POLITECNICO DI TORINO
Repository ISTITUZIONALE

Urban Building Energy Modeling: A Comparative Study of Process-Driven and Data-Driven Models

Original
Urban Building Energy Modeling: A Comparative Study of Process-Driven and Data-Driven Models / Montazeri, A., Usta, Y., Mutani, G.. - In: MATHEMATICAL MODELLING OF ENGINEERING PROBLEMS. - ISSN 2369-0739. - ELETTRONICO. - 11:10(2024), pp. 2615-2624. [10.18280/mmep.111003]

Availability:
This version is available at: 11583/2994030 since: 2024-10-31T12:51:23Z

Publisher:
IIETA

Published
DOI:10.18280/mmep.111003

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

25 November 2026

IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 11, No. 10, October, 2024, pp. 2615-2624
Journal homepage: http://iieta.org/journals/mmep

Urban Building Energy Modeling: A Comparative Study of Process-Driven and Data-Driven Models

Ahad Montazeri, Yasemin Usta, Guglielmina Mutani*

Department of Energy, Politecnico di Torino, Torino 10129, Italy

Corresponding Author Email: Guglielmina.Mutani@Polito.it

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/mmep.111003

Received: 20 August 2024
Revised: 3 October 2024
Accepted: 9 October 2024
Available online: 31 October 2024

Keywords:
urban building energy modeling, process-driven models, data-driven models, machine learning (ML), placed-based approach

ABSTRACT

This study investigates the predictive capabilities of process-driven (PD) energy modeling and Machine Learning techniques, specifically Light Gradient Boosting Machine (LGBM) and Random Forest (RF) algorithms, in analyzing building energy consumption patterns. Leveraging a comprehensive dataset encompassing diverse building characteristics, energy-related variables, and operational configurations, the comparative performances of these methodologies is explored. Results reveal that while all approaches demonstrate promising predictive accuracies, LGBM exhibits a slight advantage over RF and the process-driven model. Moreover, the process-driven model showcases efficacy in colder seasons and for buildings of extreme ages, while encountering limitations in accurately modeling energy consumption for structures constructed during 1970s to 1990s. Conversely, Machine Learning models demonstrate consistent performance (with relative errors of 5-10%) across varied building ages, underscoring their adaptability and potential for capturing nuanced energy dynamics. However, a notable constraint lies in the availability of sufficient data for training Machine Learning models, posing challenges for model testing. These findings contribute to advancing our understanding of energy modeling methodologies at urban scale and offer insights for optimizing building energy efficiency strategies for a sustainable development of urban environments.

1. INTRODUCTION

The present global urban population accounts for 57%, with a foreseen rise of 15% by 2050 according to United Nations projections [1]. Moreover, cities are responsible for 70% of worldwide greenhouse gas (GHG) emissions and consume 60-80% of global energy [2]. Hence, urban areas must intensify their efforts to meet climate neutrality objectives [3].

Urban areas, despite facing challenges such as limited renewable energy adoption and retrofit constraints, are very important players in the clean energy transition. Cities need to reduce their dependence on fossil fuels and shift towards Renewable Energy Resources (RES) to mitigate the use of finite resources and their negative impacts on the environment. Besides, residents must adopt energy-efficient practices that save energy [4]. The establishment of a sustainable urban energy framework requires both decreasing end-user energy consumption and integrating efficient, sustainable energy generation within urban environments.

Within the EU, buildings over 50 years old contribute to 35% of the building sector and overall, 75% of buildings are not energy efficient. Cultural and historical buildings (more specifically those with higher protection priority), should not be included in the poll for energy efficiency interventions. Preserving these structures poses challenges in implementing RES and energy-saving measures, resulting in only 5-6% of energy savings and 5% of GHG emissions reduction. These percentages can be meaningful considering that new buildings, constructed according to modern energy efficiency standards, need 50% less energy compared to the buildings of the 1980s [5].

Though the objectives for reducing GHG emissions are typically established nationally, significant efforts are required at the city level. This is primarily due to cities' access to extensive energy consumption data, which aids in identifying economically feasible options for improving energy efficiency and consequently reducing GHG emissions within the city scale. To effectively manage and reduce building energy consumption and GHG emissions, it is essential to comprehend not just the current state of building energy use, but also its historical patterns and future projections [6].

City-scale energy modeling of buildings primarily aims to assess the energy performance of buildings within an urban context across various spatial-temporal resolutions. Additionally, these models serve as valuable tools for urban planning and the development of both existing and planned areas. By understanding the daily and seasonal energy consumption patterns across different locations within a city, authorities gain deeper insights into balancing energy supply and demand, thereby mitigating potential instabilities and shortages in the energy system. Furthermore, these models facilitate scenario planning and benchmarking for building

retrofits and the integration of renewable energy solutions into urban energy systems [7]. Consequently, numerous energy modeling approaches have emerged in recent decades to analyze energy flows of buildings within the urban environment and, ultimately, to identify the energy supply sources at the city scale [6].

Addressing the challenges posed by urbanization, climate change, and energy conservation demands immediate action through the implementation of robust policies and design criteria to establish sustainable energy systems in urban areas. To close the gap between current practices and a more sustainable urban future, there is a need for innovative and tailored approaches, by using modeling to make buildings and cities more energy-efficient and clean. City-scale energy modeling of buildings serves as an effective tool for informing stakeholders, city planners, and decision-makers about urban energy systems, enable them to develop energy strategies, propose sustainable initiatives, and implement constructive policies [7].

Following the statements before, this paper aims to thoroughly investigate the most known methodologies utilized in urban-scale building energy modeling. It involves a detailed examination of both process-driven and data-driven approaches, with the goal of clarifying their complex mechanisms and real-world uses. Our analysis goes beyond simple description, instead offering a critical evaluation of the main goals, inherent strengths, and notable limitations of these methods. Additionally, we explore recent advancements in the field, highlighting emerging trends and promising developments that drive its continuous evolution.

After the literature review in section 2, section 3 describes process-driven and data-driven modeling; the case study of the residential buildings of the city of Turin is presented in section 4 and section 5 illustrates the results and the comparison of the energy models.

2. LITERATURE REVIEW AND RESEARCH GAP

Given the fact that cities and buildings play a pivotal role in the reduction of global energy consumptions and carbon emissions, it is essential to explore approaches and procedures that can aid to comprehend the current and future energy consumptions at various scales. Urban Building Energy Modeling (UBEM) offers a versatile approach with its diverse predictive engines, including process-driven, data-driven, and hybrid models, that enables researchers to conduct detailed analyses of energy consumptions from individual buildings to an urban scale.

However, the emphasis in demand-side energy studies has typically been on building simulation and physical models of building technologies, rather than on city-scale empirical models; a relatively lower attention has been given to emerging data-driven studies of urban energy dynamics [6]. Moreover, comparative analyses between process-driven and data-driven models have also been rare, hindering efforts to assess the scope and accuracy of these models in predicting urban energy consumption.

For instance, a work by Li et al. [6] seeks to offer an overview of energy modeling categories applicable to urban buildings, outlining the fundamental process of physics-based, bottom-up models and their role in simulating urban-scale building energy consumption. Additionally, it presents an evaluation of the strengths and weaknesses of these models. Subsequently, the paper delves into the complexities surrounding model preparation and calibration, addressing associated challenges.

In 2020, Chen et al. [8] conducted a comprehensive literature review focusing on building energy prediction models. They categorize and introduce three commonly utilized prediction approaches: building physical energy models (referred to as white box models), data-driven models (known as black box models), and hybrid models (referred to as grey box models). Their review delves into the principles, advantages, limitations, and practical applications of each model. Drawing from this examination, they underscore research priorities and outline future directions in the field of building energy prediction.

In 2020 Mutani et al. [9] introduced a dynamic urban-scale energy model founded on an energy balance approach, tailored to incorporate local climate conditions and morphological parameters at the urban scale. The objective was to introduce an engineering methodology applicable to clusters of buildings, leveraging existing urban databases. The proposed method effectively handles diverse data types across different scales, providing precise spatial-temporal insights into building energy performance. While detailed heat balance methods are typically employed at the building level to estimate heating loads, the urban-scale model serves as a decision support tool for urban design investigations and policymaking Additionally, initial estimates of constant indoor temperatures were refined by correlating them with climatic variables and improving the accuracy of the model.

In 2022, Todeschi et al. [10] integrated a hybrid model to describe the energy consumption of buildings in Geneva. In this study, process-driven modeling was initially employed; however, its precision was enhanced through the integration of adjustments obtained from the RF algorithm. The modeling was also used to evaluate the different use of consumption during the Covid-19 Pandemic.

In 2017, Kontokosta and Tull [11] developed a predictive model of energy use at the building, district, and city scales using training data from energy disclosure policies and predictors from available property and zoning information. They employed statistical models to predict the energy use of buildings in New York City, leveraging the physical, spatial, and energy use attributes of a subset derived from buildings. In this work, they fitted Ordinary least Square (OLS), RF, and support vector regression (SVM) algorithms to the city's energy benchmarking data and subsequently utilized to predict electricity and natural gas use for every property in the city. Model accuracy was assessed and validated at the building level and zip code level using actual consumption data from calendar year 2014.

Boghetti et al. [12] presented two distinct models that utilize morphological parameters at the urban scale to enhance their performance, considering the interactions between buildings and their surroundings. In this study, for the two models several urban parameters were extracted and utilized as input alongside building-scale features. Their first model adopted a bottom-up engineering approach to assess the energy balance of residential buildings, incorporating variables at the block-of-buildings scale. Then in the second model they employed a machine learning approach based on the bootstrap aggregating (bagging) algorithm, utilizing the same parameters as inputs to estimate the hourly energy consumption of each building.

Following the investigated literature of the field, the research gap lies in the limited number of the studies that

directly compare data-driven and process-driven approaches for urban building energy modeling across various time steps and spatial scales. Additionally, existing studies may use different Machine Learning algorithms, making it open to explore other algorithms suitable for the needs of the literature to understand which algorithm best fits for the energy studies at urban scale.

By addressing these gaps, the aims of the paper are to provide valuable insights into the relative performance of these modeling approaches across different time intervals (e.g., hourly, daily, monthly) and spatial scales (e.g., individual buildings, neighborhoods, city), to identify the strengths and weaknesses of each modeling approach in accurately predicting energy consumption patterns, considering multiple factors like climatic, building's geometric and non-geometric parameters, and inform future research and practical applications in urban energy management. Energy performance assessments of buildings need to complain about recent regulations to reduce energy consumptions. According to the Decree 383 of 6/10/2022, in Turin (in the Italian climate zone E) space-heating can operate between 22 October and 7 April, with a maximum of 13 hours per day and with an internal temperature of 19+2°C [13].

3. PROCESS-DRIVEN AND DATA-DRIVEN MODELS FOR BUILDINGS AT URBAN SCALE

This section outlines the implementation of two distinct methodologies for developing building energy models on an urban scale. In Figure 1 the methodology of this research is described. After a first phase of data collection with GIS [14] which can process massive amount of data using new methods to generate accurate predictions [15], a geo-database was created containing all information about the buildings and the surrounding context (that influence the heat exchange with the buildings).

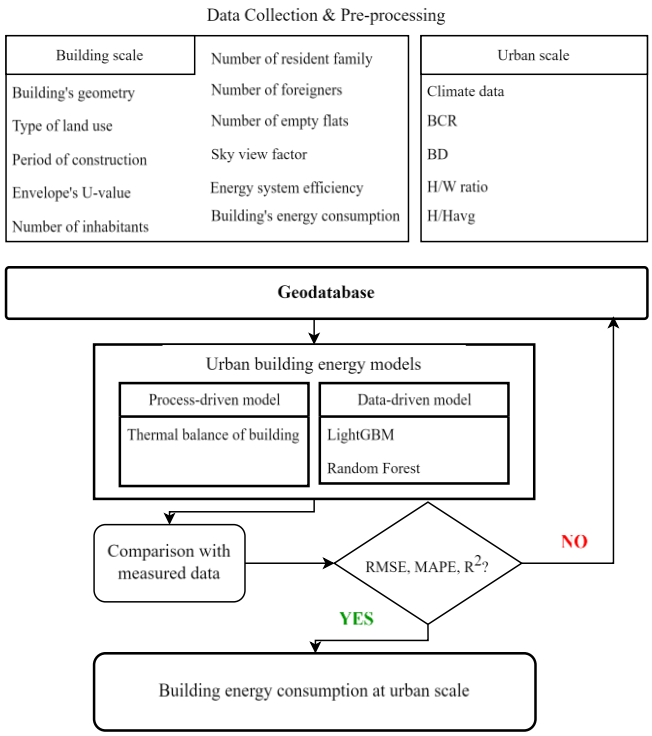


Figure 1. Flow chart of the methodology

About the energy modeling, the initial approach involves a process-driven model based on hourly thermal balance calculations. Conversely, the second approach employs data-driven modeling utilizing two Machine Learning algorithms: a bagging algorithm (Random Forest) and a gradient boosting algorithm (LightGBM). Currently, Machine learning algorithms are widely used for urban scale simulations [16].

Taking cue from Mutani et al. [9] and adhering to ISO standards 52016-1:2017 and ISO 52017-1:2017, the energy balance of for each building using three thermodynamic systems were examined. This UBEM is based on several assumptions: the use of block-scale variables to replace parameters that are not known at the building scale, uniform temperatures within thermodynamic systems, one-dimensional heat transmission through building elements, and latent energy exchanges are neglected (as humidification is not controlled by the heating systems).

Three thermodynamic systems used to describe the heat fluxes between each building and the outside environment are:

The opaque envelope (E), encompassing all opaque components dividing the internal heated volume from external or unheated spaces.

The glazing (G), encompassing transparent components separating the internal heated volume from external or unheated spaces.

The internal part of the building (B), encompassing internal structures, furnishings, peoples, appliances, and air.

For each building, the general energy balance equation was obtained using data about buildings' geometry, urban environment, and climate condition. The general equation (Eq. (1)) for the three thermodynamic systems TS (i.e., G, E, and B) considers:

$$C_{TS}\frac{dT_{TS}}{dt} = \emptyset_{sol} + \emptyset_I + \emptyset_H - (\emptyset_T + \emptyset_V) \quad (1)$$

where, for each thermodynamic system (TS), the variables represent: C is heat capacity (JK^{-1}), T is the temperature of TS, t is the time (s), $\emptyset_{sol}$ is the heat flow rate from solar gains (W), $\emptyset_I$ is the heat flow rate from internal gains (W), $\emptyset_H$ is the heat flow rate released from the heating system (W), $\emptyset_T$ is the heat flow rate lost by transmission (W), $\emptyset_V$ is the heat flow rate lost by ventilation (W). A description of the equations can be found in reference [9].

From the system of three Eq. (1) (for the three TSs) and assuming the internal temperature of the building at 19°C, the temperatures of the glazing, the temperature of the envelope and the heat flow released by the heating system $\emptyset_H$ were derived.

For data-driven modeling, two algorithms were employed: the first, RF, utilizes the principles of bootstrap aggregating (bagging) and the second, LGBM, is grounded in gradient boosting techniques.

RF is an ensemble learning method that belongs to the class of decision tree algorithms [17]. RF builds a "forest" with multiple decision trees by resampling the training data with replacement (bootstrap samples). It then combines the predictions of these trees through averaging or voting. At each split in a decision tree, RF selects a random subset of features to consider for splitting [18], which introduces randomness and reduces overfitting. For regression tasks, it takes the average prediction from all trees.

The second algorithm LGBM is a gradient boosting framework that falls under the category of ensemble learning

methods. LGBM builds decision trees sequentially, each one focusing on the errors made by the previous trees. It minimizes the loss function using gradient descent. Unlike traditional depth-wise tree growth, LGBM grows trees leaf-wise. It selects the leaf with the maximum delta loss to grow, which can lead to faster convergence [19]. This algorithm applies a gradient-based method for sampling instances, where it keeps the instances with larger gradients and randomly drops instances with smaller gradients. This improves the training efficiency without sacrificing accuracy [20].

Differences between utilized algorithms are:

-LGBM uses a leaf-wise growth strategy with Gradient-based One Side Sampling (GOSS) technique for instance sampling [21], while RF typically uses depth-wise growth and samples data by bootstrapping with replacement;

-LGBM calculates feature importance based on the number of times a feature is used in decision trees and the average gain of splits that use the feature, whereas RF calculates feature importance based on the mean decrease in impurity.

In this work, the aim to use process-driven and data-driven (with bagging and gradient boosting algorithms) models and to conduct a comparative analysis of their performance on the dataset or problem and provide a more thorough analysis, increase model robustness, and potentially lead to better predictive performance.

4. CASE STUDY

Turin is the fourth largest city in Italy, it is located in the North-West and has a temperate climate with cold winters and hot-humid summers. It is characterized by about 44,290 buildings and blocks of building, alongside a large District Heating Network (DHN) that supplies 2500 GWh/year to approximately 73.2 Mm^3 buildings and 650,000 residents through 726 km of double pipeline. As Italy's leading district-heated city and one of Europe's with the more extensive DHN, Turin is characterized by predominantly large and compact condominiums, with approximately 83% of residential buildings constructed prior to 1970. Buildings built between 1971 and 2005 account for 16%, while those built after 2005 make up just 1% of the total.

In this work, the hourly consumption data for the heating season 2022-23 of a sample of 110 residential buildings were used to calibrate and validate the energy modeling for space-heating and to describe the different characteristics of the models in predicting energy consumptions. Table 1 shows the weather data in Turin during the heating season 2022-23 in Turin: form November 3rd to April 7th.

Table 1. Climate date for Turin in 2022-23

	Nov-22	Dec-22	Jan-23	Feb-23	Mar-23	Apr-23
T°C	9.4	3.7	4.9	6.9	11.4	13.7
IRR Wh/m^2	73.2	51.1	58.5	100.1	161.5	211.6

Table 2 and Figure 2 reveals that the 110 sampled residential buildings share common characteristics with the entire DH network-connected building stock, and in most aspects, they closely resemble the overall building characteristics in the DH area of Turin. Additionally, median values of some energy-related factors such as S/V (surface-to-volume ratio), BCR (building coverage ratio), H/W (height-to-width ratio), and SVF (sky view factor) were computed and reported in Table 2 for the buildings in the DH area, DH-connected buildings, and 110 sample buildings. Moreover, the 110 sample of buildings for the modeling were selected to represent all periods of construction. These findings show that the sampled buildings effectively represent the broader building stock of Turin in the DH area, thus suggesting that models developed and validated using this sample could reliably predict energy consumption patterns for buildings on a larger scale.

Table 2. Comparison of some characteristics of residential buildings in three clusters of residential buildings

Median Values	S/V, m^{-1}	BCR, -	H/W, -	SVF, -
DH area	0.43	0.32	1.40	0.93
DH connected buildings	0.37	0.31	1.50	0.94
110 sample buildings	0.36	0.32	1.56	0.93

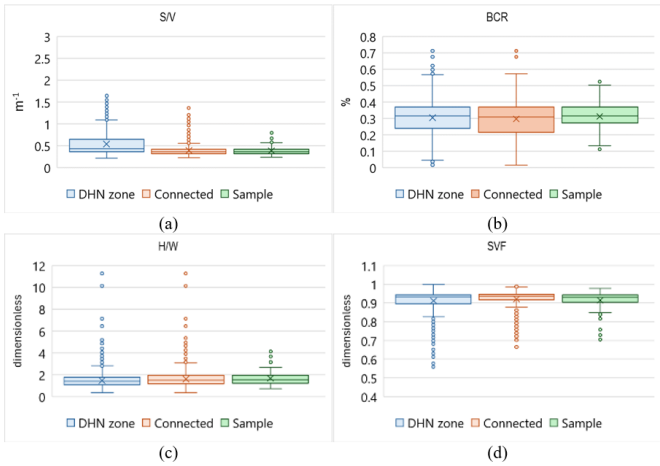


Figure 2. Statistical analysis of the buildings characteristics (a) S/V, (b) BCR, (c) H/W, and (d) SVF

The collection and analysis of data are facilitated by the use of geographic information systems (GIS), which has allowed the creation of a geo-database through the exploitation of these resources:

-BDTRE: the Technical Map and Database of the Piedmont Region.

-ISTAT census database: the socio-economic database of the Piedmont Region.

-Climate data recorded by Politecnico di Torino weather station, including air temperature, direct and diffuse solar radiation, relative humidity, and wind velocity and direction.

-Digital surface model, that describes the 3D built environment with the precision of 0.2 meter.

-Hourly energy consumption data for 110 buildings for the heating season 2022-2023 (provided by the district heating company operating in Turin).

Table 3. Frequency of the variables utilized in the energy-use modeling

Typology of Variable	Frequency
Building geometry	16
Temporal	13
Climatic	3
Socio-economic	4
Building environment	4

For the sample of buildings, a notable effort is invested to collect as much as data is available to build a solid geo-

database, that is fundamental for conducting urban building energy modeling at a urban scale. Overall, 40 variables, as detailed in Table 3, are collected and geo-localized using GIS and made it easy to train models specifically for the case of the use ML algorithms.

5. RESULTS AND DISCUSSION

This section presents the primary findings of this research. Leveraging a substantial dataset containing a complete range of variables, we integrated these variables into Machine Learning and process-driven models for a comprehensive energy-use assessment. Within Table 4, the most significant variables affecting energy consumptions, along with their respective weights, are outlined within the scope of the current study (only seven variables are not affecting the energy-use). Reflecting on the hourly simulation undertaken in this research, it becomes evident that climatic variables, time variables and fundamental building geometries play a pivotal role in analyzing energy consumption patterns. Additionally, socio-economic parameters contributed to enhancing the predictive capabilities of the models. RF model has similar results.

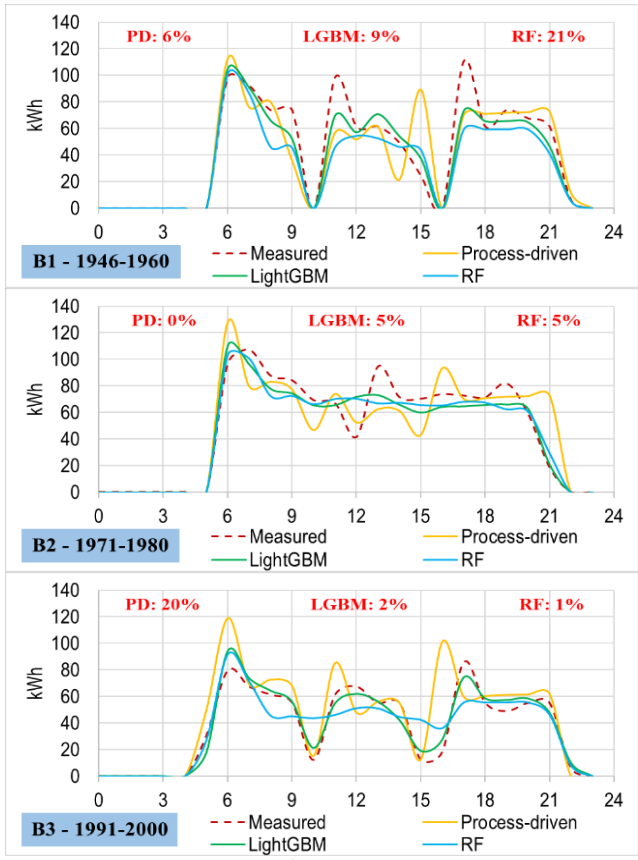
Table 4. Significant variables for LGBM energy-use model

Variable	Significance	Variable	Significance
Hour	10125	Wall area (SE)	674
Air temperature	9174	Walls SVF (NW)	671
Solar radiation	6602	Walls SVF (SW)	660
Day of the year	6304	U-wall	654
Wind velocity	6278	H/H_{avg}	641
Month of the year	4560	Empty dwellings	622
Building height	3802	BD	639
Building surface	2894	H/W	563
Building S/V	1662	BCR	547
Number of inhabitants	1091	Number of strangers	484
Roof SVF	1074	SVF_{avg}	461
Walls area (NE)	915	Windows surface	366
Walls SVF (NE)	880	Building NHS	352
Building volume	875	Number of families	323
Walls SVF (SE)	862	U-window	234
Wall area (SW)	773	U-roof	120
Wall area (NW)	758		

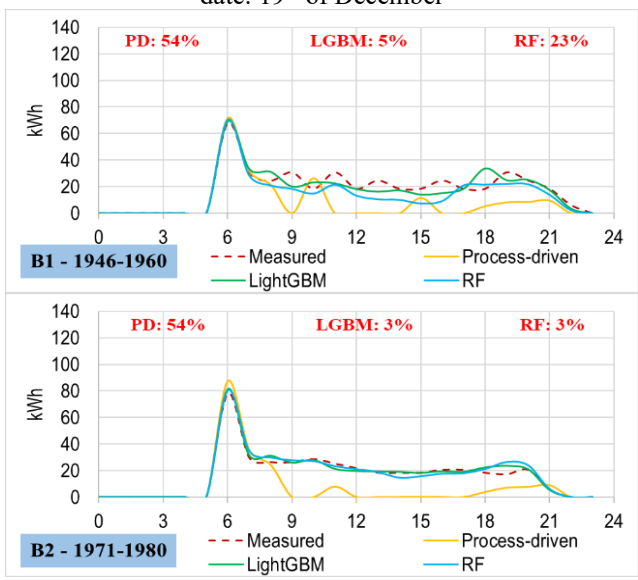
Considering energy-related variables, three energy consumption models were developed utilizing thermal balance equations and LGBM and RF algorithms. To analyze the model's performance, three buildings were selected due to their similar geometric features and different period of construction: B1 built in 1946-1960; B2 built in 1971-1980; and B3 built in 1991 - 2000.

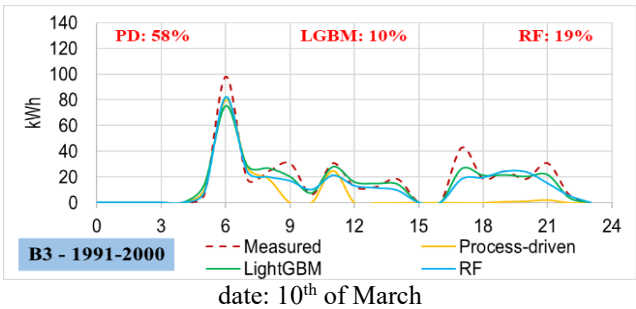
The creation of the three energy consumption models for residential buildings involved a thorough examination of the recent operating configuration of the DHN. Notably, the analysis conducted for the recent heating season from November 2022 to April 2023. Additionally, recent changes in the DHN operation (in agreement with condominium administrators) included two hours of system shutdown within the daily operational timeframe to respect the 13-hour operational limit. These shutdown periods primarily occurred at 9-10 am and 2-3 pm, with potential variations in some instances. However, for the process-driven model, these hours were treated as fixed to shorten simulation time.

The hourly energy consumption profiles of three selected buildings can be analyzed in Figure 3 considering a cold day in December (on the left) and a warm day in March (on the right); these buildings were chosen based on their similar volumes of about 5000 m^3, facilitating meaningful comparisons. The analysis of the graphs in Figure 3 reveals that the simulation models effectively predicted energy consumption trends. More specifically, in the cold months all these models performed remarkably with the average relative error falling below 10%. However, during warmer months the errors increase for the process-driven model, as this model is more influenced by climate variations then different types of regulation. Whereas ML algorithms can learn and comprehend the different energy consumptions in the warmer months more effectively with errors not exceeding 23%.



date: 19th of December





date: 10th of March

Figure 3. Hourly profile of energy consumption with process-driven, LGBM, and RF models for 3 typical buildings



Figure 4. Comparison of monthly measured and predicted energy consumption for three typical buildings

The hourly energy consumption predictions were also aggregated by month to compare the performance of the three models during the year (in Figure 4). The results suggest that, overall, the implemented models are in a high predictive performance. According to the monthly results, process-driven and ML models are underestimating the energy consumptions with a mean annual relative error of respectively 12% (PD), 6% (LGBM) and 8% (RF). Although, the greater error for the process-driven model is mainly due to the underestimation that occurs in the warmer months of March and April.

In Figure 5, plotting the cumulative curves of energy consumptions from the November 1st reveals when these models begin to underestimate space-heating consumptions. The cumulative curves for Machine Learning algorithms closely follow the measured consumption curve, with only a slight underestimation toward the end of the heating season for building B1. In contrast, the process-driven model starts underestimating consumption from the mid of February-March onward. This discrepancy arises because the model is based on an indoor temperature of 19°C, put into force by the regulations; with higher outside air temperatures, the model predicts lower energy-use, while users still consume energy. This indicates that the process-driven model needs to take into account more human behavior factors to better understand the energy-use during the warmer months of the heating season.

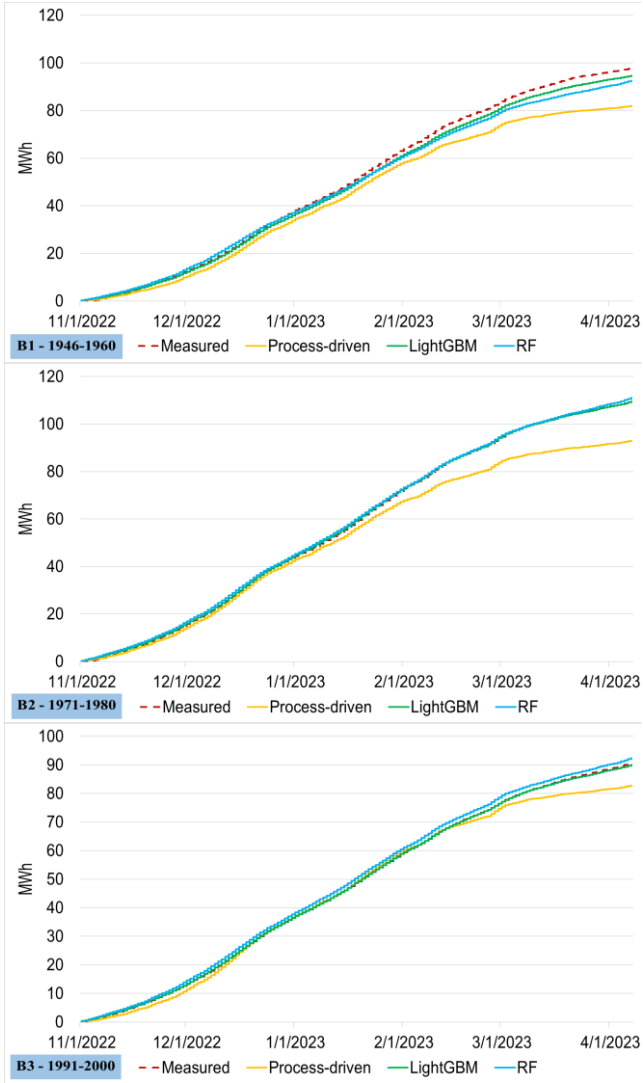
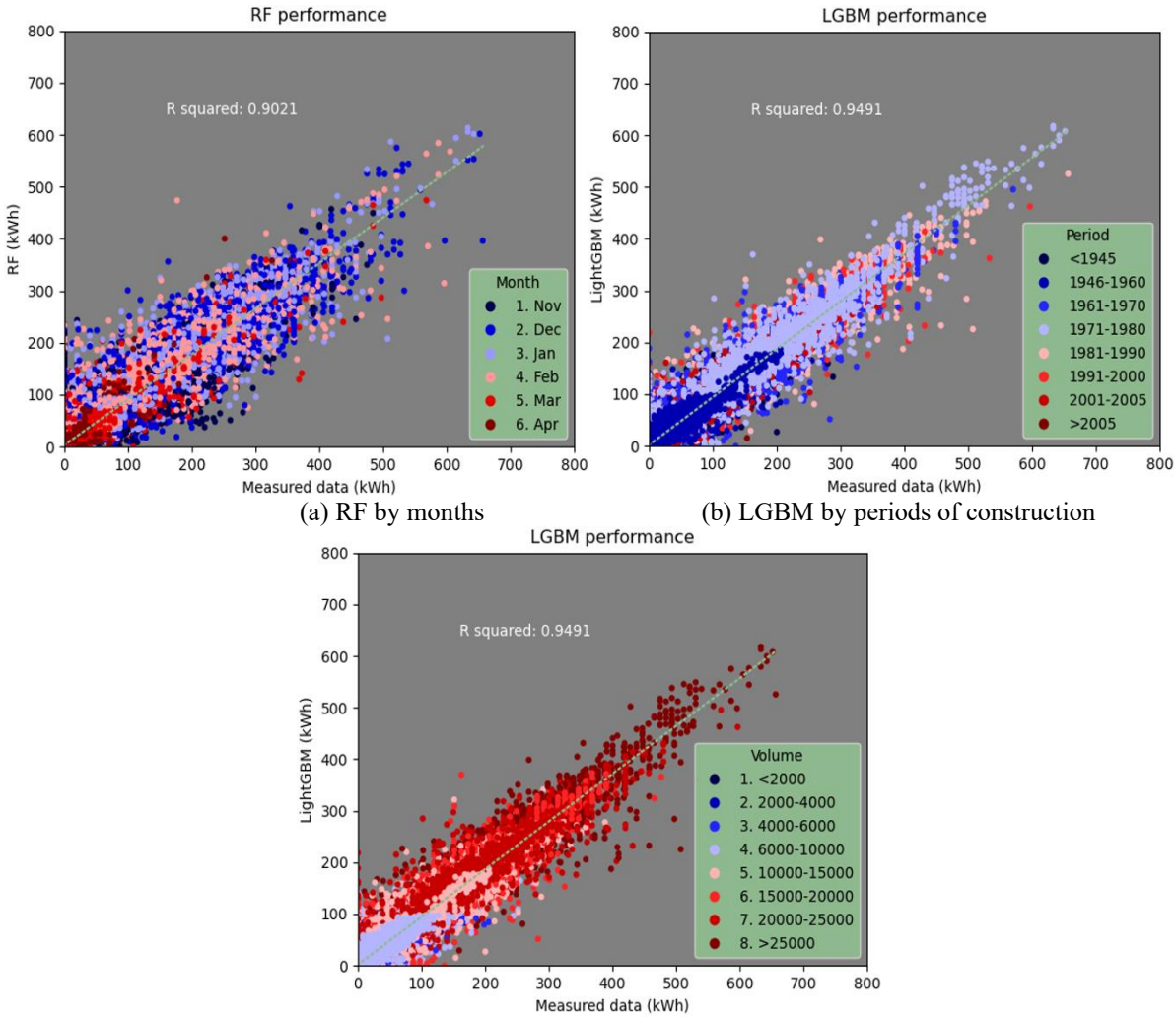


Figure 5. Cumulative curve for energy consumptions comparison for the three sample buildings

To support the last statements, linear regressions depicted in Figure 6, are plotted to check the performance of the modeling by the energy-related variables. Machine learning models have a very high coefficient of determination R2 of 0.90-0.95 with both ML algorithms. Figure 6(a) shows for RF model that the energy-use increases from the warmer to the colder months. In the warmer months of March and April, more inaccuracies are expected from the modeling because the use of energy could depend more by human behavior.

Figures 6(b) and (c) show for LGBM model the influence of period of construction and volume of buildings on their energy-uses. Lower consumptions can be observed by old buildings built before 1960; then, buildings built in '60, '70 and '80 have higher consumptions. Newer buildings have a lower consumption, but they are not visible because of their small number.



(a) RF by months (b) LGBM by periods of construction

(c) LGBM by building volume

Figure 6. Linear regression of ML energy-use models

Table 5. Mean absolute percentage error of energy-use models by month for the sample of 110 buildings with different periods of construction

Period	Model	Nov	Dec	Jan	Feb	Mar	Apr
	PD	16%	9%	8%	13%	34%	28%
Before 1919	LGBM	11%	5%	4%	5%	7%	6%
	RF	9%	5%	4%	3%	6%	13%
	PD	20%	11%	11%	20%	36%	37%
1919- 1945	LGBM	10%	4%	4%	6%	10%	15%
	RF	8%	7%	7%	5%	11%	15%
	PD	16%	17%	13%	24%	46%	37%
1946- 1960	LGBM	13%	10%	6%	10%	15%	14%
	RF	10%	13%	9%	8%	12%	17%
	PD	21%	13%	13%	24%	41%	36%
1961- 1970	LGBM	14%	7%	6%	7%	9%	10%
	RF	13%	9%	6%	4%	7%	20%
	PD	28%	16%	20%	32%	48%	42%
1971- 1980	LGBM	14%	7%	6%	8%	12%	12%
	RF	12%	9%	7%	6%	10%	13%
	PD	23%	15%	19%	30%	50%	41%
1981- 1990	LGBM	9%	3%	4%	5%	8%	7%
	RF	10%	8%	8%	7%	10%	9%
	PD	26%	16%	17%	23%	47%	46%
1991- 2000	LGBM	13%	6%	5%	6%	11%	14%
	RF	10%	6%	5%	4%	9%	11%
	PD	18%	6%	7%	19%	40%	33%
2001- 2005	LGBM	11%	4%	4%	4%	7%	10%
	RF	12%	9%	6%	3%	11%	17%

Also, the dimension of the building is one of the main energy-related factors that can be expressed with the floor area, the volume or the heigh of buildings. Figure 6(c) shows the correlation with the volume that is positively correlated with the energy-use.

Afterward, the analysis extended to realize how the trained models treat buildings of a different period of constructions. Buildings constructed in different periods may show distinct energy consumption trends since they utilize diverse construction technologies and follow other energy standards and directives. To this end, the energy-use models were tested to check their accuracy in predicting energy consumption by period of construction.

The results of Table 5 show the monthly mean absolute percentage errors of the three models for the 110 sample buildings. The errors are lower in the colder months of December and January for all process-driven and data-driven models. There are also some differences by the type of building in terms of period of construction. Process-driven models are suitable in predicting energy consumptions of buildings built before 1970 and after 2001. Moreover, for all buildings, this model has low performance in warm months. However, heeding to data-driven models with ML algorithms, both are fitting for the utilization in predicting energy consumption of buildings across all periods of constructions. Despite this, LGBM outperform slightly RF mostly in cold months, though RF have a consistent performance in all months across different buildings. It is also recognizable that ML algorithms are similarly struggle in predicting energy consumption in warmer months, while they have more solid output in winter.

6. CONCLUSION

In this study, UBEMs with process-driven and data-driven with Machine Learning techniques, specifically LGBM and RF algorithms, were employed to analyze and predict energy consumption in buildings for space-heating. Through rigorous evaluation and comparison, it is observed that all approaches had promising performances, each offering unique strengths and insights.

Our findings indicate that LGBM slightly outperformed RF and the process-driven model in terms of overall predictive accuracy. This suggests that the Machine Learning algorithms effectively captured complex relationships within the data and provided more precise forecasts of energy consumption patterns.

Moreover, distinct performance trends across different building characteristics are noted. The process-driven model demonstrated particular efficacy during colder seasons and exhibited robust performance in both older and more recent buildings. However, it encountered challenges in accurately predicting energy consumption for buildings built between the 1970s and 1990s. This discrepancy opens space for further research to be conducted for physics-based models in capturing the nuanced energy dynamics of buildings from specific architectural periods of construction. On the contrary, the use of physical-based equations allows us to better understand the heat fluxes and predict the effect of variables' variations on the consumptions.

Conversely, LGBM and RF models showcased consistent performance across varying building periods, indicating their versatility and adaptability to diverse architectural contexts. This suggests that Machine Learning techniques, with their capacity to perceive intricate patterns from large datasets, offer a promising chance for enhancing energy modeling accuracy and applicability across a broad ranges of building types and ages, and considering the specific urban context.

Despite the notable advantages of ML models in energy consumption prediction, they are not without limitations. One significant constraint lies in the availability and quality of data. ML algorithms require substantial amounts of data for training to effectively capture underlying patterns and relationships. In the context of urban energy consumption, this poses a potential limit for future scenarios, for example concerning retrofitted buildings or different outside (climate changes) or inside air temperatures (energy savings regulations). With data-driven modeling the consumptions are strictly connected to the independent variables; a change in the variables determines the mandatory development of a new model.

In closing, this comprehensive analysis underscores the complementary nature of process-driven and ML modeling in predicting building energy consumption. Hybrid models can take advantage of the best features of both modeling using process-driven models to guide the physical-based relations and ML algorithms such as LGBM and RF in capturing complex relationships and achieving more accurate forecasts. This research contributes to advancing our understanding of energy modeling methodologies and provides valuable insights for optimizing building energy efficiency strategies in diverse urban environments. UBEM can be used further to find the optimal combination of energy consumptions with low-carbon systems and RES production based on low cost and high reliability [22-24].

ACKNOWLEDGMENT

We wish to extend our gratitude to Iren Group for their support in providing detailed information on space-heating energy consumption of buildings. Their collaboration has greatly enhanced the depth and quality of this paper and has been instrumental to the success of this research endeavor.

REFERENCES

[1] Handbook of Statistics 2023. https://unctad.org/publication/handbook-statistics-2023.

[2] The Strategic Plan 2020-2023. https://unhabitat.org/the-strategic-plan-2020-2023.

[3] Alpagut, B., Gabaldon, A., Zhang, X., Hernandez, P. (2022). Digitalization in urban energy systems – Outlook 2025, 2030 and 2040. https://cinea.ec.europa.eu/publications/digitalization-urban-energy-systems_en.

[4] Urban Energy - Overview. https://unhabitat.org/topic/urban-energy.

[5] Overview - Energy efficiency in historic buildings: A state of the art. https://build-up.ec.europa.eu/en/resources-and-tools/articles/overview-energy-efficiency-historic-buildings-state-art.

[6] Li, W.L., Zhou, Y.Y., Cetin, K., Eom, J., Wang, Y., Chen, G., Zhang, X.S. (2017). Modeling urban building energy use: A review of modeling approaches and procedures. Energy, 141: 2445-2457.

https://doi.org/10.1016/j.energy.2017.11.071
[7] Johari, F., Peronato, G., Sadeghian, P., Zhao, X., Widén, J. (2020). Urban building energy modeling: State of the art and future prospects. Renewable and Sustainable Energy Reviews, 128: 109902. http://doi.org/10.1016/j.rser.2020.109902
[8] Chen, Y.B., Guo, M.Y., Chen, Z.S., Chen, Z., Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. Energy Reports, 8: 2656-2671. https://doi.org/10.1016/j.egyr.2022.01.162
[9] Mutani, G., Todeschi, V., Beltramino, S. (2020). Energy consumption models at urban scale to measure energy resilience. Sustainability, 12(14): 5678. https://doi.org/10.3390/su12145678
[10] Todeschi, V., Javanroodi, K., Castello, R., Mohajeri, N., Mutani, G., Scartezzini, J.L. (2022). Impact of the COVID-19 pandemic on the energy performance of residential neighborhoods and their occupancy behavior. Sustainable Cities and Society, 82: 103896. https://doi.org/10.1016/j.scs.2022.103896
[11] Kontokosta, C.E., Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. Applied Energy, 197: 303-317. https://doi.org/10.1016/j.apenergy.2017.04.005
[12] Boghetti, R., Fantozzi, F., Kampf, J., Mutani, G., Salvadori, G., Todeschi, V. (2020). Building energy models with Morphological urban-scale parameters: A case study in Turin. In 4th IBPSA-Italy Conference Bozen-Bolzano, pp. 131-139. https://doi.org/10.13124/9788860461766
[13] Ministry of Environment and Energy Security. Decreto Ministeriale del 6 ottobre 2022, n.383, in Italian (2022). https://www.mase.gov.it/content/decreto-ministeriale-del-6-ottobre-2022-n-383-piano-nazionale-contenimento-dei-consumi-di.
[14] Jimenez-Palomino, W.H., Soto-Juscamayta, L.M., Ccatamayo-Barrios, J.H., Bendezú-Prado, J.L., Berrocal-Argumedo, K., Esparta-Sanchez, J.A., Maldonado-Llacua, G.M., Mayorga-Rojas, J.C., Romero-Baylon A.A. (2024). Implementation of a GIS for the conservation of irrigation canals: Using ArcGIS and Python for automation. Mathematical Modelling of Engineering Problems, 11(9): 2337-2346. https://doi.org/10.18280/mmep.110907
[15] Shree, P, Suvvari, S (2024). Parallel memory-based collaborative filtering for distributed big data environments, International Journal of Computational Methods and Experimental Measurements, 12(3): 217-225, https://doi.org/10.18280/ijcmem.120303
[16] Khan, I.U., Ullah, M., Tripathi, S., Sahu, M., Zeb, A., Faiza, Kumar, A. (2024). Machine learning for Markov modeling of COVID-19 dynamics concerning air quality index, PM-2.5, NO_2, PM-10, and O_3. International Journal of Computational Methods and Experimental Measurements, 12(2): 121-134. https://doi.org/10.18280/ijcmem.120202
[17] Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Sciences, 43(6): 1947-1958. https://doi.org/10.1021/ci034160g
[18] IBM. (2024). What is Random Forest? In IBM TechXchange Conference, Las Vegas, USA. https://www.ibm.com/cloud/learn/random-forest.
[19] Ke, G.L., Meng, Q., Finley, T., Wang, T.F., Chen, W., Ma, W.D., Ye, Q., Liu, T.Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, California, USA, pp. 3149-3157.
[20] Fan, J.L., Ma, X., Wu, L.F., Zhang, F.C., Yu, X., Zeng, W.Z. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agricultural Water Management, 225: 105758. https://doi.org/10.1016/j.agwat.2019.105758
[21] Chugani, V. (2024). Exploring LightGBM: Leaf-Wise Growth with GBDT and GOSS. https://machinelearningmastery.com/exploring-lightgbm-leaf-wise-growth-with-gbdt-and-goss/.
[22] Altayf, A., Trabelsi, H., Hmad, J., Benachaiba, C. (2024). (2024). Multi-criteria decision-making approach to the intelligent selection of PV-BESS based on cost and reliability. International Journal of Energy Production and Management, 9(2): 83-96, https://doi.org/10.18280/ijepm.090203
[23] Todeschi, V., Mutani, G., Baima, L., Nigra, M., Robiglio, M. (2020). Smart solutions for sustainable cities—The re-coding experience for harnessing the potential of urban rooftops. Applied Sciences, 10(20): 7112. https://doi.org/10.3390/app10207112
[24] Bressan, M., Campagnoli, E., Ferro, C.G., Giaretto, V. (2022). Rice straw: A waste with a remarkable green energy potential. Energies, 15(4): 1355. https://doi.org/10.3390/en15041355

NOMENCLATURE

BCR	Building Coverage Ratio, %
C	Heat capacity, JK^{-1}
H/W	Height to width, m/m
S/V	Surface-to-Volume ratio, m^{-1}
SVF	Sky View Factor, -
t	Time, s
T	Temperature, °C
TB	Temperature of Building, °C
TE	Temperature of Envelope, °C
TG	Temperature of Glazing, °C
U	Thermal transmittance, $Wm^{-2}K^{-1}$

Greek symbols

Ø	Heat flow rate, W

Subscripts

DHN	District Heating Network
GHG	Greenhouse Gas
GOSS	Gradient-based On Side Sampling
H	Space Heating
I	Internal Gains
LGBM	Light Gradient Boosting Machine
ML	Machine Learning
NHS	Net Heated Surface
OLS	Ordinary Least Square

PD	Process-Driven
RF	Random Forest
RES	Renewable Energy Source
sol	Solar Gains
SVM	Support Vector Machine
T	Transmission
TS	Thermodynamic System
UBEM	Urban Building Energy Modeling
V	Ventilation