

Article

Urban Buildings Energy Consumption Estimation Leveraging High-Performance Computing: A Case Study of Bologna

Aldo Canfora ^{1,2,*} , Eleonora Bergamaschi ² , Riccardo Mioli ² , Federico Battini ³ , Mirko Degli Esposti ¹ ,
Giorgio Pedrazzi ²  and Chiara Dellacasa ² 

¹ Department of Physics and Astronomy, University of Bologna, Via Irnerio 46, 40126 Bologna, Italy; mirko.degliesti@unibo.it

² HPC Department, Cineca, Via Magnanelli 6/3, 40033 Casalecchio di Reno, Italy; e.bergamaschi@cineca.it (E.B.); r.mioli@cineca.it (R.M.); g.pedrazzi@cineca.it (G.P.); c.dellacasa@cineca.it (C.D.)

³ IFAB, Via Galliera 32, 40121 Bologna, Italy; federico.battini@ifabfoundation.org

* Correspondence: aldo.canfora2@unibo.it

Abstract

Urban building energy modeling (UBEM) is crucial for assessing energy consumption patterns at the city-scale and for supporting data driven planning and decarbonization strategies. However, its practical deployment is often hindered by the need to balance detailed physics-based simulations with acceptable computation times when thousands of buildings are involved. This work presents a large-scale real world UBEM case study and proposes a workflow that combines EnergyPlus simulations, high-performance computing (HPC), and open urban datasets to model the energy consumption of the building stock in the Municipality of Bologna, Italy. Geometric data such as building footprints and heights were acquired from the Bologna Open Data portal and complemented by aerial light detection and ranging (LiDAR) measurements to refine elevations and roof geometries. Non-geometrical building characteristics, including wall materials, insulation levels, and window properties, were derived from local building regulations and the European TAB-ULA project, enabling the assignment of archetypes in contexts where granular information about building materials is not available. The pipeline's modular design allows us to analyze different combinations of retrofitting scenarios, making it possible to identify the groups of buildings that would benefit the most. A key feature of the workflow is the use of Leonardo, the supercomputer hosted and managed by Cineca, which made it possible to simulate the energy consumption of approximately 25,000 buildings in less than 30 min. In contrast to approaches that mainly reduce computation time by simplifying the physical model or aggregating representative buildings, the HPC-based workflow allows the entire building stock to be individually simulated (within the intrinsic simplifications of UBEM) without introducing further compromises in model detail. Overall, this case study demonstrates that the combination of open data and HPC-accelerated UBEM can deliver city-scale energy simulations that are both computationally tractable and sufficiently detailed to inform municipal decision-making and future digital twin applications.



Received: 12 November 2025

Revised: 12 December 2025

Accepted: 16 December 2025

Published: 20 December 2025

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

Keywords: HPC simulations; UBEM; urban digital twin; EnergyPlus; ray

1. Introduction

One of the most urgent challenges in the global fight against climate change is the reduction of energy demand and carbon emissions from buildings. Worldwide, the build-

ing sector accounts for nearly 30% of total greenhouse gas emissions, making it a critical target for decarbonization efforts [1]. Similar conclusions were reached by Chen et al. (2017), who showed that buildings in major cities consume between 30 and 70% of total primary energy, highlighting the retrofit of existing building stock as a central strategy for mitigating global warming [2]. Therefore, in densely populated urban environments, understanding how much energy each building consumes—and how these consumption patterns are distributed spatially—is essential for optimizing renovation strategies, prioritizing investments, and designing effective climate-action policies. Against this backdrop, UBEM has emerged as a crucial field aimed at simulating the energy performance of entire building stocks. Unlike traditional single-building models, UBEM relies on geospatial data and building archetypes to produce bottom-up simulations of energy demand, emissions, and retrofit scenarios across thousands of buildings [3,4]. Hong et al. (2020) emphasized that modeling urban building stocks collectively can significantly improve energy efficiency, resilience, and long-term sustainability, especially when supported by HPC and rich datasets [5].

Over the past decade, many UBEM studies have been developed for real cities, consolidating the method's role in climate mitigation and adaptation planning. A landmark contribution came from the MIT Sustainable Design Lab, Cerezo et al. (2016), which demonstrated a replicable workflow that used publicly available GIS data and archetype libraries to model over 83,000 buildings in Boston [6]. Their work illustrated how UBEM can assess citywide interventions such as photovoltaic deployment or district energy networks, while also stressing the importance of archetype accuracy, calibration, and integration with empirical data. Since then, UBEM has been applied in diverse urban contexts. Mutani et al. (2020) mapped building energy demand in Turin and revealed that older, poorly insulated buildings—often in lower-income neighborhoods—had the highest consumption, thus linking energy modeling with social equity [7]. Buckley et al. (2021) achieved similar insights for Dublin, identifying retrofit strategies capable of reducing emissions by up to 60% by 2030 [8]. Studies in Geneva [9], Barcelona [10], and several Mexican cities [11] have further shown how UBEM can capture climatic, morphological, and socio-economic differences across urban areas, strengthening its role as a decision-support tool for planners.

The evolution of UBEM has been accompanied by methodological diversification. Early models were grounded purely in physical principles, but later approaches began integrating statistical techniques to better capture phenomena not explicitly represented by physics. Foucquier et al. (2013) showed that hybrid grey-box models, which combine physical and empirical methods, offer a favorable balance between computational efficiency and accuracy—particularly advantageous for large-scale applications [12]. Foundational reviews of the field [13] further distinguished UBEM from earlier top-down energy models, advocating for bottom-up simulations integrating heat transfer dynamics, occupancy patterns, and microclimatic influences. Ali et al. (2021), in a comprehensive review, reported that bottom-up, physics-based methods still dominate the field, representing nearly four times as many studies as purely data-driven approaches [14]. They also identified the growing need for hybrid models that maintain physical realism while enhancing scalability and predictive power.

More recently, machine learning (ML) has become increasingly important in UBEM. Yet, its success depends heavily on the availability of high-quality datasets. A pivotal study by Kontokosta et al. (2017) demonstrated that ML models can reliably predict the energy use of over 1.1 million buildings in New York City, offering valuable support for policy design and emissions reduction [15]. Todeschi et al. (2021) compared ML-based and engineering models in Fribourg, showing complementary strengths and supporting the case for hybrid UBEM approaches [16]. Additional representative studies confirmed the

high predictive accuracy of neural networks, support vector machines, and hybrid physical data-driven models [17,18].

Microclimatic effects are of significant importance and should be duly considered given their influence on energy consumption. Nevertheless, they remain challenging to model at large scale. In the UBEM context, some researchers have explored coupling building energy simulations with microscale urban climate models to capture local wind, radiation, and temperature effects more realistically. Obstacle-resolving models such as MITRAS provide detailed representations of airflow and heat exchange around buildings, but their high computational cost limits applications to small urban areas [19]. Large-eddy simulation frameworks like PALM-4U offer even finer turbulence-resolving capabilities and integrated urban surface physics, yet remain too demanding for citywide UBEM workflows [20]. To overcome these scale constraints, mesoscale approaches such as COSMO-CLM coupled with an urban parameterization and a building energy model (DCEP-BEM) have been tested, successfully reproducing mean radiant temperature patterns in Berlin but with some systematic biases due to coarser grid resolution [21]. Overall, these studies show the potential of climate-UBEM coupling, while highlighting that computational burden and scale mismatch remain major barriers to widespread adoption.

To support UBEM deployment, numerous software tools have been developed. While their features and levels of detail vary, comparative analyses show that differences in accuracy, usability, and computational cost remain significant and that the lack of standardization hinders widespread adoption [3]. Malhotra et al. (2022) found that EnergyPlus is the most commonly used simulation engine and CityGML the most common data format, though interoperability remains a challenge [22]. In parallel, platforms such as CityBES [23] and the City Energy Analyst (CEA) [24] have democratized access to urban energy modeling by automating UBEM generation and enabling integrated simulation of demand, renewable supply, and district systems. Recently, digital twin oriented tools—such as BIPV-city [25]—have extended these capabilities to real-time applications and renewable energy systems.

Despite significant progress, important challenges persist. UBEMs often require high computational resources, rely on heterogeneous and incomplete datasets, and struggle to capture occupant behavior. Large scale sensitivity analyses show that data quality and accessibility are major determinants of model reliability [26]. Furthermore, as highlighted by Johari et al. (2020), hybrid approaches hold promise but remain difficult to validate due to limited availability of measured data [4]. Efforts such as the rapid calibration method proposed by Chen et al. (2020) demonstrate that efficient, data-informed workflows are possible even when only limited information is available [27].

Building on these insights, this article presents a case-study workflow designed to perform large-scale energy consumption simulations for the building stock of the Municipality of Bologna, leveraging HPC resources. First, we provide a detailed overview of the urban datasets used as the basis for constructing and parameterizing the models. Next, we describe the methodological framework and computational infrastructure employed to scale the simulations to approximately 25,000 individual buildings. Finally, we present the key results, discuss limitations, and outline future pathways to enhance the scalability, accuracy, and practical adoption of the proposed UBEM workflow. A central motivation for this work lies in two tightly connected challenges that currently limit the practical deployment of UBEM as a decision-support system for cities. First, UBEM simulations remain computationally demanding, especially when thousands of buildings must be evaluated across multiple scenarios or when optimization loops are required. As recent studies emphasize, even simplified or reduced-order UBEM approaches can become slow when applied to real urban districts, making HPC essential for enabling timely, scenario-based planning and

interactive urban-scale analyses [28]. Many different approaches were already adopted and implemented to reduce simulation time [13], such as reduction of the level of detail in building modeling [29], simulation of representative buildings out of the entire building stock [30], and simplification algorithms [31]. However, these proposed strategies will inevitably introduce approximations. By contrast, HPC enables parallel UBEM simulations without introducing new modeling techniques, scientific insights, or energy-modeling capabilities, facilitating the simulation of the entire building, which still operates within the inherent simplifications of UBEM, without additional compromises in model detail: computation time is reduced without resorting to overly simplified representations that may lead to errors or discrepancies in the results, but through workflow parallelization. Second, despite methodological advances, real large city-scale UBEM applications remain scarce, particularly those relying solely on open data and operating in contexts with limited building information. Beyond early and now dated examples such as Boston, only a few recent works have demonstrated full city or large-district energy modeling—such as the EURECA platform, tested on small- and medium-sized districts but not yet at full city-scale [32]; the Energies open-data UBEM workflows applied to European municipalities [33]; and the Springer CLIMA 2025 contribution showing neighborhood-scale UBEM integration into decarbonization studies [34]. The limited number of such studies confirms that constructing operational UBEMs for an entire city, especially using heterogeneous, incomplete, or fully open datasets, remains a non-trivial task. In this context, the Bologna case study provides a relevant contribution by combining open municipal datasets with an HPC-accelerated workflow capable of simulating tens of thousands of buildings in minutes, demonstrating both methodological scalability and practical feasibility.

This work is carried out within the “Bologna Digital Twin” initiative, a collaborative project promoted by the Municipality of Bologna and coordinated by Fondazione Bruno Kessler (FBK), in partnership with Cineca, the University of Bologna, and the Fondazione IU Rusconi Ghigi.

2. Building Data and Input Parameters

The accuracy of a UBEM primarily depends on the quality and quantity of available data that are synthesized in the input file of the simulations, such as building geometries, construction materials, installed HVAC systems, occupancy types, and historical energy consumption data. When implementing a city-scale model, it becomes necessary to find a balance between the number of buildings modeled and the level of detail of the available data. Our prototype energy model has limited detail on individual buildings because there are no consistent, in-depth data freely available on the entire urban building stock. For this case study, only available open data were employed and physical properties were assigned to the buildings using a system of archetypes based on the year of construction. This approach allowed us to simulate a very large number of buildings and provide a wide representation of urban consumption and its distribution across the territory. However, the use of open data necessarily introduces a degree of simplification in the representation of the building stock. For instance, publicly available datasets do not provide information on past renovation activities (e.g., window replacements or the installation of external thermal insulation systems) carried out on buildings or on individual dwelling units. Such details are typically documented only in cadastral or land-registry records, to which access is usually restricted.

There are several tools for energy building simulations; among the various alternatives, the choice is to use EnergyPlus [35], one of the most widely used simulation engines in major energy simulation tools such as CityBES [23] and UMI [36]. EnergyPlus requires

input files (.idf files) that specify the characteristics of each building. At the end of the simulation, output files describing the “energy behavior” of the building are produced.

Geometric information is extracted from open data provided by the Municipality of Bologna and digital terrain model (DTM) and digital surface model (DSM) obtained from LiDAR data. Information on construction materials is obtained from the TABULA project and legal regulations. The following paragraphs present the data used to prepare the input files for the simulations.

2.1. Open Data Municipality of Bologna

The Open Data Municipality of Bologna portal [37] managed by the Municipality of Bologna, provides a comprehensive repository of datasets related to different aspects of the city. Among the available datasets, we utilize the following key resources:

- **Building Parcels:** This dataset offers detailed information on individual building parcels within Bologna, including attributes such as plan area and Geo Shape of buildings. Figure 1 shows a visualization of the buildings’ geometries extracted from this dataset.
- **Building Volumes:** This dataset provides three-dimensional representations of buildings, capturing volumetric data such as height, plan area, and Geo Shapes of buildings.
- **Address Numbers:** This dataset contains the geolocated addresses (house numbers) throughout the city and a variable that has been used as proxy for construction year. This variable corresponds to the year in which the building was added to the cadastral records, implicitly assuming that registration occurs shortly after construction. This assumption does not account for subsequent renovations at either the building level (e.g., external insulation) or the individual dwelling level (e.g., window replacements). Nevertheless, since this is the only open-data source enabling the assignment of a building archetype, it represents the best available approximation for assigning construction materials and thermal properties to each building.



Figure 1. Snippet image of building footprint polygons in shapefile format, derived from the Building Parcels dataset, illustrating the spatial representation of some buildings used as input data in this work.

2.2. TABULA

In our study, we also integrate information from TABULA (Typology Approach for Building Stock Energy Assessment) [38], an European research project aimed at creating a standardized typology of residential buildings across EU countries. The project provides reference data for building components, including typical wall constructions, insulation

levels, and thermal transmittance values (U-values) based on building age, typology, and country-specific construction practices.

We use the Italian TABULA dataset [39] to estimate parameters such as wall thickness, thermal conductivity, and overall heat transfer coefficients for the building stock in Bologna. This information is essential for urban energy simulations, particularly when detailed architectural or construction data are not available for each individual building. By leveraging TABULA, we ensure a consistent and replicable approach to modeling the thermal performance of buildings at scale across the urban fabric.

2.3. Legal Regulations

In addition to TABULA, we also refer to Italian national building codes and energy efficiency regulations to verify thermal parameters. These normative references are crucial for aligning our assumptions with the current legal framework, especially for newly built or renovated structures where regulation-based requirements may differ from historical averages provided by TABULA.

By combining empirical typologies with regulatory standards, we aim to ensure both accuracy and legal consistency in our energy modeling approach. This is a list of used regulations:

- Building materials and products–Hygrothermal properties–Procedure for determining the design values (UNI 10351:2015 [40]);
- Building components and building elements–Thermal resistance and thermal transmittance–Calculation method (ISO 6946:2007 [41]).

Both regulations were used to derive the material properties and the calculation procedures for determining the thermal transmittance of the layers characterizing the different archetypes. However, these are not the only sources available to modelers. One additional notable reference is UNI/TR 11552 [42], which provides a catalog of typical thermophysical parameters for building materials and opaque envelope assemblies, offering other possible standardized reference values when product specific data are not available.

2.4. Energy Plus and Climate File

EnergyPlus is an open-source building energy simulation software that has been produced by the U.S. Department of Energy since 1997. It models heat transfer through the building envelope, the operation of HVAC systems, and natural or mechanical ventilation, while also accounting for solar gains, infiltration, and occupant behavior. To perform these simulations, EnergyPlus requires two essential inputs:

1. a building description file (i.e., the `.idf` file) containing information such as geometry, materials, internal loads, and system configurations. Since open data do not provide precise information on the cadastral use category of individual buildings, it was not possible to define detailed, use-specific system schedules for each building based on its category of use. Therefore, a uniform schedule, described in Section 2.5, was applied across all buildings in the simulation;
2. a climate file (`.epw`) that provides meteorological data representative of the study area.

The `.epw` file (EnergyPlus Weather file) is a standardized format that contains hourly data on temperature, humidity, solar radiation, and other climatic variables. By using typical meteorological year (TMY) datasets, EnergyPlus is able to simulate energy performance under realistic local conditions. Our prototype incorporates the TMY 2009–2023 for Bologna [43] This allows the model not only to estimate annual consumption and emissions but also to test “what-if” scenarios, compare retrofit strategies, and evaluate

building performance under different climate assumptions. In this way, the climate file plays a crucial role in ensuring that simulations are both location-specific and reproducible.

2.5. Input Parameters

EnergyPlus simulations require several parameters to be explicitly defined, including indoor setpoint temperatures, HVAC operation schedules, ventilation rates, and internal heat gains. Since these inputs are rarely available at city scale, UBEMs necessarily rely on standardized and averaged assumptions to ensure consistency across the building stock.

In this work, the buildings are conditioned using `ZoneHVAC:IdealLoadsAirSystem`, which provides a simplified representation of heating and cooling demand without specifying detailed system components.

The heating and cooling setpoints are defined as 20 °C and 26 °C, respectively, according to the values provided in UNI/TS 11300 [44]. Ventilation is represented using `ZoneInfiltration:DesignFlowRate` set to 0.3 ACH (air changes per hour), matching the UNI/TS 11300 reference air-renewal rate for naturally ventilated residential buildings. Although EnergyPlus interprets this as infiltration (uncontrolled leakage), in our study, it is used as an equivalent combination of infiltration and natural ventilation rate, which is a standard simplification for UBEM applications and remains separate from any mechanical outdoor-air supply.

Due to data availability constraints, the input files created in the workflow do not model internal gains. We acknowledge that UNI/TS 11300 prescribes reference internal load values by use category; the omission is a simplification tied to the available dataset and the UBEM approach. In future work, if the data allow, we plan to include new archetypes based on building types. This would allow for differentiation in terms of occupancy and therefore for modeling certain internal gains based on this information.

3. Methodology

This workflow is not a methodological innovation but rather a case-study application of standard UBEM practices. Our strategy involves collecting and integrating geospatial and typological data from multiple sources to characterize the geometrical and thermal properties of buildings in the urban context. Some properties, such as height and plan area, are assigned directly to individual buildings, while building materials and window characteristics are specified after partitioning buildings into archetypes and were based on TABULA values. Window sizes, instead, are defined according to regulatory requirements stipulating that the ratio between window area and floor area must be 1:8. This constitutes a modeling simplification that is nonetheless necessary, as determining the actual window dimensions would require access to cadastral data for each building or the application of facade segmentation models to street-level images of the buildings. Such an approach would raise scalability issues, since it would depend on the availability of street-level imagery for every building and would fail whenever facades are not visible from the street.

Following the data collection step, we implemented a pipeline to automatically generate a specific `.idf` file containing all the information for each building. The `.idf` files are then used as input for simulations carried out with EnergyPlus. Figure 2 shows the visualization of an input file that was used for the simulation.

Leveraging the Leonardo HPC infrastructure, it was possible to perform the simulation of approximately 25,000 buildings in less than 30 min using 1120 CPU cores. After the simulation of the current buildings state, we analyzed possible retrofitting scenarios using optimization techniques such as analysis of the Pareto front and identified the archetypes with the higher benefits concerning energy consumption.

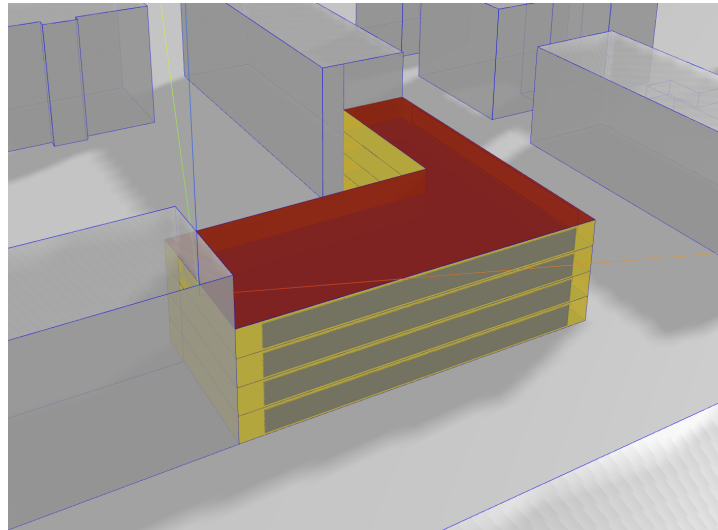


Figure 2. Visualization of an input building geometry. The figure shows a 2.5D (LoD1) building representation generated from footprint vertices included in the .idf file and LiDAR-derived height data, used as input in the proposed workflow. The windows were added respecting the legal regulation ratio $\frac{windowed_{area}}{plan_{area}} = \frac{1}{8}$.

The following sections describe in depth how the homogeneous dataset for the building stock was created.

3.1. Data Integration

Geometrical data were primarily sourced from the Open Data Municipality of Bologna portal, which provides detailed footprints and approximate building heights. LiDAR data were used to fill gaps in the Open Data, providing accurate height information.

The dataset Building Parcels and the dataset Building Volumes have different identification codes (since a single parcel may contain one or more volumetric units); for this reason, it was necessary to perform a spatial join in order to integrate them. The Address Numbers dataset, on the other hand, shares the same identification code as the Building Parcels, although it has significantly fewer unique codes than the latter. The integration of these different sources made it possible to build a comprehensive dataset of approximately 25,000 building constructions, using the parcel building code as the identifier. The information contained in this dataset for each unit includes geometry, surface area, height, and year of construction. As the Building Parcels dataset provides only the building footprint, each floor in the EnergyPlus .idf file is modeled as a single zone. This approximation is unavoidable when performing UBEM using bottom-up approaches based entirely on open data, as the internal floor plans of the buildings are usually non-open data. It should also be acknowledged that EnergyPlus is not inherently designed to capture single-zone approximations, large-scale shading, or urban microclimate effects. Nevertheless, using EnergyPlus as the simulation engine allows for the incorporation of additional details, such as multiple internal zones, once such data becomes available. Flexibility and modularity are the main reason for selecting EnergyPlus as the simulation engine even in UBEM scenarios like the one detailed by this study.

3.2. Building Height Extraction

When data from the Building Volumes dataset were not available, building heights were extracted by using polygon shapefiles representing building footprints obtained from the Open Data Municipality of Bologna portal. To accurately determine building heights, we calculated the difference between digital surface model (DSM) and digital terrain model

(DTM), thus excluding ground elevation. Given that EnergyPlus considers buildings as 2.5D structures (box-shaped without roofs, so LoD1 buildings), only the perimeter of each building polygon was considered. Polygons were buffered to ensure the exclusion of roof heights, and the maximum DSM-DTM difference along each buffered perimeter was selected to represent the building's height.

3.3. Building Archetypes

In order to further differentiate buildings, we grouped buildings into archetypes according to a range of construction years referencing the TABULA classification. Eight types of archetypes were identified based on the year of construction (<1900, 1901–1920, 1921–1945, 1946–1960, 1961–1975, 1976–1990, 1991–2005, >2005) distributed across the areas of Bologna, as illustrated in Figure 3. For each category, we enrich the corresponding input files with specifics extracted from TABULA on building components, including wall materials, insulation, thermal transmittance values (U-values), and window composition.

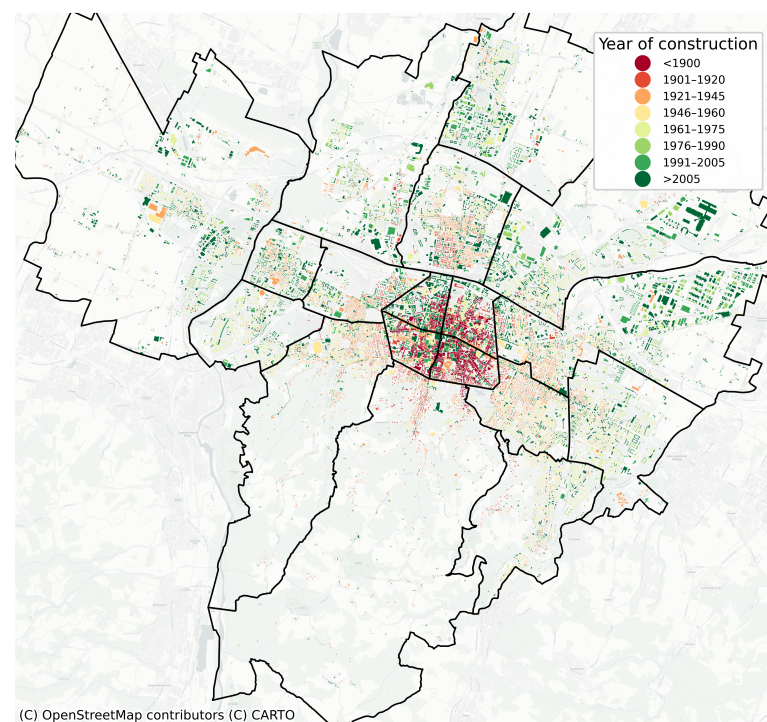


Figure 3. Spatial distribution of archetypes in Bologna. The figure shows building footprint polygons colored by year-of-construction classes and overlaid on a Bologna basemap. Administrative zone boundaries are highlighted to distinguish neighborhoods.

4. Results

The simulation pipeline comprises two sequential stages: the generation of the EnergyPlus input file (.idf) and the execution of the EnergyPlus simulation itself. Managing the parallel execution of this pipeline across tens of thousands of buildings poses substantial orchestration challenges as simulations must be dynamically queued and dispatched as computational resources and “logical” dependencies become available (i.e. a simulation cannot be started before the corresponding .idf file is created), ensuring continuous utilization of CPU cores without manual intervention.

Each building simulation is largely independent from other construction simulations, with the only interdependence arising from shadow interactions between buildings. These interactions are handled by embedding the geometry of neighboring buildings directly within the input file of the target building. Beyond this preprocessing step, simulations operate in complete isolation, without requiring inter-process communication or synchro-

nization. This independence classifies the task as an embarrassingly parallel problem—one that is particularly well suited for HPC environments. Once the .idf files are generated, the workload can be readily distributed across multiple compute nodes, significantly reducing total simulation time.

To address these orchestration and scalability challenges, we adopted the Ray distributed computing framework, implementing each step as an independent Ray task. This architecture enables fault-tolerant and resource-efficient orchestration while providing excellent scalability: by simply adding compute nodes to the Ray cluster, the system can seamlessly execute a larger number of simulations in parallel, thereby reducing total computation time in proportion to the available resources.

For optimal resource utilization, we assigned one CPU core per task. This decision was guided by the current design of EnergyPlus, which does not support multithreading within a single simulation. As a result, allocating multiple cores to a single simulation yields no performance gains. Consequently, distributing the workload such that each simulation runs on a single core maximizes the resource usage.

The EnergyPlus simulations were executed on the Leonardo HPC cluster at Cineca. Each compute node provides 112 CPU cores and 512 GB of RAM, offering a solid foundation for single-node parallelization of the simulation workload. By leveraging 10 compute nodes—equivalent to 1120 cores—the simulation of approximately 25,000 buildings was completed in roughly 30 min. For further discussion of scalability aspects, see Appendix A.

Figure 4 shows the result of the simulation step.



Figure 4. Simulated annual energy consumption (kWh/m^2) for a portion of the building stock. Buildings are colored according to their annual energy use per square meter. Administrative zone and street names are shown in Italian for geographic reference.

4.1. Calibration

To evaluate the reliability of the archetypes and the predictive accuracy of the resulting simulations, a calibration phase must be conducted using real energy consumption data from the residential buildings. However, consumption data with daily or monthly resolution are generally difficult to access due to privacy constraints and regulatory restrictions, making their use challenging unless the data are properly anonymized and aggregated. Moreover, such datasets are often held by multiple private entities, further complicating efforts to collect the comprehensive information required for robust calibration.

To compensate for the lack of direct observational data on energy consumption in Bologna, we performed an approximate validation using the heating needs values reported in the TABULA project for Italian residential archetypes.

The following analysis does not constitute a proper calibration or validation of the model; rather, it provides a qualitative comparison to TABULA archetype values to assess the general consistency of the simulated results.

TABULA heating needs represent the annual energy required for space heating of a building archetype, calculated per unit floor area. Each archetype corresponds to a typical building in Turin with a reference floor area. According to [38], the heating needs reported in the TABULA web tool are computed using the UNI/TS 11300 quasi-steady-state procedure, which estimates energy flows as constant over time based on building geometry, envelope properties, internal gains, and typical operation schedules. For each TABULA archetype, we selected buildings from our Bologna dataset with floor areas similar to those of the corresponding Turin archetype. These buildings serve as proxies for the archetypes. Although the selected buildings are located in Bologna, we simulated them using the Turin weather file. This ensures that the heating demand is estimated under the same climate conditions as TABULA. For each archetype, we averaged the heating energy demand of the area-matched Bologna buildings simulated under Turin climate and compared these averages with the TABULA values, as shown in Figure 4. The figure also includes the simulated heating demand for Bologna buildings with an area similar to the one used as reference in TABULA archetype.

A few studies [45,46] show that the UNI/TS 11300 procedure yields higher estimated energy needs than those simulated using EnergyPlus. This is primarily due to the different methodologies used to calculate consumption. Indeed, EnergyPlus uses a dynamic method that simulates energy consumption by accounting for continuous changes that influence energy use.

Consistent with these findings, Figure 5 illustrates that EnergyPlus simulations predict lower absolute heating needs than those reported in TABULA.

However, the overall trends across archetypes remain consistent, indicating that the simulations provide reasonable estimates for city-scale UBEM applications. Nevertheless, since this remains a qualitative assessment, the model is not validated, and the following results are intended to illustrate qualitative trends rather than provide accurate quantitative predictions.

4.2. Scenario Analysis

In contexts where a proper calibration is missing, simulations can still be leveraged for qualitative “what-if” scenario analyses. These analyses explore the potential energy impact of hypothetical interventions on the building stock, although they do not provide quantitative forecasts.

Typical scenarios analyzed include the following:

- Neighborhoods benefit: which neighborhoods would benefit most from efficiency policies.
- No renovation vs. full retrofit: assessing the citywide performance improvement based on building’s envelope improvement and window replacement;

While qualitative, these scenarios provide valuable foresight for long-term urban planning, particularly in districts under redevelopment pressure or policy-driven transformation goals.

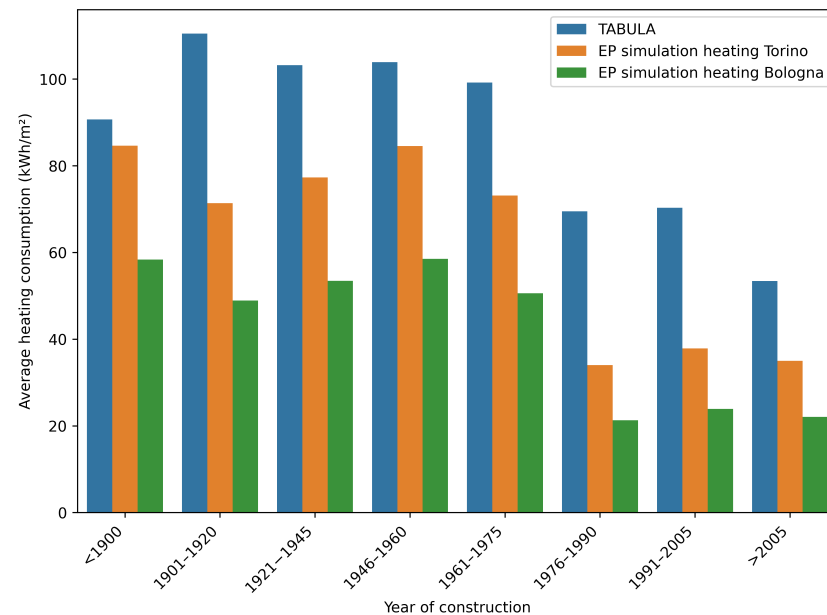


Figure 5. Heating needs per archetype. Bars show (1) TABULA heating needs for the Turin archetype, (2) average EnergyPlus-simulated heating demand of Bologna buildings with similar floor areas using the Turin weather file (for comparison with TABULA), and (3) average simulated heating demand of the same Bologna buildings using Bologna weather (shown for reference).

4.2.1. Impact of Building Renovations on Neighborhood Energy Performance

For each archetype, TABULA proposes two types of refurbishment. We chose the standard one and applied it to all the archetypes; then, for each neighborhood, we calculated the average total energy savings percentage in kWh/m² achievable across the buildings in that area. As can be seen in Figure 6, there is no restricted subset of neighborhoods that would benefit the most in terms of energy savings. This is because old buildings are evenly spread across all the neighborhoods.

4.2.2. City-Scale Effects of Building Retrofits

Figure 7 shows the energy consumption of buildings before and after the refurbishment. Each point represents a building. The x-axis shows the energy consumption of the existing state, and the y-axis shows the energy consumption after standard refurbishment. The closer a building is to the red line, the less impact the refurbishment will have on it. As can be seen, the most recent buildings are very close to the red line because the construction materials used are already efficient enough. On the other hand, archetypes from the middle of the past century offer the best advantages in terms of consumption improvement.

Refurbishment planning often involves trade-offs between multiple conflicting objectives, such as reducing energy consumption, minimizing investment costs, and limiting the number of buildings targeted for intervention.

In such multi-objective contexts, the Pareto front is a valuable decision-making tool. A Pareto front represents the set of non-dominated solutions (i.e., scenarios in which no objective can be improved without worsening another); in other words, rather than focusing on a single optimal solution, the front highlights a range of optimal trade-offs. In the context of refurbishment, this might mean identifying scenarios where maximum energy savings are achieved for a given cost or where the highest return is obtained from refurbishing the fewest buildings.

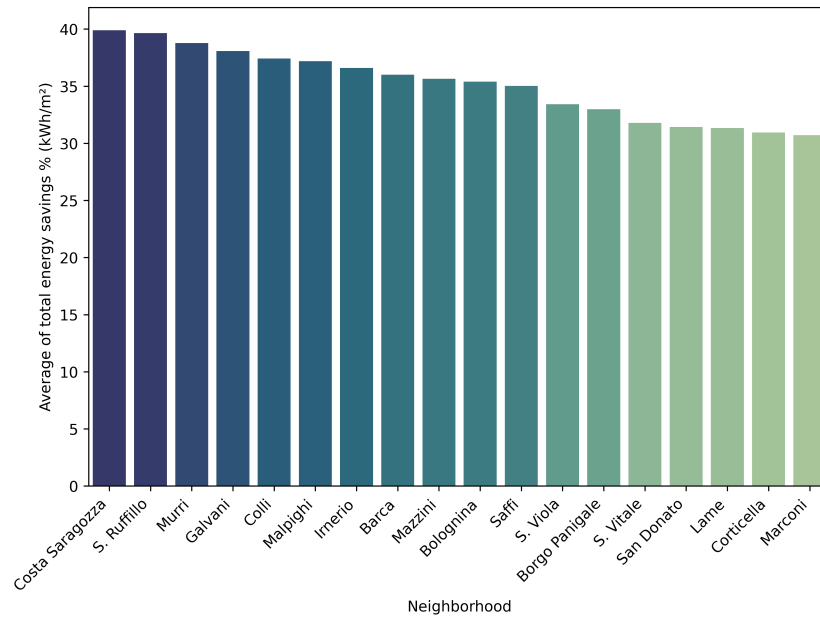


Figure 6. Total energy savings percentage (in kWh/m²) grouped by neighborhoods.

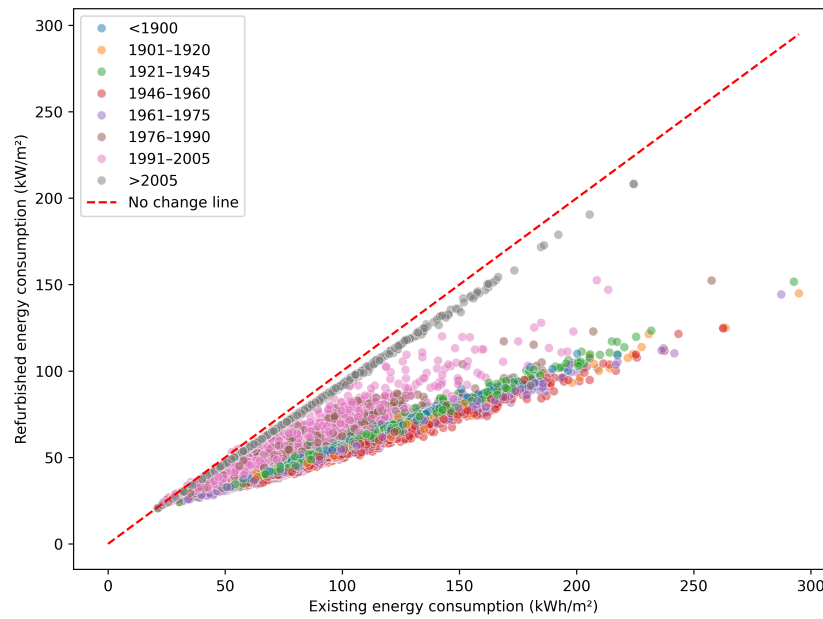


Figure 7. Energy consumption of buildings before (x-axis) and after (y-axis) standard refurbishment. Each point represents a building; points farther below the red line indicate greater refurbishment-induced savings.

For our purpose, the optimization problem consists of minimizing the following objectives:

- number of buildings refurbished;
- total energy consumption of the city.

The explored scenarios are based on combinations of refurbished buildings based on the construction periods. Specifically, in each scenario, all buildings from certain periods are retrofitted (i.e., 0% or 100% of buildings made in 1950).

Figure 8 shows the results of the refurbishment scenario selection. Each point represents a specific scenario: the red points describe optimal scenarios in terms of minimal number of buildings retrofitted and the total energy per m² used; the light blue points represent non-optimal scenarios.

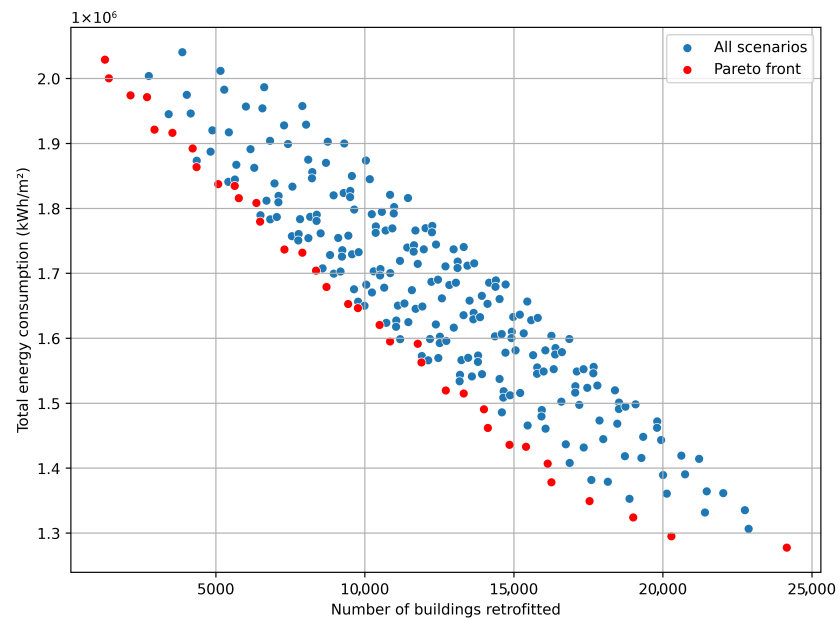


Figure 8. Pareto front for refurbishment scenarios selection. Red points denote Pareto-optimal scenarios minimizing both the number of retrofitted buildings and total energy use per m², while blue points represent non-optimal scenarios.

In Figure 9, each column of the binary map represents a scenario on the Pareto front. A scenario is defined as a combination of archetypes that have been retrofitted and archetypes that have been left unchanged. There is therefore a one-to-one correspondence between the red dots in Figure 8 and the columns of the matrix in Figure 9. Each row refers to a specific archetype: a blue cell indicates that the archetype in the corresponding row has been completely retrofitted in the scenario in the corresponding column. A white cell indicates that the corresponding archetype has not been modified in that scenario. This helps visually detect patterns, for example: “Buildings from 1946 to 1960 are almost always retrofitted in solutions that turned out to be optimal” or “>2005 buildings are never retrofitted in optimal scenarios”.

As evidenced from the binary map, the archetypes that are often included in the Pareto front are as follows: 1946–1960 (27 times), 1901–1920 (20 times), <1900 (19 times), and 1961–1975 (18 times). Buildings corresponding to these archetypes thus constitute the key priority areas for decision-makers. Figure 10 shows the energy intensity savings after applying the retrofitting scenario for archetype 1946–1960, the most prevalent one on the Pareto front.

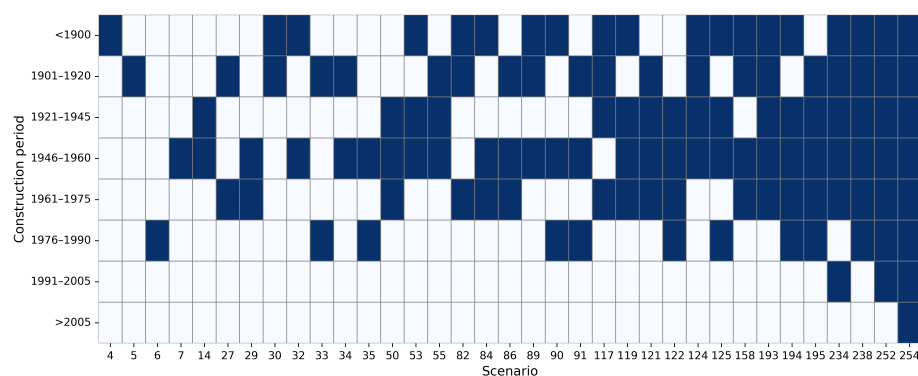


Figure 9. Binary representation of Pareto-optimal refurbishment scenarios. Columns correspond to individual scenarios and rows to building archetypes. Blue cells denote archetypes fully retrofitted in a scenario, while white cells indicate no retrofit.

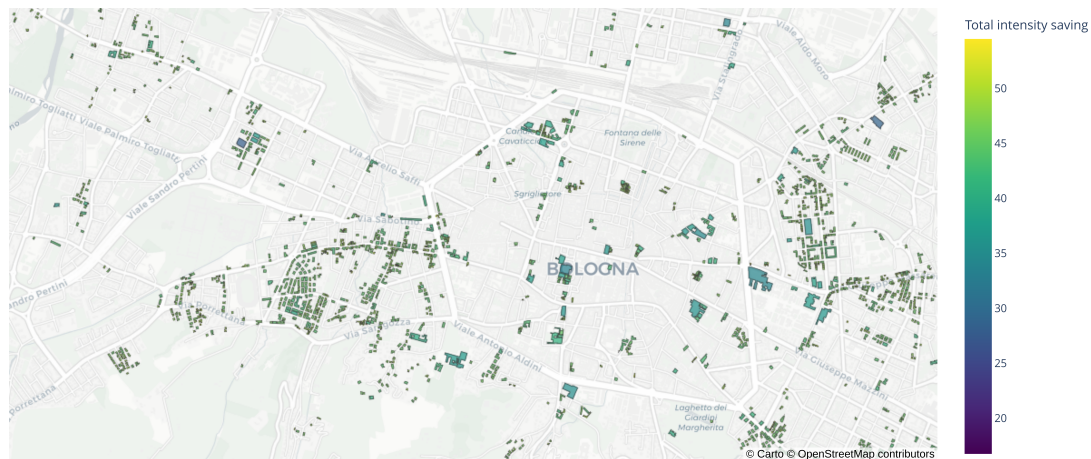


Figure 10. Energy-intensity savings under the retrofitting scenario for a portion of the “1946–1960” building archetype, which is the most frequent on the Pareto front. Administrative zone boundaries and street labels are shown in Italian for geographic reference.

5. Discussion

The use of geometric data, complemented with LiDAR information and TABULA archetypes, enabled the development of city-scale simulations starting from publicly accessible datasets. The HPC infrastructure proved effective in handling the computational load associated with large-scale simulations, providing a scalable solution that can be replicated in other urban contexts. However, several methodological considerations and limitations must be acknowledged, especially when compared with the broader UBEM literature.

A significant body of research has focused on reducing simulation time by introducing deliberate simplifications. Cerezo et al. (2016) [6] noted that computational constraints often require reductions in model detail; Dogan et al. (2016) [29] proposed geometry simplifications to increase tractability; Ferrando et al. (2022) [30] suggested simulating only representative buildings rather than entire stocks; and Zarrella et al. (2020) [31] developed simplified thermal models to reduce computation time. While these strategies do lower computational demands, they inevitably introduce approximations that may obscure urban-scale heterogeneity or alter predicted energy balances.

By contrast, the HPC approach adopted here enables simulation of the entire building stock—still within the inherent simplifications of UBEM but without additional compromises in model detail. HPC reduces simulation time without requiring further abstraction of geometry, physical properties, or shading interactions, thereby preserving the accuracy of building-level heterogeneity and supporting robust citywide scenario analysis. This represents a further development of the earlier UBEM literature, where computational limitations dictated model simplifications.

Despite methodological progress, true large-scale UBEM applications at full city scale remain rare, especially those relying exclusively on open data and operating in contexts with limited building information. Beyond early examples such as Boston [6], only a few recent works have approached similar scales: the EURECA platform [32] has been applied mainly to small- or medium-sized districts; open-data UBEM workflows published in Energies have been demonstrated across European municipalities but often with partial building coverage; and the Springer CLIMA 2025 [34] contribution integrates UBEM at neighborhood scale but not for entire cities. In this context, our work provides an example of a full city, open-data UBEM implementation applicable to typical European data environments. Nonetheless, the lack of detailed building-level metadata introduces limitations. Archetypes remain a necessary abstraction, but studies such as Cerezo et al. (2017) [47] and Oraopoulos et al. (2022) [48] show that archetype-based models can exhibit substantial

errors when heterogeneity in renovation history or envelope assemblies is large. This challenge is present in Bologna, where key information—such as refurbishment records or HVAC system details—is limited or absent in open datasets. The modular structure of the workflow also ensures that additional data or new renovation classes can be incorporated in the future.

A particularly critical limitation concerns occupancy modeling, one of the most influential yet uncertain elements in UBEM. Occupant schedules, internal gains, thermostat use, and ventilation behaviors strongly affect heating and cooling loads. However, due to GDPR and national privacy regulations, real occupancy patterns, smart-meter data, or appliance-load profiles cannot be collected or shared at building resolution. As noted by Johari et al. (2020) [4] and Hong et al. (2020) [5], this structural privacy barrier prevents the use of empirical occupancy datasets, forcing UBEM practitioners to rely on generalized or archetypal schedules. As a result, energy estimates may diverge from real behavior, particularly in mixed-use buildings or dwellings with atypical usage patterns.

Calibration poses another methodological barrier. As is widely emphasized in the literature, calibrated UBEMs exhibit substantially improved accuracy; yet the lack of anonymized or aggregated consumption data prevented direct calibration in our case. The surrogate comparison with TABULA values offers a partial consistency check but also highlights methodological differences between dynamic (EnergyPlus) and quasi-steady-state (UNI/TS 11300) procedures, consistent with findings by Corrado et al. (2016) [45] and Ballarini et al. (2018) [46].

Overall, this study confirms several themes highlighted in recent UBEM research:

- HPC can reshape the traditional trade-off between model detail and tractability, enabling full city simulations without introducing additional simplifications.
- Archetype-based UBEMs remain constrained by data availability, especially in open-data contexts.
- Occupancy and operational behavior represent persistent, privacy-driven uncertainties that cannot currently be resolved at scale.
- Full city UBEM applications remain uncommon, reinforcing the relevance of the proposed approach.
- The appendix includes practical scalability reference values that can support the adoption of similar workflows and help researchers to estimate the computing resources needed for simulations of similar scale.

6. Conclusions

This study demonstrates the feasibility and scalability of large-scale UBEM by integrating publicly available geospatial datasets, regulatory information, real-world archetypes from TABULA, and the computational capabilities of a HPC infrastructure. Through the combination of LiDAR-derived DSM/DTM, building footprints, volumetric data, and archetypal characteristics, we developed a robust workflow capable of simulating the energy performance of approximately 25,000 buildings in Bologna with a high level of detail. Powered by a parallel architecture based on Ray and the use of EnergyPlus for dynamic thermal simulation, the pipeline efficiently orchestrated thousands of independent simulations. The computation of the entire city required under 30 min, demonstrating that full city UBEM—traditionally considered extremely time consuming—can be executed rapidly when supported by modern HPC environments.

Compared with previous real city implementations, such as the Boston UBEM study [6], which required several days despite employing coarser resolution and more aggregated modeling steps, our approach achieves significantly shorter runtimes while preserving detailed, per-building modeling. Other recent attempts at large-scale UBEM,

including EURECA-based [32] implementations or open-data workflows applied to European municipalities [7,8,16], either reduce geometric fidelity or simulate only subsets of the building stock. In contrast, our method processes the entire city without relying on building selection or excessive simplifications, highlighting the advantages of HPC-driven UBEM in maintaining model richness without compromising computational feasibility.

The GIS-based nature of the workflow further enhances its practical value. The model directly produces georeferenced outputs that can be visualized as thematic maps, enabling spatial identification of patterns and hotspots, such as buildings with the highest energy intensities or retrofit needs. These maps provide actionable insights for municipal planners and policymakers, supporting prioritization of interventions and integration into digital twin frameworks. This spatial perspective—visible also in the figures included throughout the manuscript—shows how UBEM can inform energy strategies at the neighborhood and district scale, aligning technical modeling with real-world planning needs.

Finally, we would like to emphasize that this is currently a case study and there is no actual calibration; therefore, all analyses remain purely qualitative. Looking ahead, several directions can enhance the robustness and applicability of the workflow, including the following:

- Integrating anonymized real-world consumption data would enable model calibration and improve quantitative accuracy.
- Introducing temporal dynamics, including variable occupancy patterns, behavioral factors, and seasonal variations, would increase realism in operational energy predictions.
- Coupling this framework with additional urban layers—such as socio-economic indicators, mobility data, and urban microclimate models—would support more comprehensive and cross-sectoral energy planning.

Our results position this study among the few real-world, full-city UBEM applications leveraging HPC, and the reported runtimes offer a practical benchmark that can guide future research without necessitating extensive simulation testing.

In conclusion, this research presents a scalable, open-data-driven UBEM methodology that demonstrates how HPC resources can transform large-scale building energy simulation into a tractable, citywide analysis tool. The workflow provides a replicable blueprint for cities aiming to steer their building stock toward sustainable, energy-efficient, and climate-resilient futures.

Author Contributions: Conceptualization, F.B.; Methodology, A.C., E.B., R.M. and F.B.; Software, E.B. and R.M.; Validation, E.B., R.M. and F.B.; Formal analysis, E.B. and R.M.; Investigation, A.C., E.B. and R.M.; Data curation, E.B.; Writing—original draft, A.C.; Writing—review and editing, E.B. and R.M.; Visualization, A.C., E.B. and R.M.; Supervision, F.B., M.D.E., G.P. and C.D.; Project administration, M.D.E., G.P. and C.D. All authors have read and agreed to the published version of the manuscript.

Funding: CINECA Consorzio Interuniversitario.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The LiDAR data are hosted on the Bologna Digital Twin Data Platform <https://dashboard.bolognadigitaltwin.it/> (accessed on 2 December 2025) and are not publicly available. All other datasets can be accessed freely at the Open Data Comune di Bologna <https://opendata.comune.bologna.it/> (accessed on 2 December 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Simulation Performance Evaluation

Appendix A.1. Exploring the Impact of Shading Buffers and Cluster Size on Total Simulation Time

In dynamic energy simulations, the representation of shading interactions between constructions is crucial for reliably estimating a building's energy consumption, particularly for heating and cooling loads. EnergyPlus allows for the detailed modeling of shading geometries. For our purposes, we used the `Shading:Building:Detailed` object, which describes shading elements that are external to the building. Therefore, we modeled nearby buildings using their geoshapes and included them into the `.idf` file of the building to be simulated.

The choice of shading radius, which defines the spatial extent of nearby buildings included in the simulation, affects both the accuracy and complexity of energy simulation: a smaller radius may underestimate shading effects, while a larger one captures more accurate interactions, especially in dense urban areas. However, increasing the radius also raises the number of surfaces involved in ray-tracing and shadow projections computations, substantially increasing simulation times.

Our simulation pipeline therefore requires tuning of the shading radius parameter. Figure A1 presents the execution times associated with the simulation of the entire building stock while varying the number of nodes utilized for parallelization and adjusting the radius used to model surrounding buildings. As evidenced by the plot, an increase in the inclusion radius leads to an increase in total simulation time. This behavior manifests as a vertical translation along the time axis in curves corresponding to different radii.

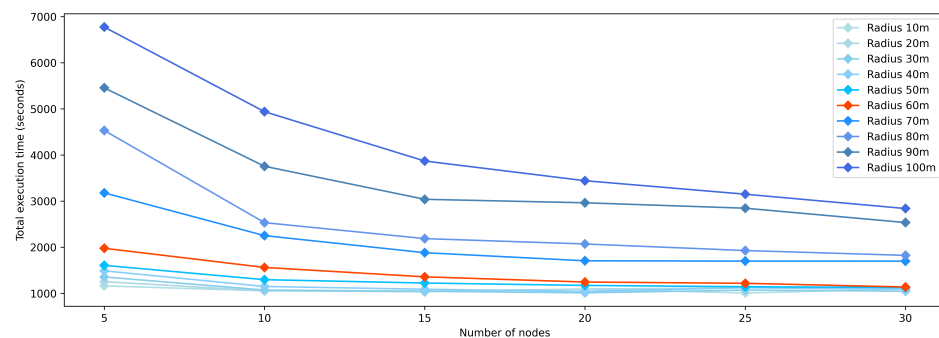


Figure A1. Total execution time variation reducing the shading radius and adding more nodes to the Ray cluster. Increasing the shading radius raises execution time due to more complexity in the simulation, whereas increasing the number of nodes enables more parallel simulations, reducing total computation time.

The total execution time associated with a given radius configuration can be reduced by increasing the number of simulations executed in parallel. This is accomplished by augmenting the pool of Ray workers through the addition of computational nodes. The Ray scheduler will then autonomously allocate the available CPU cores within the cluster to execute multiple simulations concurrently, thereby reducing the number of buildings remaining in queue waiting for resources. However, even when using a cluster with 24,537 CPU cores—matching the total number of buildings to be simulated and thus achieving perfect one-to-one parallelism—the overall execution time remains constrained by the duration of the longest individual building simulation. Consequently, each radius configuration is associated with a baseline execution time that cannot be further reduced through parallelization.

We evaluated the potential benefits of increased parallelization across different combinations of nodes and radii by subtracting the mean simulation time of the ten slowest-running buildings from the total simulation time. This provides an estimate of the time

savings achievable by executing a larger number of building simulations in parallel. The resulting speed-ups obtained by progressively increasing the number of worker nodes in the computing cluster are illustrated in Figure A2.

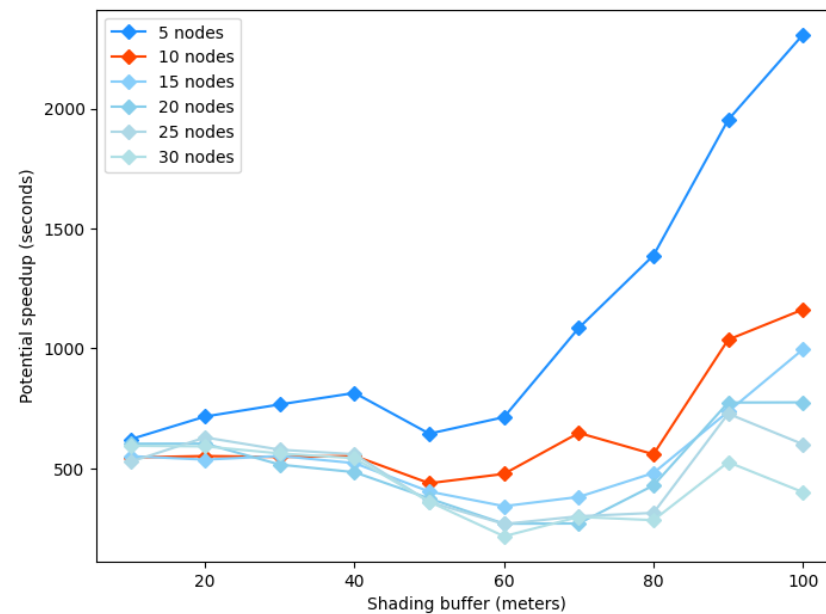


Figure A2. Potential speed-up obtainable by increasing the number of concurrent building simulations. Each line corresponds to a cluster configuration ranging from 5 nodes to 30 computing nodes. For each setup tested, we show the difference between the 10 longest running buildings simulations and the total execution time. A smaller gap indicates limited room for further improvement through additional parallelization.

As shown in Figure A2, increasing the number of nodes in the Ray cluster results in a downward shift along the vertical axis, as more buildings can be simulated concurrently, thereby reducing the overall execution time. The spread between the curves is also impacted by variations in buffer size: larger buffers require a greater number of shade projections to be computed, which increases the baseline execution time and consequently diminishes the relative gains achievable through additional parallelization.

With a cluster configuration of 10 computing nodes, the performance improvement over the previous node configuration is maximized. Although a further increase in the number of nodes could theoretically shorten the gap between the baseline and total execution times, the resulting gains are marginal when considering the additional resource consumption and the increase in scheduling and communication overhead. Specifically, only negligible reductions in execution time are observed, especially for radii configurations smaller than 60 m. This point represents a qualitative balance between increasing computational cost and only marginal gains in shading accuracy. Moreover, across the entire explored range (10 m to 100 m), the absolute variation in total simulated energy is approximately 1000 kWh, which corresponds to a difference of less than 1%. This suggests that the sensitivity of the UBEM results to the shading-radius parameter is relatively limited. For these reasons, the selection of 60 m should be interpreted as a practical compromise derived from the observed behavior of the simulations rather than as a climatologically grounded optimum. Consequently, the configuration with 10 nodes and 60 m buffer represents a balanced and robust option, achieving efficient utilization of computational resources while preserving sufficiently low total execution times.

Appendix A.2. Sensitivity of Energy Demand to Shading Buffers

The shadow buffer radius not only affects the simulation time but also the total simulated energy consumption. In this section, we investigate how varying the shading buffer radius affects simulated energy consumption.

The plot in Figure A3 displays the percentage change in average energy intensity (kWh/m^2) across all buildings as a function of the shading buffer radius. On the x-axis are the various shading buffers considered in the simulations. For each radius, the corresponding value on the y-axis represents the relative change in mean energy intensity compared to the simulation with the minimum shading radius of 10 m. This value is calculated using the following formula:

$$\left(\frac{E_r}{E_{min}} - 1 \right) \times 100 \quad (\text{A1})$$

where

- E_r is the mean energy intensity (kWh/m^2) at a given radius r .
- E_{min} is the mean energy intensity (kWh/m^2) at the smallest radius of 10 m.

This normalization allows us to evaluate how sensitive the energy demand is to changes in shading radius, independent of absolute values.

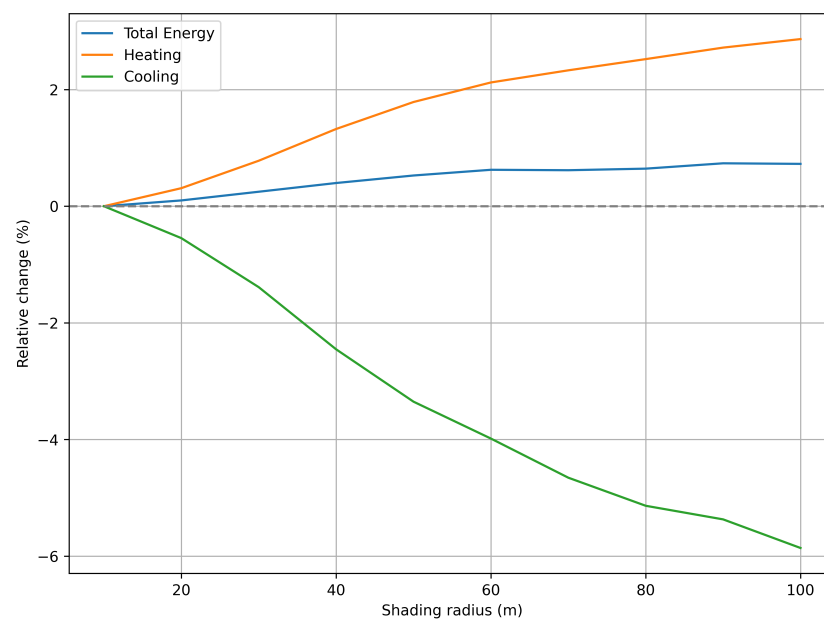


Figure A3. Relative change in mean energy intensity (kWh/m^2) across different shading radii. As the shading radius increases, the heating demand increases while the cooling demand decreases. Overall consumption increases but remains low (below 1%), indicating a limited impact of shading radius on total consumption. The dashed horizontal line at 0% represents the baseline, indicating no relative change compared to the 10 m scenario.

The results show that as the shading radius increases,

- Heating demand increases by up to +3%, due to reduced solar heat gains caused by the additional shading from surrounding buildings. This indicates that larger buffers limit passive solar heating, especially in winter.
- Cooling demand decreases by up to -6% , as increased shading helps mitigate solar overheating during warmer months.
- The overall energy intensity change is positive, reflecting the net effect of increased heating demand outweighing the cooling energy savings in the simulated climate.

However, the percentage differences across radii are relatively small (e.g., within $\pm 5\text{--}6\%$). Therefore, in addition to the considerations in the previous section, a mid-range radius provides a good balance between simulation accuracy and computational cost. Moreover, if we specifically analyze the average total energy as a function of the shading radius, as shown in Figure A4, we can observe an “elbow” around a radius of 60 m. Overall, the total energy demand exhibits only moderate variation across different radii.

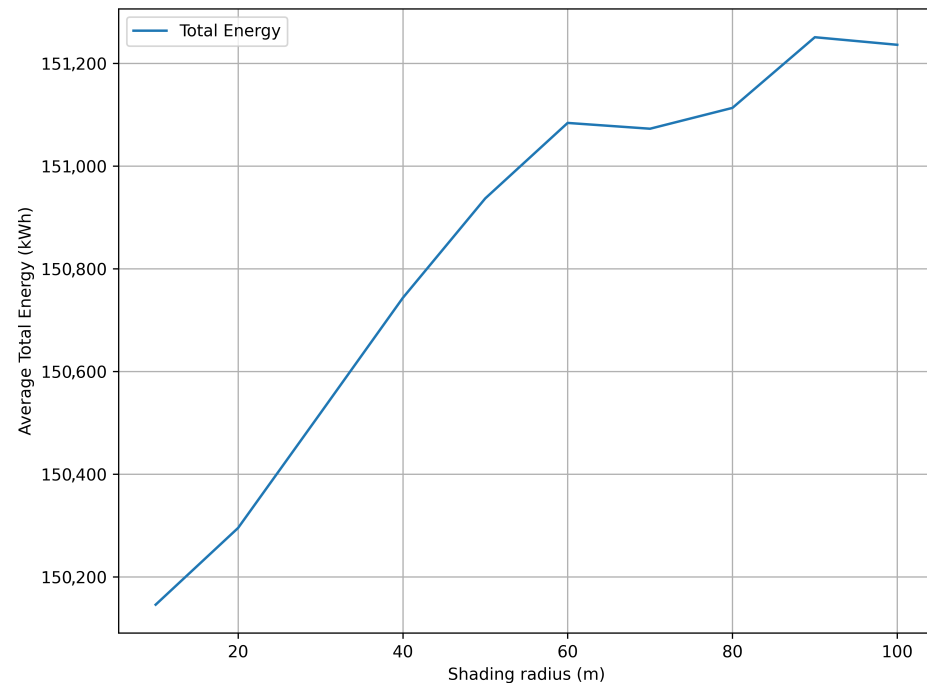


Figure A4. Average total energy as a function of the shading radius. As the shading radius increases, so does the total energy consumption. However, once the radius exceeds 60 m, the percentage change is less than 1%.

Appendix B. Urban Energy Profile Assessment: Analysis of Simulated Energy Consumption for the Current Building Stock

This section presents the results obtained by simulating the energy consumption of the existing building stock. To better understand the diversity in building energy consumption, we first explore the distribution of total energy use per square meter (kWh/m^2) grouped by the archetypes. This allows us to assess how energy needs vary with building characteristics.

As shown in Figure A5, the histograms reveal differences in energy intensity (kWh/m^2) across buildings from different construction periods. Older buildings tend to show broader and right-skewed distributions, indicating a higher variability and, generally, greater energy demands. This is likely due to poorer thermal performance, limited insulation, and outdated envelope standards typical of early 20th-century construction. In contrast, more recent buildings, particularly those built after 1975, show narrower and left-shifted distributions, reflecting improved energy performance likely due to stricter building codes and more efficient construction techniques. The progressive shift toward lower energy intensity over time highlights the impact of energy regulations and retrofitting practices on reducing energy demand per square meter.

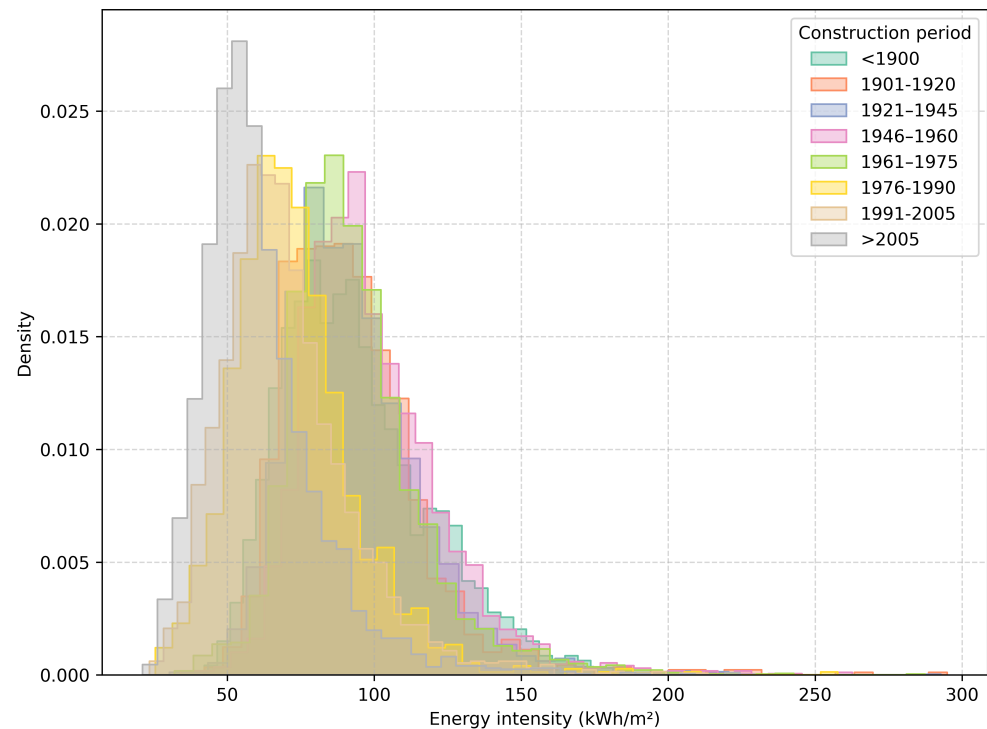


Figure A5. Energy intensity distribution by period of construction.

Pareto analysis helps to identify the most significant factors contributing to an outcome, allowing for prioritized actions where they matter most. In the context of building energy consumption, Pareto analysis is often used to identify which buildings or zones consume the most energy, allowing retrofits and policy interventions to be focused on high-impact areas. Typically, about 80% of the effects stem from 20% of the causes. While our situation yields different results, they can still be useful for targeted action.

The Pareto plot in Figure A6 reveals that almost 70% of buildings account for 80% of the total energy consumption in terms of total energy consumption. This demonstrates a distribution flatter than expected in typical urban energy datasets and suggests that high energy consumption is not highly concentrated in a small subset of buildings, but rather widely distributed across much of the stock. As a result, energy efficiency policies must be broadly targeted, instead of focusing only on a few outliers.

Figure A7 shows that most of the buildings with the highest energy consumption obtained with the Pareto analysis are the oldest ones or were built in the middle of the past century, when construction laws were less restrictive in terms of efficiency.

After simulating the building's energy consumption, it is possible to estimate its CO₂ emissions. For this purpose, we can compute an approximation using the emission factor, which is a conversion factor that depends on the fuel used to generate energy. This CO₂ estimate does not consider other greenhouse gases; therefore, we cannot talk about the equivalent CO₂, but it is a good approximation. We used the emission factor provided by ISPRA in document [49] for the year 2023, which is 59.182 tCO₂/TJ, and obtained an average CO₂ emission of 12.2 kg/m². This value likely underestimates the actual value because it only considers CO₂ emissions from gas consumption, not from electricity or other greenhouse gases.

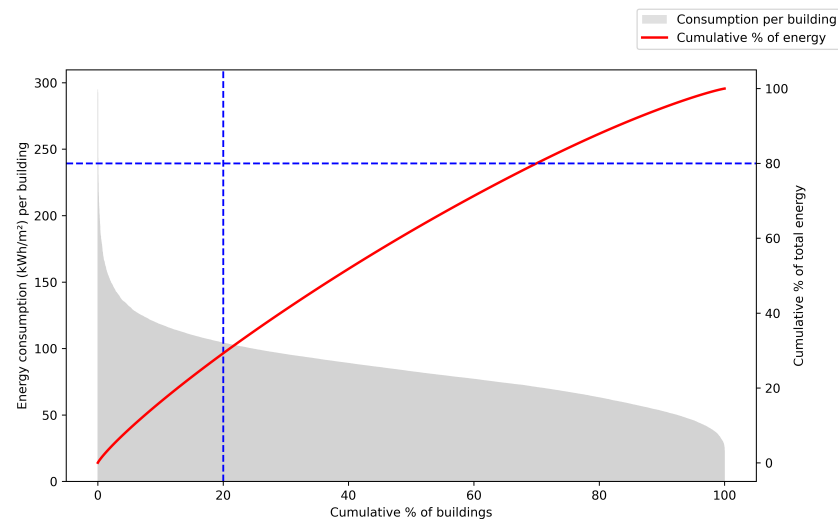


Figure A6. Pareto diagram of building energy intensity consumption for the entire building stock simulated. Gray bars show individual building energy intensity ordered from highest to lowest, while the red curve represents the cumulative percentage of total energy consumption. Dashed blue reference lines illustrate the 80/20 principle, highlighting the fraction of buildings responsible for the majority of energy consumption.

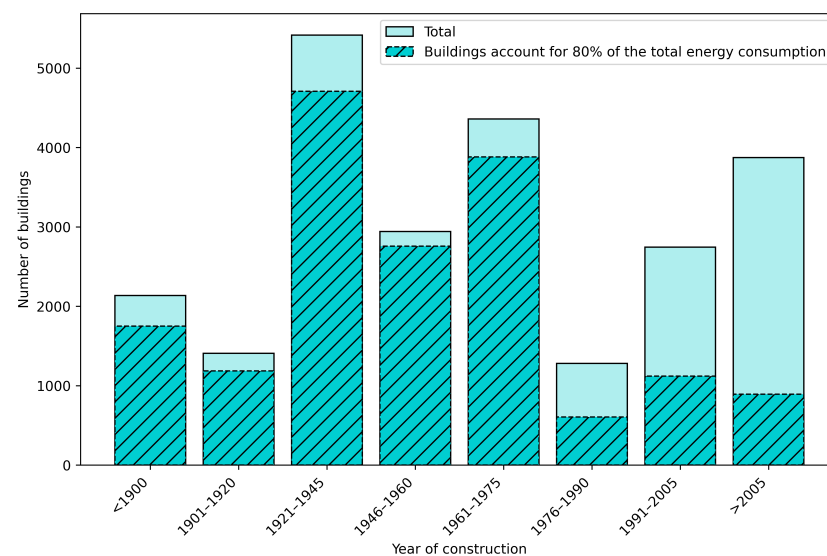


Figure A7. Histogram with the archetype distribution for the 70% of buildings (dashed part of the bars) that account for 80% of total energy consumption. The histogram shows that proportionally the majority of buildings constructed in the mid-20th century are in the group of the high-impact buildings.

References

1. Agency, I.E. Buildings Sector Overview. Available online: <https://www.iea.org/energy-system/buildings> (accessed on 2 December 2025).
2. Chen, Y.; Hong, T.; Piette, M.A. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Appl. Energy* **2017**, *205*, 323–335. [CrossRef]
3. Ferrando, M.; Causone, F.; Hong, T.; Chen, Y. Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. *Sustain. Cities Soc.* **2020**, *62*, 102408. [CrossRef]
4. Johari, F.; Peronato, G.; Sadeghian, P.; Zhao, X.; Widén, J. Urban building energy modeling: State of the art and future prospects. *Renew. Sustain. Energy Rev.* **2020**, *128*, 109902. [CrossRef]

5. Hong, T.; Chen, Y.; Luo, X.; Luo, N.; Lee, S.H. Ten questions on urban building energy modeling. *Build. Environ.* **2020**, *168*, 106508. [[CrossRef](#)]
6. Cerezo Davila, C.; Reinhart, C.F.; Bemis, J.L. Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. *Energy* **2016**, *117*, 237–250. [[CrossRef](#)]
7. Mutani, G.; Todeschi, V.; Beltramino, S. Energy Consumption Models at Urban Scale to Measure Energy Resilience. *Sustainability* **2020**, *12*, 5678. [[CrossRef](#)]
8. Buckley, N.; Mills, G.; Reinhart, C.; Berzolla, Z.M. Using urban building energy modelling (UBEM) to support the new European Union’s Green Deal: Case study of Dublin Ireland. *Energy Build.* **2021**, *247*, 111115. [[CrossRef](#)]
9. Deng, Z.; Javanroodi, K.; Nik, V.M.; Chen, Y. Using urban building energy modeling to quantify the energy performance of residential buildings under climate change. *Build. Simul.* **2023**, *16*, 1629–1643. [[CrossRef](#)]
10. Garcia-Navado, E.; Lopez-Besora, J.; Besuievsky, G. Including Open Balconies in Housing Retrofitting: A Parametric Analysis for Energy Efficiency. *Urban Sci.* **2025**, *9*, 439. [[CrossRef](#)]
11. Espino-Reyes, C.A.; Ortega-Avila, N.; Lucero-Álvarez, J.; Rodríguez-Muñoz, N.A. Energy Consumption in Mexican Homes: Using a Reference Building as a Launchpad for Achieving Nearly Zero Energy. *Urban Sci.* **2025**, *9*, 113. [[CrossRef](#)]
12. Fouquier, A.; Robert, S.; Suard, F.; Stéphan, L.; Jay, A. State of the art in building modelling and energy performances prediction: A review. *Renew. Sustain. Energy Rev.* **2013**, *23*, 272–288. [[CrossRef](#)]
13. Reinhart, C.F.; Cerezo Davila, C. Urban building energy modeling—A review of a nascent field. *Build. Environ.* **2016**, *97*, 196–202. [[CrossRef](#)]
14. Ali, U.; Shamsi, M.H.; Hoare, C.; Mangina, E.; O’Donnell, J. Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. *Energy Build.* **2021**, *246*, 111073. [[CrossRef](#)]
15. Kontokosta, C.E.; Tull, C. A data-driven predictive model of city-scale energy use in buildings. *Appl. Energy* **2017**, *197*, 303–317. [[CrossRef](#)]
16. Todeschi, V.; Boghetti, R.; Kämpf, J.H.; Mutani, G. Evaluation of Urban-Scale Building Energy-Use Models and Tools—Application for the City of Fribourg, Switzerland. *Sustainability* **2021**, *13*, 1595. [[CrossRef](#)]
17. Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1027–1047. [[CrossRef](#)]
18. Blanco, L.; Alhamwi, A.; Schiricke, B.; Hoffschmidt, B. Data-driven classification of Urban Energy Units for district-level heating and electricity demand analysis. *Sustain. Cities Soc.* **2024**, *101*, 105075. [[CrossRef](#)]
19. Salim, M.H.; Schlünzen, K.H.; Grawe, D.; Boettcher, M.; Gierisch, A.M.U.; Fock, B.H. The microscale obstacle-resolving meteorological model MITRAS v2.0: Model theory. *Geosci. Model Dev.* **2018**, *11*, 3427–3445. [[CrossRef](#)]
20. Maronga, B.; Gross, G.; Raasch, S.; Banzhaf, S.; Forkel, R.; Heldens, W.; Kanani-Sühring, F.; Matzarakis, A.; Mauder, M.; Pavlik, D.; et al. Development of a new urban climate model based on the model PALM ? Project overview, planned work, and first achievements. *Meteorol. Z.* **2019**, *28*, 105–119. [[CrossRef](#)]
21. Jin, L.; Schubert, S.; Fenner, D.; Salim, M.H.; Schneider, C. Estimation of mean radiant temperature in cities using an urban parameterization and building energy model within a mesoscale atmospheric model. *Meteorol. Z.* **2022**, *31*, 31–52. [[CrossRef](#)]
22. Malhotra, A.; Bischof, J.; Nichersu, A.; Häfele, K.H.; Exenberger, J.; Sood, D.; Allan, J.; Frisch, J.; van Treeck, C.; O’Donnell, J.; et al. Information modelling for urban building energy simulation—A taxonomic review. *Build. Environ.* **2022**, *208*, 108552. [[CrossRef](#)]
23. Hong, T.; Chen, Y.; Lee, S.H.; Piette, M. CityBES: A Web-based Platform to Support City-Scale Building Energy Efficiency. In Proceedings of the 5th International Urban Computing Workshop, San Francisco, CA, USA, 13 August 2016.
24. Fonseca, J.A.; Nguyen, T.A.; Schlueter, A.; Marechal, F. City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts. *Energy Build.* **2016**, *113*, 202–226. [[CrossRef](#)]
25. Kasmeridis, I.K.; Skandalos, N.; Dimitriou, T.; Dimakopoulos, V.V.; Karamanis, D. Digitized Energy Systems and Open-Access Platforms: Accelerating Cities’ Transition to Carbon Neutrality. *Urban Sci.* **2025**, *9*, 364. [[CrossRef](#)]
26. Goy, S.; Maréchal, F.; Finn, D. Data for Urban Scale Building Energy Modelling: Assessing Impacts and Overcoming Availability Challenges. *Energies* **2020**, *13*, 4244. [[CrossRef](#)]
27. Chen, Y.; Deng, Z.; Hong, T. Automatic and rapid calibration of urban building energy models by learning from energy performance database. *Appl. Energy* **2020**, *277*, 115584. [[CrossRef](#)]
28. Ferrando, M.; Causone, F.; Banfi, A.; Corrado, V.; Ballarini, I.; Piro, M.; Zarrella, A.; Carnieletto, L.; Borgato, N.; Evola, G.; et al. Developing an Italian Library of Reference Buildings for Urban Building Energy Modeling (UBEM): Lessons Learnt from the URBEM Project. *Energies* **2025**, *18*, 6026. [[CrossRef](#)]
29. Dogan, T.; Reinhart, C.; Michalatos, P. Autozoner: An algorithm for automatic thermal zoning of buildings with unknown interior space definitions. *J. Build. Perform. Simul.* **2016**, *9*, 176–189. [[CrossRef](#)]
30. Ferrando, M.; Ferroni, S.; Pelle, M.; Tatti, A.; Erba, S.; Shi, X.; Causone, F. UBEM’s archetypes improvement via data-driven occupant-related schedules randomly distributed and their impact assessment. *Sustain. Cities Soc.* **2022**, *87*, 104164. [[CrossRef](#)]

31. Zarrella, A.; Prativiera, E.; Romano, P.; Carnieletto, L.; Vivian, J. Analysis and application of a lumped-capacitance model for urban building energy modelling. *Sustain. Cities Soc.* **2020**, *63*, 102450. [CrossRef]
32. Prativiera, E.; Romano, P.; Carnieletto, L.; Pirotti, F.; Vivian, J.; Zarrella, A. EURECA: An open-source urban building energy modelling tool for the efficient evaluation of cities energy demand. *Renew. Energy* **2021**, *173*, 544–560. [CrossRef]
33. Nardelli, C.; Colombo, R.; Banfi, A.; Ferrando, M.; Shi, X.; Causone, F. A Comparative Analysis of Two Urban Building Energy Modelling Tools via the Case Study of an Italian Neighbourhood. *Energies* **2025**, *18*, 2618. [CrossRef]
34. Banfi, A.; Nardelli, C.; Ferrando, M.; Shi, X.; Causone, F. Using UBEM to Study Decarbonisation Strategies of Urban Neighbourhoods: A Case Study in Milan. In Proceedings of the 15th REHVA HVAC World Congress—CLIMA 2025, Milan, Italy, 4–6 June 2025; pp. 823–833, ISBN 978-3-032-06806-4.
35. U.S. Department of Energy. *EnergyPlus*, Version 9.4.0; National Renewable Energy Laboratory: Golden, CO, USA, 1997. Available online: <https://energyplus.net> (accessed on 2 December 2025).
36. Reinhart, C.; Dogan, T.; Jakubiec, J.; Rakha, T.; Sang, A. Umi—An Urban Simulation Environment For Building Energy Use, Daylighting And Walkability. In Proceedings of the 13th Conference of International Building Performance Simulation Association, Chambéry, France, 26–28 August 2013. [CrossRef]
37. Municipality of Bologna. Open Data Comune Di Bologna. Available online: <https://opendata.comune.bologna.it/pages/home/> (accessed on 10 December 2025).
38. Corrado, V.; Ballarini, I.; Corgnati, S.; Talà, N. *Building Typology Brochure-Italy Fascicolo Sulla Tipologia Edilizia Italiana-Tabula eU Project*; Politecnico di Torino: Turin, Italy, 2011; pp. 1–117, ISBN 978-88-8202-065-1.
39. TABULA WebTool. TABULA WebTool Building Typologies. Available online: <https://webtool.building-typology.eu/> (accessed on 10 December 2025).
40. *UNI 10351:2015*; Materiali e Prodotti per Edilizia—Proprietà Termoigrometriche—Procedura per la Scelta dei Valori di Progetto. Ente Italiano di Normazione: Milano, Italy, 2015.
41. *ISO 6946:2007*; Building Components and Building Elements—Thermal Resistance and Thermal Transmittance—Calculation Method. International Organization for Standardization: Geneva, Switzerland, 2007.
42. *UNI/TR 11552:2014*; Abaco Delle Strutture Costituenti l’involucro Opaco Degli Edifici—Parametri Termofisici. Ente Italiano di Normazione: Milano, Italy, 2014.
43. Climate.OneBuilding.Org WMO Region 6—Europe. Available online: https://climate.onebuilding.org/WMO_Region_6_Europe/ITA_Italy/index.html (accessed on 10 December 2025).
44. *UNI/TS 11300-1:2014*; Prestazioni Energetiche Degli Edifici—Parte 1: Determinazione del Fabbisogno di Energia Termica Dell’edificio per la Climatizzazione Estiva ed Invernale. Ente Italiano di Normazione: Milano, Italy, 2014.
45. Corrado, V.; Ballarini, I.; Dirutigliano, D.; Murano, G. Verification of the New Ministerial Decree about Minimum Requirements for the Energy Performance of Buildings. *Energy Procedia* **2016**, *101*, 200–207. [CrossRef]
46. Ballarini, I.; Primo, E.; Corrado, V. On the limits of the quasi-steady-state method to predict the energy performance of low-energy buildings. *Therm. Sci.* **2018**, *22*, 1117–1127. [CrossRef]
47. Cerezo, C.; Sokol, J.; AlKhaled, S.; Reinhart, C.; Al-Mumin, A.; Hajiah, A. Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City. *Energy Build.* **2017**, *154*, 321–334. [CrossRef]
48. Oraipoulos, A.; Howard, B. On the accuracy of Urban Building Energy Modelling. *Renew. Sustain. Energy Rev.* **2022**, *158*, 111976. [CrossRef]
49. Romano, D.; Bernetti, A.; Caputo, A.; Cordella, M.; Lauretis, R.D.; Cristofaro, E.D.; Fiore, A.; Gagna, A.; Gonella, B.; Moricci, F.; et al. Italian Greenhouse Gas Inventory 1990–2022 National Inventory Document 2024. Annual Report for Submission Under the UNFCCC and the Paris Agreement. Available online: <https://unfccc.int/sites/default/files/resource/National%20Inventory%20Document%202024%20-%20Italy.pdf> (accessed on 2 December 2025).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.