Assessing data uncertainty in large-scale energy model

Assessing data uncertainty in large-scale energy model

Matteo Piro [1,*], Ilaria Ballarini [1], Vincenzo Corrado [1]
[1] Politecnico di Torino, Turin, Italy
[*]*Corresponding Author:* matteo.piro@polito.it

Abstract

Under the recently approved EPBD (EU 2024/1275), European countries have to develop national building renovation plans relying on Urban Building Energy Modelling (UBEM) and building archetypes.

This paper highlights the role of buildings archetypes by analysing the impact of key input parameters. Using a real case study from Aosta (Italy)—composed by eighteen residential buildings segmented into three archetypes—this work assesses the influence of seven independent parameters through multiple linear regression. Results show that older buildings are most affected by wall thermal transmittance and heating set-point, while newer ones are mainly influenced by heating set-point and infiltration rate.

Key Innovations

- Informatisation of building stock energy performance, for guiding large-scale developers and policymakers
- Demonstration of a scalable workflow for supporting building renovation plan development

Practical Implications

Probabilistic building archetypes mitigate uncertainty in urban building energy modelling (UBEM) by characterising building stock energy properties, defining variation ranges for input data. Our study helps UBEM developers in identifying key influential variables in the building archetype schema, and guiding optimisation toward the most impactful parameters.

Introduction

Background analysis

National building renovation plans are a central element of the newly approved Energy Performance of Buildings Directive (European Commission, 2024). This plan will provides a comprehensive overview of the energy and environmental performance of both residential and non-residential buildings. To effectively map the energy *status* of the urban building stocks, it is essential to analyse building typologies based on climatic zones, use categories, and construction periods.

Urban Building Energy Models (UBEMs) rely on vast datasets, which are often limited or inaccurate. UBEMs can be classified as *bottom-up* or *top-down* models (Swan and Ugursal, 2009). The suitability of one technique over the other depends on the analysis objectives, input data granularity, and correlation with various factors. Bottom-up models are further divided into *physical-based* (engineering), *hybrid*, or *statistical-based* models (data-driven) approaches. The archetype-based approach—where large-scale energy model are developed using building typologies—belongs to the bottom-up engineering category.

Several factors influence the UBEM inaccuracies, including model simplifications, microclimate variations, and the lack of real measurement data, resulting in uncalibrated or non-validated models. Piro et al. (2023) provided a detailed description of UBEM simplifications derived from Building Energy Models (BEMs). Worthy et al. (2025) investigated methodologies and tools from the literature for integrating microclimate modeling into UBEMs to account for urban heat island effects. Oraiopoulos and Howard (2022) highlighted that only 9 % of UBEM studies in the literature were validated, primarily due to the high computational effort required and the lack of high-resolution spatial and temporal data. From this perspective, privacy regulations further limit access to city-wide building information (HosseiniHaghighi et al., 2022; Johari et al., 2023).

To address data uncertainty in large-scale energy analysis, building archetypes (BAs) created by collected data are commonly used. The BA approach offers a balance between model complexity and accuracy in large-scale energy assessment. In this context, energy performance certificates (EPCs) serves as a key data source to bridge uncertainty and are widely used for creating representative buildings. BAs encapsulate average technological characteristics and represent specific segments of the building stock, typically categorised by climatic zone, construction period, and building use category. Non-geometric dataset included in the BA schema are then assigned to the actual 3D building stock geometry. This information can be represented either deterministically—as single fixed values—or probabilistically, using value ranges to reflect data variability (Sokol et al., 2017).

Assessing uncertainty at the urban scale has become a crucial topic and, as mentioned, includes various aspects of Urban Building Energy Modeling. Prataviera et al.

 https://doi.org/10.26868/25222708.2025.1706

(2022) assessed UBEM simulations reliability through uncertainty and sensitivity analyses. Wang et al. (2020) combined Morris and Monte Carlo methods in a UBEM tool to conduct sensitivity analyses. Demir Dilsiz et al. (2023) performed a global sensitivity analysis to assess the impact of various input sets on space heating and cooling demand across different building forms and climates.

Aim of the research

Input data uncertainty in single-building energy models is significantly amplified in large-scale analyses. The archetype-based approach in UBEM is widely used to mitigate this uncertainty by providing representative datasets for defining building envelope and technical building systems characteristics. However, it remains essential to assess how variations in key inputs—within the BA schema—affect the overall energy performance of the building stock.

BAs consist of non-geometric parameters assigned to real building stocks to characterise their energy performance. In this context, probabilistic BA values can assume a discrete uniform distribution, where each value has an equal probability of occurrence.

This paper aims to support and guide UBEM developers in optimising key BA parameters during model creation. The development of a large-scale energy model, even using BAs, is a time-consuming process that benefits from clear recommendations and guidelines. The study focuses on identifying the most influential metrics within the BA schema, mainly derived from the EPCs of the Aosta Valley Region (Italy), taken as a case study. The probabilistic BAs are sourced from the Italian research project Urban Reference Buildings for Energy Modelling (URBEM, 2024), which provides first, second, and third quartiles for each dataset variable.

These data sheets prompt UBEM developers to consider which value should be selected to characterise energy performance, and how this choice influences the final output. Using a three-step interval (first, second and third quartiles), a discrete uniform distribution was applied, and all possible combinations of BA parameters were tested using CitySim (Robinson et al., 2009) to evaluate their impact on space heating thermal energy need.

The parameters analysed—extracted from EPC statistical data—include the thermal transmittance of opaque (walls, floor, and roof) and transparent envelope components, as well as the window-to-wall ratio. Additionally, space heating set-point temperature and infiltration rate were also analysed.

Overall, this study ranks the most impactful metrics within a BA group and compares predominant parameters across different BA schemas.

Methods

Urban Reference Buildings for Energy Modelling

Urban Reference Buildings for Energy Modelling (URBEM, 2024) is an Italian national research project funded under the PRIN 2020 Programme. Its goal is to develop a comprehensive digital library of representative buildings to be used by UBEM tools, accurately reflecting the national building stock to reduce uncertainties in energy simulations. The project ultimately aims to enhance the understanding of building energy performance, improve energy efficiency, and contribute to the long-term sustainability goals.

Italian building archetypes are segmented based on four criteria: climatic zone, building use category, construction period, and, for residential buildings, size and shape. For instance, single-family houses (_SINGLE) are distinguished from units in apartment blocks or multi-family houses (_APPBLOCK).

The BA schema incorporates multiple data sources, including EPCs, local and national databases, questionnaires, and other relevant datasets. Data within the archetype schema are structured into key categories, such as building geometry, thermal envelope characteristics, internal heat gains and ventilation, and technical building system properties. Each category is composed by several qualitative and quantitative variables, where numerical metrics are represented using mean values, standard deviation, and the first, second, and third quartiles.

Methodology

The methodology aims to assess the influence of key parameters within the BA on urban energy modelling using an UBEM archetype-based approach, supporting developers during the model creation phase. The proposed methodology comprises the following steps:

1. extract relevant input data from the BA schema,
2. define the interval of variation and the probability function for each selected independent variable,
3. apply a statistical or data-driven approach to assess the relative influence of each input on the outcome, and finally
4. use a UBEM tool to perform scenario analyses with different input variations, assessing the output variability.

Multivariate regression analysis is a common technique for assessing the influence of input parameters on the building energy need. The general formulation for a multivariate linear regression is given in Equation (1), where Y_i represents the outcome of the large-scale energy model, β_0 is the intercept, β_j denotes the regression coefficients for independent variables x_{ip}, and ε_i is the residual error for the i-th observation.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_j x_{ip} + \varepsilon_i \qquad i = 1, \ldots, n \quad (1)$$

https://doi.org/10.26868/25222708.2025.1706

A common way to interpret the relative importance of each predictor is to calculate the weight w_j for each regression coefficient β_j. This can be defined as the ratio of an individual coefficient to the sum of all coefficients, as reported in Equation (2).

$$w_j = \frac{\beta_j}{\sum_{k=1}^{p} \beta_k} \qquad j = 1, \dots, n \qquad (2)$$

Application

The probabilistic BAs simulated in CitySim were generated using data from the EPC database of the Aosta Valley Region (Italy). The influence of seven independent variables on space heating energy need was assessed for a residential block in the municipality of Aosta, highlighting the significant impact of key UBEM inputs on overall building stock performance. Depending on their construction period, three BA schemas were assigned to characterise the residential block, using archetypes derived from the URBEM project (URBEM, 2025).

CitySim

CitySim, developed by the Solar Energy and Building Physics Laboratory at EPFL (École Polytechnique Fédérale de Lausanne), is a widely recognised and utilised UBEM tool (Robinson et al., 2009). Its calculation engine, CitySim Solver (CitySim Solver, 2024), underpins the graphical user interface developed by the KAEMCO company. CitySim operates as a dynamic hourly energy model, employing a Resistive-Capacitive system that discretise opaque and transparent building envelope components into temperature nodes, thermal resistances, and heat capacities (Emmanuel and Kämpf, 2015).

Case study

The case study focuses on a real residential block located in the municipality of Aosta (583 m a.s.l., heating degree days: 2876 °C·day), comprising 18 buildings. Their construction periods were identified following D'Alonzo et al. (2020). Figure 1 provides an overview of the analysed residential building stock, detailing building codes, construction period ranges, and compactness ratios. Three age ranges were considered: 1919-45, 1946-60, and 1961-70. The _APPBLOCK BA schema was assigned accordingly.

The Level of Detail (LoD) of the urban geometry model was progressively refined through multiple steps. Initially, the building footprint (LoD0) was derived from the OpenStreetMap database. Footprints were then extruded to generate a shoebox representation (LoD1), followed by the incorporation of roof slopes (LoD2). Additional refinements were applied to enhance accuracy, including converting incorrect volumes into shading elements (see Figure 2).

The first (Q_1), second (Q_2), and third (Q_3) quartiles of key parameters—window-to-wall ratio (WWR), thermal transmittance of walls (U_{wl}), roof ($U_{fl;up}$), floor ($U_{fl;lw}$), and windows (U_w)—are presented in Table1a-b. These values were statistically derived from the EPCs of the Aosta Valley Region within the UBEM research framework (URBEM, 2025). Additionally, two independent variables were added across all BA groups: space heating temperature set-point ($\theta_{H;set}$) and infiltration air flow rate (n). According to technical standards EN 12831-1 (CEN, 2017) and EN 16798-1 (CEN, 2018), these variables were assumed to have the following discrete intervals: $n = \{0.30; 0.40; 0.50\}$ and $\theta_{H;set} = \{18.0, 20.0, 21.0\}$, respectively. A discrete uniform distribution function was applied to all inputs.

Figure 1: Satellite view of the city block (GoogleMaps).

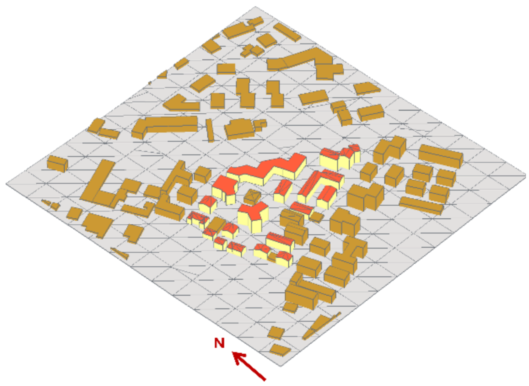

Figure 2: Urban scene imported in CitySim Pro.

Table 1a: BA characteristics per construction period.

	1919-45			1946-60		
	Q_1	Q_2	Q_3	Q_1	Q_2	Q_3
WWR %	0.08	0.11	0.14	0.10	0.13	0.16
U_{wl} W·m^{-2}·K^{-1}	0.87	1.40	1.91	0.82	1.13	1.27
$U_{fl;up}$ W·m^{-2}·K^{-1}	0.88	1.32	1.45	1.09	1.34	1.45
$U_{fl;lw}$ W·m^{-2}·K^{-1}	1.07	1.15	1.36	1.10	1.15	1.20
U_w W·m^{-2}·K^{-1}	2.28	2.61	2.85	2.09	2.69	3.08

https://doi.org/10.26868/25222708.2025.1706

Table 1b: BA characteristics per construction period.

	1961-70		
	Q_1	Q_2	Q_3
WWR %	0.11	0.14	0.16
$\boldsymbol{U_{\mathrm{wl}}}$ W·m⁻²·K⁻¹	0.89	1.12	1.23
$\boldsymbol{U_{\mathrm{fl;up}}}$ W·m⁻²·K⁻¹	1.10	1.32	1.42
$\boldsymbol{U_{\mathrm{fl;lw}}}$ W·m⁻²·K⁻¹	1.11	1.14	1.19
$\boldsymbol{U_{\mathrm{w}}}$ W·m⁻²·K⁻¹	2.16	2.77	3.32

The Typical Meteorological Year file, elaborated by the Italian Thermotechnical Committee (CTI, 2024), was used in simulations. Internal heat gain profiles, and intensities (occupants, appliances, and lighting) were derived from the draft of the Italian National Annex of UNI EN 16798-1 (CTI, 2022).

Implementation of UBEMs

The influence of seven independent parameters on space heating thermal energy need was tested for each BA group using a three-step interval and a discrete uniform distribution. As mentioned, the three discrete values are equivalent to the first, second, and third quartiles of the BA schema. This analysis highlights the significant impact of key UBEM inputs on the overall performance of the building stock (see Figure 3).

The algorithm for testing all possible input sets in combination with CitySim is available (CMA-ES/HDE, 2025).

Three simulation groups were conducted for each BA schema: residential buildings (_APPBLOCK) constructed during 1919-45, 1946-60, and 1961-70. Each simulation group comprised $3^7 = 2187$ runs, for a total of $3^8 = 6561$ simulations.

Within each simulation group, e.g. the first BA group (BA_1), the thermal characteristics of the second (BA_2) and third (BA_3) representative building groups remained constant, while features of the target archetype were varied. To accurately assess short-wave and long-wave exchanges in the urban environment using CitySim, the other two BA simulations were thermally simulated without considering them as shading objects.

Subsequently, the physical values were converted into logical ones {0, 1, 2} to ensure an apple-to-apple comparison when generating the regression line for each building and calculating the weight w_j, as described in Equation (2).

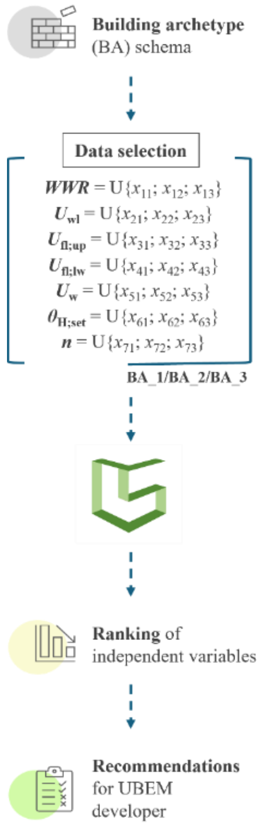

Figure 3: Workflow diagram.

Results and discussion

The municipality of Aosta experiences a climate where space heating represents the dominant energy use. Consequently, only the variation in space heating energy need for different building configurations was evaluated.

For each simulated building, a multivariate regression model was developed to correlate the seven independent variables with space heating thermal energy need. Figure 4 presents the simulation results, expressed in terms of normalised regression coefficients for the 18 residential buildings within the assessed urban city block.

The residential apartment blocks are categorised into three construction periods: 1919-45 (buildings A3 to F5), 1946-60 (buildings B3 to G5), and 1961-70 (buildings A4, A5, and B2).

The space heating temperature set-point ($\theta_{\mathrm{H;set}}$) is the most influential variable across all construction periods, particularly in newer apartment blocks, where it accounts for up to 42 % of the variation. The thermal transmittance of walls (U_{wl}) plays a crucial role in older buildings but decreases in importance for constructions from 1946-60 to 1961-70. In winter condition, window-related parameters—window-to-wall ratio (WWR) and thermal transmittance (U_{w})—have relatively low influence.

The thermal transmittance of the roof ($U_{\mathrm{fl;up}}$) is more significant than that of the floor ($U_{\mathrm{fl;lw}}$). Additionally, infiltration rates are becoming increasingly important.

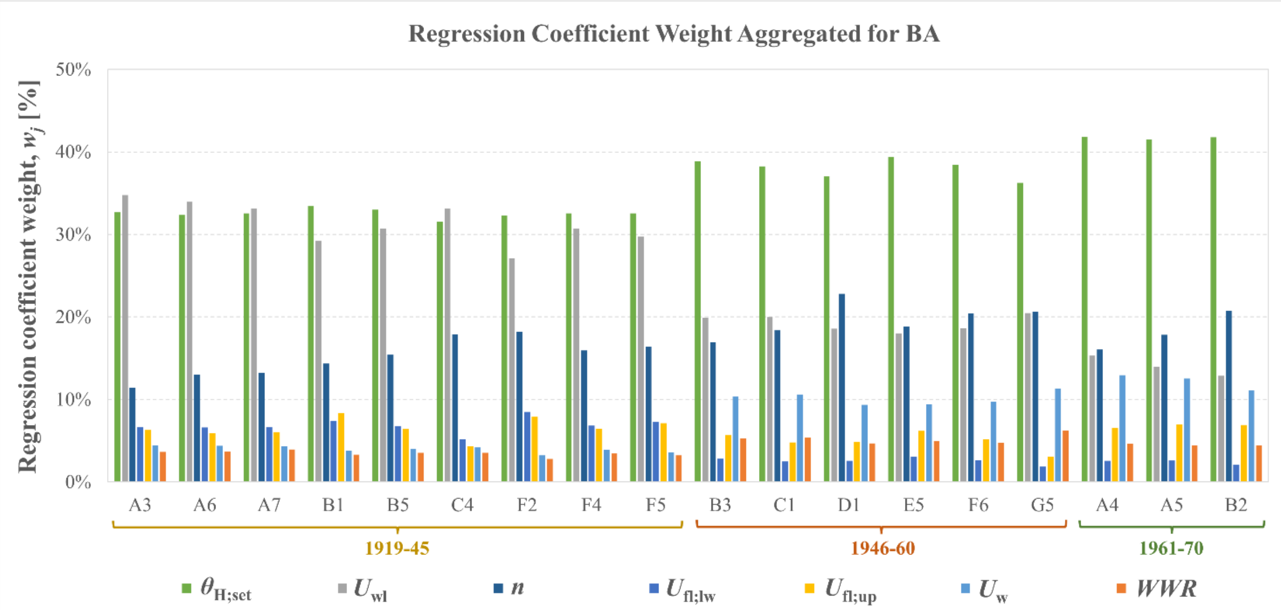

Figure 4: Regression coefficient weight aggregated for BA.

These findings highlight the effectiveness of segmenting the building stock into building archetypes. This is reflected in differences between representative construction, the common trends within the same building age groups, and the varying degrees of influence among independent variables.

In this specific climate context, parameters that require less optimisation during large-scale energy model creation are those related to transparent building envelope components and floor thermal transmittance.

Overall, the proposed methodology provides recommendations for UBEM developers, called to assess the building stock energy performance using building archetypes.

Conclusion

This study highlights the effectiveness of the archetype-based approach in mitigating input data uncertainty within UBEMs. Using a case-study in Northern Italy, the predominant BA parameters were assessed to support the development of the national building renovation plan, as outlined in the recently approved EPBD (EU 2024/1275).

Findings reveal that the heating temperature set-point is the most impactful variable across all construction periods, particularly in more recent buildings. Thermal transmittance of walls is crucial for older buildings, but its impact diminishes in newer constructions. Infiltration rates are increasingly relevant, whereas *WWR* and window thermal transmittance have a minor impact on thermal energy need in this climate.

These results validate the segmentation of the building stock into archetypes, providing a structured framework for large-scale energy modelling. The insights gained can guide UBEM developers in optimising key input parameters, improving energy performance assessments, and informing urban energy planning strategies.

Future work will focus on implementing and validating the developed BAs within UBEMs. Further research will expand scenario analyses by incorporating a larger dataset, enhancing the robustness of the input data uncertainty analysis.

Acknowledgement

The study was developed within the framework of the project URBEM “Urban Reference Buildings for Energy Modelling” (PRIN, code: 2020ZWKXKE), which has received funding from the Italian Ministry of University and Research (MUR).

Nomenclature

Symbols

n	infiltration air flow rate (h^{-1})
U	thermal transmittance ($W{\cdot}m^{-2}{\cdot}K^{-1}$)
w	regression coefficient weight (–)
WWR	window-to-wall ratio (%)
β	regression coefficient (–)
ε	residual error (–)
θ	temperature (°C)

Subscripts

fl	floor	up	upper
H	heating	w	window
lw	lower	wl	wall
set	set-point		

https://doi.org/10.26868/25222708.2025.1706

References

CitySim Solver. (2025). CitySim Solver Official Code web-site. https://github.com/kaemco/CitySim-Solver. Accessed 28 February 2025.

CMA-ES/HDE. (2025). Evolutionary Algorithm - CMA-ES/HDE. https://github.com/jkaempf/EA. Accessed 28 February 2025.

D'Alonzo, V., Novelli, A., Vaccaro, R., Vettorato, D., Albatici, R., Diamantini, C., et al. (2020). A bottom-up spatially explicit methodology to estimate the space heating demand of the building stock at regional scale. *Energy and Buildings 206*.

Demir Dilsiz, A., Ng, K., Kämpf, J., and Nagy, Z. (2023) Ranking parameters in urban energy models for various building forms and climates using sensitivity analysis. *Building Simulation 16*, 1587–1600.

Emmanuel, W., and Kämpf, J. H. (2015). A verification of CitySim results using the BESTEST and monitored consumption values. *Proceedings from BSA2015: 2nd IBPSA-Italy Conference*. Bozen (Italy), 4-6 February 2015.

European Commission (2024). *Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the Energy performance of buildings (recast).*

European Committee for Standardization (CEN) (2017). *Energy performance of buildings -- Method for calculation of the design heat load -- Part 1: Space heating load, Module M3-3 (EN 12831-1).*

European Committee for Standardization (CEN) (2019). *Energy performance of buildings -- Ventilation for buildings -- Part 1: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics - Module M1-6 (EN 16798-1).*

HosseiniHaghighi, S., de Uribarri, P. M. Á., Padsala, R., and Eicker, U. (2022). Characterizing and structuring urban GIS data for housing stock energy modelling and retrofitting. *Energy and Buildings 256*.

Italian Thermotechnical Committee (CTI). (2022). *Draft of the Italian National Annex to EN 16798-1 (document no. 024100360)*. Restricted document. CTI: Milan (Italy).

Italian Thermotechnical Committee (CTI). (2025). Typical Metereological Year (2015). https://try.cti2000.it. Accessed 28 February 2025.

Johari, F., Shadram, F., and Widén, J. (2023). Urban building energy modeling from geo-referenced energy performance certificate data: Development, calibration, and validation. *Sustainable Cities and Society 96*.

Oraiopoulos, A., and Howard, B. (2022). On the accuracy of Urban Building Energy Modelling. *Renewable and Sustainable Energy Reviews 158*.

Piro, M., Ballarini, I., and Corrado, V. (2023). From Building Energy Models (BEM) to Urban Building Energy Models (UBEM): input data and modelling approaches. *Proceedings from BS2023: 18th Conference of IBPSA*. Shanghai (China), 4-6 September 2023.

Prataviera, E., Vivian, J., Lombardo, G., and Zarrella, A. (2022). Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis. *Applied Energy 311*.

Robinson, D., Haldi, F., Kämpf, J., Leroux, P., Perez, D., Rasheed, A., et al. (2009). CITYSIM: Comprehensive Micro-Simulation of Resources Flows for Sustainable Urban Planning. *Proceedings from BS2009: 11th International IBPSA Conference*. Glasgow (Scotland), 27-30 July 2009.

Sokol, J., Cerezo Davila, C., and Reinhart, C. F. (2017). Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings 134*, 11–24.

Swan, L. G., and Ugursal, V. I. (2009). Modeling of enduse energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews 13*(8), 1819–1835.

URBEM. (2025). Urban Reference Buildings for Energy Modelling web-site. https://www.urbem.polimi.it/en/. Accessed 28 February 2025.

Wang, C. K., Tindemans, S., Miller, C., Agugiaro, G., and Stoter, J. (2020). Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam. *Journal of Building Performance Simulation 13*(3), 347–361.

Worthy, A., Ashayeri, M., Marshall, J., and Abbasabadi, N. (2025). Bridging the simulation-to-reality gap: A comprehensive review of microclimate integration in urban building energy modeling (UBEM). *Energy and Buildings 331*.

https://doi.org/10.26868/25222708.2025.1706