POLITECNICO DI TORINO
Repository ISTITUZIONALE

Advancing Urban Building Energy Modeling: The Role of Hybrid Energy Modeling in Enhancing Energy Consumption Predictions

Original
Advancing Urban Building Energy Modeling: The Role of Hybrid Energy Modeling in Enhancing Energy Consumption Predictions / Montazeri, A., Zhou, X., Mutani, G.. - In: CHALLENGES IN SUSTAINABILITY. - ISSN 2297-6477. - ELETTRONICO. - 13:4(2025), pp. 525-534. [10.56578/cis130404]

Availability:
This version is available at: 11583/3005158 since: 2025-11-13T17:28:46Z

Publisher:
Acadlore

Published
DOI:10.56578/cis130404

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)

06 September 2026

Challenges in Sustainability
https://www.acadlore.com/journals/CIS

Advancing Urban Building Energy Modeling: The Role of Hybrid Energy Modeling in Enhancing Energy Consumption Predictions

Ahad Montazeri*, Xuan Zhou, Guglielmina Mutani

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

* Correspondence: Ahad Montazeri (ahad.montazeri@polito.it)

Received: 09-26-2025 **Revised:** 10-24-2025 **Accepted:** 11-02-2025

Citation: Montazeri, A., Zhou, X., & Mutani, G. (2025). Advancing urban building energy modeling: The role of hybrid energy modeling in enhancing energy consumption predictions. *Chall. Sustain.*, *13*(4), 525–534. https://doi.org/10.56578/cis130404.

© 2025 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Urban building energy modeling (UBEM) is essential for understanding energy consumption and developing sustainable policies at the city scale. However, current UBEM approaches overlook spatial and temporal interactions and lack generalizability across diverse urban contexts. This study introduces a hybrid framework that integrates physics-based simulations with machine learning based residual learning to enhance prediction accuracy using real energy consumption data. The methodology incorporates GIS-supported data collection and processing. Multiple ML models were applied to predict monthly consumption and validate their performance. Meanwhile, a physics-based model is used to simulate hourly energy consumption. The best performing ML model was later used for daily residual learning to calibrate physics-based simulation outputs. The framework was tested on residential buildings connected to the District Heating Network in Turin, Italy. Results showed LGBM achieved the highest performance with a R^2 of 0.883 and a MAPE below 15% in most months. Residual learning reduced daily prediction error in 80% of cases, with up to 75% improvement in extreme cases. After model calibration, 65% of buildings achieved a daily MAPE below 30%, and 55% fell below 20%, demonstrating consistent error reduction across varied building types and consumption levels. This confirms the effectiveness of the hybrid approach in enhancing accuracy and reliability at the urban scale.

Keywords: Urban building energy modeling; Process-driven models; Data-driven models; Hybrid models; Machine learning; Space heating

1. Introduction

Urban building energy modeling (UBEM) is a key tool for understanding urban energy consumption and developing energy-saving strategies (Patil et al., 2025), especially since buildings account for nearly 40% of total urban energy use (Pérez-Lombard et al., 2008). Traditionally, there are two main approaches in UBEM: the first is the bottom-up physical simulation method (Ali et al., 2021); it uses building physical features and weather data to simulate energy use through tools such as EnergyPlus (Li et al., 2023); the second is the data-driven method (Ali et al., 2021), which could use machine learning (ML) to explain energy patterns from historical consumption data. While physical models are clear and explainable, they need calibration, take time, and are hard to scale up to the city level (Li et al., 2023). In contrast, data-driven models are faster and often more accurate, but they lack explainability and depend heavily on the quality of training data (Ali et al., 2021). They are also less reliable when applied to city-wide cases.

To solve these problems, hybrid models (Li et al., 2023) have become more popular in recent years. These models combine the explainability of physics-based methods with the predictive power of ML, thereby enhancing both interpretability and consistency with real-world building energy performance (EP) (Oraiopoulos & Howard, 2022). Most studies still use a bottom-up UBEM method, starting from individual buildings and expanding to blocks or neighborhoods (Kavgic et al., 2010). Well-known examples include the Data-driven Urban Energy Simulation (DUE-S) frameworks that were studied by Nutkiewicz et al. (2017), Nutkiewicz et al. (2018), and Nutkiewicz et al. (2021). These studies combine physical simulation (EnergyPlus) with deep learning models like Residual Neural Network (ResNet) and Long Short-Term Memory (LSTM) to predict the difference between

simulated and real energy use. They also add network-based ML methods to model buildings' interactions, enabling scenario testing without retraining.

Similarly, Mui et al. (2021) proposed a hybrid model combining EnergyPlus and Artificial Neural Network (ANN) in Hong Kong. The model considered different typical public housing block layouts in high-density urban environments. And recently, Dai et al. (2025) developed the CityTFT model, which integrates CitySim pro with a Temporal Fusion Transformer (TFT) to create a time-series energy prediction framework, trained on 114 campus buildings. While both models showed good performance in terms of speed and accuracy, they did not comprehensively consider interactions between buildings and did not include real measured electricity or smart meter data for supervision or calibration.

Other researchers have explored the performance of different ML algorithms in predicting energy consumption. For example, Kamel et al. (2020) tested several models, including eXtreme Gradient Boosting (XGBoost) and Simple Linear Regression (SLR). Similarly, Mehdizadeh Khorrami et al. (2024) evaluated 768 residential buildings, comparing Decision Trees (DT), Logistic Regression (LR), and ANN. Quan (2024) developed a prediction model for annual electricity use in residential buildings using data from 2,078 neighborhoods in Chicago, which systematically compared five ML models: DT, ANN, Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Gradient-Boosted Decision Trees (GBDT). Their results highlighted the trade-off between model accuracy and explainability.

Some studies focus more on dynamic features. Wang et al. (2019) used LSTM networks to predict indoor heat gains based on data such as Wi-Fi connection counts, device usage, and occupant schedules. This research showed the potential of behavioral features to improve the accuracy of hourly-level predictions. Although previous research (such as Deng et al., 2018) has pointed out the impact of system design, management, and user behavior on EP gaps, most models still fail to include these uncertainty characteristics.

Recently, many researchers have attempted to use existing ML tools embedded within Geographic Information Systems (GIS) for building-level energy prediction or spatial regression. The reason is that large-scale energy prediction requires a significant amount of data collected, preprocessed, and stored, which, with the support of GIS tools, is more easily manageable. However, there are still notable limitations in ML integration with GIS platforms. QGIS plugins like Dzetsaka (Karasiak, 2021) and the Semi-Automatic Classification Plugin (Congedo, 2024) mainly work for satellite image classification and spatial tasks, which are not suitable for prediction. ArcGIS Pro includes more tools, like Forest-based Classification and Regression, Generalized Linear Regression, and Geographically Weighted Regression. These tools can handle vector data and simple predictions (Esri, n.d.). Nowadays, they cannot easily include time-based data, many custom features, or other model structures, and their modeling workflows are fixed and difficult to customize, which makes it difficult for complex tasks like UBEM.

Consequently, this research contributes to the advancement of urban-scale hybrid UBEM supported by GIS-based data processing. This study leverages GIS as a foundational tool for large-scale spatial data integration that enables the inclusion of high-resolution building attributes, contextual urban features, and time-dependent inputs (Mutani & Todeschi, 2017). Moreover, this study emphasizes the importance of hybrid models, integrating advanced ML models to capture non-physical behavioral and environmental influences that are difficult to represent in purely physics-based models. Additionally, the framework supports transfer techniques to ensure scalability and adaptability across diverse urban contexts.

2. Research Gap and Objectives

Current UBEM studies use a wide range of ML methods. These include simple tree models, deep learning, and ensemble methods. But most of them share some common problems:

(1) Insufficient data realism: Most UBEMs do not use real consumption data but rely on simulated consumption data as the target variable. These simulated values depend on many input assumptions and cannot fully reflect how buildings perform under real operating conditions.

(2) Lack of urban-scale modeling: A lot of studies are limited to single buildings, neighborhoods, or specific types of buildings. They often do not include spatial interaction between buildings, which is important for modeling energy use at the urban scale.

(3) Weak temporal modeling: Few models use time-based models like LSTM. Most are static and cannot capture the strong daily, weekly, or seasonal changes in energy use.

(4) Limited generalization: Many datasets are from one city or one type of building. They lack diversity in building types and climates, so models are hard to apply in other places.

Even though these models often have good accuracy, they cannot yet fully model the complex relationship between space, time, and human behavior. These features are key to predicting building consumption across a whole urban scale. This research uses several ML models to improve energy consumption prediction. Additionally, it runs physics-based simulations and compares the simulation results with real energy data at the daily level to analyze the difference and identify key factors that cause it. This helps improve the model and better understand how different variables affect urban building energy consumption.

3. Methodology

This study introduces a hybrid framework that integrates physics-based simulations with ML techniques to predict building energy consumption at the urban scale. The aim is to combine physics-based simulations with data-driven residual correction across the daily timescale to reduce prediction errors. The framework is specifically designed to address spatial, temporal, and behavioral uncertainties. The methodology comprises several stages shown in Figure 1, detailed as follows.

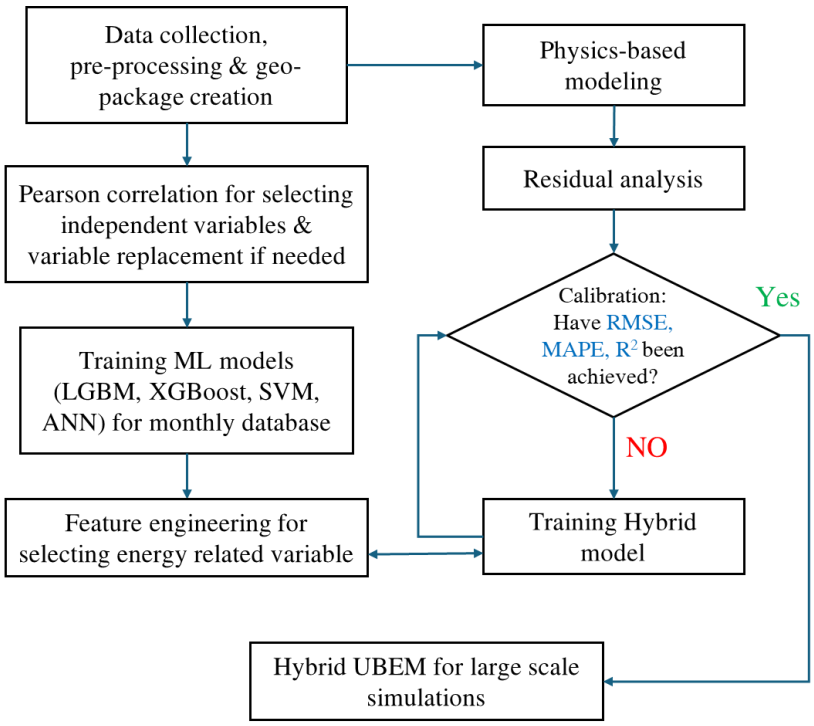

Figure 1. Flowchart of the applied methodology

3.1 Data Collection and Preprocessing

A wide range of datasets was compiled to support the hybrid modeling approach. This step is supported by GIS-based analysis for data collection and creating site-specific geo-packages. Building data included geometric attributes (such as footprint, height, volume, surface, and S/V ratio), thermal properties (such as thermal transmittances and thermal bridge heat losses), and occupancy information. Spatial characteristics such as adjacency relationships, Building Coverage Ratio (BCR), Building Density (BD), Sky View Factor (SVF), and neighborhood average height were also incorporated. Contextual data like climatic variables, socio-economic characteristics of population, and temporal features, including hourly timestep and day-of-week indicator, enriched the dataset.

For energy consumption, monthly metered data were used to train the ML models, while high-resolution hourly readings supported daily residual learning. In this phase, feature engineering played a crucial role in model performance. This step made it possible to select more energy-related variables while excluding the correlated parameters. For this reason, SHAP (SHapley Additive exPlanations) values were used to perform sensitivity analysis and identify the most influential features across the models.

3.2 Monthly Energy Prediction at the Urban Scale

To predict monthly energy consumption, four different ML regression models were trained based on the monthly EP of buildings:

- ANN used to learn intricate consumption patterns influenced by factors like occupancy behavior, weather variability, and building characteristics;
- XGBoost and Light Gradient Boosting Machine (LGBM), both of which are well-suited to train large-scale, heterogeneous urban data; and
- SVM for its effectiveness in handling high-dimensional datasets, especially valuable when training data is limited.

Model performance was assessed using common evaluation metrics such as Median Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and R-squared (R^2), with comparisons made against actual monthly consumption data.

3.3. Daily Residual Learning for Calibration

To generate high-resolution energy predictions, hourly simulations were carried out using a process-driven model (Mutani et al., 2020). Outputs consisted of hourly energy consumption for sample buildings in the study area. The study enhanced physics-based predictions through residual learning. Residuals, defined as the difference between measured and simulated hourly values, were aggregated to daily averages to mitigate noise. A LGBM algorithm was then trained to model these residuals. The model was tuned with hyperparameter optimization using Optuna with 100 trials. The LGBM architecture was chosen due to its ability to capture non-linear patterns in the residuals, particularly those arising from spatial-temporal effects on energy consumption.

3.4 Validation Framework

In each step that ML models utilized, after carefully tuning hyperparameters and training and testing their performance, their generalization power was validated over a sample of randomly selected buildings. Monthly ML models were evaluated through MAPE comparisons, with acceptable variability set at less than 20%. The performance of the residual model was benchmarked against the measured data with a target MAPE of less than 30% if the MAPE was improving with respect to physics-based simulation.

This research contributes several methodological innovations:

- A spatial-temporal hybrid framework that integrates urban-scale physics simulations with ML-based residual learning.
- A transferable modeling framework that leverages large-scale training for enhanced adaptability across diverse urban contexts.

The proposed workflow balances computational efficiency with accuracy, highlighting the trade-offs and dependencies inherent in hybrid modeling approaches, particularly the reliance on the quality of the physics-based model to establish realistic residual bounds.

4. Case Study

The proposed methodology is applied to the residential buildings connected to the District Heating Network (DHN) of Turin, for which monthly or hourly energy consumption data are available. From 4888 residential buildings or blocks of buildings connected to DHN, the monthly energy consumption data of 3917 were available for the 2022–2023 heating season. However, after preprocessing of consumption data for outliers and missing information, the buildings included in the investigations were 2453 buildings. For daily residual analysis, the hourly consumption of 136 buildings was available for the same heating season, from which 9 buildings were excluded from the assessments because the energy consumption of those buildings was not completely recorded.

Table 1 reports the statistics of the sample buildings for each timestep analysis.

Table 1. Characteristics of the buildings with available energy consumption data

POC	Buildings with Monthly Consumption			Buildings with Hourly Consumption		
	n	S/V	EP	n	S/V	EP
Pre–1919	72	0.33	18.43	13	0.33	18.45
1919–1945	346	0.33	17.88	10	0.33	20.90
1946–1960	652	0.34	18.90	12	0.35	43.47
1961–1970	895	0.33	19.53	34	0.36	23.35
1971–1980	403	0.33	20.56	25	0.32	23.57
1981–1990	67	0.34	21.07	16	0.31	28.49
1991–2000	10	0.33	23.41	9	0.34	29.03
2001–2005	7	0.32	19.53	5	0.31	25.00
2006–2012	-	-	-	1	0.31	26.39
2013–2015	1	0.60	27.14	2	0.30	21.70

For new buildings built after 2006, and as highlighted in the table, the available data were limited, and they were excluded from this analysis. Consequently, the models tested in this paper can be applied to old buildings built before that particular year.

For analysis in each timestep, 80% of the buildings were randomly selected for training and testing of the ML models, and the remaining 20% of them were reserved for model calibration. The amount of data available in this

study easily supported the selection of the buildings' sample for validation. The validation dataset acted as an unseen building to assess the generalization power of the ML models. The training and testing datasets were also split with 80% and 20% rates, respectively.

5. Discussion and Results

This section provides a detailed explanation of the applied methodology with results extracted. The proposed workflow offers a flexible and replicable approach to urban-scale energy modeling, combining the strengths of physics-based simulation and ML to create a hybrid model. The flexibility and replicability of the methodology are guaranteed with GIS-based data acquisition and geo-package creation for a large-scale context (Mutani et al., 2025) that is the backbone of UBEMs. This phase is fundamental for appropriate UBEM.

Before creating any simulation modeling, the input features were categorized and selected based on statistical correlation analysis, physical relevance, and modeling objectives. Pearson correlation coefficients were used to detect multicollinearity, while semantic overlap among variables was evaluated to eliminate redundant information. Priority was given to features that represent key dimensions such as building geometry, thermal performance, environmental exposure, occupancy behavior, and construction background. Geometric and construction-related variables were largely consolidated, resulting in a final set of 23 input features.

Correlation analysis revealed strong associations between household-related or economic indicators (e.g., number of residents, households, and employment rate) and geometric attributes. This suggests a structural coupling in interrelated variables. However, to ensure the model captures variations in energy consumption driven by non-geometric factors rather than solely by building characteristics, certain non-geometric features were retained after normalization or ratio transformation. For instance, we introduced composite variables such as the number of inhabitants per family, income per resident, and the ratio of temporary to permanent residents to reduce multicollinearity between behavioral and structural features. These indicators better reflect differences in population distribution and occupancy structure, enhancing the model's ability to explain behavioral impacts on energy consumption.

Importantly, while redundant variables were removed, the selection process ensured that each major energy-influencing factor, like principal geometric variables, user behavior, and environmental conditions, remained represented. This preserved the model's capacity to learn key determinants of urban building energy use. Compared to the unfiltered feature set, the refined version significantly reduced training time and improved convergence. Moreover, the streamlined model demonstrated more stable feature importance rankings on the test set. Ultimately, the final model achieved a balance between predictive accuracy and interpretability.

In the current research, four ML models, namely, ANN, LGBM, XGBoost, and SVM, are tested to understand how they are performing in the energy consumption task. In the training and testing phase of the study, the hyperparameters of each model were tuned using the Optuna optimizer with the integration of 5-fold cross-validation. Only the SVM was tuned using Grid Search optimization to reduce the computation cost of the model tuning.

To further enhance ML models' prediction capabilities, while understanding how the model predicts building energy consumption, SHAP (SHapley Additive exPlanations) values were used to quantify the contribution of each input feature. Using the results of this step, more energy-related variables were selected for final model training and testing. The mean absolute SHAP values were generated and reported in Table 2 to illustrate the average impact of each variable on the model's output across the entire sample. It is also important to note that the feature importance analysis was not concluded for SVM because of its high computational cost.

The analysis reveals fundamental differences in how the three algorithms respond to input features. ANN clearly captures seasonal dynamics through strong SHAP responses to climate and temporal variables; tree-based models display different behaviors. Although feature importance rankings in LGBM and XGBoost indicate that variables such as Month and Outdoor Temperature were involved in some splitting paths, their marginal SHAP contributions suggest these features had little actual influence on prediction outcomes. Instead of learning seasonal changes directly, tree-based models rely more on static features (such as S/V, BD, SVF) to classify energy consumption levels. Their ability to predict monthly energy consumption in unseen buildings may come from indirect patterns, where time-related effects are reflected through their association with these structural variables. However, if future buildings have different climate conditions or physical characteristics, the model's performance may decrease due to this limited flexibility.

The non-negligible importance of the Height-to-Width ratio (H/W) implies that spatial configurations at the urban scale, such as building spacing and orientation, may influence microclimatic interactions through mechanisms like solar access, shading, and ventilation in urban canyons.

Social and household-related features, such as the number of families, heated floor area per family (NHS), population density, and occupancy rate, also exhibit measurable impacts. These variables suggest that the model partially accounts for the influence of usage behavior on heating demand. This suggests that more research needs to be done to carefully account for behavioral proxies in energy consumption.

Interestingly, the thermal transmittance of the building components shows very low SHAP values. This is primarily due to their derivation: instead of being measured or calibrated, U-values in the dataset were assigned based on building age. This limited variability across samples restricts their informativeness in the model. Although these parameters are physically meaningful, their low variance means they contribute little to the model's learning process in a ML context.

Table 2. Feature importance analysis using SHAP

Training Feature	ANN	LGBM	XGBoost
Outdoor Air Temperature	0.7115	0.0354	0.0507
Solar Radiation	0.3438	-	-
Month	0.3325	-	-
N. Days of Month	0.2264	0.0041	-
S/V	0.1691	0.0617	0.0756
Wind Speed	0.1323	-	-
Height	0.0879	0.0233	0.0332
NHS per Family	0.0404	0.0171	0.0178
BD	0.0276	0.0124	0.0115
U glazing	0.0258	0.0000	0.0001
Occupancy %	0.0235	0.0115	0.0061
H_W	0.0225	0.0172	0.0170
U wall	0.0218	0.0293	0.0223
Income per Inhabitant	0.0218	0.0213	0.0219
H_Havg	0.0212	0.0058	0.0064
SVF roof	0.0187	0.0168	0.0225
SVF S_wall	0.0181	0.0138	0.0136
Stranger per Inhabitant	0.0156	0.0048	0.0064
Area	0.0152	0.0115	0.0107
U ground	0.0150	0.0004	0.0065
SVF E_wall	0.0124	0.0069	0.0079
Families	0.0084	0.0041	0.0038
SVF N_wall	0.0063	0.0085	0.0061
SVF W_wall	0.0055	0.0124	0.0054
U roof	0.0031	-	-

Figure 2 presents the predicted EP index values by the ANN and LGBM models, plotted against actual measured data across different months. Both models effectively capture seasonal variations in energy use, with notably higher accuracy during peak demand periods from December to February. The ANN model achieves an R^2 of 0.877, indicating a strong predictive relationship, though it tends to underestimate consumption at higher levels (above 5 kWh/m^3/m). Its predictions are more concentrated during the winter months, while greater variability appears particularly at lower consumption levels in November, March, and April. LGBM, with a slightly higher R^2 of 0.883, shows a marginally improved fit and better alignment with the ideal prediction line, especially in mid-to-high consumption ranges. It also exhibits reduced scatter and more consistent predictions compared to ANN, particularly for lower energy consumed periods.

Figure 3 illustrates the average monthly MAPE for the ML model predictions. Errors for the ANN and SVM models for April are omitted from the graph, apart from SVM in November, which is also excluded. These omissions are due to their significantly higher error values, which distorted the scale and reduced the readability of the graph for other models and months.

The results indicate that SVM often produces MAPE values around or above 30%, except in February, highlighting its limited ability to capture the complex relationships between input features and the EP. In contrast, the other models generally maintained MAPE below 15% across most months, except for ANN and XGBoost, which exceeded 20% in April. These findings suggest that ANN, LGBM, and XGBoost offer acceptable accuracy for urban-scale monthly EP prediction, with LGBM performing more consistently. This may be attributed to LGBM's strength as a tree-based algorithm, which allows it to capture patterns associated with temporal features, even without explicitly modeling sequential dependencies. Unlike ANN and other deep learning models, which often require a rich set of time-structured or lagged features to perform well, LGBM can leverage static encodings of time-related variables with relatively simple inputs.

The LGBM model, which showed the best performance in temporal EP prediction, was chosen for the residual learning phase. This phase builds a hybrid modeling approach by integrating LGBM with a validated process-driven simulation to correct prediction residuals. Hourly simulation outputs from the physics-based model were compared against measured data to calculate residuals, which were then aggregated at the daily level for analysis.

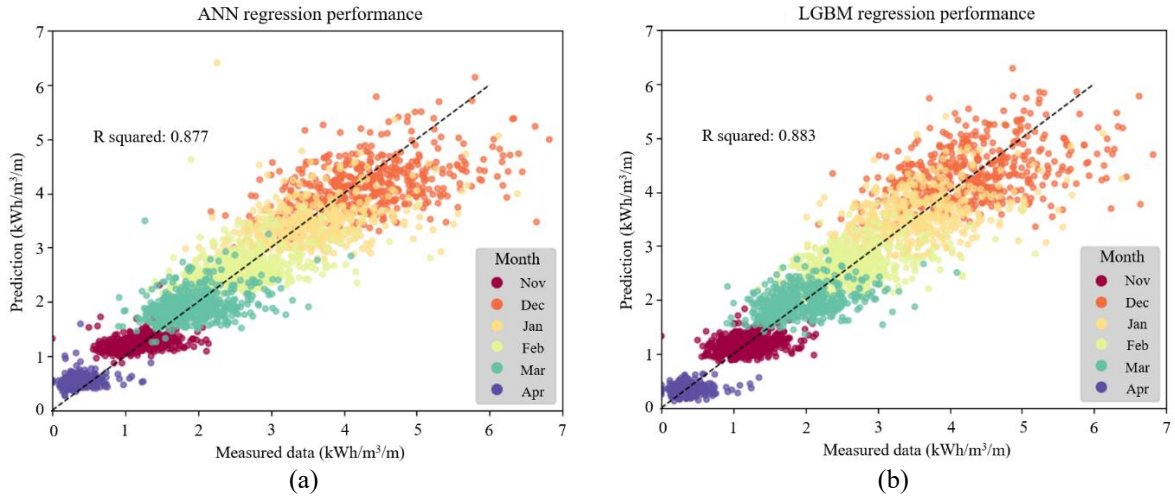

Figure 2. Measured and predicted EP indexes with (a) ANN and (b) LGBM models

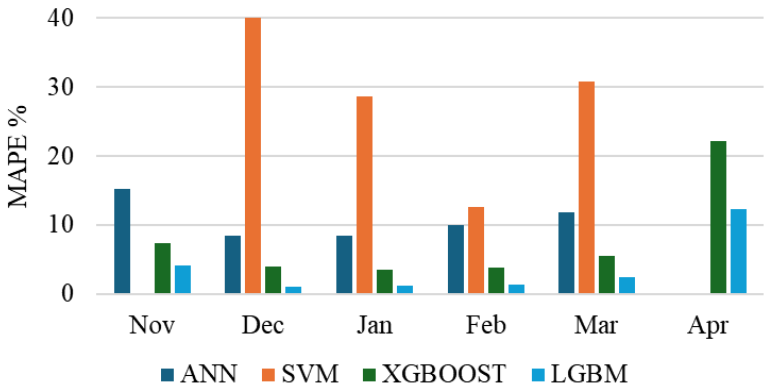

Figure 3. Flowchart of the applied methodology

Before training, 20 buildings were randomly selected to serve as the validation set, while the remaining buildings were divided into training and testing sets. The LGBM model was then fine-tuned using hyperparameter optimization and trained on the residuals. Once trained, it was applied to the validation buildings for residual correction. The results of this calibration indicate that in 4 cases, the correction increased the error. However, for the remaining 16 buildings, the error was reduced, achieving post-calibration errors below 30% and 20%, in 65% and 55% of cases.

Figure 4 illustrates the residual correction results for two sample buildings. As shown in the graph, the trained model successfully reduced the residuals by approximately 30% on average. In the most extreme case, the residual correction achieved an improvement of up to 75%.

In closing, the results of the suitability of ML models for monthly building energy consumption reveal that each model presents distinct strengths, weaknesses, opportunities, and threats (SWOT) within this domain.

ANN shows powerful capabilities in capturing nonlinear relationships among climate, temporal, and behavioral variables, offering advantages in identifying seasonal patterns. However, it requires extensive preprocessing, cannot handle missing or categorical data natively, and depends heavily on large, high-quality datasets. With improved interpretability tools such as SHAP, ANN shows potential for future applications in user behavior-based modeling, although its generalization remains limited in data-scarce or unstable contexts.

SVM is simple in design and works well with small (maximum 10,000 training data), high-dimensional datasets. However, it cannot process missing values, categorical data, or time-related features, and it does not capture interactions between features, making it less adaptable for complex energy consumption. It may serve as a lightweight tool for early-stage modeling, but it has problems with scaling up to urban-level data.

XGBoost handles interactions well, is not sensitive to input scale, and supports missing and categorical data. Nonetheless, its interpretability is limited, and it tends to prioritize dominant variables, that potentially ignoring climate features with low variance. While adding lag features or time windows could improve its ability to model changes over time.

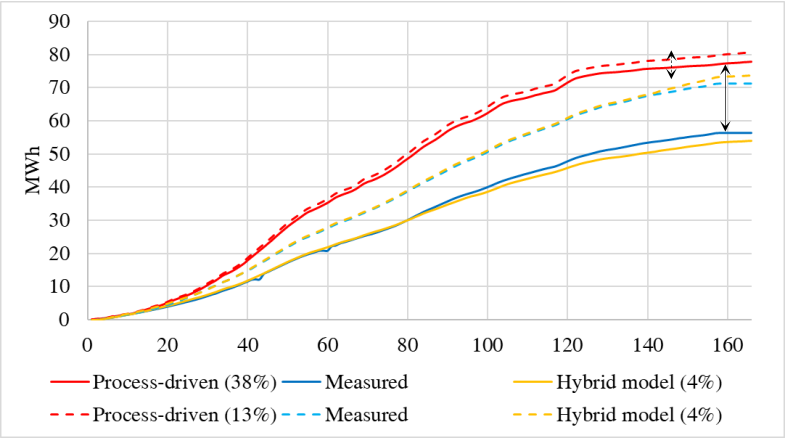

Figure 4. Flowchart of the applied methodology

LGBM performs efficiently on sparse, unbalanced, or high-dimensional datasets, with fast training and low preprocessing needs. Its strengths make it ideal for integration into urban GIS-based platforms. However, it mostly uses static features and responds weakly to changes in climate or time, which may limit its accuracy and generalization when applied to new regions or under future climate conditions.

In sum, these models show complementary SWOT characteristics. In practical applications, model selection should be based on data characteristics and modeling goals or combined in ensemble frameworks to improve both prediction accuracy and interpretability.

Despite their complementary strengths, these ML models rely solely on historical data and lack physical constraints, which can limit their interpretability and generalization to unseen conditions. Moreover, process-based simulations, while physically modeling, often exhibit systematic discrepancies from real-world measurements due to input uncertainties or modeling assumptions. To address these limitations, this study developed a hybrid approach that integrates LGBM with process-based physical simulation to correct prediction residuals.

Future research could explore residual learning strategies with optimized temporal structures and evaluate the transferability of this approach across different regions. In addition, the residual feedback mechanism of the hybrid model could help adjust input parameters of physical simulations, improving their reliability. This would also support multi-scenario analysis before and after retrofitting, offering valuable insights for building upgrades and regional energy-saving policies.

6. Conclusions

This study presented a hybrid UBEM framework that integrates GIS-supported spatial data processing, physics-based simulations, and ML to improve the accuracy and generalizability of energy consumption predictions at the urban scale. The methodology was applied to a comprehensive dataset of 2453 residential buildings in Turin, Italy, all connected to the DHN, with hourly energy data available for 127 buildings.

Among the four tested ML models, ANN, LGBM, XGBoost, and SVM, LGBM consistently outperformed the others in terms of accuracy and robustness. It achieved the highest R^2 value of 0.883, compared to 0.877, 0.888, and 0.889 for ANN, XGBoost, and SVM, respectively, and maintained a monthly MAPE below 15% in most months, with better prediction stability even during low consumption periods. In contrast, SVM exhibited MAPE values exceeding 30% in several months, highlighting its unsuitability for this task. ANN and XGBoost surpassed the 20% MAPE threshold in April, revealing occasional temporal sensitivity.

The feature selection process, guided by SHAP values, confirmed the dominant influence of outdoor temperature (SHAP = 0.7115 in ANN) and solar radiation on energy use, while highlighting the relatively minor role of U-values due to their low variability in the dataset. Variables such as S/V ratio, building height, and occupancy metrics also contributed significantly to the model's output, reinforcing the importance of integrating behavioral and morphological data.

In the residual learning phase, the LGBM model was trained on daily residuals from physics-based simulations, against real hourly energy data. After tuning with Optuna over 100 trials, the model was validated on a 20-building sample. Results showed error reduction in 16 out of 20 buildings, with 65% achieving post-calibration MAPE below 30%, and 55% below 20%. In extreme cases, residual correction improved prediction accuracy by up to 75%. These findings demonstrate the hybrid framework's ability to compensate for simulation inaccuracies caused by behavioral or temporal uncertainties. This research contributes to a scalable hybrid UBEM workflow adaptable to diverse urban contexts and reveals that the integration of ML with process-driven simulation provides a powerful pathway toward more accurate, explainable, and transferable urban energy models.

Author Contributions

Conceptualization, G.M. and A.M.; methodology, G.M. and A.M.; software, A.M. and X.Z.; validation, and A.M. and X.Z.; writing—original draft preparation, A.M.; writing—review and editing, G.M.; supervision, G.M. All authors have read and agreed to the published version of the manuscript.

Funding

This work is funded by the PNRR-NGEU project (Grant No.: MUR–DM118/2023).

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

Ali, U., Shamsi, M. H., Hoare, C., Mangina, E., & O'Donnell, J. (2021). Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. *Energy Build.*, *246*, 111073. https://doi.org/10.1016/j.enbuild.2021.111073.

Congedo, L. (2024). *Semi-Automatic Classification Plugin (SCP): User Manual*. https://semiautomaticclassificationmanual.readthedocs.io/en/latest/

Dai, T. Y., Niyogi, D., & Nagy, Z. (2025). CityTFT: A temporal fusion transformer-based surrogate model for urban building energy modeling. *Appl. Energy*, *389*, 125712. https://doi.org/10.1016/j.apenergy.2025.125712.

Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. *Energy Build.*, *163*, 34–43. https://doi.org/10.1016/j.enbuild.2017.12.031.

Esri. (n.d.). *An overview of the Modeling Spatial Relationships toolset*. ArcGIS Pro Documentation. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/an-overview-of-the-modeling-spatial-relationships-toolset.htm

Kamel, E., Sheikh, S., & Huang, X. (2020). Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. *Energy*, *206*, 118045. https://doi.org/10.1016/j.energy.2020.118045.

Karasiak, N. (2021). *Dzetsaka: Classification tool* [QGIS plugin]. GitHub. https://github.com/nkarasiak/dzetsaka

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., & Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Build. Environ.*, *45*(7), 1683–1697. https://doi.org/10.1016/j.buildenv.2010.01.021.

Li, Z., Ma, J., Tan, Y., Guo, C., & Li, X. (2023). Combining physical approaches with deep learning techniques for urban building energy modeling: A comprehensive review and future research prospects. *Build. Environ.*, *246*, 110960. https://doi.org/10.1016/j.buildenv.2023.110960.

Mehdizadeh Khorrami, B., Soleimani, A., Pinnarelli, A., Brusco, G., & Vizza, P. (2024). Forecasting heating and cooling loads in residential buildings using machine learning: A comparative study of techniques and influential indicators. *Asian J. Civ. Eng.*, *25*(2), 1163–1177. https://doi.org/10.1007/s42107-023-00834-8.

Mui, K. W., Wong, L. T., Satheesan, M. K., & Balachandran, A. (2021). A hybrid simulation model to predict the cooling energy consumption for residential housing in Hong Kong. *Energies*, *14*(16), 4850. https://doi.org/10.3390/en14164850.

Mutani, G. & Todeschi, V. (2017). Space heating models at urban scale for buildings in the city of Turin (Italy). *Energy Procedia*, 122, 841–846. https://doi.org/10.1016/j.egypro.2017.07.445.

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.

Mutani, G., Tundo, A., & Capezzuto, P. (2025). Renewable energy communities in Italy: A national framework for sustainable cities. *Chall. Sustain.*, *13*(3), 398–411. https://doi.org/10.56578/cis130306.

Nutkiewicz, A., Choi, B., & Jain, R. K. (2021). Exploring the influence of urban context on building energy retrofit performance: A hybrid simulation and data-driven approach. *Adv. Appl. Energy*, *3*, 100038. https://doi.org/10.1016/j.adapen.2021.100038.

Nutkiewicz, A., Yang, Z., & Jain, R. K. (2017). Data-driven Urban Energy Simulation (DUE-S): Integrating machine learning into an urban building energy simulation workflow. *Energy Procedia*, *142*, 2114–2119.

https://doi.org/10.1016/j.egypro.2017.12.614.
Nutkiewicz, A., Yang, Z., & Jain, R. K. (2018). Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Appl. Energy*, *225*, 1176–1189. https://doi.org/10.1016/j.apenergy.2018.05.023.
Oraiopoulos, A. & Howard, B. (2022). On the accuracy of urban building energy modelling. *Renew. Sustain. Energy Rev.*, *158*, 111976. https://doi.org/10.1016/j.rser.2021.111976.
Patil, G., Tiwari, C., Kavitkar, S., Makwana, R., Mukhopadhyay, R., Aggarwal, S., Betala, N., & Sood, R. (2025). Optimising Energy Efficiency in India: A sustainable energy transition through the adoption of District Cooling Systems in Pune. *Chall. Sustain.*, *13*(1), 1–17. https://doi.org/10.56578/cis130101.
Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy Build.*, *40*(3), 394-398. https://doi.org/10.1016/j.enbuild.2007.03.007.
Quan, S. J. (2024). Comparing hyperparameter tuning methods in machine learning based urban building energy modeling: A study in Chicago. *Energy Build.*, *317*, 114353. https://doi.org/10.1016/j.enbuild.2024.114353.
Wang, Z., Hong, T. Z., & Piette, M. A. (2019). Data fusion in predicting internal heat gains for office buildings through a deep learning approach. *Appl. Energy*, *240*, 386-398. https://doi.org/10.1016/j.apenergy.2019.02.066.

Nomenclature

ANN	Artificial Neural Network
BCR	Building Coverage Ratio, m^2/m^2
BD	Building Density, m^3/m^2
DHN	District Heating Network
DT	Decision Tree
DUE-S	Data-driven Urban Energy Simulation
E	East
EP	Energy Performance Index, $kWh/m^3/y$
GBDT	Gradient Boosting Decision Tree
GIS	Geographic Information System
GRU	Gated Recurrent Units
H_Havg	Ratio of the height of the building to the average height of surrounding buildings, m/m
H_W	Height-to-Width ratio of urban canyon, m/m
KNN	K-Nearest Neighbor
LGBM	Light Gradient Boosting Machine
LR	Logistic Regression
LSTM	Long Short-Term Memory
MAPE	Median Absolute Percentage Error, %
ML	Machine Learning
n	number, -
N	North
NHS	Net Heated Surface, m^2
ResNet	Residual Neural Network
RSME	Root Mean Squared Error (on energy performance of buildings), $kWh/m^3/y$
R^2	Coefficient of determination or R-squared, -
S	South
SHAP	SHapley Additive exPlanation
SLR	Simple Linear Regression
SVF	Sky View Factor, -
S/V	Surface-to-Volume ratio, m^{-1}
SVM	Support Vector Machine
TFT	Temporal Fusion Transformer
U	Thermal Transmittance, $W/m^2/K$
UBEM	Urban Building Energy Modeling
XGBoost	eXtreme Gradient Boosting
W	West