dwelling type
Type of glass	Facade insulation
Room insulation	Floor insulation
Heating system type	How water supply type
Ventilation system	Solar panels and solar water heater

Due to data availability and accessibility limitation, the major role of the current calibrated UBEM would be providing decision support or assisting scenario urban planning and building renovation. One practical use case could be large scale building performance mapping and labeling. In the Netherlands, registration of a definitive energy label when selling, releasing or delivering a house is enforced by law since January 2015. Based on the definite energy label, the house owner can take suggested building retrofit measures if necessary. Provisional energy label, calculated based on publicly registered data such as construction year, etc., is no longer sufficient after January 2015[1] [7]. However, registering definite energy label, which is calculated based on ten building characteristics as listed in Table 7.1, requires an authorized expert to evaluate individual house, which is a labor-intensive and time demanding task. This might explain why more than 50% of buildings do not have a definite energy label by the end of 2017 yet (Figure 7.3), and consequently slowing down the building renovation process. Following the building renovation rates achieved over the years 2010-2014, Filippidou [4] points out attaining the short-term goals of upgrading to an average energy label B in the non-profit Dutch housing stock by the end of 2020 is not probable. Based on this fact and the urgent need, calibrated UBEM is a powerful and versatile alternative to perform large-scale building performance mapping and labeling and comes with the capability to carry out retrofit measures assessment and scenario analysis. In addition, when the calibrated UBEM is developed into a decision support environment, visualizing energy consumption pattern and retrofit saving potentials could potentially increase citizen engagement, which is one of the key factors to ensure a successful energy transition.

Except the above general discussion, the following paragraphs will discuss other UBEM aspects in separated sections.

7.1.2 Data preparation and data harmonization

UBEM is a bottom-up approach to study energy performance at urban scale. As a consequence, extensive search for available open-source datasets from diverse sources and data harmonization works are indispensable in the very beginning.

- *What open-source GIS layers, statistics data, and technical datasets related to Amsterdam are available for UBEM development? and how to integrate these datasets in a sensible way to build up a simulation ready 3D city model?*

The study of the above question is fully discussed in Chapter 4. Table 4.5 provides an overview of the required and adopted datasets for the Amsterdam UBEM development. This also includes data period, source and

1 https://www.ep-online.nl/ep-online/Default.aspx

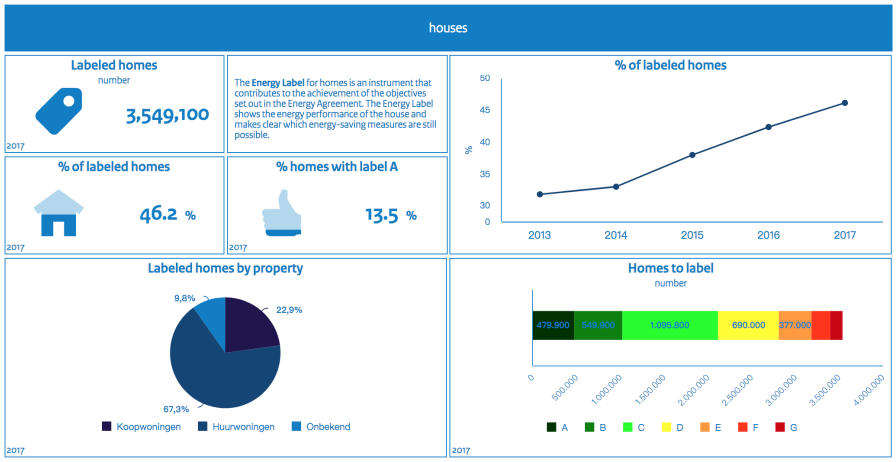


Figure 7.1: Energy label dashboard provides a quick overview of the current state of the Dutch building stock. Accessed from https://energiecijfers.databank.nl/dashboard/Energielabels/ [October 2018]

remark regarding data quality. Additional information regarding each parameter value, uncertainty range can refer to Table 4.6, Table 4.7, and Table 4.8 respectively. The open-source data collections for this project may not be guaranteed as the best datasets available, for instance, open-source CBS postcode 6 population is rather old compared with other datasets, and SHAERE database could be a very valuable source if accessibility is authorized. Nevertheless, it is believed that the summary made in Table 4.5 is a starting point for further UBEM projects applicable to the Netherlands.

At the time of writing this section (October, 2018), the author has also noticed that the stakeholders involved in energy transition in the Netherlands, from government bodies to utility companies have gathered together to discuss energy and spatial information standards and tried to make existing but decentralized national and international energy data collections into collected inventories, which is available at: https://www.geonovum.nl/themas/energie (in Dutch).

In addition to data preparation, integrating multi-datasets with different spatial-temporal resolutions is often recognized as a time-consuming and sometimes a discouraging task. Although the data integration workflow and specific data model tailed for the project requirement have been made from scratch (Figure 4.6 and Figure 4.7), to avoid reinventing the wheel and to increase the reusability of the 3D city model, it is worth to think about maintaining the city model in a standardized format such as CityGML [58] with Energy ADE [59] support, which is designed to facilitate data exchange and interoperability. More future works would be needed to understand and test how the newly released Energy ADE version 1.0 (January, 2018) can support multi-scales and multi-domains simulation and how general can the standardized data model fits diverse energy simulation engines and simulation applications.

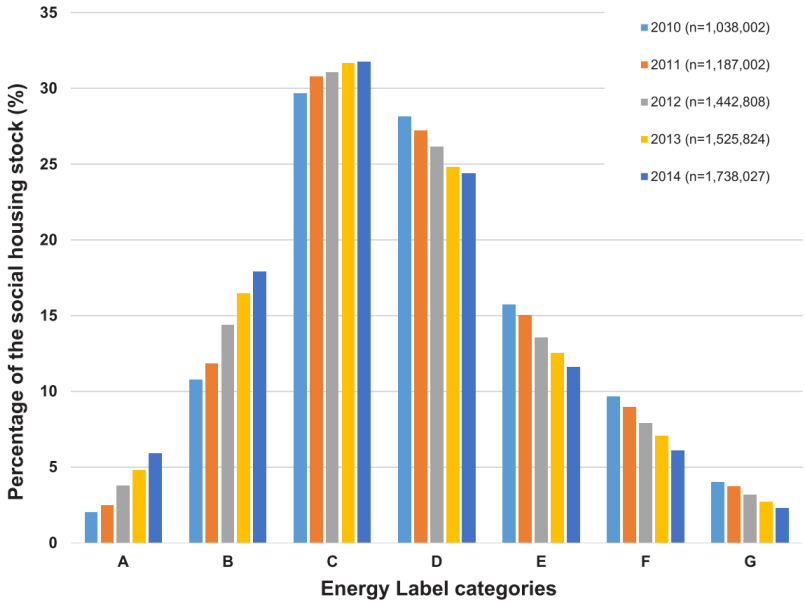


Figure 7.2: Filippidou [4] monitored the energy labels distribution of the non-profit rented buildings in the Netherlands from 2010 to 2014. The author concluded the renovation pace is too low to fulfill the ambitious goals of the national Covenant agreed in 2012 or reach the EU goals for energy efficiency.

7.1.3 CitySim simulation characteristics and sensitivity analysis

The theoretical considerations and practical requirements of adopting CitySim as a simulation engine are introduced in section 2.3 and section 2.4 in Chapter 2. Chapter 5 takes a closer look at CitySim simulation characteristics with the following sub-questions:

- *What is the added value and appropriateness of adopting CitySim as a simulation engine in this research scope?*
- *Given a number of simulation parameters with the associated uncertainty ranges, which ones are the key parameters affecting annual heating EUI calculation and which ones are minimal and even ignorable, according to the sensitivity analysis?*

Both experiments demonstrated in section 5.1 have clearly shown that CitySim can take geometry obstruction, radiation shadowing effect into account when a 3D city (building) model and local weather file are given. It does not come with the capability to simulate convective energy exchange occurred in the built environment. Coupling with other simulation models might be a solution [54], but it is not within the scope of the research. Besides, the experiments also show the positive correlation between building shape coefficient (S/V ratio) and annual heating EUI (kWh/m^3) calculation, which is expected in moderate to cold climate zones.

In addition, the experimental results reveal that the EUI difference caused by inter-building interactions (when only radiation aspect is considered) is comparatively modest. It is worth rethinking the necessity of adopting a

Figure 7.3: More than 50% buildings in the Netherlands do not have definite energy labels by the end of 2017, as can be seen from the National Energy Atlas Accessed from https://www.nationaleenergieatlas.nl/en/kaarten [October 2018]

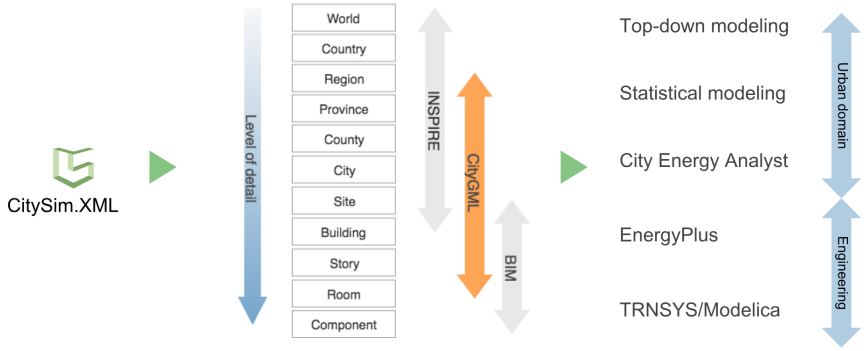


Figure 7.4: From stand-alone data model to standardized semantic data model can potentially increase data interoperability and facilitate data exchange to make multi-scale and multi-domain simulations and applications easier. Adapted from [59]

simulation engine with a detailed radiation model for UBEM development. If a simplified radiation model could significantly increase simulation speed without compromising too much accuracy, it would have positive impacts on UBEM scalability, usability as well as decrease development time.

The second sub-question can be answered by the sensitivity analysis result based on Morris method (see Figure 5.6). The user behavior (operation) parameter: minimum thermostat setting ($Tmin$) is shown to be the most influential input affecting annual heating EUI (kWh/m^3) calculation under the defined uncertainty range (15 to 20 °C) and the geometry dimension given to CitySim. This is followed by building envelope parameters such as $Uwall$, $Ninf$ and so on. On the other hand, highly uncertain parameters: building surface shortwave reflectance (SW), ground surface shortwave reflectance (GSW), and window openable ratio (WOR) are insignificant. The ignorable result of SW and GSW might be caused by the experiment setting that there is no surrounding building in this case. Conducting a sensitivity analysis when the neighboring buildings are presented is thus recommended in order to obtain a more thorough parameter ranking result in the future work. In addition, building height estimation uncertainty has a comparatively insignificant influence on annual heating EUI (kWh/m^3) calculation. This indicates that LOD1 city model should be sufficient for annual heating de-

mand simulation, which is consistent with the result concluded by Nouvel et al. [63]

7.1.4 Archetype modeling

Building construction periods and dwelling types have been widely used to classify building stocks of many European countries, one example is TABULA project [6]. This classification scheme is also appeared in Dutch national reference home standard, but with finer classification in dwelling types [84, 85]. The deterministic classification scheme applied in this project is adapted from these references, where the Amsterdam residential building stock is classified into 6 construction periods and 3 dwelling types, see Figure 5.7. To complete the entire archetype modeling, the following to sub-questions have to be answered.

- *How to infer building dwelling type in order to assist deterministic archetype classification?*
- *Which key uncertain parameters should be applied to probabilistic archetype modeling?*

As dwelling type information is not available from the existing data collections, additional GIS processing is required. The first approach attempts to use geometry features, shape coefficient (S/V ratio) and relative compactness for the classification purpose [18, 97]. However, it is found that a number of buildings are misclassified if only these two features are used.

The final implemented classification rule is based on analyzing number of address per building and building footprint topology relation (how many neighbors does the building have), see Figure 4.9.

Sensitivity analysis based on Morris method provides an insight for the uncertain parameter importance ranking. Two key uncertain parameters: $Tmin$ and $Ninf$, divided into 5 sections respectively and assumed uniform prior probability distributions, are adopted to expand the existing archetype. This probabilistic archetype model is visualized in Figure 5.7. Although the author believe the best practice of modeling a probabilistic archetype classification would require running sensitivity analysis for each existing deterministic archetype individually, namely, conducting 18 sensitivity analysis with each one is provided with the corresponding uncertainty ranges (some parameters are construction year dependent) and geometry (example dwelling type geometry). The probabilistic archetype modeling is then modified accordingly. However, this will inevitably increase the complexity of the project and thus leave it as a future work.

7.1.5 Bayesian inference and model calibration

The effectiveness, limitation, and potential of applying Bayesian inference for model calibration have been discussed in the beginning of the chapter and thus do not repeat here. Nevertheless, it should be reminded that in the Bayesian inference and calibration framework, using Gaussian normal distribution to describe the likelihood function is an assumption since the explicit function does not exist. The current implementation might fall short

to accurately describe the real likelihood distribution for the respective input combination, for instance, the simulation variability caused by other input uncertainties are not incorporated in this formulation. This is an issue to be addressed in future work.

Furthermore, possible future works can investigate how to optimize this framework to compute more efficiently and effectively. In this project, training the model (2178 buildings) with six years of gas consumption data on a single 4 cores personal computer requires approximately two weeks. How effective would the calibration process be if the model is trained with even fewer observation data? How to avoid over-fitting issue and how to efficiently scale the framework to high dimension calibration to incorporate more parameters? How this approach can possibly reduce more subjective uncertainties in order to fill spatial and energy data gap? would be some future work questions in order to unlock the full potential of UBEM.

7.2 CONCLUSION

Energy simulation performance gap is a common issue and can happen at any spatial temporal modeling scale especially when input data quality is poor. To fill data gap and minimize data uncertainty at urban scale and to increase heating simulation reliability and usability, the methodology applied in this project has successfully carried out an urban scale heating demand modeling based on a LOD1 3D city model of the partial districts in Amsterdam and calibrated at least 84 residential postcodes based on Bayesian approach, provided with six years of gas consumption data. The effectiveness of the Bayesian calibration framework is validated when comparing the baseline and the calibrated heating demand simulation results, see Figure 6.3 and Figure 6.4. After the model calibration phase, postcodes with absolute percentage error less than 10% have increased from 23.8% to 75% in 2016 (number of calibrated postcodes = 84) and from 32.9% to 78.8% in 2017 (number of calibrated postcodes = 85).

To ensure an effective model calibration, performing sensitivity analysis is well-advised. The result derived from the efficient and effective Morris method indicates that thermostat setting has the most significant impact on annual heating demand simulation in terms of Amsterdam building stock, followed by building construction parameters such as U-values and infiltration rate. Besides, LOD1 city model should be sufficient to produce acceptable annual heating results based on CitySim.

Modeling the key uncertain parameters in a probabilistic way can appropriately depict the dynamic nature of the urban environment. With the help of Bayesian inference and adequate observation data, parameter uncertainties can be further reduced, and consequently establishing a more reliable UBEM and opening more applications. Following this framework and adjusting according to the local context, calibrated bottom-up heating demand energy model can be developed in most cities in the world as long as sufficient spatial, non-spatial data is provided, and the utilities are willing to disclose partial energy consumption data.

The calibrated urban building energy model would be most needed by the municipality, urban planners, utilities and engineering consultancy who might show keen interest to perform energy policy assessment, scenario analysis. It also has a potential to perform large-scale building performance mapping and labeling in order to prioritize building retrofit targets and to accelerate building stock renovation and energy transition.

7.3 CONTRIBUTIONS

In this interdisciplinary research, many efforts have been made in sourcing available energy related datasets from diverse sources in oder to characterize building properties with deterministic values, and to quantify the associated uncertain ranges. These spatial, non-spatial data collections are in the end integrated into a 3D city model in a CitySim XML format to carry out an urban scale heating demand simulation. The collected datasets and the developed workflows for data integration and geometry preparation should serve as a good starting point for any future UBEM project, which are particularly applicable to the cities in the Netherlands.

Conducting an urban scale heating demand estimation is not new in spatial science field; however, simulation performance gap either caused by simulation engine inadequacy or insufficient input data quality have always been challenging and not fully solved issues. On the contrary, in building energy simulation field, parameter uncertainty, sensitivity analysis, and model calibration are more widely discussed but also more confined at the building or component scale. Hopefully by doing such interdisciplinary study, more dialogues and knowledge exchanges can happen between both fields, and consequently come up with a more integrated solution to tackle the sustainability challenges.

Furthermore, the major contribution is developing and validating a calibrated 3D urban energy model in partial districts in Amsterdam, which is a more versatile tool than statistical method and possible to perform diverse scenario analysis with a better credibility. For instance, the government and urban planners can start to use the calibrated urban energy model to carry out heating consumption analysis given long-term climate change scenario and formulate energy transition roadmaps. The calibrated urban heating demand model can also assist in identifying retrofit hot spots and performing building retrofit assessment, helping the government and stakeholders to find the most cost-effective building renovation measures.

From a technical perspective, the developed Python scripts: allowing overwriting CitySim XML file from the queried data and executing simulation automatically, FME script: automating data cleaning and integration workflow, and Grasshopper script: preparing building geometry for CitySim energy simulation as well as visualizing simulation results. These outputs are made available on the GitHub repository[2]. Follow the instruction should allow anyone with certain technical experience to develop an UBEM for other areas of interest.

2 Amsterdam CitySim UBEM: https://github.com/ckwang25/Amsterdam_CitySim_UBEM

7.4 FUTURE WORK AND RECOMMENDATION

The possible future works with respect to the different aspects will be listed and discussed below.

7.4.1 Data preparation and uncertainty quantification

- Uncertainties are everywhere in building energy simulation. The collected datasets and the assumptions made can not guarantee these uncertainties are best minimized. A call for high-quality data is always needed, especially related to operation and system data categories. The latest CBS census data, local occupancy profiles (possibly derived from smart meter observation) rather then a standardized profile, and SHAERE database could be valuable inputs to the future UBEM development.

- Preparing simulation ready city model is one of the most resource and time demanding tasks in this project. It is worth to think about and test transforming the existing data model to the standardized data model such as CityGML and Energy ADE, so the city model can be reused for multi-scale and multi-domain simulations.

7.4.2 Sensitivity analysis

- Morris method is the adopted sensitivity analysis method because of its efficiency. It is capable of giving parameter importance ranking within acceptable computation time and easily interpretable. Nevertheless, it is worth to experiment alternative methods such as Sobel method, in order to cross validate the results and to test how can it assist uncertainty quantification.

- Conducting individual sensitivity analysis for each existing deterministic archetype. This could lead to a more specific parameter ranking. Probabilistic modeling would be reflected in a more diverse but precise way, and could potentially lead to a more effective calibration.

7.4.3 UBEM development in general

- Since the calibration is an over-specified and under-determined problem, if high spatial-temporal resolution energy measurements (monthly or even weekly consumption data) are accessible, this could help high dimension calibration (calibrating more key uncertain parameters at once) become more feasible and also makes the result more stable (seasonally variations would also be tested).

- Running calibrating simulations is a computationally intensive process especially when the city model becomes too complex, distributed computing or cloud architecture might be an alternative solution.

- Besides transferring to other simulation architecture, developing UBEM with different simulation engines with simplified radiation model can be tested. Since the results shown in section 5.1 demonstrates that annual EUI difference is rather modest in this case study when geometry

obstruction is the only variable. Be noted the result could be different if tested under different weather conditions.

- How to more accurately formulate the likelihood function would need further studies.
- Investigating how the Bayesian approach can be optimized to carry out an efficient and effective calibration. For instance, train the model with fewer measured energy data or examine the simulation performance when applying the parameter posteriors to run energy simulation on the untrained postcodes (buildings).
- The current implementation is only focus on residential building heating demand simulation. Scaling up to electricity consumption, cooling demand simulation or even commercial buildings in the UBEM would be another research possibility.
- Further development is needed to make this calibrated UBEM into a decision support and planning assessment platform. However, extreme care should be taken on the user experience, workflow simplification, and application life cycle as according to the interview with one of the developers of the ambitious UBEM project, TRANSFORM (six European cities involved). The developed application became less actively used after the end of the 3 years project because the users found it comprehensive but too complicated.

BIBLIOGRAPHY

[1] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, 2009. ISSN 13640321. doi: 10.1016/j.rser.2008.09.033.

[2] Yeonsook Heo. *Bayesian Calibration of Building Energy Models for Energy Retrofit Decision-Making under Uncertainty*. 2011.

[3] Tian Wei. A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, 20:411–419, 2013. ISSN 13640321. doi: 10.1016/j.rser.2012.12.014. URL http://dx.doi.org/10.1016/j.rser.2012.12.014.

[4] Faidra Filippidou. *Energy performance progress of the Dutch non-profit housing stock: a longitudinal assessment*. 2018. ISBN 9789463660471.

[5] K Leidelmeijer and P Van Grieken. *Wonen en energie - Stook en ventilatiegedrag van huishoudens*. 2005.

[6] Tobias Loga, Nikolaus Diefenbach, Britta Stein, Constantinos A Balaras, Otto Villatoro, and Kim B Wittchen. *Typology Approach for Building Stock Energy Assessment . Main Results of the TABULA project – Final Project Report*. 2012.

[7] Rijksdienst voor Ondernemend Nederland. *Energielabel woningen*. https://www.rvo.nl/onderwerpen/duurzaam-ondernemen/gebouwen/wetten-en-regels-gebouwen/bestaande-bouw/energielabel-woningen [Accessed: October 2018].

[8] European Commission. *EU Building Stock Observatory*. https://ec.europa.eu/energy/eubuildings [Accessed: March 2018].

[9] Carsten Petersdorff, Thomas Boermans, Jochen Harnisch, Ole Stobbe, Simone Ullrich, and Sina Wartmann. *Cost-Effective Climate Protection in the EU Building Stock*. 2005.

[10] Department of Economic United Nations and Population Department Social Affairs. *World Urbanization Prospects*, volume 12. 2014. ISBN 9789211515176. doi: 10.4054/DemRes.2005.12.9.

[11] International Energy Agency. *Transition to Sustainable Buildings - Strategies and opportunities to 2050*. 2013. ISBN 978-92-64-20241-2. doi: 10.1787/9789264202955-en. URL http://www.iea.org/publications/freepublications/publication/Building2013_free.pdf%5Cnwww.iea.org/etp/buildings.

[12] James Keirstead, Mark Jennings, and Aruna Sivakumar. A review of urban energy system models: Approaches, challenges and opportunities. *Renewable and Sustainable Energy Reviews*, 16(6):3847–3866, 2012. ISSN 13640321. doi: 10.1016/j.rser.2012.02.047. URL http://dx.doi.org/10.1016/j.rser.2012.02.047.

[13] Christoph F. Reinhart and Carlos Cerezo Davila. Urban building energy modeling - A review of a nascent field. *Building and Environment*, 97: 196–202, 2016. ISSN 03601323. doi: 10.1016/j.buildenv.2015.12.001. URL http://dx.doi.org/10.1016/j.buildenv.2015.12.001.

[14] Alessio Mastrucci, Olivier Baume, Francesca Stazi, and Ulrich Leopold. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. *Energy and Buildings*, 75:358–367, 2014. ISSN 03787788. doi: 10.1016/j.enbuild.2014.02.032. URL http://dx.doi.org/10.1016/j.enbuild.2014.02.032.

[15] Jonas Allegrini, Kristina Orehounig, Georgios Mavromatidis, Florian Ruesch, Viktor Dorer, and Ralph Evins. A review of modelling approaches and tools for the simulation of district-scale energy systems. *Renewable and Sustainable Energy Reviews*, 52:1391–1404, 2015. ISSN 18790690. doi: 10.1016/j.rser.2015.07.123.

[16] Joshua Bergerson, Ralph T Muehleisen, B O Rodda, Joshua A Auld, Leah B Guzowski, Jonathan Ozik, and Nicholson Collier. *LakeSIM Integrated Design Tool For Assessing Short- And Long-Term Imapcts Of Urban Scale Conceptual Designs*. 2015.

[17] B. Howard, L. Parshall, J. Thompson, S. Hammer, J. Dickinson, and V. Modi. Spatial distribution of urban building energy consumption by end use. *Energy and Buildings*, 45:141–151, 2012. ISSN 03787788. doi: 10.1016/j.enbuild.2011.10.061. URL http://dx.doi.org/10.1016/j.enbuild.2011.10.061.

[18] Sara Torabi Moghadam, Jacopo Toniolo, Guglielmina Mutani, and Patrizia Lombardi. A GIS-statistical approach for assessing built environment energy use at urban scale. *Sustainable Cities and Society*, 37: 70–84, 2018. ISSN 22106707. doi: 10.1016/j.scs.2017.10.002. URL https://doi.org/10.1016/j.scs.2017.10.002.

[19] Romain Nouvel, Claudia Schulte, Ursula Eicker, Dirk Pietruschka, and Volker Coors. CityGML-Based 3D City Model for Energy Diagnostics and Urban Energy Policy Support. *Proceedings of BS2013: 13th Conference of International Building Performance Simulation Association*, pages 218–225, 2013.

[20] Jimeno A. Fonseca and Arno Schlueter. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. *Applied Energy*, 142:247–265, 2015. ISSN 03062619. doi: 10.1016/j.apenergy.2014.12.068. URL http://dx.doi.org/10.1016/j.apenergy.2014.12.068.

[21] Carlos Cerezo, Julia Sokol, Saud AlKhaled, Christoph Reinhart, Adil Al-Mumin, and Ali Hajiah. Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City. *Energy and Buildings*, 154:321–334, 2017. ISSN 03787788. doi: 10.1016/j.enbuild.2017.08.029. URL https://doi.org/10.1016/j.enbuild.2017.08.029.

[22] Daniel Coakley, Paul Raftery, and Marcus Keane. A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37:123–141, 2014. ISSN 13640321.

doi: 10.1016/j.rser.2014.05.007. URL http://dx.doi.org/10.1016/j.rser.2014.05.007.

[23] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001. ISSN 1369-7412. doi: 10.1111/1467-9868.00294. URL http://doi.wiley.com/10.1111/1467-9868.00294.

[24] A Saltelli, M Ratto, T Andres, F Camplongo, J Cariboni, D Gatelli, M Saisana, and S Tarantola. *Global Sensitivity Analysis: The Primer*. 2008. ISBN 9780470059975. URL http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470059974.html.

[25] T.R. Oke. *Boundary layer climates, 2nd ed.* 1987.

[26] Viktor Dorer, Jonas Allegrini, Kristina Orehounig, Peter Moonen, Govinda Upadhyay, Jérôme Kämpf, and Jan Carmeliet. Modelling the urban microclimate and its impact on the energy demand of buildings and building clusters. *13th Conference of International Building Performance Simulation Association, Chambery, France, August 26-28*, pages 3483–3489, 2013.

[27] Loïc Frayssinet, Lucie Merlier, Frédéric Kuznik, Jean Luc Hubert, Maya Milliez, and Jean Jacques Roux. Modeling the heating and cooling energy demand of urban buildings at city scale. *Renewable and Sustainable Energy Reviews*, 81:2318–2327, 2018. ISSN 18790690. doi: 10.1016/j.rser.2017.06.040. URL https://doi.org/10.1016/j.rser.2017.06.040.

[28] Darren Robinson, Frédéric Haldi, Jérôme Henri Kämpf, P Leroux, D Perez, a Rasheed, and U Wilke. CITYSIM: Comprehensive micro-simulation of resource flows for sustainable urban planning. *International IBPSA Conference*, pages 1083–1090, 2009.

[29] International Energy Agency. *Energy Technology Perspectives: Scenarios & Strategies To 2050*. 2010. ISBN 9789264085978. doi: 10.1049/et:20060114. URL http://www.oecd-ilibrary.org.ezproxy.library.uq.edu.au/energy/energy-technology-perspectives-2010_energy_tech-2010-en.

[30] WBCSD. *Transforming the Market: Energy Efficiency in Buildings*. 2015. URL http://www.wbcsd.org/.

[31] International Energy Agency. *World Energy Outlook 2014*. 2014. ISBN 9789264208056. doi: 10.1787/weo-2014-en. URL http://www.oecd-ilibrary.org.proxy1.cl.msu.edu/docserver/download/6112251e.pdf?expires=1364602454&id=id&accname=ocid177642&checksum=AF7E0799FCF65B15C93CCE111D065B9E.

[32] Intergovernmental Panel on Climate Change. *Climate Change 2007 Synthesis Report*. 2007. ISBN 9291691224. doi: 10.1256/004316502320517344.

[33] Ralph Evins. A review of computational optimisation methods applied to sustainable building design. *Renewable and Sustainable Energy Reviews*, 22:230–245, 2013. ISSN 13640321. doi: 10.1016/j.rser.2013.02.004. URL http://dx.doi.org/10.1016/j.rser.2013.02.004.

[34] Hyunwoo Lim and Zhiqiang John Zhai. Review on stochastic modeling methods for building stock energy prediction. *Building Simulation*, 10 (5):607–624, 2017. ISSN 19968744. doi: 10.1007/s12273-017-0383-y.

[35] Pekka Koponen, Antti Mutanen, and Harri Niska. Assessment of some methods for short-term load forecasting. *IEEE PES Innovative Smart Grid Technologies, Europe*, pages 1–6, 2014. doi: 10.1109/ISGTEurope.2014.7028901. URL http://ieeexplore.ieee.org/document/7028901/.

[36] Clayton Miller, Zoltán Nagy, and Arno Schlueter. A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. *Renewable and Sustainable Energy Reviews*, pages 1–13, 2016. ISSN 18790690. doi: 10.1016/j.rser.2017.05.124. URL http://dx.doi.org/10.1016/j.rser.2017.05.124.

[37] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74: 902–924, 2017. ISSN 18790690. doi: 10.1016/j.rser.2017.02.085. URL http://dx.doi.org/10.1016/j.rser.2017.02.085.

[38] Miguel Molina-Solana, María Ros, M. Dolores Ruiz, Juan Gómez-Romero, and M. J. Martin-Bautista. Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews*, 70: 598–609, 2017. ISSN 18790690. doi: 10.1016/j.rser.2016.11.132. URL http://dx.doi.org/10.1016/j.rser.2016.11.132.

[39] Geoffrey K.F. Tso and Kelvin K.W. Yau. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768, 2007. ISSN 03605442. doi: 10.1016/j.energy.2006.11.010.

[40] Marta Braulio-Gonzalo, Pablo Juan, María D. Bovea, and María José Ruá. Modelling energy efficiency performance of residential building stocks based on Bayesian statistical inference. *Environmental Modelling and Software*, 83:198–211, 2016. ISSN 13648152. doi: 10.1016/j.envsoft.2016.05.018.

[41] David Hsu. Identifying key variables and interactions in statistical models ofbuilding energy consumption using regularization. *Energy*, 83: 144–155, 2015. ISSN 03605442. doi: 10.1016/j.energy.2015.02.008. URL http://dx.doi.org/10.1016/j.energy.2015.02.008.

[42] Francis Moran, Sukumar Natarajan, and Marialena Nikolopoulou. Developing a database of energy use for historic dwellings in Bath, UK. *Energy and Buildings*, 55:218–226, 2012. ISSN 03787788. doi: 10.1016/j.enbuild.2012.09.016. URL http://dx.doi.org/10.1016/j.enbuild.2012.09.016.

[43] Thomas Olofsson, Staffan Andersson, and Jan Ulric Sjögren. Building energy parameter investigations based on multivariate analysis. *Energy and Buildings*, 41(1):71–80, 2009. ISSN 03787788. doi: 10.1016/j.enbuild.2008.07.012.

[44] Claudia Sousa Monteiro, Carlos Costa, André Pina, Maribel Y. Santos, and Paulo Ferrão. An urban building database (UBD) supporting a smart city information system. *Energy and Buildings*, 158:244–260, 2018. ISSN 03787788. doi: 10.1016/j.enbuild.2017.10.009. URL http://dx.doi.org/10.1016/j.enbuild.2017.10.009.

[45] Clayton Miller, Daren Thomas, Jérôme Kämpf, and Arno Schlueter. Urban and building multiscale co-simulation: case study implementations on two university campuses. *Journal of Building Performance Simulation*, pages 1–13, 2017. ISSN 19401507. doi: 10.1080/19401493.2017.1354070.

[46] Y. Heo, R. Choudhary, and G. A. Augenbroe. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47:550–560, 2012. ISSN 03787788. doi: 10.1016/j.enbuild.2011.12.029. URL http://dx.doi.org/10.1016/j.enbuild.2011.12.029.

[47] A. T. Booth, R. Choudhary, and D. J. Spiegelhalter. Handling uncertainty in housing stock models. *Building and Environment*, 48(1):35–47, 2012. ISSN 03601323. doi: 10.1016/j.buildenv.2011.08.016. URL http://dx.doi.org/10.1016/j.buildenv.2011.08.016.

[48] René Buffat, Andreas Froemelt, Niko Heeren, Martin Raubal, and Stefanie Hellweg. Big data GIS analysis for novel approaches in building stock modelling. *Applied Energy*, 208:277–290, 2017. ISSN 03062619. doi: 10.1016/j.apenergy.2017.10.041.

[49] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W. F. Buhl, Y. Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319–331, 2001. ISSN 03787788. doi: 10.1016/S0378-7788(00)00114-6.

[50] Carlos Cerezo Davila, Christoph Reinhart, and Jamie Bemis. Modeling Boston: A workflow for the generation of complete urban building energy demand models from existing urban geospatial datasets. *Energy*, 117:237–250, 2016. ISSN 03605442. doi: 10.1016/j.energy.2016.10.057. URL http://dx.doi.org/10.1016/j.energy.2016.10.057.

[51] Julia Sokol, Carlos Cerezo Davila, and Christoph F. Reinhart. Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings*, 134:11–24, 2017. ISSN 03787788. doi: 10.1016/j.enbuild.2016.10.050. URL http://dx.doi.org/10.1016/j.enbuild.2016.10.050.

[52] Jonas Allegrini, Viktor Dorer, and Jan Carmeliet. Influence of the urban microclimate in street canyons on the energy demand for space cooling and heating of buildings. *Energy and Buildings*, 55:823–832, 2012. ISSN 03787788. doi: 10.1016/j.enbuild.2012.10.013. URL http://dx.doi.org/10.1016/j.enbuild.2012.10.013.

[53] Emmanuel Walter and Jérôme Henri Kämpf. A verification of CitySim results using the BESTEST and monitored consumption values. *Proceedings of the 2nd Building Simulation Applications conference*, pages 215–222, 2015. URL https://infoscience.epfl.ch/record/214754.

[54] Dasaraden Mauree, Silvia Coccolo, Jérôme Kaempf, and Jean Louis Scartezzini. Multi-scale modelling to evaluate building energy consumption at the neighbourhood scale. *PLoS ONE*, 12(9), 2017. ISSN 19326203. doi: 10.1371/journal.pone.0183437.

[55] R. Judkoff and J. Neymark. International Energy Agency building energy simulation test (BESTEST) and diagnostic method. *National Renewable Energy Laboratory*, page Size: 296 pages, 1995. ISSN 14771535. doi: 10.2172/90674. URL http://www.osti.gov/scitech/biblio/90674%5Cnhttp://www.nrel.gov/docs/legosti/old/6231.pdf.

[56] TRANSFORM development Team. *TRANSFORM DST Deployment Guide*. 2015. URL http://urbantransform.eu/wp-content/uploads/sites/2/2013/04/DSE-Deployment-Guide-v1.0.pdf.

[57] Luís De Sousa, Christopher Eykamp, Ulrich Leopold, Olivier Baume, and Christian Braun. iGUESS - A web based system integrating Urban Energy Planning and Assessment Modelling for multi-scale spatial decision making. *International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet*, page 8, 2012. doi: 10.13140/2.1.3913.7284. URL http://www.iemss.org/society/index.php/iemss-2012-proceedings.

[58] Gerhard Gröger, Thomas Kolbe, Claus Nagel, and Karl-Heinz Häfele. OGC City Geography Markup Language (CityGML) En-coding Standard. *Ogc*, pages 1 – 344, 2012. ISSN 18632246. doi: OGC12-019. URL https://portal.opengeospatial.org/files/?artifact{_}id=47842.

[59] Giorgio Agugiaro, Joachim Benner, Piergiorgio Cipriano, and Romain Nouvel. The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations. *Open Geospatial Data, Software and Standards*, 3(1):2, 2018. ISSN 2363-7501. doi: 10.1186/s40965-018-0042-y. URL https://opengeospatialdata.springeropen.com/articles/10.1186/s40965-018-0042-y.

[60] Tatjana Kutzner and Thomas H. Kolbe. Extending semantic 3D city models by supply and disposal networks for analysing the urban supply situation. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation e.V.*, page 13, 2016. ISSN 0942-2870.

[61] virtualcitySystems. https://www.virtualcitysystems.de/en/citygml-training/3dcitydb-workshop [Accessed: October 2018].

[62] Tianzhen Hong, Yixing Chen, Hoon Lee, and Mary Ann Piette. CityBES : A Web - based platform to support city - scale building energy efficiency. *5th International Urban Computing Workshop*, 2016. doi: http://dx.doi.org/10.1145/12345.67890.

[63] Romain Nouvel, Maryam Zirak, Volker Coors, and Ursula Eicker. The influence of data quality on urban heating demand modeling using 3D city models. *Computers, Environment and Urban Systems*, 64:68–80, 2017. ISSN 01989715. doi: 10.1016/j.compenvurbsys.2016.12.005. URL http://dx.doi.org/10.1016/j.compenvurbsys.2016.12.005.

[64] R Nouvel, K-H Brassel, M Bruse, E Duminil, V Coors, U Eicker, and D Robinson. SIMSTADT, A new worflow-driven urban energy simulation platform for CityGML city models. *CISBAT 2015 - Lausanne, Switzerland*, pages 889–894, 2015.

[65] Enrico Fabrizio and Valentina Monetti. Methodologies and advancements in the calibration of building energy models. *Energies*, 8(4):2548–2574, 2015. ISSN 19961073. doi: 10.3390/en8042548.

[66] Kathrin Menberg, Yeonsook Heo, and Ruchi Choudhary. Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information. *Energy and Buildings*, 2016. ISSN 03787788. doi: 10.1016/j.enbuild.2016.10.005.

[67] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018. URL http://dx.doi.org/10.1016/j.cpc.2009.09.018.

[68] Thierry A. Mara and Stefano Tarantola. Application of global sensitivity analysis of model output to building thermal simulations. *Building Simulation*, 1(4):290–302, 2008. ISSN 1996-3599. doi: 10.1007/s12273-008-8129-5. URL http://link.springer.com/10.1007/s12273-008-8129-5.

[69] Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice*. 2004. ISBN 0470870931.

[70] Janelle S. Hygh, Joseph F. DeCarolis, David B. Hill, and S. Ranji Ranjithan. Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, 57:165–175, 2012. ISSN 03601323. doi: 10.1016/j.buildenv.2012.04.021. URL http://dx.doi.org/10.1016/j.buildenv.2012.04.021.

[71] Ilaria Ballarini and Vincenzo Corrado. Analysis of the building energy balance to investigate the effect of thermal insulation in summer conditions. *Energy and Buildings*, 52:168–180, 2012. ISSN 03787788. doi: 10.1016/j.enbuild.2012.06.004. URL http://dx.doi.org/10.1016/j.enbuild.2012.06.004.

[72] Yusuf Yildiz, Koray Korkmaz, Türkan Göksal özbalta, and Zeynep Durmus Arsan. An approach for developing sensitive design parameter guidelines to reduce the energy requirements of low-rise apartment buildings. *Applied Energy*, 93:337–347, 2012. ISSN 03062619. doi: 10.1016/j.apenergy.2011.12.048. URL http://dx.doi.org/10.1016/j.apenergy.2011.12.048.

[73] Pieter de Wilde and Wei Tian. Identification of key factors for uncertainty in the prediction of the thermal performance of an office building under climate change. *Building Simulation*, 2(3):157–174, 2009. ISSN 19963599. doi: 10.1007/s12273-009-9116-1.

[74] Pieter de Wilde and Wei Tian. Preliminary application of a methodology for risk assessment of thermal failures in buildings subject to climate change. *Building Simulation*, pages 2077–2084, 2009.

[75] Per Heiselberg, Henrik Brohus, Allan Hesselholt, Henrik Rasmussen, Erkki Seinre, and Sara Thomas. Application of sensitivity analysis in design of sustainable buildings. *Renewable Energy*, 34(9):2030–2036, 2009. ISSN 09601481. doi: 10.1016/j.renene.2009.02.016. URL http://dx.doi.org/10.1016/j.renene.2009.02.016.

[76] Vincenzo Corrado and Houcem Eddine Mechri. Uncertainty and sensitivity analysis for building energy rating. *Journal of Building Physics*, 33 (2):125–156, 2009. ISSN 17442591. doi: 10.1177/1744259109104884.

[77] Houcem Eddine Mechri, Alfonso Capozzoli, and Vincenzo Corrado. Use of the ANOVA approach for sensitive building energy design. *Applied Energy*, 87(10):3073–3083, 2010. ISSN 03062619. doi: 10.1016/j.apenergy.2010.04.001. URL http://dx.doi.org/10.1016/j.apenergy.2010.04.001.

[78] Clara Spitz, Laurent Mora, Etienne Wurtz, and Arnaud Jay. Practical application of uncertainty analysis and sensitivity analysis on an experimental house. *Energy and Buildings*, 55:459–470, 2012. ISSN 03787788. doi: 10.1016/j.enbuild.2012.08.013. URL http://dx.doi.org/10.1016/j.enbuild.2012.08.013.

[79] Roberto Ruiz Flores, Stephane Bertagnolio, Vincent Lemort, Roberto Ruiz, Stephane Bertagnolio, and Vincent Lemort. Global Sensitivity Analysis applied to Total Energy Use in Buildings. *International High Performance Buildings Conference*, (2004):1–10, 2012. URL http://docs.lib.purdue.edu/ihpbc%5Cnhttp://docs.lib.purdue.edu/ihpbc/78%5Cnhttp://orbi.ulg.ac.be/handle/2268/129192.

[80] Pieter De Wilde and Wei Tian. Predicting the performance of an office under climate change: A study of metrics, sensitivity and zonal resolution. *Energy and Buildings*, 42(10):1674–1684, 2010. ISSN 03787788. doi: 10.1016/j.enbuild.2010.04.011. URL http://dx.doi.org/10.1016/j.enbuild.2010.04.011.

[81] Wei Tian and Pieter de Wilde. Uncertainty and sensitivity analysis of building performance using probabilistic climate projections: A uk case study. *Automation in Construction*, 20(8):1096–1109, 2011. doi: 10.1016/j.autcon.2011.04.011.

[82] ESRI. *ESRI Shapefile Technical Description*, volume 16. 1998. ISBN 3931516482. doi: 10.1016/0167-9473(93)90138-J. URL http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf.

[83] P. Prempraneerach, F. S. Hover, M. S. Triantafyllou, C. Chryssostomidis, and G. E. Karniadakis. Sensitivity analysis and low-dimensional stochastic modeling of shipboard integrated power systems. *PESC Record - IEEE Annual Power Electronics Specialists Conference*, pages 1999–2003, 2008. ISSN 02759306. doi: 10.1109/PESC.2008.4592237.

[84] Agentschap NL - Ministerie van Economische Zaken. *Voorbeeldwoningen 2011: Bestaande bouw*, volume 6. 2011.

[85] Agentschap NL - Ministerie van Economische Zaken. *Referentie woningen nieuwbouw 2013*. 2013.

[86] Hugo Ledoux and Martijn Meijers. Topologically consistent 3D city models obtained by extrusion. *International Journal of Geographical Information Science*, 25(4):557–574, 2011. ISSN 13658816. doi: 10.1080/13658811003623277.

[87] Filip Biljecki, Hugo Ledoux, and Jantien Stoter. Height references of CityGML LOD1 buildings and their influence on applications. *Proceedings of the ISPRS 3D GeoInfo 2014 conference*, pages 1–17, 2014. doi: 10.4233/uuid:09d030b5-67d3-467b-babb-5e5ec10f1b38.

[88] Diane Perez. *A framework to model and simulate the disaggregated energy flows supplying buildings in urban areas*, volume 6102. 2014. doi: 10.5075/EPFL-THESIS-6102.

[89] C. N. Bramiana, A. G. Entrop, and J. I.M. Halman. Relationships between Building Characteristics and Airtightness of Dutch Dwellings. *Energy Procedia*, 96:580–591, 2016. ISSN 18766102. doi: 10.1016/j.egypro.2016.09.103. URL http://dx.doi.org/10.1016/j.egypro.2016.09.103.

[90] Wanyu R. Chan, William W. Nazaroff, Phillip N. Price, Michael D. Sohn, and Ashok J. Gadgil. Analyzing a database of residential air leakage in the United States. *Atmospheric Environment*, 39(19):3445–3455, 2005. ISSN 13522310. doi: 10.1016/j.atmosenv.2005.01.062.

[91] D'Ambrosio F.R. Alfano, M. Dell'Isola, G. Ficco, and F. Tassini. Experimental analysis of air tightness in Mediterranean buildings using the fan pressurization method. *Building and Environment*, 53:16–25, 2012. ISSN 03601323. doi: 10.1016/j.buildenv.2011.12.017. URL http://dx.doi.org/10.1016/j.buildenv.2011.12.017.

[92] Brunsting Suzanne, Daniels Bert, Schoots Koen, and Hammingh Pieter. *National Energy Outlook 2015*. 2015. URL http://www.pbl.nl/sites/default/files/cms/publicaties/pbl-2015-dutch-national-energy-outlook-2015-summary-02325.pdf.

[93] Simona D'Oca and Tianzhen Hong. Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, 88:395–408, 2015. ISSN 03787788. doi: 10.1016/j.enbuild.2014.11.065. URL http://dx.doi.org/10.1016/j.enbuild.2014.11.065.

[94] Olivia Guerra Santin, Laure Itard, and Henk Visscher. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings*, 41(11):1223–1232, 2009. ISSN 03787788. doi: 10.1016/j.enbuild.2009.07.002.

[95] O. Guerra-Santin and S. Silvester. Development of Dutch occupancy and heating profiles for building simulation. *Building Research and Information*, 45(4):396–413, 2017. ISSN 14664321. doi: 10.1080/09613218.2016.1160563.

[96] Casper Tigchelaar and Kees Leidelmeijer. *Energiebesparing: Een samenspel van woning en bewoner - Analyse van de module Energie WoON 2012*. 2013.

[97] Eugénio Rodrigues, Ana Rita Amaral, Adélio Rodrigues Gaspar, and Álvaro Gomes. How reliable are geometry-based building indices as thermal performance indicators? *Energy Conversion and Management*, 101:561–578, 2015. ISSN 01968904. doi: 10.1016/j.enconman.2015.06.011.

[98] Jon Herman and Will Usher. SALib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9), jan 2017. doi: 10.21105/joss.00097. URL https://doi.org/10.21105/joss.00097.

A ATTRIBUTE TABLES

Table A.1: Attribute table of BAG.pand

Attribute	Attribute (raw)	Database	Description
identification	identificatie	v	Building identification
construction_year	bouwjaar	v	-
status	status	v	-
surface_min	oppervlakte_min	-	-
surface_max	oppervlakte_max	-	-
number_of_objects	aantal_verblijfsobjecten	v	Number of addresses in the building

v: Attribute is required and stored in the database
-: Self explanatory term

Table A.2: Attribute table of BAG.verblijfsobject

Attribute	Attribute (raw)	Database	Description
identification	identificatie	v	Address identification
surface	oppervlakte	v	-
status	status	v	-
function	gebruiksdoel	v	-
public_space	openbare ruimte	-	-
house_number	huisnummer	v	-
addition	toevoeging	v	-
postcode	postcode	v	-
residence	woonplaats	-	e.g. Amsterdam
construction_year	bouwjaar	v	-
building_identification	pandidentificatie	v	The corresponding building contains this address
building_status	pandstatus	v	-

v: Attribute is required and stored in the database
-: Self explanatory term

Table A.3: Attribute table of 2014 CBS gas consumption data

Attribute	Attribute (raw)	Database	Description
gid	gid	-	-
Avg house consumption	Gemiddelde aardgaslevering woningen	-	-
PC6 total consumption	Totale aardgaslevering woningen	-	-
Number of supplied households	Aantal toegewezen leveringsadressen aardgas woningen	-	-
postcode	postcode	-	-

v: Attribute is required and stored in the database
-: Self explanatory term

Table A.4: Attribute table of Liander energy data

Attribute	Attribute (raw)	Database	Description
NETWORK OPERATOR	NETBEHEERDER	-	The EAN-code of the regional network operator
NETWORK_AREA	NETGEBIED	-	-
STREET_NAME	STRAATNAAM	-	-
POSTCODE_FROM	POSTCODE_VAN	v	The postal code in the format of 4 digits and 2 letters
POSTCODE_TO	POSTCODE_TOT	v	Merged postcode is presented if less than 10 connections are existed within the postcode_from area. Otherwise, it remains the same
RESIDENCE	WOONPLAATS	-	-
COUNTRY_CODE	LANDCODE	-	-
PRODUCT_TYPE	PRODUCTSOORT	v	ELK (electricity) or GAS (gas)
CONSUMPTION SEGMENT	VERBRUIKSSEGMENT	-	Small consumption: connection value of electricity connection is not greater than 3x80 ampere and gas not greater than G25
CONNECTIONS NUMBER	AANSLUITINGEN AANTAL	v	The number of connections in the relevant postcode area for the relevant energy type
DELIVERY DIRECTION PERC	LEVERINGSRICHTING PERC	v	The percentage gets lower if more return delivery takes place (e.g. via self generation)
PHYSICAL_STATUS PERC	FYSIEKE_STATUS PERC	v	The percentage of connections that are in operation
TYPE_INERATION PERC	SOORT_AANSLUITING PERC	-	Indicating connection capacity
TYPE_OF CONNECTION	SOORT_AANSLUITING	-	Name of the most common connection type
SJV_GEMIDDELD	SJV_GEMIDDELD	v	The average annual netted consumption of electricity (kWh) and gas (m^3)
SJV_LAAG TARIEF_PERC	SJV_LAAG TARIEF_PERC	-	The percentage of connections with a double rate
SMART_METER PERC	SLIMME_METER PERC	-	The percentage of smart meter

v: Attribute is required and stored in the database
-: Self explanatory term

Table A.5: Attribute table of CBS postcode 6 data

Attribute	Attribute (raw)	Database	Description
postcode 6	pc6	v	e.g. 1094 NA
municipality code	gemeentecode	-	-
Oad2010	Oad2010	-	-
Sted	Sted	-	-
population	aantal_inwoners	v	Total population within the postcode
number_men	aantal_mannen	-	-
number_women	aantal_vrouwen	-	-
perc_00_14	perc_00_14	-	Percentage of population within these ages
perc_15_24	perc_15_24	-	-
perc_25_44	perc_25_44	-	-
perc_45_64	perc_45_64	-	-
perc_65_74	perc_65_74	-	-
perc_75_older	perc_75_older	-	-
non-Western immigrants	nietwestersallochtoon	-	-
gem household size	gemhuishoudensgrootte	-	-
single-person household perc	eenpersoonshuishoudenperc	-	-
single-parent household perc	eenouderhuishoudenperc	-	-
morespecialchildren perc	meerpzonderkinderen perc	-	-
two-parent household_perc	tweeouderhuishouden_perc	-	-
home	woningvrd	-	-
house value	gemwoningwaarde	-	-
number_parthh	aantal_parthh	-	-
Income recipients	Inkomensontvangers	-	-
Low income	Laaginkomen	-	-
High income	Hooginkomen	-	-
Benefit recipients	Uitkeringsontvangers	-	-
Self-employed	Zelfstandigen	-	-
Fiscal monthly income	Fiscaalmaandinkomen	-	-

v: Attribute is required and stored in the database
-: Self explanatory term

B U-VALUES FROM DIFFERENT DATA SOURCES

Table B.1: Construction U-values in different construction periods from [6]

Parameters	Pre 1965	1965-1974	1975-1991	1992-2005	2006-2014	Post 2014
Uroof	1.68	0.89	0.64	0.36	0.23	0.16
Uwall	1.76	1.45	0.64	0.36	0.27	0.21
Ufloor	1.75	2.09	0.94	0.35	0.27	0.27
Uwindow	2.90	2.90	2.90	1.80	1.80	1.80

Table B.2: Construction U-values in different construction periods from [8]

Parameters	Pre 1945	1945-69	1970-79	1980-89	1990-99	2000-2010	Post 2010
Uroof	2.6	1.8	1.1	0.6	0.6	0.4	-
Uwall	2.55	2.15	1.9	1.38	1.23	0.8	-
Ufloor	2	1.7	1.5	1	0.9	0.4	-
Uwindow	3.8	3.4	3.4	3.4	2.9	2	-

Table B.3: Construction U-values in different construction periods from [9]

Parameters	Pre 1975	1975-90	1991-2002	2003-2006	Post-2006
Uroof	1.5	0.5	0.4	0.25	0.23
Uwall	1.5	1	0.5	0.41	0.38
Ufloor	1.2	0.8	0.5	0.44	0.41
Uwindow	3.5	3.5	2	1.84	1.68

C DATA INTEGRATION WORKFLOW

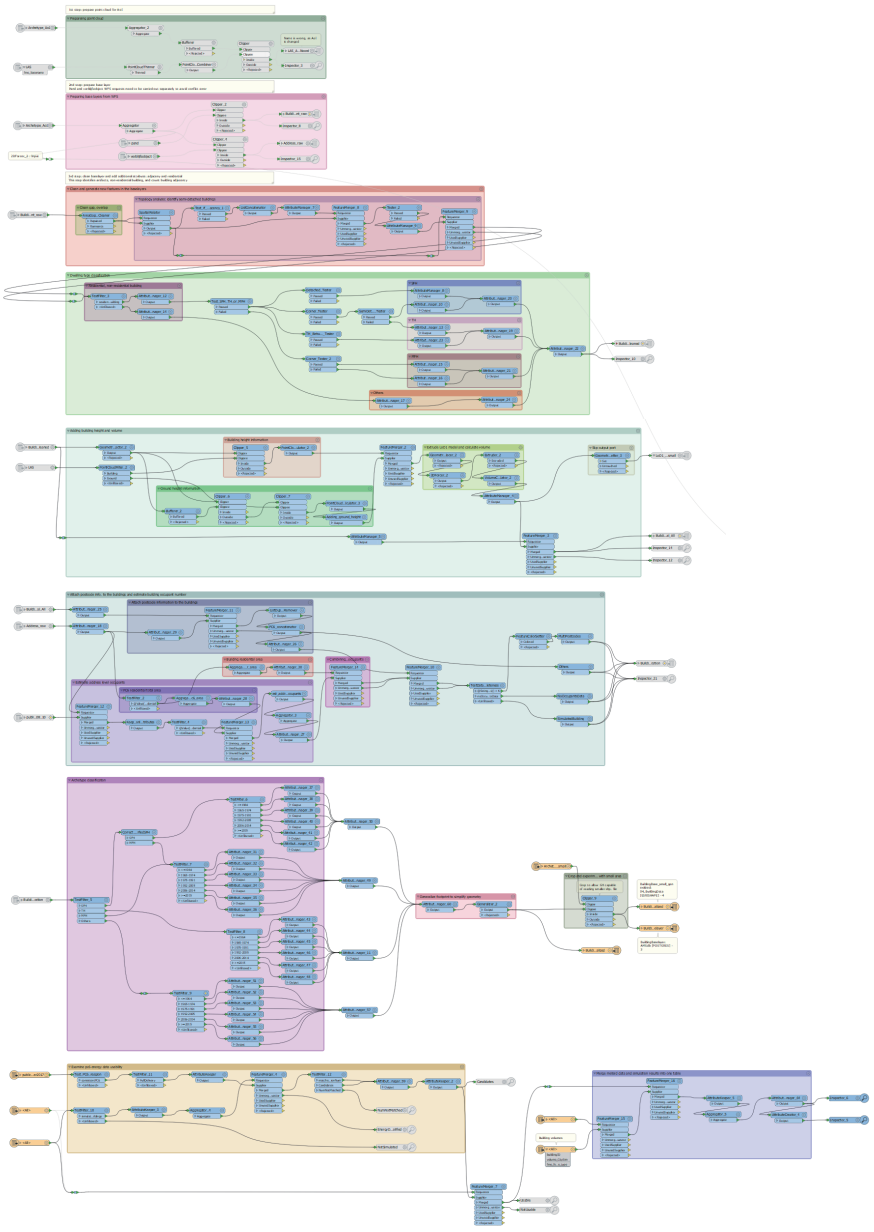


Figure C.1: Complete data integration workflow developed in FME platform

D GEOMETRY PREPARATION WORKFLOW

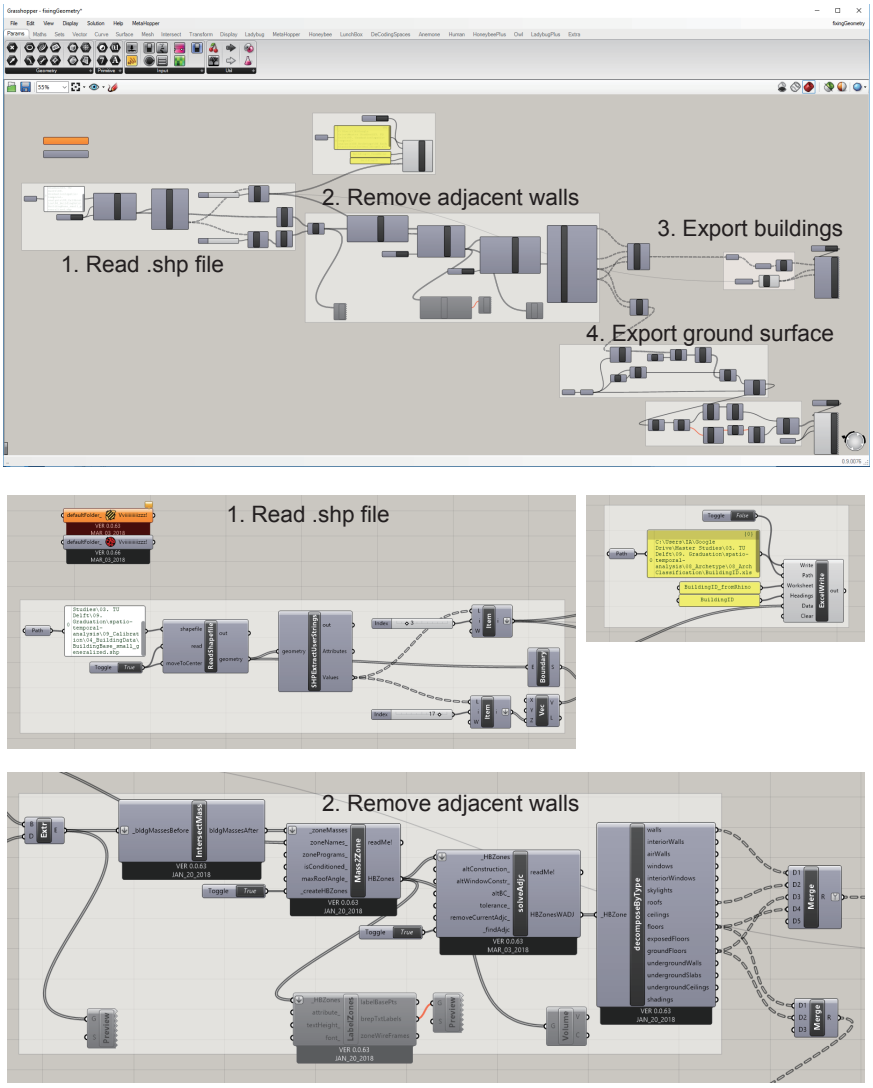


Figure D.1: Complete Grasshoper script developed to prepare a CitySim ready geometry from a .SHP file. Part 1

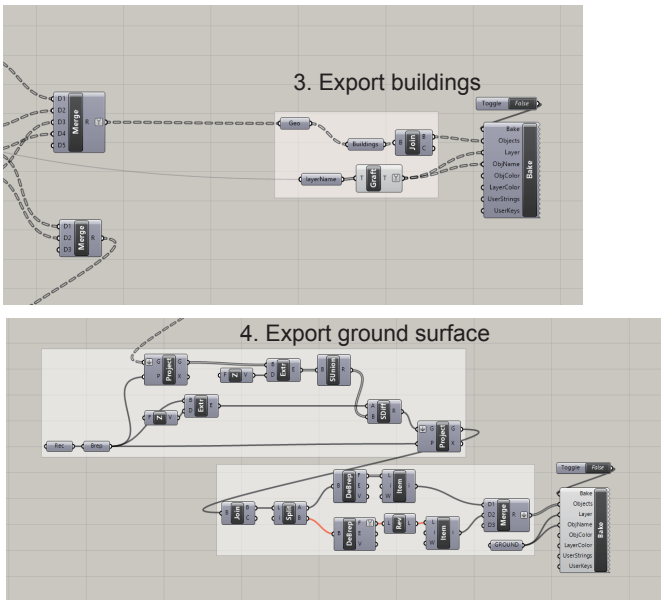


Figure D.2: Complete Grasshoper script developed to prepare CitySim ready geometry from a .SHP file. Part 2

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from André Miede.

TU Delft