Characterizing and Structuring Urban Data for Housing Stock Energy Modeling

SeyedehRabeeh HosseiniHaghighi

A Thesis

in

The Department

of

Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Building engineering) at

Concordia University

"Montreal, Quebec, Canada"

April 2021

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **SeyedehRabeeh HosseiniHaghighi**

Entitled: **Characterizing and Structuring Uban Data for Housing Stock Energy Modeling**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Building engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

Dr. F. Nasiri.. Chair

Dr. N. Bouguila...CIISE Examiner

Dr. F. Nasiri..BCEE Examiner

Dr. M. Ouf...BCEE Examiner

Dr. U. Eicker..Supervisor

Approved by

Dr. Ashutosh Bagchi, Chair
Department of Building, Civil and Environmental Engineering

Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

April 12, 2021

Abstract

Characterizing and Structuring Urban Data for Housing Stock Energy Modeling

By: SeyedehRabeeh HosseiniHaghighi

Setting Canada on a net-zero emission pathway by 2050 encouraged the current research to identify a workflow to focus on housing stock energy modeling by addressing the gap between the accessible urban datasets and the parameters required for building energy performance simulation. Urban building energy modeling (UBEM) supports policymaking for retrofitting buildings toward location-specific, low energy, and low carbon possibilities. The study deals with UBEM data management based on real urban datasets and characterizing archetype profiles to develop a bottom-up model capable of evaluating district energy scenarios. The different origins of datasets from multiple sources and scales required using a range of tools for data processing and analysis to synchronize with the UBEM specification and the CityGML geometry standard. The simulation platform SimStadt was used to integrate the city model with the developed archetypes in eight identified vintages for low-rise housing and estimate the district heating load.

A case study was carried out for the city Kelowna in BC /Canada. The average simulated heating energy use intensity showed almost 10% deviation from the national measured data in British Columbia. Furthermore, the model calculated the equivalent CO2 emission as well as the district's potential for on-site power generation combining photovoltaics with demand-side management. On the urban data side, the key barriers were the low compatibility between different urban datasets, lack of building identifiers shared between documentation conveying building information and deriving urban building geometry from data sources with low accuracies. Concerning the building energy input, the availability of characterized archetypes and local metered data would enhance the developed workflow and outputs toward effective energy policy and bottom-up retrofit planning.

Acknowledgments

I would like to express my deepest gratitude to my research advisor Ursula Eicker for her guidance, patience, kindness, and inspiration.

To all the talented CERC researchers and students who have shared their time and knowledge generously, especially Raman, Rushi, Pilar, and Guillermo, patiently helped me bring this work to an end. I appreciate my colleague in the Canadian End-Use Mapping Supplementary project, Sanam and Karthik, for their good teamwork and helpful discussion. Thanks to my brilliant friends in the CERC group that made the past two years memorable, and to Jurgen, whose spirit is still here in the CERC team.

Lastly, deepest of thanks to my mother and my husband for their endless support. And to my little son, Hafez, without whom I wouldn't be strong enough for all the magic.

Contents

List of Tables

List of Figures

Chapter 1

1. Introduction

1.1. Motivation

The five different sectors within Canada: residential, commercial/institutional, industrial, transportation, and agriculture industries used 8,786.4 PJ in 2016 [1]. The residential sector has been responsible for 16.6% of energy consumption and 12.9% of GHG emissions, as presented in Figure 1. However, from 1990 to 2016, the residential sector has successfully reduced 30.2 Mt CO2 (27% of total) by enhancing building codes, applying minimum energy performance standards for appliances, improving energy monitoring systems and home retrofits. Under the Paris Agreement, Canada committed to reducing its GHG emissions up to 30% below the 2005 level by 2030. Moreover, the government announced a plan to set Canada on a net-zero emissions pathway by 2050 [2]. Canada's 2030 GHG emissions target is 511 Mt CO2 eq, compared to the 2015 level at 815 Mt CO2 eq. Of the nine principal categories, buildings are committed to 47 Mt CO2 eq reduction demonstrated in Figure 2. The key priorities come from efforts to increase clean electricity, greener buildings, communities, and nature-based climate solutions[2].

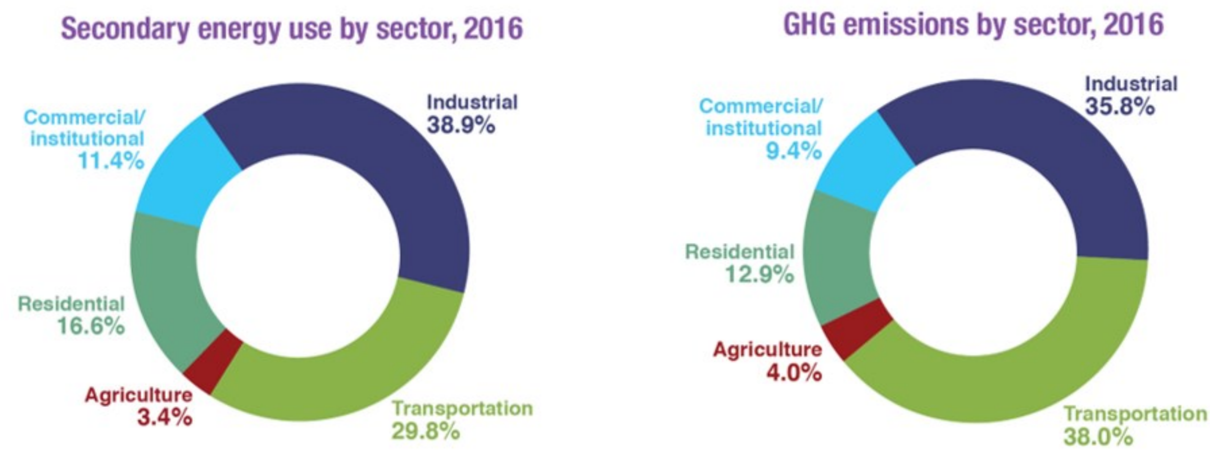


Figure 1: Energy use and GHG emission by five primary sectors in Canada, 2016 [1]

Residential building energy use resulting from stock growth has had a minimum increase of only 2% from 1990 to 2016, see Figure 3. It seems a low increase over the long lifespan of residential buildings, and consequently, there is less potential for energy saving in developing strategies for new buildings in the short- and medium-term. However, existing residential buildings have significant potential for maximizing energy-saving and transition to the low-carbon milestone. Urban building energy modeling (UBEM) provides a robust mechanism to approach the large-scale building energy modeling and evaluation to enable cities for retrofit planning and GHG mitigation. It also allows capturing the benefits of community-scale opportunities for raising energy renovation rates through the district photovoltaic (P.V.) systems or district heating systems.

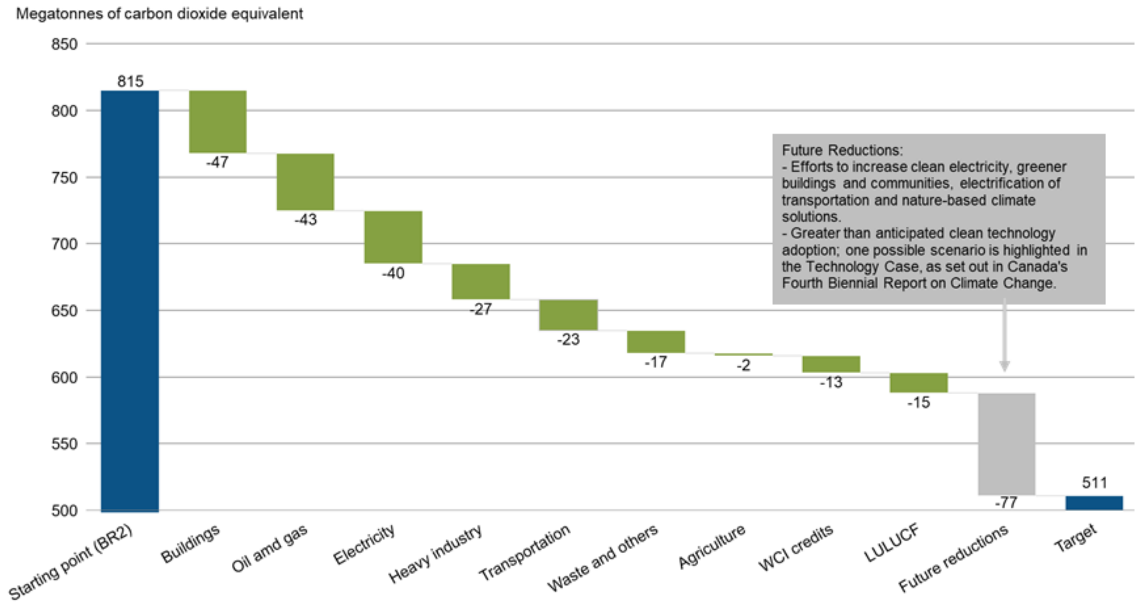


Figure 2: Contribution of different sectors in projected Mt CO2 eq reduction in 2030 relative to 2015 [2]

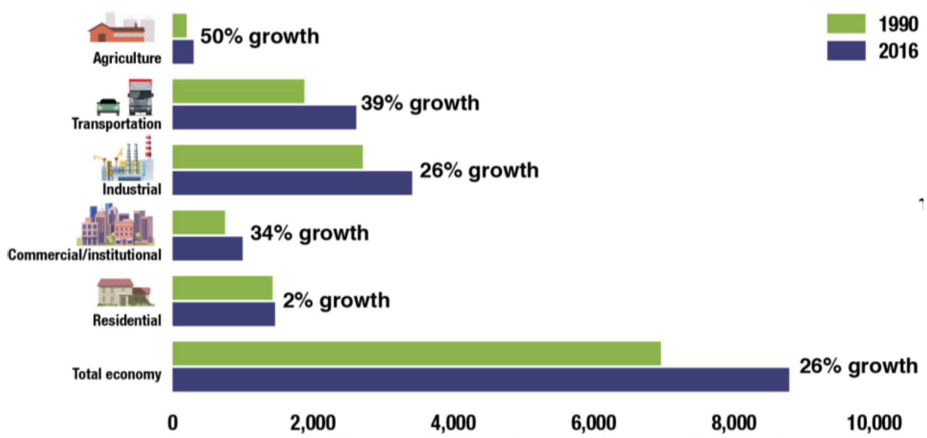


Figure 3: Total energy use growth by sector(petajoules), 1990 vs. 2016 [1]

Annex 70 is a worldwide collaboration between academics, industry, and governments to develop pathways toward improving energy demand in the building stock. It focuses on studying and modeling the building stock at scale and addresses the gap between available data and energy performance evaluation. It aims to facilitate the use of real energy data in policymaking and industry development toward geo-specific, low energy, and low carbon possibilities [3]. The domain of Annex 70, the investigation of housing and building energy data, and their energy performance and end-use spatial distribution in terms of supporting stakeholder groups, is the research concern of CanmetENERGY-Ottawa[1] titled "the Canadian Energy End-use Mapping" (CEE Map) project. The CEEMap project aims to identify a harmonized data incorporation, characterization, and energy performance evaluation, projected in the coordinate system and consistent with the online-authorized GIS platform. Preserving the coordinates of any data point is the priority of this project to access a contextual resilient, and climate responsive design.

Following the CEEMap approach, the primary step is understanding the building stock's data classification in segmentation factors such as age, type, floor area, and then grouping them into typical archetypes representing the majority of buildings in the urban area. On the other side, energy use is available through various sources and methods, including metered and simulated datasets. The recent advancement of CanmetENERGY-Ottawa lab in building energy modeling and simulation[2], Housing Technology Assessment Platform (HTAP), and Building Technology Assessment Platform (BTAP), has enabled the generation of the baseline end-use as well as providing a significant number of building optimization packages at varying cost increments on buildings across all Canadian climate zones respecting the building typology and year of construction. Integrating the developed archetypes and the evaluated results into an online map based on an authoritative decision platform makes data accessible and interactive for a broader domain of expertise, aiding policy development and decision making. The Building Energy Mapping and Analytics Concept Development Study (BEMA-CDS project)[3] uses the knowledge of CEEMap to assess and support the development of energy mapping and analytics to specify the retrofit-based efficiencies, renewables technologies, and environmental opportunities in the building stock. It is intended for geo-oriented energy planning to facilitate shifting to a low-carbon community using utility conservation potential and demand-side management programs at the building stock level. It also helps developing policy, standards, and codes at territorial, provincial, and federal levels [4].

The CEEMap approach has low dependency on the urban dataset by relying on the archetype identification from the target context and using their characteristics to find similar cases in the

[1] CanmetENERGY-Ottawa, an energy R&D division within the Energy Technology Sector of Natural Resources Canada (NRCan) has a mandate to lead the development of energy science and technology solutions for the environmental and economic benefit of Canadians.
[2] This data is developed in support of the National Building Code, energy efficiency programs and equipment standards led by NRCan's Office of Energy Efficiency.
[3] Another project is leading besides the CEEMAP project managed by Natural Resources Canada.

HTAP dataset. The output of HTAP simulation will be generalized for urban buildings based on the identified archetypes. This methodology does not maintain the GIS dataset and does not deal with the urban dataset issues and inaccuracies. However, it results in the inconsistency between the real building geometry and estimated performance. It causes the buildings' individual properties to become lost and off-track for efficient bottom-up intervention. It also results in missing the buildings' radiation interaction with the built environment, particularly in the dense urban structure.

Kelowna city, located in British Columbia/Canada, is the pilot case study to cover the prime concerns of the CEEMap project in terms of data preparation and residential stock characterization. This project, on which a section was done in collaboration with Concordia University, is the thesis's motivation to follow the end-use mapping while actively maintaining the GIS datasets. The thesis aims to show the challenges and workflows associated with the preparation of urban datasets for integrated, data-driven urban building energy modeling and mapping.

1.2. Thesis Goals and Outline

This research aims to address the data preparation required for integrated, data-driven UBEM simulation and mapping. It addresses the limitation and inconsistency of urban datasets using an improved bottom-up approach that employs various data-analysis operations to reduce the data insufficiencies. It explores the extent to which a UBEM can be refined when limited urban data is available and the resolution of input data is not enough. The procedural objectives are:

1. Demonstrating the urban data challenges and issues for UBEM purposes
2. Presenting an urban data-based methodology for the creation and refinement of archetype and building data modeling for an integrated, data-driven UBEM application
3. Estimating and mapping urban building energy demand

Chapter two has two parts. The first provides a background on energy mapping experiences based on different approaches using metered data analysis or bottom-up UBEM simulation. The second part defines the UBEM data requirements and structure. Chapter three discusses the methodology centering on the case study. It challenges the available dataset's complexity and processes required for building data preparation. Chapter four consolidates the data preparation by mapping, characterizing, and modeling the processed building datasets and archetypes. It also calculates the thermal load and CO2eq for a selected region to evaluate the urban building energy performance and discusses the output validation and occurrence of the alternative scenarios based on the roof surfaces for potential photovoltaic power. Chapter five highlights the limitations, recommendations for improvement, the research's primary outcomes, and future work suggestions.

Chapter 2

2. Urban data analysis and integration

2.1. Urban data energy modeling and mapping experiences

The reference building-stock energy mapping allows a deep understanding of building energy performance versus cities' social, economic, physical, and environmental characteristics. This topic has been the concern of many studies in Spain [5], Greek [6], Italy [7], Sweden [8], Ireland [9], and the U.S. cities such as San Francisco [10] and Boston [11]. An intent to support urban energy systems, policies, environmental purposes, as well as access to substantial open data portals has motivated cities to model the spatiotemporal pattern of energy consumers. Gangolells et al. (2015) used the Energy Performance Certificate (EPC) database to map and analyze the distribution of building stock energy performance in the north of Spain. EPC is a rating system to indicate the energy efficiency of European buildings. It rates the building between A (Very efficient) to G (Inefficient), mandating buildings for update every ten years. Spain's EPC dataset evaluates the carbon dioxide (CO_2) emissions per square meter generated by buildings throughout a year. The energy mapping of the existing buildings gave a clear insight into the building stock distribution and energy performance based on the building type, construction period, and specified end-uses. This provides governments with the potential to better direct future energy policies related to refurbishment, including funding mechanisms and future revisions of building regulations and codes [5]. However, the dataset demands improvement in considering embodied energy as a critical factor for measuring building energy performance and geospatial coordination for visualized data mapping and confronting them with other urban datasets.

Hjortling et al. (2017) used the EPC dataset to map and analyze the baseline energy end-use of commercial buildings in Sweden. The dataset provides a useful classification of the buildings based on the use-type, construction period, climate zones, and energy use [8]. This study provides an understanding of Swedish building stock and identifies valuable insights into energy investment potential within various building types. As a future challenge, the research asks for an EPC labeling system based on carbon footprint to measure the environmental effect of building energy performance. Also, it emphasizes the importance of the data record's geo-spatializing to empower the EPC analysis to meet the multidisciplinary urban dataset for improving the energy-related policy and practices.

Delmastro et al. (2017) represented a GIS-based urban building energy model aiming to identify cost-optimized renovation practices and prioritizes building stock considering the socio-economic feasibility [7]. The research leveraged the combination of census data, real energy data from utility bills, archetype dataset extracting from the TABULA [12] project, and characterized geometry data using the GIS datasets. In data preparation, the integration of census and geometry data provided the initial condition for archetype definition by categorizing the segmentation factors such as the building shape, age, and climate zone. The real monitored end-use dataset's availability played a critical role in filling the gap between the simulated dataset and the actual consumption reducing the dataset's inconsistency. However, the lack of sufficient building energy data led to applying the same fuel type for all facilities. The research procedure was designed to support decision-makers in local energy planning by providing a methodological approach for analyzing the impact of future building stock retrofitting scenarios and moving toward a suitable policy from an energetic perspective. Integrating the urban dataset in a geo-coordinated UBEM allowed them to review the possibility of a detailed intervention scenario from a bottom-up perspective on the building level and a top-down vision with providing suggestions for improving urban policy investment.

The building stock in Dublin city has also been energetically mapped on a district scale using EPC and urban datasets on a GIS-based bottom-up UBEM [9]. By relying on GIS datasets, census surveys, and the national level building stocks dataset, they defined a model for processing geometry data and determined a customized archetype categorized by the building typology and year of construction across the region. The customized archetype was validated against the TABULA library for Ireland housing and apartment. The simulated demands were verified in dealing with the energy report and surveys for various archetypes. The highlighted difficulty of data processing came to reality by associating the significant number of buildings and entities in multiple datasets with multiple structures, standards, and deterministic identifiers. The hierarchy of data preparation, integration, simulation, and validation of the study is demonstrated in Figure 4.

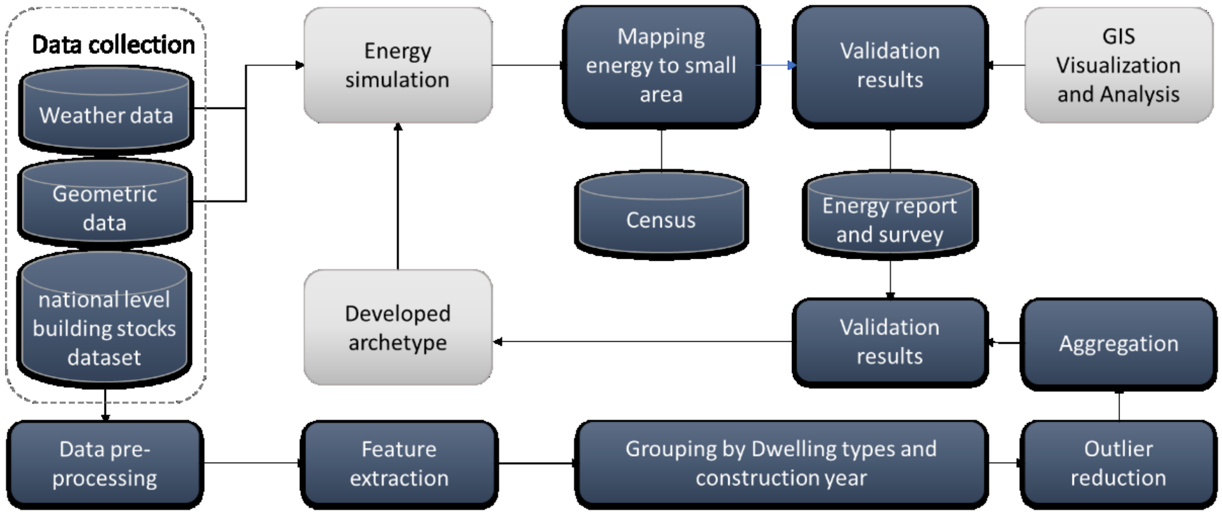


Figure 4: Data-processing for archetype development and GIS-based residential building energy modeling at the district scale [9]

Quan et al. (2016) also used the GIS platform for UBEM and Mapping, relying on the Urban-EPC simulation engine, a modified Energy Performance Calculator engine [13]. Their model uses a comprehensive urban dataset in various structures and formats, including weather data, geometry, topographic, and socio-demographic plus the U.S. reference building archetype to support the low level of building data and missing energy attributes. Figure 5 illustrates the presence, distribution, and interconnection of various datasets. However, they required measured data to verify the model's accuracy against the missing parameters filled with the supplementary information and assumption. Also, the experience of data preparation, modeling, and simulation on the GIS-based UBEM demonstrated the tool's low ability in dealing with detailed data processing and simulation on a large scale.

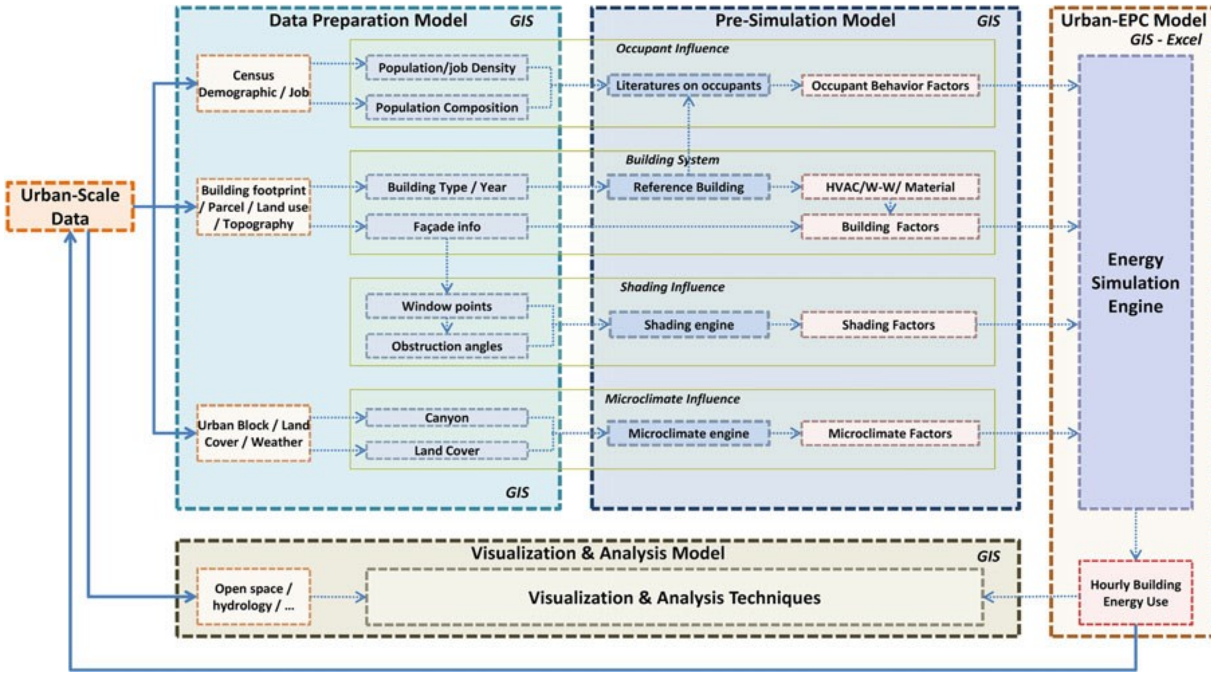


Figure 5: Structure and data flow of the urban building energy system [13]

In collaboration with the Boston Redevelopment Authority, Davila et al. (2016) developed a UBEM using the urban databases, including tax parcels, building footprint, tax record lite and full, and a customized-building archetype prepared regarding the building age and use-type [11]. The model was established on the Rhinoceros platform for the simplified 3D model generation of data and assigning the XML version of archetypes to each building, presented in Figure 6. Using the EnergyPlus Weather (EPW) dataset, the model was simulated through the EnergyPlus engine. Data preparation wise, the lack of a systematic and standard way for data identification and description for various urban datasets resulted in difficulties for interconnection and operation of datasets to provide a consistent building database. The inadequacies of widely available archetype templates and measured energy data were the other barriers for UBEM data development and validation. Therefore, they used the simulated U.S. reference building prototype models [14] to verify the archetypes and results. Besides, they recommended a more automated model for data conversion and visualization better equipped for more complicated urban geometry modeling and processing.

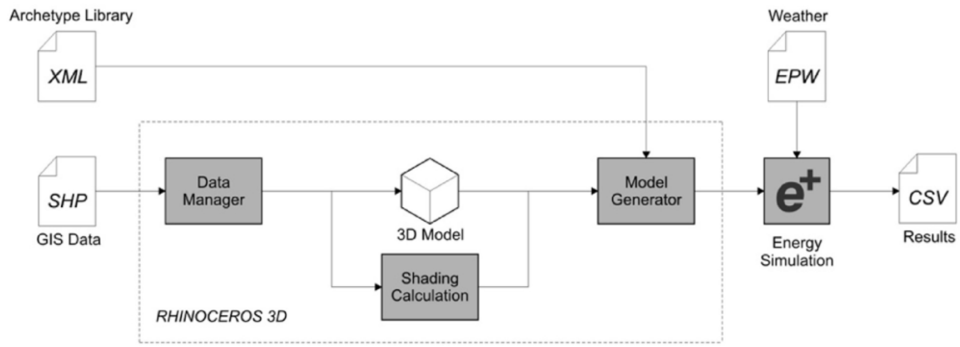


Figure 6: Urban building energy modeling workflow [11]

Focusing on urban data preparation to serving UBEM and mapping, Chen et al. aimed to define the minimum data requirement and introduce a standard way of the data structure for UBEM in 2018 [10]. They used the GIS-based urban datasets, including footprint, land use, and assessor data, to integrate various data sources in a unified master building dataset presented in Figure 7. The processed data is specified with the standard terminology using the building energy data exchange specification (BEDES [15]), modeled following the CityGML structure as the standard format of visualizing urban objects and representing the topological relation of spatial data. Using the U.S. reference building prototype model [14], the archetypes were characterized, respecting the CityGML Energy ADE (Application Domain Extension) specification as the standard extension of the CityGML model.

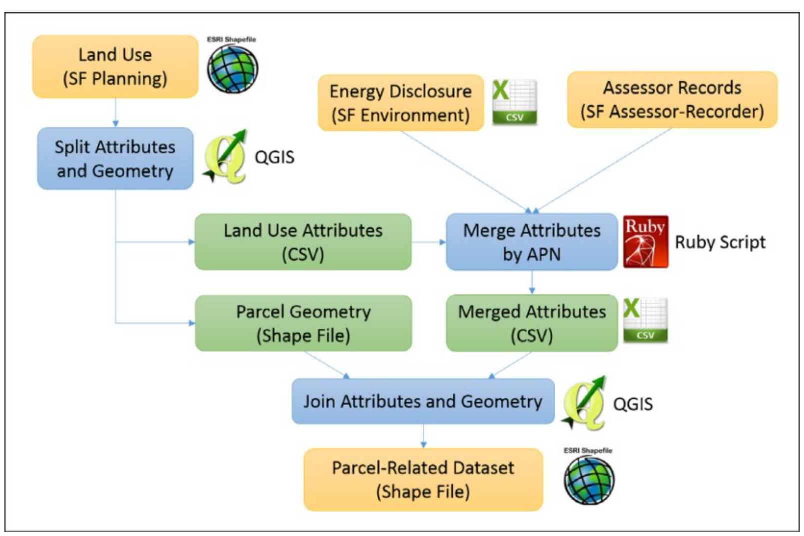


Figure 7: The workflow of parcel-related dataset consolidation[10]

Building energy mapping using bottom-up UBEM has multiscale data preparation and integration, as presented in Figure 8. On one side, it needs big data processing to provide a specified profile for individual buildings with geometry and segmentation data. On the other side, it seeks detailed energy information to embody the building energy data. Data understanding and preparation for UBEM are fundamental steps in monitoring urban system performance across time and space and discovering the pattern of energy resource consumption and GHG emission [16]. City-data portals are a significant source of UBEM data collection. However, the data is not stacked up with UBEM development in mind, and thus, it is subject to substantial amounts of missing data, inconsistencies, and inaccuracies. The prevailing reasons are the lack of standard data-gathering methodology, characteristics terminology, and characterized identifiers shared between urban datasets. Also, the large volume of datasets collected with different tools affects the dataset resolution. An example is extracting the building height from the Kelowna orthophoto data that supports only the maximum height.

There is no agreed method to improve the dataset's accuracy and consistency. Many efforts showed the possibility of reducing issues and increasing the capability of urban datasets ([7],[9],[11],[13]). However, the geospatial characteristics of urban datasets and GIS-based tools provide multi-operational platforms for data processing and integration to synchronized urban data with UBEM requirements. On developing the building energy profile, most of the studies are satisfied using the archetype modeling based on the building stock information. Cases having metered data or energy performance certificates could successfully close the gap between the designed/simulated model and the buildings' real energy performance and approach a more realistic simulation model. In the absence of local monitored data, the models have leveraged the national and regional measured datasets either for archetype development or result validation.

This multiscale data provision and connection allows for two-level benefits of intervention, bottom-up and top-down. The bottom-up level increases the potential of an effective retrofitting program, building code upgrade, and new technology assessment. The top-down draws up the opportunities of district-scale technologies, such as community-scale photovoltaic systems or combined heat and power (CHP) systems [10], besides increasing the possibility of urban policy management and development. Table 1 shows the classification of datasets and methodologies discussed in the urban data energy modeling and mapping experiences

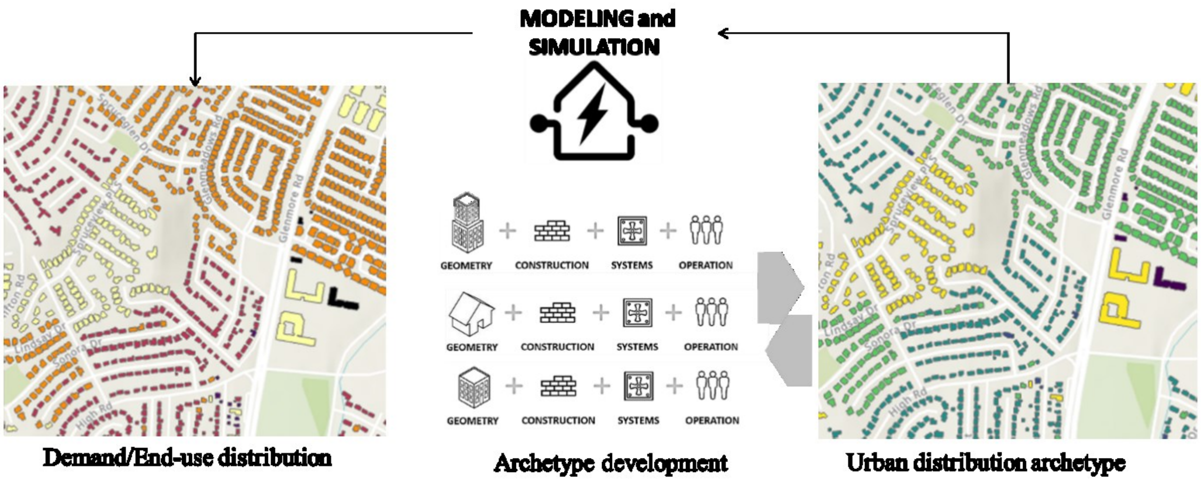


Figure 8: Multi-Scale Approach of Urban Building Energy Modeling and Mapping

Table 1: The review of peer-reviewed papers in the field of urban building energy mapping

City/Country	scale	Dataset accessibility	Methodology	UBEM/ visualization Platform	Energy Metered dataset	Urban dataset	Archetype	Limitation on data preparation and mapping
Spain [5]	Region	Depends on region	Statistical Analysis	No/No	Yes	No	No	No visualization on a GIS map, Dataset inconsistency
Sweden [8]	Region	Access free	Statistical Analysis	No/No	Yes	No	No	No visualization on the GIS map
Greece [6]	Region	No Access	Statistical Analysis	No/No	Yes	Census survey	TABULA library	No visualization on a GIS map, Inadequate data on the existing building stock, Little number of registered buildings for EPC
The U.S./ San Francisco [10]	City	Access free	Data Modeling/ Mapping	No/GIS	Yes	Building Footprint, Land Use, Accessory	The U.S. reference building archetype	Lack of standard for data description and building identification, urban dataset inaccuracy, inconsistency between building a dataset and supplementary urban dataset, Limited information about building characteristics in the urban dataset, static features of the urban dataset
The U.S./ Boston [11]	Neighbor hood	Access free	Simulation	Rhinoceros/No	No	Tax Parcels, Building footprint, Tax Record Lite, Tax Record Full	The U.S. reference building archetype/Customized model	Lack of widely available archetype templates, lack of metered energy data, lack of projected coordinate system in the model and no possibility of visualization on the GIS map, lack of using standardized terminology for urban data identification, description and documentation, geometry simplification, inconsistency between building a dataset and supplementary urban dataset

The U.S./ New York [13]	Neighbor hood	Access free	Simulation/ Mapping	GIS/GIS		Building footprint, Land use, Topography data, Population distribution data, Job distribution data, Tree canopy	The U.S. reference building archetype	Inconsistency between building dataset and supplementary urban dataset, low-level detail of building information in urban datasets, missing data
Italy [7]	City	Depends on region	Simulation/ Mapping	GIS/GIS	Yes	Census survey, Footprint outline	TABULA library	low-level detail of building information in urban datasets, data inconsistency
Ireland/ Dublin [9]	City	Access free	Simulation/ Mapping	GIS/GIS	Yes	GIS dataset, Census surveys data	TABULA library/ Customized model	lack of using standardized way for urban data identification, inconsistency between building dataset and supplementary urban dataset

2.2. Urban building energy model; data requirements

Weather and building datasets are two principal categories of urban building energy model(UBEM) data requirements [17]. As a critical dataset available in different resolutions, the weather file relies on the historical data for recent decades. Typical Meteorological Year 3 (TMY3) [18] supporting local climate and EnergyPlus Weather format (EPW) [19] considering the local and climate change effects are two widely used weather datasets provided at hourly timesteps and cover many locations in the world. The other efforts are ongoing to focus the linkage of microclimate and heat-island effect on the building energy modeling[20],[21],[22],[23]. Building datasets are divided into two categories, geometry and non-geometry parameters. The geometry dataset represents the building characteristics for modeling. It is mostly accessible in the typical two-dimensional GIS data format of Shapefile (SHP) or GeoJSON on the city scale. The primary urban datasets consist of building footprints and parcel data with additional complementary attributes attached as DataBase table Format (DBF) or accessible in Comma Separated Value (CSV) and other supplementary datasets, such as census survey, property tax assessment, or building permit issuance.

Table 2 demonstrates the available urban datasets in the city open-data portals of three Canadian metropolitan areas, which are applicable for supporting building datasets. The building footprint is generally provided with the height attributes appropriate for generating 3D urban models. As an alternative, light image detection and ranging (LiDAR) is another data source for building heights and geometry creation with a higher accuracy level. The tax property layout typically offers considerable information about buildings at parcel levels, such as building floor area, number of bedrooms, year of construction, number of stories, building typology, zoning class, population, and properties' financial values. The building permit is another critical data source reflecting the building retrofitting, refurbishment, and renovation over the building lifecycle. The information could be embedded as attributes on the building footprint through GIS techniques such as spatial join or attribute join. Interoperability between software environments is essential at this stage since datasets originate from different sources in various formats and scales. However, using multiple identifiers in datasets and assigning appropriate entities to buildings is challenging. It sometimes hinders providing a consistent and accurate dataset. For example, the urban building data gathering is typically aggregated at the parcel scale. Therefore, the unique identifier targets the parcel, not the buildings shared in the lots (parcels), and attachment of information in some cases results in misleading building formation. In the other case related to geometry generation, the building footprint outline is extracted from orthophoto, Cadastre, or LiDAR data. In these cases, the attached buildings are derived as an aggregated geometry, while the permit or tax data gathering is based on ownership division, reducing the accuracy of the building data gathered.

The non-geometry dataset includes construction data, usage systems, and schedules [17]. When structuring an existing building energy model, these characteristics are typically gathered by reviewing constructional and mechanical properties and an individual energy audit [17]. This procedure is impractical for an urban model with a considerable number of buildings. Hence, energy modelers resort to a generalization level called archetype modeling. The usual way to develop the archetype is separating the building stock into homogeneous groups based on either geometrical or segmentation data and assign the same building properties to the group buildings. This method assumes that specific building properties could represent the energy use variation between groups of buildings and display the group's prototype model. A notable effort to generate country-wide archetypes for 25 nations is the European research project TABULA [12], which targets residential buildings. Parekh (2005) [25] has focused on housing stock characterization in Canada. In other countries such as Japan [24] United Kingdom [25], and the United States [26], the national archetype has been provided for both residential and commercial building stock. The most often used urban segmentation data to differentiate buildings by the group/archetype include age (most common), shape and size, use type, structure types, and heating system[9][17].

Table 2: The variety of urban dataset conducting building information and useful for UBEM data supporting in three Canadian metropolitans

Open data portal/City	Dataset title	Shape format	Available format	Relevant Attributes	Scale of availability	Address
Montreal	Building	Polygon/ Point	SHP/GeoJSON	Object ID/ Footprint area/elevation	Building	[27]
Montreal	Construction, transformation, and demolition permits	Point	SHP/ GeoJSON/CSV	Permit ID/Applied and finished date/permit type/ work description/ address	Building	[28]
Vancouver	Building footprints 2009	Polygon	GeoJSON/KML CSV/JSON/EXCEL	Object ID/ Footprint area/roof type/min,max,avg elevation	Building	[29]
Vancouver	Property parcel polygons	Na	GeoJSON/KML CSV/JSON/EXCEL	Object ID /Civic number/Tax Coord	Lot / parcel	[30]
Vancouver	Issued building permits	Na	CSV/JSON/EXCEL	Permit ID/Applied date/ work description/ Construction type/address	Building	[31]
Vancouver	Property tax report	Na	CSV/JSON/EXCEL	Parcel id/legal type/zone category/address/year built	Building	[32]
Toronto	Building Outlines	Polygon	SHP/DEM	Object ID /Height/footprint area	Building	[33]
Toronto	Property Boundaries	Polygon	SHP	No access	Lot / parcel	[34]
Toronto	Building Permits		CSV/JSON/EXCEL/XML	Permit id/Date/ work type/description/address/status/ construction type/	Building	[35]

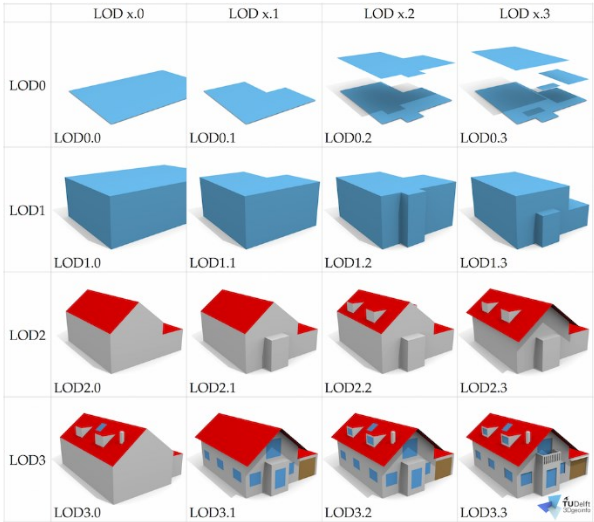
The object ID is a unique general id that GIS automatically assigns to each data point.

Addressing the error resulting from urban data modeling in predicting consumption, in most cases, relies on comparing the measured energy demand with the aggregated summation of the simulated annual energy use for an entire neighborhood or city. The reported low errors for the aggregated models have been quantified between 4% to 21%[11], 1% to 15% [17], and 1% to 19% [36]. However, it does not mean the individual buildings in the test set conform to the aggregated error. It might reach 99% per single-structure[17]. Nouvel et al. 2017 have measured the differences for individual buildings between 2% – 30% in the district level [37], Fonseca et al. 2015 reached 4-66% [36] by scaling down to single buildings in an urban block. Davila et al. 2016 found 5% to 94%[11] when modeling buildings on a city-scale.

2.2.1. The geometry data structure for UBEM

Shapefile/FileGDB and GeoJSON in 2D and CityGML in 3D are in data modelers' priority to support the geometry dataset. Shapefile (SHP) is developed and regulated by ESRI as an open specification for ESRI and other GIS software products' data interoperability [38]. It is a digital vector for storing geometric location and associated attributes in DBF format. However, it cannot store topological information. A 2D dataset is applicable in primitive geometric shapes like points, lines, and polygons, such as building footprint or lot parcel carrying the complementary attributes' table assigned to each object [39]. GeoJSON is a superseded GIS-format for SHP developed and maintained by a web developer group that follows JavaScript Object Notation (JSON) for encoding geographic data structures[40]. The GeoJSON data structure does not support a standard such as the ESRI data model. Also, the ArcGIS tool could not manipulate files in GeoJSON. GeoJSON is easy to handle but essentially designed to be loaded into the memory fully and at once. However, both GeoJSON and Shapefile could not support a schema to develop a constant dataset for the building properties.

CityGML is an open-source, XML-based data model representing 3D urban objects following the Open Geospatial Consortium (OGC) standard [41]. It supports classes and relations for the most relevant topographic features in city models concerning geometry, topology, semantic, and appearance properties [42]. The strength of CityGML for display and exchange digital (3D) models of cities and the capability for the combination of geometry and databases made it the file format of choice for many UBEM projects [43][44]. CityGML identifies urban objects, including building components, water bodies, streets, transportation properties, and vegetation, with a significant range of details from zero to four. In the case of buildings, the level of zero represents the gray layout of the outline. The first level is the extruded building outline with a flat roof, and the second level has the roof type. The third level covers the envelope details, such as windows and doors. Finally, the fourth level supports the building's full interior details. Figure 9 depicts the different LOD visualizations of a building and a wall sample code in CityGML format. Depending on the availability of urban data format, various possibilities are available to put urban datasets in CityGML format. One of the more accurate options is using LiDAR born point clouds dataset. By converting the points to a 3D model using ArcGIS or 3Dfire[45], it is conceivable to use Feature Manipulation Engine (FME) tool to convert the model to CityGML. Another option with a lower accuracy level is utilizing the building footprint and available building height or story number to generate the 3D city model using 3D modeling application such as ArcGIS CityEngine and then transform to CityGML using FME.



(a)

```
<bldg:boundedBy>
        <bldg:WallSurface gml:id="GML_d38cf762-c29d-4491-88c9-bdc89e141978">
          <gml:name>Outer Wall 2 (South)</gml:name>
          <bldg:lod2MultiSurface>
            <gml:MultiSurface>
              <gml:surfaceMember>
                <gml:Polygon gml:id="PolyID7351_1722_416019_316876">
                  <gml:exterior>
                    <gml:LinearRing gml:id="PolyID7351_1722_416019_316876_0">
                      <gml:pos>458889 5438353 3.43094010767585 </gml:pos>
                      <gml:pos>458877 5438353 3.43094010767585 </gml:pos>
                      <gml:pos>458877 5438353 -0.2 </gml:pos>
                      <gml:pos>458889 5438353 -0.2 </gml:pos>
                      <gml:pos>458889 5438353 3.43094010767585 </gml:pos>
                    </gml:LinearRing>
                  </gml:exterior>
                </gml:Polygon>
              </gml:surfaceMember>
            </gml:MultiSurface>
          </bldg:lod2MultiSurface>
        </bldg:WallSurface>
      </bldg:boundedBy>
      <bldg:boundedBy>
```

(b)

Figure 9: (a)An improved LOD specification for 3D building models [40], (b) A sample of CityGML code for the southern wall from FZKHouse LoD2[41]

2.2.2. Non-geometry data structure, Energy ADE

To develop the archetype module on the various urban scales, the Energy Application Domain Extension, Energy ADE [46], is the standard OGC approved extension for the CityGML following XML schema. Energy ADE stores the relevant energy-related input data assisting bottom-up energy assessment from a single building to a city-wide and regional scale. Energy ADE's design respects the buildings' data specifications prescribed within the INSPIRE Directive of the European Parliament (2007/2/CE). It also follows the concept defined by the U.S. Building Energy Data Exchange Specification (BEDES) concerning the building requirements such as usage, construction year, and the number of dwellings and residents [47]. Energy ADE enables data operations and dynamic simulations for urban energy applications following the energy balance methods comprised within the ISO 13790 standard. It is able to do so even at sub-hourly resolutions. The core of Energy ADE consists of four principal modules connected to the modeled building with the CityGML semantic features. It includes (i) the building physics supporting single- or multi thermal zone elements, (ii) the occupant's behavior modeling the behavior of occupants, (iii) the material and construction covering the physical properties of materials bounded the thermal zones, and (iv) the energy system representing the energy conversion, distribution, storage, and emission devices within a building and their associated energy flows[47][48]. The interaction of modules within the core of the building in CityGML is illustrated in Figure 10.

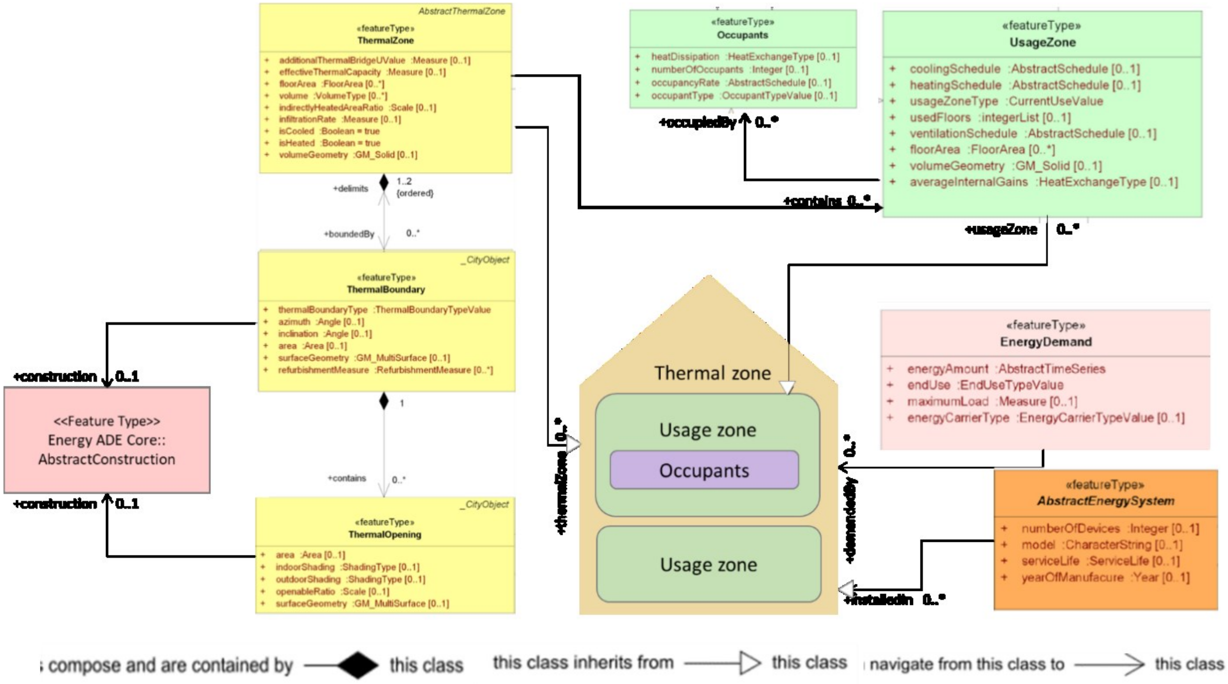


Figure 10: Principal Energy ADE modules interaction in a building core following CityGML format

The material and construction module enables modeling the building construction features using the physical, thermal, and radiative properties associated with the assemblies' materials. This module's specification establishes the roadmap of transferring the National Renewable Energy Laboratory (NREL) construction data, which has precise usage, to the Canada Excellence Research Chair[4] data repository, whereby it can be expanded for more effective use for all of North America. This module allows specifying the thermal transmittance (uValue) and optical properties (OpticalProperties) of the various thermal boundary types, such as interior, exterior, and shared walls, attic and flat roofs, intermediary floor, ground slab, basement ceiling, and fenestration in detail.

Within the module of material and construction presented in Figure 11, the option of the OpticalProperties is listed alongside the uValue. The OpticalProperties specifies the three radiative

[4] Canada Excellence Research Chair for Smart, Sustainable and Resilient Cities and Communities

characteristics of construction assemblies as separate classes: emissivity[5], reflectance[6], and transmittance[7], alongside the glazing ratio representing the ratio of transparent construction surfaces (against the total construction surface). The model takes the absorptance as being equal to the emittance following Kirchoff and Lambert's law. Therefore, it does not cover the absorptance as a separate attribute. The radiative characteristics of construction are classified into separated classes. Supporting information related to the factors such as the fraction of use, the collision surface side (inside/outside), and the wavelength range for three radiation beams, solar, visible, and infrared, are also often stored in the separate classes.

The thermal properties of construction are specified with two methods, either using the heat transmission coefficient (uValue) for steady-state thermal modeling or considering a list of materials whose properties are laid out in AbstractMaterial/ SolidMaterial class for dynamic modeling. SolidMaterial provides a unique profile for opaque materials covering thermal conductivity, density, and the specific heat coefficient. More physical characteristics of materials are also often supported like porosity, permeance, embodied carbon and embodied heat to measure building energy performance. However, they are not required for the thermal load calculation. LayerComponent is an intermediate class between SolidMaterial and Construction to manage the thickness of materials and their fraction of use within the surfaces as they occur. Materials such as air gaps that do not represent significant thermal capacity are characterized by their thermal resistance (R-Value) and essential information regarding ventilation. Figure 11 illustrates the UML diagram, and Figure 12, the XML sample of construction and material.

[5] The ratio of the long-wave radiation emitted by an object to the black body emittance
[6] fraction of incident radiation which is reflected by an object
[7] fraction of incident radiation which passes through an object

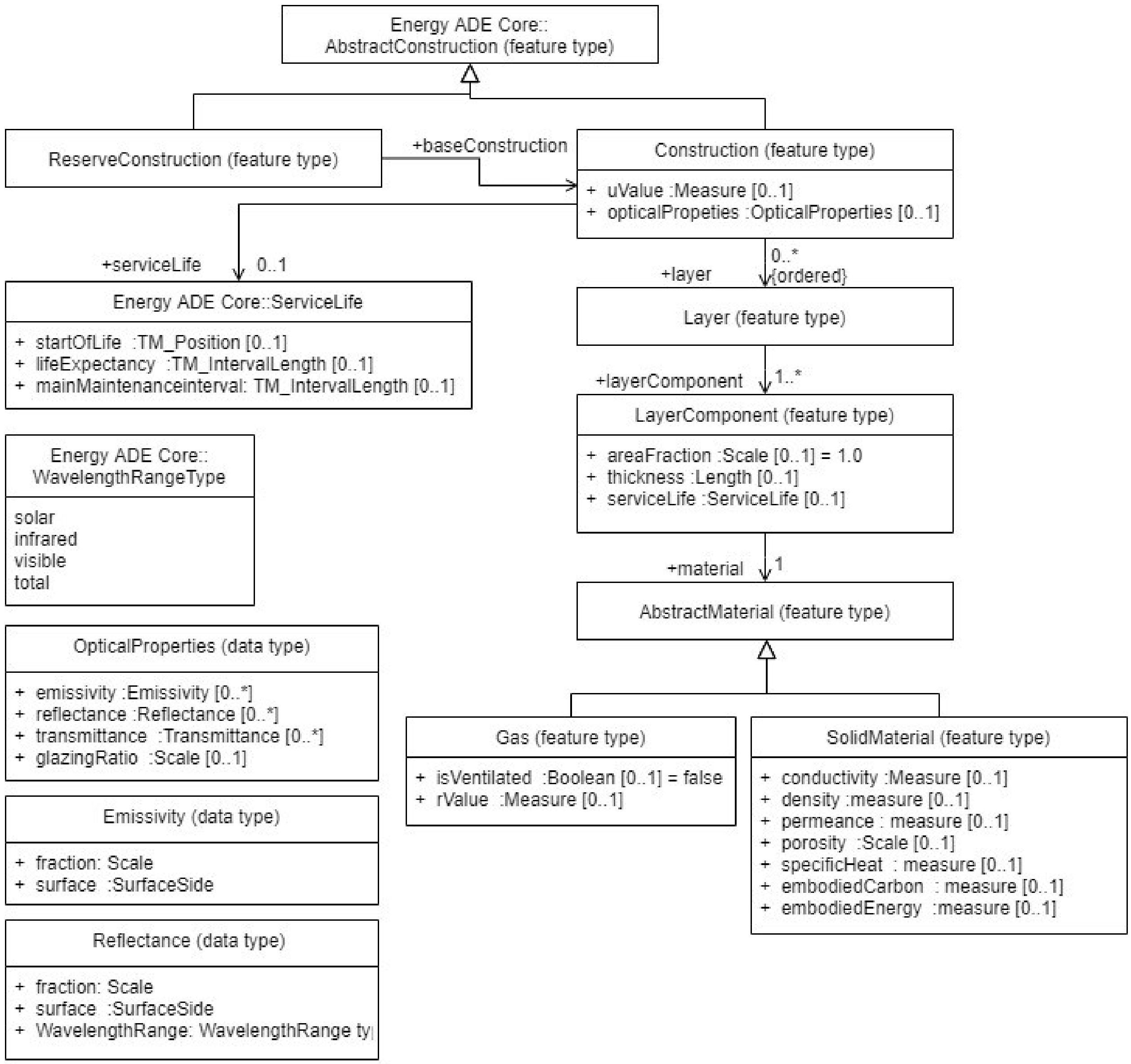


Figure 11: The UML diagram of the material and construction module in the EnergyADE model [47]

Table 3: UML Diagram Legend

0..1	No instances, or one instance	this class inherits from → this class
1	Exactly one instance	
0..*	Zero or more instances	you can navigate from this class to → this class
1..*	One or more instances	

Construction with two layers of materials

```xml
<energy:Construction gml:id="TwoLayerConstruction">
  <energy:uValue uom="W/K*m2">0.558</energy:uValue>
  <energy:layer>
    <energy:Layer gml:id="Layer1">
      <energy:layerComponent>
        <energy:LayerComponent gml:id="Layer1Component">
          <energy:areaFraction uom="scale">1</energy:areaFraction>
          <energy:thickness uom="m">0.2</energy:thickness>
          <energy:material xlink:href="#MediumMaterial" />
        </energy:LayerComponent>
      </energy:layerComponent>
    </energy:Layer>
  </energy:layer>

  <energy:layer>
    <energy:Layer gml:id="Layer2">
      <energy:layerComponent>
        <energy:LayerComponent gml:id="Layer2Component">
          <energy:areaFraction uom="scale">1</energy:areaFraction>
          <energy:thickness uom="m">0.049</energy:thickness>
          <energy:material xlink:href="#InsulationMaterial" />
        </energy:LayerComponent>
      </energy:layerComponent>
    </energy:Layer>
  </energy:layer>
</energy:Construction>
```

SolidMaterial

```xml
<energy:SolidMaterial gml:id="InsulationMaterial">
  <energy:conductivity uom="W/(K*m)">0.04</energy:conductivity>
  <energy:density uom="kg/m3">30</energy:density>
  <energy:specificHeat uom="kJ/K*kg">1</energy:specificHeat>
</energy:SolidMaterial>
```

Figure 12: The XML format of construction with two layers of materials and solid material data [48]

2.2.2.1. Road-mapping NREL construction data to the CERC platform

The U.S. Department Of Energy (DOE) has developed two archetype datasets in Input Data file (IDF) format serving EnergyPlus, and supporting eight ASHRAE climate zones in sixteen regions. The IDF is an intermediate data format, an ASCII file, translating the building geometry, construction, electrical, and mechanical system information for simulation in EnergyPlus [49]. The specified version and structure of the datasets have limited their usage in other building simulation platforms. Due to this issue, NREL has provided a program to make the dataset compatible with OpenStudioApplication. The steps listed in the subsequent paragraphs aim to draw the current data structure and map a plan to derive, transfer, and transform the construction data for use with the Energy ADE specification. This gives additional versatility to the file format by making it compatible with the UBEM data requirements.

IDF is the official format of building archetypes specified for the U.S. national building stock, including "the Commercial Reference Building Models[14] " provided as the reference stock model covering the primary building vintages, and "the Commercial Prototype Building Models[50] " generated for supporting technical analyses of new technologies and building codes focused on ASHRAE 90.1[51] and IECC[52]. Each IDF provides information on all significant complexities of building geometry, construction, individual thermal loads, detailed heating ventilation, air conditioning (HVAC), and more. The specific version of the archetypes in IDF format limited users to apply them in platforms such as OpenStudio, enabling the building energy modeling (BEM) with EnergyPlus. Thus, NREL developed the OpenStudio-Standard gem, which uses the Ruby programming language. It helped in the creation of the OpenStudio versions of the U.S. Commercial Reference/Prototype Building Models [53]. The NREL investigation generated two sets of data to cover a wide range of building simulation applications at different levels, "measures" and "components."

The measures are a set of energy-saving scripts designed for applying energy conservation measures on a model[54]. For instance, one possibility is adding overhangs to all south-facing windows in the model to measure overhangs' effect on the building energy saving. The next collection of the dataset is relevant to building components (construction assemblies), enabling users to generate archetypes script in IDF, gbXML, and SDD using Open Studio Application. This dataset allows developing archetypes based on building use-type, vintage, and standard code. In this case, using the Building Component Library (BCL) API[55] and connecting from Open Studio Application to NREL library, it is possible to provide various optimization and renovation scenarios over the reference archetypes. In the building component library, the construction assemblies are divided into four subcategories: walls, floors, roofs as opaque components, and fenestration as transparent components. The opaque elements are organized in two-fold datasets. The XML format with the information level suits the steady-state thermal simulation models. The text formats of IDF, OSM, and OSC with highly detailed construction data are fit for creating dynamic models using bottom-up engineering data modeling and simulation engines such as EnergyPlus. An explanation of the dataset characteristics is as follows.

1. The XML file contains the component's thermal resistance (R-value) in addition to some information applicable for classifying them in a large data library. Each XML component provides information on construction category (wall, roof, and floor), construction type (e.g. for walls include wooden, still, mass, ...), the effective R-value (m^2.K/W), the minimum insulation R-value (ft2.F.h/Btu), standard (reference code), standard type (residential or non-residential), and climate zone.
2. The second data set within IDF, OSM, and OSC formats contain information on the construction components (with detailed ingredient materials organized with a hierarchy of locations from outside to inside). Physical and optical properties of materials are provided in S.I. units. The information regarding the gas material between cavity spaces is usually defined as part of the opaque elements. This information offers nominal thermal resistance (m^2*K/W) for a given thickness. Figure 13 shows a sample of wall construction and material properties in OSM format, and Figure 14 illustrates the input data for the XML and the IDF version of opaque components.

```
OS:Construction,
  {fb78c073-3b0f-4fc5-ad0e-439d621685c0}, ! Handle
  189.1-2009 Nonres 1A Ext Wall Mass,     ! Name
  ,                                       ! Surface Rendering Name
  {0d27118f-2752-4791-ab27-e1c3920aed20}, ! Layer 1
  {095f9d21-18c0-4f22-b74c-14c944547750}, ! Layer 2

OS:Material,
  {0d27118f-2752-4791-ab27-e1c3920aed20}, ! Handle
  1IN Stucco,                             ! Name
  Smooth,                                 ! Roughness
  0.0253,                                 ! Thickness {m}
  0.6918,                                 ! Conductivity {W/m-K}
  1858,                                   ! Density {kg/m3}
  837,                                    ! Specific Heat {J/kg-K}
  0.9,                                    ! Thermal Absorptance
  0.92,                                   ! Solar Absorptance
  0.92;                                   ! Visible Absorptance

OS:Material,
  {095f9d21-18c0-4f22-b74c-14c944547750}, ! Handle
  8IN Concrete HW,                        ! Name
  MediumRough,                            ! Roughness
  0.2033,                                 ! Thickness {m}
  1.7296,                                 ! Conductivity {W/m-K}
  2243,                                   ! Density {kg/m3}
  837,                                    ! Specific Heat {J/kg-K}
  0.9,                                    ! Thermal Absorptance
  0.65,                                   ! Solar Absorptance
  0.65;                                   ! Visible Absorptance
```

Figure 13: The sample of wall construction in OSC data format [56]

There are also two data sets supporting the transparent (fenestration) subcategory. The dataset in XML format informs the reduced-order model. The IDF, OSM, and OSC formats of datasets cover the dynamic model.

1. The XML format comprises classifying parameters, thermal properties, and optical values. They include construction (component's type as fenestration then, e.g., window, skylight, etc.), construction type (frame construction type), overall U-factor (W/m^2 K), Solar Heat Gain Coefficient (SHGC), Visiting Light Transmittance (VLT), minimum/maximum glazing fraction, façade direction, standard type (residential and non-residential), standard (reference code), and climate zone.
2. The IDF, OSM, and OSC formats of the dataset provide more details, such as the unique identifier, type (subcategory), description (as explained in the standard code), minimum/maximum glazing fraction, effective U-value (W/m^2 K), SHGC, VLT, glazing material, and the detailed physical and optical properties of the used glazings in the model concerning the hierarchy of their installation from outside to inside. Figure 14 shows data arrangement in the XML and IDF versions of window components.

In parallel with the construction component datasets, another dataset is dedicated to opaque materials in XML, IDF, OSM, and OSC formats. The XML format provides detailed information on the materials and their classification, including the category of fabric (masonry, insulation, building board, and siding, framing with the cavity, solid wood, roofing, finishing flooring, and plaster), thickness, conductivity, resistance, specific heat, and typical application. The IDF format of material contains a unique name for the material, roughness, the used thickness of materials, conductivity, resistance, specific heat, thermal absorbtance, solar absorbance, and visible absorbance. The physical and thermal properties listed are the same for both data versions. Looking at the differences, the XML version carries some data for classification as opposed to the IDF/OSC version, while the IDF, OSC, and OSM formats support radiative information and roughness for each material which is missing in the former.

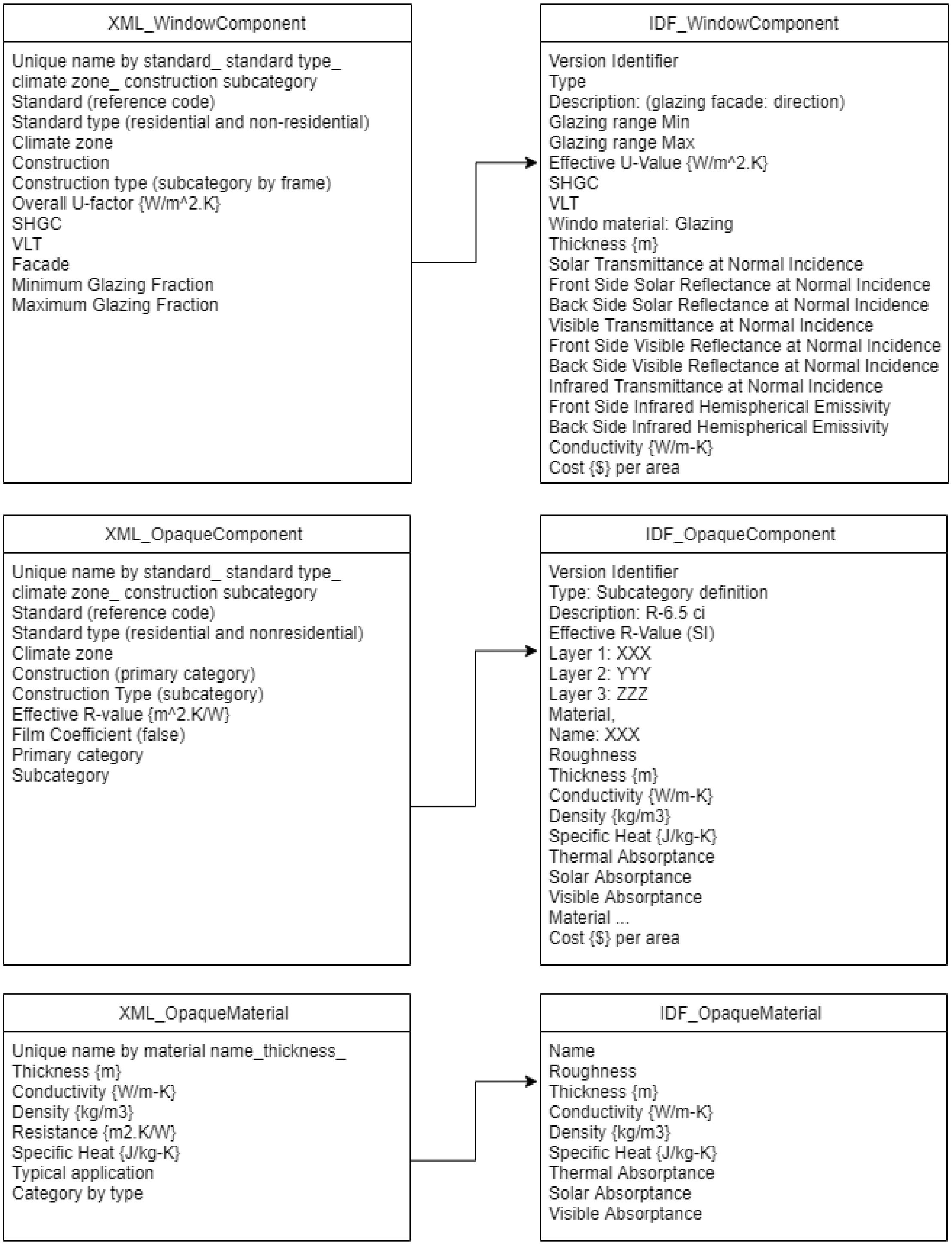


Figure 14: The available information in the opaque components, window (transparent)components, and opaque materials in the XML and IDF versions of elements. It is essential to keep the association between the XML with one of the text formats of IDF, OSM, or OSC.

Overall, the NREL dataset allows reaching a significant number of building construction assemblies, 2,460 opaque components (further divided into979, 905, and 576 for the wall, floor, and roof ceiling types, respectively) 1453 fenestration types, and 392 opaque materials. However, this does not mean that the components and materials listed are unique; their uniqueness is highly dependent on other classifying information such as climate zone, standard (reference code), and standard type (residential and non-residential)

The first step for developing the Construction and Material module adopted to Energy ADE data structure and UBEM application is transferring the dataset from the NREL server to the CERC repository with the defined association between the datasets as demonstrated in Figure 14. It means each XML version of the component needs to be connected to an IDF version of itself, and each XML version of the material is connected to an IDF version of itself. This process was done using the JSON format suitable for the data transmission between online servers. The transferred dataset is currently accessible through the following link and capable of providing building components for the last vintage of the U.S. so-called New Construction, which follows the ASHRAE 90.1-2004, ASHRAE 90.1-2007, ASHRAE 189.1-2009.

The accessible link to building components providing an association between XML and the text formats of IDF, OSM, OSC

https://binarycat.org/concordia/index.php?climazone=ASHRAE+2004%3A1A&activity=hospital&standard=ASHRAE+Std189&construction=Exterior+Wall

If the link does not work, please contact (guillermo.gutierrezmorote@concordia.ca)

In terms of organizing the internal CERC repository according to datasets' future applications, the second objective is building an active connection between materials and components with the possibility of edits in the future. This connection enables prospective users to test and implement energy-saving measures, retrofit scenarios, and archetype development by managing the materials and improving the reference constriction properties. The first category is the fenestration group. The XML format of fenestration available in the NREL_stock has the primary requirements for the thermal load calculation and the general information for components classification. On the other side, the IDF format consists of the window's construction layers (glazing and gas) with thermal and optical properties in detail. Both datasets have critical information that needs to be integrated to increase the dataset's functionality and used in the CERC_library.

The recommendation for developing the fenestration dataset in the CERC_library begins by extracting and storing the physical and optical information of individual glazings and gases from IDF and storing them into an XML format (XML_GlazingOpticalProperty) identified by their unique name (glazing and gas name) as illustrated in Figure 15. In the next step, the current XML format of fenestration in the NREL_stock should be programmed to circulate in the associated IDF format for taking the construction layers of each window component specified by their unique names. Each window could get between one and seven layers of glazing and gas spacing arranged from outside to inside. The result will be collected in a separate class as XML_WindowComponent. The third stage is essential to define the association between two designed XML classes based on the defined construction layers (glazing and gas name). This association's focus is connecting the optical properties of glazing to its thermal and classification factors to deliver complete window parameters; Ready to be managed and used depending on the simulation model and data requirements.

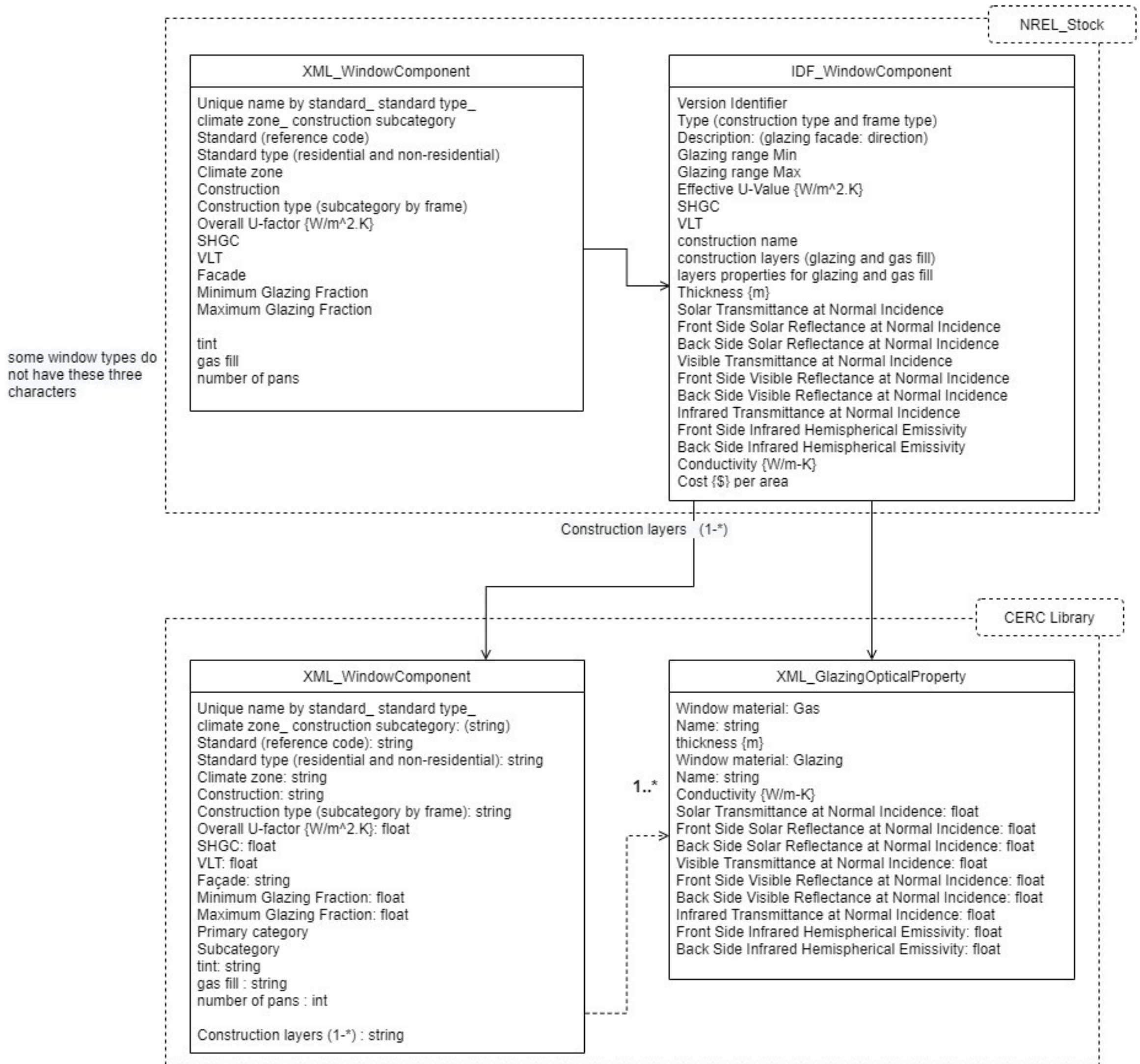


Figure 15: The proposed process of window component development for storing in the CERC platform (prospective UBEM), it is noted that three attributes of tint, gas fill, and number of pans are available in some window types that have more than one panel

The second group of components belongs to the opaque elements, which need to be strongly connected with the CERC library, "material," presented in Figure 16. The process starts with developing the material library identified by specified name and classified by type. The type shows the material category by their application, including masonry, insulation, building board-siding, framing with the cavity, solid wood, roofing, finish flooring, and plaster. The XML version of opaque material available in NREL_stock is good enough to provide the physical properties of materials for thermal load calculation and information such as typical application and category by type for classification. In this case, the thickness is unnecessary information that can be eliminated from the dataset. Only unique materials with unique names need to be stored. For covering the value of absorptance for materials (visible, thermal, and solar), it is possible to programming the XML file to go over the associated IDF version of the material for taking the relevant information for individual materials. However, it is possible to organize it in a separate class.

After developing the material library in the XML_CERC_OpaqueMaterial, the second objective is to generate a collection of components using the CERC material library (based on the NREL components). In this case, we use the XML format of opaque components available in the NREL_stock as the leading XML and program to go over the IDF version of the component for picking up the construction layer information, consisting of the name and thickness of layers. The output data will be programmed to circulate into the XML_CERC_OpaqueMaterial to find the matching materials by name, put them in the determined spots in the leading XML dataset, and calculate the effective U_value based on the thickness and the physical property of materials presented in Figure 16. The advantage of such component creation is having an interconnected dataset of building components and materials that enables users to parameterize the design and optimization of the construction component by changing the material's thickness and type. The classified information in the dataset allows for the generation of building construction assemblies based on the U.S. building standards, containing ASHRAE 90.1-2004, ASHRAE 90.1-2007, ASHRAE 189.1-2009, as well as the climate zone and building type, including residential and non-residential.

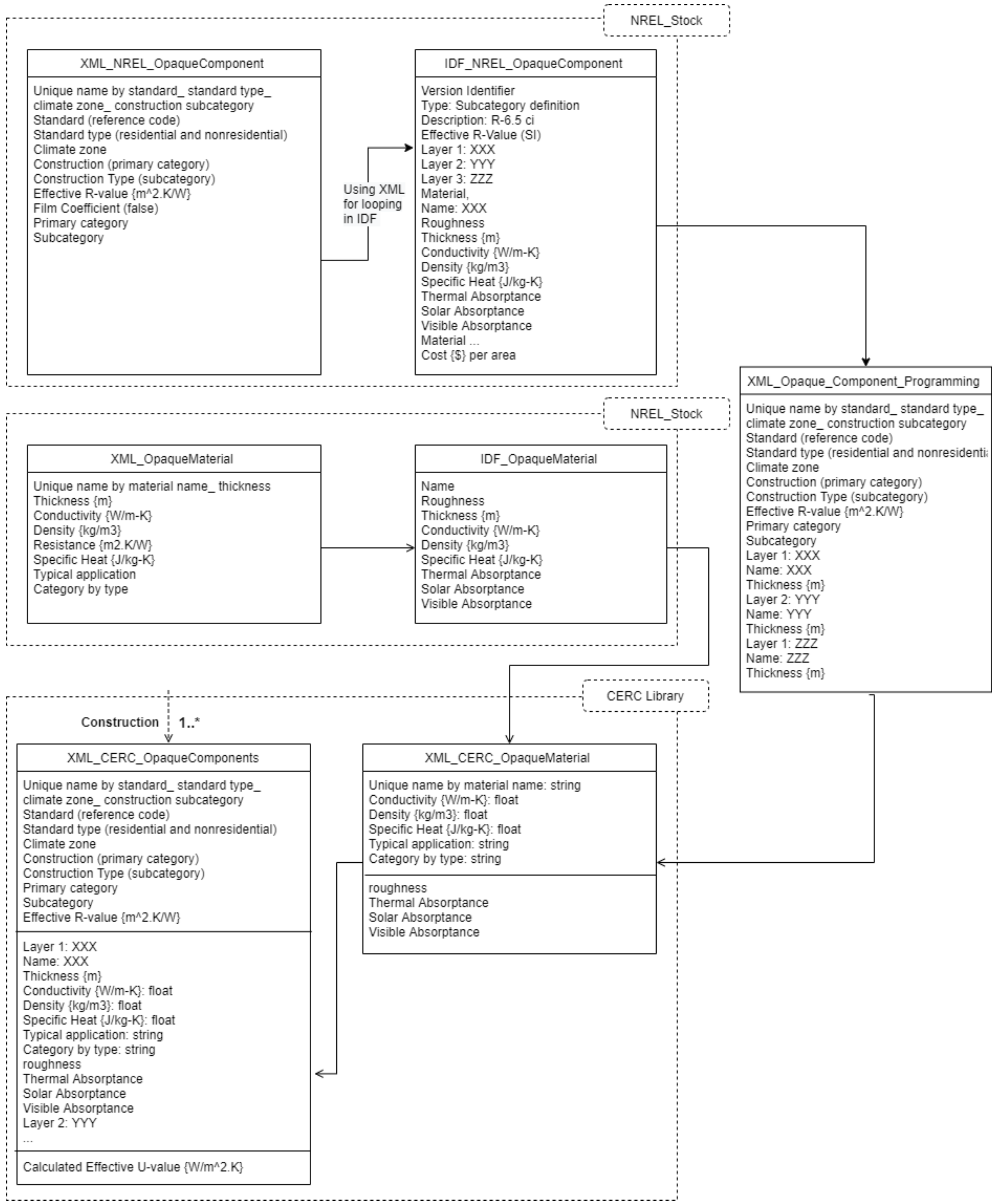


Figure 16: The proposed process of opaque component development for storing in the CERC platform

Overall, the prospective dataset in the CERC_library, Figure 17, would be composed of two parallel datasets for the opaque constructions supporting floors, roofs, walls, and the transparent construction (fenestrations) covering the windows, skylights, and doors. Each category comprises three interconnected sections. The thermal transmittance of construction assemblies comes alongside many classification factors. The AbstarctMaterial (Opaque/Transparent) supports the physical and optical properties of materials in two subcategories. The XML_OpaqueMaterial covers the physical and radiative properties of opaque materials. The XML_GlazingOpticalProperty provides the required physical and optical characteristics of glazings consisting of absorptance, emittance, and reflectance in respect to three beam's wavelengths, including solar, visible, and infrared on two surface sides. The third section is XML_LayerComponents that manages the layers' thickness and hierarchy for each construction component.

Having classifying data for each component overloads the size of the construction library. However, it provides added value for the user's application. It allows for filtering the components based on the defined frameworks for building energy data in North America, such as ASHRAE climate zones (8 zones), ASHRAE standards, and building use-type. The other factors, such as construction and construction type, are somehow mandated for categorizing data when there are significant numbers of construction assemblies. The defined method for providing the dataset tried to keep as much information for each construction assembly to be self-descriptive, identified, capable of participating in the steady and dynamic simulation, and applicable for various climate regions based on the specified standard.

The first difference between the provided model and the Energy ADE data structure is the considerable number of the description as classifying parameters alongside materials and constructional components. Having them in the library section is demanded and facilitates data management and archetype generation regarding the mentioned varieties. The next difference is the integration of radiative parameters alongside the material properties. The radiation elements are separated into six classes in the Energy ADE model, while the radiative properties of opaque materials in the CERC library would come across thermal and physical properties in the same class. The glazing optical properties are also non-segregated for three radiation beams on double surface-sides in the fenestration category. Although such integration overloads the material datasets, it reduces the number of links in the model since each material layer is composed of complete information.

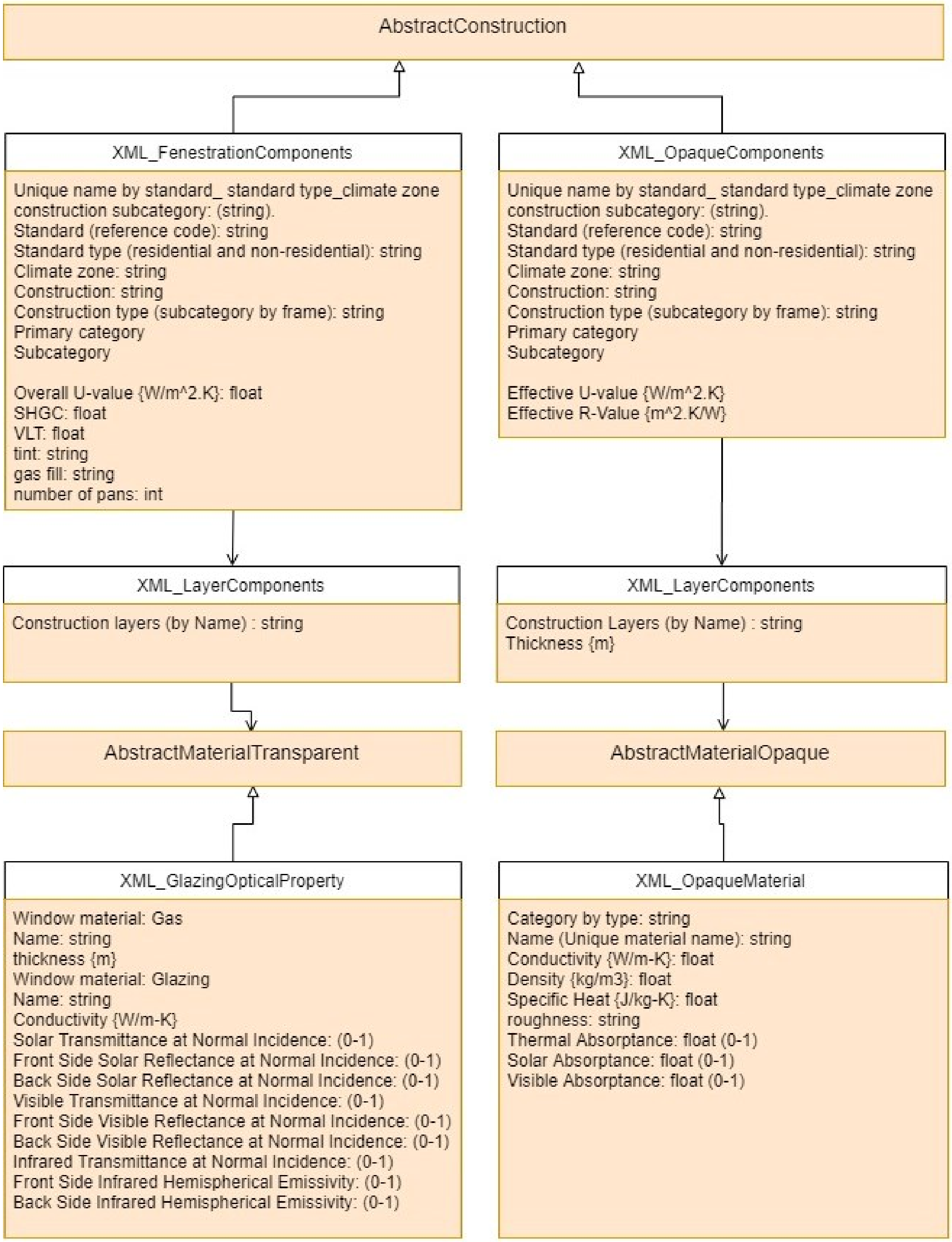


Figure 17: The interconnected Construction and Material module based on NREL construction Components and Materials

Chapter 3

3. Methodology to create an urban modeling workflow

The designed workflow is articulated in three principle data operation categories: the data description, preparation, simulation, and mapping demonstrated in Figure 18. In the beginning, the study deals with the variety of data sources, scales, and formats entailed for the discovery of their potential and challenges to satisfy the UBEM requirements. The next step focuses on data preparation, which consists of multi-step data processing and mapping to consolidate the master building dataset in both levels of geometry and non-geometry. The prepared geometry dataset follows two steps of the 3D city model generation and conversion to adhere to the CityGMLschema, the required geometry standard for the used UBEM in the current study. Simultaneously, the archetype models are characterized by providing energy data for Kelowna housing and delivering them into two datasets of building physics and building usage. The last step involves acquiring and using the generated databases for UBEM simulation, output validation versus the national measured end-use surveys, and mapping results. The project starts out on city-scale data analysis and preparation, then narrows the scale down to the Rutland neighborhood for building modeling and energy simulation.

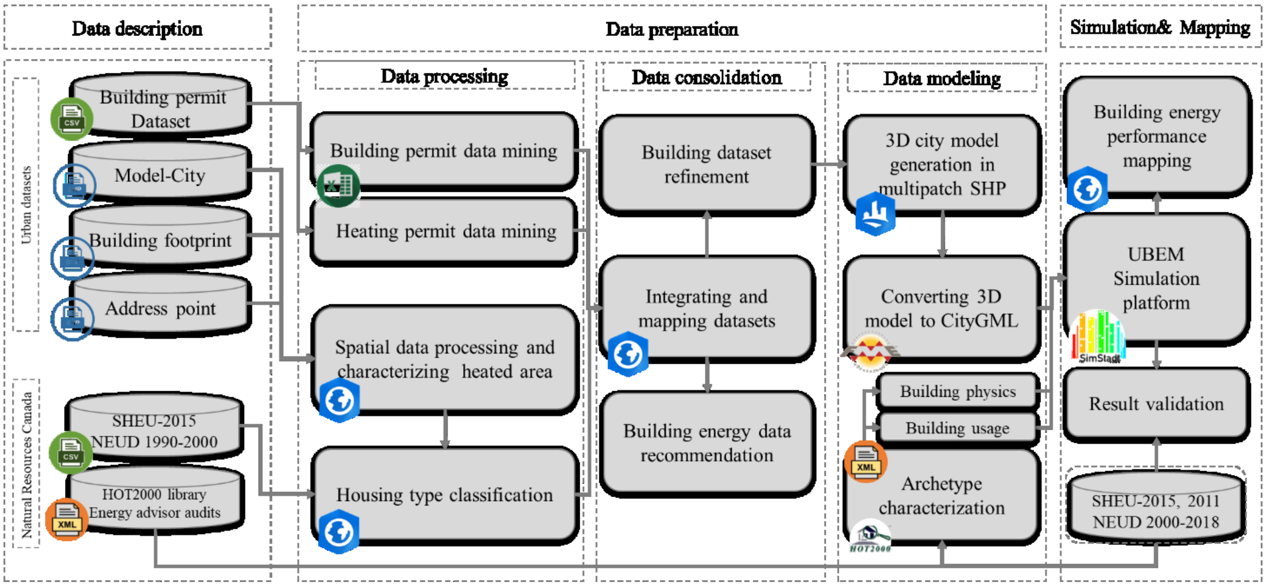


Figure 18: The workflow of urban building energy data preparation, modeling, and mapping using urban datasets

3.1. Data understanding and case study introduction

The low-rise residential stock in Kelowna city is the pilot case study for the CEEMap project and the work within this thesis. Kelowna is located in the southern part of British Columbia, Canada, and is categorized as possessing a humid continental climate by the Köppen climate classification system. In the coldest months, Kelowna's average temperature is slightly above −3.0 °C and below 0 °C with cold, cloudy weather, and the summer days are dry, hot, and sunny. Following the ASHRAE climate zone, Kelowna is located in zone 5 with 3000-3999 Heating Degree Days (HDD); see Figure 19 (b). Rutland is the largest residential neighborhood of Kelowna by far. It covers a 4.5 km2 area in the middle of the city. The Rutland neighborhood area sits at the valley's foot on a flat site with fringes continuing up into the hills. The regional location of Kelowna and Rutland neighborhood are demonstrated in Figure 19 (a and b).

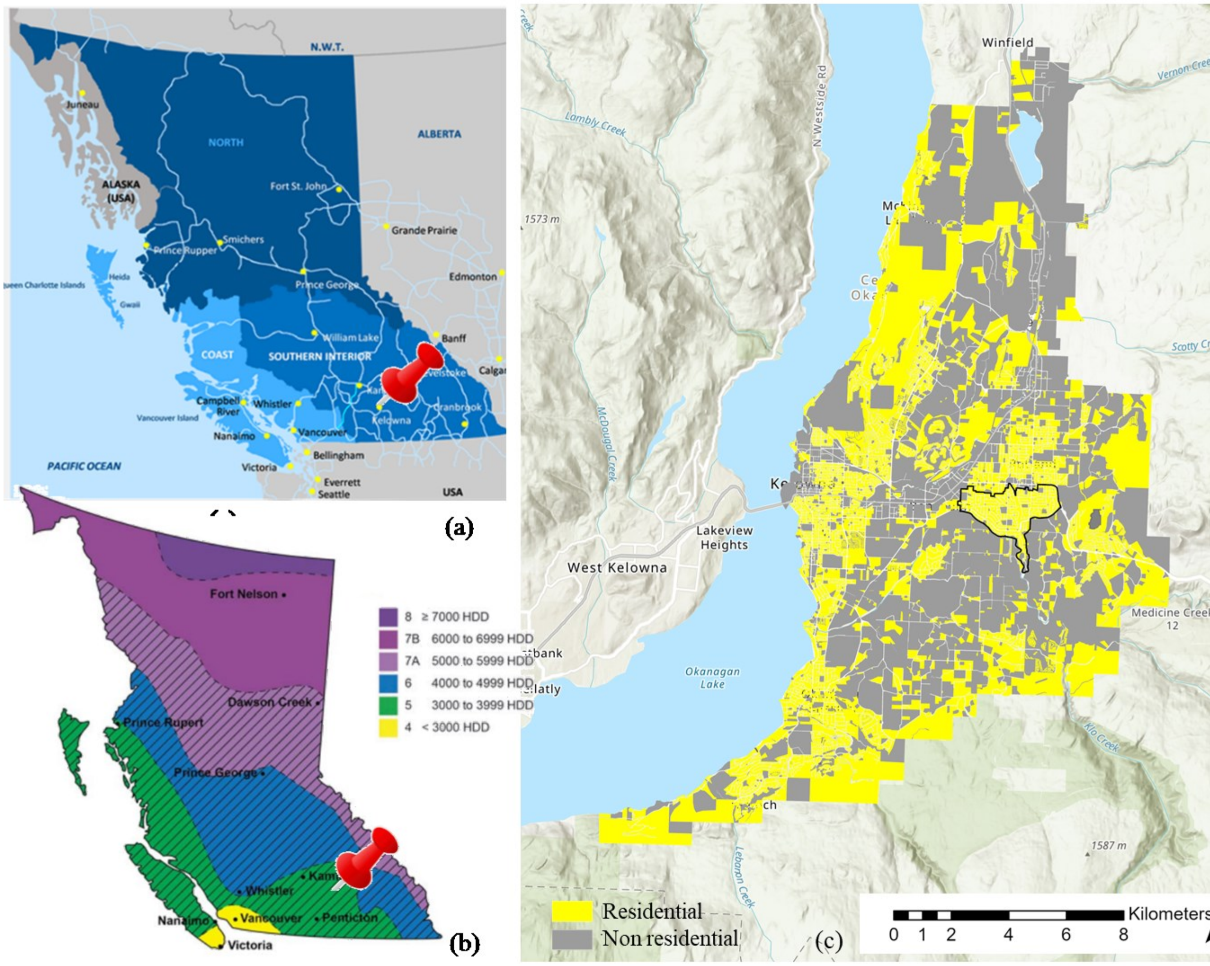


Figure 19: (a) Kelowna City in the southern interior of the British Colombia province, (b) Classification of climate zones for British Colombia and Kelowna city, (c) Residential land use distribution in Kelowna city and Rutland neighborhood as the study region for modeling and simulation

3.1.1. Model City dataset

ModelCity is an integrated SHP dataset projected in the coordinate system NAD 1983 UTM Zone 11N developed by Kelowna city to boost corporate performance and support informed business decisions. It includes (35,063) records with 48 attributes assigned to each lot (parcel). The attributes consist of a unique identifier named Kelowna IDentity (KID), zoning bylaw, and regulation for six principal categories of land-use: "residential," "commercial," "civic, institutional and recreational," "farm," "industrial," and "transportation, communication, and utility." The dataset also covers the building information such as typologies, gross floor area, year of construction for old and new buildings constructed in the lot, story number, bathroom and bedroom numbers, the gross value of buildings, and land at the parcel level. The dataset has been updated twice a year since 2017, when it was developed from corporate data. The essential attributes for improving the building dataset are listed in Table 4. Although Model City provides complete details about buildings, it does not individually target living spaces (buildings) at the tax lots. This results in confusion where there is more than a building in the parcel. Another issue is the value of zero for critical building characteristics, such as total floor area, year of construction, and story number. Table 4 represents the challenges within attributes that are suitable for building modeling. Table 3 also shows the statistical features of ModelCity required for building data processing.

Table 4: ModelCity attributes and their description and limitation of use for association with the building dataset

No	Attributes	Description	challenges
1	Actual_Use_Code	Building archetype in specified 3digits code	Confusion in recognition of the principal building for parcels with more than one building
2	Actual_Use_Classes	Land use classification	
3	Actual_Dse_Description	Building archetype with some physical features	
4	Nbr_of_res_units	Number of units in the building or parcel	The area is aggregated for parcels if they have more than one building
5	Total_ strata_Res_UnitArea_sqf	Total gross floor area for strata type of residential buildings	
6	Res_House_Total_Area	Total gross floor area for the non-strata type of residential buildings	
7	Res_ House_Stories	The maximum number of building stories in the parcel	Confusion in recognition of the principal building for parcels with more than one building
8	Oldest_building_year	The primary year of construction,	
9	Newest_building_year	The newest year of construction, in case of new construction	
10	KID	Unique identifier for parcels	It is not connected to the primary building of each parcel
11	Address	Unique information for parcels introducing street name, number, and legal address	

Table 5: ModelCity statistical facts

Characteristics					Records count			
Total parcels in ModelCity					35,063 (100%)			
Total parcels for low-rise residential stock					29,963 (86%)			
Residential parcels with zero floor area					4481 lower than 46m2 (500 ft^2) (4%)			
Residential parcels with zero as the year of construction					408 (1% of Total lots for the low-rise residential stock)			
Building types available at the parcel level								
Building type	Single Family Dwelling	Duplex	Tri/Fourplex	Multifamily house	Strata/ Condominium	Town/ Rowhouse	Mobile home	Not applicable (NA)
Actual Use Code (AUC)	000, 002, 032, 060	033, 034, 035, 036	047, 049, 053	050, 054	030	039	037, 038, 063	001, 020, 051,052,061 and more than 063
Record number	27921	1222	72	152	222	284	90	1412

The dataset introduces seven building types within low-rise residential buildings with 25 subcategories. Figure 20 shows the distribution of building types based on the record number. Notably, Single Family Dwelling (SFD) covers most low-rise residential building records distributed in Kelowna city; see Figure 21. The other important segmentation factor is the construction year that directly affects the building envelope's thermal properties. Figures 20 and 22 demonstrate the distribution of the building's construction year following the housing vintage applied in the Survey Household End-Use (2015). The majority of construction is related to three vintages of 1961-1977, 1984-1995, and 2001-2010 by supporting 66% of the total parcels.

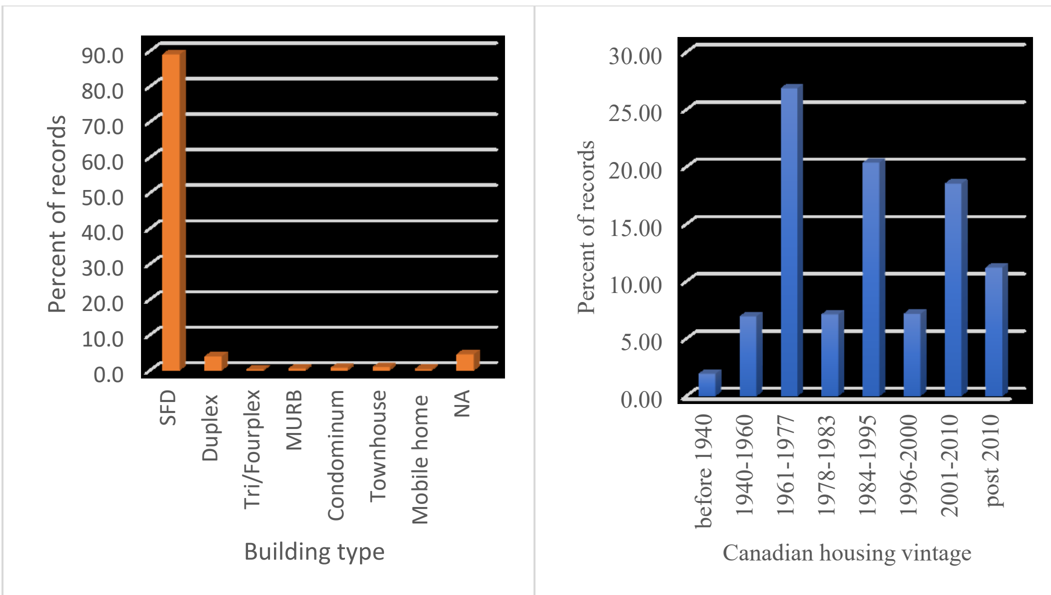


Figure 20: (a) Classification of building records based on identified typology, (b) Classification of building records based on principle housing vintage in Canada

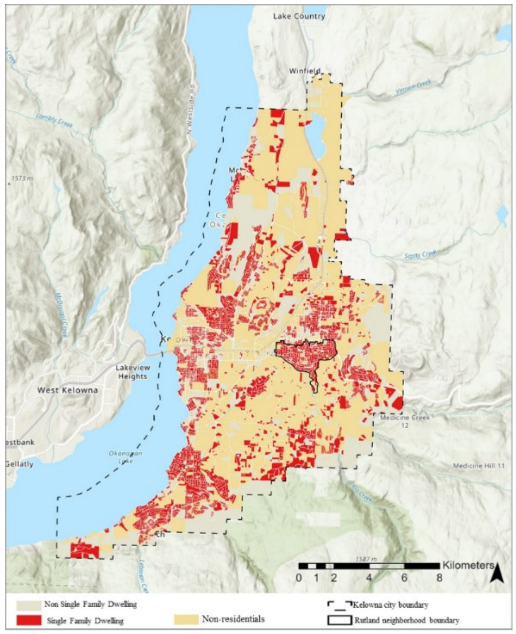


Figure 21: The spatial distribution of Single-Family Dwellings in Kelowna city and Rutland neighborhood

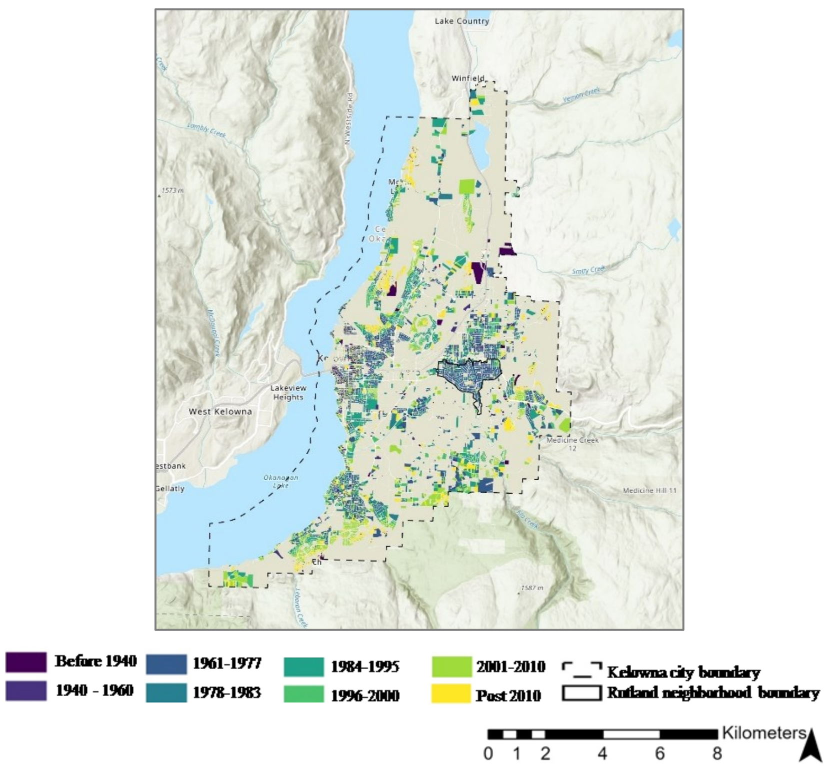


Figure 22: The distribution of vintages between the low-rise residential stock of Kelowna city

3.1.2. Building Footprint dataset

The building footprint is spatial data in SHP format derived from the orthophoto flown in May 2019 and consists of (43,928) building records. Only buildings larger than 47 m2 are included in this dataset. The projected coordinate system is NAD 1983 UTM Zone 11N, the same as other spatial datasets in Kelowna's open data portal. The dataset carries important attributes consisting of the ground elevation and the maximum elevation of buildings. The differences between them represent the buildings' maximum height. Another feature recently added to the dataset is "building_PK." P.K. stands for permanent key, identifying each building with a unique code. The building_PK association to building footprints follows the rule BLD10000X. X was the OBJECTID when the footprint map was released for the first time. For instance, the BLD100001 represents the OBJECTID 1, and the BLD142671 represents the OBJECTID 42671. Adding the unique building identifier is a new idea for mapping building footprints in Kelowna and not officially defined for the rest of the urban datasets provided at building scale. Therefore, it could not help assign entities to the building footprints where there is more than one building in the parcel. The final attribute is the footprint area, which is the geometry processing outcome in the GIS tools. It is calculated by default when adding two-dimensional spatial data (polygon) at scale.

In general, the dataset covers the building height between zero to more than 83m. As demonstrated in Figure 23, the normal distribution of height is between 3.9m and 8.7m. This indicates that the majority of buildings in the city are single and two-story buildings. The mean and median of building height is 6.2m and 5.9m, respectively, with a standard deviation of 2.26m. Figure 23 depicts the significant intervals for building heights. 0 to 2.6m is the first and minimum building height range, and 360 records have lower than 2.60m, and 92 buildings have zero value. Comparing the maximum building height and GoogleMap data for random buildings shows a discrepancy in some cases and demand accuracy checks with other supplementary datasets. In general, the geometry extraction from the orthophoto is subject to error if no additional database supports it [57]. One of the reasons is related to the sources of orthophoto data. When processing orthophoto for the building height detection, the lack of access to the sensor data leads to unprocessed data exploration. If the original data is available, the other issue could be a wrong selection of points on the ground for measurement of height (which is usually measured from base to the top, perpendicular from the point of measurement). This methodology used in orthophoto data processing, introduced by ESRI[58], is not reliable if buildings have a lean, angle, or setback.

In this case, the LiDAR dataset is a preferred choice to validate the building height and support multi-measures for complicated building geometries. Nevertheless, the lack of such a processed dataset leads the project to use the other supplementary sources such as ModelCity discussed in the spatial data processing.

Figure 24 shows the building footprint distribution, including garages located under the roof, which ranges between 47m^2 to 11,139m^2. The majority of records (near 98% of entire buildings) are lower than 400m^2. The average footprint area is 231m^2, the median and standard deviation are 200m^2 and 219m^2, respectively. The building footprint is not without error as well. It covers the spaces below the roof as the footprint. Hence, attached garages, storerooms, or empty rooms below the roofs would be recognized as the footprint.

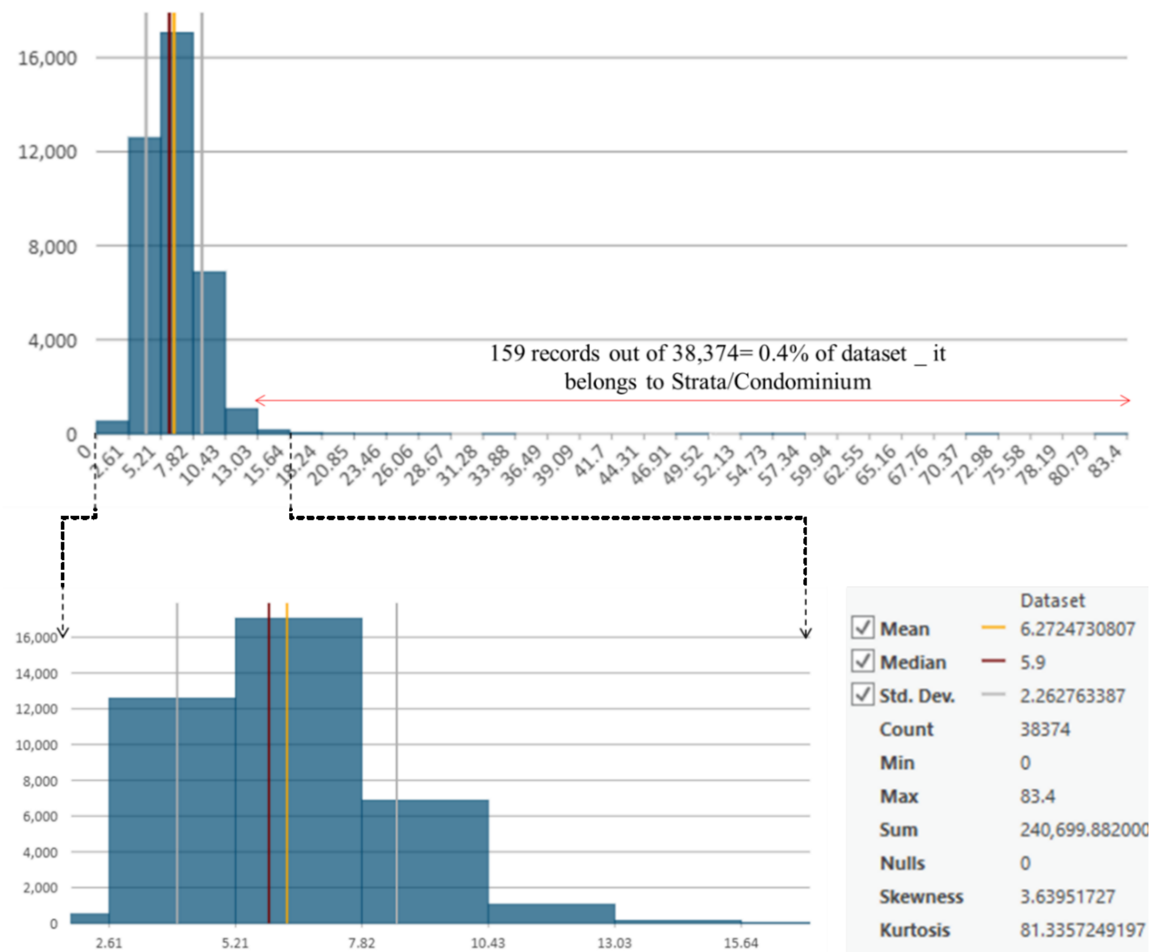


Figure 23: Distribution of residential building height records in the city of Kelowna

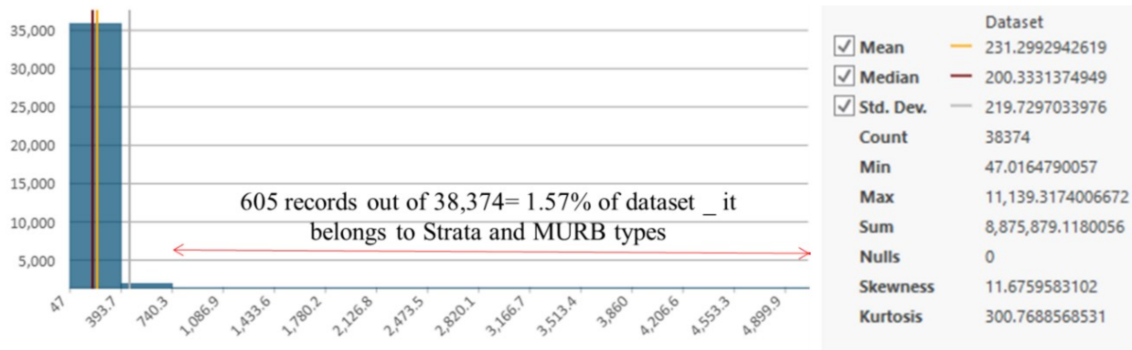


Figure 24: Distribution of residential building footprint in the city of Kelowna

3.1.3. Address Point dataset

Address point is the point dataset in SHP format available in Kelowna open portal [59]. It carries the address of buildings by the apartment. The priority of the dataset is the geospatial determination of the living area. However, it does not support facilities by targeting their unique identifier. The geographical characteristics of the Address Point dataset provide valid location-specific points to identify living areas or conditioned spaces within the buildings. The mapping of spots has been done manually by Kelowna city. As a result, human errors have resulted in felling points outside the building outline in limited cases. The maximum distance between the points marked outside the building footprint and the closest building has been measured to be up to 5m without intersecting with other building outlines.

3.1.4. Permit dataset

The permit dataset consists of 88,830 unique permit-numbers demarcated into four categories: "heating,” "construction,” "plumbing," and the last group relevant to the inspection and other sorts of activities mandated for permit requests. The start of systematic permit data gathering goes back to 1995, and before that, the process was manual. Although the older datasets exist, there is a very sparse data description to show the permits' detail in that period (before 1995). The dataset prepared in CSV format includes eleven attributes described in Table 6. The last three attributes, 9 to 11, were recently added based on the address column consisting of parcel ID (KID) and coordinate system of parcels defined as longitude and latitude. Adding these attributes resulted in a considerable number of data duplication (84,390 duplicated records), which is filtered based on the unique permit-numbers. The dataset also has some practical issues. The first and foremost is the lack of a standard to monitor the dataset's accuracy, the absence of supporting documents clarifying the work descriptions, and explaining many missing values. The dataset is also prepared with very little focus on quantitative measurements. It also provides a low guarantee for data accuracy/validity, even for necessary information such as the floor area and building unit information. The absence of building identifiers is another issue that limits the connection of data to appropriate buildings. Although the dataset has added KID's attribute (Kelowna ID in ModelCity), it still creates confusion for parcels with more than one establishment. The absence of GIS tools in data collection and addressing facilities is another feature that is missing within the dataset. This has not been entirely offset by the addition of the KID attribute and parcel coordinate system.

Table 6: Data understanding of permit attributes

No	Attributes	Description	Limitation
1	permit type	It refers to four categories of permits, including heating, building, plumbing, and other.	-
2	permit status	It refers to five possible permits' status, including closed, open, new, expired, and canceled.	-
3	permit subtype	It refers to the category of issued permits.	Not cleared definition for some shorten words and abbreviation
4	work description	It describes the activities planned within buildings.	ambiguous words for some exercises, minimum support of data with quantitative measures, non-clear shorten form of terms and abbreviations
5	permit number	The unique number of each issued permit	-
6	address	It determines the address of the parcel applied for the permit	-
7	dwelling units	The number of units in a building or parcel applied for the same permit at the same time	The numbers need some explanation, 0 means 1, 1 means 2, 2 means 3, and so on.
8	gross floor area	The measure of the area added or reduced in the process of practice	There is a considerable number of missing values. Moreover, it mentioned the values are unreliable.
9 10 11	KID Longitude Latitudes	Unique identifier for parcels Defining the coordinate system of the parcel (X) Illustrating the coordinate system of the parcel (Y)	These three attributes are merged from other datasets and led to the replication of permits.

3.2. Data processing

3.2.1. Spatial data processing

As explained in the dataset's introduction, building data is patchy and needs to be aggregated in a master dataset representing all attributes and values in a unified format. The three datasets of ModelCity, Building Footprint, and Address Point should complement and overlap each other to remove the previously discussed gaps and issues and provide a consistent dataset ready for 3D city modeling in CityGML format. As demonstrated in Figure 25, the first step of spatial data processing starts out with assigning the parcel information to the building through the spatial join of ModelCity to the building footprint. This has been explained in the upcoming section 3.2.1.1. The result was subsequently connected with the Address Point to further analyze and determine the living area (heated area) of the primary buildings in each parcel. This has been detailed in section 3.2.1.2. In the third sub-section, section 3.2.1.3., the focus is on reviewing and classifying the building types with reference to their architectural typology and vintages following the Survey of Household Energy-Use (SHEU-2015) [60]. This has been done with the intention to identify the archetype characteristics that could represent the majority of low-rise residential housing in Kelowna.

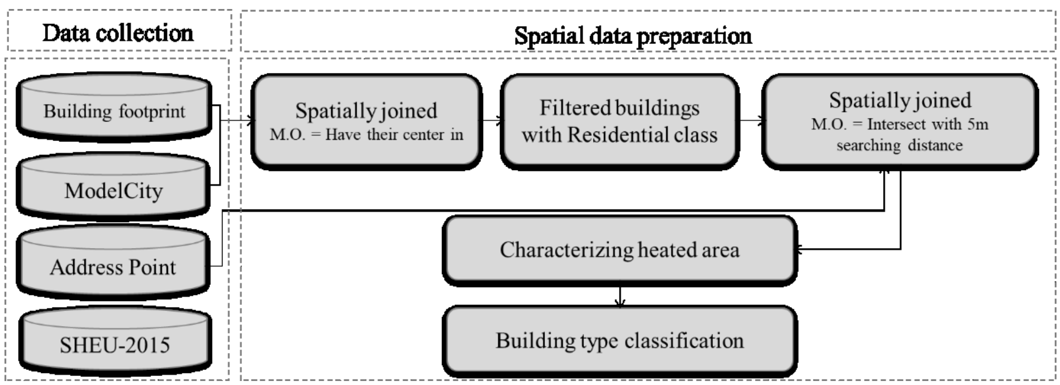


Figure 25: The workflow of spatial data processing (M.O. stands for Match Option as an option of spatial join procedure in ArcGIS Pro

3.2.1.1. Mapping the building footprint datasets with Model-City

Mapping building footprint with Model City information in the GIS platform needs a minimum requirement, either having a unique identification attribute in common between two maps or having the same or closed geospatial coordinate. The lack of individual and standard identifiers shared between two datasets resulted in using spatial join, leveraging the similarity of the coordinate system between datasets. Since the dataset's polygons have not been matched based on the tax registered criteria, the overlying layers are subject to errors for less than 2% of the dataset. That means if one footprint meets more than two parcels then, one of them is randomly picked. To eliminate such a problem, the toolbox of “SpatialJoin” in ArcGIS Pro (10.4) proposes the match option of "having center in,". It allows the model to consider the center of the footprint within the parcel as a principal condition of matching. This method provides almost a proper association between buildings and parcel information with one option for each building footprint. Mapping of building footprint and Model City prepares a large building dataset. However, the output of spatial-join is problematic for parcels with more than one building. In this case, buildings in a parcel would receive the same parcel KID and information, which might not be accurate. Such error addresses 26% of the building footprint dataset covering 8,226 low-rise residential buildings (that receives data from 4,453 parcels).

Another problem discovered through the joining of building footprint and ModelCity refers to building height. For 46% of building records, there is a mismatch between the building height measured through orthophoto processing in the building footprint dataset and the expected measure for the specified story number added from ModelCity. The possible errors in orthophoto processing are mentioned in section 3.1.2. Another reason might be assigning the same parcel information to structures within the same parcel. It has occurred for 15% (4,928) of buildings.

Table 7 provides an approximate number of residential building records out of the expected range based on their story number. There is a large difference between the minimum and maximum building heights in the specified story number categories shown in Figure 26(a). However, more than 50% of building heights respect the estimated measure based on the story number. The average building height in these groups is very close to the estimated range and is used for improving the mismatched groups. The story number in ModelCity has been listed as zero for 5% of buildings. In reality, this (zero) signifies a single-story building, which has an average height of 4.4m.

Table 7: The record number of buildings having out of range height based on story number lower than 5

Story number	# Total records	Avg. height of the group(m)	Height range in the group (m)	Expected measure based on story number (m)	# Records out of range	% of out of range in the dataset
0	1,754	4.40	0-27	3 – 5.7	119	0.5
1	24,474	6.20	0-71.5	3 – 5.7	12,282	37
2	6,517	8.26	0-19.79	6-8.85	2,650	8
3	245	10.06	0-22	9-10.65	189	1
4	142	13.23	0-56	12-14.5	97	0
Total	33,132*				15337	46.5

* total building records includes story number lower than four regardless of the typology of buildings

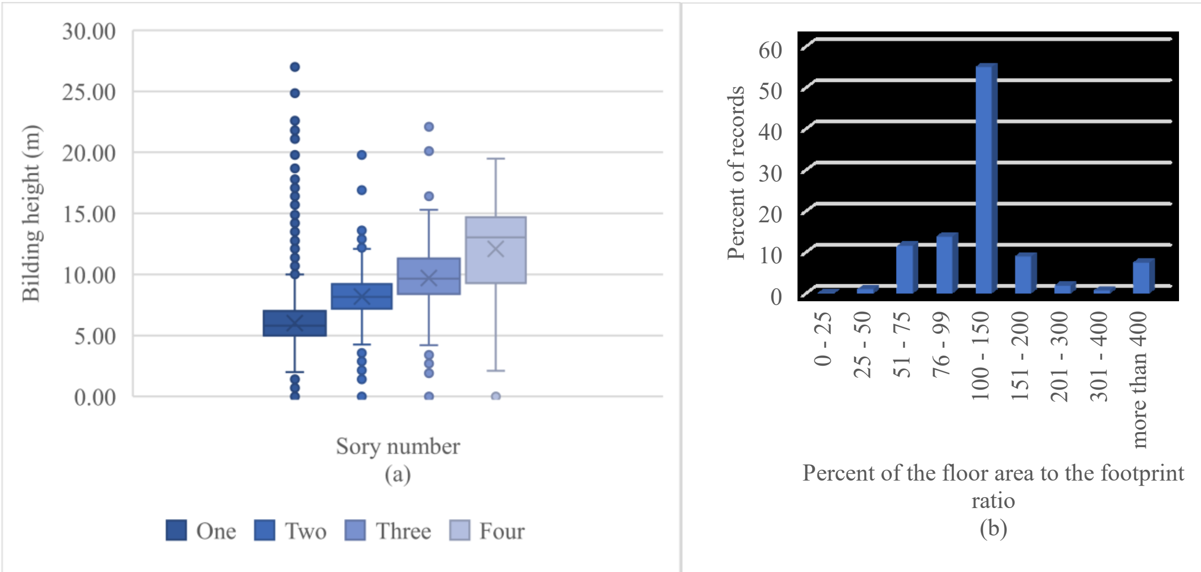


Figure 26: (a) Distribution of building height in each story number category, (b) Distribution building floor area to the footprint

Figure (b) illustrates the distribution of the building floor area to footprint ratio in Kelowna city, which basically represents the level of compact design in a piece of land. The first deduction from this figure is that the majority of buildings in the City of Kelowna are between one and one-and-a-half story. The second case is the possibility of error in near 20% of the derived building footprint area, assuming the assigned floor area in ModelCity is reliable. The building floor area to footprint ratio of lower than 1 represents the built area does not satisfy the minimum compactness for filling the derived footprint outline. As mentioned in section 3.1.2., one of the reasons is considering the rooftop as the footprint outline. Therefore, the garage or any empty space below the roof is included for drawing the border of building footprints, while the measured floor area excluded them. The other possibility might be the error of color classification for detecting the roof borders [57].

3.2.1.2. Mapping the building footprint with address point

The residential building footprints are composed of two building types, primary buildings considered as living areas and accessory structures such as sheds, detached parking, and storeroom. Accessory building structures are not subject to heating; therefore, it needs to be detected and labeled as unconditioned spaces. However, there are no codes to differentiate them from the primary building of the parcel. Therefore, to recognize the heated area, joining the building's overlay with ModelCity information with Address-Point would provide a georeferenced solution. In this case, the geospatial character of the Address-Point dataset helps determine the living area in parcels.

The lack of shared attributes between the Address-Point layer and building footprint led to the use of spatial join. As mentioned in previous sub-sections, some points are not within the building boundaries and are falling at a close distance away from the building footprint. Hence the spatial join between two datasets occurs with the possibility of intersection for points within 5m distance from the building footprint. This strategy helped find 167 more buildings compared to using the spatial join without the option of 5m searching distance. Overall, mapping the footprint with Address Point resulted in detecting 33,970 records, 88% of the dataset, as the referenceable living area. The rest of the 4709 buildings were classified as having a high potential for being an accessory and non-heated installation in the parcels. However, they are not removed from the dataset. They will be modeled besides the other buildings to keep the radiation effect of building on each other. Figure 27 depicts the overlay of datasets conducting the building information and shows the eligibility of primary facilities at the parcels.

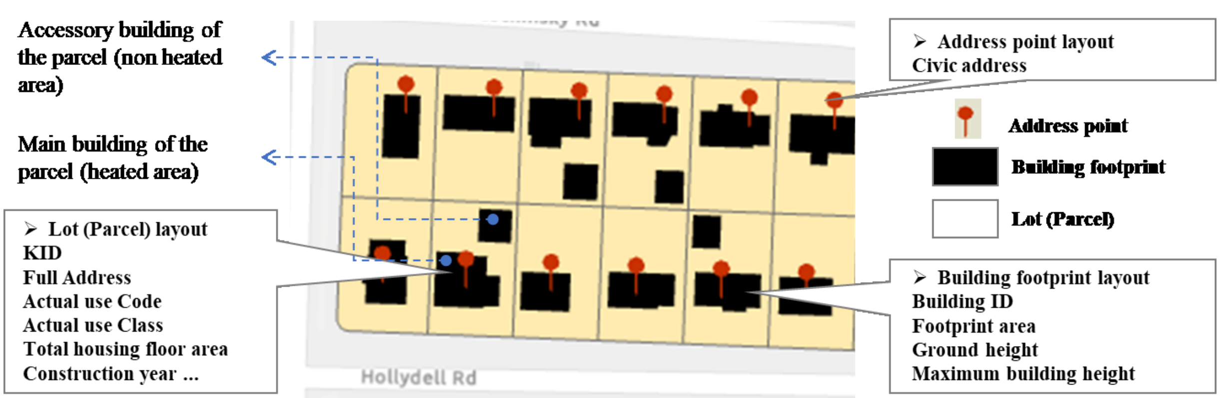


Figure 27: The overlay of various building-related datasets, including the building footprint with City-Model, Address Point, and Permit dataset

3.2.1.3.Building type classification

The Survey of Household Energy Use (SHEU– 2011[61], 2015[60]), conducted as a joint project between Statistics Canada and Natural Resources Canada (NRCan), is a leading survey focused on gathering information on end-user housing and influential factors of household energy use. The SHEU-2015 building classification covers the private dwellings of the vast majority of the

Canadian population. The SHEU-2015 building classification can be broadly broken down into two super-categories: low-rise housing and high-rise housing. The low-rise housing types are further classified into four groups: single-detached dwelling (equivalent to a single-family dwelling), semi-detached and town/rowhouse, low-rise apartments (duplex, or dwelling in a building with no more than four stories), and mobile home. Each group is categorized in eight vintages: before 1940, 1940-1960, 1961-1977, 1978-1983, 1984-1995, 1996-2000, 2001-2010, and post 2010.

It is essential to note that the building types vary from region to region and might not suit precisely fit the introduced classes and may need to be adjusted based on each class's definition. Statistics Canada(statcan.gc.ca) has described each building type [62]. Structural features and ownership are two main factors to classify buildings. Generally, the classification of building types in the SHEU dataset has had some changes toward lowering the number of classes with broader involvement while the periods of vintages are growing. Referring to SHEU-2015 makes it possible to re-classify the Kelowna housings based on the building typology and vintage. The master building dataset resulted from the overlay of datasets provides the essential data to categorize the available accommodation in Kelowna city.

The various building types of Kelowna are displayed in Figure 28. Using the SHEU classification allows to re-classify them into four principal groups, which are further divided by eight vintages, as shown in Figure 29. In the Kelowna dataset, the single-family dwellings, town/row houses, and mobile homes have similar categories and definitions as the SHEU-2015 dataset. The other sorts of housings with more than a unit/household in a building and the story number lower than four are categorized as low-rise apartments. This category includes the Duplex (support two families), Tri/Fourplex (cover three to four families), Condominium, and Multi-family houses. Figure 30 shows the distribution of residential records with reference to the SHEU-2015 housing classification based on typology and eight vintages.

Figure 28: Some examples of building types in Kelowna city (*source: GoogleMap of Kelowna city, **2018 BC Step Code [63])

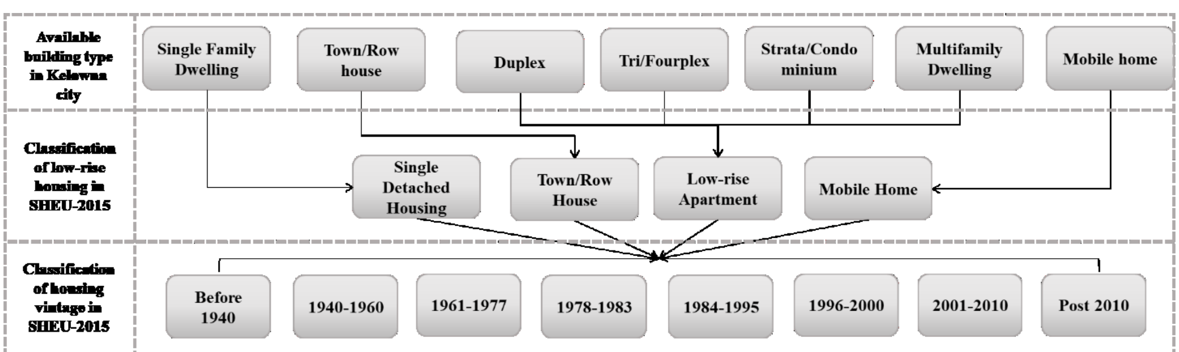


Figure 29: Reclassification of Kelowna building types to four principal low-rise dwelling type in eight vintages

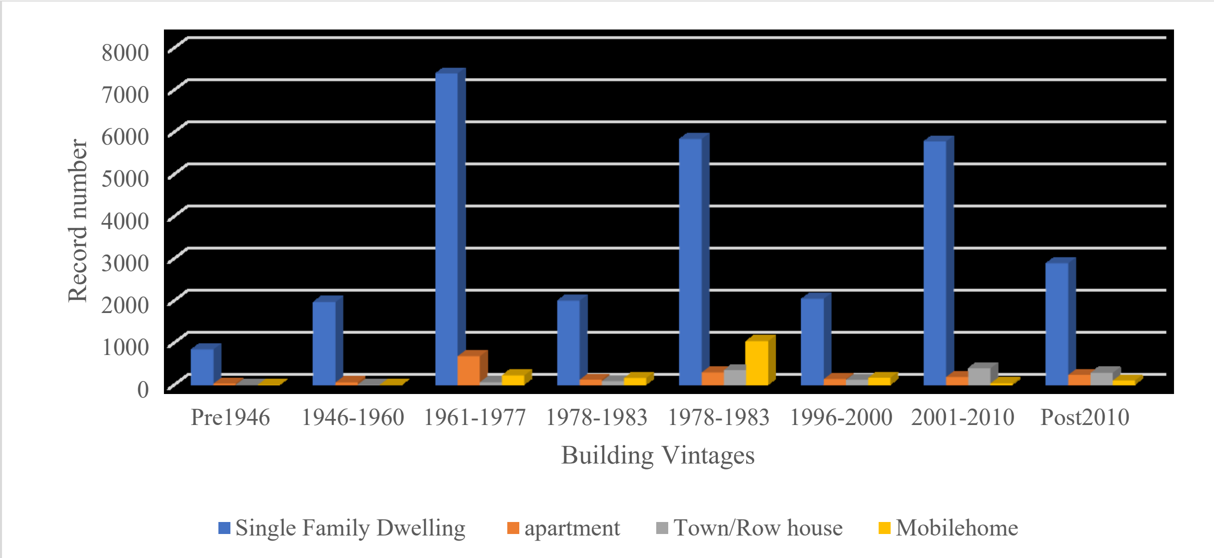


Figure 30: The distribution of building types per eight vintages

As expected, the single-family dwelling is the leading housing type in eight vintages by covering 89 percent of the dataset. Hence, it could be considered the primary building type in Kelowna city for archetype development. The other option for selecting the primary archetypes is going over each building type separately and picking out the vintages with the highest frequencies. However, it does not mean the total housing coverage of the second option would be more than the first one. Given the energy data available for low-rise buildings in Canada are better informed by the building age rather than the type of structure. Therefore, the study made a strong case for modeling the archetypes following the eight vintage classifications, noticing that single-family dwellings are the priority of any retrofit planning.

3.2.2. Permit data processing

The methodology of permit data processing is segmented into five steps as structured in Figure 31. The permit dataset has descriptive characteristics and provides the primary classification on permit status, type, and subtype. They facilitate the dataset's organization and management in the first step. At this level, the permits with the closed, new, and open condition, which indicates the operation has been implemented or currently is in progress, were considered. Then, filtering the dataset based on their permit types allows access to two primary groups of building permits and heating permits with 24,648 and 26,762 records, respectively, ready for data processing. Each dataset consists of many missing elements in the work description column. The qualitative nature of research for permit data processing does not call for filling missing values. The immediate need is to clarify keywords and enrich the dataset with a more profound knowledge of the dataset information. The processing section is followed by various data mining methods and techniques to draw the critical factors affecting the building energy behavior through the heating system alteration or building physics intervention. The derived parameters and characteristics from the permit dataset will be mapped with ModelCity based on the parcel identifier (KID), the common attribute between the two datasets. Consolidation of the permit processing outputs with the building data information determines the direction in which the permit data is enabled to enhance the initial urban building data either through urban building data refinement (available in GIS dataset) or building energy model improvement.

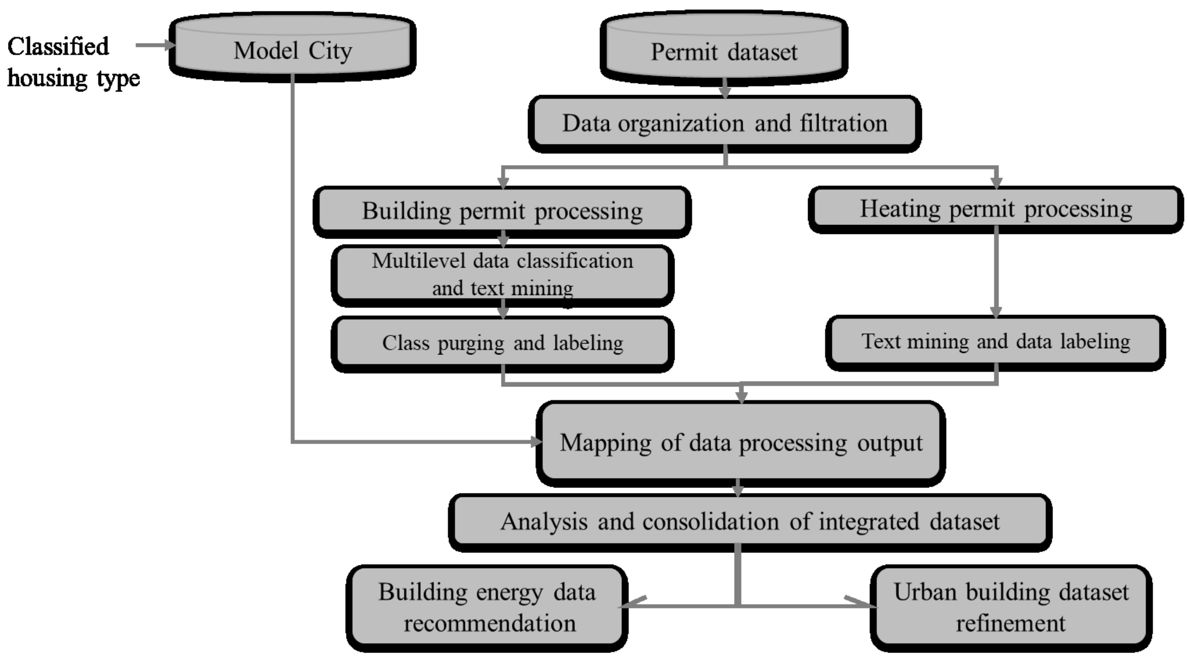


Figure 31: The workflow of Permit data processing

3.2.2.1. Building permit

Data on building permits deliver distinguished factors influencing the building energy performance. The variety of constructional subjects covered in the building permit subtype demands multi-level data classification and class purging based on the work description interpretation. Besides, text mining is the other technique applied to go over the work description and derive the building energy-related components from the non-categorized textual data. Focusing on the permit subtype, it provides a functional sort that allows organizing the dataset into three main groups; (1) the issued permits characterized based on housing type, (2) the licenses established based on the practice type, and (3) the non-relevant to low-rise housing permits. Each category has some subcategories demonstrated in Figure 32. Table 8 also provides the description and statistical information of subcategories.

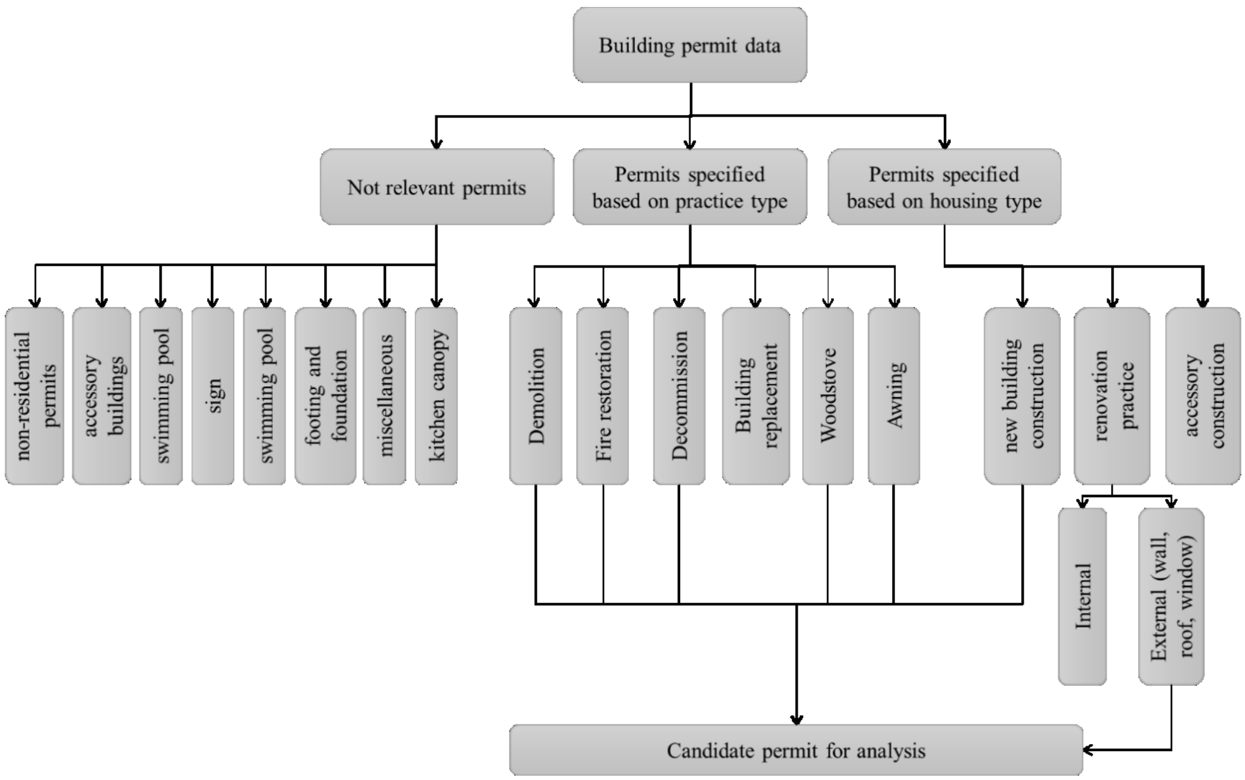


Figure 32: Classification of building permits based on permit subtype

The primary group covering 30% of building permits supports three principal practices divided by building type, including the renovation of buildings, construction activities on the accessory structures, and construction of new buildings in the lot. The first subcategory, named renovation, is divided into interior building assemblies and envelope construction. The priority of data deriving is the exterior construction assemblies, which could help define the buildings' refurbishment level. Since the absence of renovation type subdivision, text mining in the work description of the group is followed to explore any maintenance based on the construction assembly types that occurred outside of the building. The result shows that the work description sentences are not well-structured. Also, there is a wide variety of terms with possible typo errors for a certain building component or location. Therefore, it needs an extensive trial to detect the appropriate keywords came alongside to put them under the envelope improvement class.

In the second subcategory, Kelowna's accessory facilities consist of non-heated spaces, such as a garage, shed, and deck, which are subject to elimination from the candidate dataset influential on the building energy performance. The construction of new dwellings is the critical subclass that updates the building dataset concerning the number of housing in the lot, housing type, and buildings' construction year based on the permits' applied-date. Although the building permit registers the building intervention in real-time, the lack of integration between the permit and the GIS dataset covering the building information inquires the building data refreshing.

The second group of permits established based on activity type fills 11% of the dataset and represents six practices that could influence the buildings' energy performance. Demolition refers to the destruction of habitation in the parcel and comes alongside a measurable floor area in the dataset representing the living areas that do not consume energy anymore. Fire restoration indicates the count of buildings repaired because of fire damage. The work description displays different sorts of renovation and reconstruction that were treated with text mining techniques to be classified as the new construction or envelope improvement (with the classification of assembly types). Decommission class introduces the illegal spaces, particularly suites that have been come to a halt. In this subcategory, there is no reduced gross floor area. That means the suite's functionality as a separate space has been withdrawn. At the same time, it is still a livable part of the house and does not reduce the total heated area.

Woodstove, awning, and building replacement are the final three practices that might affect the building’s energy performance in the dataset. Woodstove represents a specific type of heating system in buildings with wood or electrical energy sources that could be reviewed versus other heating systems frequent in the heating permit dataset. Another important subcategory is the awning or canopy installed to shade windows and transparent surfaces. This factor controls the heat transfer through radiation. However, there is insufficient information to show clearly the application of awnings. The last case as building placement shows mobile homes' moving in a lot without any description related to construction renovation or insulation installation on the slabs (platform). So, it does not clearly affect the factors that might change the building energy consumption. Finally, the third group of permits, called the non-relevant to low-rise housing, covers 54% of the dataset and targets non-residential buildings and subjects, such as the sign, swimming pool, footing, and foundation, etc. This group is explained in Table 10 and entirely ignored from the candidate dataset.

Table 8: The distribution of records based on the defined classification of permit subtype

Principal dataset classification	permit subtype	# record	building renovation		accessory construction	new building	applicable records	record check
			internal	envelope				
1_ permits specified based on housing type	apartment	702	459	31	29	183	214	702
	SFD	4754	1620	81	819	2234	2315	4754
	SFD with suite	1145	651	1	73	420	421	1145
	four families	133	21	3	2	107	110	133
	three families	14	4	0	1	9	9	14
	two families	320	87	0	11	222	222	320
	townhouse	318	114	2	9	193	195	318
	mobile home	240	51	0	60	129	129	240
	boarding house	22	8	4	1	9	13	22
	carriage house	189	0	0	0	174	174	174
	congregate house	17	3	4	1	9	13	17
			Description					
2_ permits specified based on action	demolition	1527	The parameter of demolition alongside the floor area attribute shows the amount of building cleared off in parcels. It has been provided for all housing types such as single-family dwelling, single-family, mobile home, townhouse, etc.					
	fire restoration	250	fire restoration as a means of building renovation includes a broad range of construction activities from a small part of the building component to a complete building replacement.					
	decommission	597	withdraw some spaces, mostly illegal suite, from service. The data is not supported with the quantitative measure of areas converted to non-heated just in 6 records with an average of 60m2.					
	building placement	91	It mostly refers to the moving of mobile homes					
	woodstove	292	the attribute is valuable in terms of representing the other sort of heating system in buildings.					
	Awning	361						
3_ non-relevant to low-rise housing permits	non-residential permits	7954	It refers to permit related to agricultural, institutional, industrial, and commercial groups					
	accessory buildings	2871	It refers to accessory structures such as a deck, shed, garage, porch, etc.					
	swimming pool	2851						
	sign		Graphical signs using for indicating the presence of something					
	footing and foundation		it addresses the construction practices footing and of foundation					

	miscellaneous	it refers to boarding and retaining wall practices
	kitchen canopy	The attribute mostly covers non-residential buildings such as restaurants.
	fireplace	A few records (#2) are available; the subject is covered under the heating permit umbrella.
total records in all status		26189
not applicable records (status Expired (E), Canceled (L))		1541
records with no work description		7527
candidate records for text mining after removing irrelevant records		10611

In addition to six identified operations, two keywords of the suite and basement are used with high frequency in various building permits such as new construction, decommission, fire restoration, etc. Both the suite and the basement are critical spaces in any building type that change the building energy consumption depending on their use-type. Given the Kelowna municipality information, suites and basements are applied as living areas, typically for temporary rent. Hence, they were considered separate attributes in the dataset for further analysis where the processed building permit output meets the building information (in ModelCity dataset).

3.2.2.2. Heating permit

In Kelowna, 31,731 building records have applied for the heating permit. The permits can be categorized into four principal categories: “commercial,” “kitchen canopy,” “standards,” and “residential.” The commercial and kitchen canopy is focused on non-residential buildings. The standard subcategory with 979 records and 692 missing work descriptions covers the mandated standard for installing heating devices. The residential subtype with 26,762 unique permit numbers supports heating permits for residential buildings in three applicable statuses: closed, open, and new. The absence of work descriptions for 17,093 records, besides the lack of documents, to support such missing values resulted in departing more than half of the heating permit dataset. However, the other records with work descriptions do not follow any standard or classification to represent the variety of permit subjects. Each record might be a collection of abbreviations, shorten form of words, numbers, hashtag (#), at sign (@) without a dictionary to clarify them. Hence the frequency of meaningful texts in the work description is the only source for data mining for the 13,050 candidate records (the records with any words).

Regarding the context of the permit category, a variety of keywords in the field of heating system fuels and types were searched. Besides, the work description delivered some hot words which basically referred to a specific operation. The most frequent word, repeated 6020 times in the dataset, is “re & re,” which was written in the other forms as well, such as “re and re,” or “re/re.” The proximity of this keyword with specific characters, such as “HWT,” the acronym of “hot water tank,” and the “furnace,” which were also the following significant characters in the dataset, highlighted the importance of “re & re” as an essential operation. Searching the work description with detailed information proved the convergence of “re & re” with two practices of “remove” and “replace or reinstall.” These facts are essential data for the interpretation of the other found characters. However, searching the arbitrary and irrelevant letters for describing a certain action needs a trained algorithm with huge time for text exploration, considering it might not lead to an accurate result.

The list of heating system-related keywords with high frequency are listed in Table 9. The provided explanation relies on the facts available in a few work descriptions and does not necessarily reflect the general characteristics of the detected word. The deriving attribute process was followed by associating them with the permit dataset to generate a reduced, processed dataset ready for mapping with the master building dataset. It allows discovering the pattern of characters in relation to each other, and also versus the variety of building data like types and vintages for further processing.

Table 9: Influential keywords in heating permits with the derived facts

No	Searched characteristics	# record	Description (the supplementary facts are derived from a low number of the work-description available in the data set)
2	Furnace	4,313	Gas furnace, natural gas furnace, or only furnace have a high frequency in the dataset accompanied with re & re. Many records show furnace and HWT candidate for re & re at the same permit. The frequently mentioned heat rating for a furnace is 60000 BTU/hour. The mentioned efficiencies are 80% and 96%.
3	HWT (hot water tank)	4597	HWT or the hot water tank is the other important object coming alone or with a furnace alongside re & re. The frequently mentioned tank capacity is 40, 50, and 60 gallons.
4	Gas	2496	Gas is the commonly used word to support the heating source coming alongside the furnace or HWT. It is also used besides BBQ and fireplace and in all cases represents the primary source of heating system.
5	Water Heater (tank)	936	It refers to a heating system type. In some cases, the used fuel of gas has been mentioned, without pointing out the quantitative information.

6	Fireplace	806	It's a heating system type supported primarily with gas. Though, in some cases, the gas has converted to electricity as the primary source.
7	Boiler	213	It is a heating system mentioned in the dataset with low frequency and is used chiefly for pools.

Chapter 4

4. Result and Discussion

4.1. Data consolidation

4.1.1. Mapping the master building dataset and building permit output

The building permit analysis output concluded in a processed dataset with eight practical attributes influential on building energy performance, displayed in Figure 33. The relatively high frequencies for the new building construction, suite, and basement put them at the candidate dataset's priority and worth considering as a characterized information for mapping with the current building dataset (ModelCity). Two decommission and demolition attributes are essential factors to update the building dataset regarding the constructional alteration done over the last updated date. Awning, woodstove, and envelope assembly improvement are critical to developing or refining the archetype model. However, they comprise less than three percent of the candidate dataset. Therefore, it is sufficient to display typical features or a refurbishment level for any identified archetypes in Kelowna city.

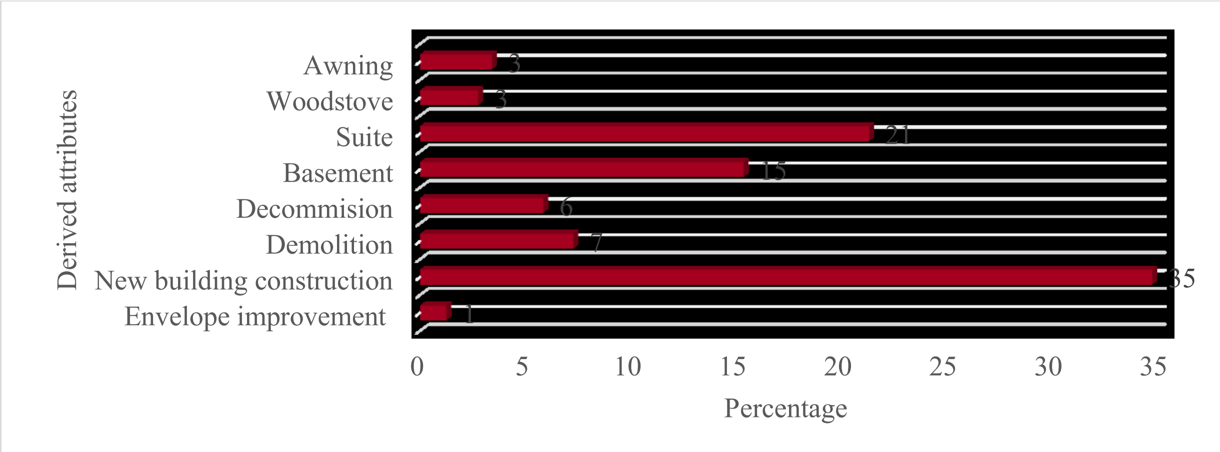
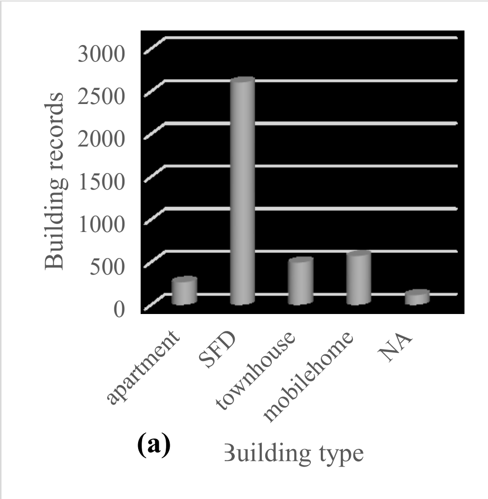


Figure 33: Percentage distribution of identified attributes relative to the total relevant records

Mapping the processed dataset with ModelCity (the GIS source of the building information) shows the effect of five selected attributes on the current buildings and boosts the initial resolution of the building dataset. Adding the new construction and their applied date attributes allowed updating the construction year for 1053 buildings. It also filled the missing construction year for 269 building records and qualified them for being categorized based on vintage, the critical parameter for Kelowna archetype development. Figure 34 shows the residential parcel distribution variation in each vintage category before and after mapping the new construction resulted from building permit analysis. Updating the construction year gave almost a 20% rise to the total building records of post-2000 vintages and reduced 16% from the entire buildings before-2000. Shifting numbers between categories did not affect the selected archetypes' priority mentioned in section 5.1. By Focusing on the building typology, most new construction records were belonged to the SFD type and almost were matched with the indicated type in the dataset. The considerable difference was related to 109 buildings, classified as non-low-rise housing (named as not applicable (NA)) in ModelCity, which switched their class to SFD for 109 cases.



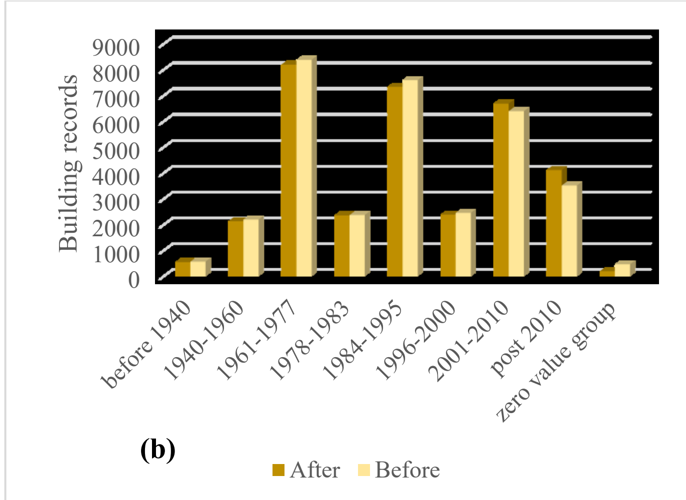


Figure 34: (a) The distribution of the newly constructed housing per the primary building types, (b) The distribution of buildings based on their vintage before and after updating the construction year

Demolition is the next identified attribute that was evaluated versus the building gross floor area available in ModelCity. As mentioned, demolition refers to the total floor area that does not consume energy anymore. From 680 demolished recorded mapped with building dataset, 309 cases have been subject to the new construction, meaning they are still active. There is no document to show the rest of the 371 building records are replaced with a living area. The summation of the gross floor area for 371 buildings provided in the permit dataset and ModelCity shows a considerable measure of heated area reduction, 17,000 and 100,000 m^2, respectively. The lower total floor area in the permit dataset has eventuated from many missing values for the demolished floor area. However, it also needs local authentication to confirm the calculated measure in the ModelCity dataset.

Facing the decommission attribute and buildings with the suite attribute resulted in deactivating 522 suites explored in the dataset. Decommission only limits the suites' usage as a separated unite and does not lead to the floor area reduction. The elimination of deactivated suites resulted in 1342 active suites mapped approximately with SFDs in ModelCity. Comparing the population of SFDs with a suite versus SFDs without a suite provided an impressive result about the occupant's density. The output, demonstrated in Figure 35, indicates the average population of an SFD with a suite is almost two persons more than an SFD without a suite. The average, mean, and standard deviation of an SFD with a suite are 5, 4, and 2.1, respectively. SFDs without suit represent 3, 2, and 1.6 for the average, median, and standard deviation, in order. Respecting the EnerGuide rating system's assumption for the occupant's density, three people per household in SFDs, the permit data confirms if the suite is not considered within the building. Where the suite is included, SFDs in Kelowna city is more populated.

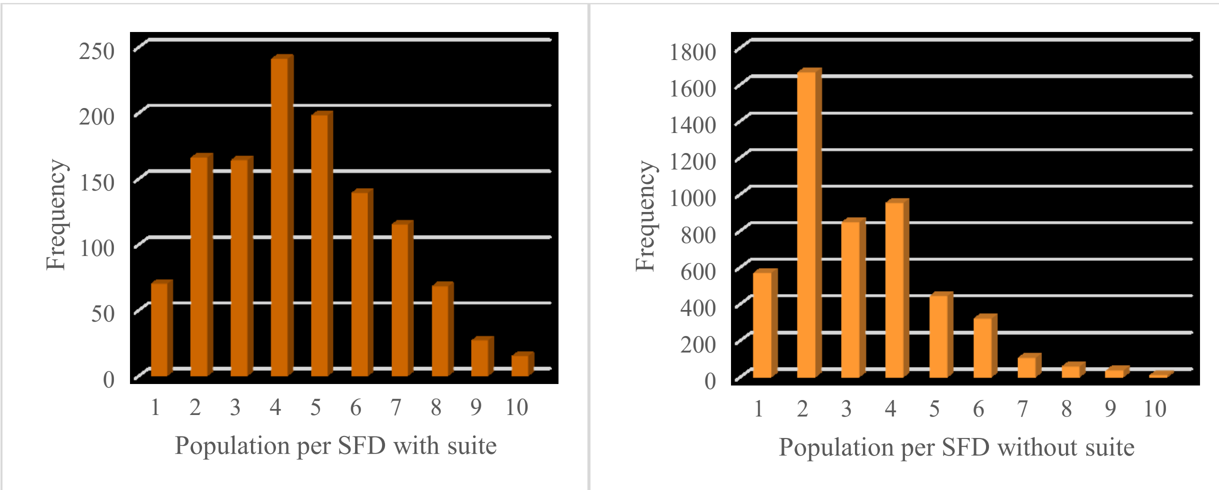


Figure 35: Distribution of population per single-family dwellings (SFD) (a) with a suite, (b) without a suite

Comparing the basement's attribute against various building types notices the basement as a characterized structure of SFDs by receiving more than 95% of the basement records. In 43% of cases, the basement is considered a suite (conditioned space) with an average gross floor area of 275 m^2. Kelowna municipality announced that most basements are subject to living and potential for being used as a suite. Although the basement's availability as a typical structure of SFDs is confirmed in the BC Step Code 2018 documents [63], the approval of being heated in Kelowna city SFDs needs further survey from supplementary sources, such as the construction permit attachments, including the architectural characteristics of buildings.

4.1.2. Mapping master building dataset and heating permit output

The heating permit analysis output yielded a reduced data set with six characters related to the heating system and a high-frequency keyword (re & re) representing the system improvement if it was mentioned next to any characterized attributes. Figure 36 displays the frequency of identified features. Furnace and HWT are at the front of the applied heating permits. The highly repeated proximity of them with the hot word “re& re” supports the interpretation of improving HWT and the furnace as a primary heating system in the dataset. Gas as the fourth meaningful character addresses the fuel type of heating systems and is majorly used as a supplementary description alongside the other attributes. The rest of the characters are focused on the heating systems consisting of the boiler, fireplace, and water heater, making a shallow portion of the applicable heating permits.

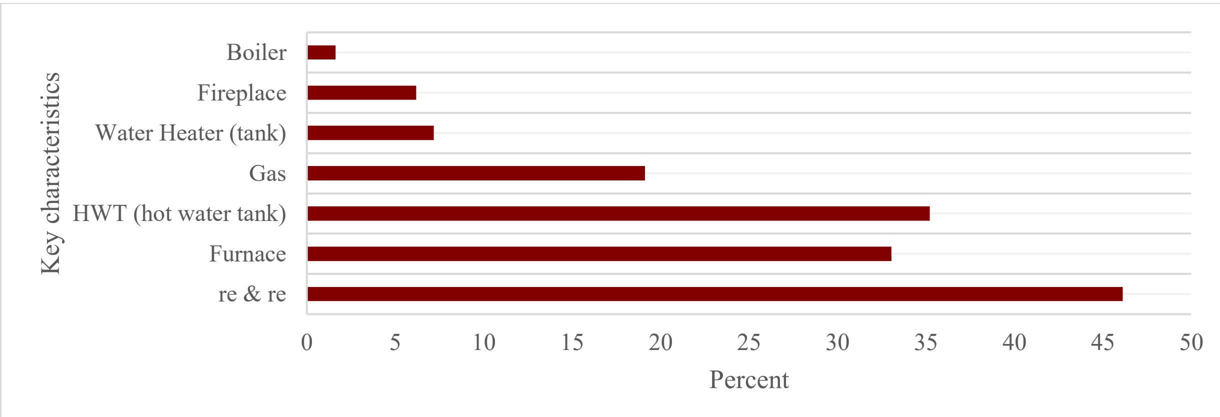


Figure 36: The ratio of explored characteristics frequency to the total candidate records

Mapping the heating permit output with the master building dataset brings a good insight into the heating system distributed between the low-rise residential building types of Kelowna. Since most keywords are focused on heating systems and there is low quantitative support for HWT, the study eliminates the analysis of HWT mapping and considers it as a potential field for future investigation. The distribution of the permits consisting of the gas attribute is entirely within low-rise housing groups (4500 buildings). In half of the cases, it came across the furnace, corroborating the gas's assumption as the furnace's primary fuel source.

Figure 37 shows the spread of applied date for the furnace permit beginning in 2000, increasing at a moderate rate over the years and rising steeply after 2015. The proximity of furnace and “re & re” was occurred for almost 67% of the buildings that received the furnace character. Therefore, the inference of this proximity is removing and replacing the furnace for those buildings. That means the age of the furnace, and consequently, the efficiency of that has changed.

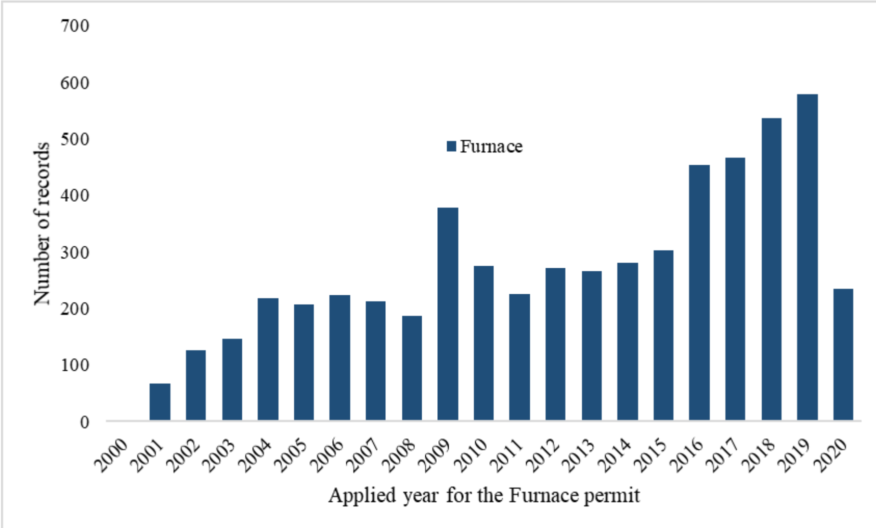


Figure 37: Temporal distribution of permits applied for furnace improvement

Figure 38 shows the percentage of buildings applied for furnace permits per primary building group divided by eight vintages. The graph's first implication is the widespread use of the furnace as the heating system between four building categories consisting of SFDs, mobile homes, townhouses, and low-rise apartments. The second message that the graph emphasizes is the ratio of heating system improvement with respect to the building type and vintage. In the SFD category, almost fifteen percent of buildings belonged to the 1960-2000 vintage (which makes up 67% of the total number of SFDs) have updated their furnace with a post-2000 furnace model. It also essential to notice, 26% of buildings was constructed after 2000 and, as a result, have a post-2000 furnace model. In the apartment category, the vintages of 1984-2000, including 45% of low-rise apartments, have applied for the post-2000 furnace permit. Mobile homes are at the top of furnace improvement, with nearly 96% of the building records with a post-2000 furnace model. Townhouses are the second building type with an improved furnace for most buildings. Regarding the distribution of most facilities and the percentage of applied records in post-1978 vintages, more than 60% of buildings have at least a post-2000 furnace model. The update of the furnace as the primary heating system for nearly one-fifth of the city's low-rise housing, as well as the consideration of the post-2000 construction for nearly 28% of buildings, are a critical indicator for revising the minimum furnace efficiency value for the City of Kelowna if the EnerGuide rating standard is used as the default.

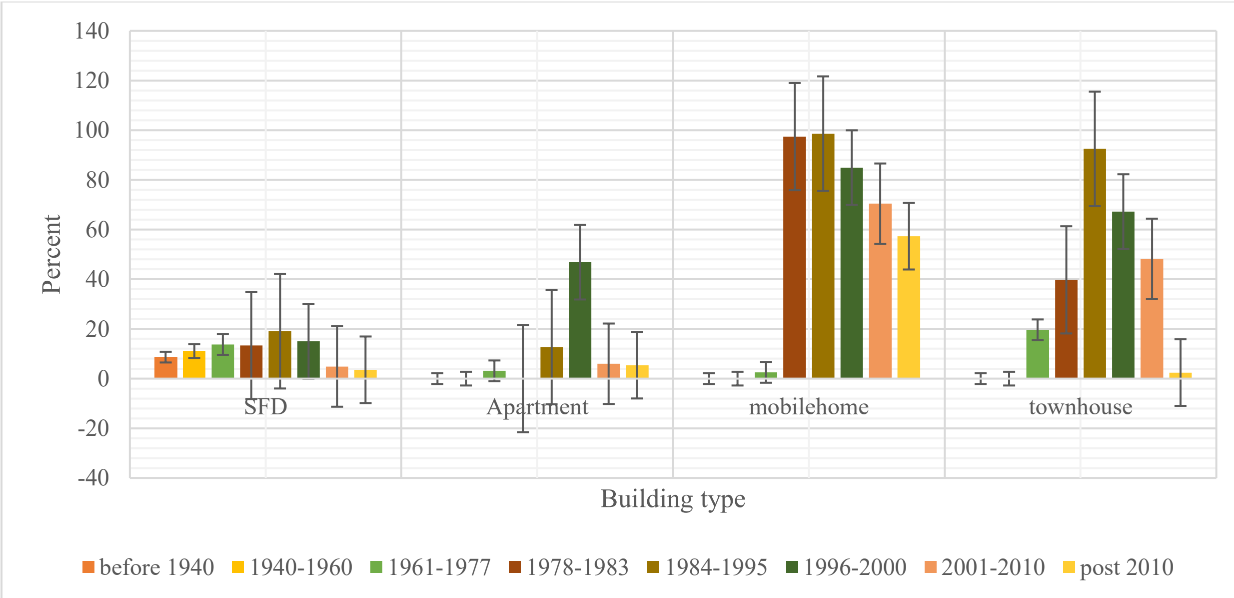


Figure 38: The ratio of applied furnace permit in each building vintage divided by building type

Examining the water heater and fireplace distribution provided a sparse mapping to the variety of building types and vintages, which does not allow us to draw a concrete concept that the archetype applies them as the primary heating system. However, almost 30% of buildings that received fireplaces were applied4 for the furnace. That means one of these two has been used as a supplementary heating system.

4.2. Data modeling

4.2.1. Characterization of geometry data

The work on 3D city modeling and providing geometry for UBEM energy simulation is focused on the most populated neighborhood in Kelowna named Rutland, which contained 10% of the residential buildings of Kelowna. The process of city modeling leverages the ArcGIS City Engine to generate the three-dimensional geometry at the level of detail 2 (LOD2) using the master building dataset. The Feature Manipulation Engine (FME) tool will then be utilized to transform the model to standard CityGML (lod2solid) format, synchronized with the UBEM requirement.

Creating the 3D model of the selected region relies on some essential geometry input. The master building dataset supports initial data like footprint and height for generating the buildings at LOD1 while creating the sloped roof of buildings in LOD2 needs more geometry information. The additional attributes to fill in the available gap in the current data set based on the following step requirements are the building roof type, slope, roof rise, and eave height. Respecting the geometry model proposed by BC Step Code for the B.C. archetype, the roof type of one and two-story housing follows the conventional slope of (4/12), and buildings with more than two stories are considered flat. Assigning the slope size to the qualified sloped-roof housing allows scaling up the roof rise and then the eave height by subtracting the roof rises from the maximum building heights. The provided data for individual buildings will be used to generate the 3D city model systematically.

4.2.1.1. City modeling using ArcGIS CityEngine

The model uses ArcGIS CityEngine for developing 3D city modeling of the Rutland neighborhood. ArcGIS CityEngine is a commercial 3D city information modeling (3DCIM) application that is capable of creating the built environment quicker and smarter than traditional modeling techniques [64]. CityEngine is a part of the ESRI suite that deals with geospatial vector data using a procedural modeling approach to generate city models via predefined rulesets created through ce.py (python interface) as a built-in CityEngine library. The rules are defined through a Computer Generated Architecture (CGA) shape grammar system enabled for complex parametric modeling either via ready-to-use rule packages or custom-designed packages created by users [65]. Figure 39 shows the schematic 3D modeling workflow in CityEngine used in the current project. The process of 3D modeling in this project is initiated by importing the selected region's footprint in shapefile format. The shapefile is enriched with additional attributes that facilitate the model generation in CityEngine and also simulation in the next step.

The required modeling characteristics consist of ridge height, eave height, roof type, and the building footprint. The other supplementary attributes essential for UBEM simulation were added, including building function, construction year, gross floor area, and the updated furnace efficiency. By determining the projected coordinate system of the 2D-model in CityEngine, NAD 1983 UTM Zone 11N, data importing is completed. The second step focuses on moving data from 2D to 3D using the downloaded rule set of "Building_From_Footprint.cga [66]" to create the building geometry. This rule allows applying the provided attributes for generating a model with a second Level of Detail (LOD2) leveraging the footprint extruding and sloped roof creation based on the footprint plan, roof rise, and eave height attributes.

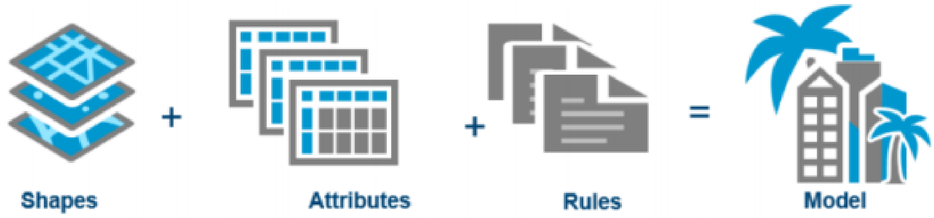


Figure 39: 3D city modeling workflow in ArcGISCityEngine using the contribution of 2D model, supplementary attributes, and predefined rules

The last step exports the provided model to a required format for the data transformation process. CityEngine supports all Esri products, Collada, Autodesk FBX, 3DS, Wavefront OBJ, RenderMan RIB, mental ray MI, and e-on software's Vue. This research uses the ESRI file GBD format to export the 3D model for Esri applications and FME data transformation. Although the GBD file is readable by FME, the exported data is converted to a multipatch shapefile using ArcGIS Pro to remove the shapes and textures attached through modeling.

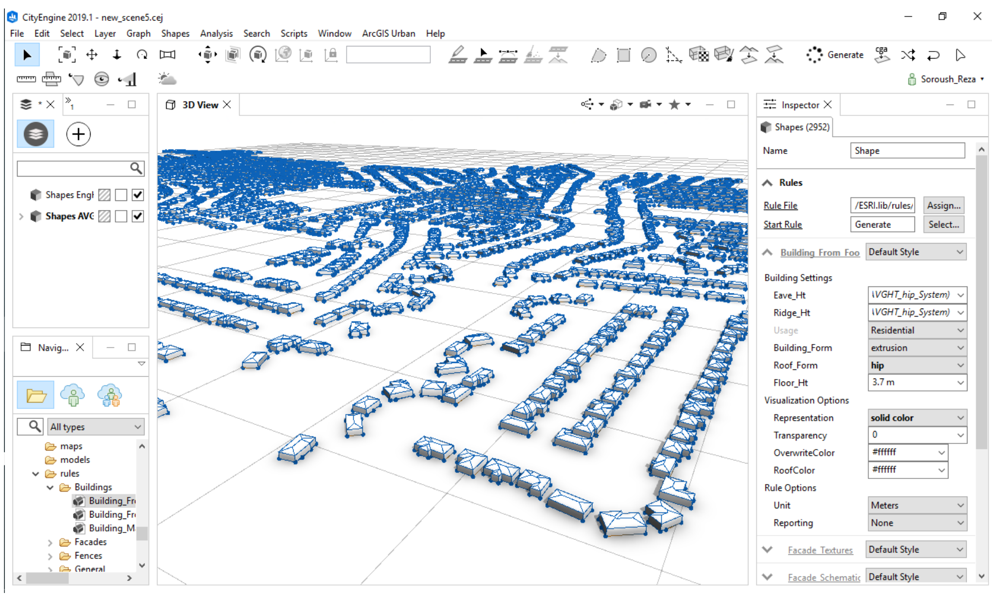


Figure 40: The generated 3D city model using ArcGIS CityEngine and "Building_From_Footprint.cga" predefined ruleset to model the buildings in LOD2 using the eave height, ridge height (maximum building height), and roof type attributes alongside the imported footprint

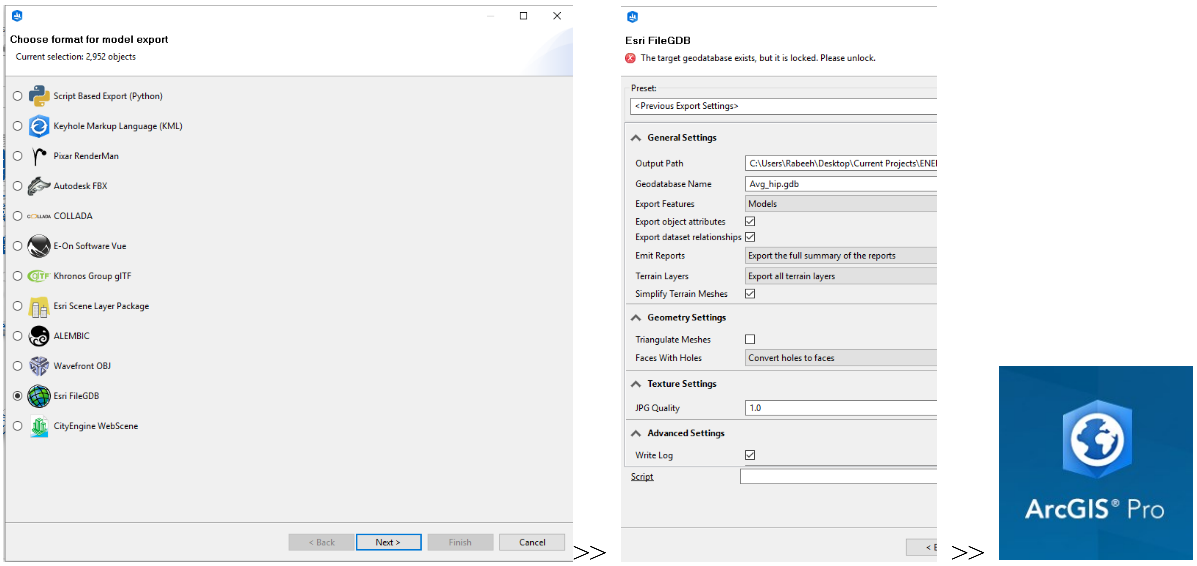


Figure 41: Exporting the 3D model to GBD format and using "models" as export features, and using ArcGIS Pro to covert GBD to Multipatch shapefile

4.2.1.2. City model transformation to CityGML using FME

Feature Manipulation Engine (FME) is the chosen software within this study to convert the 3D multipatch shapefile to CityGML. FME is a data integration platform supporting spatial data worldwide [67]. FME facilitates data conversion and reduces compatibility issues through pre-built transformation codes and algorithms. FME allows customized workflows regarding the level of data requirements and validates the output to ensure the quality of the datasets. This dataset utilized within this project is a 3D multipatch shapefile containing geometry features without supporting semantic information such as the wall, roof, or ground surfaces. CityGML is a flexible data model, and there is no obligation for semantics if the geometry is valid enough[68]. However, the explained transformation process aims to provide an FME based CityGML LoD2 conversion process following the CityGML guidelines[41]. One crucial element, particularly for energy demand simulations using CityGML LoD2, is building solid. For making elements of CityGML in its LoD2, LoD2Solid is more recommended relative to LOD2multisurface due to its potential for delivering accurate volume[69][70]. LoD2MultiSurface geometry is a geometry made by connecting multiple surfaces, which may or may not form a watertight geometry. In contrast, LoD2Solid guarantees a watertight geometry, critical for any building volume calculations. Figure 42 shows the differences between both geometry models, multisurface and solid.

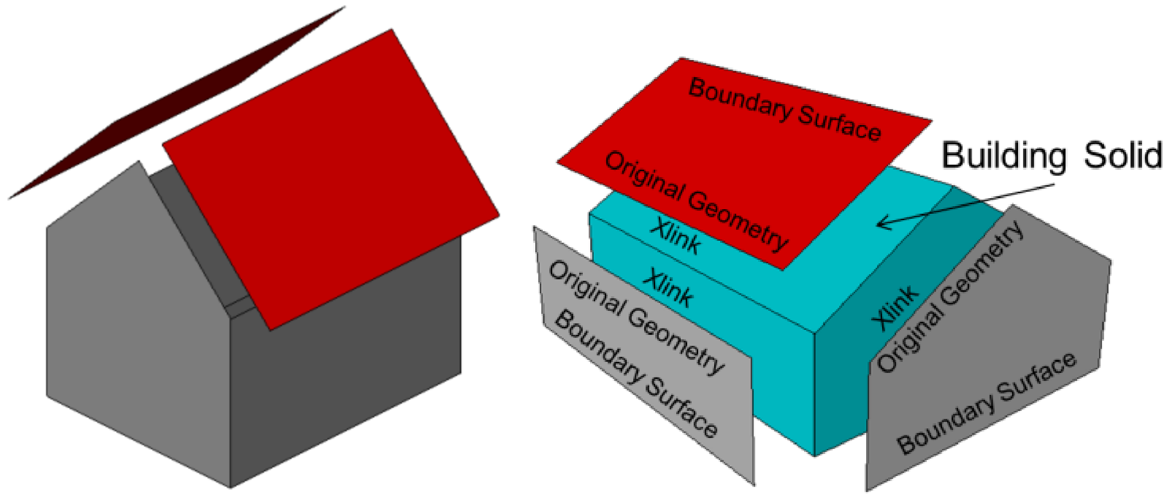


Figure 42: (left) LOD2Multisurface vs. (right)LOD2Solid

LoD2solid contains external references (Xlinks) to the bounding geometry such as wall surfaces, roof surfaces, ground surfaces[70]

The UML diagram of CityGML's building model is designed by the schematic extension module and presented by Groger et al .2012 [41]. CityGML is an application schema that stands for Geography Markup Language in XML format. Data conversion begins with the import of the multipatch shapefile of the geometry. The FME based conversion process from 3D multipatch shapefile to CityGML LoD 2 is developed by Padsala et al. (2020) and openly available on the GitLab[71].

The first phase is characterizing and consolidating the building geometry in harmonized polygons. The process starts with labeling each building with a unique identifier (gml_parent_ID) using AttributeCreator. Applying two levels of GemetryCoercer first turns the geometry to the brep-solid to provide a solid-volume in three-dimensional space specified by a collection of tightly closed surfaces. Then it shifts the layers to a composed surface, allowing members of the composite to be oriented consistently with reference to their neighbors. The composite surfaces are then de-aggregated into individual polygons for which a child id (gml_ID) is assigned. However, when the data is de-aggregated, certain surfaces based on the nature of input geometry (here 3D multipatch) need to be converted to a mesh with respect to the X-Y plane. Therefore, the utilization of the GeometryPartExtractor helps to extract these surfaces as TriangleFan geometries. The Triangulator, followed by geometry coercers, is used to convert the triangle fans to mesh and composite surfaces. In the next level, all processed polygons use two transformers of UUIDGenerator and StringConcatenator to receive a unique gml_id. The double use of GeometryPropertySetter prepares surfaces to take an LoD name of LoD2MultiSurface and get the primary condition for obtaining semantic roles. The hierarchy of transformers is shown in Figure 43.

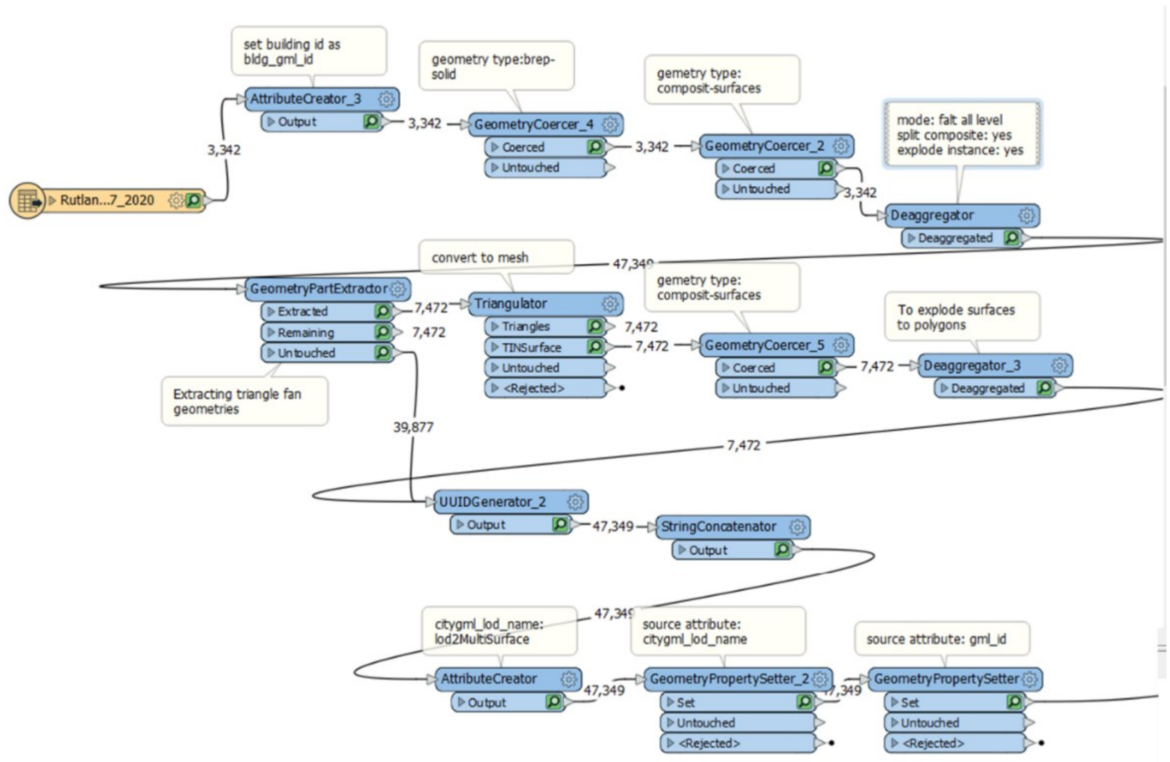


Figure 43: Reading multipatch shapefile, geometry modification, labeling buildings and surfaces, and putting geometry traits to LOD2multisurfaces

The second phase focuses on correctly classifying the semantics of composite surfaces. It begins with extracting the bounding coordinate of all the polygons illustrated in Figure 44. This helps identify each set of building features (wall, roof, and ground surfaces) and makes them ready for geometry processing using GeometryValidator. GeometryValidator gives the option of calculating the Z_normal of vertices applicable as the surface-normal. The vertex normal is stored as a geometry measure on vertices. It is the role of the MeasureExtractor to take the Z_normal of vertices and places them in attributes. Finally, the transform of Tester evaluates the Z_normal of vertices and classifies polygons into the correct semantics of wall, roof, and ground surfaces along with the feature role of "BoundedBy." This workflow successfully writes the CityGML document of the roof, wall, and ground surfaces. In the tester classification, the vertices with the absolute value of Z_normal, lower than 0.1, were classified as wall surfaces. The rest of the range [-1, 1] is shared between the roof and ground surfaces by receiving the Z_normal of (> 0) and (< 0), respectively.

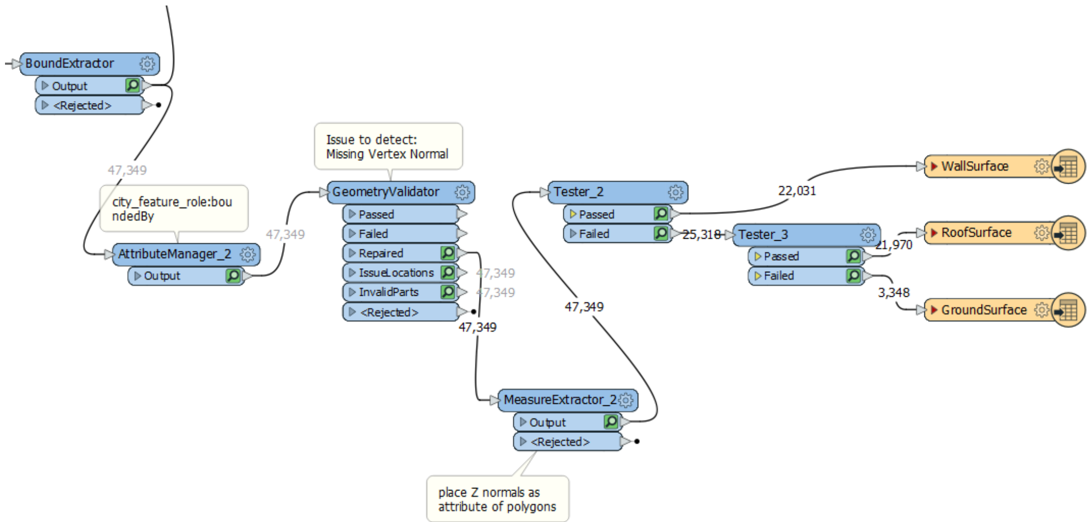


Figure 44: Processing polygons to categorize them as the wall, roof, or ground surfaces

The last step is defining the building as a solid composed of composite surfaces. According to the CityGML guidelines, building solids must be determined using the xlink connections with the respective boundary surfaces, as shown in Figure 45. To achieve this, first GeometryRemover is used to remove buildings' geometry properties. AttributeCreator sets citygml_lod_name as lod2solid and Xlink_href as gml_id, which will extract all the gml_ids of boundary surfaces (multisurface) and replaces them as xlink_href. GeometryPropertySetter will assign xlink_href as a reference link to all surfaces. In the second level, features are aggregated under the hierarchy of childID (gml_id for polygons) and parentID (gml_parent_id for buildings) using the Aggregate transformer and changed back to fme_brep_solid. Using AttributeRenamer allows for the renaming of the determined building identifier, bldg_gml_id to gml_id, and GeometryPropertySetter set geometry trait as lod2solid. Depending on the critical attributes (coming from the GIS dataset) which need to be associated with the building, it is possible to use AttributeManager to modify or manipulate them before connecting to the Building CityGML Writer. The hierarchy of transformers is shown in Figure 45.

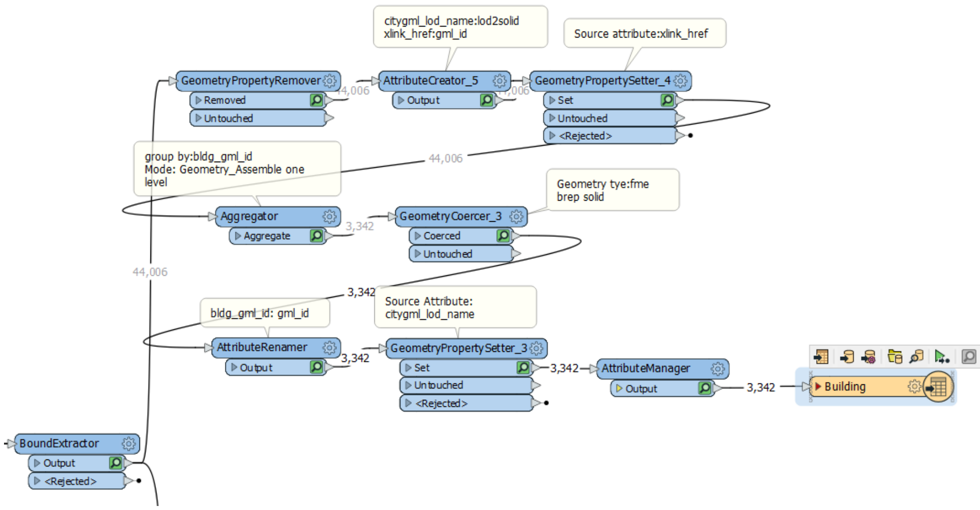


Figure 45: Writing Xlink reference for connecting polygons to relevant buildings

The output, CityGML LoD2 dataset, can then be visualized in the FME data inspector, as shown in Figure 46. As available in the input 3D multipatch shapefile, the coordinate system will be written to the output CityGML file.

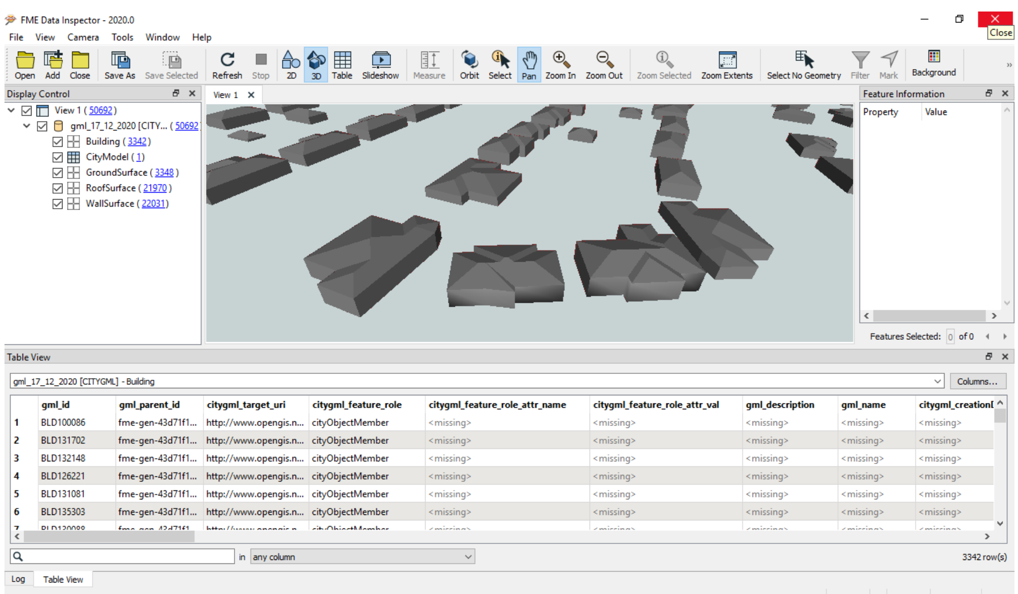


Figure 46: Visualization of the model using the FME inspector

4.2.2. Characterization of archetype parameters

The energy profile of eight identified archetypes in Kelowna city was modeled with the support of the HOT2000 data library, Kelowna energy audit received from National Resources Canada

(NRCan), and the research outputs provided by HosseiniHaghighi et al. [72], focusing on housing energy data in Kelowna city. HOT2000 is an energy simulation suite focused on low-rise residential buildings developed by NRCan. It also provides a reference housing energy data for different vintages starting from 1920 on. HOT2000 presents specified values for the thermal resistance of building construction assemblies differentiated by climate zone. The proposed values support 5-time series, including before 1940, 1950-9, 1960-79, 1980-1999, and post 2000. The available data needs adjustment regarding the local metered data. In this case, the energy audits provided by the energy advisors in the city of Kelowna offer a good insight into the housing energy data. However, the received data is related to the limited number of buildings that have been registered for an energy audit and only represent the average of surveyed data for SFDs divided by vintages.

Characterizing the thermal resistance of the construction data for identified archetypes was done through multiple simulation-analysis utilizing the UBEM application. In this case, relying upon the district geometry, the derived parameters from the HOT2000 data library were assigned to buildings based on their vintages category. Comparing the simulated energy demand versus the B.C. measured data and using the Kelowna energy audits resulted in improving the construction thermal resistance and approaching the B.C. measured thermal demand per identified vintages.

The archetype usage input also utilizes the Kelowna energy audits besides the research output on the Kelowna housing energy data [72]. The Kelowna housing energy data uses the thermostat data analysis to find the essential measures regarding the setpoint temperature and occupancy presence at home. This study uses the clustering method to identify the critical ranges for SFDs dwellings, of which the average value of the most populated cluster was selected to model the Kelowna archetype usage, listed in Table 10. The energy audit provides data on the heating system fuel, type, and efficiency per building vintages.

Table 10: The building usage input data applied for eight vintages

Parameter	Unit	Value	Resource
Occupancy			
Occupant Density	Person per sqm	0.03	[72]
Usage Days Per Year	Days	365	[72]
Usage Hours Per Day	Hours	14	[72]
The internal gain resulted by Persons and home appliances			
Average Intern Gain Per Sqm		4.3	Standard DIN V 18599-2[73]
Convective Fraction	Ratio from 1	0.4	Standard DIN V 18599-2[73]

Radiant Fraction	Ratio from 1	0.5	Standard DIN V 18599-2[73]
Latent Fraction	Ratio from 1	0.1	Standard DIN V 18599-2[73]
Heating and cooling temperature			
Heating Setpoint Temperature	C	19	[72]
Heating Setback Temperature	C	18	[72]
Cooling Setpoint Temperature	C	21	[72]
No Ventilation			
Heating system			
Natural gas furnace	*AFUE	92%	Kelowna Energy Audit

*AFUE, Annual fuel utilization efficiency, is an efficiency rating for furnaces based on the seasonal measure of fuel efficiency regardless of the operational electrical energy.

4.3.UBEM data processing and simulation (SimStadt)

The heating load prediction in the modeled district utilizes the urban simulation platform SimStadt. SimStadt is an ongoing project in the University of Applied Sciences HFT Stuttgart, Germany[74] accessible through Simstadt.ericduminil.com [75]. SimStadt is a workflow-based model established on a series of programmed modules using JAVA 8 that integrates the two domains of urban energy simulation and 3D Geographical Information System[74]. Each workflow sets up different data operations that measure the parameters for approaching low-carbon energy strategies. Two of the workflows employed within this thesis are solar and photovoltaic potential analysis and heating demand analysis. The heating demand analysis consists of five data operation steps, starting with four input dataset acquisitions presented in Figure 47. The weather file providing the monthly average air temperature and global radiation flux comes from the INSEL [76] weather data repository. The repository contains data for many locations globally, and Summerland, particularly in Canada, matches the Kelowna climate zone. The next dataset is the building location and geometry data generated at LOD2, CityGML format enriched with building function and construction year. The building physics is the third dataset representing the identified archetype's constructional data based on Canadian housing vintage. The building usage information as the final dataset indicated the internal load, occupant density, presence, and temperature schedule in the residential house.

The second step focuses on data preprocessing, including data parsing, data quality check, and data reorganizing. In this stage, the coordinate system across the geometry dataset allows access to the relevant weather data for providing the ambient temperature and global and diffuse radiation following the Hay sky model, which suits the low-density urban area [74]. The processed geometry data will be mapped to the related physics and usage data based on the building construction year and function to define the building models. The final step concentrates on modeling individual buildings' thermal load and sending them for building demand simulation using the German standard DIN V 18599-2 [73] (analogous to the international norm ISO 13790 [77]) for monthly balance calculation.

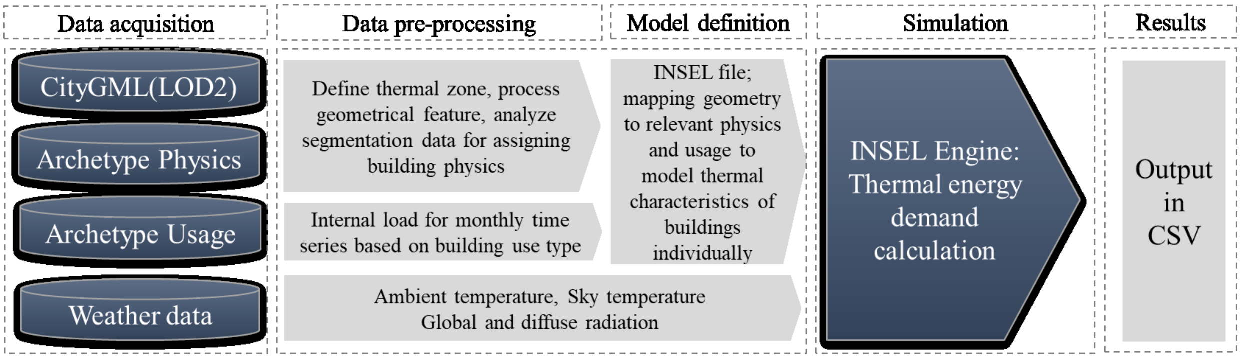


Figure 47: The UBEM data processing using INSEL engine

4.3.1. Simulated model and comparison to measured data, SHEU-2011, SHEU-2015, and NEDB

Following the SimStadt thermal model, the annual heating energy use intensity (HEUI) was calculated per Kelowna buildings. Since the model estimated demand for all available facilities in the selected region, the consumption of buildings with unheated status (discussed as the accessory and additional structure) was eliminated from the output. According to the building vintages, the estimated heating demand is supplied by the natural gas furnace with an efficiency of 92% AFUE. Applying the appropriate efficiency level to the building demands based on their archetype category delivers the building energy consumption.

Figure 48 represents the distribution of HEUI sorted by housing vintage and compares them to the average HEUI of the national measured datasets collected in British Columbia consist of the SHEU-2015 and 2011 and NEDB (2000-2018). The estimated HEUII ranges from a minimum level, 55 kWh/m^2 for the post-2010 dwellings, to a maximum level, 141 kWh/m^2 for the vintage 1946-1960. The mean estimated heating use goes approximately 2% and 22% over the SHEU-2011 and SHEU-2015 datasets, respectively, and 14% below the NEDB (2000-2018). The average of simulated HEUI fluctuates between -6% to +14% relative to the average of the national measured data per vintages. This finding confirms the UBEM performance, given almost a similar deviation was also measured in Section 4.2.1. to compare the primary prototype relative to the average of measured data. Furthermore, considering that 60% to 65% of total housing consumption is heating demand, the measured average for low-rise housing in B.C. provided by Canada Green Building Council[79], is almost close to the average HEUI simulated for the low-rise buildings in Kelowna, as shown in Figure 49.

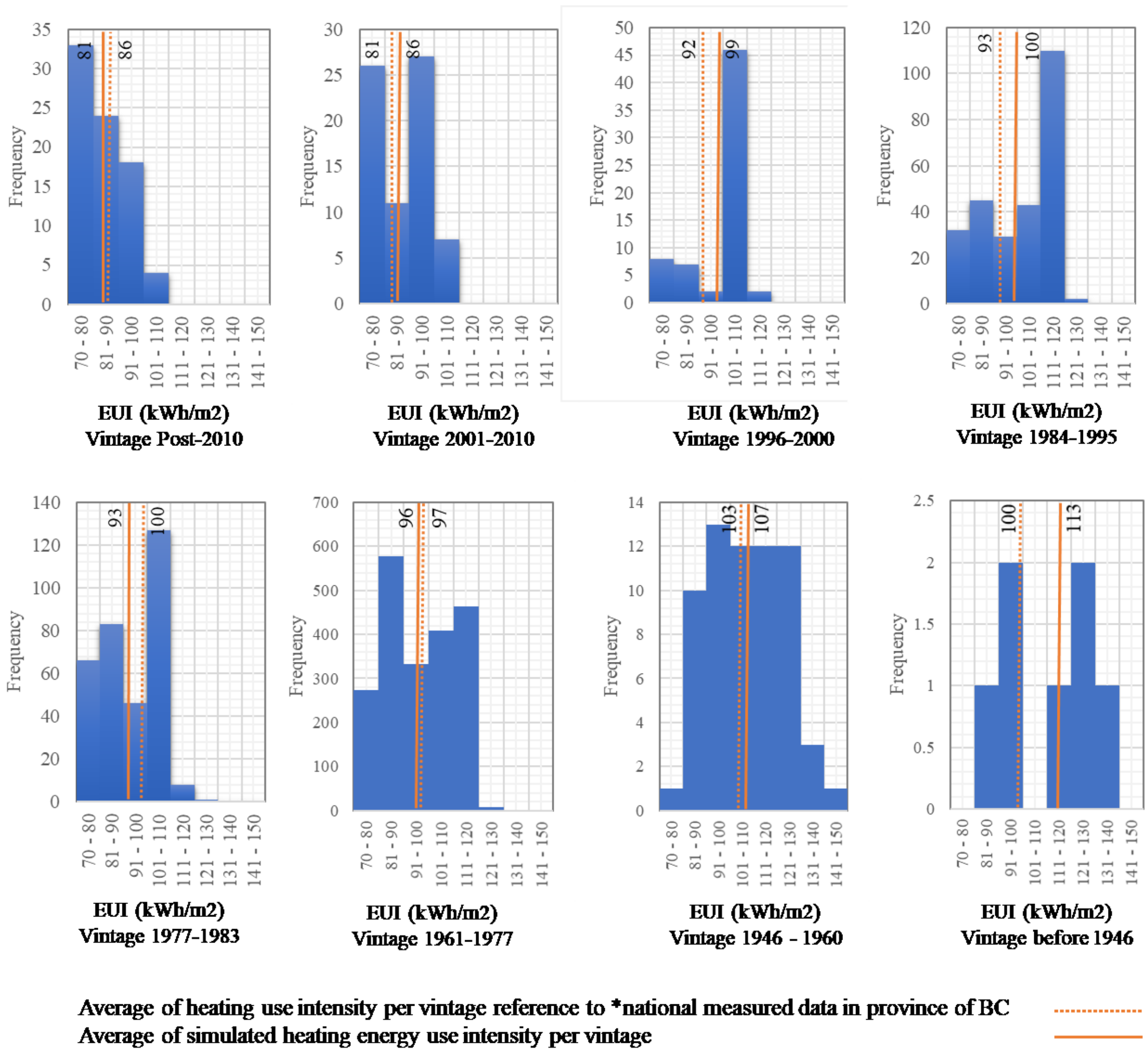


Figure 48: The frequency of estimated building's heating energy use intensity (HEUI) compared to the average HEUI of British Columbia in SHEU-2015, 2011 and NEDB per eight vintages

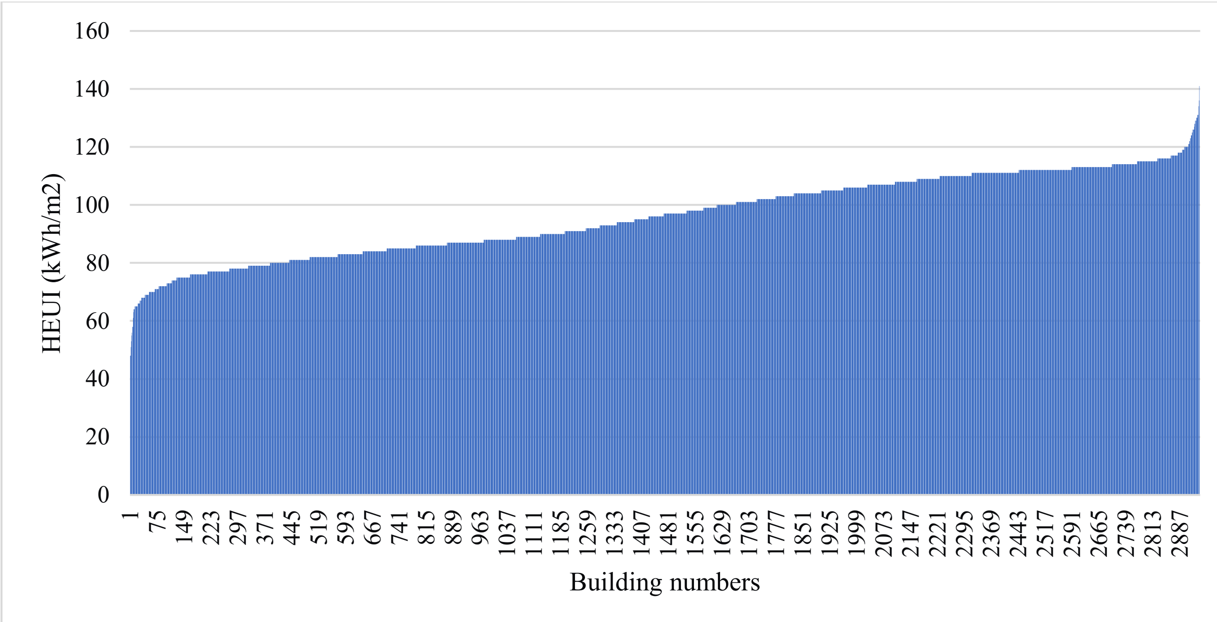


Figure 49: Distribution of the heating energy use intensity (kWh/m^2. yr.) per individual buildings and comparison of the estimated average HEUI in Kelowna (96 kWh/m^2. yr.) and the measured low-rise housing in B.C. (93 -99 kWh/m^2. yr.) provided by Canada Green Building Council

The building geometry characteristics were counted as an influential parameter on the building heating demand in Kelowna. Two ratios of the building floor area to the footprint and the building envelope surface to the volume were instrumental in determining the intensity of heating energy use by providing a strong correlation (R2) of 0.83 and 0.54, respectively, shown in Figure 50. Increasing the building envelope to the volume ratio gives rise to the heating energy use intensity proportionately. In this case, sharing more surfaces for heat transfer through conduction seems more critical than the absorbed radiation through the cold periods and results in more energy consumption. The other factor, the floor area to the footprint ratio, has the inverse effect; when the ratio increases, the energy use intensity decreases. This shows for a determined piece of land, the more the number of stories, the less energy consumption. Both defined geometry factors represent the state of compact design, which originates indirectly from the zoning bylaw. Hence, they are worthy of being considered for further research as influential factors on energy-efficient urban zoning and land management.

A known model limitation is the equal distribution of windows on the building façade regardless of the orientation priorities. The other case is the archetype physics that considers the same construction properties for the basement walls (below ground) as the above-grade walls.

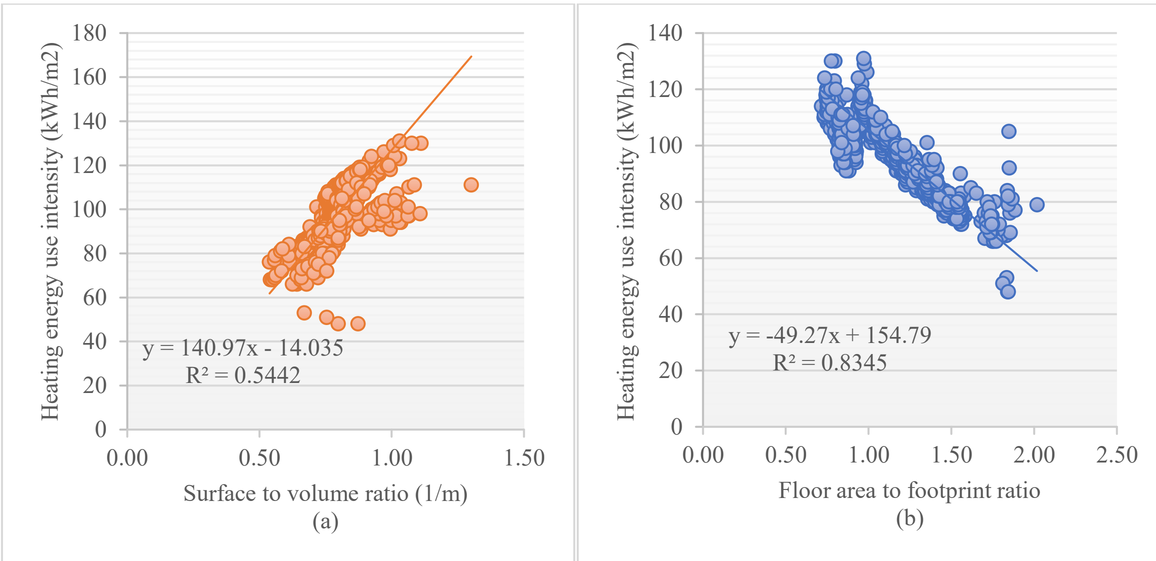


Figure 50: The impact of building geometry on the heating energy use intensity (HEUI) (a) The correlation between the building HEUI and the ratio of the building envelope surface to the volume, (b) The correlation between the building HEUI and the ratio of floor area to the footprint

4.3.2. Mapping of the annual heating energy intensity and carbon footprint

The summation of simulated heating consumption in the Rutland neighborhood (2945 buildings) is 62.6 GWh equals 225,360 GJ. Producing this energy from natural gas with an emission factor of 49.75 (kg/GJ) [78] releases almost 11.21 megaton CO2 equivalent per year. Figures 51 illustrate the spatial distribution of the annual HEUI per individual buildings. The figure shows 40% of buildings have an HEUI above 100 kWh/m^2, 40% of housings have an HEUI between 80 to 90 kWh/m^2 (the average of heating consumption in B.C.), and less than 5% have an HEUI below 50kWh/m^2, which is the B.C. benchmark for the zero-carbon low-rise housing [79]. (the value of 50kWh/m^2 covers the total end-use per building).

The distribution of CO2eq intensity (kg CO2-eq/m^2. yr.) resulting from the heating consumption is presented in Figure 52 and elucidates that more than 80% of buildings are below 20 kgCO2eq/m^2. yr. This number accounts for the measured average of the CO2eq intensity for the B.C. low-rise housing, including all sorts of consumed energy in the building [79]. The mapping of the HEUI and the CO2eq intensity reveals the repartition of the intensive energy users and facilities across the community. Also, it allows prioritizing the active retrofit planning versus the building properties and characters, between which heating system type and fuel improvement are at the front in Kelowna city.

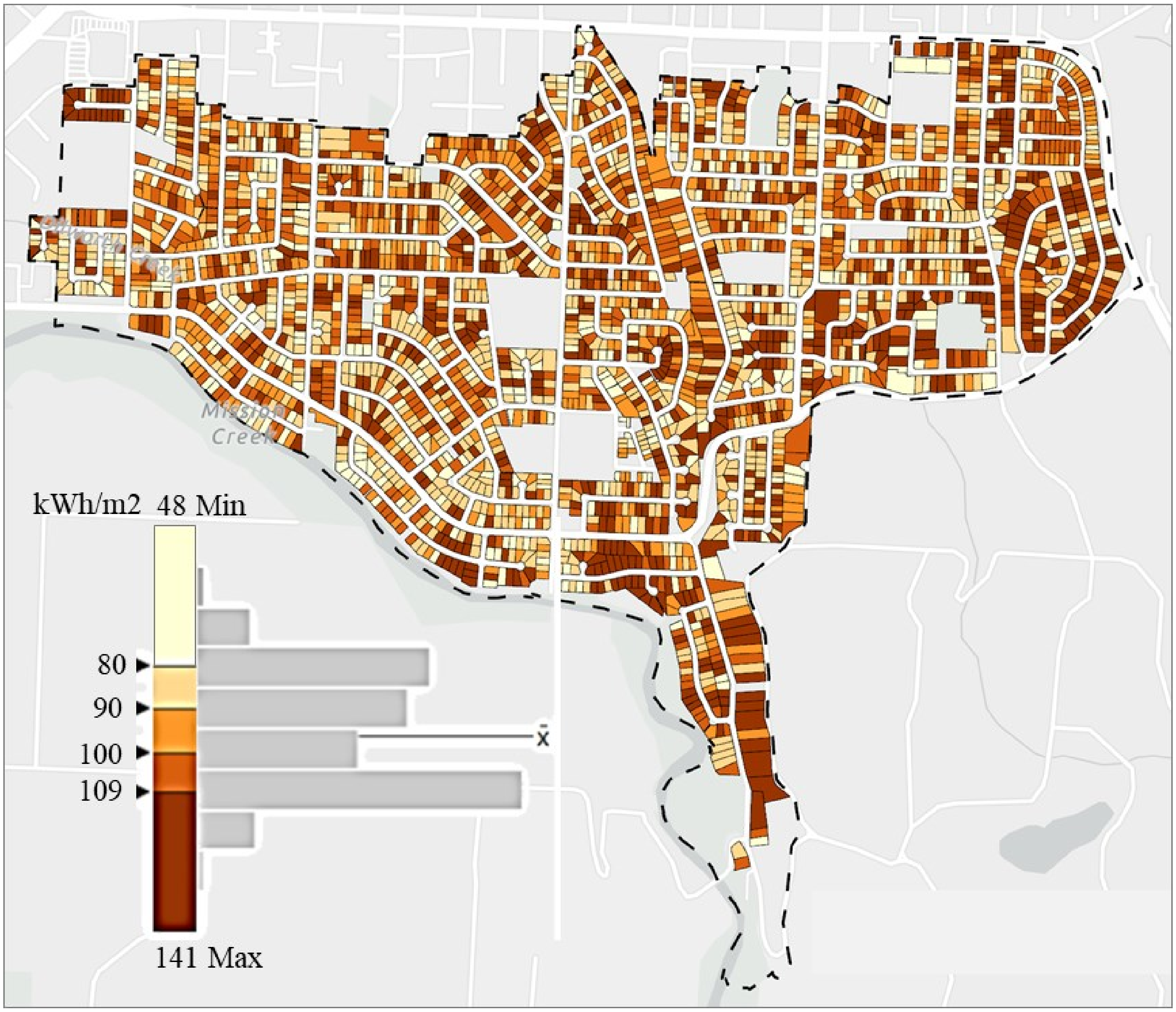


Figure 51: The distribution of heating energy use intensity per individual buildings in the neighborhood

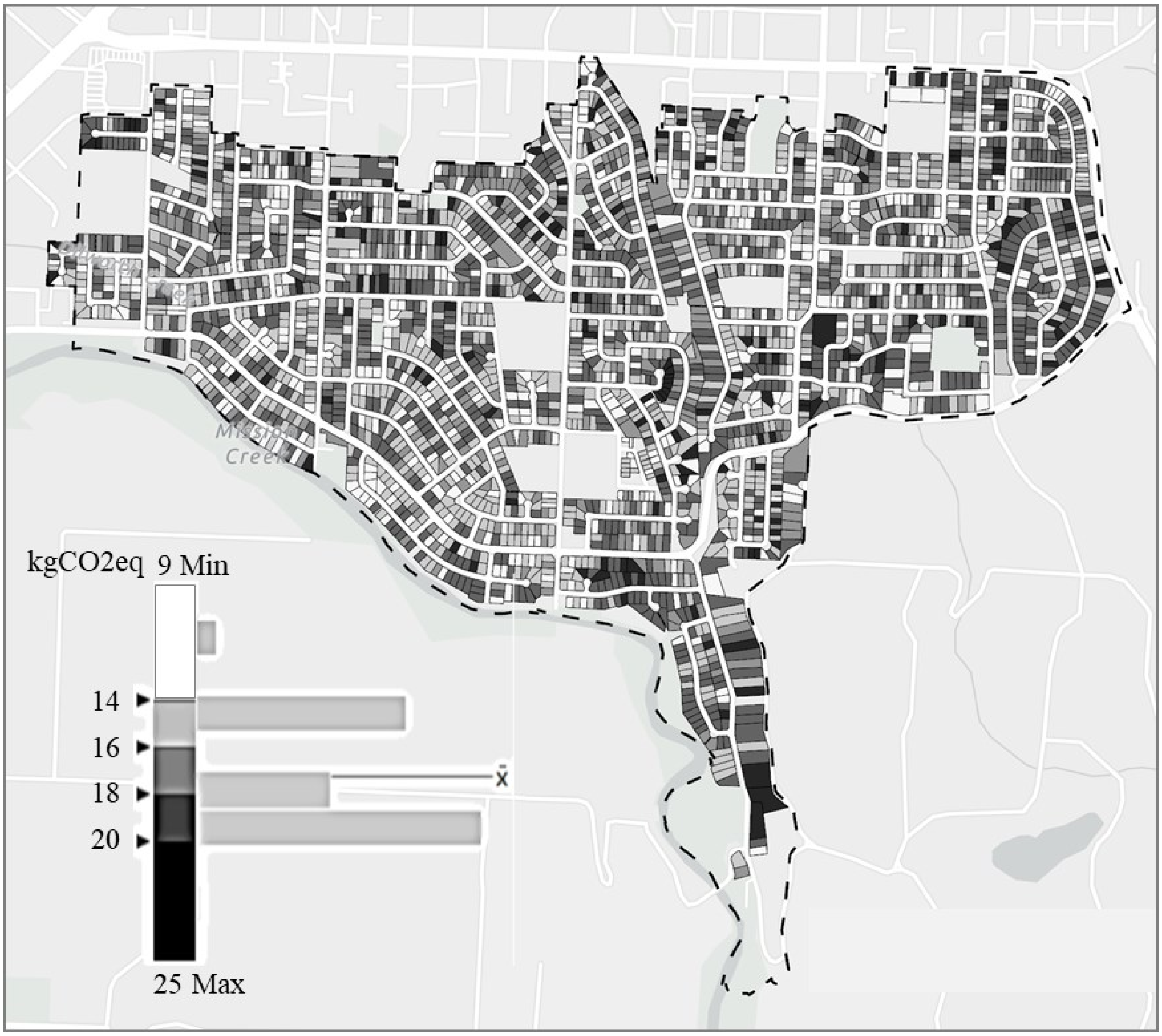


Figure 52: The distribution of annual CO2eq intensity per individual buildings in the neighborhood (the primary fuel is natural gas with emission factor 49.75 kgCO2/GJ)

4.3.3. Photovoltaic (PV) potential

One of the neighborhood's potential for onsite energy generation is the many sloped roofs' surfaces with minimum mutual shading due to the low urban density. Showing the capability of UBEM, the monthly solar potential of individual buildings could be calculated for the roof surfaces. SimStadt allows for picking up the proper roof surfaces conditional to certain area limitations. Therefore, assuming the constraint of a minimum surface area needed for an economically viable PV system of 40m^2 and the usability of 70% of the individual roof surfaces, a total roof area of 313,052 m^2 was obtained for the PV assessment (near 50% of the entire roofs available in the community). Taking the PV panels' efficiency, i.e., the solar panel's electrical power (kWp) divided by the panel area at nominal operating conditions of 1kW/m2 irradiance and 25^{0}C, and a performance ratio of 85%, the total potential energy was calculated as 33 GWh annually. Following Canada's milestone for the GHG reduction solution, the solar potential is a significant source for on-site power generation and carbon depletion. Such solar photovoltaic systems produce clean energy and can satisfy nearly 57% of the community's annual thermal demand assuming the electric heat generator with 100% efficiency is applied. Coupled with the PV potential, the heating systems' retrofitting would highly affect the community's aim toward GHG mitigation. The various system types for heat pumps offer a significant range of coefficients of performance (COP). For instance, two heat pump options with the possible sources of air and water could provide average COPs of 2 and 3.5, respectively. The usage of each option, the heat pump with COP 2 and COP 3.5, could reduce the total electricity demand for heating demand up to 54% and 28% of the reference level of a direct electrical heating system, respectively, as presented in Figure 53. In the case of using the heat pump with COP 2, the annual PV potential will be sufficient to support the total heating demand. In the second scenario, using the heat pump with a COP of 3.5, the potential photovoltaic energy not only satisfies the total heating demand of the community but also provides a surplus of nearly 16 GWh, which supports the other sectors running on electricity. However, the applicability of such potential is dependent on neighborhood infrastructure to save PV power through storage facilities or share it via the local grid system.

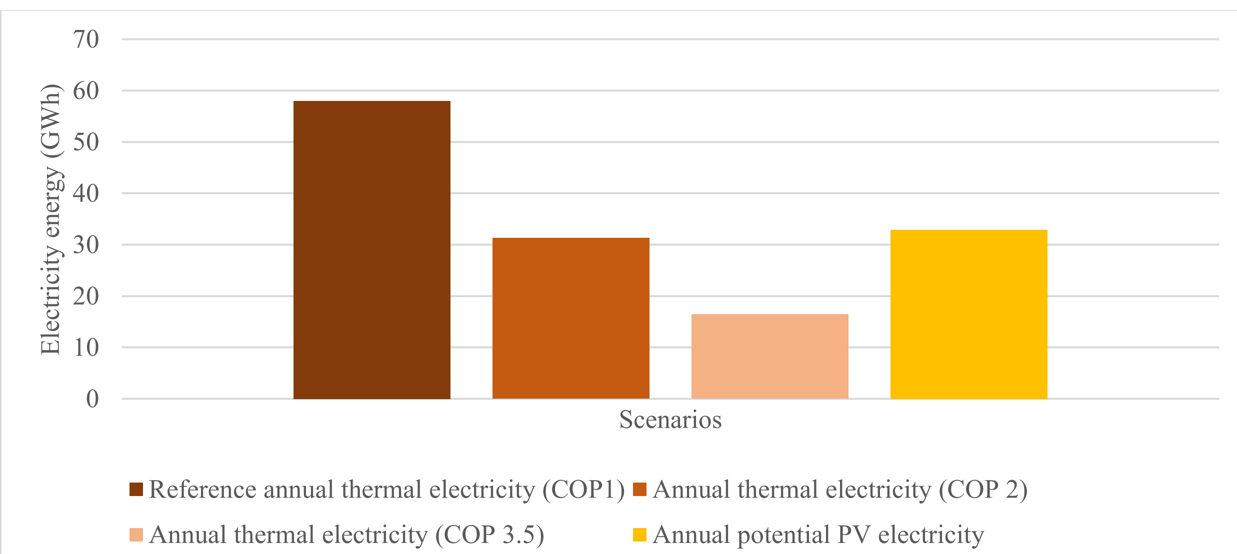


Figure 53: Annual thermal electricity required for the reference scenario (COP=1) and two heat pump systems with COP of 2 and 3.5 with the potential PV energy from rooftops

Figure 54 tries to combine on-site PV power generation and the possibility of the heating system retrofitting using a heat pump with COP 3.5 per individual buildings. In this scenario, almost 70% of buildings can satisfy their total heating demand, and 65% could support the other energy users either within the building or the community. Such a map discloses the possibility of on-site renewable energy potential relative to the district opportunities and the required annual thermal load versus the possible retrofitting scenarios. It bridges the urban community layout to reference and alternative energy assessment strategies for context-sensitive supply-demand management.

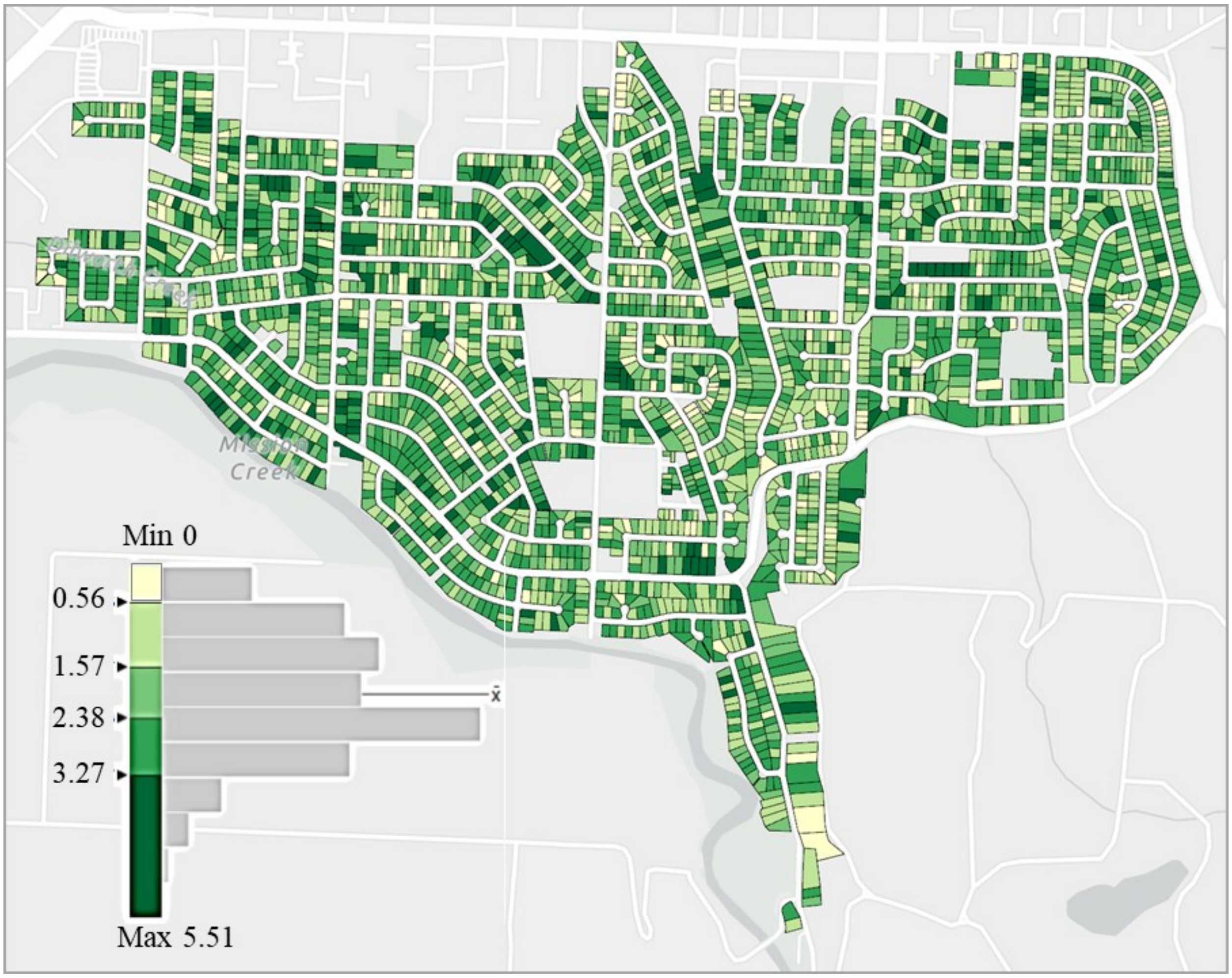


Figure 54: The fraction of potential PV power to the total heating demand generated with heat pump with (COP 3.5) for individual buildings in the neighborhood

Chapter 5

5.Limitations, Recommendations, and Conclusions

5.2.Limitations and Recommendations

Using building information from different sources and with different purposes requires multi-scale data processing, integration, and synchronization with UBEM. In this case, using a shared building identifier is of priority when utilizing multiple datasets conveying building information. It facilitates the aggregation of various datasets and reduces the uncertainty resulting from the dataset's inconsistencies. The recommended methodology in statcan.gc.ca [80] for identifying structures is based on a linkage between the building and the census subdivision. This is done by determining the census subdivision's unique identifier plus an ordinal 6-digit code that is special for individual buildings in that census subdivision. This way provides a geo-referenced identity for buildings, stabilizes building data collection efforts, and assists data processing on the national scale.

A second issue that needs amelioration is the urban building geometry provided by Kelowna city. In the case of Kelowna, the building footprint and height were derived only from the orthophoto dataset, which is basically suitable for use as a background map or measuring horizontal distances with 1meter resolution. Using the orthophoto to detect building footprints is concentrated at the roof boundary. It means the only visible part of the building from the top vision (as an image captured in orthophoto) is the roof, and the boundary of the roof is considered as the footprint. The correct detection of roof boundary is subject to an error between 1% to 48% based on the color quality of objects, proximity of elements in the image, and color classification methods[57]. Another problem is the difference between the roof and the footprint where accessory buildings such as garages are attached, or the building has multi-level setbacks.

The other error occurs where the building elevation is measured. In this case, the possibility of deviation between the measured and real data increases if there is no access to the original sensor data from a satellite or aerial camera. The next concern is accurately deriving building height, which is mostly measured from the structure's base to the top. In this case, the top height will be detected perpendicular to the chosen spot on the ground (building footprint), assuming no angle in the building. Therefore, such measurement is not accurate for objects and elements that taper, lean, or have a setback from the base[58]. Hence, improving urban geometry using other supplementary sources, preferably in three dimensions such as the LiDAR dataset, is strongly suggested.

The third point focuses on permit dataset quality, which strongly needs initiatives to collect and share documentation related to various permit types. The dataset needs to define a systematic frame and workflow to reduce human errors that currently resulted in numerous missing and ambiguous work descriptions in Kelowna permit data. Twenty-five years of experience on the accumulated information in the permit dataset and an analysis of the influential factors on building energy performance rating in Canada are significant sources informing a detailed classification of required data for enhancing the quality of permit data collection and raising its functionality for realistic building energy planning and energy policy.

The lack of accessible defined archetypes for low-rise housing in Canada led the project to use the HOT2000 default datasets for the Summerland climate zone similar to Kelowna city. The absence of local metered energy also limited archetypes' potential for being calibrated. Hence, it is crucial to access regional energy surveys or utility bills to customize developed archetypes and better integrate demand and supply-side modeling. However, archetype templates' availability following the building types, sizes, principal vintages, and climate zones in Canada are the other valuable missed energy datasets that systematically accelerate the evaluation of urban building energy performance.

On the technical side, the geometry conversion process from multipatch shapefile to CityGML using FME requires a deep understanding of CityGML schema (for building modeling) and the FME transformers that enable multi-step data conversion to reach the standard format. The generated surfaces in CityEngine do not follow the same quality for transformation. Hence, it needs multiple operations to harmonize surfaces and prepare them for semantic roles. The complicated workbench of FME data transformation is limited to a multipatch shapefile generated by CityEngine and does not support datasets emanated from other sources. On the other hand, the research leverages multiple platforms to draw a 3D city model from the building footprint and convert it to CityGML format of lod2solid. This process needs multidisciplinary expertise from geoinformatics, urban design, and building energy modeling. Combining those steps to create an automated workflow facilitates the use of the model in urban planning and supporting energy-sensitive decision-making.

5.3.Conclusion

The Canadian government's objective to reach net-zero emissions by 2050 was one of the primary motivators behind the current research. Identifying a workflow to focus on building stock energy modeling by addressing the gap between the accessible urban data and energy performance evaluation would greatly help allocate necessary resources towards the net-zero emissions goal. The study employed multi-level data-operation and analysis to fill in the recognized inconsistencies and inaccuracies of datasets and increase the urban building energy model (UBEM) input data resolution. The provided multiscale workflow successfully manifested urban datasets' challenges and generated an urban building energy model based on actively maintained GIS datasets. The output of building data processing was synchronized to create a master building dataset ready for procedural 3D city modeling and automated transformation to the semantic model following the CityGML schema. The integration of CityGML with the developed archetypes in the simulation platform of SimStadt concluded the workflow by estimating the single building's heating load, CO2eq intensity, and on-site energy generation, analyzing the solar and photovoltaic potential using 70% of the building's roof surfaces.

The average deviation between the simulated annual heating use intensity and the B.C. measured data was ranged between -6% to +14% per various vintages. The building geometry factors representing the compactness of structures, such as the building floor area to the footprint ratio and the envelope surface to the volume ratio, have been recognized as critical indicators for demand-side management, which are accessible through developing energy-efficient zoning bylaw. The community thermal load was simulated at approximately 225TJ, resulting in 11 megaton CO2eq emission. Reviewing the district potential for solar access and on-site electricity generation illustrated a significant alternative for CO2eq reduction if new strategies and retrofitting plans for heating systems and potential power management are investigated.

The building geometry analysis and the intervention study on the district PV potential for on-site generation disclosed the importance of input data resolution to bridge the community's actual energy performance and energy planning and policy toward context-based GHG mitigation. Study outcomes have further potential for integration with the online map to facilitate interaction with a broader expertise domain. The data model's barriers include the low compatibility between different urban datasets that affect the consistency of the master building dataset. It is possible to increase the functionality of building stock datasets by using a modulus identifier and also implementing systematic initiatives to collect and share datasets conducting the building information. The geometry dataset analysis also highlighted the need for supplementary sources such as LiDAR to improve the complexity and accuracy of the building models. Regarding the non-geometry data, it is crucial to access the local utility bills and broader energy audits to fill the simulated and real consumption gaps in the archetype development and UBEM performance evaluation in terms of bottom-up retrofit planning.

References

[1] "Energy Use in Canada: Trends Publications | Natural Resources Canada." [Online]. Available: https://oee.nrcan.gc.ca/publications/statistics/trends/2016/totalsectors.cfm. [Accessed: 19-Dec-2020].

[2] "Progress towards Canada's greenhouse gas emissions reduction target - Canada.ca." [Online]. Available: https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/progress-towards-canada-greenhouse-gas-emissions-reduction-target.html. [Accessed: 19-Dec-2020].

[3] "IEA EBC Annex 70 – Building Energy Epidemiology – 'Building Energy Epidemiology: Analysis of Real Building Energy Use at Scale.'" [Online]. Available: https://energyepidemiology.org/. [Accessed: 19-Dec-2020].

[4] "Building Energy Mapping and Analytics CDS | OGC." [Online]. Available: https://www.ogc.org/projects/initiatives/bdgenergycds. [Accessed: 19-Dec-2020].

[5] M. Gangolells, M. Casals, N. Forcada, M. MacArulla, and E. Cuerva, "Energy mapping of existing building stock in Spain," *J. Clean. Prod.*, vol. 112, no. June 2013, pp. 3895–3904, 2016, doi: 10.1016/j.jclepro.2015.05.105.

[6] K. G. Droutsa, S. Kontoyiannidis, E. G. Dascalaki, and C. A. Balaras, "Mapping the energy performance of hellenic residential buildings from EPC (energy performance certificate) data," *Energy*, vol. 98, no. 2016, pp. 284–295, 2016, doi: 10.1016/j.energy.2015.12.137.

[7] C. Delmastro, G. Mutani, and S. P. Corgnati, "A supporting method for selecting cost-optimal energy retrofit policies for residential buildings at the urban scale," *Energy Policy*, vol. 99, pp. 42–56, 2016, doi: 10.1016/j.enpol.2016.09.051.

[8] C. Hjortling, F. Björk, M. Berg, and T. af Klintberg, "Energy mapping of existing building stock in Sweden – Analysis of data from Energy Performance Certificates," *Energy Build.*, vol. 153, pp. 341–355, 2017, doi: 10.1016/j.enbuild.2017.06.073.

[9] O. Conference, "GIS-Based Residential Building Energy Modeling at District Scale Usman Ali , Mohammad Haris Shamsi , Cathal Hoare , James ODonnell School of Mechanical and Materials Engineering , Energy Institute University College Dublin (UCD), Belfield , Dublin 4 , I," no. September, pp. 11–12, 2018.

[10] Y. Chen, T. Hong, X. Luo, and B. Hooper, "Development of city buildings dataset for urban building energy modeling," *Energy Build.*, vol. 183, pp. 252–265, 2019, doi: 10.1016/j.enbuild.2018.11.008.

[11] C. Cerezo Davila, C. F. Reinhart, and J. L. Bemis, "Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets," *Energy*, vol. 117, pp. 237–250, 2016, doi: 10.1016/j.energy.2016.10.057.

[12] "Europe: Typology Approach for Building Stock Energy... - Google Scholar." [Online]. Available: https://scholar.google.com/scholar_lookup?title=Typology Approach for Building Stock Energy Assessment. Main Results of the TABULA Project%3A Final

Project Report&publication_year=2012&author=TABULA Team. [Accessed: 19-Dec-2020].

[13] S. J. Quan, Q. Li, G. Augenbroe, J. Brown, and P. P. J. Yang, “Urban data and building energy modeling: A GIS-based urban building energy modeling system using the urban-EPC engine,” in *Lecture Notes in Geoinformation and Cartography*, 2015, vol. 213, pp. 447–469, doi: 10.1007/978-3-319-18368-8_24.

[14] “Commercial Reference Buildings | Department of Energy.” [Online]. Available: https://www.energy.gov/eere/buildings/commercial-reference-buildings. [Accessed: 19-Dec-2020].

[15] “BEDES Dictionary | BEDES.” [Online]. Available: https://bedes.lbl.gov/bedes-online. [Accessed: 20-Dec-2020].

[16] L. Aelenei *et al.*, “Smart City: A Systematic Approach towards a Sustainable Urban Transformation,” in *Energy Procedia*, 2016, vol. 91, pp. 970–979, doi: 10.1016/j.egypro.2016.06.264.

[17] C. F. Reinhart and C. Cerezo Davila, “Urban building energy modeling - A review of a nascent field,” *Building and Environment*, vol. 97. Elsevier Ltd, pp. 196–202, 15-Feb-2016, doi: 10.1016/j.buildenv.2015.12.001.

[18] “Select TMY3 Files.” [Online]. Available: http://weather.whiteboxtechnologies.com/TMY3. [Accessed: 19-Dec-2020].

[19] “Weather Data | EnergyPlus.” [Online]. Available: https://www.energyplus.net/weather. [Accessed: 19-Dec-2020].

[20] B. Bueno, L. Norford, G. Pigeon, and R. Britter, “A resistance-capacitance network model for the analysis of the interactions between the energy performance of buildings and the urban climate,” *Build. Environ.*, vol. 54, pp. 116–125, Aug. 2012, doi: 10.1016/j.buildenv.2012.01.023.

[21] J. Liu, M. Heidarinejad, S. Gracik, and J. Srebric, “The impact of exterior surface convective heat transfer coefficients on the building energy consumption in urban neighborhoods with different plan area densities,” *Energy Build.*, vol. 86, pp. 449–463, Jan. 2015, doi: 10.1016/j.enbuild.2014.10.062.

[22] L. Merlier, L. Frayssinet, K. Johannes, and F. Kuznik, “On the impact of local microclimate on building performance simulation. Part I: Prediction of building external conditions,” *Build. Simul.*, vol. 12, no. 5, pp. 735–746, Oct. 2019, doi: 10.1007/s12273-019-0507-7.

[23] M. Mosteiro-Romero, D. Maiullari, M. Pijpers-van Esch, and A. Schlueter, “An Integrated Microclimate-Energy Demand Simulation Method for the Assessment of Urban Districts,” *Front. Built Environ.*, vol. 6, p. 165, Sep. 2020, doi: 10.3389/fbuil.2020.553946.

[24] Y. Yamaguchi, Y. Shimoda, and M. Mizuno, “Proposal of a modeling approach considering urban form for evaluation of city level energy management,” *Energy Build.*, vol. 39, no. 5, pp. 580–592, May 2007, doi: 10.1016/j.enbuild.2006.09.011.

[25] R. Choudhary, “Energy analysis of the non-domestic building stock of Greater London,” *Build. Environ.*, vol. 51, pp. 243–254, May 2012, doi: 10.1016/j.buildenv.2011.10.006.

[26] M. Deru *et al.*, “U.S. Department of Energy commercial reference building models of the national building stock,” *Publ.*, Feb. 2011.

[27] “Buildings - Dataset.” [Online]. Available: https://donnees.montreal.ca/ville-de-montreal/batiment-2d. [Accessed: 19-Dec-2020].

[28] “Permis de construction, transformation et démolition - Jeu de données.” [Online]. Available: https://donnees.montreal.ca/ville-de-montreal/permis-construction. [Accessed: 19-Dec-2020].

[29] “Building footprints 2009 — City of Vancouver Open Data Portal.” [Online]. Available: https://opendata.vancouver.ca/explore/dataset/building-footprints-2009/information/. [Accessed: 19-Dec-2020].

[30] “Property parcel polygons — City of Vancouver Open Data Portal.” [Online]. Available: https://opendata.vancouver.ca/explore/dataset/property-parcel-polygons/information/. [Accessed: 19-Dec-2020].

[31] “Issued building permits — City of Vancouver Open Data Portal.” [Online]. Available: https://opendata.vancouver.ca/explore/dataset/issued-building-permits/information/. [Accessed: 19-Dec-2020].

[32] “Property tax report — City of Vancouver Open Data Portal.” [Online]. Available: https://opendata.vancouver.ca/explore/dataset/property-tax-report/information/. [Accessed: 19-Dec-2020].

[33] “City of Toronto Building Outlines | Map and Data Library.” [Online]. Available: https://mdl.library.utoronto.ca/collections/geospatial-data/city-toronto-building-outlines. [Accessed: 19-Dec-2020].

[34] “Property Boundaries - City of Toronto Open Data Portal.” [Online]. Available: https://open.toronto.ca/dataset/property-boundaries/. [Accessed: 19-Dec-2020].

[35] “Building Permits - Pool Enclosures - City of Toronto Open Data Portal.” [Online]. Available: https://open.toronto.ca/dataset/building-permits-pool-enclosures/. [Accessed: 19-Dec-2020].

[36] J. A. Fonseca and A. Schlueter, “Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts,” *Appl. Energy*, vol. 142, pp. 247–265, Mar. 2015, doi: 10.1016/j.apenergy.2014.12.068.

[37] R. Nouvel, M. Zirak, V. Coors, and U. Eicker, “The influence of data quality on urban heating demand modeling using 3D city models,” *Comput. Environ. Urban Syst.*, vol. 64, pp. 68–80, 2017, doi: 10.1016/j.compenvurbsys.2016.12.005.

[38] “ESRI Shapefile Technical Description,” 1998.

[39] “Shapefile - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/Shapefile. [Accessed: 19-Dec-2020].

[40] "GeoJSON - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/GeoJSON. [Accessed: 19-Dec-2020].

[41] E. Standard, T. H. Kolbe, and C. Nagel, "Open Geospatial Consortium OGC City Geography Markup Language (CityGML) En- coding Standard," 2012.

[42] "Basic Information - CityGML Wiki." [Online]. Available: http://www.citygmlwiki.org/index.php?title=Basic_Information. [Accessed: 19-Dec-2020].

[43] U. Eicker, R. Nouvel, E. Duminil, and V. Coors, "Assessing passive and active solar energy resources in cities using 3D city models," in *Energy Procedia*, 2014, vol. 57, pp. 896–905, doi: 10.1016/j.egypro.2014.10.299.

[44] R. Kaden and T. H. Kolbe, "CITY-WIDE TOTAL ENERGY DEMAND ESTIMATION OF BUILDINGS USING SEMANTIC 3D CITY MODELS AND STATISTICAL DATA."

[45] "GitHub - tudelft3d/3dfier: The open-source tool for creating of 3D models." [Online]. Available: https://github.com/tudelft3d/3dfier. [Accessed: 19-Dec-2020].

[46] "CityGML Energy ADE - CityGML Wiki." [Online]. Available: http://www.citygmlwiki.org/index.php/CityGML_Energy_ADE. [Accessed: 19-Dec-2020].

[47] G. Agugiaro, J. Benner, P. Cipriano, and R. Nouvel, "The Energy Application Domain Extension for CityGML: enhancing interoperability for urban energy simulations," *Open Geospatial Data, Softw. Stand.*, vol. 3, no. 1, 2018, doi: 10.1186/s40965-018-0042-y.

[48] J. Benner, "CityGML Energy ADE V. 1.0 Specification," 2018.

[49] "EnergyPlus File Extensions: Getting Started — EnergyPlus 8.2." [Online]. Available: https://bigladdersoftware.com/epx/docs/8-2/getting-started/energyplus-file-extensions.html. [Accessed: 19-Dec-2020].

[50] "Commercial Prototype Building Models | Building Energy Codes Program." [Online]. Available: https://www.energycodes.gov/development/commercial/prototype_models. [Accessed: 19-Dec-2020].

[51] "Standard 90.1." [Online]. Available: https://www.ashrae.org/technical-resources/bookstore/standard-90-1. [Accessed: 19-Dec-2020].

[52] "IECC - ICC." [Online]. Available: https://www.iccsafe.org/products-and-services/i-codes/2018-i-codes/iecc/. [Accessed: 19-Dec-2020].

[53] Y. Ye, W. Zuo, and G. Wang, "A Comprehensive Review of Energy-Related Data for U.S. Commercial Buildings," *Energy Build.*, vol. 186, pp. 126–137, 2019, doi: 10.1016/j.enbuild.2019.01.020.

[54] "Building Component Library." [Online]. Available: https://bcl.nrel.gov/nrel/types/measure. [Accessed: 19-Dec-2020].

[55] "Building Component Library." [Online]. Available: https://bcl.nrel.gov/. [Accessed: 19-

Dec-2020].

[56] “Search results | Building Component Library.” [Online]. Available: https://bcl.nrel.gov/search/site/?f%5B0%5D=im_field_component_tags%3A127&f%5B1%5D=im_field_component_tags%3A130. [Accessed: 19-Dec-2020].

[57] A. D. Schlosser, G. Szabó, L. Bertalan, Z. Varga, P. Enyedi, and S. Szabó, “Building extraction using orthophotos and dense point cloud derived from visual band aerial imagery based on machine learning and segmentation,” *Remote Sens.*, vol. 12, no. 15, pp. 1–28, 2020, doi: 10.3390/RS12152397.

[58] “Measure feature height in imagery—ArcGIS Pro | Documentation.” [Online]. Available: https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/measuring-height-in-imagery.htm. [Accessed: 07-Jan-2021].

[59] “City of Kelowna.” [Online]. Available: https://www.kelowna.ca/. [Accessed: 04-Mar-2021].

[60] “2015 Survey of Household Energy Use (SHEU-2015) Data Tables | Natural Resources Canada.” [Online]. Available: http://oee.rncan-nrcan.gc.ca/corporate/statistics/neud/dpa/menus/sheu/2015/tables.cfm. [Accessed: 19-Dec-2020].

[61] N. Resources Canada, “Survey of Household Energy Use 2011 – Detailed Statistical Report.”

[62] “Dictionary, Census of Population, 2016 - Structural type of dwelling.” [Online]. Available: https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/dwelling-logements013-eng.cfm. [Accessed: 20-Dec-2020].

[63] BC Housing, “2018 Metrics Research Full Report Update STEP CODE,” 2018.

[64] “Advanced 3D City Design Software | ArcGIS CityEngine.” [Online]. Available: https://www.esri.com/en-us/arcgis/products/arcgis-cityengine/overview. [Accessed: 20-Dec-2020].

[65] “CityEngine - Wikipedia.” [Online]. Available: https://en.wikipedia.org/wiki/CityEngine. [Accessed: 20-Dec-2020].

[66] “CityEngine_cga/Building_From_Footprint.cga at master · jbdong/CityEngine_cga · GitHub.” [Online]. Available: https://github.com/jbdong/CityEngine_cga/blob/master/Building_From_Footprint.cga. [Accessed: 20-Dec-2020].

[67] “FME | Data Integration Platform | Safe Software.” [Online]. Available: https://www.safe.com/fme/. [Accessed: 20-Dec-2020].

[68] F. Biljecki, H. Ledoux, and J. Stoter, “An improved LOD specification for 3D building models,” *Comput. Environ. Urban Syst.*, vol. 59, pp. 25–37, Sep. 2016, doi: 10.1016/j.compenvurbsys.2016.04.005.

[69] V. Coors, M. Betz, and E. Duminil, “A Concept of Quality Management of 3D City

Models Supporting Application-Specific Requirements,” *PFG - J. Photogramm. Remote Sens. Geoinf. Sci.*, vol. 88, no. 1, pp. 3–14, Feb. 2020, doi: 10.1007/s41064-020-00094-0.

[70] “Modeling Guide for 3D Objects - Part 1: Basics (Rules for Validating GML Geometries in CityGML) - SIG3D Quality Wiki EN.” [Online]. Available: https://en.wiki.quality.sig3d.org/index.php?title=Modeling_Guide_for_3D_Objects_-_Part_1:_Basics_(Rules_for_Validating_GML_Geometries_in_CityGML)#gml:Solid. [Accessed: 20-Dec-2020].

[71] “Rushikesh Padsala / shp-Rhino-CityEngine2CityGML · GitLab.” [Online]. Available: https://transfer.hft-stuttgart.de/gitlab/rushikesh.padsala/rhino-cityengine2citygml. [Accessed: 23-Dec-2020].

[72] U. HosseiniHaghighi SeyedehRabeeh, Panchabikesan Karthik, Dabirian Sanam, Webster Jessica, Ouf Mohammed, Eicker Ursula, 2021, Discovering, Processing and Consolidating of Building Stock and Smart Thermostat Data in Support of Energy End-use Mapping and Housi, “No Title.”

[73] “DIN V 18599-2 - Energy efficiency of buildings - Calculation of the net, final and primary energy demand for heating, cooling, ventilation, domestic hot water and lighting - Part 2: Net energy demand for heating and cooling of building zones | Engineering360.” [Online]. Available: https://standards.globalspec.com/std/13053482/DIN V 18599-2. [Accessed: 31-Dec-2020].

[74] R. Nouvel *et al.*, “SIMSTADT, a New Workflow-driven Urban Energy Simulation Platform for CityGML City Models,” *CISBAT Int. Conf.*, no. February 2017, pp. 889–894, 2015.

[75] “SimStadt - SimStadt Documentation.” [Online]. Available: https://simstadt.ericduminil.com/index.html. [Accessed: 31-Dec-2020].

[76] “INSEL - Homepage - INSEL en.” [Online]. Available: https://www.insel.eu/en/. [Accessed: 31-Dec-2020].

[77] “ISO - ISO 13790:2008 - Energy performance of buildings — Calculation of energy use for space heating and cooling.” [Online]. Available: https://www.iso.org/standard/41974.html. [Accessed: 31-Dec-2020].

[78] Victoria, “2014 B.C. BEST PRACTICES METHODOLOGY FOR QUANTIFYING GREENHOUSE GAS EMISSIONS INCLUDING GUIDANCE FOR PUBLIC SECTOR ORGANIZATIONS, LOCAL GOVERNMENTS AND COMMUNITY EMISSIONS Ministry of Environment,” 2014.

[79] CaGBC, “Making the case for building to zero carbon: Executive summary,” 2019.

[80] “The ODB - an experimental Open Database of Buildings Metadata document: concepts, methodology and data quality.” [Online]. Available: https://www.statcan.gc.ca/eng/lode/databases/odb/metadata#a6. [Accessed: 06-Jan-2021].