Machine-learning assisted optimization of residential building envelopes across contrasting Italian climates: energy, thermal comfort, and carbon footprint

Energy & Buildings 363 (2026) 117532

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Machine-learning assisted optimization of residential building envelopes across contrasting Italian climates: energy, thermal comfort, and carbon footprint

Alireza Ghomimoghadam [a,*], Mostafa Adibian [b], Luca Barbierato [c,e], Daniele Salvatore Schiera [d,e], Samad ME Sepasgozar [f]

[a] Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy
[b] Department of Environmental and Land Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy
[c] Department of Control and Computer Engineering, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy
[d] Energy Center Lab, Politecnico di Torino, via Paolo Borsellino 38/16, 10152 Turin, Italy
[e] Department of Energy (DENERG), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[f] School of Built Environment, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Keywords:
Building energy performance
Automation
CatBoost
Artificial neural networks (ANN)
Artificial intelligence (AI)
Carbon footprint
Sensitivity analysis

ABSTRACT

Energy efficiency, thermal comfort, and life-cycle carbon are central to residential design, yet practical automated workflows that scale across climates remain rare. This study aims to optimize residential envelopes in two contrasting Italian climates, identify cross-climate solutions that remain competitive in both contexts, and then re-rank the shortlist with life-cycle carbon (A1–A3, B6) to reveal energy–carbon interactions. This study develops and validates an automation pipeline around EnergyPlus that explores 14,400 envelope options and simulates 2,880 space-filling cases (Latin Hypercube with corner-point coverage) to model EUI, heating/cooling loads, and comfort (PMV/PPD). Eight algorithms are benchmarked as surrogates under a uniform tuning budget. CatBoost delivers the best fidelity for EUI and heating, while a tuned artificial neural network performs best for cooling and PMV; both achieve $R^2 > 0.96$ with low errors across targets. Re-simulation and Pareto-front identification yield five top-performing envelopes per city and reveal a robust overlapping subset that remains competitive in both climates. Although the most energy-efficient design (Wall 12) achieved top simulation scores, its high embodied carbon reversed its overall ranking; the Wall 14 configuration emerged as the most balanced low-carbon solution across climates, as confirmed by a component-level sensitivity analysis identifying walls as the dominant carbon contributors. The workflow demonstrates cross-climate portability by accurately predicting building performance in an unseen climate (Rome) using local weather descriptors and envelope U-values, incorporates boundary-aware sampling, identifies high-accuracy surrogate models for each key performance indicator (EUI, heating, cooling, PMV), and compares carbon-informed re-rankings with energy and comfort-based rankings of building envelopes.

1. Introduction

Buildings continue to be a major energy consumer globally and a source of carbon emissions, with residential housing making a substantial contribution because of ongoing needs for heating and cooling [1,2]. Early-stage envelope decisions regarding walls, roofs, floors, and glazing strongly influence a building's operational energy use, occupants' thermal comfort, and life-cycle carbon footprint [3,4]. However, current design guidance and surrogate modeling frameworks exhibit critical limitations.

Energy performance models are often developed in one climate and then applied elsewhere without validation, leading to degraded accuracy under different weather conditions [1,5]. Similarly, common simulation sampling methods (e.g., Latin Hypercube designs) rarely target extreme corner combinations, undermining model fidelity exactly where high-performance envelope solutions tend to lie [2,4]. Design workflows are often fragmented and rely on manual, GUI-driven steps, which hinder automation, reproducibility, and rigorous error checking

* Corresponding author.
E-mail address: alireza.ghomimoghadam@studenti.polito.it (A. Ghomimoghadam).

https://doi.org/10.1016/j.enbuild.2026.117532
Received 24 October 2025; Received in revised form 3 April 2026; Accepted 23 April 2026
Available online 30 April 2026

[6,7].

Finally, carbon is typically embedded as a concurrent optimization objective or analyzed at aggregated scales rather than used to test how carbon reshapes energy comfort shortlists. For example, He and Wang [3] formulate envelope optimization with simultaneous objectives for energy, embodied carbon, and cost, while Hasan et al. [8] estimate sectoral energy use and associated carbon at the national scale; neither examines whether adding life-cycle carbon (A1–A3, B6) reorders near-*iso*-energy, comfort-feasible envelope candidates across distinct climates. Data-driven operational-carbon models further clarify drivers but usually omit embodied stages and cross-climate re-ranking [4].

These gaps, specifically poor cross-climate transferability, a lack of boundary-aware sampling, non-automated design pipelines with minimal validation, and limited comfort and carbon integration motivate the present study. Accordingly, this research develops and evaluates an automation, uncertainty-aware surrogate modeling framework for early-stage residential envelope design. The approach is demonstrated in two climate zones in Italy (a heating-dominated northern zone and a cooling-prone southern zone) to test cross-climate robustness. A parametric library of realistic envelope options is simulated in EnergyPlus via a fully automated pipeline (with batch execution and failure-recovery) to ensure scalable, reproducible data generation.

The design space is sampled using Latin Hypercube designs augmented with explicit corner-point combinations, improving coverage of extreme envelope configurations. Surrogate models from multiple algorithm families (ensemble methods and neural networks) are trained under equal tuning conditions and benchmarked with standard accuracy metrics, including ASHRAE Guideline 14 criteria (CVRMSE and NMBE). We quantify model uncertainty and introduce physics-informed explainability via SHAP (Shapley Additive Explanations) to ensure surrogate predictions align with principles and expectations. In this study, the multi-objective analysis treats operational energy (site EUI for heating/cooling) and thermal comfort (PMV) as key decision criteria during surrogate-based screening. Within that set, comfort acceptability is verified via PPD. The resulting shortlist is then re-ranked using life-cycle carbon (A1–A3, B6) as a post-optimization filter to test whether carbon reshapes near-*iso*-energy, comfort-feasible envelopes and to assess cross-climate robustness. The trained pooled surrogate is subsequently applied to a third, unseen climate (Rome) using its local weather descriptors and envelope thermo-physical properties to verify cross-climate portability before carbon assessment for Turin and Naples.

Building on the above discussion, this study addresses four key research questions linked to the identified gaps. First, how stable are envelope surrogate models when transferred across markedly different climates (for example, from a heating-dominated region to a cooling-dominated one), and can they maintain accuracy under such domain shifts? Second, to what extent do rigorous, boundary-aware sampling strategies, including corner-point simulations, improve surrogate fidelity in the high-performance regions of the design space where critical trade-offs occur? Third, which machine learning algorithms (e.g., ensemble methods versus neural networks) offer the best trade-offs between accuracy and uncertainty under equal tuning conditions, and how should surrogate models be validated and selected using standardized metrics to ensure confidence in their guidance? Fourth, how can explainable surrogates and transparent Pareto-front analyses inform envelope design decisions that balance energy use, and how often do post-process checks on life-cycle carbon alter the ranking of top envelope solutions?

This study proposes a simulation-grounded, climate-aware workflow for early-stage building envelope design that integrates surrogate modeling, uncertainty-aware validation, and post-optimization decision analysis. Unlike many prior works that focus on predictive accuracy within a single climatic context, the proposed framework explicitly evaluates surrogate robustness across contrasting climates. Boundary-aware sampling is introduced to improve surrogate fidelity at performance extremes, where optimal envelope solutions typically reside. Rather than embedding carbon directly into the optimization objective, life-cycle carbon impacts are applied as a post-optimization re-ranking filter to test whether near-*iso*-energy, comfort-feasible envelope solutions change their ordering across climates. Through this structured combination of simulation, surrogate benchmarking, and decision-oriented re-ranking, the study provides transferable insights into how climate, uncertainty, comfort, and carbon jointly shape envelope design decisions.

2. Literature Review

Improving the energy performance of building envelopes is pivotal for achieving energy savings, enhancing indoor comfort, boosting overall efficiency, and advancing the development of nearly zero-energy buildings [9–11]. Most data-driven envelope models are developed and validated within a single climatic context, raising concerns about their applicability and reliability when transferred across different climates [1,5]. Even recent optimization work that considers climate effects tends to remain confined to one region or to only minor climate variations [5,12]. As a result, true cross-climate transferability of envelope designs remains largely unproven, motivating Aim 1 of this study: to evaluate the robustness of surrogate model performance across contrasting climatic contexts.

Design-space sampling plays a critical role in surrogate modeling, particularly when addressing boundary-aware sampling in high-dimensional envelope design spaces, where space-filling methods such as Latin Hypercube Sampling (LHS) are commonly adopted [4,13]. Despite their widespread use, LHS-based studies rarely include envelope parameter corner cases corresponding to extreme design combinations, leading to insufficient coverage of high-performance boundaries [2,13]. As a result, surrogate models often exhibit reduced surrogate fidelity at extremes, even if average predictive performance appears satisfactory. This lack of boundary-aware sampling motivates Aim 2 of this study, which focuses on improving surrogate fidelity at performance extremes by explicitly considering boundary cases.

Many building simulation workflows remain partially manual, limiting scalability and reproducibility when hundreds of simulations are required [6,14]. Recent studies therefore emphasize fully scriptable and automated pipelines [7], including BIM-enabled approaches by Najjar et al. [15] and data-driven optimization frameworks by Verma et al. [16], which have demonstrated substantial energy savings [17] and improved efficiency across design and operational phases [18,19]. However, surrogate validation practices remain inconsistent: while goodness-of-fit metrics such as RMSE or R^2 are commonly reported, standardized calibration criteria (e.g., ASHRAE CVRMSE and NMBE) and uncertainty-aware evaluation are often neglected. Recent work highlights the importance of quantifying prediction uncertainty, for instance, through quantile regression, to support more robust model assessment [20]. These limitations motivate Aim 3 of this study, which focuses on establishing an automated workflow coupled with standardized surrogate validation.

Researchers have explored a wide range of machine-learning algorithms as surrogate models for building performance [21,22], with data-driven approaches increasingly adopted to provide scalable and accurate solutions for building energy management [23,24]. Several studies have demonstrated the effectiveness of surrogate models for building energy prediction using different algorithms and sampling strategies, including quadratic regression by Lee and Quan [25], k-nearest neighbors by Liang et al. [21], and artificial neural networks by Ghalehkhondabi et al. [26], achieving high predictive fidelity across urban and residential applications [27]. Hybrid optimization–machine-learning approaches by Hasan et al. [8] and ensemble tree-based surrogates by Ahmad et al. [28] have shown strong predictive performance compared to single models and kernel-based methods, often achieving higher accuracy [29] and improved computational efficiency [30,31], although

no single algorithm consistently outperforms others [32]. Shirzadi et al. [22] evaluated surrogate models for building design, while Baqer and Rashidi-Khazaee [33] showed that XGBoost can achieve high fidelity in predicting residential heating and cooling loads; Ali et al. [34] further report that, in some extreme-climate cases, simpler models can deliver comparable accuracy with substantially lower computational cost. Meaningful comparison between surrogate-modeling studies remains challenging because results are sensitive to algorithm choice, input variables, and sample size, and because studies differ in modeling scope, outputs, and validation setups [7,22,27,35]. Data-driven methodologies employ machine-learning techniques as black-box models trained on historical data to evaluate system performance and adapt dynamically to specified objectives [36]. To enhance transparency and trust, several studies incorporate explainable artificial intelligence techniques, such as SHAP or LIME, to identify influential input features and support interpretable surrogate benchmarking [35,37]. Such interpretability can increase trust in the model if the highlighted influences align with known physics (for instance, confirming expected effects of insulation or solar gain on energy use).

Reviews by Duan et al. [38], and Li et al. [39], how that recent studies increasingly adopt multi-objective formulations that elevate thermal comfort and carbon considerations alongside energy performance, often relying on evolutionary algorithms such as NSGA-II and, in some cases, surrogate-assisted and explainable frameworks [37]. However, most prior work either co-optimizes carbon with energy in a limited manner [3] or estimates carbon impacts at aggregated scales [8], while operational-carbon analyses frequently omit embodied stages and cross-climate re-ranking effects [4]. Consequently, few studies explicitly examine whether incorporating life-cycle carbon alters the ranking of energy–comfort Pareto-optimal envelope solutions across different climates. This gap motivates Aim 4 of this study, which integrates comfort constraints and life-cycle carbon as post-optimization decision filters to reassess Pareto-optimal envelope designs and their cross-climate implications.

3. Methodology

This study employs a quantitative approach to evaluate, predict, and optimize the energy performance of a set of selected and simulated residential buildings in Italy. As depicted in Fig. 1, the methodology is divided into five phases: (1) model creation and data collection, (2) data generation with automation, (3) machine learning and model development, (4) model validation, and (5) carbon footprint analysis and final results. By combining architectural modeling, standardized data collection, automated energy simulations, advanced machine learning algorithms, multi-objective optimization, and Carbon footprint analysis, the research establishes a robust framework for identifying optimal building envelope configurations across different climate conditions.

3.1. Baseline architectural model and stratigraphy data collection

This section outlines the foundational methodology, detailing the development of the baseline architectural model in Revit and the compilation of material stratigraphies from the ACCA software catalog. At this stage, the baseline includes only the building geometry and architectural space definitions, and generic envelope surfaces. Operational and control assumptions required for EnergyPlus simulations are standardized later in Section 3.2.1. The architectural foundation for this study is a two-story, detached residential building archetype, representative of common Italian typologies, with a total Gross Area (GA) of 478.8 m^2 (Fig. 2). The building's envelope was characterized by a fixed Window-to-Wall Ratio (WWR), ranging from 3.4% to 24% across its facades. The model was exported to EnergyPlus IDF format via a BIM-to-IDF workflow, with architectural spaces that are geometrically enclosed. Envelope constructions were initially assigned as generic placeholders and later replaced with the curated stratigraphy alternatives directly in the IDF editor prior to running the automated pre-processing script (Section 3.2.1). To evaluate the optimization framework's robustness, two municipalities representing a significant climatic gradient were selected: Turin (Climate Zone E, cold continental) and Naples (Climate Zone C, temperate-warm Mediterranean). The contrasting climates, defined by their Heating Degree Days (DPR 412/93), provided a suitable testbed for climate-specific design strategies using Typical Meteorological Year (TMY) weather data. Finally, material and construction data were sourced primarily from ACCA Software's Edificius catalog (widely adopted in Italian practice), and all properties were compared to the Italian technical report UNI/TR 11552:2014, and then selected to define the discrete construction alternatives implemented in EnergyPlus. The detailed stratigraphy of these assemblies is reported in Table 10 in the Appendix.

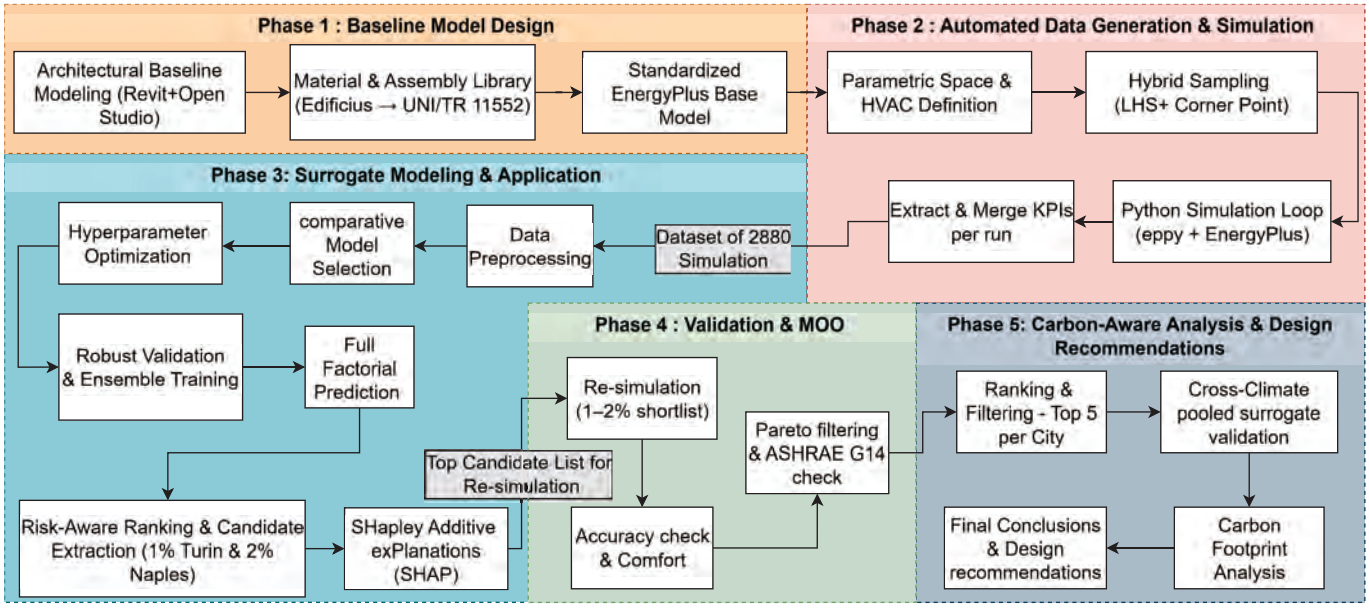


Fig. 1. Methodology workflow.

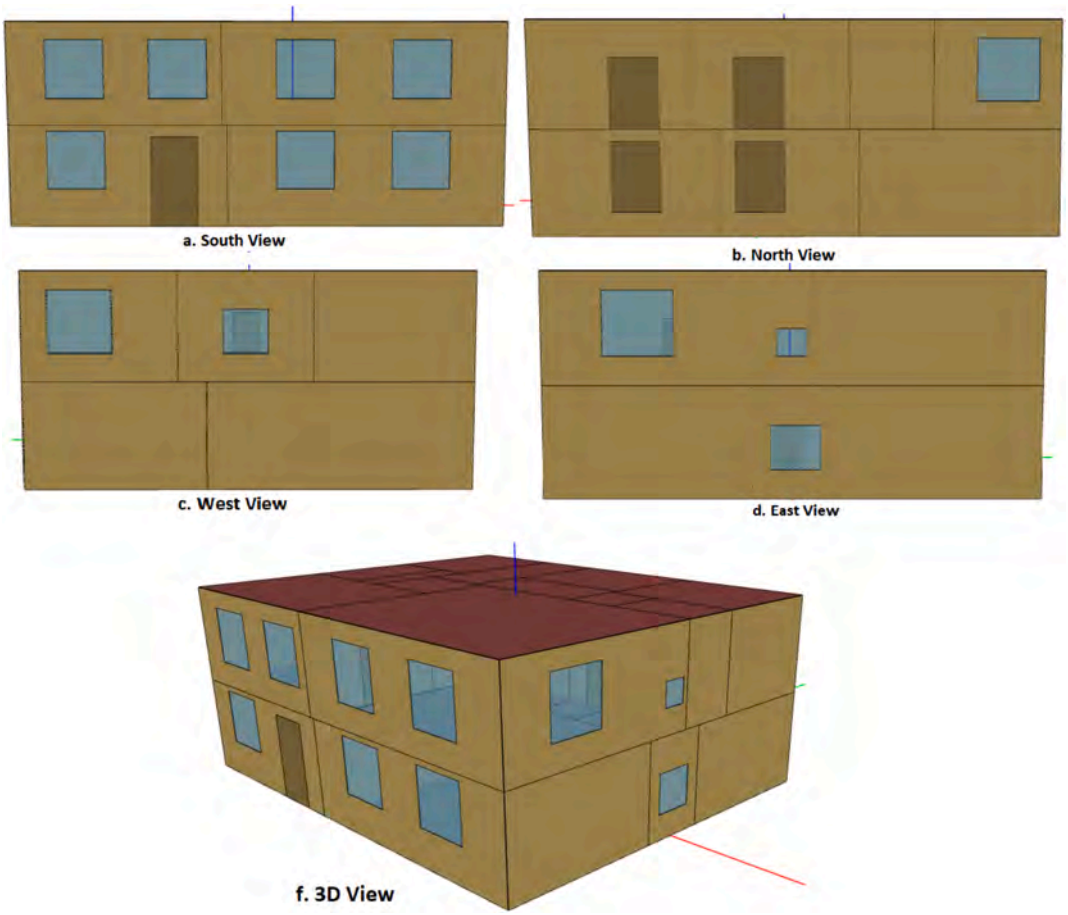


Fig. 2. Models of selected detached residential buildings in Italy.

3.2. Automation data generation & simulation

The large-scale evaluation of 2,880 unique building envelope configurations across climatic contexts required the development of an automated framework capable of consistently handling thousands of simulations. Implemented in Python, the framework was structured around two complementary scripts: a pre-processing routine that standardized the baseline model and defined uniform operational parameters and a main orchestration script that managed parametric sampling, batch execution of EnergyPlus runs, and automated extraction of multi-scale performance metrics. This modular structure ensured reproducibility, scalability, and resilience across the computational workflow. The overall logic of automated data generation and simulation is summarized in Fig. 3.

3.2.1. Simulation framework and pre-processing

To ensure reproducibility and scalability, the parametric workflow was supported by an automated simulation framework integrating model pre-processing and robust batch execution. The pre-processing stage, implemented by a dedicated Eppy-based script, first standardized the exported IDF model. The script enforced a high-granularity zoning strategy in which generic architectural spaces were mapped 1:1 to thermal zones, enabling zone-level comfort risk indicators to be evaluated at the finest available spatial resolution. Cryptic zone identifiers were converted into consistent labels, enabling reliable targeting of surfaces during parametric modifications. Occupancy and equipment schedules were defined in accordance with EN 16798–1:2019 (Annex C), a fixed clothing level of 0.5 clo in summer and 1.0 clo in winter was assumed, together with a metabolic rate of 119 W/person (Table 11). Regarding air exchange, a constant rate of 0.5 ACH was applied. This value is consistent with the operational infiltration and ventilation rates observed in the Italian building stock, which typically range from 0.5 to 1.0 ACH [40], and aligns with the requirements of the national standard UNI/TS 11300–1. The building was modeled with its main façade facing south to maximize passive solar gains. By idealizing the HVAC system, the study explicitly focuses on the building envelope's thermal performance, independent of specific HVAC system efficiencies. Consequently, each zone was equipped with an IdealLoadsAirSystem (20 °C heating, 26 °C cooling) with active dehumidification (max humidity ratio 0.015 kg/kg) to address latent loads.

The script also programmatically inserted internal gains and all necessary ScheduleTypeLimits, equipment lists, and output variables, yielding a uniform, simulation-ready baseline. Governing this standardized model, a modular automation framework orchestrated the execution of thousands of simulations. It encapsulated tasks, such as model editing, simulation calls, and result extraction, into discrete functions, thereby enhancing readability and reusability. EnergyPlus was executed through command-line calls, with results consolidated in a central CSV file for post-processing analysis. A built-in restart logic allowed interrupted batches to resume seamlessly by excluding already-completed cases, ensuring efficient use of computational resources. Simulations were conducted over a full annual period (8,760 h) with a 10-minute time step (6 per hour) to capture dynamic thermal inertia effects accurately. The framework included an intelligent orchestration layer with proactive input validation that screened invalid combinations prior to execution. As a result, all scheduled runs completed successfully, and no simulations were discarded due to runtime errors.

3.2.2. Hybrid sampling strategy and KPI extraction

Exploring the five-dimensional parameter space (20 walls, 4 interior

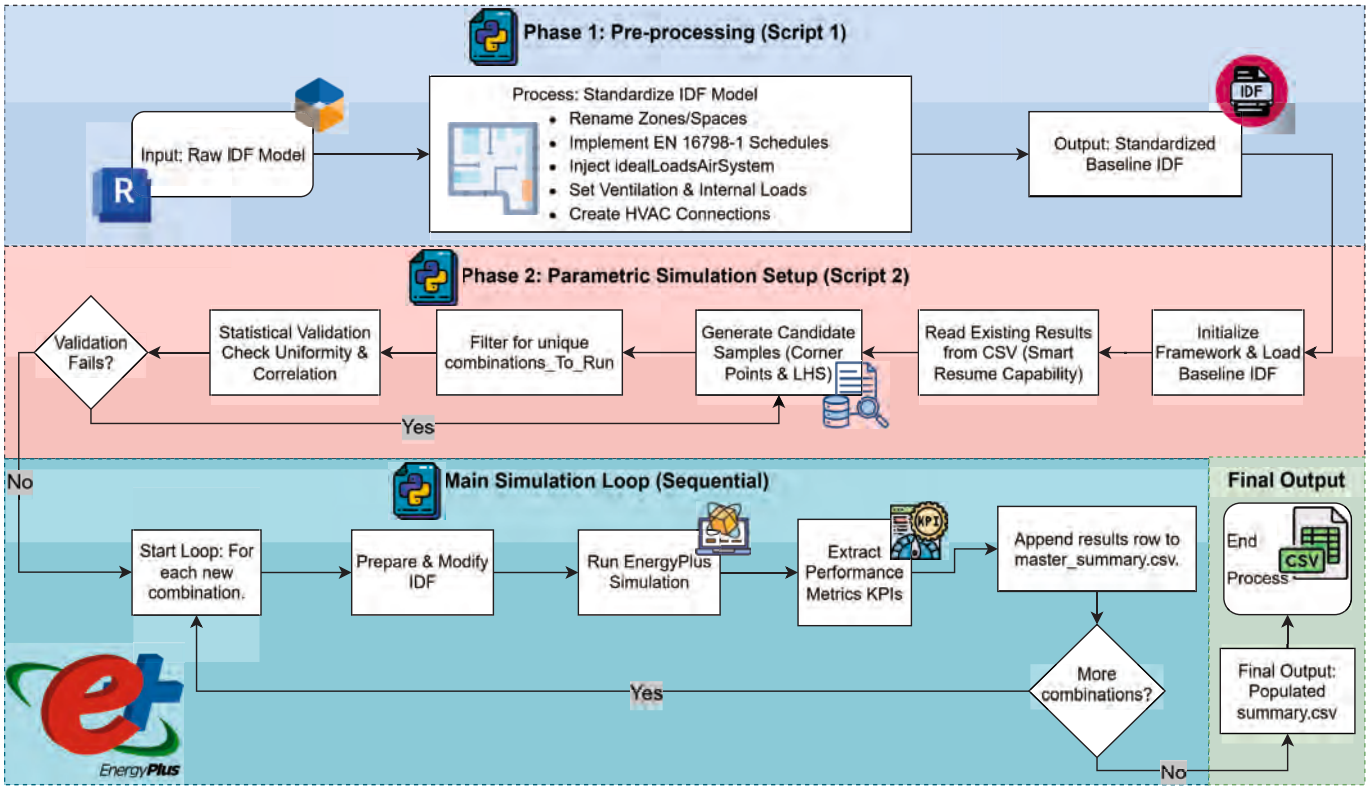


Fig. 3. Automation Data Generation & Simulation.

floors, 6 ground floors, 10 roofs, and 3 window options; 14,400 total combinations) through a full factorial design was computationally prohibitive. To balance coverage and feasibility, a hybrid sampling approach was adopted.

First, all corner point combinations were generated by systematically selecting the first and last levels of each parameter, ensuring representation of extreme design scenarios. Second, the interior of the design space was explored using Latin Hypercube Sampling (LHS), implemented with the scipy.stats.qmc. LatinHypercube engine, to provide a stratified, space-filling distribution of points. Together, these two methods produced 2,880 unique parameter sets. To enhance computational efficiency, the framework applied adaptive filtering: before launching new runs, the script compared candidate samples with the central summary file and excluded any already-completed cases, thereby enabling a "smart resume" function for interrupted batches.

Finally, each sample set was subjected to a statistical validation routine to confirm diversity and coverage. If a validation failed, the generator automatically produced a new candidate until the quality criteria were satisfied. This hybrid strategy ensured that the final dataset captured both boundary and interior behaviors of the design space, while the integrated validation and error-handling routines guaranteed robustness and efficiency across thousands of simulations. Following each simulation, an automated post-processing routine extracted key performance indicators (KPIs) and consolidated them into a master dataset. The extracted metrics included annual heating and cooling demand, whole-building Site Energy Use Intensity (EUI), and occupant comfort evaluated through the Fanger Predicted Mean Vote (PMV).

3.3. Surrogate modeling & application

The computational cost associated with high-fidelity EnergyPlus simulations prevents exhaustive exploration of the 14,400 possible envelope configurations. To overcome this, a representative subset of 2,880 configurations, which is 20% of the total design space, was selected for simulation instead of modeling all combinations, which would take weeks. A suite of surrogate models was developed and trained on this dataset and then deployed across the full 14,400-combination design space, enabling comprehensive analysis without additional high-fidelity simulations. The methodological framework for surrogate modeling comprised sequential stages: (i) comparative evaluation of multiple learning algorithms, (ii) hyperparameter optimization of the top performers, (iii) ensemble-based training and cross-validation to ensure robustness and quantify uncertainty, (iv) exhaustive deployment to the 14,400 combinations with multi-objective ranking, (v) interpretability analysis through SHAP, and (vi) re-simulation of selected candidates in EnergyPlus for final validation. This structured pipeline, shown in Fig. 4, provides a comprehensive overview of the entire methodological framework for the ML process. This approach ensured both predictive accuracy and scientific transparency, enabling surrogate models to serve as reliable proxies for physics-based simulations. The data usage and splitting strategy across all stages are summarized in Table 1.

Recent studies suggest that the train–test split should reflect the model's effective complexity; a commonly used heuristic is a ratio of $\sqrt{p}:1$ [43], where p denotes the effective number of input features. Here, categorical envelope descriptors were one-hot encoded, resulting in $p \approx 38$(excluding the intercept), which corresponds to an approximately 86/14 train–test split. For Stage 1 model-family screening, an 80/20 split was adopted to provide a more stringent holdout for comparative evaluation.

3.3.1. Model selection and comparative analysis

A comparative analysis of eight regression algorithms was conducted to identify the most suitable surrogates for four key performance indicators (KPIs): site Energy Use Intensity (EUI), total heating demand, total cooling demand, and average Predicted Mean Vote (PMV). The algorithm set covered a broad methodological range, including two linear baselines (Ordinary Least Squares Regression and ElasticNet), one

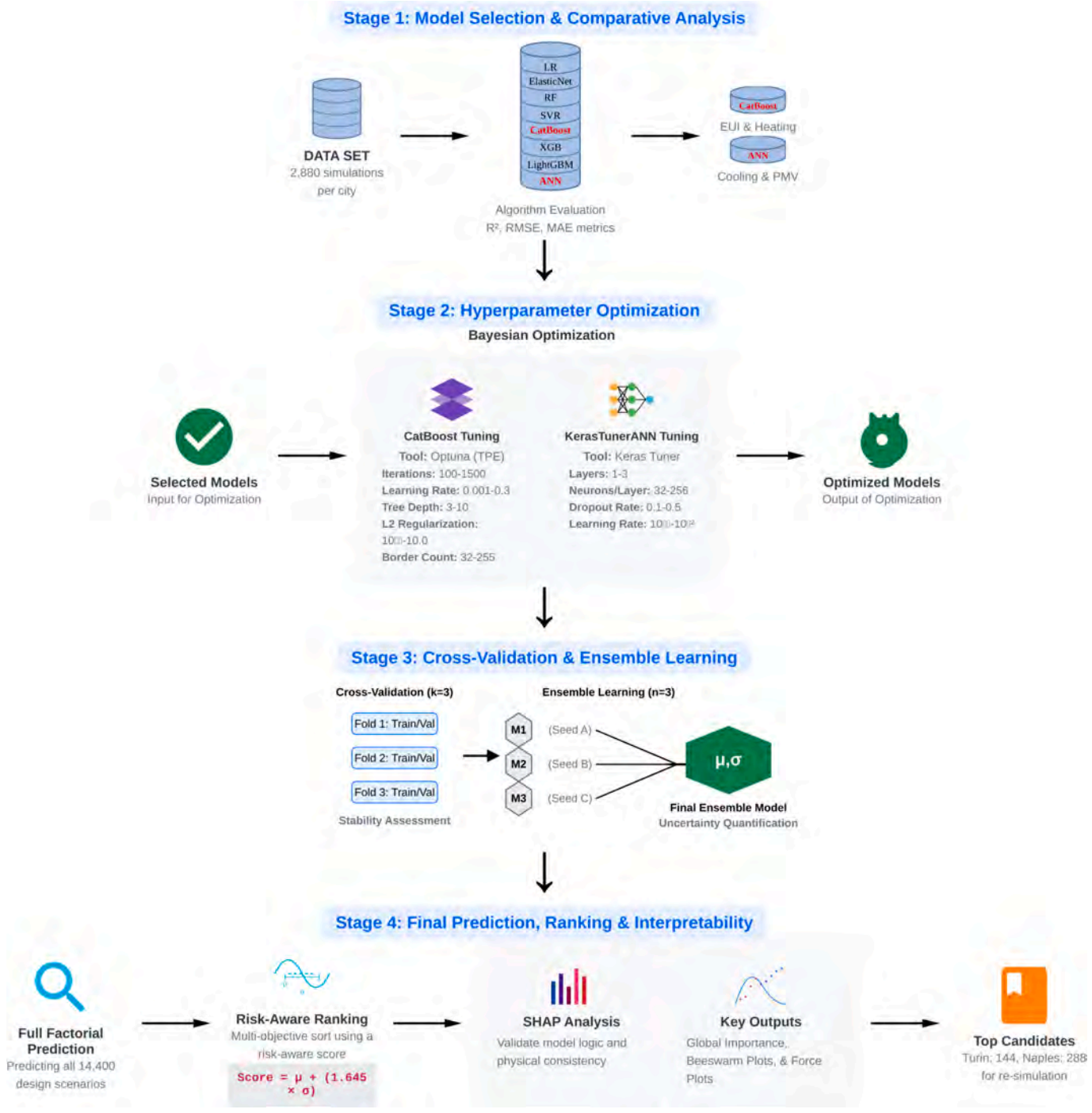


Fig. 4. The complete Machine Learning Methodology Workflow.

kernel-based method (Support Vector Regression, SVR), four ensemble tree-based algorithms (Random Forest, XGBoost, LightGBM, and CatBoost), and a fully connected Artificial Neural Network (ANN) representing deep learning. All models were trained on the same standardized dataset to ensure a fair comparison. Preprocessing involved one-hot encoding of categorical features and min–max scaling for algorithms sensitive to feature magnitude (e.g., SVR, ANN), while tree-based models used unscaled features. Samples were split into training and validation subsets separately for each climate (Turin and Naples) to evaluate potential context-specific performance. Each algorithm was tuned under an equal hyperparameter search budget. Model performance was evaluated using the Coefficient of Variation of the Root Mean Square Error (CVRMSE, %), computed on the validation set for each KPI and each climate. This single, standardized metric enabled a consistent comparison across models and performance targets.

The comparative results are presented in Section 4.2 and summarized in Table 2 and Table 3.

3.3.2. *Hyperparameter optimization strategy*

Recognizing that surrogate model accuracy is highly sensitive to hyperparameter choices, tailored optimization strategies were designed for CatBoost and ANN. The ANN was tuned using Bayesian optimization with the KerasTuner library. The search explored architectural depth (1–3 layers), neurons per layer (32–256), dropout rate (0.1–0.5), and

Table 1
Data-splitting strategy across the four-stage workflow (Turin, Naples).

Stage (Fig. 4)	Purpose	Data source	Turin / Naples	Rome	Output used in the next stage
Input data	DOE ground truth for surrogate modelling	EnergyPlus DOE simulations	2880	–	Candidate pool for all learning steps
Stage 1 *	Model selection & comparative analysis	DOE pool	2,304 train + 576 holdout	–	Selected model family per KPI (e.g., CatBoost vs ANN)
Stage 2	Hyperparameter optimization (Bayesian)	DOE pool (internal resampling)	80/20 split per Optuna trial; KerasTuner uses validation_split = 0.2 during search	–	Optimized hyperparameters per target/model
Stage 3	Cross-validation (K = 3) & ensemble learning (n = 3)	DOE pool (rotating folds) + full refit	~1,920/~960 per fold; then full fit on 2,880 / same	–	Final ensemble surrogate (mean μ and uncertainty σ)
Stage 4	Full-factorial prediction & ranking	Design space (surrogate inference only)	14,400 predictions / 14,400 predictions	14,400 predictions	Ranked candidates (risk-aware score, SHAP/interpretability)
Stage 4 * (Holdout)	Targeted independent validation (ASHRAE G-14)	New EnergyPlus re-simulations	144 re-simulated / 288 re-simulated	144 re-simulated	Reported final validation metrics
Stage 5	Cross-climate validation in an unseen city (Rome)	Rome EPW-derived climate descriptors (HDD_{18}, CDD_{26}, annual GHI/DHI) + identical envelope-variable schema; design grid)	–	Uses Stage 4: 14,400 inferred + Stage 4 hold-out: 144 re-simulated	External (unseen-city) performance confirmation; in-domain (interpolation) in climate-feature space

*Holdout in Stage 1 is used only for algorithm screening and should not be interpreted as the final test set. The independent hold-out is created by re-simulating the Top-K (Fig. 13) selected in Stage 4 of Fig. 4.

Table 2
Comparative Performance of Machine Learning Models for the Turin Climate.

Model	CVRMSE Site EUI (%)	CVRMSE Heating (%)	CVRMSE Cooling (%)	CVRMSE PMV (%)
CatBoost	2.63	3.25	6.22	7.72
KerasTunerANN	4.01	7.05	4.38	6.09
LinearRegression	54.31	53.54	24.955	25.18
XGBoost	8.47	13.97	8.89	12.90
LightGBM	11.27	18.87	13.92	17.47
SVR	19.40	34.73	13.58	27.21
RandomForest	29.89	48.05	24.90	37.77
ElasticNet	61.70	112.65	81.21	275.36

Table 3
Comparative Performance of Machine Learning Models for the Naples Climate.

CatBoost	3.11	4.46	8.30	13.55
KerasTunerANN	4.58	9.17	6.57	10.20
LinearRegression	8.48	10.5	11.75	22.22
XGBoost	9.20	16.45	13.70	24.93
LightGBM	13.14	14.16	32.18	52.02
SVR	20.19	49.17	17.22	46.06
RandomForest	24.17	36.27	25.33	52.02
ElasticNet	68.11	158.55	81.85	482.84

learning rate (10^{-4}–10^{-2}), with early stopping callbacks (patience = 15 epochs) to prevent overfitting. The full hyperparameter ranges and sampling strategies are reported in Annex Table 14. CatBoost models were optimized using Optuna's Tree-structured Parzen Estimator (TPE). Five critical hyperparameters, iterations, learning rate, tree depth, L2 regularization, and border count were simultaneously tuned, with Optuna's pruning strategy terminating unpromising trials early. This yielded regressors with strong predictive accuracy while minimizing computational burden. The complete search space and parameter ranges are also documented in Annex Table 14. To quantify epistemic uncertainty, ensembles of three models were trained for each target variable. For CatBoost, ensemble diversity arose from random initialization of feature sampling; for ANN, diversity was induced by varying random seeds and weight initialization. Ensemble predictions were averaged to yield mean estimates, while the standard deviation across models served as a proxy for uncertainty. This approach provided not only improved predictive accuracy but also explicit confidence measures, essential for risk-aware candidate ranking in later stages.

3.3.3. *Cross-Validation and ensemble learning framework*

Ensuring that the surrogate model generalization beyond the training dataset was a central requirement. A three-fold cross-validation protocol was employed during optimization. Each dataset was split into three equal folds, with training and validation subsets across runs. For each model, mean values and 95% confidence intervals of R^2 and RMSE were calculated. Ensemble CatBoost models typically achieved R^2 values above 0.97 with narrow confidence intervals (<0.01), while ANN models maintained R^2 values of 0.94–0.96 across targets. The tight variance confirmed that performance was stable and not dependent on a particular data split. Cross-validation also provided estimates of uncertainty bounds for surrogate predictions. For configurations with similar mean errors but higher variance across folds, confidence intervals widened, signaling areas of potential instability. These insights directly informed the subsequent risk-aware ranking strategy.

3.3.4. *Multi-Objective ranking and candidate selection*

After training and tuning the selected surrogates, the full set of envelope combinations was evaluated in both climates. For each configuration and performance target, the ensemble provided a predictive mean and μ an uncertainty estimate σ, capturing both expected value and uncertainty. Candidate selection then applied a lexicographic ranking scheme aligned with practical design priorities. Site Energy Use Intensity (EUI) was treated as the leading objective, reflecting its central role in policy targets and building codes. Heating and cooling demands followed as climate-dependent sub-objectives, with heating given higher priority in Turin's cold context, and cooling in Naples' mixed, summer peaking climate. Absolute PMV served as the final criterion, during surrogate-based shortlist ranking, ensuring that shortlisted designs preserved comfort before re-simulation. The final Pareto analysis was then conducted only on the re-simulated energy KPIs (site EUI, total heating, and total cooling), with PPD used as a comfort-feasibility check rather than as an optimization axis. Ranking proceeded hierarchically. Configurations were first ordered by EUI; any ties were resolved using the next objective in sequence. To incorporate model uncertainty, each objective was expressed as a risk-adjusted value based on Eq. (1). This adjustment penalized options with high epistemic uncertainty, favoring designs whose performance was both strong and dependable. To determine the optimal verification budget, K was selected via a

sensitivity–stability analysis of the surrogate ranking. Ensemble predictions were perturbed with bootstrap noise and Top-K set stability was quantified using the Jaccard index, and marginal changes in a composite performance score were monitored against computational cost. K was chosen at the earliest stability plateau and point of diminishing returns; see Fig. 13 in the Appendix, identifying the final set of high-confidence candidates advanced to detailed EnergyPlus verification.

For each envelope design vector x evaluated under a given climate (city) c, i.e., $(x, c) \in \mathscr{D}$, where $\mathscr{D}$ denotes the feasible full-factorial envelope design grid (14,400 configurations), the ensemble surrogate predicts each KPI k by reporting its mean response $\mu_k(x;c)$ and the associated predictive uncertainty $\sigma_k(x;c)$ (standard deviation) across ensemble members. Here, λ is a risk-aversion weight that controls the penalty on predictive uncertainty relative to the mean. It is set to the 95th percentile of the standard normal distribution ($\lambda \approx 1.645$), where Φ denotes the cumulative distribution function (CDF) of the standard normal. This yields an approximate 95% upper confidence bound under a normality assumption.

Designs are ranked using the conservative risk-adjusted objective defined in Eq. (1):

$$\widetilde{f}_k(x;c) = \mu_k(x;c) + \lambda\,\sigma_k(x;c),\ \lambda = \Phi^{-1}(0.95) \approx 1.645 \tag{1}$$

Comfort is handled as an absolute deviation objective, as defined in Eq. (2):

$$\widetilde{f_{|PMV|}}(x;c) = |\mu_{PMV}(x;c)| + \lambda\,\sigma_{PMV}(x;c) \tag{2}$$

City-specific lexicographic ranking is then applied. For the heating-dominated climate of Turin, the minimization order is defined in Eq. (3).

$$\min_{\text{lex}} \left[\widetilde{f_{site\ EUI}}(x;c), \widetilde{f_{Heating\ load}}(x;c), \widetilde{f_{Cooling\ load}}(x;c), \widetilde{f_{|PMV|}}(x;c)\right], x \in \mathscr{D} \tag{3}$$

For the cooling-priority climate of Naples, the lexicographic order is modified as shown in Eq. (4).

$$\min_{\text{lex}} \left[\widetilde{f_{site\ EUI}}(x;c), \widetilde{f_{Cooling\ load}}(x;c), \widetilde{f_{Heating\ load}}(x;c), \widetilde{f_{|PMV|}}(x;c)\right], x \in \mathscr{D} \tag{4}$$

3.3.5. Model Interpretability: SHapley Additive exPlanations (SHAP)

Model interpretability was used to verify that the selected surrogates encode physically meaningful relationships. We applied SHapley Additive exPlanations (SHAP) to both model families. For the tree-based CatBoost, we used TreeExplainer; for the ANN, we used KernelExplainer with background samples drawn from the training distribution. Two levels of interpretation were generated. At the global level, feature importance was summarized by the mean absolute SHAP value per input, and distributional effects were visualized with beeswarm plots to indicate how variations in each parameter shift predictions across the dataset. At the local level, force plots were produced for representative cases (e.g., best, median, and challenging configurations within the validation folds or the ranked shortlist) to decompose each prediction into additive feature contributions. All SHAP analyses were computed on the same preprocessed inputs used for training and evaluated per climate. The outputs served two purposes: (i) a physics-consistency check that surrogate attributions align with first-principles expectations, and (ii) decision support by clarifying which envelope inputs most influence each KPI.

3.4. Validation Framework, Multi-Objective optimization

Following surrogate model development and interpretability analysis, a final validation phase was conducted to ensure predictive reliability and to establish the set of optimal design solutions. Phase 4 (Fig. 1) creates a shortlist of top-ranked candidates from surrogate predictions, re-simulates these candidates in EnergyPlus, evaluates comfort metrics, identifies Pareto-optimal solutions, and performs accuracy checks consistent with ASHRAE Guideline 14.

3.4.1. Validation workflow and Pareto-Optimal candidate identification

In this validation stage, surrogate-based ranking is used to select high-confidence candidates for EnergyPlus verification. The selected candidates were then re-simulated in EnergyPlus under an identical automated workflow to ensure methodological consistency, and a dedicated post-processing script extracted key performance indicators (KPIs) to assemble a high-fidelity dataset for final validation.

Thermal comfort metrics were then finalized on this dataset. PMV was taken directly from the EnergyPlus outputs. PPD was derived from PMV using the ISO 7730 / ASHRAE 55 (Eq. (5)):

$$\text{PPD} = 100 - 95\cdot\exp\left(-0.03353\cdot\text{PMV}^4 - 0.02179\cdot\text{PMV}^2\right) \tag{5}$$

Within the shortlist, PPD served as a feasibility check rather than an optimization axis. Pareto fronts were identified on the re-simulated EnergyPlus dataset across the primary energy objectives (site EUI, total heating, and total cooling) by standard dominance comparisons. A design was retained as non-dominated if no alternative achieved equal or better values in all objectives with a strict improvement in at least one. The resulting non-dominated sets were flagged and prioritized within the dataset and provided the basis for the multi-criteria analysis reported in Section 4. Predictive accuracy was assessed under ASHRAE Guideline 14–2014 using a re-simulated holdout excluded from training and tuning. For each climate, Normalized Mean Bias Error (NMBE) and the Coefficient of Variation of RMSE [CVRMSE] were computed for site EUI, total heating, and total cooling via automated post-processing.

Here, c indexes the city, and $\mathscr{D}_c$ denotes the city-specific re-simulated shortlist (with $N_{\text{Turin}} = 144$ and $N_{\text{Naples}} = 288$). Each $x^{(i)} \in \mathscr{D}_c$ is a single envelope design in that shortlist. Performance is represented by the objective vector $f(x)$ in Eq. (6), with components $f_m(x)$ for $m \in \{1, 2, 3\}$ corresponding to site EUI, heating load, and cooling load, all minimized. The symbol $\prec$ denotes standard Pareto dominance within $\mathscr{D}_c$: a design dominates another if it is no worse in all objectives and strictly better in at least one. The Pareto set $\mathscr{P}_c$ is defined as the subset of non-dominated designs in $\mathscr{D}_c$, and Is_Pareto(x) is a binary indicator equal to 1 for $x \in \mathscr{P}_c$ and 0 otherwise.

$$\begin{gathered}
\min_{x \in \mathscr{D}_c} \boldsymbol{f}(x) = [f_1(x), f_2(x), f_3(x)] \\
\mathscr{D}_c = \{x^{(i)}\}_{i=1}^{N_c},\ x^{(i)} \in \mathscr{D}_c, c \in \{\text{Turin}, \text{Naples}\} \\
x^{(j)} \prec x^{(i)} \Leftrightarrow \left(\forall m \in \{1,2,3\} : f_m\left(x^{(j)}\right) \le f_m\left(x^{(i)}\right)\right) \wedge \left(\exists m \in \{1,2,3\} : f_m\left(x^{(j)}\right) < f_m\left(x^{(i)}\right)\right) \\
\mathscr{P}_c = \left\{ x^{(i)} \in \mathscr{D}_c \middle| \neg\exists x^{(j)} \in \mathscr{D}_c : x^{(j)} \prec x^{(i)} \right\} \\
\text{Is_Pareto}(x) = \begin{cases} 1, x \in \mathscr{P}_c \\ 0, \text{otherwise} \end{cases}
\end{gathered} \tag{6}$$

The overall validation and decision workflow proceeds through the following sequential steps:

(1) surrogate-based screening via risk-aware lexicographic ranking, followed by the selection of Top-K candidates (detailed in Section 3.3.4); (2) re-simulation of these candidates in EnergyPlus; (3) Pareto analysis and PPD checking based exclusively on the re-simulated KPIs; (4) selection of the Top 5 Pareto-optimal designs in each climate; and (5) life-cycle carbon assessment and final re-ranking.

3.4.2. Cross-climate pooled surrogate validation in an unseen city (Rome)

This study developed a climate-aware pooled surrogate to evaluate design-space performance in a new climatic context (Rome). The surrogate ensemble (CatBoost + ANN) was trained on pooled data from two contrasting climates, Turin (cold-temperate) and Naples (warm-temperate) under an identical feature schema. Inputs included (i) macro-level climate descriptors extracted from EPW weather files (HDD_1^8, CDD_2^6, annual GHI/DHI) and (ii) thermo-physical envelope properties (U-values for walls, floors, roofs, and windows). After training, the pooled surrogate was applied to Rome, using Rome's EPW-based climate descriptors together with the same set of envelope variables. Predictions were generated for the entire 14,400-case design grid. To assess real design-decision reliability, a risk-aware ranking (mean uncertainty penalty) was applied, and the top 144 high-performance designs were then re-simulated in EnergyPlus to provide a ground-truth reference at the optimal frontier. This targeted validation confirms the surrogate's accuracy exactly where performance differences are most critical, avoiding the cost of full-grid simulation. Rome's HDD, CDD, and GHI values lie within the range spanned by Turin and Naples; thus, this evaluation represents interpolation in climate-feature space rather than extrapolation. Accuracy was assessed using R^2, RMSE, MAE, and CVRMSE/NMBE, consistent with in-climate validation. Additionally, the monthly dry-bulb profile of Rome remains within the Turin–Naples envelope (Fig. 5), supporting joint feature-space coverage.

3.5. Carbon footprint analysis

The carbon footprint assessment integrates both embodied and operational emissions to evaluate the life-cycle implications of envelope design choices. The boundary case is defined according to EN 15978 and includes embodied carbon associated with material production (modules A1–A3) and operational energy-related emissions (B6), while excluding construction (A4–A5), other use-phase processes (B1–B5, B7), and end-of-life stages (C1–C4). This boundary captures the dominant emission sources (73%) and ensures consistency in the comparative evaluation across design alternatives (Based on Fig. 6).

Embodied carbon factors were obtained from the Inventory of Carbon and Energy (ICE database V4.0, Dec 2024) [41], which provides emission coefficients derived from Environmental Product Declarations (EPDs) and is widely used for benchmarking building material impacts. Operational emissions were calculated using the Italian national electricity emission factor of 0.276 $kgCO_2e/kWh$, based on the ISPRA (Italian Institute for Environmental Protection and Research) National Inventory Report (2023) [42]. These sources were selected to ensure

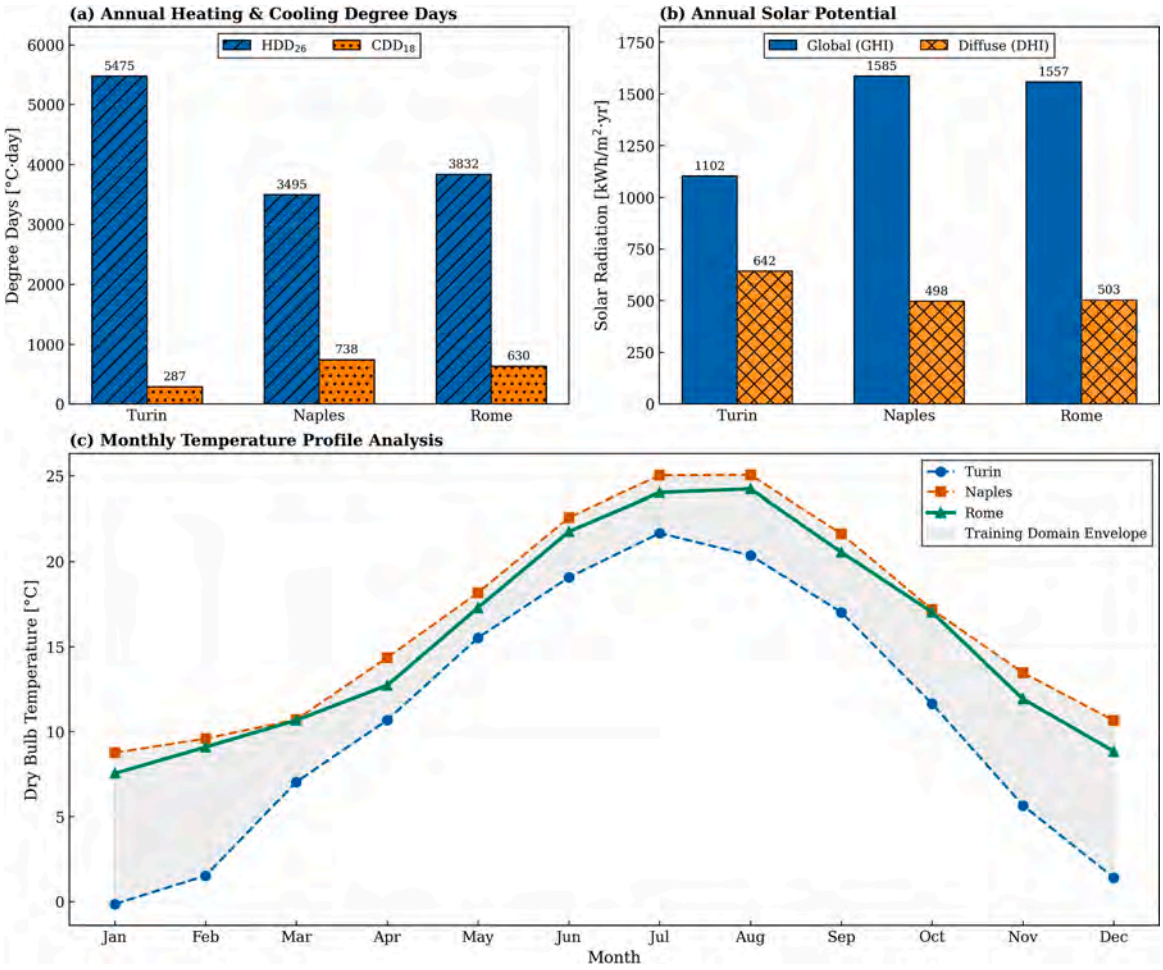


Fig. 5. Climatic analysis for the cross-climate approach: (a) Annual heating and cooling degree days; (b) Solar radiation potentials; and (c) Monthly temperature profiles demonstrating Rome's interpolation within the Turin-Naples envelope.

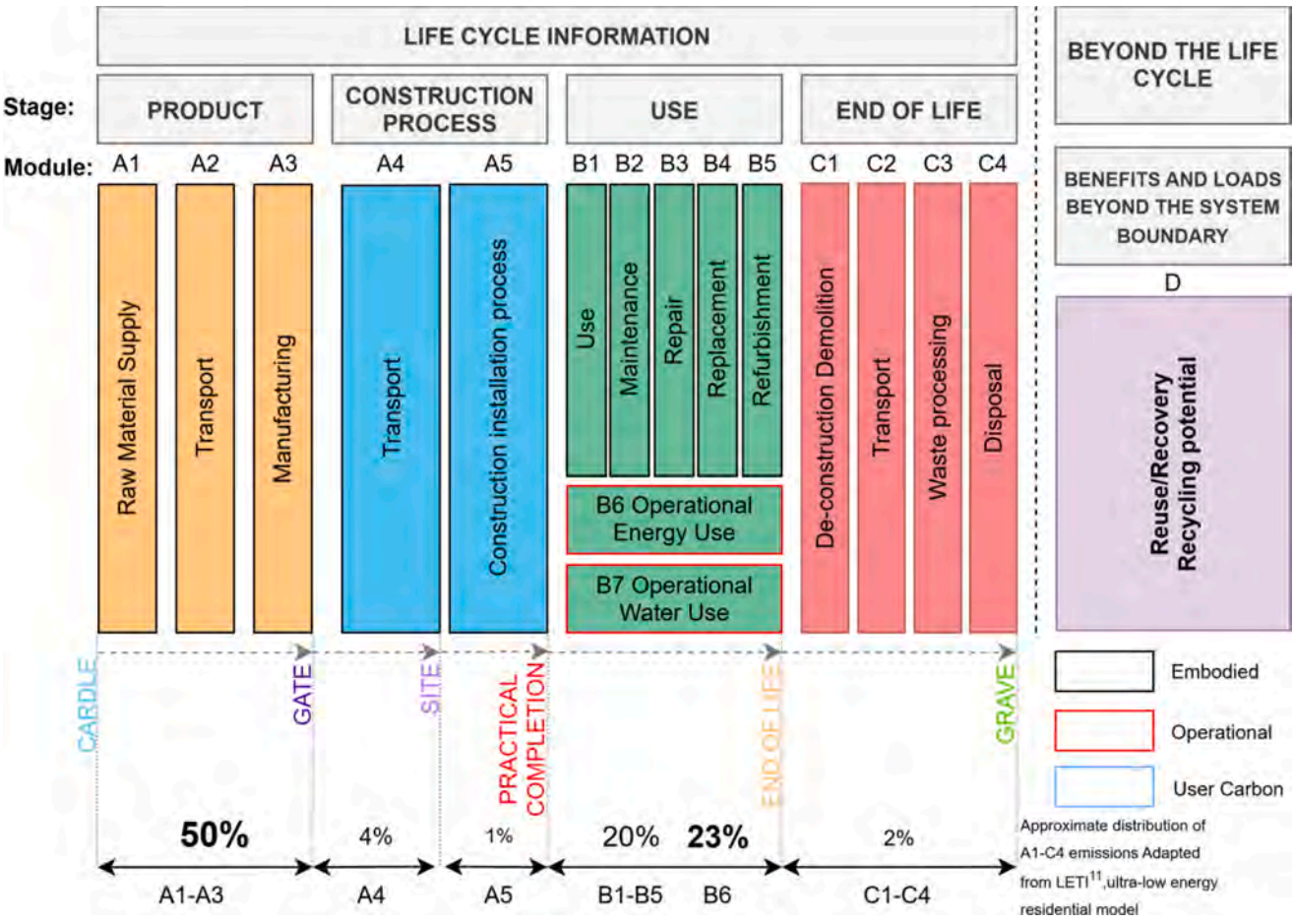


Fig. 6. Life cycle stages defined in BS EN 15978, with an illustrative distribution of emissions across Modules A1–C4 for a medium-scale residential building; operational carbon modules are highlighted in red Ref: The Institution of Structural Engineers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

methodological consistency and regional relevance, with ICE providing standardized and widely validated embodied carbon coefficients, and ISPRA offering nationally representative electricity emission factors aligned with the Italian context.

Carbon assessment was applied only to the re-simulated top five Pareto-optimal design solutions in each climate. This post-optimization application ensures that carbon does not influence the initial surrogate-based screening but instead acts as a secondary evaluation criterion to examine how life-cycle emissions alter the ranking of energy-based design solutions.

Embodied Carbon Estimation

Embodied carbon was calculated by multiplying the mass of each material by its corresponding emission factor ($kgCO_2e/kg$) for the selected life-cycle modules, as shown in Eq. (7).

$$\text{Embodied carbon } (kgCO_2e) = \text{material quantity } (kg) \times \text{carbon factor } (kgCO_2e/kg) \quad (7)$$

Carbon factors were assigned to each envelope component based on representative construction typologies and material compositions.. The analyzed assemblies include a range of configurations, from lightweight timber-based systems to high-mass double-insulated constructions (e.g., Wall 17), capturing variability in both thermal performance and material intensity.

Operational Carbon Estimation

Operational carbon emissions were calculated by converting simulated site energy use intensity (EUI) into equivalent emissions using the Italian national electricity emission factor of 0.276 kg CO_2e/kWh, based on a four-year average of national grid emission factors reported in the ISPRA National Inventory Report (2023). Operational carbon is not treated as an optimization objective, but is evaluated in conjunction with embodied emissions to quantify the total life-cycle impact and to support post-optimization comparisons of design alternatives.

4. Results

This section reports the study's findings across the main stages of the analysis pipeline. It first verifies the design-space representativeness of the reduced experiment using the LHS + Corner scheme by comparing the sampled and full-factorial categorical distributions via Total Variation Distance (TVD). This establishes that subsequent modeling rests on a low-bias dataset. Next, it presents a comparative benchmarking of surrogate models across the target variables in the two climates, using CVRMSE as the principal accuracy metric, and identifies the top-performing surrogates adopted for downstream use. To interpret model behavior, a SHAP analysis is then reported, highlighting climate-dependent feature effects that are physically consistent with building heat transfer mechanisms. Model credibility is further addressed through formal validation against accepted criteria (e.g., ASHRAE Guideline 14), confirming robust calibration across climates. With these foundations, the section proceeds to multi-objective optimization outcomes, reporting the top envelope configurations per city and discussing trade-offs among energy use, heating, cooling loads, and comfort. Finally, a life-cycle carbon assessment integrates embodied and operational impacts, complemented by sensitivity checks on envelope subsystems, to derive cross-climate design insights and actionable

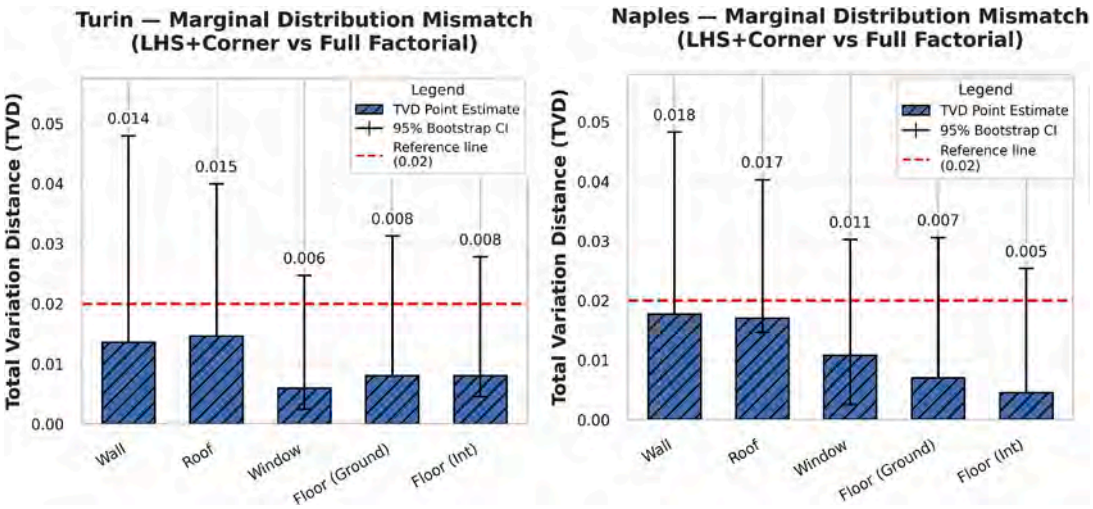


Fig. 7. TVD between the LHS + Corner sample (n = 2,880) and the full-factorial space (n = 14,400) for Turin and Naples, computed on marginal categorical distributions of five envelope variables. Bars show TVD point estimates; error bars are 95% bootstrap confidence intervals. The dashed line (0.02) is a reference for practical tolerance.

recommendations.

4.1. LHS + Corner coverage (TVD)

Fig. 7 quantifies the representativeness of an LHS + Corner sampling (2,880 runs) against the full-factorial design space (14,400 combinations) for Naples and Turin. For each city, bars report on the Total Variation Distance (TVD) between the marginal categorical distributions in the sample and in the full-factorial space for five envelope variables. TVD is a standard distributional metric on [0,1] (0 = identical distributions); For categorical variables, TVD upper-bounds the absolute deviation in marginal probability for any single category; hence TVD = 0.02 implies at most a 2-percentage-point deviation in any category's marginal frequency (≈ 0.02 × 2,880 ≈ 58 observations in expected count). Across both cities, all marginal TVD values remain below 0.02 (Naples: 0.018–0.005; Turin: 0.015–0.006). To avoid relying on an arbitrary cut-off, TVD was calibrated against a random-sampling baseline at the same budget (N = 2,880) by generating 5,000 uniform random subsets from the full-factorial space (Table S1, Table S2, Fig. S21, Fig. S22); this calibration reveals that finite-sample noise alone typically yields TVD values significantly higher than the observed values. Specifically, the observed global worst-case TVD (0.0146) is substantially lower than the random sampling noise ceiling (95% FWER≈0.038), confirming that the LHS + Corner strategy achieves coverage systematically superior to random chance.

Finally, to address potential blind spots in joint parameter interactions, pairwise co-occurrence heatmaps are reported for all envelope-variable pairs in the Supplementary Material (Figs. S1–S20), confirming that all pairwise category combinations are observed, including extreme pairings (e.g., most insulating walls with least insulating windows). Overall, these results demonstrate that the strategy preserves the design space statistics with high fidelity while reducing the simulation burden by approximately 80%.

4.2. Comparative model performance and selection

This section presents the performance metrics of the eight machine learning models evaluated for each of the four target variables across the two climatic contexts. These results, summarized in Table 2 (Turin) and Table 3 (Naples), form the basis for selecting the top-performing models used in the subsequent phases of the study. Across both climates, predictive accuracy was assessed using the Coefficient of Variation of the Root Mean Square Error (CVRMSE), which reflects each model's ability to approximate the ground-truth outputs from EnergyPlus simulations. Across both climates, the gradient boosting models (CatBoost, XGBoost, LightGBM) and the deep learning model (KerasTunerANN) demonstrated superior fidelity, achieving the lowest CVRMSE values among all candidates. In contrast, traditional linear regressors and single-tree ensembles showed substantially higher error levels, indicating limited capacity to capture nonlinear interactions among envelope parameters.

Final model selection followed a minimum CVRMSE criterion applied per target. Based on this evaluation, CatBoost achieved the best performance for site EUI and total heating, while the KerasTunerANN outperformed others in total cooling and PMV prediction. These two surrogates, CatBoost for energy-related targets and ANN for comfort-related targets, were therefore adopted for the optimization and validation phases.

4.3. Comparative SHAP analysis results

The SHAP analysis provides a clear, climate-dependent hierarchy of feature importance, indicating that the surrogate models capture the underlying building physics. The results are reported with emphasis on the Site EUI for Turin and Naples to highlight the climate adaptability of the models. The bar charts (Fig. 8a; Fig. 9f) show the ground-floor and roof assemblies as the principal contributors to Site EUI in the surrogates, with walls and windows exerting comparatively smaller effects. This underscores the critical role of heat transfer through the ground and roof across climates. However, the beeswarm plots, (Fig. 8b; Fig. 9g), reveal a crucial difference in the sign of these effects between the two climates. In Turin, a heating-dominated climate, high-thermal-mass ground-floor constructions (Floorground3 and 1) and single-clear glazing exhibit positive SHAP values. This is consistent with greater transmission losses and higher heating demand, and hence a higher Site EUI. Conversely, in Naples, a mixed-mode climate, these same ground-floor constructions show negative SHAP values, indicating a reduction in the predicted Site EUI. This reversal indicates that the surrogates capture the beneficial role of thermal capacity under warm conditions, where it helps mitigate cooling loads. The force plots further illustrate this phenomenon by showing how specific combinations of components lead to a given prediction. In both cities, the model's predictions arise from the collective contributions of all features, where a detrimental impact from one component (e.g., a low-performing window) can be offset by a beneficial one (e.g., a high-performance roof). The difference lies in which components increase or decrease the predicted Site EUI, a relationship that reverses between the two climatic contexts, (Fig. 8c–e;

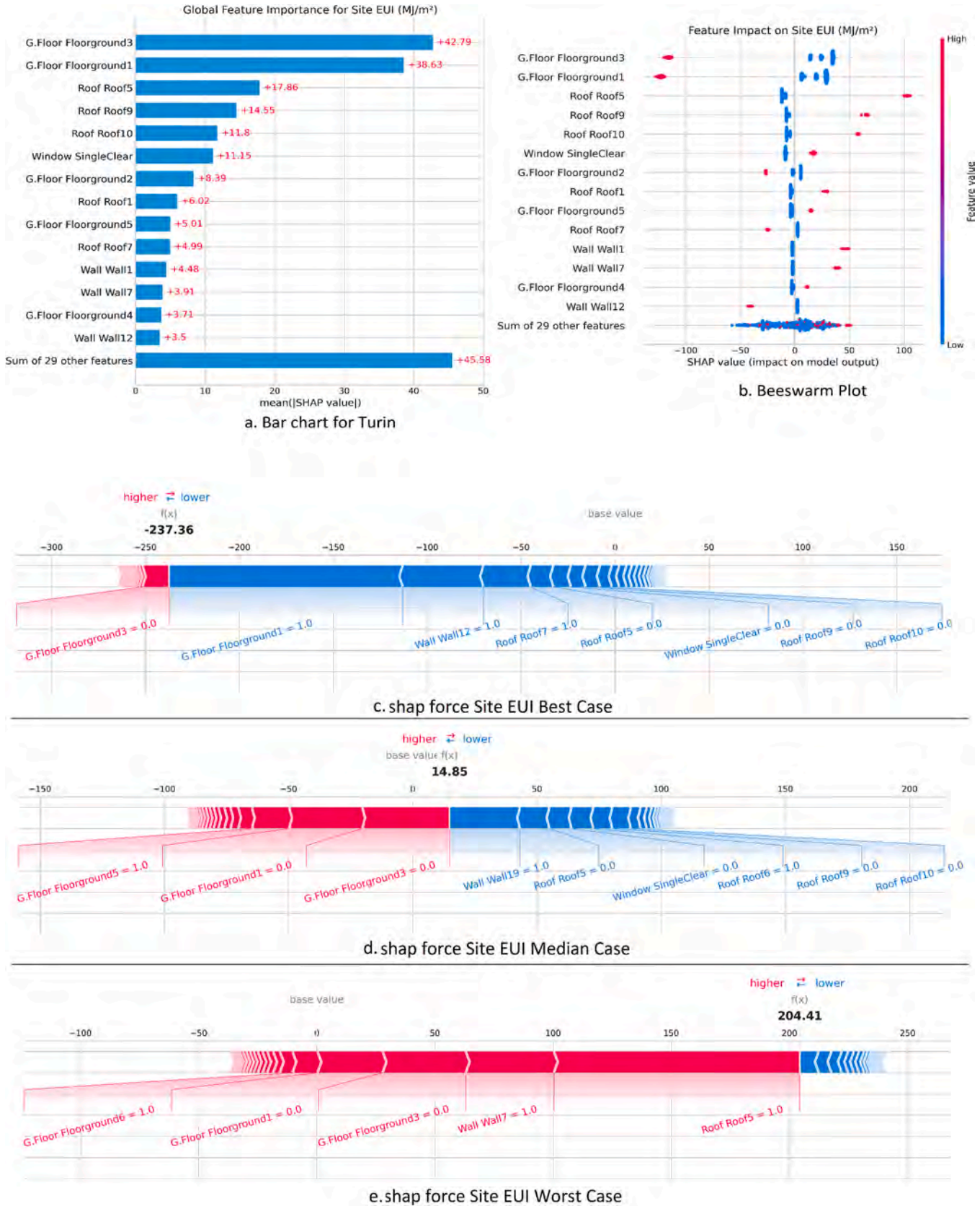


Fig. 8. SHapley Additive ExPlanations for Turin.

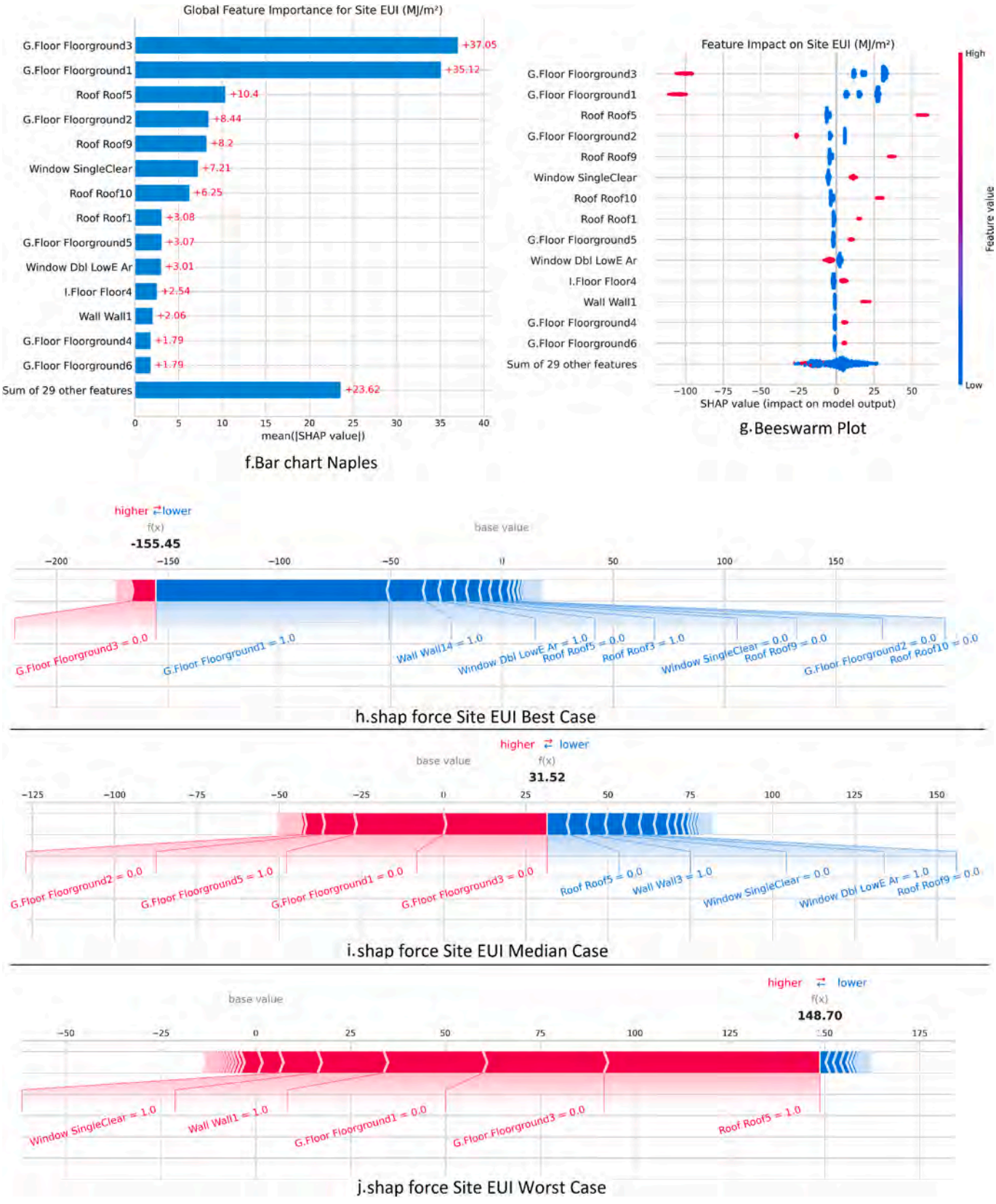


Fig. 9. SHapley Additive ExPlanations for Naples.

Table 4
Turin Validation Results.

Output Variable	NMBE (%)	NMBE Criteria	CVRMSE (%)	CVRMSE Criteria
Site EUI (MJ/m^2)	−0.201 (Pass)	<= 10%	0.287 (Pass)	<= 30%
Total Heating (GJ)	−0.712 (Pass)	<= 10%	1.056 (Pass)	<= 30%
Total Cooling (GJ)	0.197 (Pass)	<= 10%	0.529 (Pass)	<= 30%

Table 5
Naples Validation Results.

Output Variable	NMBE (%)	NMBE Criteria	CVRMSE (%)	CVRMSE Criteria
Site EUI (MJ/m^2)	−0.158 (Pass)	<= 10%	0.344 (Pass)	<= 30%
Total Heating (GJ)	0.222 (Pass)	<= 10%	1.211 (Pass)	<= 30%
Total Cooling (GJ)	0.415 (Pass)	<= 10%	0.508 (Pass)	<= 30%

Fig. 9h–j). This analysis confirms that the surrogate models are interpretable and physically consistent, providing climate-adaptive predictions suitable for the optimization study.

4.4. Assessment of predictive accuracy based on ASHRAE Guideline 14 criteria

The results of the formal validation for both Naples and Turin climates, summarized in Table 4 and Table 5, demonstrated that the surrogate models successfully satisfied the Guideline's stringent calibration criteria, which require an $|NMBE| \leq 10\%$ and a $CVRMSE \leq 30\%$. For both cities, the error values for all energy metrics were exceptionally low, confirming the high fidelity and trustworthiness of the models for use in the subsequent optimization analysis. The CVRMSE scores consistently remained well below 1.3% for all metrics, significantly exceeding the guideline's threshold.

4.5. Ranking and filtering of top 5 optimal designs

The top five design alternatives for each climate were selected from the Pareto-optimal sets obtained after re-simulating the surrogate-screened Top-K candidates in EnergyPlus. These re-simulated non-dominated solutions balance Site EUI, total heating, and total cooling, ensuring that the architectural comparison is based on verified simulation outputs rather than on surrogate predictions alone. For instance, in Turin, a re-simulated Site EUI of 281.28 MJ/m^2 was predicted, with an exceptionally close value of 281.36 MJ/m^2, confirming the high accuracy of the predictive models near the performance boundary. The architectural composition of the selected designs, summarized in Fig. 10A, reveals both convergent and divergent optimization strategies. Several convergent patterns emerged, indicating shared high-performance design features across both climates. Across all ten Pareto-optimal designs, the following components were consistently adopted: the Roof7 assembly, the Floor2 intermediate floor, and a limited subset of the most insulated ground floor options (Floorground1 and Floorground3). All designs also exclusively utilized the Double Low-E Argon window system, underscoring its robust contribution to energy efficiency in both climatic zones. Conversely, the most notable divergence appeared in the selection of external wall assemblies, strongly influenced by climate-specific demands. For Turin (a heating-dominated climate), all optimal solutions featured the three most thermally resistant wall types (Wall12, Wall14, and Wall17), reflecting a clear design priority toward maximizing envelope insulation and minimizing heat loss. In contrast, for Naples (a mixed-mode climate), the optimal set exhibited greater diversity. While high-performance Wall12 and Wall14 remained prominent, the inclusion of the more conventional Wall8 (a cavity wall with lower thermal resistance) signaled a deliberate trade-off. This choice reflects a balanced, climate-adaptive strategy where a modest reduction in heating performance is offset by a significant improvement in cooling efficiency, ultimately enhancing overall annual performance under warmer conditions. These climate-driven architectural choices are quantitatively reflected in the normalized performance comparison shown in Fig. 10B. The analysis highlights a stark contrast in energy demand: Turin has a 144.2% higher heating requirement compared to Naples, resulting in a 31.2% increase in the overall Site EUI. Conversely, the warmer climate in Naples drives a 48.3% higher cooling demand. Despite significant variations in energy loads, both sets of optimal designs maintain consistent optimal thermal comfort levels

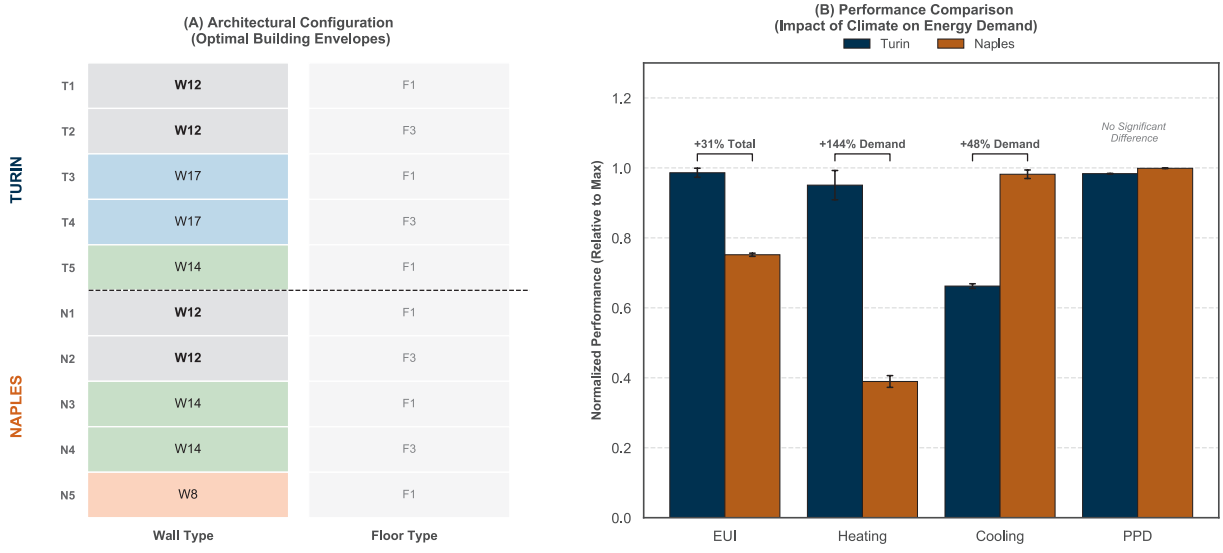


Fig. 10. Cross-climate comparison of top-performing building envelopes. (A) Architectural composition highlighting shared high-performance envelope features versus climate-specific solution patterns. (B) Normalized performance metrics illustrating the climate-driven trade-off between heating (+144% in Turin) and cooling (+48% in Naples) demands. Total Site EUI remains 31% higher in Turin, despite comparable thermal comfort (PPD) across the reported top-performing designs.

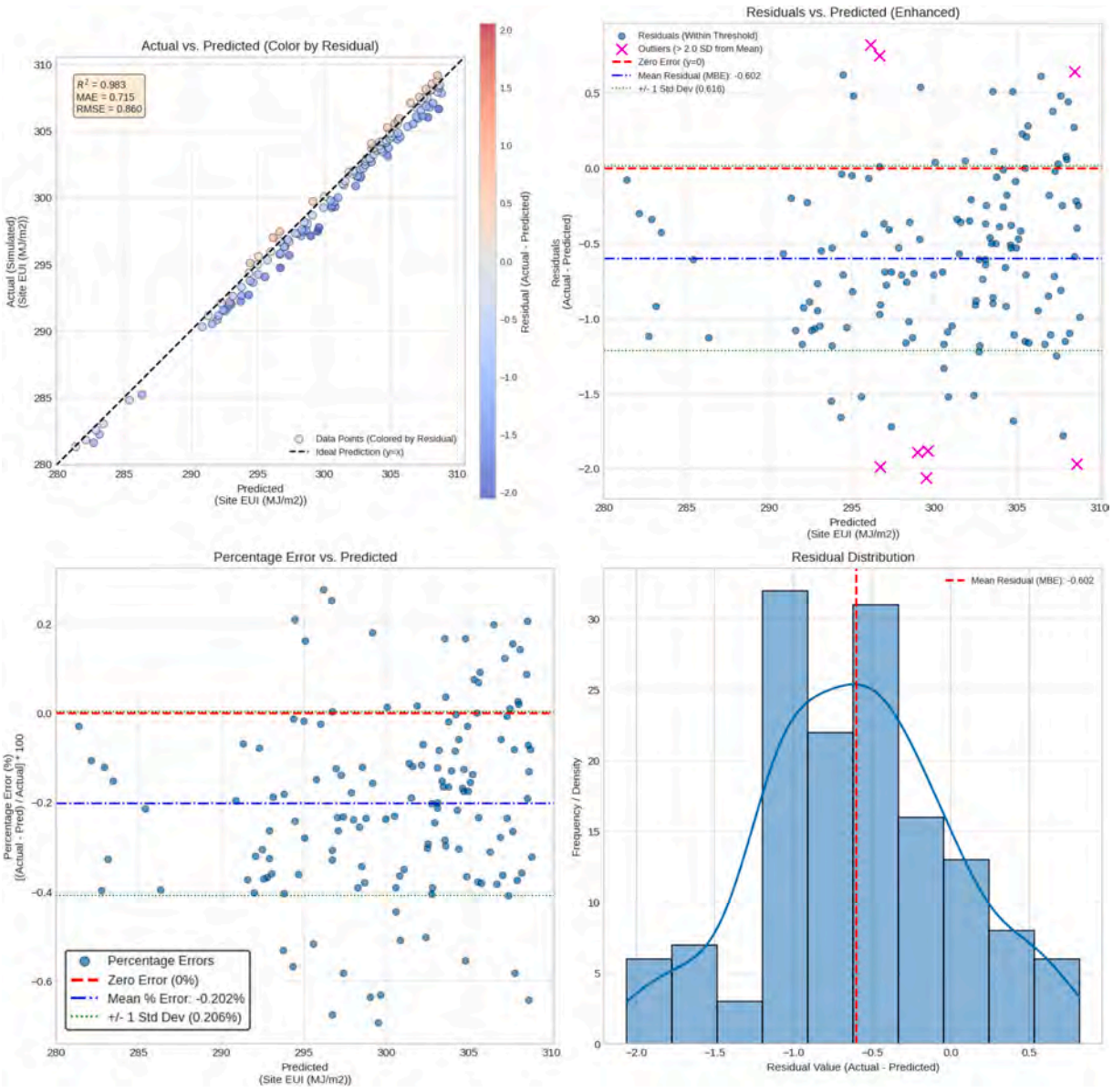


Fig. 11. Site EUI (MJ/m^2) for Turin.

(PPD), demonstrating the effectiveness of climate-specific envelope tuning.

4.6. Surrogate model validation

The final validation on an unseen dataset confirmed that the surrogate models achieved consistently high predictive accuracy for both Naples and Turin. For the primary performance target, EUI (e.g., Fig. 11 and Fig. 12), the surrogates demonstrated a close alignment of predictions with the 1:1 line. In Naples, the model reached $R^2 = 0.965$, with low absolute errors (MAE = 0.57 MJ/m^{-2}, RMSE = 0.79 MJ/m^{-2}). The residuals were approximately normally distributed, centered near zero, and the mean bias was small and negative (MBE = − 0.36 MJ·m^{-2}), indicating a mild tendency toward underestimation. In Turin, the surrogate achieved an even higher explained variance (R^2 = 0.983), although with slightly larger absolute errors in energy terms (MAE = 0.72, RMSE = 0.86 MJ·m^{-2}) compared to Naples. The bias was again negative (MBE = − 0.60 MJ·m^{-2}), implying a stronger underestimation. Percentage errors were tightly clustered around a mean close to zero with a smaller spread than in Naples. This trade-off illustrates a common pattern: Turin offers stronger variance explanation and tighter relative error dispersion, whereas Naples yields smaller absolute errors in energy units. In both climates, the residual plots showed no systematic structure, and the error distributions were homoscedastic and nearly Gaussian, confirming robust calibration. The analysis of the remaining performance targets, reported in the Supplementary Material (Figs. S23–S32), reinforces these findings. For total heating, both climates exhibited very high R^2 values (up to 0.983 in Turin) and low error magnitudes; in Naples, residual percentage errors were slightly more dispersed at low loads, a typical effect of small denominators. For total cooling, the models again showed excellent agreement, with Naples achieving $R^2 = 0.991$ and a small positive bias, while Turin delivered the lowest absolute errors and negligible bias. For PMV, residuals were tightly distributed around zero, and the horizontal banding visible in the residual plots reflects the quantization of simulation outputs (and, for PPD, its derivation from PMV) rather than any modeling deficiency. Taken together, the visual and numerical diagnostics confirm that the surrogate models are accurate, unbiased in practice, and robust across both climates, providing a reliable foundation for subsequent optimization analyses. The complete numerical results for all targets are

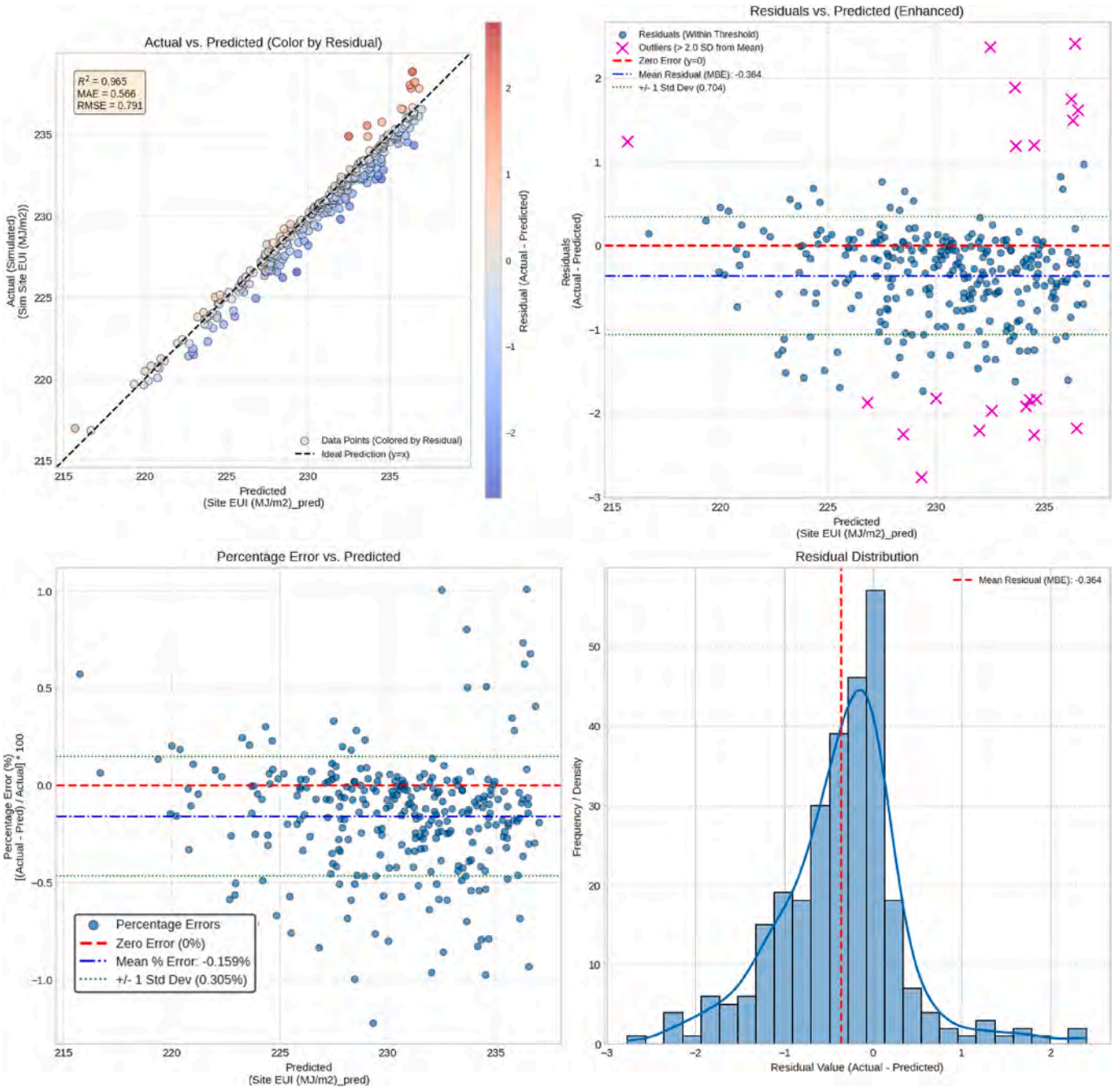


Fig. 12. Site EUI (MJ/m^2) for Naples.

Table 6
Hold-out validation metrics by target and climate. Errors are reported as MAE, RMSE, and mean bias error (MBE).

Target	Climate	R^2	MAE	RMSE	MBE
EUI (MJ·m^{-2})	Naples	0.965	0.566	0.791	−0.364
	Turin	0.983	0.715	0.860	−0.602
Total Heating (GJ)	Naples	0.970	0.384	0.434	−0.292
	Turin	0.983	0.152	0.219	+0.040
Total Cooling (GJ)	Naples	0.991	0.306	0.335	+0.290
	Turin	0.988	0.203	0.239	+0.007
PMV (–)	Naples	0.962	1.366×10^{-3}	1.732×10^{-3}	$+9.861 \times 10^{-4}$
	Turin	0.979	2.286×10^{-3}	2.930×10^{-3}	$+1.967 \times 10^{-3}$

summarized in Table 6. A further validation was conducted to assess the cross-climate performance of the pooled surrogate in an unseen location. When the Turin + Naples surrogate was applied to Rome, using only its local climate descriptors together with the same envelope thermo-physical parameters (U-values for walls, floors, roofs, and windows), the model maintained a very high level of predictive accuracy across all

Table 7
Building Envelopes Carbon Footprint Benchmarks.

Items	Embodied Carbon (kg CO2e / m2)	Source
Wall 8	164.5	ICE database
Wall 12	97.3	ICE database
Wall 14	48.9	ICE database
Wall 17	59.2	ICE database
Roof 7	41.6	ICE database
Floor Ground 1	56.8	ICE database
Floor Ground 3	117.9	ICE database
Floor Intermediate	53.1	ICE database
Italian National Grid Emission Factor (kg CO2e / kWh)	0.276	ISPRA

performance targets. For Site EUI, it achieved $R^2 = 0.986$ with MAE = 0.962 MJ·m^{-2}; comparable fidelity was observed for Total Cooling ($R^2 = 0.989$), Total Heating ($R^2 = 0.944$), and Average PMV ($R^2 = 0.998$). Residual errors remained near zero, showing no systematic bias. Because no retraining or retuning was performed and Rome's climatic indicators lie within the range covered by Turin and Naples, this evaluation represents an interpolative cross-climate test rather than extrapolation. The strong agreement across all metrics confirms that the surrogate demonstrates validated cross-climate transferability within the trained climatic range, reinforcing the model's robustness when exposed to a new but climatologically consistent context.

4.7. Carbon footprint

To complement energy-based optimization, life-cycle carbon was evaluated by combining embodied emissions (A1-A3) with operational emissions (B6). Table 7 summarizes the embodied carbon intensities of the main envelope components considered, while operational impacts were calculated using the Italian national grid emission factor (0.276 kg CO_2e/kWh). Building on this dataset, the five best-performing configurations in Turin and Naples were assessed in terms of life-cycle carbon. This post-optimization carbon assessment (Section 3.5) enables identification of cases where carbon-based re-ranking diverges from energy-based Pareto rankings, particularly where energy differences among candidate solutions are marginal but embodied carbon varies significantly.

4.7.1. Turin climate (Heating-Dominated)

In the Turin climate, the top five simulation-based solutions exhibit a narrow range of energy performance, with Site EUI values ranging from 281 to 291 MJ/m^2 (representing a 3.5% range, based on Table 8). However, their total carbon emissions span from 490 to 590 kg CO_2e/m^2 (20.5% difference). This disparity indicates that carbon introduces a stronger differentiation among solutions than energy performance alone.

The lowest-carbon configuration combines Wall 14, Floor Intermediate 2, Floor Ground 1, and Roof 7, (490 kg CO_2e/m^2). In comparison, configurations using Wall 12 with Floor Ground 3, despite achieving slightly lower EUI, result in substantially higher total emissions due to increased embodied carbon.This leads to a shift in preferred design: highly insulated assemblies optimized for heating reduction do not remain optimal when life-cycle emissions are considered. Instead, configurations with moderately lower thermal resistance but significantly reduced material-related emissions achieve better overall performance.

4.7.2. Naples climate (Cooling-Dominated)

In Naples, where cooling loads dominate, energy variation among the top solutions is even smaller, with Site EUI ranging from 216.8 to 220.5 MJ/m^2 (representing a 1.7% range, based on Table 9). However, their life-cycle carbon ranges from 419 to 535 kg CO_2e/m^2 (27.7% span). With minimal energy differentiation, embodied carbon becomes the dominant factor in ranking. The configuration using Wall 14 again yields the lowest emissions (419 kg CO_2e/m^2), while the inclusion of Wall 8, despite competitive energy performance, results in the highest carbon footprint (535 kg CO_2e/m^2) due to material intensity.

This result highlights that in cooling-dominated climates, where operational differences are limited, material selection governs the final decision outcome.

4.7.3. Cross-Climate insights

Across both climates, a consistent decision pattern is observed. Configurations that minimize operational energy, particularly those based on Wall 12, are systematically displaced by lower-embodied-carbon alternatives such as Wall 14 once life-cycle emissions are considered. This divergence between energy-optimal and carbon-optimal rankings of the top five Pareto-optimal envelope configurations (see Table 8 and Table 9) indicates that single-metric optimization is insufficient for envelope selection under carbon-constrained design criteria. This implies that envelope selection based solely on thermal performance can lead to suboptimal life-cycle outcomes, particularly when energy differences among candidate solutions are marginal.

4.7.4. Sensitivity analysis of carbon footprint

The sensitivity analysis, conducted by selectively excluding the embodied emissions of external walls and ground floors, provides insight into the contribution of individual subsystems to total carbon performance. In Turin, as shown in Table 8, removing wall emissions shifts the optimal solution from Simulation 5 to Simulation 1, whereas excluding ground floor emissions does not alter the ranking. This indicates that walls are the dominant driver of life-cycle carbon in heating-dominated conditions.

In Naples, as shown in Table 9, removing wall emissions again alters the ranking, while excluding ground floor emissions produces a smaller effect, with Simulation 4 approaching the top position. This suggests a more balanced contribution of subsystems, although walls remain the primary driver.

Table 8
Turin Building Envelopes Carbon Footprint and Sensitivity Analysis.

S.ID	W	FI	FG	R	Site EUI (MJ/m2)	Embodied (A1-A3) (kg CO2e)	Operational (B6) (kg CO2e)	Total Emission (kg CO2e)	Rank	Wall Removed Total Emission (kg CO2e)	New Rank	FG Removed Total Emission (kg CO2e)	New Rank
1	12	2	1	7	281.3	248.8	279.5	528	3	431	**1**	471	4
2	12	2	3	7	281.8	309.9	280.0	590	5	493	4	472	5
3	17	2	1	7	290.3	210.7	288.4	499	2	440	2	442	2
4	17	2	3	7	290.8	271.8	289.0	561	4	502	5	443	3
5	14	2	1	7	291.1	200.4	289.3	490	**1**	441	3	433	**1**

S.ID: Simulation ID W: Wall FI: Floor Intermediate FG: Floor Ground R: Roof

Table 9
Naples Building Envelopes Carbon Footprint and Sensitivity Analysis.

S.ID	W	FI	FG	R	Site EUI (MJ/m2)	Embodied (A1-A3) (kg CO2e)	Operational (B6) (kg CO2e)	Total Emission (kg CO2e)	Rank	Wall Removed Total Emission (kg CO2e)	New Rank	FG Removed Total Emission (kg CO2e)	New Rank
1	12	2	1	7	216.8	248.8	215.4	464	2	367	**1**	407	2
2	12	2	3	7	217.0	309.9	215.6	525	4	428	4	408	3
3	14	2	1	7	219.9	200.4	218.5	419	**1**	370	2	362	**1**
4	14	2	3	7	220.1	261.5	218.6	480	3	431	5	362	**1**
5	8	2	1	7	220.5	316.0	219.0	535	5	371	3	478	4

S.ID: Simulation ID W: Wall FI: Floor Intermediate FG: Floor Ground R: Roof

Overall, the results confirm that subsystem influence is climate-dependent, with walls consistently exerting the strongest impact on life-cycle carbon outcomes.

5. Discussion

This study addressed key gaps in envelope optimization by demonstrating cross-climate transferability, within the trained climatic range, and by integrating comfort (PMV/PPD) and life-cycle carbon (A1–A3, B6) into decision-making. Building on the Methods and Results, this discussion focuses on implications for design decisions, limitations, and transferability. In particular, the workflow's value lies in enabling repeatable multi-criteria screening (energy–comfort–carbon) at early design stages, where manual GUI-driven processes typically constrain scalability [6].

Surrogate benchmarking results underline the importance of evaluating multiple algorithms under equal conditions. The eight candidate models exhibited heterogeneous performance, similar to findings in other building surrogate studies [27,35]. In our case, an ensemble method (CatBoost) achieved the lowest prediction errors for site EUI and heating load. In contrast, a tuned neural network (ANN) minimized errors for cooling energy and PMV. As shown in the Results, both selected surrogates achieved high agreement with simulation and met relevant calibration criteria for their intended screening use. This level of accuracy indicates an unbiased, reliable fit for optimization purposes. It also aligns with broader trends: comprehensive reviews report that tree-based ensembles often excel in building energy prediction while ANNs perform strongly for complex, nonlinear outputs [31]. Our findings reinforce these patterns and demonstrate that no single algorithm is universally best, underscoring the value of surrogate benchmarking to identify the optimal model for each performance metric. This supports output-specific surrogate selection for each performance metric under consistent tuning budgets.

Applying the proposed screening-and-validation workflow in two climates yielded insightful comparisons. Pareto fronts were constructed on the re-simulated operational energy (site EUI, heating and cooling). As expected, the top envelope designs varied by climate, confirming that envelope performance is highly climate-specific. In cold Turin, every Pareto-optimal solution employed a highly insulating external wall (types 12, 14, or 17) to minimize heat loss during winter. By contrast, in warm Naples (cooling-dominated), the Pareto set included a design with a less insulated wall (type 8), indicating a deliberate trade-off: a modest increase in heating demand was accepted in exchange for significantly reducing summer cooling loads. This climate-adaptive choice reflects the priority shift toward mitigating overheating in the south. Notably, all Pareto-optimal candidates maintained acceptable comfort levels (average PPD $\approx$ 5%, within the acceptable comfort range see Section 4.5, Fig. 10), demonstrating that energy efficiency was achieved without sacrificing thermal comfort. The inclusion of comfort (PMV/PPD) as an explicit objective thus proved essential, as it prevented overly insulating solutions in Naples that would have exacerbated discomfort. Overall, the multi-objective results captured the intuitive divergence in envelope strategy between a heating-driven and a cooling-driven climate, while ensuring all shortlisted designs met basic comfort criteria.

Introducing life-cycle carbon (A1–A3, B6) as a post-optimization filter revealed substantial shifts in what constitutes an "optimal" design. Even among the top energy-and-comfort solutions, embodied carbon differences of over 100 kg CO_2e/m^2 were observed. In Naples, for example, one Pareto-optimal envelope featured a conventional cavity wall (Wall 8) that ranked among the top five in operational performance, yet this design incurred the highest total carbon footprint ($\sim$535 kg CO_2e/m^2) due to its materials. Another solution using a moderately insulated wall (Wall 14) achieved a much lower footprint ($\sim$419 kg CO_2e/m^2) and ultimately became the lowest-carbon option, despite having slightly higher energy use. A similar pattern emerged in Turin. Consequently, after carbon accounting, the Wall 14-based configuration (paired with Floor 2, Floor 1, Roof 7, and low-e glazing) stood out as the most sustainable envelope in both climates. This outcome highlights an important cross-climate insight: a carefully balanced envelope (moderate insulation with low embodied carbon) can deliver strong lifecycle performance across divergent climates. In contrast, the ultra-insulated Wall 12 designs that dominated purely energy-based rankings were consistently overtaken by Wall 14 once embodied emissions were considered. These findings underscore that optimizing only for operational energy can be misleading; an envelope seemingly optimal for energy may perform worse when embodied carbon is factored in. Our results validate the need to integrate embodied and operational perspectives in early design, echoing recent studies that call for carbon-inclusive building optimization [3,4]. Further validation was conducted to assess cross-climate performance in an unseen location. The Rome test further supports transferability under interpolation within the trained Turin–Naples climate-feature range (without retraining).

Few prior studies have examined such transferability. While some researchers have explored future climate scenarios for a single location by adjusting designs for projected warming; [5], others have applied automated ML optimization in a single region [12]. The present study uniquely evaluates envelope robustness across two contrasting current climates. This contributes to the growing multi-objective building design literature by showing how climate-specific trade-offs and life-cycle impacts can be systematically balanced.

From a practical design standpoint, the results translate into clear guidance for residential envelopes. Certain components emerged as universally beneficial "low-E" measures. A highly insulated roof (Roof 7), a lightweight intermediate floor (Floor 2), and high-performance double low-e windows consistently delivered significant energy savings in both the heating- dominant and cooling-dominant contexts. These elements form a robust baseline across climates and should be prioritized in early design decisions. By contrast, the external wall demanded a climate-adaptive, carbon-sensitive approach. In the cold climate (Turin), the optimal envelopes exclusively adopted the most thermally resistive walls (e.g. Walls 12, 14, 17) to curb heat loss. In the warm climate (Naples), however, top designs balanced insulation with thermal mass, notably by including Wall 8, a less-insulated cavity wall, to limit overheating. This difference highlights the importance of tailoring wall assemblies to seasonal demand profiles. Moreover, the carbon analysis showed that among Naples optimal walls, the heavy-insulated option (Wall 12) had such high embodied emissions that it

was outperformed in total carbon by a slightly less efficient but much lower-carbon wall (Wall 14). In practice, this means designers should avoid blindly maximizing insulation in milder climates; instead, a moderate insulation level coupled with low-carbon materials can achieve superior whole-life performance.

Overall, our automation pipeline and surrogate modeling framework enabled this nuanced evaluation of envelopes. The approach aligns with the growing adoption of multi-objective optimization methods for building energy performance at early design stages [38]. By transparently incorporating energy, comfort, and carbon considerations, the workflow provides a scalable template for designing high-performance, climate-responsive, and sustainable building envelopes.

Limitations and Future Work.

Despite its contributions, this study has certain limitations. First, it analyzed a single residential building archetype; the optimal solutions and surrogate model accuracy might differ for other building types or geometries. Second, only two climate zones were examined (a heating-dominated northern climate and a cooling-dominated southern climate in Italy). These provide a strong contrast, but further studies should test additional climates, including more extreme and moderate zones to generalize the findings. Third, the simulations used an ideal HVAC system (EnergyPlus IdealLoadsAirSystem) with fixed setpoints and standard occupancy schedules. This idealized approach ensured consistency but does not capture real HVAC performance limits or adaptive occupant behaviors. Fourth, the life-cycle assessment was limited to production (A1–A3) and operation (B6) stages, excluding construction, maintenance, and end-of-life phases. While the chosen stages represent the majority (~70%) of total emissions, omitting later-life stages means the absolute carbon figures are conservative. Fifth, the embodied carbon assessment relies on the ICE database, which is widely adopted in building LCA studies but primarily reflects UK manufacturing conditions. Consequently, absolute embodied-emission values may differ from those associated with Italian-specific supply chains. Nonetheless, the ICE database offers a consistent, transparent basis for comparative evaluation across alternative envelope configurations. Operational carbon emissions were estimated using a single Italian grid emission factor representative of current conditions. Future grid decarbonization or shifts in the electricity fuel mix are expected to reduce operational emissions, potentially increasing the relative importance of embodied impacts. The findings should therefore be interpreted within the context of the adopted database and the present grid scenario.

Future work should extend the life-cycle boundary (e.g., include material replacement and end-of-life disposal) and consider diverse building configurations. It would also be valuable to apply the automation, multi-objective pipeline to other building categories (such as offices or retrofits) and to investigate adaptive comfort criteria for naturally ventilated scenarios. These extensions will further test the cross-climate robustness of the approach and enhance its applicability as a practical design tool for sustainable buildings.

6. Conclusion

This study advanced an automated, machine-learning–assisted framework to optimize residential building envelopes across two contrasting Italian climates by jointly targeting operational energy (site EUI; heating/cooling), thermal comfort (PMV/PPD), and life-cycle carbon (A1–A3, B6). The framework enabled reliable early-stage screening of envelope alternatives under consistent modeling assumptions, supporting decision-making across multiple performance metrics.

Pareto fronts were constructed for energy objectives, site EUI, total heating, and total cooling. At the same time, comfort acceptability was verified within the shortlist (PMV/PPD), and life-cycle carbon (A1–A3, B6) was applied afterwards as a post-optimization re-ranking filter. This screening exposed climate-dependent trade-offs that are often obscured in energy-only studies. All shortlisted envelopes preserved acceptable comfort (average PPD ≈ 5%), yet the composition of Pareto sets pivoted with climate: in heating-dominated Turin, every optimal solution adopted the most insulating external walls to suppress winter losses; in cooling-dominated Naples, a less-insulated cavity wall also appeared in the Pareto set to mitigate summertime overheating. The pooled surrogate trained on Turin and Naples was evaluated in an unseen climate (Rome) using local weather descriptors and identical envelope U-values. This Rome evaluation supports cross-climate transferability under interpolation within the trained climatic range.

Introducing life-cycle carbon as a post-optimization filter reshaped these energy-and-comfort shortlists. Among near-*iso*-energy finalists, total footprints diverged by more than 100 kg CO_2e/m^2, with external walls acting as the principal driver of whole-life impact. When A1–A3 and B6 were applied, a balanced, lower-embodied-carbon wall consistently displaced ultra-insulated alternatives in both climates; notably, the configuration with Wall 14 + Floor-Intermediate 2 + Floor-Ground 1 + Roof 7 + double low-E argon glazing emerged as the lowest-carbon choice in Turin and Naples. This carbon-informed re-ranking is the central decision insight: optimizing for energy and comfort alone can misidentify "winners" when materials are assessed over the life cycle.

For practice, the results translate into clear early-stage guidance: high-performance roof insulation, a lightweight intermediate floor, and double low-E argon glazing proved effective across climates and should anchor baseline envelope choices. By contrast, external walls demand climate-aware, carbon-sensitive selection: maximize insulation in cold contexts to curb transmission losses; in warm contexts balance insulation with thermal mass and embodied-carbon profiles to avoid whole-life penalties. These findings should be interpreted within the context of the adopted embodied-carbon database and the current Italian electricity grid emission scenario, noting that alternative datasets or future grid decarbonization may affect the quantitative outcomes.

Overall, this study shows that an interpretable ML-surrogate layer, coupled with boundary-aware sampling and post-optimization life-cycle assessment, can reveal cross-climate solutions and expose energy–comfort–carbon trade-offs that conventional workflows overlook. By aligning envelope selection with life-cycle carbon while maintaining energy performance and comfort, the study provides a transparent, scalable route to climate-responsive, low-carbon residential envelopes, and a reproducible, automated workflow suitable for early-stage adoption.

CRediT authorship contribution statement

Alireza Ghomimoghadam: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mostafa Adibian:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Luca Barbierato:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Daniele Salvatore Schiera:** . **Samad ME Sepasgozar:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Appendix

.

Table 10
Library of Standardized Building Envelope Components for Parametric Analysis.

Cross Section		Layer No.	Thickness (mm)	Material
	W1			
		1	10	Internal plaster
		2	10	Concrete wall (Internal)
		3	30	Glass fibre (wool) insulation board
		4	300	Concrete wall (external)
		5	20	External plaster – cp 1000
	Total		370	U-value ($W/m^2 \cdot K$) = 0.6708
	W2			
		1	10	Generic plywood
		2	40	Glass fibre (glass wool) insulation board
		3	300	Lightweight_Brick
		4	20	External plaster – cp 1000
	Total		370	U-value ($W/m^2 \cdot K$) = 0.4583
	W3			
		1	10	Internal plaster
		2	120	Hollow clay brick
		3	80	80 mm vertical air cavity
		4	80	Hollow clay brick – strength class 0.23
		5	10	External plaster – cp 1000
		6	100	Expanded polystyrene (EPS) board, block-moulded, UNI 7891 – density 30
		7	20	External plaster – cp 840
	Total		420	U-value ($W/m^2 \cdot K$) = 0.2888
	W4			
		1	15	Internal plaster
		2	80	Hollow brick (strength 0.23)
		3	80	Vertical air layer (8 cm)
		4	30	Expanded polystyrene beads
		5	120	Lightweight bricks ($\rho \approx 800$ kg/m^3)
	Total		325	U-value ($W/m^2 \cdot K$) = 0.6407
	W6			
		1	20	Internal plaster
		2	120	Hollow brick
		3	30	Vertical air layer (3 cm)
		4	40	Basic insulation layer
		5	120	Hollow brick
		6	20	Plaster (finish)
	Total		350	U-value ($W/m^2 \cdot K$) = 0.7008
	W7			
		1	20	Internal plaster
		2	120	Hollow brick
		3	50	Vertical air layer (5 cm)
		4	250	Hollow brick
		5	20	External Plaster
	Total		460	U-value ($W/m^2 \cdot K$) = 0.6247
	W8			
		1	20	Internal plaster
		2	80	Hollow brick (strength 0.20)
		3	120	Cavity with insulation
		4	250	Clay brick
		5	20	External Plaster
	Total		490	U-value ($W/m^2 \cdot K$) = 0.2694
	W9			
		1	20	Internal plaster
		2	80	Hollow brick (strength 0.20)
		3	40	Cavity insulation
		4	380	Solid_Dense_Brick (1800)
		5	20	External Plaster
	Total		540	U-value ($W/m^2 \cdot K$) = 0.5443
	W10			
		1	10	Generic plywood
		2	40	Glass-wool insulation panel
		3	300	Clay brick
		4	20	External Plaster
	Total		370	U-value ($W/m^2 \cdot K$) = 0.4583
	W11			

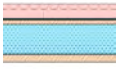
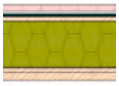
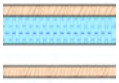
(continued on next page)

Table 10 (*continued*)

Cross Section		Layer No.	Thickness (mm)	Material
		1	15	Lime-gypsum plaster
		2	300	Hollow clay block (Poroton P800 BIO)
		3	5	Scratch/base coat (lime or lime-cement)
		4	120	Rock wool rigid boards (dual density)
		5	7	Reinforced skim coat
		6	3	Finishing coat + primer (1800 kg/m3)
	Total		450	U-value (W/m^2·K) = 0.2014
	W12			
		1	10	Lime-gypsum plaster
		2	50	Cement-bonded particle board
		3	20	OSB board
		4	1	PVC vapor control layer
		5	220	Cellulose fiber fill
		6	20	OSB board
		7	4	Bituminous felt
		8	60	Glass-fiber insulation panel (Isover T100)
		9	50	Vertical air cavity
		10	12	Ceramic tiles
	Total		447	U-value (W/m^2·K) = 0.1197
	W13			
		1	15	Lime-gypsum plaster
		2	100	Hollow clay block (10 × 25 × 25)
		3	0.1	Kraft vapor retarder
		4	99.9	Rock wool rigid boards
		5	10	Scratch/leveling coat
		6	200	Hollow clay block (20 × 25 × 25)
		7	15	Lime or lime-cement plaster
	Total		440	U-value (W/m^2·K) = 0.2081
	W14			
		1	12.5	Gypsum board (Density 800 kg/m3)
		2	50	Wood-fiber board (GUTEX) (Density 150 kg/m3)
		3	15	OSB panel
		4	0.2	Vapor retarder
		5	120	Mineral wool
		6	15	OSB panel
		7	3	ETICS adhesive
		8	80	EPS “80” panel
		9	4	Base coat for EPS
		10	3	Textured finish plaster
	Total		302.7	U-value (W/m^2·K) = 0.1538
	W15			
		1	15	Knauf Vidiwall gypsum-fiber board
		2	45	Rockwool Panel 226
		3	155	X-LAM (CLT) panel
		4	120	Rockwool Panel 226
		5	8	Lime-gypsum plaster
	Total		343	U-value (W/m^2·K) = 0.1603
	W16			
		1	15	Lime-gypsum plaster
		2	80	Hollow clay brick (250 × 80 × 250)
		3	160	Rubble / crushed stone infill
		4	80	Hollow clay brick (250 × 80 × 250)
		5	40	XPS board (unskinned, 30 kg/m^3)
		6	15	Lime-cement plaster
	Total		390	U-value (W/m^2·K) = 0.5497
	W17			
		1	9.5	Gypsum plasterboard
		2	62	Graphite EPS board
		3	165	Medium-density concrete (~1800 kg/m^3)
		4	123	Graphite EPS board
		5	15	Lime-cement plaster
	Total		374.5	U-value (W/m^2·K) = 0.1577
	W18			
		1	15	Lime/gypsum plaster
		2	80	Hollow clay brick (F825)
		3	0.1	Kraft paper vapor retarder
		4	99.9	Rock wool rigid board, medium density
		5	10	Lime/cement base coat
		6	300	Poroton P800 BIO clay block
		7	15	Lime/cement plaster
	Total		520	U-value (W/m^2·K) = 0.2039
	W19			

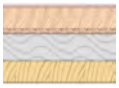
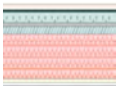
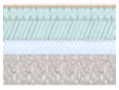
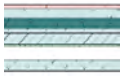
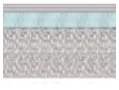
(*continued on next page*)

Table 10 (*continued*)

Cross Section		Layer No.	Thickness (mm)	Material
		1	20	Interior plaster
		2	80	Hollow clay brick
		3	50	River pebble layer
		4	250	Solid clay bricks (dense)
		5	20	Exterior plaster
	Total		420	U-value ($W/m^2 \cdot K$) = 1.1914
	W20			
		1	20	Pure gypsum plaster
		2	300	Lightweight hollow clay bricks (density = 600 kg/m3)
		3	60	Expanded cork board (density = 160 kg/m3)
		4	20	Cement/sand plaster
		5	6	ETICS skim coat
		6	3	Finish coat
	Total		409	U-value ($W/m^2 \cdot K$) = 0.4013
	Internal Wall			
		1	5	Ytong RY25 skimming for internal masonry
		2	150	Y-PRO blocks
		3	5	Ytong RY25 skimming for internal masonry
Total			160	U-value ($W/m^2 \cdot K$) = 0. 6934
	Roof1			
		1	50	Polystyrene insulation panel (EPS)
		2	20	Cement mortar (screed)
		3	40	Reinforced concrete
		4	160	Hollow clay floor block
		5	20	Interior plaster
	Total		290	U-value ($W/m^2 \cdot K$) = 0.6094
	Roof2			
		1	70	Terracotta roof tiles
		2	5	Bituminous sheet
		3	3	Bituminous sheet
		4	20	OSB board
		5	140	EPS 100 rigid insulation
		6	4	Expanded polyethylene sheet
		7	35	Timber roof decking
	Total		277	U-value ($W/m^2 \cdot K$) = 0.2005
	Roof3			
		1	25	Terracotta roof tiles
		2	10	Ventilated horizontal air layer
		3	8	Bituminous sheet
		4	4	Bituminous sheet
		5	18	OSB board
		6	1	Breathable membrane
		7	180	Mineral wool insulation
		8	2	Vapor barrier
		9	18	OSB board
		10	32	Timber roof decking
	Total		298	U-value ($W/m^2 \cdot K$) = 0.1695
	Roof4			
		1	70	Lightweight concrete screed
		2	4	Asphalt (vapor barrier)
		3	50	Wood-wool panels – density 400 (CELENIT)
		4	50	Wood-wool panels – density 400 (CELENIT)
		5	4	Asphalt (vapor barrier)
		6	60	XPS board – extruded polystyrene (unskinned)
		7	40	Ordinary screed
		8	225	Insulated RC–hollow clay floor slab
		9	10	Interior plaster
	Total		513	U-value ($W/m^2 \cdot K$) = 0.2571
	Roof5			
		1	60	Terracotta tiles
		2	40	Ordinary screed
		3	225	Insulated RC–hollow clay floor slab
		4	10	Interior plaster
	Total		335	U-value ($W/m^2 \cdot K$) = 1.1339
	Roof6			
		1	20	Timber roof boarding
		2	50	Polystyrene insulation panel
		3	40	Horizontal air layer (4 cm, upward)
		4	20	Timber roof boarding

(*continued on next page*)

Table 10 (*continued*)

Cross Section		Layer No.	Thickness (mm)	Material
	Total		130	U-value ($W/m^2 \cdot K$) = 0.5856
	Roof7			
		1	10	Terracotta roof tiles
		2	100	Pressed wood-fibre board (ρ≈50)
		3	20	Generic timber (ρ≈450)
		4	120	ROCKWOOL panel 226 (mineral wool)
		5	100	X-LAM (CLT) panel
	Total		350	U-value ($W/m^2 \cdot K$) = 0.1387
	Roof8			
		1	50	Polystyrene insulation panel (EPS)
		2	20	Cement mortar
		3	40	Reinforced concrete
		4	240	Hollow clay floor block (strength 0.31)
		5	20	Internal plaster
	Total		370	U-value ($W/m^2 \cdot K$) = 0.5949
	Roof9			
		1	30	External paving (clinker tiles)
		2	10	Bitumen (vapor barrier)
		3	60	Ordinary screed
		4	20	Cement mortar
		5	40	Reinforced concrete
		6	240	Hollow clay floor block (strength 0.31)
		7	20	Internal plaster
	Total		420	U-value ($W/m^2 \cdot K$) = 1.3735
	Roof10			
		1	10	Bitumen (vapor barrier)
		2	60	Ordinary screed
		3	20	Cement mortar
		4	40	Reinforced concrete
		5	240	Hollow clay floor block – resistance 0.31
		6	20	Internal plaster
	Total		390	U-value ($W/m^2 \cdot K$) = 1.4594
	FloorGround1			
		1	10	Generic plywood
		2	65	Ordinary screed
		3	100	Ordinary concrete
		4	0.3	Polyvinyl chloride (PVC)
		5	75	Insulation
		6	0.3	Synthetic sheet (vapor barrier)
		7	150	Pebbles & crushed stones (ballast)
	Total		400.6	U-value ($W/m^2 \cdot K$) = 0.42
	FloorGround2			
		1	60	Ordinary screed, 60 mm
		2	180	Lightweight screed "Mix Them Light"
		3	300	Reinforced-concrete slab
	Total		540	U-value ($W/m^2 \cdot K$) = 0.2208
	FloorGround3			
		1	16	Terracotta tiles
		2	4	Paste adhesive – Bonding 11
		3	40	Lightweight concrete screed
		4	50	STIFERITE GTE (rigid insulation board)
		5	20	Standard screed
		6	50	Expanded-clay concrete, closed structure – density 1600
		7	10	PVC membrane (vapor barrier)
		8	60	Horizontal air layer, 6 cm – upward
		9	50	Concrete sub-base
	Total		720	U-value ($W/m^2 \cdot K$) = 0.35
	FloorGround4			
		1	15	Interior flooring
		2	30	Cement mortar
		3	100	Lightweight concrete
		4	200	Coarse gravel (no clay)
	Total		345	U-value ($W/m^2 \cdot K$) = 1.4899
	FloorGround5			
		1	15	Interior flooring
		2	30	Cement mortar
		3	100	Lightweight concrete
		4	300	Coarse gravel (no clay)

(*continued on next page*)

Table 10 (*continued*)

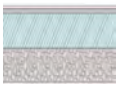
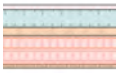
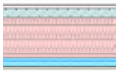
Cross Section		Layer No.	Thickness (mm)	Material
	Total FloorGround6		445	U-value ($W/m^2 \cdot K$) = 1.3253
		1	15	Interior flooring
		2	30	Cement mortar
		3	200	Lightweight concrete
		4	200	Coarse gravel (no clay)
	Total Floor1		445	U-value ($W/m^2 \cdot K$) = 1.0265
		1	15	Ceramic tiles
		2	45	Ordinary screed
		3	230	Clay-concrete slab (ribbed slab), resistance 1.22
		4	10	Internal plaster
	Total Floor2		300	U-value ($W/m^2 \cdot K$) = 0.6858
		1	15	Interior flooring
		2	60	Lightweight concrete screed
		3	30	Timber boarding (planking)
		4	100	Secondary joists — cavity
		5	10	Generic plywood
	Total Floor3		215	U-value ($W/m^2 \cdot K$) = 1.3695
		1	15	Ceramic tiles
		2	40	Ordinary screed
		3	175	Clay–concrete flat slab (latero-cement)
		4	40	Medium insulation
		5	10	Lime-gypsum plaster
	Total Floor4		280	U-value ($W/m^2 \cdot K$) = 0.7974
		1	12.5	Flooring (finish)
		2	85	Screed
		3	40	Structural slab
		4	120	Insulation
		5	12.5	Plaster
	Total		270	U-value ($W/m^2 \cdot K$) = 0.27

Table 11
Occupancy pattern and people assumptions (Residential, Detached House; EN 16798–1:2019 Annex C –aligned).

Occupancy Pattern	Number of People (method & density)	Activity level	Presence schedule (fractions)
Standard Residential Pattern *(aligned with EN 16798–1:2019 Annex C)*	Per zone, computed as Zone floor area / 42.5 $m^2 \cdot person^{-1}$	119 W/person *(derived as 2.8 W/m^2 × 42.5 m^2/person from EN 16798–1:2019 Annex C)*	Weekdays: 00:00–06:00 → 1.0; 06:00–08:00 → 0.5; 08:00–12:00 → 0.1; 12:00–16:00 → 0.2; 16:00–19:00 → 0.5; 19:00–22:00 → 0.8; 22:00–24:00 → 1.0. Weekends: 00:00–06:00 → 1.0; 06:00–22:00 → 0.8; 22:00–24:00 → 1.0.

Table 12
The materials of windows.

No	Windows	U-factor ($W/m^2 \cdot K$)	Solar Heat Gain Coefficient	Visual Light Transmittance
W1	Single Clear 3 mm Window	5.906	0.861	0.890
W2	Double Clear 3 mm Air 13 mm Window	2.725	0.763	0.797
W3	Double Low-E Argon Window	1.834	0.594	0.666

Table 13
Library of Standardized Building Envelope Components for Parametric Analysis.

Assembly / Element	Inside → Outside stratigraphy (thickness, mm)	Key Purpose / Notes
Wall 8	Internal plaster (20) • Hollow brick, $f_\beta \approx 0.20$ (80) • Insulated cavity (120) • Clay brick (250) • External plaster (20)	Cavity masonry wall that couples high thermal mass (250 mm clay brick) with a continuous insulated cavity to reduce steady-state heat flow and improve decrement factor/phase shift under summer loads. Plasters act as air-tight layers and provide surface durability. Reported U ≈ 0.269 W/m^2·K, total ≈ 490 mm.
Wall 12	Lime-gypsum plaster (10) • Cement-bonded particle board (50) • OSB (20) • PVC air/vapour control layer (1) • Dense cellulose fibre infill (220) • OSB (20) • Bituminous felt (4) • Mineral-wool (Isover T100) external insulation (60) • Ventilated air cavity (50) • Ceramic tile cladding (12)	High-performance timber/boarded wall: primary airtightness via taped OSB + PVC AVCL; capillary-active cellulose provides heat capacity and hygrothermal buffering; exterior ventilated rainscreen controls moisture and solar load; external MW layer limits thermal bridges. U ≈ 0.120 W/m^2·K, total ≈ 447 mm.
Wall 17	Gypsum plasterboard (9.5) • Graphite EPS (62) • Medium-density concrete core (~1800 kg/m^3) (165) • Graphite EPS (123) • Lime-cement render (15)	Massive retrofit façade with double-sided EPS to reach low U while retaining structural mass for time-lag and indoor temperature stability. Interior EPS partially decouples mass—useful for winter efficiency but reduces direct thermal storage; exterior render provides durability. U ≈ 0.158 W/m^2·K, total ≈ 374.5 mm.
Wall 14	Gypsum board (12.5) • Wood-fibre board (50) • OSB (15) • Vapour retarder (0.2) • Mineral wool in stud zone (120) • OSB (15) • ETICS adhesive (3) • EPS "80" (80) • Base coat (4) • Textured finish (3)	Timber-frame with ETICS: interior service/installation layer (wood-fibre + OSB) protects AVCL and enables airtight detailing; MW in studs gives low λ and acoustic damping; external EPS-ETICS provides continuous insulation and thermal-bridge mitigation. U ≈ 0.154 W/m^2·K, total ≈ 302.7 mm.
Roof 7	CLT panel (100) • Mineral-wool (ROCKWOOL 226) (120) • Generic timber layer (20) • Wood-fibre board (100) • Terracotta roof tiles (10)	Warm roof over CLT with dual, vapour-open insulation (MW + wood-fibre) for excellent phase shift, fire-robust external layer, and drying potential to the exterior; roof tiles act as UV/rain shield. U ≈ 0.139 W/m^2·K, total ≈ 350 mm.
Floor 2 (intermediate)	Interior flooring (15) • Lightweight concrete screed (60) • Timber boarding (30) • Secondary joists—cavity (100) • Plywood soffit (10)	Lightweight inter-story floor balancing service voids with limited mass; suitable where dynamic loads and acoustic treatment are required. Add resilient layers if impact sound control is critical. U ≈ 1.370 W/m^2·K, total ≈ 215 mm (heat flow direction: room → room).
Floor ground 1 (slab on grade)	Plywood (10) • Ordinary screed (65) • Ordinary concrete (100) • PVC layer (0.3) • Thermal insulation (75) • Synthetic vapour barrier (0.3) • Pebble/crushed-stone base (150)	Conventional slab-on-grade: insulation above vapour control limits ground losses; granular base ensures capillary break and drainage. Coordinate edge-insulation to control perimeter thermal bridges. U ≈ 0.420 W/m^2·K, total ≈ 400.6 mm.
Floor ground 3 (insulated slab)	Terracotta tiles (16) • Paste adhesive (4) • Lightweight concrete screed (40) • STIFERITE GTE rigid insulation (50) • Standard screed (20) • Expanded-clay concrete, closed structure ($\rho \approx 1600$) (50) • PVC vapour membrane (10) • Horizontal air layer, upward (60) • Concrete sub-base (50)	High-R slab-on-grade variant: PIR/PUR-type rigid board (GTE) above structural layers yields low heat flow; vented air space and vapour control reduce soil moisture effects; suitable for retrofit where raising floor level is possible. U ≈ 0.35 W/m^2·K, total thickness ≈ 300 mm (note: 6 cm air layer included).
Wall 8	Internal plaster (20) • Hollow brick, $f_\beta \approx 0.20$ (80) • Insulated cavity (120) • Clay brick (250) • External plaster (20)	Cavity masonry wall that couples high thermal mass (250 mm clay brick) with a continuous insulated cavity to reduce steady-state heat flow and improve decrement factor/phase shift under summer loads. Plasters act as air-tight layers and provide surface durability. Reported U ≈ 0.269 W/m^2·K, total ≈ 490 mm.

Table 14
Hyperparameter Search Space for ANN and CatBoost Models.

Model Type	Hyperparameter	Search Range / Values	Sampling Strategy
Keras ANN	Number of Layers	1 to 3	Integer
	Neurons per Layer	32 to 256 (step of 32)	Integer
	Dropout Rate	0.1 to 0.5 (step of 0.1)	Float
	Learning Rate	10^{-4} to 10^{-2}	Log-uniform
CatBoost Regressor	Iterations	100 to 1500	Log-uniform
	Learning Rate	0.001 to 0.3	Log-uniform
	Tree Depth	3 to 10	Integer
	L2 Regularization	10^{-8} to 10.0	Log-uniform
	Border Count	32 to 255	Integer

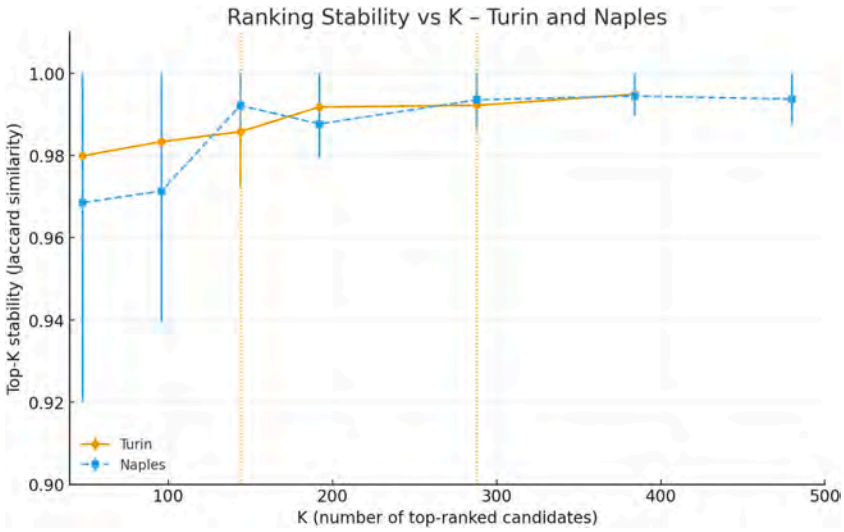


Fig. 13. Top-K Jaccard stability and composite-score gain vs. cost; vertical line = chosen K.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2026.117532.

Data availability

Data will be made available on request.

References

[1] M. Fellah, S. Ouhaibi, N. Belouaggadia, K. Mansouri, Harnessing machine learning for enhanced thermal insulation and energy efficiency in buildings worldwide, Results Eng. 25 (2025) 104086, https://doi.org/10.1016/j.rineng.2025.104086.
[2] S. Alghamdi, W. Tang, S. Kanjanabootra, and D. Alterman, "Optimising Building Energy and Comfort Predictions with Intelligent Computational Model," *Sustainability*, vol. 16, no. 8, 2024, doi: 10.3390/su16083432.
[3] L. He, W. Wang, Design optimization of public building envelope based on multi-objective quantum genetic algorithm, Journal of Building Engineering 91 (2024) 109714, https://doi.org/10.1016/j.jobe.2024.109714.
[4] C. Zhou, et al., Deciphering the nonlinear and synergistic role of building energy variables in shaping carbon emissions: a LightGBM- SHAP framework in office buildings, Build. Environ. 266 (2024) 112035, https://doi.org/10.1016/j.buildenv.2024.112035.
[5] Z. Ding, J. Li, Z. Wang, and Z. Xiong, "Multi-Objective Optimization of Building Envelope Retrofits Considering Future Climate Scenarios: An Integrated Approach Using Machine Learning and Climate Models," *Sustainability*, vol. 16, no. 18, 2024, doi: 10.3390/su16188217.
[6] Y. Liu, H. Chen, L. Zhang, Z. Feng, Enhancing building energy efficiency using a random forest model: a hybrid prediction approach, Energy Rep. 7 (2021) 5003–5012, https://doi.org/10.1016/j.egyr.2021.07.135.
[7] D. Kömürcü and E. Edis, "Machine Learning Modeling for Building Energy Performance Prediction Based on Simulation Data: A Systematic Review of the Processes, Performances, and Correlation of Process-Related Variables," *Buildings*, vol. 15, no. 8, 2025, doi: 10.3390/buildings15081301.
[8] M.S. Hasan, Md. Tarequzzaman, Md. Moznuzzaman, M.A. Ahad Juel, Prediction of energy consumption in four sectors using support vector regression optimized with genetic algorithm, Heliyon 11 (2) (2025) e41765, https://doi.org/10.1016/j.heliyon.2025.e41765.
[9] J. Huang, S. Wang, F. Teng, W. Feng, Thermal performance optimization of envelope in the energy-saving renovation of existing residential buildings, Energy Build. 247 (2021) 111103, https://doi.org/10.1016/j.enbuild.2021.111103.
[10] X. Zhen, et al., Analysis of building envelope for energy consumption and indoor comfort in a near-zero-energy building in Northwest China, Results Eng. 25 (2025) 104243, https://doi.org/10.1016/j.rineng.2025.104243.
[11] A. Abdeen, E. Mushtaha, A. Hussien, C. Ghenai, A. Maksoud, V. Belpoliti, Simulation-based multi-objective genetic optimization for promoting energy efficiency and thermal comfort in existing buildings of hot climate, Results Eng. 21 (2024) 101815, https://doi.org/10.1016/j.rineng.2024.101815.
[12] M.H. Mehraban, S.M.E. Sepasgozar, A. Ghomimoghadam, B. Zafari, AI-enhanced automation of building energy optimization using a hybrid stacked model and genetic algorithms: Experiments with seven machine learning techniques and a deep neural network, Results Eng. 26 (2025) 104994, https://doi.org/10.1016/j.rineng.2025.104994.
[13] A. Afzal, K.-Y. Kim, J.-W. Seo, Effects of Latin hypercube sampling on surrogate modeling and optimization, International Journal of Fluid Machinery and Systems 10 (Jul. 2017), https://doi.org/10.5293/IJFMS.2017.10.3.240.
[14] G. Chiesa, F. Fasano, and P. Grasso, "A New Tool for Building Energy Optimization: First Round of Successful Dynamic Model Simulations," *Energies (Basel).*, vol. 14, no. 19, 2021, doi: 10.3390/en14196429.
[15] M. Najjar, K. Figueiredo, A.W.A. Hammad, A. Haddad, Integrated optimization with building information modeling and life cycle assessment for generating energy efficient buildings, Appl. Energy 250 (2019) 1366–1382, https://doi.org/10.1016/j.apenergy.2019.05.101.
[16] A. Verma, S. Prakash, A. Kumar, A Comparative Analysis of Data-Driven based Optimization Models for Energy-Efficient buildings, IETE J. Res. 69 (2) (2023) 796–812, https://doi.org/10.1080/03772063.2020.1838347.
[17] K. Bamdad, M.E. Cholette, J. Bell, Building energy optimization using surrogate model and active sampling, J. Build. Perform. Simul. 13 (6) (2020) 760–776, https://doi.org/10.1080/19401493.2020.1821094.
[18] M.V. Shoubi, M.V. Shoubi, A. Bagchi, A.S. Barough, Reducing the operational energy demand in buildings using building information modeling tools and sustainability approaches, Ain Shams Eng. J. 6 (1) (2015) 41–55, https://doi.org/10.1016/j.asej.2014.09.006.
[19] B. Kubwimana and H. Najafi, "A Novel Approach for Optimizing Building Energy Models Using Machine Learning Algorithms," *Energies (Basel).*, vol. 16, no. 3, 2023, doi: 10.3390/en16031033.
[20] S.G. Kangalli Uyar, B.K. Ozbay, B. Dal, Interpretable building energy performance prediction using XGBoost Quantile Regression, Energy Build. 340 (2025) 115815, https://doi.org/10.1016/j.enbuild.2025.115815.
[21] Y. Liang, Y. Pan, X. Yuan, W. Jia, Z. Huang, Surrogate modeling for long-term and high-resolution prediction of building thermal load with a metric-optimized KNN algorithm, Energy Built Environ. 4 (6) (2023) 709–724, https://doi.org/10.1016/j.enbenv.2022.06.008.
[22] N. Shirzadi, D. Lau, and M. Stylianou, "Surrogate Modeling for Building Design: Energy and Cost Prediction Compared to Simulation-Based Methods," *Buildings*, vol. 15, no. 13, 2025, doi: 10.3390/buildings15132361.
[23] M. Manfren, P.A.B. James, L. Tronchin, Data-driven building energy modelling – an analysis of the potential for generalisation through interpretable machine learning, Renew. Sustain. Energy Rev. 167 (2022) 112686, https://doi.org/10.1016/j.rser.2022.112686.
[24] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier,, Modeling and forecasting building energy consumption: a review of data-driven techniques, Sustain. Cities Soc. 48 (2019) 101533, https://doi.org/10.1016/j.scs.2019.101533.
[25] D. Lee, S.J. Quan, Surrogate Modeling Strategy for Urban Building Energy simulation in Early-Stage Urban Design: a Case Study of Energy-Efficient Neighborhood Design in Seoul, E3S Web of Conferences 436 (2023) Jul, https://doi.org/10.1051/e3sconf/202343601016.

[26] I. Ghalehkhondabi, E. Ardjmand, G.R. Weckman, W.A. Young, An overview of energy demand forecasting methods published in 2005–2015, Energy Syst. 8 (2017) 411–447, https://doi.org/10.1007/s12667-016-0203-y.
[27] M. Shikatani, R. Richman, Surrogate modelling methodology for predicting annual site energy for single-family wartime bungalow archetypes in Toronto, Energy Build. 311 (2024) 114122, https://doi.org/10.1016/j.enbuild.2024.114122.
[28] M.W. Ahmad, J. Reynolds, Y. Rezgui, Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees, J. Clean. Prod. 203 (2018) 810–821, https://doi.org/10.1016/j.jclepro.2018.08.207.
[29] Z. Wang, Y. Hong, L. Huang, M. Zheng, H. Yuan, R. Zeng, A comprehensive review and future research directions of ensemble learning models for predicting building energy consumption, Energy Build. 335 (2025) 115589, https://doi.org/10.1016/j.enbuild.2025.115589.
[30] M. El Alaoui, M. Rougui, Examining the Application of Artificial Neural Networks (ANNs) for advancing Energy Efficiency in Building: a Comprehensive Reviews, Journal of Sustainability Research 6 (1) (2024) e240001, https://doi.org/10.20900/jsr20240001.
[31] G. Chen *et al.*, "A Systematic Review of Building Energy Consumption Prediction: From Perspectives of Load Classification, Data-Driven Frameworks, and Future Directions," *Applied Sciences*, vol. 15, no. 6, 2025, doi: 10.3390/app15063086.
[32] J. Huang and S. Kaewunruen, "Forecasting Energy Consumption of a Public Building Using Transformer and Support Vector Regression," *Energies (Basel).*, vol. 16, no. 2, 2023, doi: 10.3390/en16020966.
[33] N. Baqer and P. Rashidi-Khazaee, "Residential Building Energy Usage Prediction Using Bayesian-Based Optimized XGBoost Algorithm," *IEEE Access*, vol. PP, p. 1, Aug. 2025, doi: 10.1109/ACCESS.2025.3544699.
[34] A. Ali, R. Jayaraman, A. Mayyas, B. Alaifan, E. Azar, Machine learning as a surrogate to building performance simulation: predicting energy consumption under different operational settings, Energy Build. 286 (2023) 112940, https://doi.org/10.1016/j.enbuild.2023.112940.
[35] E. Markarian, et al., Informing building retrofits at low computational costs: a multi-objective optimisation using machine learning surrogates of building performance simulation models, J. Build. Perform. Simul. (2024) 1–17, https://doi.org/10.1080/19401493.2024.2384487.
[36] A. Kaplan, M. Haenlein, Siri, Siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence, Bus. Horiz. 62 (1) (2019) 15–25, https://doi.org/10.1016/j.bushor.2018.08.004.
[37] A. M. Khan, M. A. Tariq, S. K. U. Rehman, T. Saeed, F. K. Alqahtani, and M. Sherif, "BIM Integration with XAI Using LIME and MOO for Automated Green Building Energy Performance Analysis," *Energies (Basel).*, vol. 17, no. 13, 2024, doi: 10.3390/en17133295.
[38] Z. Duan, B. Li, Y. Zi, and G. Yao, "Building retrofit multiobjective optimization using neural networks and genetic algorithm three for energy carbon and comfort," Sci. Rep., vol. 15, Art. no. 38076, Oct. 2025, doi: 10.1038/s41598-025-21871-0.
[39] R. Li, Z. Shari, M.Z.A. Ab Kadir, A review on multi-objective optimization of building performance - Insights from bibliometric analysis, Heliyon 11 (4) (2025) e42480, https://doi.org/10.1016/j.heliyon.2025.e42480.
[40] F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, "Addressing Large-Scale Energy Retrofit of a Building Stock via Representative Building Samples: Public and Private Perspectives," *Sustainability*, vol. 9, no. 6, 2017, doi: 10.3390/su9060940.
[41] C. Ecology Inventory of Carbon and Energy (ICE) 2024 database (Version 4.0.
[42] ISPRA, Italian, Greenhouse Gas Inventory 1990–2021: National Inventory Report 2023, Istituto Superiore per la Protezione e la Ricerca Ambientale 2023 Rome.
[43] V.R. Joseph, Optimal ratio for data splitting, Statistical Analysis and Data Mining: the ASA Data Science Journal 15 (4) (Aug. 2022) 531–538, https://doi.org/10.1002/sam.11583.