

Regional building energy modelling: A residential building stock model for a Swedish island

Lukas Dahlström ^{a,*}, Fatemeh Johari ^b, Joakim Widén ^b

^a Division of Civil Engineering and Built Environment, Department of Civil and Industrial Engineering, Uppsala University Campus Gotland, 621 57, Visby, Sweden

^b Division of Civil Engineering and Built Environment, Department of Civil and Industrial Engineering, Uppsala University, 751 04, Uppsala, Sweden

ARTICLE INFO

Keywords:

Urban building energy modeling
Building energy simulation
Model validation
Open data
Building archetypes

ABSTRACT

This study presents the development and validation of a regional urban building energy model (UBEM) for the island of Gotland, Sweden, using openly available national datasets. The aim is to capture the diversity of the residential building stock - including both urban and rural areas - and provide a robust tool for large-scale energy planning and decarbonisation strategies. The model integrates building geometry, national construction data, and energy performance certificates (EPCs) with probabilistic approaches for infiltration and stochastic occupancy simulation. Building geometry is calibrated against EPC data to assure optimal agreement to real-world circumstances. A novel archetype methodology, based on clustering analysis, is employed to represent the heterogeneous building stock accurately with 15 archetypes for two residential building use types. Implemented with an EnergyPlus-based simulation core, the model achieves high computational efficiency. Validation of aggregated annual results against regional energy use statistics and EPC data demonstrates strong agreement on the aggregate level: for single-family buildings, the annual energy use difference is 3.3%, with a weighted mean difference in energy performance of 0.2%, while multi-family houses show a modest overestimation. These results confirm that combining open data with advanced probabilistic methods allows reliably simulating building energy dynamics at a regional scale. The framework is easily transferable and adaptable to new case studies, cities, or regions, making it a valuable resource for policymakers and urban planners aiming to enhance energy efficiency and reduce greenhouse gas emissions.

1. Introduction and background

The building sector is a cornerstone of modern energy systems, and buildings account for 40% of final energy use and 36% of energy-related greenhouse gas (GHG) emissions in the EU [1]. Up to 75% of buildings in the EU are also still energy-inefficient. Given the urgent need to mitigate climate change, the EU has set ambitious goals, including a 55% reduction in greenhouse gas emissions and a 32.5% improvement in energy efficiency by 2030 compared to 1990 levels [2]. Achieving these targets requires large-scale energy planning that integrates sustainable urban strategies with interconnected energy solutions.

Comprehensive energy modelling is essential for informed decision-making, allowing policymakers to simulate various scenarios, assess the impact of climate and urbanization trends, and design efficient, resilient energy systems. However, simulating the energy performance of thousands of buildings at high spatial and temporal resolutions presents significant data and computational challenges. To address challenges concerning data and computational resources, researchers often employ

simplified methods of abstracting the building stock into multiple categories, or *archetype buildings*. Archetype buildings are reference buildings that are representative of a group of buildings which share similar characteristics. For an aggregated simulation of the studied building stock, virtual buildings are created using cadastral data and simulated with the corresponding archetype characteristics.

This large-scale approach not only enhances model transferability and benchmarking but also provides critical insights for optimizing future urban energy systems [3–5]. By adopting holistic, data-driven energy planning, cities or regions can move toward sustainable, low-carbon development while meeting the demands of a growing population.

This study focuses on developing an urban building energy model (UBEM) for the region of Gotland, Sweden, a pilot area for smart and renewable energy systems. Using this case study the model can be validated against energy performance certificates and regional energy statistics, providing a scalable framework for urban and regional energy analysis using open data sources. By leveraging probabilistic modelling,

* Corresponding author.

E-mail address: lukas.dahlstrom@angstrom.uu.se (L. Dahlström).

<https://doi.org/10.1016/j.enbuild.2025.116178>

Received 31 March 2025; Received in revised form 27 June 2025; Accepted 20 July 2025

Available online 22 July 2025

0378-7788/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

stochastic occupancy simulations, and a novel archetype approach, the study enhances existing UBEM methodologies.

1.1. Case study area

The island of Gotland, Sweden, was selected as a case study. The island is a proposed pilot study area for smart and renewable energy systems in Sweden, and has a diverse building stock, which makes the area ideal for future studies regarding building stock-level carbon reduction strategies. The well-defined system boundaries of the area also makes it ideal for validating against statistical aggregated energy use data. The island has an area of approximately 3200 km² and has 61,173 inhabitants (as of 2022). Roughly half of them are living in the main city of the island, Visby, which makes the building stock of the island fairly distributed between urban and rural conditions. In order to distinguish between urban and rural conditions in terms of building locations in this study, urban areas are defined as areas with a population density of over 1000 persons per km². This corresponds well to buildings located within Visby or the two smaller towns of Gotland, Hemse and Slite. The island has a significant historical building stock, with a large proportion of the residential buildings being constructed before 1940, and several as far back as the 12th century.

Gotland and Visby have been used as case studies in multiple research studies, including decarbonisation scenarios [6], energy storage potential [7], and evaluation of energy renovation strategies for historic buildings [8–10], which all contribute to the understanding of the energy demand on the island.

1.2. UBEM

Urban building energy modelling (UBEM) is a relatively new approach aimed at creating bottom-up, engineering-based models of energy use in buildings across specific geographic areas, such as city districts or entire cities [11]. This modelling relies on the buildings' geometrical and physical properties, along with external inputs such as ambient temperature, solar irradiance, and internal heat gains [3,12]. While UBEM techniques are an upscaling of existing building energy modelling (BEM) methods, it introduces specific challenges. These include obtaining extensive urban-scale building data, dividing the building stock into representative building archetypes, and balancing model detail with computational efficiency. When practical and accurate, UBEMs can be powerful tools for urban planners, policymakers, and utilities to inform design decisions, evaluate policies, and forecast energy demand.

Over the last few years, a number of reviews of the UBEM field have investigated and mapped common approaches and workflows, and suggested new extensions and improvements for the research field [3,4,12–14]. While some best practices are identified, some major differences between approaches remain; mainly the types of data and modelling inputs used, the simulation approach (applied software or from-scratch mathematical energy transfer calculations), and the choice of utilising deterministic or probabilistic modelling approaches.

A multitude of UBEM studies have identified that data availability remains a key challenge for building stock modelling, especially high-quality building input data and measured energy use data for model validation [15,16]. Legal restrictions often limit the access to detailed data, especially in city-level studies [17], and many methods with access to specific data become limited and lack generalisation, making them challenging to apply to other contexts [13]. Additionally, handling large datasets presents technical challenges, including issues with data quality, which complicates the development of UBEMs [18]. To address the lack of sufficient datasets, model developers often rely on assumptions, simplifications, and more complex methodologies, thus accepting the risk of increased uncertainty and reduced accuracy in their models [19].

One promising solution to this issue is to utilise energy performance certificates (EPCs). EPCs are public domain documents issued for individual buildings with the goal of providing energy-related information, comparing buildings in terms of energy use, and to propose possible energy efficiency improvements [1]. The availability and usefulness of the EPCs in the EU, US, and the UK have made them increasingly valuable as a tool for decision-planners and real-estate analysts. In recent years, the use of EPCs to thoroughly model building stocks has also been steadily increasing [4,20]. Researchers in energy modelling has utilised EPCs to compare energy efficiency measures [21,22], optimise refurbishment processes [23,24], to perform energy planning, and to predict future energy demands [5].

1.3. Probabilistic modelling

Current literature suggests that physics-based and probabilistic computational methods out-perform statistics-based and deterministic ones, but only if the methods are properly calibrated [3,19].

Guo et al. [25] developed a fully probabilistic approach to map building stock U-values and window-to-wall ratios (WWR), using Flemish EPC datasets. Their method provides a robust framework for examining the modelling uncertainties and parameter distributions, but the method is not validated against measured data, and accuracy of the energy demand forecasts is not determined.

Cerezo et al. [26] analysed three methods for characterising UBEM archetype parameters, and calibrating against annual energy use data. They concluded that the probabilistic and calibrated methods were capable of producing a better goodness-of-fit regarding energy performance. The authors also highlighted that probabilistic modelling methods for archetype parameters needs to be further developed, given the intrinsic lack of knowledge about the diversity of the building stock.

Similarly, Sokol et al. [27] developed a model for 2263 buildings using a probabilistic approach for infiltration and occupancy parameters. The resulting model had a very high accuracy to measured energy performance but was calibrated to monthly, building-specific energy use data, which is generally unavailable in many areas.

Several reviews of UBEM methods and tools have identified the importance of building occupancy modelling for increased accuracy in UBEM [4]. Detailed approaches for stochastic and dynamic building occupancy modelling are necessary, if the goal of the UBEM is to perform peak load analyses, estimate hourly energy demand, or implement energy-efficient management strategies at the urban level [28,29]. Large city-wide UBEM have been around since the model of Boston by Cerezo Davila et al. [17]. Their streamlined model successfully simulated around 83,000 buildings with 52 archetypes. The model was validated using only sample comparisons to survey data, and the authors highlighted the future need for more accurate building archetype definitions and occupancy models, as well as openly available data for calibration and validation.

Numerous studies have since then been successful in creating accurate UBEM using deterministic methods and openly available data sets [30]. Johari et al. [31] created models for two Swedish cities, and got an absolute percentage error (APE) for the total energy use of 13% at city level, but the study did not compare the total simulated energy demand against aggregated measured energy use. This was later done by Johari et al. [32], where the simulated energy demand of 2193 buildings in a Swedish city was validated against hourly electricity use measurement data, achieving an hourly mean APE of 26%. Machine learning approaches have also been used to successfully develop (open) data-based UBEM of very large urban areas, as demonstrated by Roth et al. [33]. Recently, Song et al. [34] developed a UBEM of Shanghai, China, simulating 539,119 buildings with an aggregated electricity use that differed around 5% from measured use data (with no validation of gas use for heating).

1.4. Building infiltration

Building infiltration is the passage of outside air into the interior of a building, which has a major influence on the thermal load of a building and its internal environment [35]. Infiltration is a complex process which depends on building characteristics, outdoor conditions, and occupant behaviour. UBEM studies and reviews have suggested that probabilistic infiltration modelling methods are more accurate than deterministic ones, but are lacking in the field [16].

Outside UBEM, several studies have investigated infiltration in building stocks, and found average flow rates that correspond to specific building characteristics. As building characteristics and climate conditions are highly dependent on location, suitable literature was reviewed for findings that were of relevance for this study. Jokisalo et al. [36] showed that for Finnish detached houses, the average flow rate for sheltered (urban) conditions was 0.15 air volume changes per hour (ACH), and 0.25 ACH for exposed (rural) buildings. Pietrzyk and Hagentoft [37] found that the mean value for a log-normal distribution of infiltration rates was $\mu = 0.19$ ACH and the standard deviation was $\sigma = 0.09$ ACH for detached low-rise buildings in Gothenburg, Sweden. Wang et al. [38] found distribution values of $\mu = 0.49$ ACH and $\sigma = 0.36$ ACH for rural houses in the “cold” climate zone of China, including Beijing, which roughly corresponds to the climate zone of Gotland. Shi et al. [39] found that $\mu = 0.17$ ACH and $\sigma = 0.62$ ACH for urban residential buildings in Beijing.

Several studies also found patterns relating infiltration flow rate to UBEM input parameters, allowing for probabilistic modelling of building characteristics. Jokisalo et al. [36] also found that the building height is an important determining factor for the infiltration rate; the mean rate was found to be 0.10 ACH for one floor and 0.15 ACH for two floors. Eskola et al. [40] similarly found that the rate for two-storey buildings are approximately double that of one-storey buildings in Sweden. Persily et al. [41] found that the age of the building is a significant factor affecting the rate of infiltration. The rate for older American residential buildings was shown to be around 90–120 % higher compared to new construction. Similar values are also shown for the Dutch building stock [42]. Both studies also showed that the infiltration rate is around 25–50 % higher for single family houses than for multi-family housing.

Summarising these findings, a likely average flow rate for infiltration in the Gotland climate zone can be estimated.

1.5. Knowledge gaps

UBEM techniques have traditionally been applied to large-scale simulations of dense urban environments. To the authors’ knowledge, rural areas of the scale considered in this case study have not yet been included in such simulations. Probabilistic modelling approaches remain relatively rare, particularly in large-scale UBEMs, where deterministic parameter models or, more recently, machine learning methods are more commonly used.

While the use of national datasets in UBEMs has led to improvements in model validity, substantial uncertainty persists. This is largely due to the simplification of real-world buildings into digital and mathematical representations based on a limited set of parameters [19]. A study by Nouvel et al. [43] identified 13 key parameters in UBEM development that contribute to varying levels of uncertainty if not accounted for. “Must-have” parameters identified are building function, year of construction, state of refurbishment and residence type (main, secondary, or vacant). Ensuring the accuracy and reliability of a UBEM requires validation against actual measurement data. However, collecting such validation data - especially energy and power data at various spatial and temporal resolutions - is challenging. In many cases, the availability or accessibility of building-level data remains a barrier, making it difficult to verify UBEM results.

Kamel [16] noted open data availability, accurate building height data, stochastic modelling approaches, calibration against measured

data, interoperability features between UBEM tools, and computational complexity as significant challenges for future UBEM development.

1.6. Aim of the paper and scientific contributions

The main aim of this study is to develop an accurate UBEM for the area of Gotland, Sweden, that utilises open data and that accurately captures the diversity of the residential building stock, in regards to energy performance. In detail, this study aims to:

(1) propose a method for developing an UBEM that can be used on regional (urban and rural) level and extended to the national level, (2) propose a method for incorporating probabilistic parameters and stochastic occupancy modelling on individual building level, as well as a novel archetype approach, to increase model representativeness of real-world building stock conditions, by (3) implementing the model for a case study of the island of Gotland, Sweden, and (4) validate the simulation results against both regional energy use statistics and normalised energy performance data, to ensure ideal model accuracy.

The novel scientific contributions of this study are as follows:

1. The UBEM is regional, including rural and sub-urban areas as well as urban areas. This regional approach also allows for model validation against measured statistical energy use data, gathered for the case study region.
2. The model framework utilises a combination of open data sets and statistics to acquire building specific data, allowing for a high adaptability of the model for different geographical areas, after validation to the regional statistical data.
3. This study fills identified research gaps for UBEM simulation, as highlighted by Kamel [16] among others, and improves an already existing model. A new approach for identifying and assigning building archetypes is implemented, utilising cluster analysis and geographical patterns. Other improvements include a new method for identifying building HVAC systems, an approach for probabilistic modelling of building infiltration, and integration of stochastic occupancy modelling.
4. The case study region is interesting in that it has a high proportion of historical buildings, as well as modern construction, and the regional building stock has a significant proportion of holiday homes. Furthermore, as a regional study the geographical extent is large; the area for which the model was developed is significantly larger than in traditional UBEMs, around 3200 km², with around 19,000 buildings simulated.

The structure of this article is as follows. Section 2 gives an overview of the methodology used in this study, describing the model structure and simulation workflow. Section 3 later provides a summary of the input data used in this study, as well as explanations of additional assumptions that were considered necessary. This is followed by Section 4 which describes each step of the methodology in detail. This starts with data pre-processing and calibration, and the procedure for building archetype identification and characterisation. Subsequently, the approach for modelling building HVAC systems is described, followed by a description of the probabilistic infiltration and the integration of stochastic occupancy modelling in the simulation. The results are presented, analysed and validated in Section 5. Finally, in Section 6 the findings of this study is discussed; in particular the model accuracy, the simulation methodology, and the study limitations are evaluated.

2. Methodology overview

In this section, the UBEM approach used in this study is summarised in brief.

The simulation approach applied in this study is as follows: for every building in the case study data set, a virtual 3D building is created using cadastral data and assigned one of the identified representative

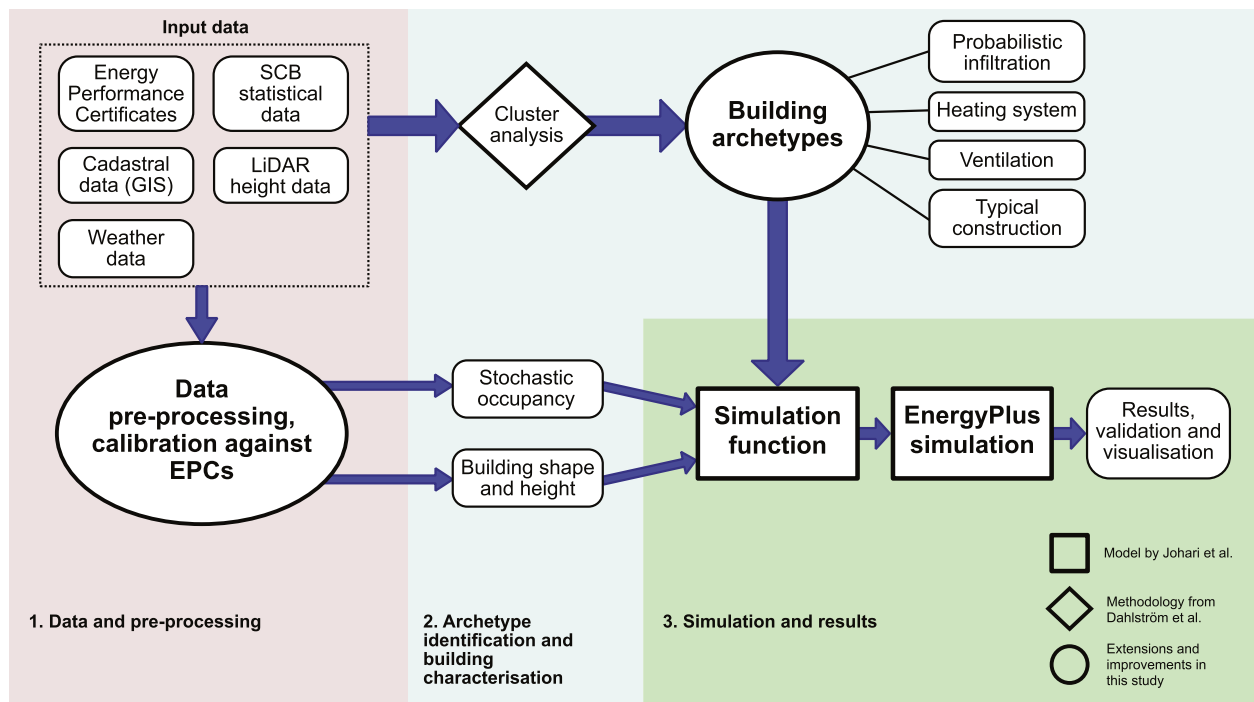


Fig. 1. Flowchart of the model and simulation framework. Arrows indicate data flows and straight lines indicate inclusion.

archetypes, which are all characterised by their own set of simulation parameters, such as construction material composition and properties. From available data, probabilistic infiltration flow rates, HVAC systems, and stochastic occupancy profiles are generated for every virtual building. Utilising all generated inputs, and additional weather data, the energy balance over the course of one year is consequently simulated for all buildings. These steps are explained in detail in Section 4, and the overall framework is visualised in Fig. 1.

The UBEM was developed in Python, as an extension and further development of the model developed and validated by Johari et al. [31] and Johari et al. [32] around the building simulation software EnergyPlus. EnergyPlus is well-regarded for its ability to provide accurate and efficient estimates of the energy performance of individual buildings. It has also been shown to outperform similar building simulation tools such as IDA ICE and TRNSYS in terms of computational time and accuracy for aggregated models [44], which is paramount for UBEM. Additionally, because it can be controlled through common languages like Python, it is also suitable for sequential energy modelling and simulations of multiple buildings. Python is a powerful and flexible programming language, and it is the basis for the package Eppy [45], which is a scripting language for EnergyPlus idf files and output files. This makes the overall framework suitable for performing city-wide energy simulations for large building datasets. In this study, a simulation interface was developed using Python 3.8.5 and Eppy 0.5.63. The automated interface processes both geometric and non-geometric data of buildings, generates building models, and runs dynamic energy demand simulations in EnergyPlus under real-world weather conditions. It then outputs results on an hourly, daily, monthly, and annual basis. The developed interface is based on a user-defined data structure built with Python classes and objects, making it flexible and adaptable for various use cases. It is not limited to any specific region, and can be applied for UBEM of any case study. The simulations were run on a desktop workstation with an Intel Xeon Silver 4216 CPU @ 2.1 GHz processor with 32 cores (64 threads) and 194 GB RAM.

The simulation core was extended to include new input data, additional features and a novel approach for identifying and characterising building archetypes. The simulation modules of the developed model

is shown in Fig. 1, which highlights the modules developed by Johari et al. [31,32], the methodology for archetype identification developed in Dahlström et al. [46], and the modules added by this study. This flowchart also describes the workflow for data processing and the simulation of one building, which is then repeated for all buildings in the data set. After the EnergyPlus simulation, results are acquired and stored. These can be used to further calibrate the model or simulate and analyse the entire dataset.

Each building is simulated as a low level of detail (LoD1) “shoebbox” model with one thermal zone per building, as seen in Fig. 2. For large scale applications, the significant decrease in computational time when using only one zone per building far outweighs the increased accuracy of higher detailed models, especially since the overall error in simulated energy demand is only around 9% [44]. Two thermal zones are used if the building has a heated basement, for which an additional thermal zone was assigned to the basement. As the presence of a heated basement is an EPC parameter, the probability of having a heated basement was added as an archetype parameter. Each building is then assigned a basement based on this probability, similar to the approach for the heating systems.

Window shading from blinds or curtains was added for the simulation. These are activated if the indoor temperature gets higher than the

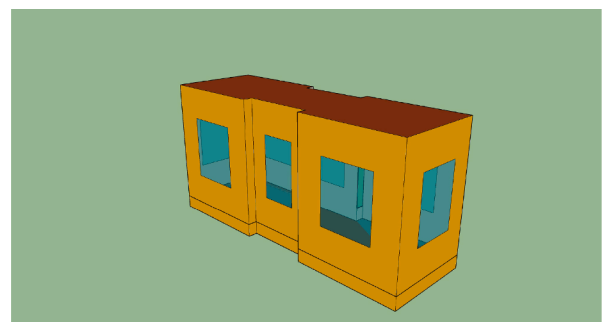


Fig. 2. 3D rendering of the “shoebbox” thermal energy model design [44].

thermostat cooling set-point (26 °C) and if the incoming solar irradiance exceeds 250 W/m². The simulation core can employ a shading calculation, where incoming sunlight on the simulated buildings is blocked and reflected by surrounding shading elements from all buildings in the cadastral data set, including accessory buildings. The shading calculation was applied for a test run of 50 buildings, and it was found that using shading only resulted in a 1 % difference in total energy demand, but a 57 % increase in computational time. This minor difference can be explained by the low impact of solar irradiance on building heating demand in the cold winter climate of Sweden, and by the exclusion of solar panels and cooling systems in the simulations, both of which would have been more strongly dependent on the solar irradiance. This finding is supported by Faure et al. [47], who found that around 2 % of the difference at district scale could be attributed to the surrounding shading environment. Therefore, the shading calculation was ultimately not applied in this study since it significantly increases computational time with only a negligible increase in accuracy.

3. Data

In this section, a summary of the data used in the study is provided.

UBEMs are dependent on detailed building-level data, both for model development and validation. One main objective of this study is to develop an UBEM from open-access data. Therefore, all of the used data has been extracted from databases that are nationally available online or upon request. The model is run using an input file for all residential buildings, divided into single family housing (SFH) and multi-family housing (MFH). This input file contains the geometry (footprint and height), construction and re-construction year, building type, property address, and socio-economic parameters for each building. The file also contains calculated living area, based on building size and estimated number of floors. For consistency, all data are for the year 2022. Including measured local weather data, this framework has access to all 13 of the relevant parameters identified by Nouvel et al. [43], although year of refurbishment is not utilised.

3.1. Cadastral data

Cadastral data, or geographic information system (GIS) data, for the studied area was acquired from the Swedish Mapping, Cadastral and Land Registration Authority [48]. This data contains a property map for the built-up area including information on building footprints, building types, and the year of construction or re-construction (if any) of the buildings. For UBEM calculations and simulation, the construction of a 3D model is necessary. Cadastral data only contain 2D footprints of buildings, and the Swedish EPCs do not provide any data on measured building height, so the height must be determined by other means. A commonly used approach is to determine the building height and elevation information from national geo-datasets, photogrammetry, or laser scanning, i.e., light detection and ranging (LiDAR) point cloud data. LiDAR data is openly available in many countries, including the EU, Norway and the UK [49].

3.2. Holiday homes

Another aspect that also makes the case study area interesting is that the island is a major vacation and tourism hub, which has the effect that a significant fraction of the building stock are holiday homes or secondary homes. Holiday homes are not permanent residencies, and are usually occupied between a few select days over the year to around three months every summer. Holiday homes were not in the scope of this particular study so they were not simulated as a part of the aggregated results, but they needed to be removed from the GIS data set so that the number of permanent dwellings was correct in the simulation.

In the available data sets, it is unfortunately not possible to ascertain which exact houses in the building stock are holiday homes. The total

number of holiday homes, however, is known to be 12,705 homes [50], which corresponds to around 40 % of all the buildings on the island, or around 42 % of all SFH. For the model simulations, 12,705 random SFH were assumed to be holiday homes, using a set randomisation seed as to assure comparable simulation runs.

3.3. Energy performance certificates

Energy Performance Certificates (EPCs) are a valuable source of data for building simulations. In Sweden, they are collected by the National Board of Housing, Building and Planning, “Boverket” [51], and the current data set is updated by the year 2022. In contrast to other European countries and the US, nearly all energy use values available in Swedish EPCs are measured, not calculated [52,53], which makes them highly suitable for model validation and calibration. Among other parameters, the EPCs provide information about building construction year, main use, heated indoor area, number of floors, and the HVAC systems installed [54]. The building energy performance (EP) of individual buildings is also an EPC parameter, which is a measure of the specific energy use (kWh/m²) that is required for space heating and hot water use, by heated indoor area, normalised to a normal weather year. This parameter excludes household and property electricity¹ and is valuable for comparing buildings as it generally gives a good understanding of the state of the building envelope and HVAC system.

3.4. Statistical data and validation

Statistical data, including socio-economic parameters, were used alongside EPC data for archetype identification. This data includes mean income per person, mean resident age, and population density, and is collected and managed by Statistics Sweden [55]. The socio-economic data is available in an aggregated format, to avoid privacy concerns and to uphold the EU regulations regarding open data distribution. The data is distributed in squares, 25 × 25 m to 100 × 100 m in area. Spatial interpolation of the aggregated values was used to assign probable values for each individual building. For a detailed description of how these parameters were integrated as building values, please refer to the methodology section of the paper by Dahlström et al. [46].

For this study, measured energy use data of high temporal resolution was unavailable as open-access data. Therefore, calibration and validation of the model has to be made against aggregated yearly energy use data, which are available. For this purpose, statistical energy use data for residential buildings in the region of Gotland were used, accessed from Statistics Sweden [55]. This data is divided by end use category and primary energy source,² which is convenient as it makes validation of the division of heating systems in the simulation possible.

While a majority of UBEM studies are validated against annual data, a low temporal resolution significantly increases the uncertainty of evaluated model accuracy [19]. However, as Swedish EPCs utilise measured energy use, they can be used as an additional data source for model calibration and validation, which increases the reliability of the validation. The resulting energy performance (EP) from simulated buildings can be compared to the corresponding value from the buildings’ EPCs. This can be done on individual building level for the buildings that have EPCs, or at an aggregated level if comparing the distribution of EP values.

3.5. Weather data

To enable calibration and validation of the model from both statistical measured data and EP values from the EPCs, two weather data files were used.

¹ Energy used for elevators, stairwell lighting etc. in apartment buildings.

² Fossil fuels are mainly heating oil, bio-energy includes firewood and pellet burners, and electricity includes both direct electric heating and heat pump systems.

Most UBEM studies utilise a centralised weather data file, and/or a typical meteorological year (TMY) which is normalised to historical data [13,16]. However, when calibrating or validating against actual measurement data, it is necessary to use measured weather data from the same simulation time period to ensure accuracy. EnergyPlus weather input data (EPW format) can be very detailed, containing up to 35 weather parameters which can be used for the simulations, but such detailed weather data for the case study area for the year 2022 was not found. Instead, a weather data file for 2022 was manually compiled from open-access weather data from the Swedish Meteorological and Hydrological Institute (SMHI), containing dry- and wet bulb temperature, atmospheric pressure, total global solar irradiance, rainfall, relative humidity, wind direction and wind speed.

As the EPC EP values are normalised to TMYs, normalised weather data must be used for fair comparison. A normalised³ historical weather data file for the case study area in EPW format was obtained from Climate.OneBuilding [56]. The data was also recorded and managed by SMHI [57].

3.6. Additional input data

EnergyPlus takes into account the thermal inertia of interior walls, doors and furniture. The surface area for the internal mass is estimated from the volume of the building, and the construction materials are assumed to be identical to the interior floor and ceiling, not including insulation. For this calculation, a ratio between internal wall surface area and the building indoor volume, $A_{int} = kV_{ind}$, is used, where $k = 0.40 \text{ m}^{-1}$ [58].

The minimum requirement for window area according to Swedish construction regulation is 10% of the floor area per room [59]. This window area is divided by the total wall area of the building to get the minimum WWR for the building. This value of WWR is then used for each wall unless the wall is adjacent to another building or if the length of the wall is less than 2 m. In this way, the WWR will follow national guidelines but still have different values for different buildings, reflecting real-world conditions. It is also assumed that much older buildings (<1910) have around 10% lower WWR.

The thermostat set-point was set to 20 °C for all buildings, which is the Swedish guideline for all residential housing [59]. No building was assumed to have any central cooling system, as this is very uncommon for Swedish residential buildings.

The ventilation rate for exhaust ventilation is set to follow the Boverket Swedish national regulations, where the minimum flow rate is set at 0.35 l/s/m².

4. Methods

This section describes the methods used in the study, with each step explained in detail. Section 4.1 explains the data pre-processing that constitutes part 1 of the methodology flowchart in Fig. 1, and Section 4.2–4.4 correspond to part 2: Archetype identification and building characterisation in the flowchart.

4.1. Data pre-processing

With a large geographical area such as Gotland, the data size of the LiDAR point cloud is very large. To comfortably deal with the point cloud data, the data is first filtered to only include points in and immediately around the building polygons. Once filtered, the building height could be determined from the point cloud coordinates within each building polygon, and can then be calibrated against EPC data to ascertain representative building heights.

³ Normalised for the years between 2009–2023. Monthly historical data are used, where each month is the normal month for the period.

A common inaccuracy that occurs when using LiDAR data for buildings is that large vegetation such as trees often grow over the buildings, obscuring the roof. These data points can not be filtered with a simple algorithm; detailed filtering approaches are available which reduces potential inaccuracies, however, they are computationally intensive. For this reason, and considering the large data set, a simplified filtering approach was developed. Height outliers that were implausibly high ($> 4\sigma$) or too low (less than 2 m in height) were replaced by a probable height value, which was determined by using the average height of the closest neighbouring buildings. To avoid noise, the point cloud median z -coordinate value was used instead of the maximum z value. For these cases, the median height value often well represented the real building height, as the roof contain extensively more data points than the obscuring vegetation data points do.

The building height and (heated) indoor area was calibrated using the EPCs, as they contain information about indoor area for all EPC buildings and number of floors for most buildings. SFH and MFH were separately calibrated by comparing calculated height for each EPC building with estimated area and height for corresponding buildings in the GIS and LiDAR data. For the EPC building height, the height was calculated as the standard indoor height-per-floor multiplied with the number of floors of the building, with 1 m added if the building has a (heated) basement. The standard indoor height-per-floor was defined at 2.65 m as 2.4 m is the required floor-to-ceiling height for Swedish buildings (2.3 m for basements) [59] and the average thickness of inner ceilings is around 0.25 m [60,61].

To calibrate against the measured building height, a correction factor c_h for the LiDAR values is determined by taking the mean value of the calculated (EPC) height divided by the LiDAR height; this factor was found to be $c_h = 0.80$ for SFH and $c_h = 0.87$ for MFH, meaning that on average the LiDAR values overestimate the actual building height.

The building indoor heated area also needs to be calibrated to EPC values. The calculated indoor area for the GIS buildings is the building footprint A_{fp} multiplied by the number of floors n_f , which is determined by the calibrated height divided by 2.65 m, rounded off to the nearest integer value.

To determine a correction factor for the building footprint, we calculate the factor c_A corresponding to the minimum difference between measured and calculated values, by performing a minimum root mean square error (RMSE) test:

$$RMSE_j = \sqrt{\frac{1}{n} \sum_{i=1}^n ((c_A A_{fp_i} n_{f_j}) - A_{EPC_i})^2}, \quad (1)$$

where the factors were found to be $c_A = 0.53$ for SFH and $c_A = 0.73$ for MFH at minimum RMSE. The low values are expected as the heated share of a building can vary between 40–100% [51], and the building footprint should be larger than the indoor area as the indoor area does not include garages, porches, and similar. The resulting RMSE curves for both SFH and MFH can be seen in Fig. 3.

4.2. Archetype identification and characterisation

When analysing the EP values in the Gotland EPCs, it becomes clear that a deterministic approach for dividing buildings by use type and construction year is not sufficient to capture the heterogeneity of EP in the building stock. For houses built in the same time period the values are significantly spread out; for SFH on Gotland built in the 1960s, the EP varies between 29 kWh/m² to 433 kWh/m², with a mean of 128 kWh/m² and a standard deviation of 70 kWh/m².

For this study, archetype buildings were identified using the methodology developed by Dahlström et al. [46]. The method involves utilising automated k -means cluster analysis of multiple building parameters to identify energy use patterns in the studied building stock, to ultimately narrow down the deviation in energy performance as much as possible.

The inputs to the cluster analysis were building use type, construction year, and heating system, which are EPC parameters; population

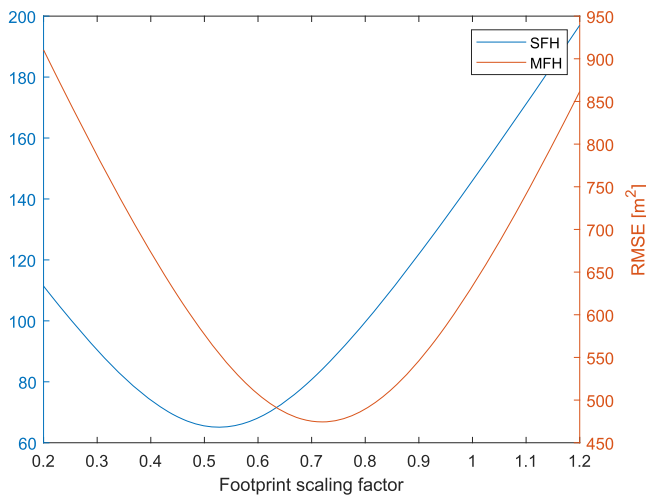


Fig. 3. RMSE for building indoor area.

density, resident age, and income per person, which were gathered from the applied socio-economic data; and finally building volume, which was calculated using GIS data and the LiDAR height. The outputs of the analysis are identified archetypes, which corresponds to the buildings closest to the centroid values of all parameters, as well as the clusters of parameters that belongs to each identified archetype. The number of clusters k corresponds to the number of archetypes, and the optimal value of k has to be determined separately when using k -means clustering. The root mean square standard deviation between all cluster parameters was compared for a range of values of k , and the optimal value was determined using the elbow method in combination with comparing the representativeness of the identified archetype to the cluster mean values of all parameters. For a detailed explanation of the procedure, as well as the applied data, the reader is referred to Dahlström et al. [46].

While nearly all MFH in the case study area have an EPC, only 2893 SFH have one, which is around 9.7% of all SFH on the island, or 16.8% excluding holiday homes. The remaining buildings must therefore be assigned the archetype that is most representative for each building. To find the archetypes of all buildings in the case study area that do not have an EPC, a novel method was developed.

Archetypes identified with the applied method are notably clustered also geographically. This is assumed to be representative of real-world conditions as buildings in certain neighbourhoods or blocks tend to be more homogeneous in terms of building characteristics, when comparing to a group of buildings that belong to different geographical areas. That is, they are generally built within the same period, have similar HVAC systems and the occupants are more socio-economically similar. Therefore, an algorithm was designed to find the building archetypes predominantly based on the location of the building. The closest neighbouring buildings are determined by a nearest-neighbour-search, and the most prevalent ($\text{mode}(x)$) archetype among them is identified. If the answer is inconclusive ($\text{size}(\text{mode}(x)) > 1$), more than one archetype is equally likely and the archetype is instead determined based on the construction year and the interpolated population density at the location of the building. A nearest-neighbour-search (knnsearch) computes the Euclidean distance $d(p, q) = \|p - q\|$ between the data point p and the query point q , then finds the k nearest data points by using a pre-defined Kd -tree algorithm.

Different algorithms were tested for the highest percentage of correct identifications, for a sample of the buildings available in the EPCs. The tested alternatives included identifying by only the closest construction year, finding the closest building considering all cluster parameters, or by only picking the geographically closest archetype. The accuracy of this method was then tested for different values of k nearest neighbours, and the optimal value turned out to be $k = 3$.

The final algorithm used in this study to identify the archetype buildings is described in Algorithm 1. The algorithm identifies the most probable archetype for one GIS building object B , for the cases where B does not have an EPC. The parameter cy denotes the building construction year and pd is the location population density.

```

P = [EPCs.coord.x, EPCs.coord.y]
PQ = [centroid(B.coord).x, centroid(B.coord).y]
K = knnsearch(P, PQ, k = 3)
NN = EPCs(K) Nearest neighbour buildings
if size(mode(NN.archetype)) = 1 then
    | archetype = mode(NN.archetype)
else
    if B.cy is not empty then
        | archetype = knnsearch(NN.cy, B.cy, k = 1)
    else
        | archetype = knnsearch(NN.pd, B.pd, k = 1)
    end
end
end

```

Algorithm 1: Algorithm to find the archetype of a building.

The identified archetypes are used to determine the construction of the buildings. The use type and construction year of the corresponding archetype building is used to determine the construction materials for each simulated building, similar to how deterministic archetypes are commonly used in UBEM studies.

For the construction materials of the archetypes, a more straightforward approach was adopted based on literature that comprehensively describe the typical construction of residential buildings in Sweden. These resources include the two reference books by Björk et al. [60] and Björk et al. [61], which highlight that the construction characteristics of Sweden's building stock are closely linked to different construction periods, and describes typical Swedish residential buildings from 1880 to 2020 in great detail. With this method, each use type and decade is assigned a construction "group" (with buildings constructed before 1910 as the first construction group). This approach makes the simulation framework adaptable to new building archetypes, for simulations of any Swedish dataset, as the identified archetypes is an input to the algorithm that assigns a construction group to the simulated building.

4.3. Heating and ventilation systems

Other simulation factors determined by the archetype building are the HVAC system, infiltration, and occupancy.

The type of heating system is largely influential on the energy use of the building, due to the inherent differences in system efficiency [31,46]. The heating systems are categorised into four energy sources: district heating, boilers (for bio and fossil fuels), electric heating, and heat pumps. By analysing the EPCs, it is clear that a considerable amount of buildings are using a great variety of energy sources for heating, often simultaneously. Each archetype is also characterised by a share of heating system types, for an accurate representation of the energy use from different sources in the simulated buildings. These shares are determined from the archetype building clusters and are shown in Table A.5. As discussed in Dahlström et al. [46], the clustering algorithm is able to divide the building stock into archetypes that are very distinct regarding the heating system type.

Knowledge about the ventilation system is important for accurate estimations of the building energy demand. A heat recovery ventilation (HRV) system can recover about 60–95% of the heat in the exhaust air, which significantly improves the building energy efficiency [62]. The HRV rate was set to 75% in the model. The probability that each building has natural or exhaust ventilation, or if it has a HRV system, is determined by the shares of each system per archetype. Although natural ventilation or HRV were not clustering parameters when identifying

Table 1
Input values for each occupancy state.

Occupancy state	Appliances [W]	Lighting [W]	Hot water [l/h/person]	Body heat [W]
Away	70	0	0	0
At home sleeping	70	0	0	80
At home awake	200	60	7.1	130

the archetypes, the distribution of ventilation systems for the archetypes are distinguished. Some archetypes has virtually no building with HRV, whereas some has a major share of HRV in the buildings (see Table A.5).

The infiltration rates for the simulated buildings was calculated with the following method. For a probabilistic approach, the infiltration rate can be determined by random sampling from a log-normal distribution, with a known mean and standard deviation (see Section 1.4). The probability density function for the log-normal distribution is:

$$p(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \quad (2)$$

where μ is the mean and σ is the standard deviation of the normally distributed logarithm of the variable.

Numbers for the infiltration rate distributions of all buildings in the model was based on the available literature. The base rate was set on $\mu = 0.2$ ACH and $\sigma = 0.3$, where μ is increased or decreased accordingly to account for building type, age and height, and rural/urban conditions (see Section 1.4).

4.4. Occupancy

A probabilistic approach for simulating building occupancy was implemented, based on the approach for upscaling of Markov chain-based occupancy models in Widén [63]. The so-called rescaled Markov chain (RMC) method was used to simulate hourly 3-state aggregated occupancy patterns for each building. The states were “Away”, “At home sleeping”, and “At home awake”, corresponding to the activity states 1, 2 and 3–9, respectively, in Widén [63]. The model parameters were estimated for weekdays and weekend days separately based on time-use data (TUD), obtained from Statistics Sweden and described in more detail in Widén [63].

The energy use and internal gains (metabolic heat gain and usable heat gain from electricity use) for each state (home awake, home asleep, and away) were assumed and then calibrated, so that the total occupant-related energy use that results from the utilised occupancy modelling algorithm produced results that were close to the 2024 guidelines of typical values for Sweden; around 21 kWh/m² for both hot water and household electricity [64]. The metabolic heat gain values per occupant and state were assumed from the EnergyPlus reference values [65]. This number represents the total heat gain per person, and an internal algorithm in the software is used to determine what fraction of the total is convective, radiant, and latent. The values used in the model is described in Table 1.

The number of occupants per building is determined from the available indoor living area. According to Statistics Sweden [66], the mean living area per person in Gotland is 47.5 m² for SFH and 39.6 m² for MFH. The number of occupants per building is the estimated living area divided by this factor, rounded off to the closest integer, minimum one occupant. The living area for each building is calculated from the building footprint and height, after the calibration against EPC data. For MFH, common areas, such as hallways, stairwells, and maintenance rooms, are not included in the living area. This “loss factor”, the ratio between available and rentable space, is assumed to be around 80–90 %, so a factor of 0.85 is used for the MFH net living area.

Assuming a constant illuminance of 150 lux in every room when any occupant is home, and using the numbers for mean living area per person we get around 5940–7125 lumen. With the luminous efficacy of modern LED lights, 6000 lumen corresponds to around 50–70 W [67]. In

Table 2
Identified building archetypes for SFH and MFH.

Archetype	Location	Const. year	Heated area [m ²]	EP [kWh/m ²]
SFH				
1	Rural	1948	166	146
2	Urban	1956	149	75
3	Rural	1970	125	117
4	Urban	1973	105	154
5	Rural	1800	108	175
6	Rural	1909	137	94
7	Rural	1982	165	72
8	Urban	1945	129	151
MFH				
1	Urban	1981	1049	137
2	Rural	1982	179	121
3	Rural	1980	367	101
4	Urban	1755	202	161
5	Urban	1905	486	164
6	Urban	1971	883	129
7	Urban	1964	216	71

Sweden, the amount of sunlight hours vary significantly over the course of the year, with dark winters and long summers days, creating drastic seasonal variations for lighting use in buildings. To adjust for Swedish conditions, with 60 W as a baseline, the lighting schedule is gradually changing with the year, from 30 W in mid summer to 90 W in mid winter. In real life scenarios, the household electricity use would also vary slightly with the month of the year as different appliances are used in different weather, among other variations. Furthermore, although there is a clear consensus that household income and resident age has an effect on the household electricity use [4], there seems to be insufficient research to define the correlation numerically.

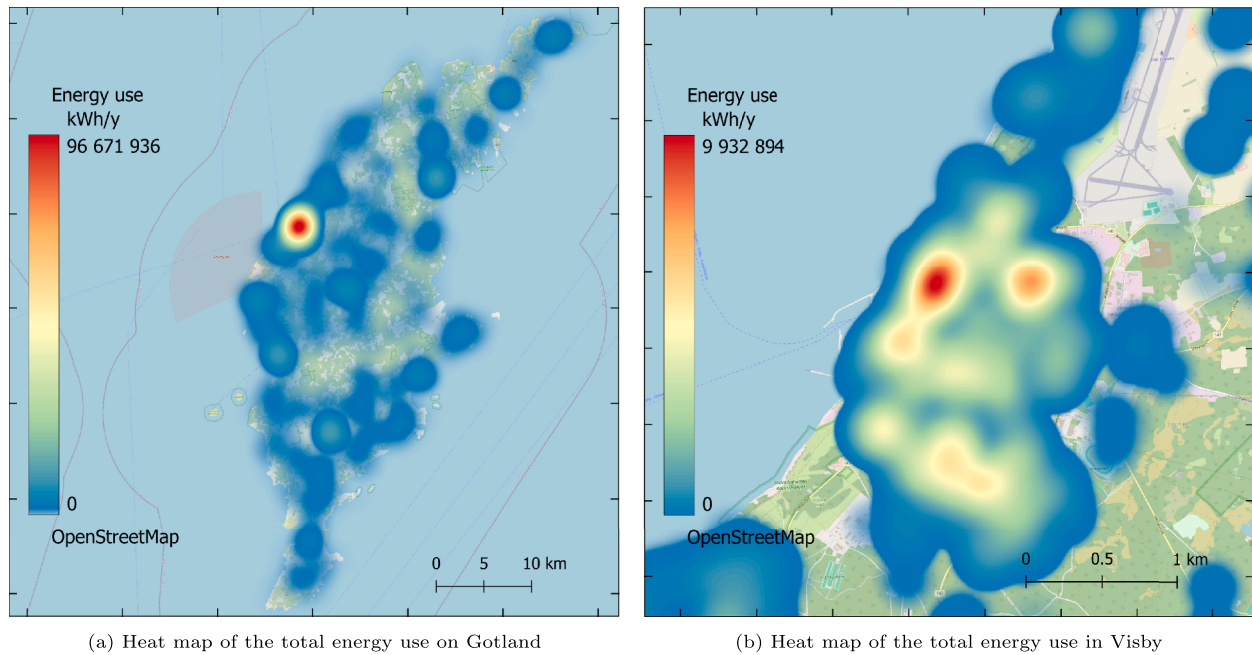
5. Results and analysis

5.1. Identified archetype buildings

The clustering algorithm identified 8 archetypes for Gotland SFH, and 7 archetypes for MFH, which can be seen in Table 2 with a limited set of parameters. The main dividing factors that define each archetypes are HVAC systems, construction year, population density (urban or rural), building volume, and ultimately energy performance. The full set of parameters for each archetype can be seen in Table A.5. The number of buildings in the case study data set that were identified to belong to each archetype cluster can be seen in Table 4.

5.2. Simulation results

The residential building stock of Gotland was simulated twice, once with each weather data file. The results from the simulation using the 2022 weather data is compared to the statistical energy use data from 2022, and the results from the simulation using normalised weather data is compared to the EP values from the EPCs, since they are normal year corrected. The simulation results were divided by the two main building use types - SFH and MFH - to better visualise the characteristics of the results. The computation time for constructing the virtual building models and to run the simulations was approximately 4.5 s per building, or around 24 h for the full data set.



(a) Heat map of the total energy use on Gotland

(b) Heat map of the total energy use in Visby

Fig. 4. Heat maps of the simulated energy use.

Table 3

Simulation results for energy use and occupancy, by heating system, end use and building use type.

Energy use [GWh]	RBEM			Statistics		
	SFH	MFH	Total	SFH	MFH	Total
Total energy use	321.9	168.9	490.8	302.1	133.2	435.3
District heating	18.4	113.2	131.6	11.5	107.5	119.0
Space heating	15.6	96.0	111.6			
Hot water	2.8	17.2	20.0			
Fuels	104.8	6.0	110.9	122.4	0	122.4
Space heating	89.2	5.1	94.3			
Hot water	15.6	0.9	16.6			
Electricity	198.7	49.7	248.4	168.2	25.7	193.9
Electric heating	94.5	19.0	113.5			
Heat pumps	35.1	2.2	37.3			
Hot water	22.7	3.8	26.5			
Household electricity	46.4	24.7	71.1			
Number of occupants	41,499	22,236	63,735	41,797	19,376	61,173
Indoor heated area [1000m ²]	1932	1035	2967	2013	938	2951

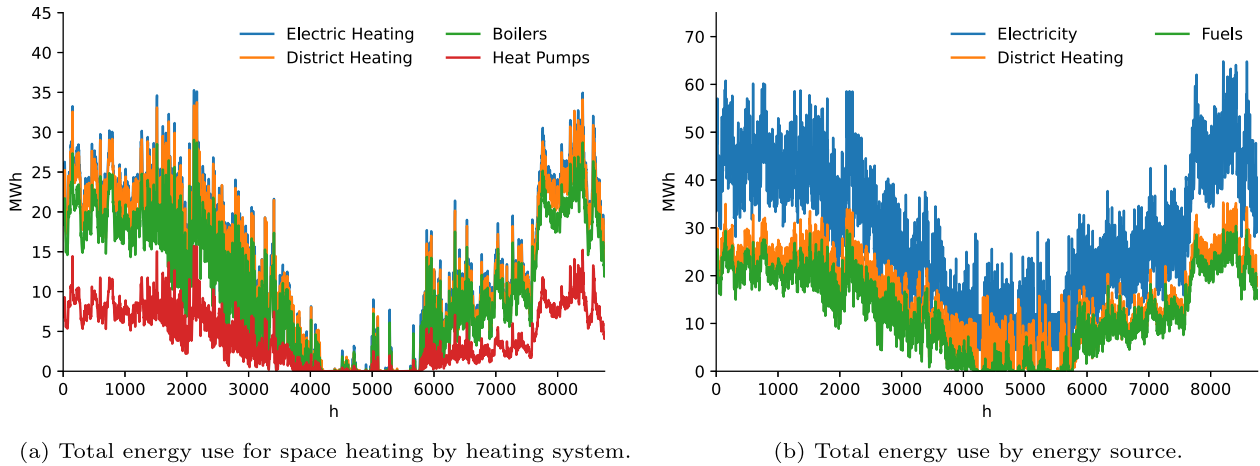
The simulation results for the 2022 weather data, by building use type, heating system and end use, is presented in Table 3. Here it is assumed that each building utilises the same energy source for hot water heating as for space heating, and the household electricity is included in the total electricity use. This enables comparison with the statistical energy use data for 2022, which is also shown in Table 3. The statistical value for MFH indoor heated area in the table comparison does not include the basement area if there are no apartments in the basement, even if it is heated. Therefore the calculated value also exclude the basement area, for a fair comparison. The simulated total energy use is presented as a heat map of the case study area in Fig. 4(a), with a zoomed in view of the city of Visby in Fig. 4(b). The heat maps were made using a kernel density function with a radius of 5 km, weighted for total energy use. Fig. 5(a) shows an hourly resolution of the simulation results for the year 2022, visualised by building heating system, and Fig. 5(b) shows an hourly resolution of the simulated total energy use by energy source.

Comparing the calculated values for living area and amount of occupants with the statistical values we can see that the values match well for SFH, with a 7% absolute error in total energy use, a 4% error for heated area, and less than 1% error for occupancy. The values for MFH are however slightly overestimated; the total energy use is overestimated by around 27%, while the calculated heated area is 10% higher and the number of occupants is 15% higher. In general, the division of energy use per energy source is similar to the statistical values, but with a significantly higher share of electricity relative to the other energy sources. Electricity use is more related to the number of occupants than other energy use, which could explain part of the overestimation. Since number of occupants was assumed from the indoor heated area, it can be viewed as a calibration parameter. If the total energy use values are adjusted to compensate for the overestimated indoor area and number of occupants the total energy use is decreased to a value closer to the statistical value, bringing the absolute error down to 10% for MFH.

5.3. Comparisons of energy performance

One of the main goals of this study was to accurately capture the heterogeneity of the residential building stock, in regards to energy performance of the buildings. A straight-forward way of evaluating if the model has been able to capture the EP diversity, is to compare these graphs with the corresponding graphs for the EPCs. As the energy performance values in the EPCs are normal year corrected, the simulation was run for a normalised weather data set to be able to do a fair comparison. The resulting energy performance of the buildings in this simulation is shown in Fig. 6(a) and (b) along with the EPC values for MFH and SFH respectively. As the slope of this curve is significant, it is clear that the EP values of the EPC data set are well spread across a wide spectrum.

Here we can see that the EP distribution follows the same shape of the curve for both use types, indicating that the model indeed captures the inherent diversity of EP in the building stock. The overestimation of energy use in the simulation can be seen also in these graphs, in and above the middle part of the distribution (corresponding to the below-median-performing buildings) for both building types.



(a) Total energy use for space heating by heating system.

(b) Total energy use by energy source.

Fig. 5. Simulated hourly energy use for Gotland.**Table 4**

Comparison between the energy performance (EP) of simulated buildings and the EP of the identified archetypes.

Archetype number	Number of buildings in simulation	Simulation mean EP [kWh/m ²]	EPC arch. EP [kWh/m ²]	Difference [%]
SFH				
1	3351	160.6	146	9.1
2	1043	57.8	75	-29.8
3	4373	165.5	117	29.3
4	1016	151.1	154	-1.9
5	358	171.6	175	-2
6	2860	79.4	94	-18.4
7	3812	61.1	72	-17.8
8	368	146.8	151	-2.9
Weighted mean difference				0.2
MFH				
1	592	139.4	137	1.7
2	161	124.6	121	2.9
3	89	127.5	101	20.8
4	76	169.7	161	5.1
5	158	168.2	164	2.5
6	372	139.6	129	7.6
7	120	70	71	-1.4
Weighted mean difference				4.3

The mean EP values of these clusters can also be compared with the EP of the identified archetypes, as shown in Table 4, and all cluster EP values are shown as box plots in Fig. 7(a) and (b). To evaluate the overall accuracy for representing the EP of the building stock as a whole, the weighted arithmetic mean difference is used. The weighted mean is defined as

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad (3)$$

where w_i is the size of each cluster i and x is the difference between the simulated results and the EPC value. If the simulated building clusters in total represents the archetypes well, \bar{x} should be zero, and if \bar{x} is positive the simulation generally overestimates the EP of all archetypes.

In this comparison we can see that while the EP of some archetypes differ by around 30%, some have a low error and the resulting weighted mean difference is very low, at 0.2% for SFH and 4.3% for MFH. This can be compared to the EP weighted means of 4.9% and 6.0% for SFH and -25.0% and -13.5% for MFH in the two case study areas simulated in Johari et al. [31], in which the archetypes were developed using a deterministic approach. These comparisons can be seen as guidelines for identifying which archetypes that should be calibrated. For this study,

the archetypes were not calibrated, as one of the purposes of the study was to examine how the national open-access data will perform when applied in the UBEM.

6. Discussion

6.1. Model accuracy

The model simulations demonstrated a high accuracy when comparing the simulated energy use of 2022 to the statistical data, divided by heating system. There is also a high agreement when comparing the EP of the normal year simulation results to the EPC values, as seen in the cumulative distribution curves of Fig. 6. Similarly, there is a high agreement for both values and spread of the EP of the archetypes, compared with the EPC archetype clusters, as seen in Fig. 7. This is especially interesting since the archetypes were not calibrated in this simulation. It is of importance to analyse the accuracy of the model prior to calibration as it accentuates the accuracy of the framework itself, and the usefulness of open data in particular. The construction and materials of the archetypes can be calibrated to be more representative of the local building stock characteristics for any case study that utilises the developed framework.

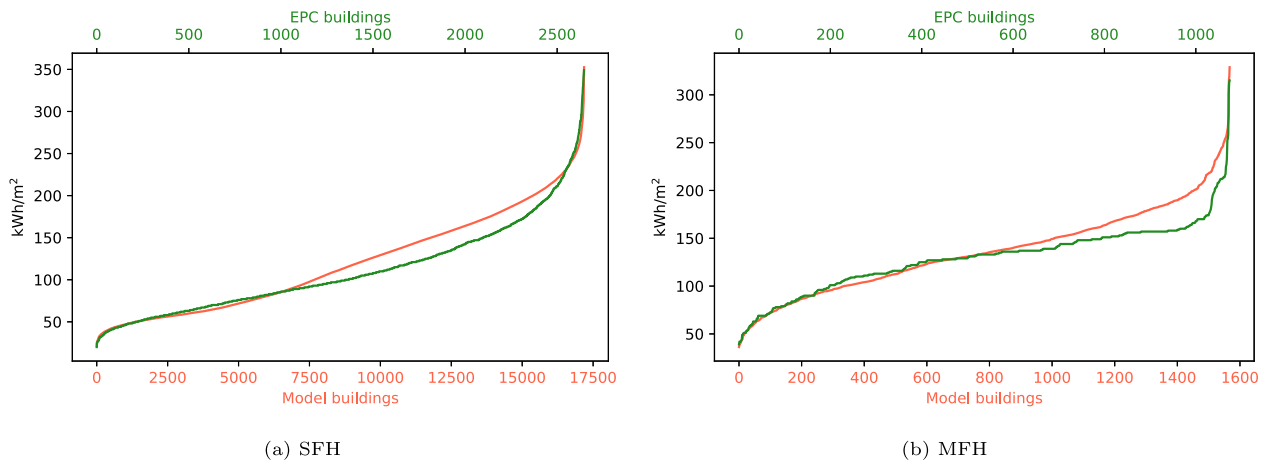


Fig. 6. Gotland energy performance distribution for model simulation and EPC buildings.

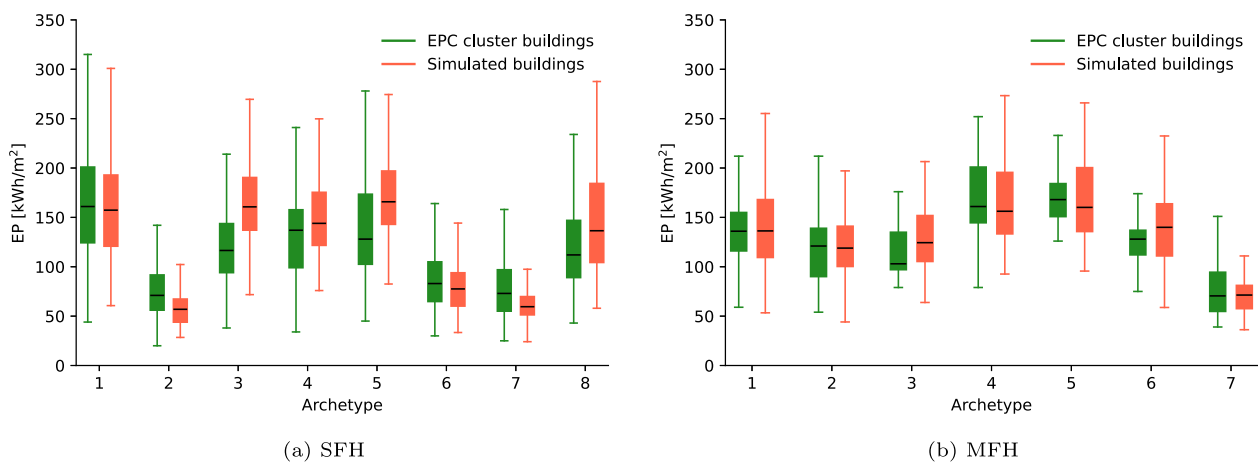


Fig. 7. Energy performance of buildings in each archetype cluster.

A general pattern to be noted in this study, which is most apparent in Fig. 7, is that the energy use for high-performing buildings (low EP value) is underestimated while the energy use for low-performing buildings (high EP value) is overestimated for both building types. The comparison to statistical data also shows an overestimation for energy use, heated area, and occupants for MFH. It is possible that the heated area is overestimated because the calibration of GIS footprint and height to EPC values did not fully capture the actual differentiating factor. The difference is also likely to be an effect of the expansive tourism on the island; no MFH are included as holiday homes, but a significant number of apartments are also rented out to tourists, effectively reducing the listed number of inhabitants and living area in MFH on the island. If this effect could be proven to be true by future research, assuming the GIS data is correct, the higher energy use predicted by the simulation could actually be more representative of the real energy use of the existing residential building stock.

Undeniably, statistical data also comes with inherent uncertainties.⁴ Worth noticing is that the statistical data does not list any energy use for MFH with bio- or fossil fuel burners, while the EPC data does so, resulting in the 6 GWh boiler energy use seen in the simulation results. This indicates some inaccuracies in the reported statistical data. Adjusting the heated area to the statistical value increases the accuracy of the

results, but the simulated energy demand for MFH is still overestimated, especially for the worst performing buildings. This is likely due to the archetypes not being calibrated to represent the typical construction of Gotland buildings. As an example, older MFH on Gotland are usually constructed out of wood and stone [9], while the general Swedish construction of those archetypes is brickwork.

The distribution of EP across the simulated building stock is well captured by the simulation, and the weighted mean of EP for the archetypes is very low. On the other hand, the spread of EP among buildings in some archetype clusters is still quite high (see Fig. 7). This indicates that the different characteristics and identified patterns of the building stock do not necessarily translate into a more homogeneous building energy use, and it is still considerably difficult to group buildings in a way so that the spread of the actual energy use in every archetype is low.

6.2. Building archetypes

A key part of the procedure is to find the correct archetype for all buildings that do not have an EPC. Different methods for identifying the archetype of all GIS buildings were tested to determine which method correctly identified most EPC building archetypes. The applied algorithm identifies the correct archetype for 94% of the MFH, due to the high EPC coverage. For test runs with the available EPC buildings, the correct archetype is identified for only 48% of all SFH in the GIS data set (that does not have an EPC). This creates uncertainties for SFH, how-

⁴ Estimations of data uncertainties was unfortunately not given for the utilised data sets.

ever, the test run of the method was only applied to a small sample of the entire building stock where, more importantly, similar buildings within the same property are only represented by one building in the EPC data set. With the applied method, the geographical clustering of different archetypes is strong; as previously discussed, it is more likely than not that buildings within the same neighbourhood or geographical area are more homogeneous. Nonetheless, detailed assessment of the accuracy of the applied method was not within the scope of this study and is left for further research.

The building materials and construction were derived from sources describing the Swedish residential building stock. A strength of this approach is that once the material IDF files have been compiled, they can be used for any Sweden-based UBEM without any further alterations. This would be true for any other nation or region, providing that similar sources exist or can be developed. With this approach it also means that local differences are not taken into account, but it enables for straightforward archetype calibration to increase representativeness of local building stock characteristics.

6.3. Simulation

The computational time of the full simulation is favourable considering the model complexity, but will certainly depend on the computer used. At 4.5 s per building it is around half that of Johari et al. [44], the study that developed the simulation core upon which the model in this study is built. The Boston model by Cerezo Davila et al. [17], using deterministic parameter modelling, run at approximately 2.6 s per building, while not including “shoebox” thermal zoning calculations.

The EnergyPlus simulations for the building heating systems in this study are ideal system calculations, using a simplified mathematical model. It is possible to use EnergyPlus to create significantly more detailed heating system simulations to mimic real-world conditions, however, the construction and thermal zone modelling of each building would similarly require much more detail. The extra time and effort required for such modelling is extensive, and would be a more relevant approach for individual BEM, not large-scale UBEM.

The variety and amount of data that becomes available for every building with the methodology of this study creates several other interesting possibilities. In future research, buildings with certain characteristics can be specifically targeted for scenarios analysing the impact of energy efficiency measures. Examples include targeting low-income areas of a city, or analysing historic and cultural heritage buildings or any other building type where measure scenarios are required to be very specific.

6.4. Limitations

It should be emphasised that the goal of the methodology presented in this paper is to capture the heterogeneity of energy performance and loads across the building stock as a whole, not to create precise, detailed models of individual buildings. Furthermore, since high-resolution (building-specific) statistical or metered data was not used, some parameters were still based on average or template values, such as the number of occupants per building. This means that the model framework is likely ill-suited for some applications that require higher accuracy and spatial resolution, or for assessing some specific patterns, like differences in energy use per capita or the relationship between energy use and population density. If detailed population data is obtained for a study area it could increase the accuracy and usefulness of the model, but at the cost of deferring from the goal of utilising only open-access data.

Only full-time residential buildings were evaluated in this study, leaving out non-residential buildings and holiday homes. The main reason for this decision was the lack of EPC coverage of these building categories, which renders the developed archetype identification approach impossible to use. Furthermore, division into end-use archetypes (schools, healthcare buildings etc.) would entail many uncertainties as

the number of buildings in each category is very limited in the case study area. Future research into the topic could elaborate on the inclusion of these building categories, which would involve developing similar archetypes for these buildings using another approach. The model used in this study could then incorporate these buildings, as the statistical validation data is available for non-residential buildings as well.

The use of EPC data for archetype development, calibration and model validation is highly beneficial for modelling practices, however, the data coverage for SFH is still low and may be subject to sample bias. It is probable that some building characteristics are more common in the SFH not covered by EPCs, effectively skewing the data. This is a probable reason for the higher share of electricity in the simulation results compared to the other energy sources. Future incentives for higher EPC coverage, or the use of auxiliary input data such as energy surveys, could potentially counteract this issue.

7. Conclusions

This study developed a regional bottom-up energy model for around 19,000 residential buildings on the island of Gotland, Sweden, using an optimised approach to utilise available UBEM techniques. Open-access national data sets, including EPC, cadastral, and statistical data, were used to identify and characterise 15 building archetypes, which were extrapolated to all buildings using a novel approach. Probabilistic approaches were used in the simulation of building infiltration and occupancy, using an EnergyPlus-based simulation core based on previously validated and published UBEM studies.

Model simulation results using uncalibrated archetypes show a high correlation to normalised energy performance data from regional EPCs, and capture the heterogeneity of energy performance in the building stock. The results also show a high correlation to statistical annual energy use data, and the difference in total annual energy use for SFH was 3.3%, and with a weighted mean of only 0.2% for archetype energy performance. The results were slightly overestimated for MFH, instead showing a total difference of 27.6% and with a weighted mean of 4.3%, likely due to misrepresentation of local building characteristics for a few archetypes, or an overestimation of the heated area in the simulation.

The results of this study prove that UBEM techniques are useful even on a regional scale; rural areas can be incorporated to create large-scale regional models, increasing the usefulness of UBEM for regional planning or forecasting. The study results also highlight the usefulness of open data sets and probabilistic approaches for archetype development and simulation parameter modelling. The resulting model framework is created to be easily transferable and adaptable to different geographical areas, and the archetypes are possible to calibrate to local characteristics, further increasing accuracy and representativeness. The comparatively low computational cost also enables the framework to be used for truly large-scale simulations while maintaining high accuracy.

CRedit authorship contribution statement

Lukas Dahlström: Writing – original draft, Methodology, Conceptualization; **Fatemeh Johari:** Writing – review & editing, Visualization, Supervision; **Joakim Widén:** Writing – review & editing, Supervision.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Table of building archetypes

Table A.5
Identified building archetypes for SFH and MFH, including relevant archetype parameters.

Archetype	Postal town	Location	Construction year	Heated area [m ²]	Nr of floors	EP [kWh/m ²]	Income [SEK/pl]	Age group	Population density [p/km ²]	Building footprint [m ²]	Building height [m]	Share of district heating	Share of boilers	Share of electric heating	Share of GSHP	Share of XAHP	Share of ASHP	Share of HRV	Share of natural ventilation	Share of basement
SFH																				
1	Tingstade	Rural	1948	166		146	90k	5.50	308.79	113.98	6.94	0.00	0.85	0.10	0.00	0.00	0.05	0.05	0.60	0.31
2	Visby	Urban	1956	149		75	145k	5.30	2221.59	98.73	6.13	0.00	0.05	0.06	0.07	0.07	0.75	0.07	0.63	0.43
3	Slite	Rural	1970	125	1	117	135k	6.88	20.63	151.94	3.22	0.00	0.09	0.85	0.00	0.00	0.06	0.14	0.45	0.13
4	Klinte-hamn	Urban	1973	105	1	154	102k	5.40	1923.18	132.32	3.88	0.96	0.01	0.03	0.00	0.00	0.00	0.06	0.29	0.18
5	Klinte-hamn	Rural	1800	108	2	175	114k	5.63	15.18	115.69	5.38	0.12	0.22	0.41	0.07	0.00	0.17	0.02	0.70	0.14
6	Visby	Rural	1909	137		94	103k	5.73	58.20	122.81	5.20	0.00	0.09	0.08	0.15	0.00	0.68	0.00	0.79	0.38
7	Visby	Urban	1982	165		72	122k	5.80	177.60	139.53	3.89	0.00	0.09	0.11	0.09	0.30	0.41	0.35	0.29	0.06
8	Visby	Urban	1945	129		151	134k	6.00	1952.00	52.16	3.43	0.01	0.14	0.78	0.00	0.00	0.08	0.09	0.48	0.41
MFH																				
1	Hemse	Urban	1981	1049	2	137	153k	6.89	2718.60	361.59	8.33	0.93	0.00	0.07	0.00	0.00	0.00	0.12	0.11	0.22
2	Havdhem	Rural	1982	179	1	121	121k	5.00	668.80	251.51	3.02	0.00	0.00	0.95	0.01	0.00	0.03	0.46	0.23	0.11
3	Tingstade	Rural	1980	367	1	101	118k	5.30	808.00	413.01	4.35	0.00	0.86	0.14	0.00	0.00	0.00	0.11	0.31	0.27
4	Visby	Urban	1755	202	1	161	205k	6.00	1338.40	128.65	3.12	0.74	0.00	0.18	0.04	0.00	0.02	0.05	0.72	0.37
5	Visby	Urban	1905	486	2	164	199k	6.00	1825.60	243.00	5.60	0.94	0.00	0.06	0.00	0.00	0.00	0.12	0.67	0.57
6	Slite	Urban	1971	883	2	129	178k	5.60	1182.40	547.87	7.64	0.90	0.00	0.10	0.00	0.00	0.00	0.17	0.08	0.25
7	Visby	Urban	1964	216	1	71	103k	6.30	1385.60	215.67	4.49	0.01	0.01	0.12	0.38	0.05	0.43	0.27	0.46	0.27

References

- [1] European Parliament, Council of the European Union, Directive (EU) 2024/1275 of the European Parliament and of the Council on the energy performance of buildings, Technical Report, Directorate-General for Energy, 2024. <https://eur-lex.europa.eu/eli/dir/2024/1275/oj>.
- [2] European Commission, 2030 climate & energy framework, 2023. Accessed: 2023-01-20, https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2030-climate-energy-framework_en.
- [3] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, J. Widén, Urban building energy modeling: state of the art and future prospects, Renew. Sustain. Energy Rev. 128 (2020) 109902. <https://doi.org/10.1016/j.rser.2020.109902>
- [4] L. Dahlström, T. Broström, J. Widén, Advancing urban building energy modelling through new model components and applications: a review, Energy Build. 266 (2022) 112099. <https://doi.org/10.1016/j.enbuild.2022.112099>
- [5] O. Pasichnyi, J. Wallin, F. Levih, H. Shahrokni, O. Kordas, Energy performance certificates—New opportunities for data-enabled urban energy policy instruments?, Energy Policy 127 (2019) 486–499. <https://doi.org/10.1016/j.enpol.2018.11.051>
- [6] A. Barney, H. Polatidis, D. Haralambopoulos, Decarbonisation of islands: a multi-criteria decision analysis platform and application, Sustain. Energy Technol. Assess. 52 (2022) 102115. <https://doi.org/10.1016/j.seta.2022.102115>
- [7] D. Sopher, C. Juhlin, T. Levendal, M. Erlström, K. Nilsson, J.P. D.S. Soares, Evaluation of the subsurface compressed air energy storage (CAES) potential on Gotland, Sweden, Environ. Earth Sci. 78 (2019). <https://doi.org/10.1007/s12665-019-8196-1>
- [8] V. Milić, K. Ekelöv, M. Andersson, B. Moshfegh, Evaluation of energy renovation strategies for 12 historic building types using LCC optimization, Energy Build. 197 (2019) 156–170. <https://doi.org/10.1016/j.enbuild.2019.05.017>
- [9] T. Broström, A. Donarelli, F. Berg, For the categorisation of historic buildings to determine energy saving, Int. J. Archit., Art Des. (2017). <https://doi.org/10.19229/2464-9309/1212017>
- [10] T. Broström, P. Eriksson, L. Liu, P. Rohdin, F. Ståhl, B. Moshfegh, A method to assess the potential for and consequences of energy retrofits in swedish historic buildings, Hist. Environ. 5 (2) (2014) 150–166. <https://doi.org/10.1179/1756750514Z.00000000055>
- [11] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling—A review of a nascent field, Build. Environ. 97 (2016) 196–202. <https://doi.org/10.1016/j.buildenv.2015.12.001>
- [12] M. Ferrando, F. Causone, T. Hong, Y. Chen, Urban building energy modeling (UBEM) tools: a state-of-the-art review of bottom-up physics-based approaches, Sustain. Cities Soc. 62 (2020) 102408. <https://doi.org/10.1016/j.scs.2020.102408>
- [13] D. Kong, A. Cheshmehzangi, Z. Zhang, S.P. Ardakani, T. Gu, Urban building energy modeling (UBEM): a systematic review of challenges and opportunities, Energy Effic. 16 (2023). <https://doi.org/10.1007/s12053-023-10147-z>
- [14] A. Malhotra, J. Bischof, A. Nichersu, K.-H. Häfele, J. Exenberger, D. Sood, J. Allan, J. Frisch, C. van Treeck, J. O'Donnell, G. Schweiger, Information modelling for urban building energy simulation—A taxonomic review, Build. Environ. 208 (2022) 108552. <https://doi.org/10.1016/j.buildenv.2021.108552>
- [15] U. Ali, M.H. Shamsi, C. Hoare, E. Mangina, J. O'Donnell, Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, Energy Build. 246 (2021) 111073. <https://doi.org/10.1016/j.enbuild.2021.111073>
- [16] E. Kamel, A systematic literature review of physics-based urban building energy modeling (UBEM) tools, data sources, and challenges for energy conservation, Energies 15 (2) (2022). <https://doi.org/10.3390/en15228649>
- [17] C. Cerezo Davila, C.F. Reinhart, J.L. Bemis, Modeling Boston: a workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250. <https://doi.org/10.1016/j.energy.2016.10.057>
- [18] T. Hong, Y. Chen, X. Luo, N. Luo, S.H. Lee, Ten questions on urban building energy modeling, Build. Environ. 168 (2020) 106508. <https://doi.org/10.1016/j.buildenv.2019.106508>
- [19] A. Oraipoulos, B. Howard, On the accuracy of urban building energy modelling, Renew. Sustain. Energy Rev. 158 (2022) 111976. <https://doi.org/10.1016/j.rser.2021.111976>
- [20] J. von Platten, C. Holmberg, M. Mangold, T. Johansson, K. Mjörnell, The renewing of energy performance certificates - Reaching comparability between decade-apart energy records, Appl. Energy (255) (2019). <https://doi.org/10.1016/j.apenergy.2019.113902>
- [21] G. Mutani, V. Todeschi, GIS-based urban energy modelling and energy efficiency scenarios using the energy performance certificate database, Energy Effic. (14) (2021). <https://doi.org/10.1007/s12053-021-09962-z>
- [22] T. Johansson, T. Olofsson, M. Mangold, Development of an energy atlas for renovation of the multifamily building stock in Sweden, Appl. Energy 203 (2017) 723–736. <https://doi.org/10.1016/j.apenergy.2017.06.027>
- [23] F. Berg, A. Donarelli, Energy performance certificates and historic apartment buildings: a method to encourage user participation and sustainability in the refurbishment process, Hist. Environ. 10 (2) (2019) 224–240. <https://doi.org/10.1080/17567505.2019.1592836>
- [24] A. Gonzalez-Caceres, J. Karlshøj, T. Arvid Vik, E. Hempel, T. Rammer Nielsen, Evaluation of cost-effective measures for the renovation of existing dwellings in the framework of the energy certification system: a case study in Norway, Energy Build. 264 (2022) 112071. <https://doi.org/10.1016/j.enbuild.2022.112071>
- [25] R. Guo, M.H. Shamsi, M. Sharifi, D. Saelens, Exploring uncertainty in district heat demand through a probabilistic building characterization approach, Appl. Energy 377 (2025) 124411. <https://doi.org/10.1016/j.apenergy.2024.124411>

- [26] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, A. Hajiah, Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): a residential case study in Kuwait City, *Energy Build.* 154 (2017) 321–334. <https://doi.org/10.1016/j.enbuild.2017.08.029>
- [27] J. Sokol, C. Cerezo Davila, C.F. Reinhart, Validation of a Bayesian-based method for defining residential archetypes in urban building energy models, *Energy Build.* 134 (2017) 11–24. <https://doi.org/10.1016/j.enbuild.2016.10.050>
- [28] S. Dabirian, K. Panchabikesan, U. Eicker, Occupant-centric urban building energy modeling: approaches, inputs, and data sources—A review, *Energy Build.* 257 (2022) 111809. <https://doi.org/10.1016/j.enbuild.2021.111809>
- [29] G. Happle, J. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, *Energy Build.* (174) (2018) 276–292. <https://doi.org/10.1016/j.enbuild.2018.06.030>
- [30] Y.Q. Ang, Z.M. Berzolla, S. Letellier-Duchesne, V. Jusiega, C. Reinhart, UBEM.io: a web-based framework to rapidly generate urban building energy models for carbon reduction technology pathways, *Sustain. Cities Soc.* 77 (2022) 103534. <https://doi.org/10.1016/j.scs.2021.103534>
- [31] F. Johari, F. Shadram, J. Widén, Urban building energy modeling from geo-referenced energy performance certificate data: development, calibration, and validation, *Sustain. Cities Soc.* 96 (2023) 104664. <https://doi.org/10.1016/j.scs.2023.104664>
- [32] F. Johari, O. Lindberg, U.H. Ramadhani, F. Shadram, J. Munkhammar, J. Widén, Analysis of large-scale energy retrofit of residential buildings and their impact on the electricity grid using a validated UBEM, *Appl. Energy* 361 (2024) 122937. <https://doi.org/10.1016/j.apenergy.2024.122937>
- [33] J. Roth, A. Martin, C. Miller, R.K. Jain, SynCity: using open data to create a synthetic city of hourly building energy estimates by integrating data-driven and physics-based methods, *Appl. Energy* 280 (2020) 115981. <https://doi.org/10.1016/j.apenergy.2020.115981>
- [34] C. Song, Z. Deng, W. Zhao, Y. Yuan, M. Liu, S. Xu, Y. Chen, Developing urban building energy models for Shanghai City with multi-source open data, *Sustain. Cities Soc.* 106 (2024) 105425. <https://doi.org/10.1016/j.scs.2024.105425>
- [35] C. Younes, C. Abishdid, G. Bitsuamlak, Air infiltration through building envelopes: a review, *J. Build. Phys.* 35 (2012) 267–302. <https://doi.org/10.1177/1744259111423085>
- [36] J. Jokisalo, J. Kurnitski, M. Korpi, T. Kalamees, J. Vinha, Building leakage, infiltration, and energy performance analyses for Finnish detached houses, *Build. Environ.* 44 (2) (2009) 377–387. <https://doi.org/10.1016/j.buildenv.2008.03.014>
- [37] K. Pietrzyk, C.-E. Hagentoft, Probabilistic analysis of air infiltration in low-rise buildings, *Build. Environ.* 43 (4) (2008) 537–549. Part Special: Building Performance Simulation, <https://doi.org/10.1016/j.buildenv.2007.01.024>
- [38] Y. Wang, S. Shi, Z. Zhou, S. Guo, B. Zhao, Air infiltration rate distribution across Chinese five climate zones: a modelling study for rural residences, *Build. Environ.* 252 (2024) 111284. <https://doi.org/10.1016/j.buildenv.2024.111284>
- [39] S. Shi, C. Chen, B. Zhao, Air infiltration rate distributions of residences in Beijing, *Build. Environ.* 92 (2015) 528–537. <https://doi.org/10.1016/j.buildenv.2015.05.027>
- [40] L. Eskola, Ü. Alev, E. Arumägi, J. Jokisalo, A. Donarelli, K. Sirén, T. Kalamees, Airtightness, air exchange and energy performance in historic residential buildings with different structures, *Int. J. Vent.* (14) (2015). <https://doi.org/10.1080/14733315.2015.11684066>
- [41] A. Persily, A. Musser, S.J. Emmerich, Modeled infiltration rate distributions for U.S. housing, *Indoor Air* 20 (6) (2010) 473–485. <https://doi.org/10.1111/j.1600-0668.2010.00669.x>
- [42] P. Wahi, T. Konstantinou, H. Visscher, M.J. Tenpierik, Evaluating building-level parameters for lower-temperature heating readiness: a sampling-based approach to addressing the heterogeneity of Dutch housing stock, *Energy Build.* 322 (2024) 114703. <https://doi.org/10.1016/j.enbuild.2024.114703>
- [43] R. Nouvel, M. Zirak, V. Coors, U. Eicker, The influence of data quality on urban heating demand modeling using 3D city models, *Comput. Environ. Urban Syst.* 64 (2017) 68–80. <https://doi.org/10.1016/j.compenvurbsys.2016.12.005>
- [44] F. Johari, J. Munkhammar, F. Shadram, J. Widén, Evaluation of simplified building energy models for urban-scale energy analysis of buildings, *Build. Environ.* 211 (2022) 108684. <https://doi.org/10.1016/j.buildenv.2021.108684>
- [45] S. Philip, eppy 0.5.63, 2022. <https://pypi.org/project/eppy/>.
- [46] L. Dahlström, F. Johari, T. Broström, J. Widén, Identification of representative building archetypes: a novel approach using multi-parameter cluster analysis applied to the Swedish residential building stock, *Energy Build.* 303 (2024) 113823. <https://doi.org/10.1016/j.enbuild.2023.113823>
- [47] X. Faure, T. Johansson, O. Pasichnyi, The impact of detail, shadowing and thermal zoning levels on urban building energy modelling (UBEM) on a district scale, *Energies* 15 (4) (2022). <https://doi.org/10.3390/en15041525>
- [48] Lantmäteriet, Lantmäteriet, 2024. Accessed: 2024-02-18, <https://www.lantmateriet.se/en/>.
- [49] G. Kakoulaki, A. Martinez, P. Florio, Non-commercial Light Detection and Ranging (LiDAR) data in Europe, Technical Report, European Commission, 2009. <https://doi.org/10.2760/212427>
- [50] SCB, Number of buildings for seasonal and secondary use owned by private persons or estates of deceased persons by region. Year 1998 - 2023, 2024. https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START_BO_BO0104_BO0104H/BO0104T08/.
- [51] Boverket, Boverket, 2024. <https://www.boverket.se/en/start/>.
- [52] C. Hjortling, F. Björk, M. Berg, T. af Klintberg, Energy mapping of existing building stock in Sweden—Analysis of data from energy performance certificates, *Energy Build.* (153) (2017) 341–355. <https://doi.org/10.1016/j.enbuild.2017.06.073>
- [53] M. Mangold, M. österbring, H. Wallbaum, Handling data uncertainties when using Swedish energy performance certificate data to describe energy usage in the building stock, *Energy Build.* 102 (2015) 328–336. <https://doi.org/10.1016/j.enbuild.2015.05.045>
- [54] Boverket, Energideklarationens innehåll, 2024. <https://www.boverket.se/sv/energideklaration/energideklaration/energideklarationens-innehall/>.
- [55] SCB, Statistics Sweden, 2024. <https://www.scb.se/en/>.
- [56] Climate.OneBuilding, Repository of Building Simulation Climate Data, 2024. Accessed: 2024-10-10, <https://climate.onebuilding.org/default.html>.
- [57] SMHI, SMHI Data, 2024. <https://www.smhi.se/data>.
- [58] B. Gepts, E. Meex, E. Nuyts, E. Knapen, G. Verbeeck, Existing databases as means to explore the potential of the building stock as material bank, *IOP Conf. Ser.* 225 (2019) 012002. <https://doi.org/10.1088/1755-1315/225/1/012002>
- [59] Boverket, Boverkets byggregler, BBR, BFS 2011:6 med ändringar till och med BFS 2020:4, Technical Report, Boverket, 2020. www.boverket.se/publikationer.
- [60] C. Björk, L. Nordling, L. Reppen, Så Byggs Villan - Svensk Villaarkitektur Från 1890 Till 2010, *Svensk Byggtjänst*, 2015.
- [61] C. Björk, L. Nordling, L. Reppen, Så Byggs hUSEN 1880–2020, *Svensk Byggtjänst*, 2021.
- [62] A. Mardiana-Idayu, S.B. Riffat, Review on heat recovery technologies for building applications, *Renew. Sustain. Energy Rev.* 16 (2) (2012) 1241–1255. <https://doi.org/10.1016/j.rser.2011.09.026>
- [63] J. Widén, Computationally efficient upscaling of Markovian occupancy models for urban-level building simulations, *J. Build. Perform. Simul.* 0 (0) (2024) 1–28. <https://doi.org/10.1080/19401493.2024.2445134>
- [64] Svebyprogrammet, Brukarindata bostäder - Remissversion, Technical Report, Sveby, 2024.
- [65] The Board of Trustees of the University of Illinois and the Regents of the University of California through the Ernest Orlando Lawrence Berkeley National Laboratory, Illinois, The Board of Trustees of the University of and Laboratory, Input Output Reference - EnergyPlus 9.2, 2020. Accessed: 2024-12-18, <https://bigladdersoftware.com/epx/docs/9-2/input-output-reference/index.html>.
- [66] Statistics Sweden, Average useful floor space per person by region, type of household and type of housing. Year 2012–2023, 2024. Accessed: 2024-10-10, https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START_HE_HE0111_HE0111A/HushallT23/.
- [67] Energimyndigheten, Från watt till lumen, 2018. Accessed: 2024-12-18, <https://www.energimyndigheten.se/effektiv-energianvandning/hushall/nar-du-ska-kopanya-produkter/belysning/fran-watt-till-lumen/>.