BUILDEVO: Designing Building Energy Consumption Forcasting Heuristics via LLM-driven Evolution

Anonymous Authors[1]

Abstract

Accurate building energy forecasting is essential, yet traditional heuristics often lack precision, while advanced models can be opaque and struggle with generalization by neglecting physical principles. This paper introduces BUILDEVO, a novel framework that uses Large Language Models (LLMs) to automatically design effective and interpretable energy prediction heuristics. Within an [evolutionary/iterative-refinement] process, BUILDEVO guides LLMs to construct and enhance heuristics by systematically incorporating physical insights from building characteristics and operational data (e.g., from the Building Data Genome Project 2). Evaluations show BUILDEVO achieves state-of-the-art performance on benchmarks, offering improved generalization and transparent prediction logic. This work advances the automated design of robust, physically grounded heuristics, promoting trustworthy models for complex energy systems.

1. Introduction

Accurate building energy forecasting is a cornerstone of modern energy management, vital for enhancing energy efficiency, ensuring grid stability, and enabling effective building operations (Senthil Kumar et al., 2024). Its widespread applications underpin demand-side response programs, the optimization of heating, ventilation, and air conditioning (HVAC), and the seamless integration of renewable energy sources (Chen et al., 2025; Mariano-Hernández et al., 2020). Modeling the complex, interacting energy dynamics inherent in buildings—influenced by their physical structure, operational schedules, and environmental conditions—is thus a central challenge for smart grid balancing and optimized facility management (Bayasgalan et al., 2024; Thomas & Kotsakis, 2020).

However, achieving consistently accurate building energy forecasts is fundamentally difficult. Energy consumption patterns are notoriously complex, exhibiting non-linear behaviors and significant stochasticity due to fluctuating weather, diverse occupant activities, varying building thermal properties, and dynamic equipment states (Lim & Zhai, 2017). Consequently, as evidenced by datasets like the Building Data Genome Project 2 (BDG2), even architecturally similar buildings can display vastly different energy profiles. Early attempts to capture these dynamics often relied on heuristic methods. These included rule-based systems derived from expert knowledge, simplified physical or resistance-capacitance (RC) models, and grey models (e.g., GM(1,1)) for trend estimation with limited data (Hu, 2020; Peña et al., 2016; Li & Wen, 2014). While offering interpretability, such handcrafted heuristics frequently suffer from limited accuracy and poor generalization, as their predefined rules and parameters are challenging to design and calibrate effectively for the diverse and dynamic nature of real-world energy consumption.

Deep learning (DL) methods have since emerged as powerful alternatives, with models like Long Short-Term Memory (LSTM) networks and Transformer-based architectures (e.g., Informer) demonstrating improved predictive accuracy (Zhou et al., 2021; Khalil et al., 2022). Despite their strengths, these DL approaches often present significant practical challenges. They can demand substantial computational resources and vast datasets for training, frequently operate as "black boxes" whose predictions lack transparency and interpretability (Runge & Zmeureanu, 2021), and may exhibit poor generalization to buildings or conditions not seen during training. This latter issue is particularly acute when standard data pre-processing overlooks or inadequately incorporates inherent physical knowledge of the built environment (Jiang & Dong, 2024; Von Krannichfeldt et al., 2024; Miller, 2019). These limitations often necessitate a difficult trade-off with the clarity and robustness offered by simpler heuristic methods, motivating our research question: *Can we automate the design of accurate and interpretable building energy forecasting heuristics that effectively leverage physical knowledge, thereby bridging the generalization and explainability gap with deep learning*

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

methods?

Inspired by the burgeoning field of automated algorithm design that combines Large Language Models (LLMs) with Evolutionary Algorithms (EAs) (Romera-Paredes et al., 2024; Guo et al., 2024), we propose a novel approach to this problem. We hypothesize that the synergy between the generative and reasoning capabilities of LLMs and the structured search and optimization power of EAs can overcome traditional limitations of manual heuristic design, yielding sophisticated, physically-grounded energy forecasting heuristics. To this end, we introduce BUILDEVO, a novel framework wherein LLMs operate within an evolutionary loop to iteratively generate, evaluate, and refine building energy forecasting heuristics directly from observational data, critically guided by underlying physical principles.

Our main contributions are as follows:

- We present BUILDEVO, a novel framework to integrate Large Language Models with Evolutionary Algorithms (EA) for the automated discovery and design of effective, explainable, and specifically physically-informed building energy consumption forecasting heuristics.
- We develop and incorporate mechanisms within BUILDEVO that enable the systematic embedding of physical building information to the LLM-driven heuristic generation and refinement process, enhancing predictive performance and ensuring the physical plausibility of the generated heuristics.
- We conduct empirical validation of BUILDEVO on public datasets (e.g., BDG2), demonstrating its capability to generate interpretable heuristics that achieve competitive forecasting accuracy and exhibit generalization capabilities across diverse building types and energy consumption patterns.

2. Related Works

Heuristic Methods for Building Energy Consumption Forecasting Heuristic approaches provide simpler, often more interpretable alternatives for building energy forecasting, especially with limited data or when expert knowledge is paramount. These include rule-based systems and expert systems (Peña et al., 2016) that leverage predefined operational logic, grey models like GM(1,1) (Hu, 2020) for trend forecasting with sparse data, and simplified physical or grey-box models (Li et al., 2021) utilizing fundamental principles with reduced complexity. While typically less precise than deep learning, their transparency and lower computational cost make them useful for initial assessments or specific control applications.

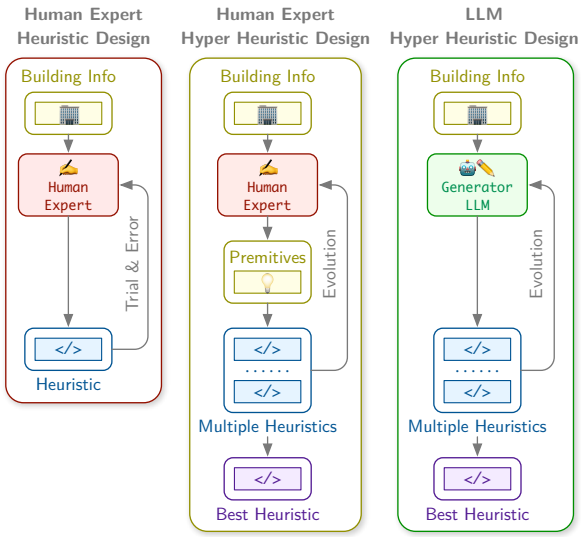

Figure 1. Comparision between human expert heuristic design, human expert hyper heuristic design, and LLM hyper heuristic design.

Learning-based Methods for Building Energy Consumption Forecasting Machine learning and deep learning have notably advanced building energy consumption forecasting, yet challenges in handling complex time series data persist. Initial statistical models like Linear Regression (LR) and machine learning methods such as Support Vector Regression (SVR) and Random Forest (RF) provided foundational capabilities but often fell short with non-linear, non-stationary energy data. Deep learning, particularly Long Short-Term Memory (LSTM) networks(Ahmed et al., 2022) and Transformer-based models like Informer (Zhou et al., 2021), offered significant improvements for long-sequence forecasting. More recently, hybrid strategies combining signal decomposition techniques like Ensemble EMD (EEMD) with deep learning predictors (e.g., EEMD-Informer) and hyperparameter optimization using methods like Particle Swarm Optimization (PSO) have become prominent. Despite these advancements, effectively managing data volatility, optimizing complex model parameters (as in PSO-Informer), and ensuring robust long-term prediction accuracy remain critical challenges, motivating our EEMD-PSO-Informer approach.

3. Methodology

3.1. Problem Definition

Task We address building-level energy consumption forecasting using deep learning models. The task is, for each building b and associated energy meter m, to predict its future energy consumption sequence $Y_{b,m} = (C_{b,m}^{T_{\text{obs}}+1}, \ldots, C_{b,m}^{T_{\text{obs}}+T_{\text{pred}}})$, where $C_{b,m}^{t} \in \mathbb{R}$ is the con-

sumption at future timestep t. This prediction is based on the observed history $H_{b,m} = (\mathcal{S}_{b,m}^1, \ldots, \mathcal{S}_{b,m}^{T_{\text{obs}}})$, where $\mathcal{S}_{b,m}^t = (C_{b,m}^t, X_b^t)$ comprises the past energy consumption $C_{b,m}^t$ and relevant exogenous features X_b^t (including building characteristics, historical weather, and calendar effects) at timestep t. Additionally, known future exogenous inputs $X_b^{\text{future}} = (X_b^{T_{\text{obs}}+1}, \ldots, X_b^{T_{\text{obs}}+T_{\text{pred}}})$ (e.g., weather forecasts, future calendar effects) are utilized. Models are trained on a training set D_{train} and evaluated on a test set D_{test}.

Metrics We evaluate the predictive performance of the deep learning models using standard regression metrics computed on the test set D_{test}. These include, but are not limited to: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These metrics collectively provide insights into the accuracy (RMSE, MAE) and scale-independent relative error (MAPE) of the forecasts.

3.2. Dataset and Preprocessing

We utilized the Building Data Genome Project 2 (BDGP2) dataset (Miller et al., 2020), containing building metadata, weather, and multi-modal energy data, which was preprocessed to create a robust dataset for precise energy prediction. After integrating and temporally aligning these diverse sources, a hierarchical, domain-knowledge-driven imputation strategy addressed missing values across meter types. This involved rule-based imputation (leveraging metadata for operational logic), model-based imputation (using machine learning with temporal, weather, and building metadata features), supplemented by donor-based methods and minimal interpolation. The resulting complete and coherent dataset features intelligently imputed values reflecting physical and operational context, preserving data integrity (details in Appendix A) and forming a solid foundation for subsequent prediction modeling.

3.3. Evolutionary Framework

Our evolutionary framework adapts the Reflective Evolution approach (Ye et al., 2024), employing Large Language Models (LLMs) for core genetic operators (initialization, crossover, mutation), guided by reflective analyses of energy forecasting heuristic performance.

Initial Population The process starts by providing the generator LLM with a task specification for building energy forecasting—including inputs (e.g., historical energy, weather, building metadata like `sqft`, `primaryspaceusage`), outputs (future energy consumption), and the objective J (e.g., minimizing RMSE)—alongside a basic energy heuristic example (e.g., a persistence model). The LLM then generates an initial population of N diverse code-based energy forecasting heuristics.

Selection for Crossover Parent heuristics for crossover are chosen from successfully executed candidates, balancing exploration (70% random selection) and exploitation (30% from elite performers with the lowest forecasting error J).

Reflections BUILDEVO utilizes two reflection types (Ye et al., 2024). *Short-term reflections* compare selected parent heuristics to guide crossover, for instance, by identifying better use of weather data. *Long-term reflections* accumulate insights across generations, pinpointing effective design patterns for energy heuristics, especially concerning the integration of physical building parameters, and generating textual "verbal gradients" (Pryzant et al., 2023) for the LLM.

Crossover Guided by short-term reflections comparing parent performance, the LLM combines elements (e.g., code segments, rules) from two parent heuristics. This aims to synthesize offspring with potentially superior energy forecasting logic, such as improved handling of building metadata or weather impacts.

Elitist Mutation The LLM mutates an elite heuristic (potentially selected by CGES, see below), informed by long-term reflections. This focuses on refining strategies, particularly for incorporating physical insights like occupancy schedules inferred from `primaryspaceusage` or weather normalization techniques.

3.4. Cross-Generation Elite Sampling for Energy Heuristics

Standard evolutionary search for complex building energy forecasting heuristics can stagnate in local optima (Osuna & Sudholt, 2018), as simple mutations often yield only minor improvements. To bolster exploration, BUILDEVO adapts Cross-Generation Elite Sampling (CGES). CGES maintains a historical archive of high-performing energy forecasting heuristics from all past generations. For mutation, instead of only using the current best, an elite heuristic is sampled from this archive via a Softmax distribution based on past forecasting errors J. Reintroducing and modifying these historically successful, diverse strategies (e.g., heuristics adept for specific building types or weather conditions) enhances exploration, aiding escape from local optima and fostering the discovery of more robust energy heuristics.

3.5. Physical Insights Feedback Loop

The aggregate forecasting error J does not reveal the efficacy of specific internal components or rules within a heuristic, especially those leveraging physical building knowledge.

Table 1. Performance results (MAPE (%) / RMSE / MAE) of different prediction methods across generalization.

Method	To College Classroom	To Public Services-B	To College Dormitory	To Office Fox	To Public Services-R
LR	12.99 / 9.76 / 6.76	16.45 / 9.02 / 6.23	84.11 / 4.95 / 2.97	21.78 / 8.11 / 4.71	65.08 / 8.36 / 5.51
SVR	9.71 / 8.20 / 5.12	9.36 / 6.00 / 3.63	84.09 / 4.96 / 2.86	14.55 / 5.71 / 3.54	62.36 / 8.13 / 5.51
RF	11.04 / 8.49 / 5.76	10.66 / 6.51 / 4.03	83.32 / 4.95 / 2.96	19.69 / 6.38 / 4.21	58.51 / 8.33 / 5.71
LSTM	10.00 / 8.09 / 5.23	9.83 / 6.42 / 3.73	79.62 / 4.88 / 2.86	14.04 / 5.28 / 3.43	61.61 / 8.25 / 5.65
Informer	8.69 / 6.88 / 4.61	8.64 / 5.84 / 3.27	29.79 / 3.84 / 2.57	10.84 / 4.60 / 2.88	47.55 / 6.47 / 4.32
PSO-Informer	8.01 / 6.64 / 4.28	8.58 / 5.78 / 3.31	27.15 / 3.75 / 2.50	10.72 / 4.53 / 2.84	46.60 / 6.35 / 4.14
- PIFL	9.10 / 7.35 / 4.75	8.90 / 5.80 / 3.38	73.50 / 4.42 / 2.60	12.75 / 4.80 / 3.10	57.00 / 7.50 / 5.12
BUILDEVO	**7.10 / 5.90 / 3.80**	**7.60 / 5.15 / 2.95**	**24.00 / 3.35 / 2.22**	**9.50 / 3.79 / 2.52**	**43.03 / 5.50 / 3.60**

BUILDEVO incorporates a Physical Insights Feedback Loop (PIFL) to address this. This loop analyzes the contribution of distinct internal logical segments or physically-informed rules—such as those for calculating base load via `sqft` and `primaryspaceusage`, applying weather-dependent adjustments, or modeling occupancy-driven variations. Statistical feedback on the empirical utility of these components is provided to the reflector and mutation LLMs. This guides them to refine how the heuristic effectively utilizes physical building insights and operational contexts for improved performance, steering the evolution towards more physically grounded and accurate energy forecasting heuristics.

4. Experiments

4.1. Experimental Setup

Baselines We compare the heuristics generated by BUILDEVO against several benchmark methods commonly used or relevant for building energy consumption forecasting. These include statistical and machine learning models such as Linear Regression (LR) (Bishop & Nasrabadi, 2006), Support Vector Regression (SVR), and Random Forest (RF); deep learning models like the Long Short-Term Memory network (LSTM) (Hochreiter & Schmidhuber, 1997) and Informer (Zhou et al., 2021), the latter being specifically designed for long-term series forecasting with a self-attention mechanism and an encoder-decoder architecture; and optimization-enhanced models including Particle Swarm Optimization (PSO) (Kennedy & Eberhart, 1995), a classical population-based stochastic optimization technique, and PSO-Informer, which utilizes PSO to optimize Informer's parameters. The performance of our proposed method is evaluated against these established approaches.

Hardware and Software All experiments were conducted on a workstation equipped with an AMD Ryzen 9 7950X 16-Core Processor and a single NVIDIA RTX 5090 GPU. The BUILDEVO framework generates building energy forecasting heuristics as executable Python code snippets in a Python 3.12 environment, employing Google's Gemini 2.0 Flash model (Google, 2025).

4.2. Main Results

We report the performance results against the benchmark methods in Table 1. The heuristics generated by BUILDEVO demonstrate competitive performance. This establishes BUILDEVO as a promising approach for generating effective energy forecasting heuristics. Note that with BUILDEVO, the algorithm doesn't have learnable parameters compared with previous neural baselines.

We also do the ablation study about the PIFL as present in Table 1. Without the PIFL, BUILDEVOstill achieves a competitive performance compared to the neural baselines. By adding the PIFL, the performance of BUILDEVOhas significantly improved, which validates the effectiveness of the PIFL.

4.3. Explainability

A key advantage of BUILDEVO is the generation of explainable Python code for its heuristics. Unlike the "black-box" nature of many DL models, the logic of the evolved heuristics can be inspected and understood. For instance, an evolved heuristic might explicitly combine rules based on time-of-day, weather inputs, and building metadata, offering transparent insights into its forecasting strategy. This interpretability is highly beneficial for building trust, debugging, and deploying in critical energy management systems.

5. Conclusion

We introduced BUILDEVO, a novel framework leveraging LLM and EAs to automatically design physically-informed building energy forecasting heuristics. BUILDEVO generates interpretable heuristics that achieve competitive accuracy and generalize well, as demonstrated on datasets like BDG2. This work offers a new path towards bridging the gap between complex "black-box" models and traditional heuristics, providing a method for creating transparent and effective forecasting tools.

References

Ahmed, D. M., Hassan, M. M., and Mstafa, R. J. A review on deep sequential models for forecasting time series data. *Applied Computational Intelligence and Soft Computing*, 2022(1):6596397, 2022.

Bayasgalan, A., Park, Y. S., Koh, S. B., and Son, S.-Y. Comprehensive review of building energy management models: Grid-interactive efficient building perspective. *Energies*, 17(19):4794, 2024.

Bishop, C. M. and Nasrabadi, N. M. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Chen, G., Lu, S., Zhou, S., Tian, Z., Kim, M. K., Liu, J., and Liu, X. A systematic review of building energy consumption prediction: From perspectives of load classification, data-driven frameworks, and future directions. *Applied Sciences*, 15(6):3086, 2025.

Google. Gemini 2.0 flash, 2025.

Guo, Q., Wang, R., Guo, J., Li, B., Song, K., Tan, X., Liu, G., Bian, J., and Yang, Y. Connecting large language models with evolutionary algorithms yields powerful prompt optimizers. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ZG3RaNIsO8.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Hu, Y.-C. Energy demand forecasting using a novel remnant gm (1, 1) model. *Soft Computing*, 24(18):13903–13912, 2020.

Jiang, Z. and Dong, B. Modularized neural network incorporating physical priors for future building energy modeling. *Patterns*, 5(8), 2024.

Kennedy, J. and Eberhart, R. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pp. 1942–1948. ieee, 1995.

Khalil, M., McGough, A. S., Pourmirza, Z., Pazhoohesh, M., and Walker, S. Machine learning, deep learning and statistical analysis for forecasting building energy consumption—a systematic review. *Engineering Applications of Artificial Intelligence*, 115:105287, 2022.

Li, X. and Wen, J. Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews*, 37:517–537, 2014.

Li, Y., O'Neill, Z., Zhang, L., Chen, J., Im, P., and DeGraw, J. Grey-box modeling and application for building energy simulations-a critical review. *Renewable and Sustainable Energy Reviews*, 146:111174, 2021.

Lim, H. and Zhai, Z. J. Review on stochastic modeling methods for building stock energy prediction. In *Building Simulation*, volume 10, pp. 607–624. Springer, 2017.

Mariano-Hernández, D., Hernández-Callejo, L., García, F. S., Duque-Perez, O., and Zorita-Lamadrid, A. L. A review of energy consumption forecasting in smart buildings: Methods, input variables, forecasting horizon and metrics. *Applied Sciences*, 10(23):8323, 2020.

Miller, C. More buildings make more generalizable models—benchmarking prediction methods on open electrical meter data. *Machine Learning and Knowledge Extraction*, 1(3):974–993, 2019.

Miller, C., Kathirgamanathan, A., Picchetti, B., Arjunan, P., Park, J. Y., Nagy, Z., Raftery, P., Hobson, B. W., Shi, Z., and Meggers, F. The building data genome project 2, energy meter data from the ASHRAE great energy predictor III competition. *Scientific Data*, 7:368, October 2020.

Osuna, E. C. and Sudholt, D. Runtime analysis of probabilistic crowding and restricted tournament selection for bimodal optimisation. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 929–936, 2018.

Peña, M., Biscarri, F., Guerrero, J. I., Monedero, I., and León, C. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Systems with Applications*, 56:242–255, 2016.

Pryzant, R., Iter, D., Li, J., Lee, Y. T., Zhu, C., and Zeng, M. Automatic prompt optimization with" gradient descent" and beam search. *arXiv preprint arXiv:2305.03495*, 2023.

Romera-Paredes, B., Barekatain, M., Novikov, A., Balog, M., Kumar, M. P., Dupont, E., Ruiz, F. J., Ellenberg, J. S., Wang, P., Fawzi, O., et al. Mathematical discoveries from program search with large language models. *Nature*, 625 (7995):468–475, 2024.

Runge, J. and Zmeureanu, R. A review of deep learning techniques for forecasting energy use in buildings. *Energies*, 14(3):608, 2021.

Senthil Kumar, S., Srinivasan, C., and Sridhar, P. Enhancing grid stability and efficiency in buildings through forecasting and intelligent energy management of distributed energy resources. *Electrical Engineering*, pp. 1–18, 2024.

Thomas, D. and Kotsakis, E. Energy management and optimal power scheduling in a smart building under uncertainty. In *Smart Cities-Their Framework and Applications*. IntechOpen, 2020.

Von Krannichfeldt, L., Orehounig, K., and Fink, O. Combining physics-based and data-driven modeling for building energy systems. *arXiv preprint arXiv:2411.01055*, 2024.

Ye, H., Wang, J., Cao, Z., Berto, F., Hua, C., Kim, H., Park, J., and Song, G. Reevo: Large language models as hyper-heuristics with reflective evolution. *arXiv preprint arXiv:2402.01145*, 2024.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

A. Appendix

Data Preprocessing and Loading Protocol

Our study utilizes the Building Data Genome Project 2 (BDGP2) dataset (Miller et al., 2020), a comprehensive collection comprising building metadata, high-frequency weather data, and multi-modal energy consumption readings (electricity, chilled water, hot water, steam, gas, solar, irrigation, and water) across a diverse set of buildings. The primary objective of our data preprocessing pipeline is to create a robust, high-quality dataset suitable for precise energy usage prediction by addressing missing data through techniques informed by built environment characteristics.

A.1. Data Integration and Temporal Alignment

The initial phase involved meticulous integration of the disparate data sources. Building metadata, including structural details (e.g., `sqft`, `yearbuilt`), operational characteristics (e.g., `primaryspaceusage`), and geographical information (e.g., `timezone`, `site_id`), was centralized. Meter readings from various cleaned CSV files, each representing a distinct energy type, were systematically processed. Weather data, including variables such as temperature, humidity, and solar irradiance, was also incorporated.

A critical step was ensuring temporal consistency across all datasets. Timestamps from meter and weather data were parsed and localized using timezone information from the metadata. Weather data, often recorded at a different frequency than meter readings, was resampled (e.g., to hourly averages or sums, as appropriate for the variable) to align with the temporal resolution of the energy consumption data. These datasets were then merged using `building_id` (for metadata) and a combination of `site_id` and the aligned `timestamp` (for weather data), creating a unified analytical dataset for each building and meter type.

A.2. Missing Data Characterization

Prior to imputation, a thorough analysis was conducted to characterize the nature and extent of missing values within each energy meter stream. This involved quantifying missing data percentages and identifying patterns, such as correlations with specific time periods (e.g., time of day, day of week, season), building archetypes (derived from `primaryspaceusage`), or particular weather conditions. This step was crucial for selecting appropriate imputation strategies.

A.3. Imputation Strategy based on Built Environment Knowledge

We employed a hierarchical imputation strategy, prioritizing methods that leverage domain-specific knowledge of building energy systems and operational characteristics over purely statistical approaches.

A.3.1. Rule-Based and Knowledge-Driven Imputation

Initial imputation was performed using predefined rules derived from building metadata and typical operational schedules. For instance, energy consumption for certain meter types (e.g., solar generation during nighttime, or HVAC in unoccupied commercial spaces during holidays inferred from `primaryspaceusage` and temporal features) was imputed with zero or near-zero values where logical. The presence or absence of specific utilities for a building, as indicated in the metadata, also guided this step.

A.3.2. Model-Based Imputation

For remaining gaps, machine learning models were developed to predict missing energy consumption values. For each meter type, regression models (e.g., Gradient Boosting Regressors or Random Forests) were trained on segments of data with complete records. Features for these models included:

- **Temporal Features:** Cyclical time features (hour of day, day of week, month), holiday indicators, and trend components.
- **Weather Features:** Current and lagged weather variables (e.g., temperature, humidity, solar irradiance, wind speed), and derived features like heating/cooling degree days.
- **Metadata Features:** Building characteristics such as `sqft`, `primaryspaceusage` (typically one-hot encoded),

and building age.

The trained models were then used to predict and fill missing consumption values, thereby incorporating the dynamic interplay between energy use, external conditions, and building properties.

A.3.3. Donor-Based Imputation

In instances where model-based imputation was challenging due to extended missing periods or unique building behaviors not well captured by general models, a donor-based approach was considered. This involved identifying ”similar” buildings (based on `primaryspaceusage`, `sqft`, `timezone`, and other relevant metadata) and using their normalized consumption patterns during comparable temporal and weather contexts to inform imputation for the target building.

A.3.4. Temporal Interpolation

As a final step, for any very short, isolated missing data points (e.g., single-record gaps) not addressed by the above methods, standard temporal interpolation techniques (e.g., linear interpolation) were applied sparingly.

This multi-faceted imputation process was applied iteratively to each of the eight distinct meter types. The specific logic for model features and rule-based imputations was tailored where necessary to the unique physical drivers of each energy stream (e.g., solar irradiance for solar generation, temperature differentials for HVAC-related meters like chilled water or steam).

A.4. Final Dataset Assembly

Upon completion of the imputation process for all meter types, the individually processed datasets were merged into a single, comprehensive time-series dataset. An overview of the imputation impact is presented in Figure 2, which showcases the substantial reduction in missing data and the preservation of data integrity (e.g., temporal trends, statistical distributions) achieved through the built-environment-informed methodology. This final dataset, characterized by its completeness and the informed nature of its imputed values, serves as the foundation for subsequent machine learning model development for energy usage prediction.

Figure 2. Visual overview of the built-environment-informed imputation process, demonstrating its effectiveness in reducing missing data while preserving key data characteristics across multiple dimensions.