POLITECNICO DI TORINO
Repository ISTITUZIONALE

A scalable, automatic, and evolutionary algorithm for calibrating urban building energy models

Original
A scalable, automatic, and evolutionary algorithm for calibrating urban building energy models / Piro, M., Kämpf, J.H., Ballarini, I., Corrado, V.. - In: SUSTAINABLE CITIES AND SOCIETY. - ISSN 2210-6707. - ELETTRONICO. - 144:(2026), pp. 1-19. [10.1016/j.scs.2026.107428]

Availability:
This version is available at: 11583/3010708 since: 2026-07-06T13:03:05Z

Publisher:
Elsevier

Published
DOI:10.1016/j.scs.2026.107428

Terms of use:

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

Publisher copyright

(Article begins on next page)

18 November 2026

Sustainable Cities and Society 144 (2026) 107428

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

A scalable, automatic, and evolutionary algorithm for calibrating urban building energy models

Matteo Piro [a,b,*], Jérôme H. Kämpf [b], Ilaria Ballarini [a], Vincenzo Corrado [a]

[a] Politecnico di Torino, Department of Energy "Galileo Ferraris", TEBE Research Group. Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[b] Idiap Research Institute, Energy Informatics Group. Rue Marconi 19, 1920 Martigny, Switzerland

ARTICLE INFO

Keywords:
Urban building energy modelling
Building archetypes
UBEM calibration
CMA-ES/HDE
Single-building energy model calibration
Building portfolio energy model calibration

ABSTRACT

The national building renovation plan, which has been established under the newly approved Energy Performance of Buildings Directive (EU 2024/1275), can be effectively supported by combining Urban Building Energy Modelling (UBEM) with building archetyping. However, due to inaccessible, unavailable, and inaccurate city-wide building energy consumption data—and the high computational cost of iteratively running large-scale energy models—most UBEM applications remain unvalidated and uncalibrated.

This work presents a novel methodology aimed at improving the credibility, transparency, and reliability of national building renovation plans. Specifically, it introduces an automatic calibration procedure for archetype-based UBEMs, using a scalable optimisation algorithm CMA-ES/HDE, which combines the Covariance Matrix Adaptation Evolution Strategy with Hybrid Differential Evolution techniques. Depending on the spatial resolution of the available energy consumption data, the procedure was applied both at the single-building and aggregated multi-building scales within the CitySim environment.

The methodology was tested on approximately 200 buildings in the municipality of Monthey (Switzerland). Calibration was achieved by minimising the coefficient of variation of the root-mean-square error, *CV(RMSE)*, between predicted and observed daily space heating and domestic hot water demands. Using a *CV(RMSE)* threshold of 25 %, calibration was successful in 43 % of cases at the individual building level; the resulting calibrated parameter ranges were then used to guide the subsequent multi-building calibration.

While the methodology does not incorporate the engineering expertise typically applied in manual calibration, it remains a practical and scalable approach for large-scale UBEM calibration, providing valuable insights for future applications.

1. Introduction

National building renovation plans play a central role in the newly approved Energy Performance of Buildings Directive (EPBD) in supporting decarbonisation of both residential and non-residential buildings (European Union, 2024). A critical step in these comprehensive plans is accurately assessing the current energy performance of the existing building stocks, in such a way to perform reliable energy refurbishment scenario analyses subsequently. To achieve this objective, the use of Urban Building Energy Modelling (UBEM), relying on building archetypes (BAs), is widely recognised by the scientific community as an effective approach. Furthermore, to accurately reflect the real operation of the building stock, the validation or calibration of UBEM outputs it is considered of utmost importance.

1.1. Urban building energy models

UBEM is an essential tool for policymakers, public administrations, urban planners, and real estate managers aiming to assess urban-scale energy efficiency (Perwez et al., 2023), energy vulnerabilities (Xia & Hu, 2024), and city-level energy planning (Li & Feng, 2025; Manso-Burgos et al., 2023). UBEM plays a central role in evaluating both the energy and environmental performance of large portfolios of buildings.

Large-scale energy models are typically classified into *top-down* and *bottom-up* models. *Top-down* approaches rely on historical data to establish macro-level relationships between the energy, technological, and economic sectors. A recent advancement of these models is

* Corresponding author.
E-mail addresses: matteo.piro@polito.it (M. Piro), jerome.kaempf@idiap.ch (J.H. Kämpf), ilaria.ballarini@polito.it (I. Ballarini), vincenzo.corrado@polito.it (V. Corrado).

https://doi.org/10.1016/j.scs.2026.107428
Received 2 October 2025; Received in revised form 10 April 2026; Accepted 21 April 2026
Available online 22 April 2026
2210-6707/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

Quantities

A	area m^2
CV(RMSE)	coefficient of variation of the root-mean-square error -
g	total solar energy transmittance -
n	air change rate h^{-1}
NMBE	normalised mean bias error -
U	thermal transmittance $W/(m^2 \cdot K)$
V	volume m^3
w	regression coefficient weight -
WWR	window-to-wall ratio -

Greek symbols

β	regression coefficient -
θ	temperature °C

Subscripts

env	envelope
fl	floor
H	heating
lw	lower
op	opaque
pers.	occupant
set	set-point
up	upper
W	domestic hot water
w	window
wl	wall

Acronyms

BA	Building Archetype
BEM	Building Energy Model
CMA-ES	Covariance Matrix Adaptation Evolution Strategy
DHW	Domestic Hot Water
EA	Evolutionary Algorithm
EPC	Energy Performance Certificate
HDE	Hybrid Differential Evolution
KPI	Key Performance Indicator
LoD	Level of Detail
UBEM	Urban Building Energy Model

described by Choi and Yoon (2024), who explored the potentialities in connecting *top-down* large-scale energy analysis with natural language model, like GPT-4o. *Bottom-up* models, on the other hand, are based on disaggregated and extensive data and are divided into *engineering* (or *physically*-based), *statistical* (or *data-driven*), and *hybrid* approaches (Hong et al., 2020; Langevin et al., 2020; Swan & Ugursal, 2009). UBEM is generally categorised as a *physically*-based, *bottom-up* model (Reinhart & Cerezo Davila, 2016).

The reliable implementation of large-scale energy analysis faces significant challenges, including the digitalisation of building stock energy performance data, the lack of monitoring data, and the high computational time required for running simulations. From this perspective, UBEMs require extensive information, which are often inaccessible, unavailable, and inaccurate (Goy et al., 2020). This raises critical issues related to data standardisation and harmonisation at municipal or urban scale, and underscores the need for interoperability between databases. As a result, UBEM developers must invest significant effort in data collection (Wang et al., 2022) or employ multi-source data fusion techniques (Choi et al., 2025) to achieve highly accurate datasets.

Another source of uncertainty in UBEM is urban mobility, i.e., the activity and behaviour of occupants across the assessed district or city. Occupant interactions within buildings affect internal heat gains, influencing energy performance predictions. Occupant behaviour can be modelled using either deterministic or stochastic methods, and analyses can be based on either person-level or space-level data (Happle et al., 2018). For instance, Yang et al. (2025) used mobile data traffic to analyse dynamic population distribution and optimise urban space-use efficiency. Samareh Abolhassani et al. (2024) used Wi-Fi sensing data to estimate occupant presence in two university buildings in Canada.

Due to data privacy concerns, building-level consumption data are considered sensitive, which obstacles the credibility of large-scale energy analyses and hinders UBEM validation or calibration processes. As noted by Oraiopoulos and Howard (2022), only 9 % of the UBEM applications found in the literature are validated against measured data. Moreover, increasing model complexity—due to a higher number of buildings or modelled surfaces in the urban scene—raises computational costs. Consequently, UBEM calibration across diverse input datasets remains a complex and unresolved challenge.

Beyond data and large-scale reliability issues, systemic deviations in UBEM can also arise from simplified assumptions used to solve urban energy balance. Insights gained from traditional Building Energy Model (BEM) has allowed the introduction of assumptions and simplifications of the heat exchanges in UBEMs, as described by Salvalai et al. (2024) and Piro et al. (2023). These simplifications involve geometry representation, zoning discretisation, envelope components and technical building systems characterisation, as well as variations in calculation modules between BEM and UBEM.

In UBEM, urban geometry is modelled using the Level of Detail (LoD) classification, originally introduced by Kolbe (2009) and subsequently updated in Kolbe et al. (2021), which ranges from LoD0 – "*highly generalised model*" – to LoD3 "*highly detailed model*". Beyond their role as purely geometric descriptors, the suitability of different LoDs for specific analytical applications has been critically assessed, highlighting that higher geometric detail does not necessarily translate into improved modelling performance for all use cases, particularly in urban-scale environmental and energy analyses (Biljecki et al., 2016). Hwang et al. (2024) investigated geometry-related biases using artificial neural network and Bayesian calibration. Regarding thermal zones subdivision, Johari et al. (2022) evaluated acceptable simplification to balance accuracy and computational efficiency. Building envelope components are typically assigned either by façade orientation or at the whole-building level, and technical building systems are generally simplified to focus on heat generator sub-systems, as described by Piro et al. (2023).

1.2. Building archetypes

To address data uncertainty in large-scale energy analyses, collected data are used to generate building archetypes (BAs). The BA approach offers a trade-off that reduces complexity while enhancing model accuracy. Building archetypes capture the representative technological characteristics of specific segments of the building stock. Specifically, BAs consist of non-geometric datasets that characterise building envelope components, technical building system properties, and pattern of building uses. The BA data schema can be defined from either a *deterministic* or *probabilistic* perspective (Sokol et al., 2017), representing either single values or parameter ranges, respectively. UBEM archetype-based approach is one of several possible *engineering* approaches, proposed by Swan and Ugursal (2009).

The concept of BA is often confused with building prototypes. According to Carnieletto et al. (2021), prototypes are linked to an abstracted representative geometry of buildings, whereas BAs are not geometry-dependant and are assigned to the real building stock.

Although a harmonised and shared methodology to generate BAs does not exist (Borges et al., 2022), the *archetyping* process generally involves two main steps: *segmentation* (or *classification*) and *characterisation* (or *parameterisation*) of the building stock (Pasichnyi et al., 2019). *Segmentation* refers to the classification of similar buildings based on different criteria, such as climatic zone, building use category, construction periods, and more. *Characterisation* involves assigning representative parameters that generalise the energy performance of the segmented clusters.

Widely used data sources for creating BAs libraries are the energy performance certificates (EPCs), as extensively documented in literature, with recent examples including Johari et al. (2023). However, EPCs present several limitations that affect BA data quality: they rely on static energy performance assessments, provide limited energy-related information, and often include low-quality parameters, as described by Piro et al. (2024).

Therefore, the development of an archetype-based UBEM usually relies on the use of a recognised and official library of representative buildings. In this regard, the Italian building typology matrix, originally developed within the pioneering TABULA project (Ballarini et al., 2014), were updated through the outcomes of the URBEM (Urban Reference Buildings for Energy Modelling) project (Ferrando et al., 2025). This latter project represents a substantial effort to characterise the energy performance of buildings across most Italian regions (URBEM, 2025). Similarly, the BUILD_ME (2025) project constitutes another important initiative, providing residential and non-residential building typologies for the territories of Lebanon, Egypt, and Jordan.

1.3. Large-scale validation and calibration

Improving the credibility of national building renovation plans is closely linked to reducing the uncertainty of UBEM outputs. Depending on the purpose of the analysis, this can be achieved through either validation or calibration of large-scale energy models. Model validation assesses whether the modelling framework and assumptions are appropriate by comparing model results with benchmark cases (e.g., standardised procedures such as BESTEST; Judkoff & Neymark, 1995) or measured data. This approach is typically used when aggregate building energy consumption data are available, or the computational cost of full calibration is prohibitive.

Recent works have shown that explicitly quantifying, propagating, and ranking the uncertainty—often through combined uncertainty and sensitivity analyses—associated with key UBEM input parameters is essential to assess model robustness and to identify the most influential assumptions governing large-scale energy simulations (Gu et al., 2023; Prataviera et al., 2022; Wang et al., 2026). Such insights are widely recognised as a necessary foundation for the subsequent parameter-adjustment phase, particularly in large-scale applications. Therefore, beyond uncertainty and sensitivity ranking, it is crucial to assess parameter identifiability, i.e., whether the selected uncertain inputs can be uniquely (or practically) inferred from the available measurements. Otherwise, calibration may yield statistically satisfactory fits while converging towards non-unique, compensatory parameter sets (Yi & Park, 2021). In this context, model calibration involves the manipulation of a vector of input parameters so that the simulation output closely matches measured data, based on statistical metrics, such as the root-mean-square error (*RMSE*), coefficient of variation of the root-mean-square error (*CV(RMSE)*), r-squared (R^2), mean bias error (*MBE*), mean absolute percentage error (*MAPE*). In the context of UBEM, where the complexity increases with the number of buildings considered, automatic calibration is generally preferred to manual fine-tuning, where the engineering judgement is safeguarded (Chen et al., 2020; Faure et al., 2024; Mattsson et al., 2025).

Large-scale energy model calibration can be performed at different spatial resolution, including the building, block, or district scale. According to scientific literature (Fabrizio & Monetti, 2015; Sokol et al., 2017), automatic calibration processes for UBEM can be categorised into three main approaches, which can be used individually or in combination:

- *Bayesian*, which incorporates prior knowledge and observed data to iteratively update the probability distributions of uncertain parameters,
- *Objective/penalty function*, which seeks to minimise the difference between simulated and measured data by systematically adjusting model inputs through deterministic or stochastic algorithms, and
- *Meta-modelling*, which involves building simplified surrogate models that approximate the behaviour of complex simulations, allowing for faster calibration through reduced computational cost.

Objective/penalty-function-based approaches and Bayesian calibration algorithms, for which practical guidelines and their integration within building energy modelling frameworks are discussed in Chong and Menberg (2018), are grouped within the optimisation-based techniques, as defined by Coakley et al. (2014). Table 1 summarises relevant research aimed at improving reliability and credibility of UBEM outputs through validation or automatic calibration approaches, specifying the simulation engine and the utilisation of the building archetypes.

1.4. Aim of the research

The credibility of national building renovation plans is undermined when unvalidated and uncalibrated large-scale energy models are used. In this context, the use of building archetypes represents a trade-off to mitigate input uncertainty in UBEM. However, challenges related to the availability, accessibility, and quality of urban data hinder the accurate informatisation of building stock energy performance. In particular, the lack of high spatial and temporal resolution, city-wide building energy consumption data poses a major barrier to the validation or calibration of large-scale model outputs. At the urban scale, validation is generally preferred over calibration due to the substantial computational resources required for iterative simulations. Consequently, without reliable models and data, the accuracy, reliability, and usefulness of decision-making initiatives driven by UBEM findings cannot be guaranteed.

With the purpose of overcoming the limitations pointed out above, this work is intended to address the following research questions:

- To what extent can optimisation-based calibration procedures, such as CMA-ES/HDE (covariance matrix adaptation evolution strategy and hybrid differential evolution algorithms) evolutionary algorithm, be effectively applied to UBEM?
- How does the calibration performance differ when applied at the single building versus the urban block scale?
- How can statistically representative trends of user behaviour be defined to inform archetype-based UBEMs?

From this perspective, this research presents a novel calibration procedure employing a hybrid CMA-ES/HDE evolutionary algorithm to minimise the *CV(RMSE)* objective function. The novelty of the proposed approach lies in maintaining a scalable approach, applying calibration directly for optimisation at the UBEM scale rather than for training inverse models. The proposed methodology enables the calibration of large-scale energy models at different spatial resolutions, namely at both the single-building and at urban block levels, depending on the spatial granularity of the available building energy consumption data. To ensure computational feasibility when calibrating hundreds of buildings, the UBEM domain is discretised into independent sub-models, thereby reducing both the computational burden and the complexity of the solution space explored by the automatic calibration algorithm. Although simplifications and assumptions are introduced in UBEM thermal models compared to detailed BEM (e.g., simplified geometries,

Table 1
Research works on UBEM validation and automatic calibration.

Reference	UBEM tool (Simulation engine)	Building archetypes	Validation	Calibration		
				Optimisation techniques		
				Bayesian	Objective/penalty function	Meta-modelling
Davila et al. (2016)	umi (EnergyPlus)	x	x			
Cerezo et al. (2017)	n/a (EnergyPlus)	x		x		
Chen et al. (2020)	CityBES (EnergyPlus)					x
Dilsiz et al. (2023)	CitySim (CitySim solver)			x		
Faure et al. (2024)	MUBES (EnergyPlus)			x		
Gholami et al. (2021)	n/a (EnergyPlus)	x		x		
Haneef et al. (2021)	CitySim (CitySim solver)	x			x	
Heo et al. (2012)	n/a (EnergyPlus)	x		x		
Kim and Kim (2024)	umi (EnergyPlus)			x		x
Lundström and Akander (2019)	n/a (ISO14N model)			x		
Nageler et al. (2017)	IDA ICE (IDA ICE engine)		x			
Nagpal et al. (2019)	n/a (EnergyPlus)	x				x
Nouvel et al. (2017)	SimStadt (SimStadt engine)		x			
Sokol et al. (2017)	n/a (EnergyPlus)	x		x	x	x
Tognoli et al. (2023)	CitySim (CitySim solver)	x			x	
Wang et al. (2020)	CitySim (CitySim solver)	x		x		

mono-thermal zoning, generalised technical building system operations, etc.), this framework enables the practical implementation of large-scale calibration while retaining a fully white-box modelling approach.

The procedure was applied to approximately 200 residential buildings in the municipality of Monthey (Switzerland), located in the Canton of Valais, modelled using an archetype-based UBEM approach and simulated in the CitySim environment (Robinson et al., 2009). This sample was selected based on the availability of measured building energy consumption data. The study emphasises the importance of urban data homogeneity in enabling the application of consistent methodologies across diverse urban contexts in different municipalities and countries.

The automatic calibration procedure minimises the *CV(RMSE)* derived by comparing simulated and measured daily building energy demand for space heating and domestic hot water (DHW). This process considers four independent input parameters: air change rate, occupied area per person, daily volume of DHW per person, and space heating set-point temperature.

The structure of this article is as follows: Section 2 theoretically describes the general workflow (2.1), the employed CMA-ES/HDE calibration methodology (2.2), and the description of the statistical model (2.3). Section 2.3 presents the CitySim UBEM tool (3.1), describes the case studies (3.2), discusses the temporal and spatial resolution of building energy consumption (3.3), and illustrates the configuration steps for setting the calibration approach (3.4). Section 4 presents the calibration results at both building (4.1) and urban block scales (4.2), provides a general discussion about the obtained findings (4.3). Finally, Section 5 presents conclusions, limitations, and future work perspectives.

2. Methods

2.1. General workflow

This section provides a theoretical overview of the automatic calibration procedure for UBEM. From this perspective, enhancing the accuracy of large-scale energy models reduces UBEM output uncertainty, thereby improving the credibility of national building renovation plans for the intended target users. The application of this procedure depends on the analysis objective, the computational time required to run the algorithm, and the availability of measurement data at either the single-building or building-portfolio scale. The approach, illustrated in Fig. 1, consists of the following steps:

1. Identifying and extracting the relevant input parameters from the BA schema.
2. Defining the variation ranges and probability distributions for each selected variable.
3. Defining the mathematical objective function of the optimisation algorithm to be maximised or minimised, and/or specifying the maximum number of iterations or the convergence threshold.
4. Using a UBEM tool to assess output variability across different input sets.
5. Calculating the relevant performance metric to meet the objective function for each simulation step.

The automatic calibration procedure employs a fully CMA-ES/HDE (covariance matrix adaptation evolution strategy and hybrid differential evolution algorithms) evolutionary algorithm to adjust model parameters and refine space heating and domestic hot water building stock energy demand estimates. This type of calibration procedure represents

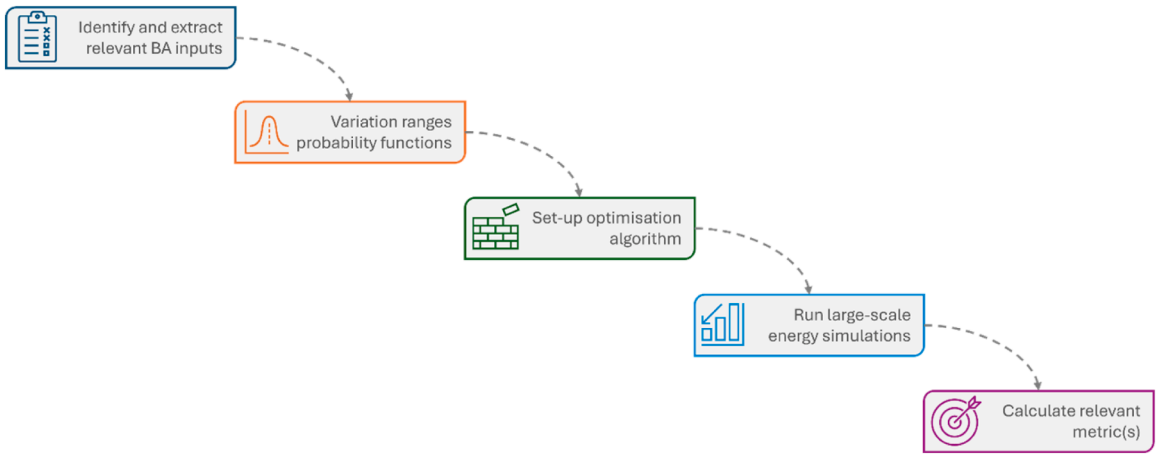


Fig. 1. Methodology for calibrating large-scale energy models.

a novel contribution to the UBEM field, enabling a more accurate alignment between simulated and measured building energy demand. Due to the varying spatial granularity of the energy consumption data, the procedure performs a calibration approach at both single-building and urban block levels, aimed to minimise the coefficient of variation of the root-mean-square error, *CV(RMSE)*. However, depending on the model's complexity—whether at the single-building or the urban-block scale—a predefined maximum number of simulations is set to limit the algorithm's runtime.

CV(RMSE) quantifies how well the simulation results match the observed data; it is expressed as a percentage and is calculated according to Eq. (1). In this formula, y_i is the daily observed or measured energy use from the district heating network, while $\hat{y}_i$ denotes the corresponding daily predicted or simulated value generated by CitySim. The term N is the total number of daily observations, and $\bar{y}$ is the mean of the observed energy use values y_i. The term $\bar{y}$ represents the temporal mean of the observed daily energy use of the same building (or urban block) under analysis and not an average across different buildings. As noted, the temporal resolution of energy consumption varies across buildings; therefore, *CV(RMSE)* is computed only when valid daily measurements are available.

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N}}}{\bar{y}} \cdot 100 \quad (1)$$

In the case of a multiple-building calibration process, the term $\hat{y}_i$ is the sum of the i th daily delivered energy for space heating and domestic hot water across all assessed buildings. This quantity is compared to the observed thermal energy demand of the urban block, y_i.

ASHRAE Guideline-14 (2023) specifies typical *CV(RMSE)* thresholds for determining whether a model can be considered calibrated, based on the temporal resolution of the observed data. Specifically, the acceptable *CV(RMSE)* limit is 30 % for hourly data and 15 % for monthly data. In this research, due to the absence of established thresholds for daily measurements, a *CV(RMSE)* value of 25 %—midway between the two limits proposed by ASHRAE Guideline-14 (2023)—was adopted.

2.2. Calibration through CMA-ES/HDE algorithm

Evolutionary algorithms (EAs) are methods based on natural evolution governing principles that have found application in different engineering fields. The algorithm begins by initialising a population of individuals. A subset is then randomly selected to generate children through crossover and mutation operators. The optimal individuals from both the parent and descendant groups are subsequently selected to form the next population generation. This evolutionary process is repeated iteratively until the termination criterion is met (e.g., a maximum number of iterations or convergence threshold).

The CMA-ES/HDE optimisation procedure, schematised in Fig. 2, belongs to the class of EAs, combining the covariance matrix adaptation evolution strategy (CMA-ES) and hybrid differential evolution (HDE) algorithms (Kämpf & Robinson, 2009; Kämpf et al., 2010). This hybridisation takes advantage from the global exploration capability of differential evolution and the fast local convergence of CMA-ES, offering robust and efficient optimisation across complex, high-dimensional, and multimodal search spaces.

Table 2 summarises the key parameters adopted for the hybridisation of the CMA-ES/HDE algorithm. Two interacting populations evolved within the hybrid CMA-ES/HDE framework. The CMA-ES population consists of λ individuals, where $\lambda = 4 + \lfloor 3\ln(n) \rfloor$ (with n being the number of continuous decision variables), and is characterised by an initial step size $\sigma = 0.2$. At each CMA-ES iteration, candidate solutions are evaluated and ranked, and the best-performing $\mu = \lambda/2$ individuals are selected to update the mean vector and covariance matrix of the sampling distribution.

These μ best solutions are then transferred to the HDE component, which operates with a population size $NP = 30$. Rather than generating new candidates randomly, the HDE population is initialised and evolved around the injected CMA-ES solutions using differential mutation and binomial crossover, with a mutation factor $F = 0.3$ and a crossover probability $Cr = 0.1$. Through this process, HDE performs a local refinement of the search space by generating improved variants of the CMA-ES solutions.

Following a fixed number of HDE evolution steps, the best-performing individuals are returned to the CMA-ES population, contributing to the subsequent update of its search distribution. This bidirectional exchange between CMA-ES and HDE continues iteratively until a termination criterion is met, such as reaching the maximum number of function evaluations, which is defined according to the spatial resolution of the available building energy consumption data.

2.3. Statistical analysis

To quantitatively assess the influence of key parameters on the simulated building energy demand, a quadratic multivariate regression model was developed. The analysis is based on simulation results generated through the iterations of the automatic calibration procedure.

Quadratic multivariate regression is a widely used technique for capturing non-linear physical relationship. The general formulation is reported in Eq. (2), where Y_i represents the simulated building energy consumption, β_0 is the intercept, and β_i, β_{ii}, and β_{ij} denote the coefficients associated with the linear, quadratic, and interaction terms of

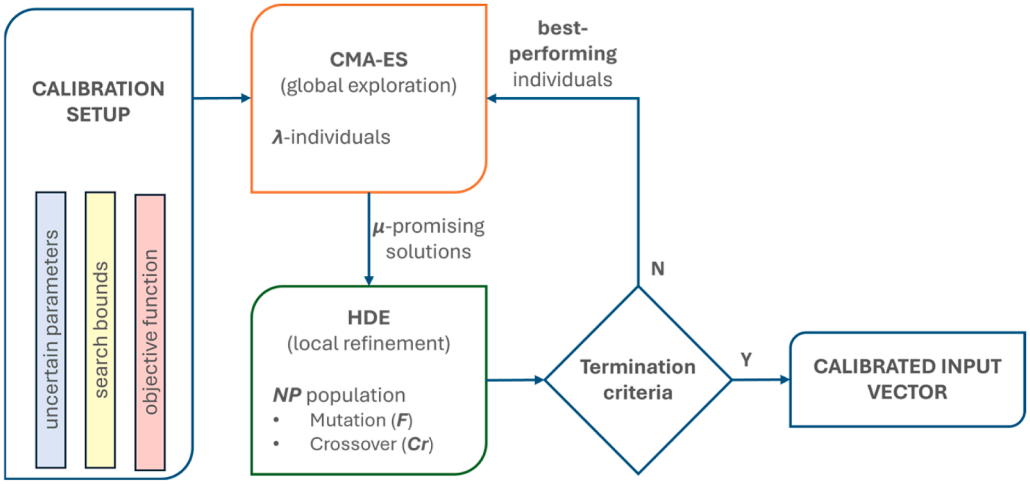


Fig. 2. Conceptual workflow of the hybrid CMA-ES/HDE calibration framework used for UBEM parameter optimisation.

Table 2
Parameters used for the hybrid CMA-ES/HDE algorithm.

Algorithm	Parameters		
CMA-ES	$\lambda = 4 + \lfloor 3\ln(n) \rfloor$	$\mu = \lambda/2$	$\sigma = 0.2$
HDE	$NP = 30$	$F = 0.3$	$Cr = 0.1$

the independent variables x_i, x_i^2, and $x_i x_j$, respectively.

$$Y_i = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j \quad (2)$$

To interpret the relative importance of the main, quadratic, and interaction terms, weights w_{x_i}, $w_{x_i^2}$, and $w_{x_i x_j}$ are computed separately. These weights are defined as the ratio between the absolute value of each coefficient ($|\beta_i|$, $|\beta_{ii}|$, and $|\beta_{ij}|$) and the sum of the absolute values of all regression coefficients, excluding the intercept.

3. Application

An automatic calibration approach employs the EA CMA-ES/HDE (2025) algorithm in a Swiss case study located in the municipality of Monthey. The methodology was applied to approximately 200 residential buildings and simulated in CitySim using an archetype-based UBEM approach, which reduces uncertainty in the characterisation of the buildings' passive energy performance and enables the assessment of energy use patterns associated with user interactions. Archetypes were assigned according to construction period, and non-geometrical attributes were assigned to real building geometries.

The calibration process minimises the *CV(RMSE)* between simulated daily space heating and DHW demands and measured district heating data from 2020. Calibration was performed at both single-building and multi-building scales, depending on the spatial resolution of the available consumption data. Four independent variables were varied within predefined ranges: air change rate, occupied area per person, daily DHW volume per person, and space heating set-point temperature.

3.1. CitySim

CitySim, the successor of SUNtool, is an urban-scale building energy simulation software developed by the Solar Energy and Building Physics Laboratory at EPFL (École Polytechnique Fédérale de Lausanne) (Robinson et al., 2009). It strikes a practical balance among various user demands: input data detail, output accuracy, and computational efficiency. The software's core calculation engine, CitySim Solver (2025), is complemented by a graphical user interface developed by the KAEMCO company (2025), becoming CitySim Pro. CitySim is a dynamic hourly energy calculation engine and its thermal model—validated using the BESTEST methodology (Emmanuel & Kämpf, 2015)—is based on a resistive-capacitive (R-C) system that discretises opaque and transparent building envelope components into temperature nodes, thermal resistances, and heat capacities (Kämpf & Robinson, 2007). The radiative model used in CitySim is based on the Simplified Radiosity Algorithm (SRA), developed by Robinson and Stone (2005). This algorithm accounts for solar and sky obstruction in both shortwave and longwave radiation exchanges. The sky vault is discretised into 145 distinct patches; each representing a specific solid angle and assigned a corresponding radiance value.

3.2. Case study

The case study is in the municipality of Monthey, in the Canton of Valais, Switzerland (430 m a.s.l.). The population of Monthey is around 16,000 inhabitants. The city comprises >2100 buildings, of which 355 were thermally simulated and included in the calibration process due to the availability of building energy consumption data. Fig. 3 reports the building undergoing thermal simulation in yellow with red roofs and the shading objects in brown. Eighty-four percent of the assessed building stock consists of residential buildings, comprising multi-family houses (80 %) and single-family houses (4 %). The remaining 16 % includes 8 % of buildings with unknown uses and non-residential buildings, such as hospitals (3 %), research centres (2 %), offices (1 %), and other use categories.

Geographical open databases of SwissBUILDINGS3D (2025) were utilised for defining the building's geometry information. Then, geometrical building data were converted into a GeoPackage format, distinguishing thermally simulated zones from shading objects (nearby constructions). The GeoPackage files were then imported into CitySim Pro for urban energy performance simulation. The Level of Detail (LoD) for the analysed urban block is LoD1, i.e., buildings represented as simple shoebox geometries. The ground surfaces were neglected during the calibration process for computational efficiency.

The thermal characteristics of both transparent and opaque building envelope components were derived from the findings of Perez (2014). Specifically, the building properties are categorised by construction period and are summarised in Table 3. The BA schema includes, for the opaque building envelope components, the thermal transmittance of walls (U_{wl}), roofs ($U_{fl;up}$), and floors ($U_{fl;lw}$). The geometrical and thermal characteristics of transparent components are represented by the window-to-wall ratio (*WWR*), the thermal transmittance of windows



Fig. 3. Whole UBEM for the city of Monthey.

Table 3
BA thermal and geometrical envelope characteristics.

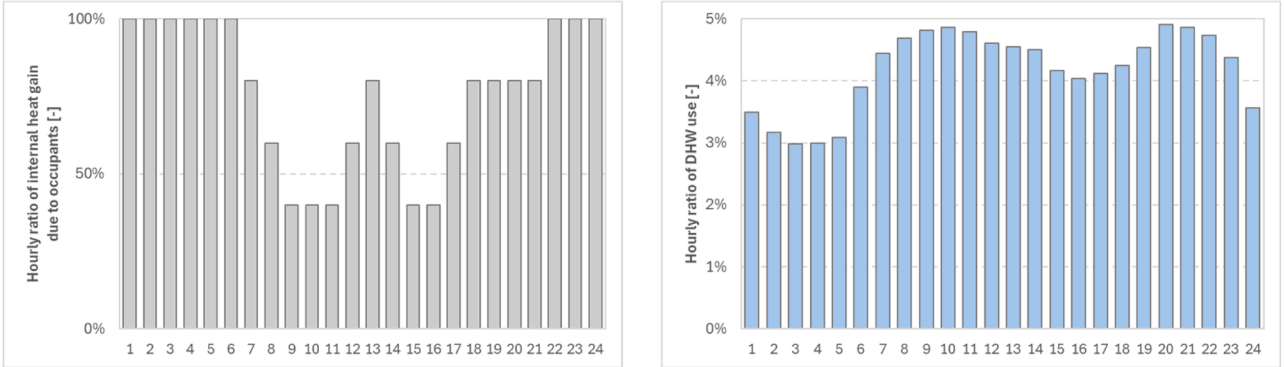
Year	Opaque envelope			Transparent envelope		
	U_{wl} [W/(m^2·K)]	$U_{fl;up}$ [W/(m^2·K)]	$U_{fl;lw}$ [W/(m^2·K)]	*WWR* [−]	U_w [W/(m^2·K)]	*g*-value [−]
1900–45	1.22	0.70	1.60			
1946–60	1.53	0.70	1.50			
1961–70	1.02	0.65	1.30	0.25	2.30	0.47
1971–80	0.55	0.60	1.10			
1981–90	0.89	0.43	0.68			
1991–00	0.69	0.31	0.49			
post-2000	0.51	0.22	0.25	0.35	1.70	0.49

(U_w), and the total solar energy transmittance of glazing (*g*-value).

Standard internal heat gains—arising from occupants, lighting equipment, and electrical appliances—are differentiated by building use category. Fig. 4 reports the hourly ratios of internal heat gains from occupants (Fig. 4a) and DHW use (Fig. 4b) for residential categories; in Fig. 4b the DHW hourly ratio is expressed as a percentage. In the simulations, both the space-heating and DHW systems were assumed to operate continuously (i.e., without intermittent schedules), and the associated user profiles were kept unchanged throughout the year: the daily DHW consumption was assumed constant for all days, and the hourly occupancy-related internal gains followed the same fixed ratio year-round.

To meet the energy demand for space heating and domestic hot water, each building is connected to the district heating network via a substation and a storage unit, both sized according to the building's specific requirements. The heat delivered to the buildings is measured on the client-side; therefore, the substation operates with an equivalent system efficiency of 1.0.

The climatic data were obtained from the weather station located in



(a) (b)

Fig. 4. Hourly ratios of internal heat gains from occupants (a) and DHW use (b).

Aigle (latitude 46.33° N, longitude 6.91° E, and altitude 381 m a.s.l.). This station, situated approximately 10 km from Monthey, was selected as the closest location with available climatic data.

During the iterative automatic EA calibration process, four parameters were adjusted to account for distinct physical phenomena affecting the building's energy balance: the air change rate (n), which affects ventilation heat losses; the occupied area per person (A_{pers}), which influences the occupancy-related internal heat gains; the daily volume of domestic hot water per person (V_W); and the heating set-point temperature ($\theta_{H;set}$), which drives the space heating demand. Due to their sensitivity, these variables are commonly used in large-scale calibration analyses (Sokol et al., 2017; Wang et al., 2020). For each parameter, a continuous variation range was defined, with lower and upper bounds derived from relevant Swiss and European technical standards. From this perspective, Table 4 summarises the four parameters, including their variation ranges and references. The air change rate considers both infiltration and natural ventilation air flow rate. The upper limit of the occupied area per person accounts for the possibility of inhabited buildings without occupants. Overall, the proposed methodology relies on an archetype-based UBEM approach, in which passive thermal properties of opaque and transparent building envelope components, as well as the *WWR*, are prescribed at the archetype level and therefore treated as less uncertain than operational parameters, for which only standardised or conventional values are typically available at the urban scale.

3.2.1. Single-building calibration

The calibration process was carried out approximately on 200 individual buildings. Fig. 5a presents the construction period distribution for multi-unit residential buildings, with a predominance of buildings constructed during 1961–70, 1981–90, and post-2000 periods. Among the individually calibrated buildings, 94 % are residential, with 96 % of those being apartment blocks. Fig. 5b displays the conditioned floor area (A_{fl}) plotted against the compactness ratio ($A_{env} \cdot V^{-1}$) for each assessed building, where this latter is calculated as the ratio between the total thermal envelope area and the building volume.

The non-residential portion, accounting for 6 % of the building stock, includes hospitals, commercial buildings, offices, hotels, and structures with unknown uses. For computational efficiency, each case was modelled using a tailored large-scale energy model that includes surrounding buildings within a 100-meter radius. Fig. 6 shows the complete UBEM scenario, discretised into single-building sub-models for large-scale energy analysis, allowing for the individual energy performance assessment of buildings M3547, M3548, and M3549—taken as an example. This procedure was applied to approximately 200 individual residential buildings subject to the automatic EA CMA-ES/HDE (2025) calibration process.

3.2.2. Urban block calibration

In 24 urban blocks, involving a total of 65 buildings, the district heating substation's energy meter is shared among multiple constructions. Each case study includes between two and five buildings, which are modelled together within a single UBEM scenario. The majority of these case studies consist of either purely residential buildings or a mix of residential and unknown-use buildings, accounting for 14 out of 24 urban blocks. The remaining cases are more heterogeneous, encompassing mixed uses such as educational, hospital, commercial, and office functions, each represented by only one or a few instances (Fig. 7a).

Fig. 7b-c illustrate two example case studies undergoing the iterative automatic calibration process. The first is a purely residential two-building-case (M2839), consisting of apartment blocks M2839 and M2968 (Fig. 7b). The second is a purely non-residential case (M2814), comprising buildings such as M2814, M2810, M2815, M4410, M2813 (Fig. 7c). While officially classified as 'Hospitals' in the geoportal, these structures appear to serve primarily as residential care homes for elderly and disabled individuals.

In both cases, the sub-UBEM model discretisation follows the same approach: the assessed buildings are modelled together with surroundings shading objects within a 100-meter radius.

Table 5 presents the general, geometrical, and thermophysical characteristics of the buildings included in the M2839 and M2814 urban blocks. Specifically, it reports each building's construction period, use category, conditioned floor area (A_{fl}), total thermal envelope area (A_{env}), gross volume (V), window-to-wall ratio (*WWR*), compactness ratio ($A_{env} \cdot V^{-1}$), and mean thermal transmittance for opaque (U_{op}) and transparent (U_w) elements.

3.3. Building energy consumption

The consumption data refer to the delivered energy demand for space heating and DHW from the district heating network. The dataset includes 314 energy meters in the municipality of Monthey, of which 290 correspond to individual buildings and 24 represent aggregated records for urban blocks composed of two to five buildings within the same UBEM scenario.

To ensure sufficient coverage of space heating days, only meters with measurements for more than two-thirds of the year were selected. This threshold guarantees adequate representation of the heating season while improving the robustness of consumption patterns by minimising the effect of impact data. Accordingly, the calibration process was carried out exclusively for days with available data. To maintain consistency with the UBEM output timestep, energy consumption on the leap day was excluded.

The building consumption data vary in both temporal and spatial granularity. In terms of temporal resolution, some meters provide readings every three hours, while others record daily values. Due to this variation and considering the operation of domestic hot water tanks, all measured energy consumption data were aggregated to a daily timestep.

Regarding spatial granularity, some meters are associated with individual buildings, while others are linked to multiple buildings. In the latter case, the calibration process was applied at the scale of the entire urban block.

3.4. Configuration steps

The calibration analysis begins with the creation of UBEM sub-models, as described in Sections 3.2.1 and 3.2.2. For each assessed building, the optimisation algorithm performs a one-year calibration using the combined district energy consumption for space heating and domestic hot water demand recorded in 2020. Consumption data, provided by the energy distributor, were converted to daily values to account for missing hourly entries and to better reflect the operation of DHW storage tanks. As detailed in Section 3.3, only buildings with at least two-thirds daily data coverage were included in the calibration process.

The EA CMA-ES/HDE (2025) algorithm begins by initialising a population and then randomly selecting a subset to produce offspring through crossover and mutation. The best individuals from both the parent and offspring groups are selected to form the next generation, and this iterative process continues until the termination criterion—defined

Table 4
Continuous ranges of the parameters used for automatic calibration.

Parameter	Symbol	Unit	Range	Reference (s)
Air change rate	n	h^{-1}	[0.5; 1.2]	CEN (2007)
Occupied area per person	A_{pers}	m^2/pers.	[28.3; 15,000]	CEN (2019)
Daily volume of DHW per person	V_W	l/(d·pers.)	[10.0; 50.0]	SIA (2006)
Heating set-point temperature	$\theta_{H;set}$	°C	[17.0; 25.0]	CEN (2019)

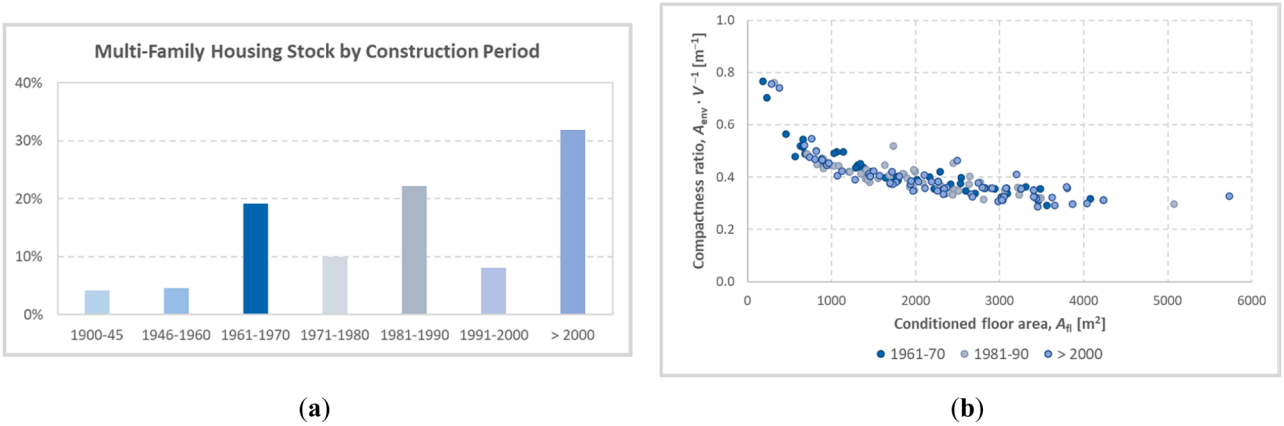


Fig. 5. Multi-family housing stock by construction period (a); relationship between conditioned floor area and building compactness ratio (b).

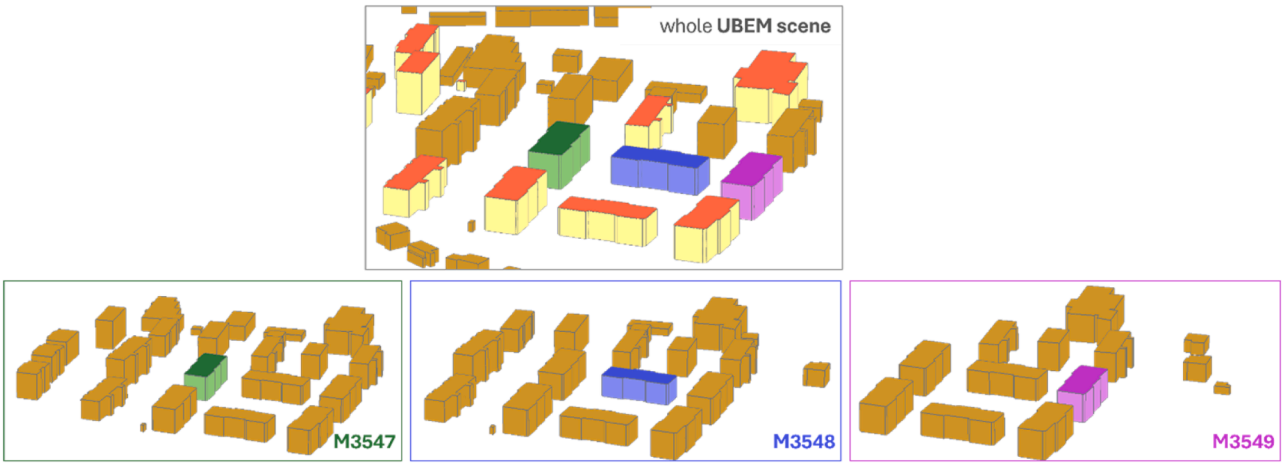


Fig. 6. Example of large-scale energy model discretisation: from the whole UBEM scene to M3547, M3548, and M3549 models with their surroundings.

as the maximum number of iterations—is reached. Increasing the number of calibration parameters, such as by including more buildings, expands the solution space, making it more challenging for the algorithm to efficiently identify the optimal solution. This also results in longer computation times, as a greater number of simulations are required to adequately explore the expanded parameter space. Consequently, depending on the degrees of freedom in the input space—determined by the number of thermally modelled buildings within a given sub-UBEM—the number of simulations needed to achieve convergence increases. Table 6 summarises the maximum number of individuals (simulations) performed for case studies involving one to five buildings.

4. Results and discussion

This section presents the results of the automatic calibration process using the EA CMA-ES/HDE (2025), based on an archetype-driven UBEM approach. The outcomes are organised according to whether buildings were calibrated at the individual or urban block scales. As mentioned in Sections 3.3 and 3.4, and depending on the spatial resolution of building energy consumption and the sub-UBEM configuration, the optimisation process minimises the *CV(RMSE)* and iterates until the predefined maximum number of simulations is reached (Table 6). In both scenarios, the objective is to capture actual occupant behaviour and interactions, aligning them with real building consumption, thereby completing the building archetype schema.

Fig. 8 shows the evolution of the *CV(RMSE)* across five different case studies during the calibration process. In the example provided, the predefined number of individuals tested is sufficient for the function to reach a plateau. Moreover, the variability observed across building and urban block IDs may reflect data-related uncertainty and large-scale modelling simplifications. In such contexts, calibration effectiveness may be hindered by simplified building geometry, uncertainty in measured consumption data, and potentially optimistic archetype-based envelope performance assumptions for specific building groups.

Section 4.1 presents the calibration results focusing on the residential multi-family houses—the most populated building archetype cluster. Section 4.2 illustrates the results obtained applying the procedure at the multiple-building scale for two case studies. Finally, Section 4.3 discusses general considerations regarding the applied methodology.

4.1. Calibration at building-scale

4.1.1. Residential buildings: calibration results

The calibration results for residential buildings constructed in 1961–70, 1981–90, and post-2000 at the building-scale are presented in Fig. 9a. Overall, the calibration process was successful in 43 % of cases, with 12 % of the tested buildings achieving a *CV(RMSE)* below 20 %. When applying the more lenient 30 % *CV(RMSE)* threshold recommended by ASHRAE Guideline-14 (2023) for hourly data, the procedure would deem 66 % of the sample as acceptable. This comparison shows that varying the *CV(RMSE)* threshold mainly affects buildings with values close to the selected limit, while the overall calibration outcome remains consistent.

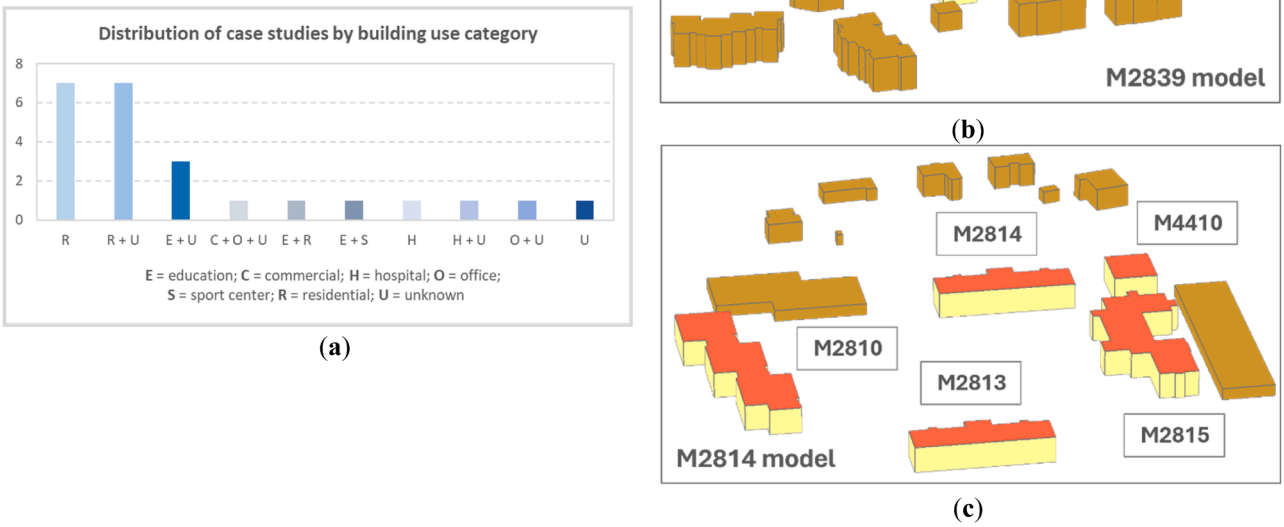


Fig. 7. Distribution of case studies by building use category (a); model M2839 composed by two residential buildings (b); model M2814 composed by five hospitals (c).

Table 5
General, geometrical, and mean thermal characteristics for opaque and transparent building envelope components for urban blocks M2839 and M2814, respectively.

Bldg. code	Age	Use (*)	A_{fl} [m^2]	A_{env} [m^2]	V [m^3]	WWR [-]	$A_{env} \cdot V^{-1}$ [m^{-1}]	U_{op} [$W/(m^2K)$]	U_w [$W/(m^2K)$]
M2839	1961–70	R	1422	1576	4265		0.37	1.00	
M2968	1971–80		2027	2341	6081		0.38	0.70	
M2814			1899	2462	5696		0.43	0.99	
M2810	1961–70		3947	4659	11,841	0.25	0.39	0.99	2.30
M2815		E	3351	4201	10,052		0.42	0.99	
M4410			799	1122	2396		0.47	1.00	
M2813	1971–80		1847	2408	5542		0.43	0.72	

(*) R = residential; E = homes for elderly and disabled people.

Table 6
Maximum no of individuals in CMA-ES/HDE per different scenario analysis.

	No. of individuals
One-building case studies	2000
Two-buildings case studies	3000
Three-buildings case studies	4000
Four-buildings case studies	5000
Five-buildings case studies	6000

Fig. 9b reports the share of case studies that meet the calibration target, i.e., *CV(RMSE)* $\leq$ 25 %, for each individual class and for three construction periods (1961–70, 1981–90, and post-2000). A clear increasing trend is observed across the classes: the frequency of successfully calibrated cases is very low for the smallest classes (around 500 individuals), rises to intermediate values around 1000 individuals, and becomes highest for the largest classes (around 1500–2000 individuals), where the acceptance rate reaches roughly 8–16 % depending on the construction period.

Across all classes, the 1961–70 and 1981–90 cohorts systematically show higher frequencies of successful calibration than the post-2000 buildings, which remain lower especially in the mid-range classes. Overall, the figure indicates that larger individual classes are associated with a higher likelihood of achieving the *CV(RMSE)* threshold, and it highlights a systematic difference in calibration success among construction periods, with newer buildings being more challenging to calibrate in this dataset, potentially due to passive thermal characteristics of their building envelope components.

Residential buildings from the construction periods 1961–70, 1981–90, and post-2000 represent the most populated clusters (Fig. 5a). Fig. 10 illustrates boxplots of the calibrated input parameters: ventilation air change rate (n), occupied area per person (A_{pers}), daily volume of DHW per person (V_W), and heating set-point temperature ($\theta_{H;set}$).

The air change rate accounts for both infiltration and natural ventilation required for hygiene purposes. As shown in Fig. 10a, most buildings converge to $n = 0.5\ h^{-1}$, with a limited dispersion and only a few outliers. This suggests that the ventilation air change rate tends to stabilise around a typical value and is not systematically used as a compensation parameter during calibration. On the other hand, the occupied area per person (Fig. 10b) shows a fairly uniform, with a median of 28.9 m^2/person, close to the conventional value recommended by EN 16,798–1 (CEN, 2019). The calibrated daily DHW volume per person (Fig. 10c) falls within a relatively narrow but high range, with a median of 46.6 l/(d·pers.), consistent with typical residential consumption. Finally, the heating set-point temperature (Fig. 10d) exhibits a median of 21 °C and an interquartile range of 3.2 °C, indicating a moderate variability across the sample. Overall, $\theta_{H;set}$ and V_W show the

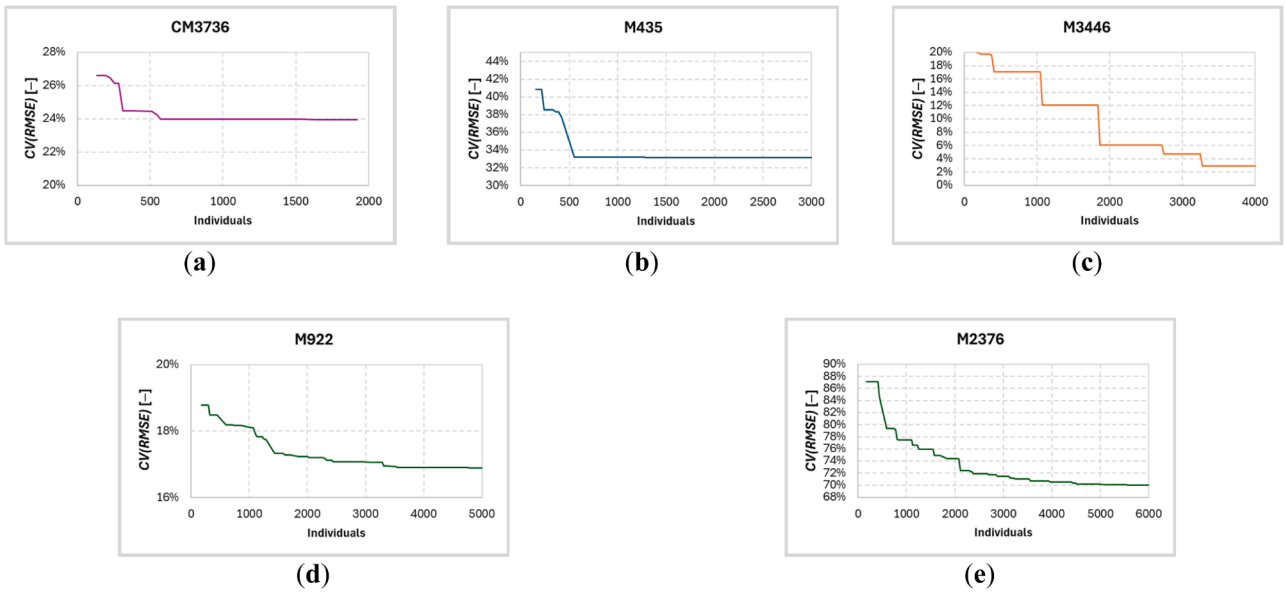


Fig. 8. Evolution of CV(RMSE) values during the calibration process for various case studies: at the building-scale (a), and at the block-scale with 2 buildings (b), 3 buildings (c), 4 buildings (d), and 5 buildings (e).

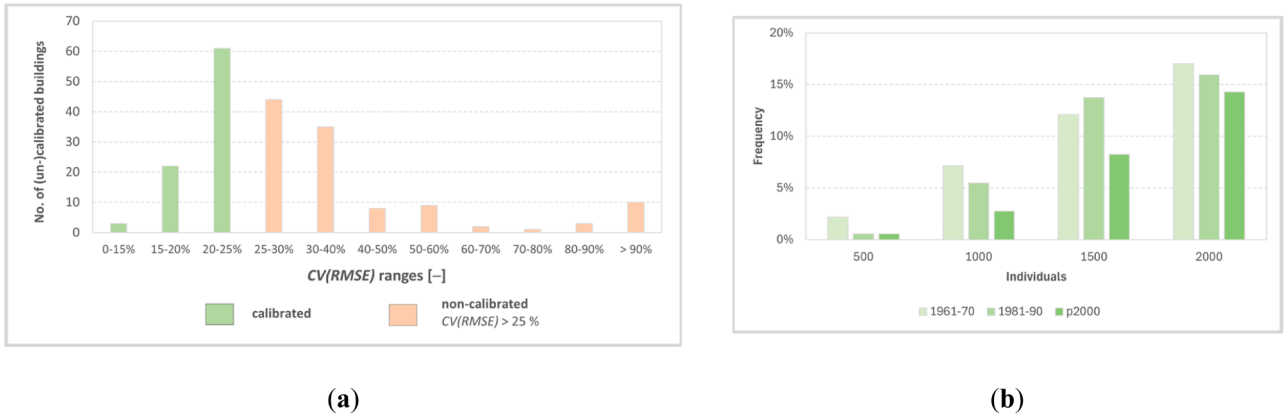


Fig. 9. Distribution of CV(RMSE) ranges across the (un-)calibrated buildings (a); frequency of case studies with CV(RMSE) ≤ 25 % across individual classes, by construction period (b).

largest dispersion among the calibrated inputs. The resulting parameter set provides a realistic representation of building-stock operation, confirming a widely adopted heating set-point around 21 °C, a typical ventilation air change rate of 0.5 h^{-1}, an occupancy density consistent with EN 16,798–1 (CEN, 2019), and DHW volume within expected residential bounds. These findings support the actual plausibility of the proposed automatic calibration procedure.

Fig. 11 shows the monthly distribution of the Normalised Mean Bias Error (*NMBE*) for the analysed dataset, which includes calibrated residential buildings built in 1961–70, computed as $(S - M)/M$, where S represents the simulated monthly energy use and M is the measured one.

In the simulations, the hourly diversity factors for occupancy and DHW consumption (Fig. 4a and b) were assumed constant throughout the year, and the thermal zone was considered continuously used (i.e., no intermittent operation of space heating and DHW technical building systems related to occupancy patterns). Overall, the model performs well during the winter months—especially January, November, and December, and also February—with *NMBE* values clustered close to zero and limited dispersion, indicating a good agreement between simulated and measured consumption when space-heating demand dominates. Conversely, the largest spread is observed in the mid-season months, particularly September and October, when heating operation is more irregular and often absent in reality; under these conditions, the model tends to overestimate consumption, leading to higher positive biases and variability. A distinct behaviour is also visible in July, which is plausibly linked to reduced occupancy during summer holidays, whereas the model maintains constant internal gains and DHW demand, potentially contributing to a mismatch with measured consumption. In this regard, modelling system intermittency could be the subject of future investigations.

4.1.2. *Multivariate quadratic regression analysis*

This section reports the results of the multivariate quadratic regression model for residential apartment blocks constructed between 1961 and 1970. The sample includes >100,000 simulations performed across 54 buildings. Specifically, Fig. 12a presents the regression coefficient weights aggregated by main, interaction, and quadratic effects, while Fig. 12b shows the disaggregated terms.

In this analysis, for each building, the simulated space heating and DHW demand computed with CitySim (Y_i) is related to the variation of four inputs considered during the calibration process: air change rate (x_1), occupied area per person (x_2), daily volume of DHW per person

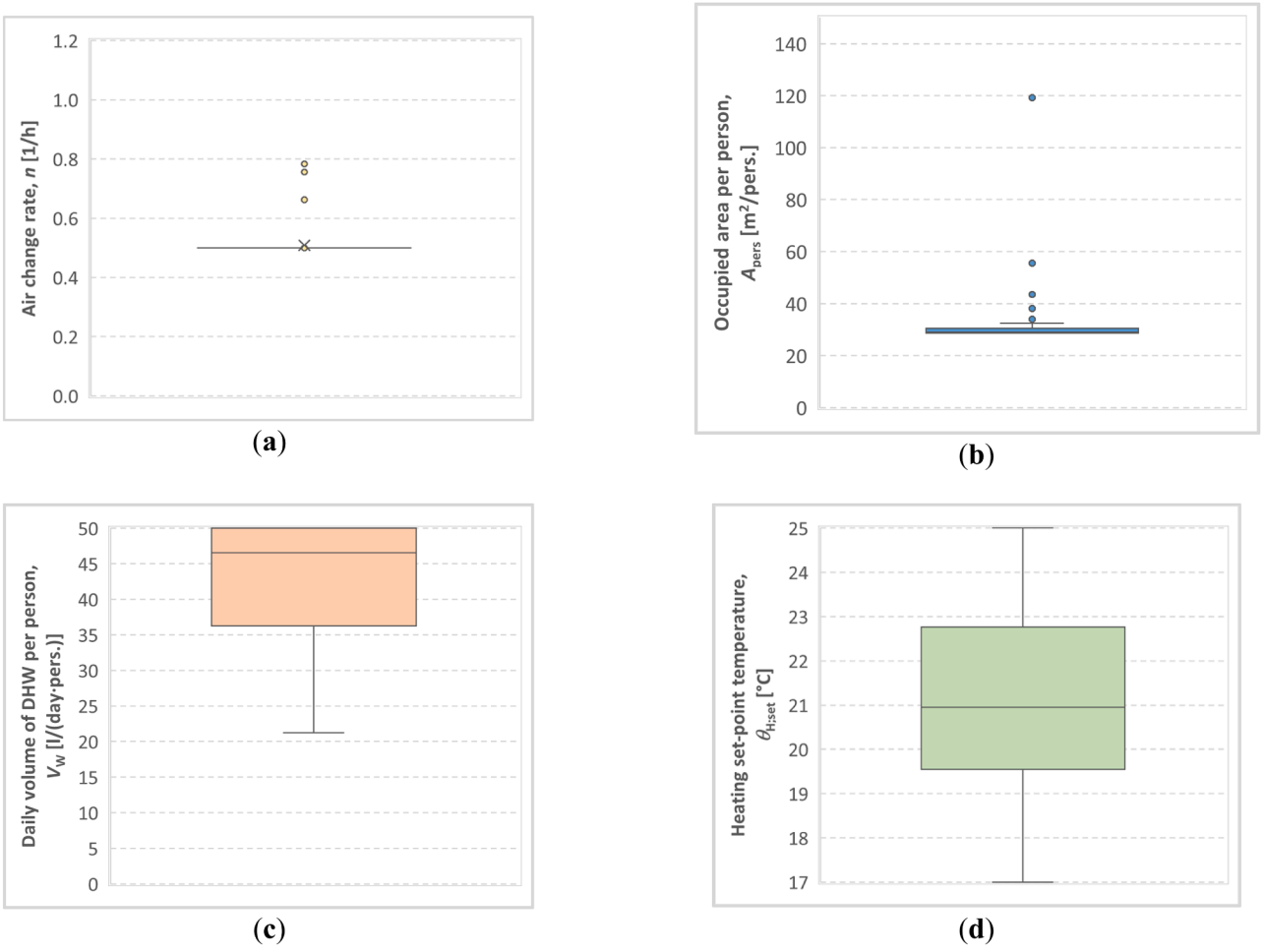


Fig. 10. Boxplots of input variables after calibration process for multi-family houses: air change rate (a), occupied area per person (b), daily volume of DHW per person (c), and heating set-point temperature (d).

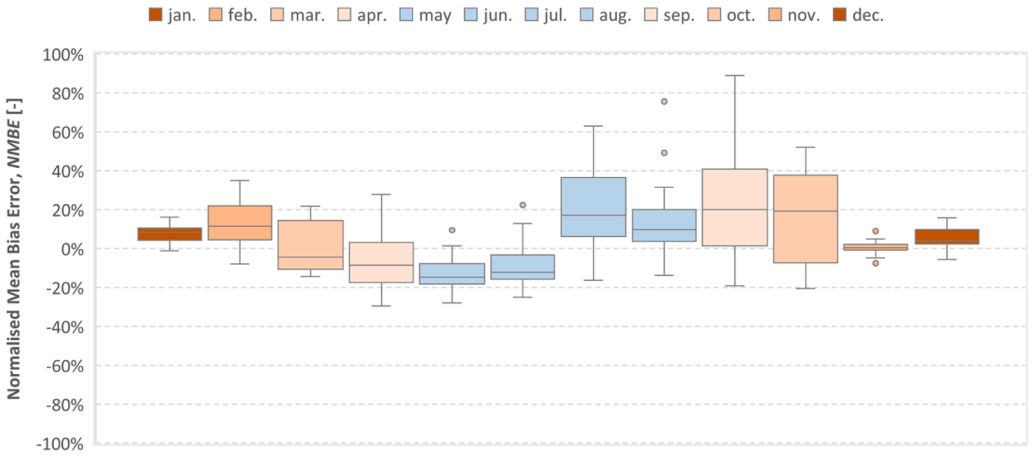


Fig. 11. Temporal distribution of monthly NMBE for calibrated residential buildings built in 1961–70.

(x_3), and the heating set-point temperature (x_4). Prior to fitting the regression model, the independent variables were standardised using a z-score approach.

Fig. 12a highlights the dominant influence of the main effects of the selected variables on the simulated energy demand. The median contribution of first-order effects is approximately 86 %, followed by interaction, which account for a median value of 11 % and reflect the combined influence of pairs of variables. Quadratic terms, representing non-linear curvature in the relationship between predictors and Y_i, exhibit a comparatively smaller contribution.

Fig. 12b shows that variations in heating set-point temperature, followed by ventilation air change rate, have the strongest impact on the overall building energy demand. Whitin this sample, the median weight w_{x_4} is approximately 45 % while w_{x_1} is around 27 %. Regarding interaction effects, the combination of set-point temperature and ventilation air change rate ($w_{x_1x_4}$) primarily governs the sensitivity of space heating

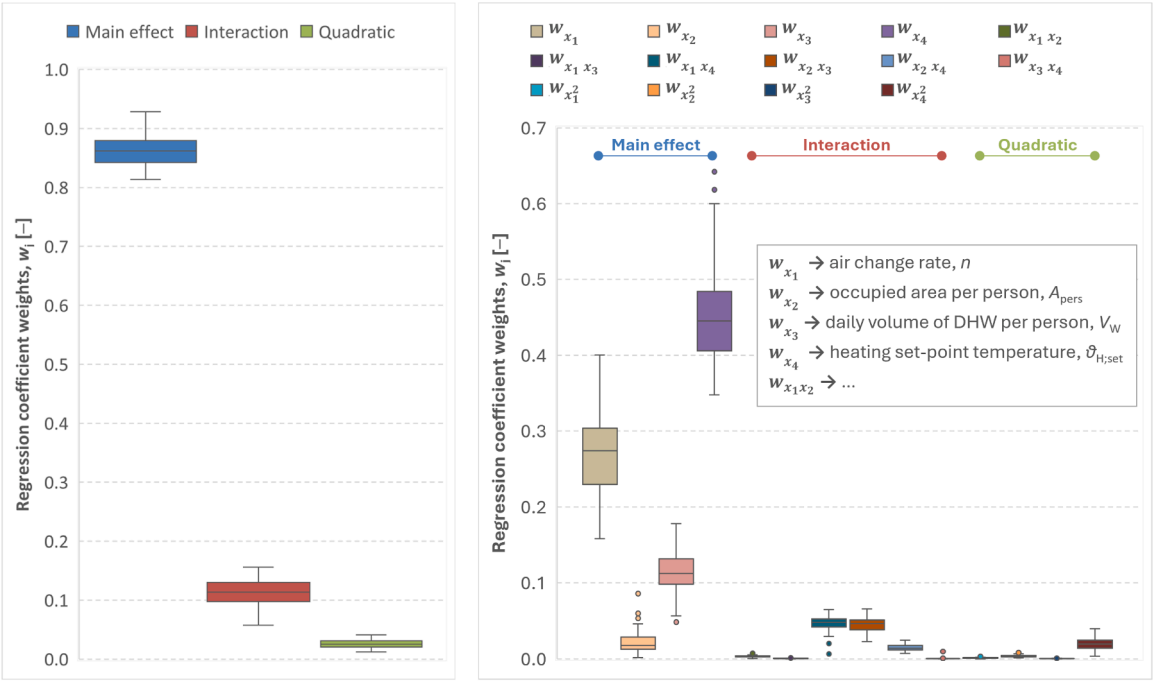


Fig. 12. Weights of regression coefficients for residential buildings constructed in 1961–70: aggregated by main, interaction, and quadratic effects (a), and disaggregated by independent variables (b).

demand, whereas the interaction between the number of occupants in the thermal zone and the daily DHW volume ($w_{x_2x_3}$) is most influential for variations in DHW demand. Among the quadratic terms, the largest contribution is associated with the non-linear effect of the $\theta_{H;set}$. In general, the quadratic regression model provides an accurate representation of the local input-output relationships, achieving a very good fit ($R^2 = 0.99$) across all buildings.

Overall, the analysis identifies the most influential variables and interactions governing building energy demand within the considered sample.

4.2. Calibration at block-scale

The complexity of the calibration approach increases with the number of buildings assessed simultaneously. In the single-building case, the combined space-heating and domestic hot water consumption is driven by four parameters. Extending the same procedure to multi-building calibrations increases the number of unknowns from eight for two buildings to twenty for five buildings. Therefore, to mitigate imbalances in urban-block calibrations—such as unrealistically high space-heating set-points for some buildings and markedly low values for others—and to reduce the solution space explored by the optimiser, the interquartile ranges of the calibrated inputs derived at the single-building scale (Fig. 10) were used to guide the urban-block calibration. Specifically, the allowable input ranges were reduced to A_{pers} = [28.3; 33.3] m^2/pers., V_W = [35.0; 50.0] l/(day·pers.), and $\theta_{H;set}$ = [19.5; 23.0] °C, from A_{pers} = [28.3; 15,000] m^2/pers., V_W = [10.0; 50.0] l/(day·pers.), and $\theta_{H;set}$ = [17.0; 25.0] °C, respectively. The ventilation air change rate was kept constant at 0.5 h^{-1}.

In this section the results after the automatic calibration process through CMA-ES/HDE are presented. Specifically, this section presents the successful results of the automatic calibration process conducted for the case studies M2839 and M2814. Parameters are reported in terms of *CV(RMSE)* for the entire assessed urban blocks, along with the calibrated input values for occupied area per person (A_{pers}), daily DHW consumption per occupant (V_W), and space heating set-point temperature ($\theta_{H;set}$).

4.2.1. Two-building case study

Table 7 reports the results of the calibrated process for the residential block M2839, composed by two apartment blocks. Achieving a *CV (RMSE)* of 16.4 %, the optimiser keeps nearly constant both the occupied area per person and the daily DHW consumption per occupant, while primarily acting on the space-heating temperature set-point. The results reveal a compensation effect within the simulated case study, i.e., tendency of the model to adjust one or more parameters to match the observed energy use. In this case, the algorithm increases the heating set-point of the building with the largest conditioned volume (M2968), while compensating by reducing the $\theta_{H;set}$ of M2839 down to 19.6 °C. Although M2968 is a relatively recent building with a lower average opaque thermal transmittance (Table 5), adjusting its winter set-point temperature proves more effective due to its larger total thermal envelope surface (A_{env}). Consequently, the overall heat transfer coefficient of building M2968 is higher than that of building M2839, making the former more influential in regulating the aggregated heating demand of the urban block.

This compensation effect is highlighted in Fig. 13a and b, where the coloured piecewise linear curves represent partial optima solutions identified by the hybrid differential evolution algorithm, i.e., the best individuals, while the red curve denotes the optimal configuration achieving a *CV(RMSE)* of 16.4 % for buildings M2839 and M2968, respectively. The partial optimal corresponds to very similar performance levels, with *CV(RMSE)* values ranging between 16.6 % and 16.4 %, highlighting the presence of multiple near-equivalent solutions.

Although all inputs are normalised between 0 and 1, clear differences

Table 7
Calibration results for urban block M2839.

Block ID	Bldg. code	*CV(RMSE)* [–]	A_{pers} [m^2/pers.]	V_W [l/(d·pers.)]	$\theta_{H;set}$ [°C]
M2839	M2839	16.4 %	32	35.0	19.6
	M2968		32	35.5	20.8

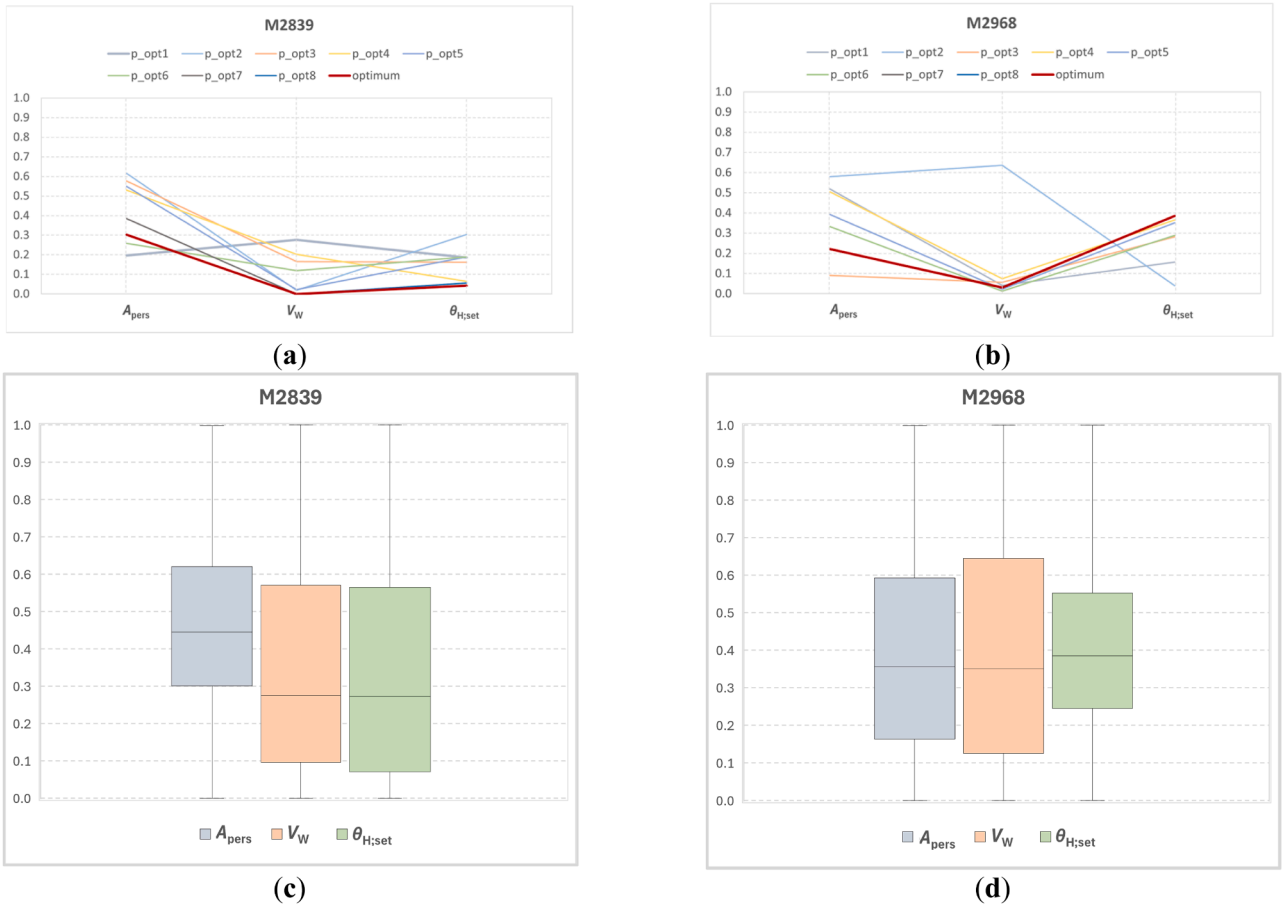


Fig. 13. Urban block M2839: partial and global optima and full distribution of simulated inputs, for M2839 (a-c) and M2968 (b-d) multi-family houses, respectively.

in parameter behaviour emerge. In both buildings, the occupied area per person exhibits the largest variability, indicating its role as a compensating parameter. However, because the admissible range was deliberately narrowed, this apparent spread corresponds to only small absolute changes. Conversely, the space-heating set-point of building M2839 shows markedly lower variability compared to the other inputs, suggesting that it acts as a constraining parameter in maintaining the overall energy balance of the urban block.

This trend is further confirmed by Fig. 13c and d, which depict the full distribution of the tested input parameters explored by the optimiser using the same normalisation criteria. The concentration of solutions around narrow ranges for selected variables, alongside broader dispersions for others, indicates that the optimisation process relies on compensatory trade-offs among inputs, with the heating set-point of building M2968 playing a dominant role in regulating the equilibrium, while the remaining variables adapt around this balance.

Fig. 14a and b compares daily measured and simulated delivered thermal energy for space heating and domestic hot water for block M2839. In Fig. 14a the model shows a strong agreement with observations (R^2 = 0.95), with data points closely distributed around the bisector, indicating consistent performance across the range of daily loads. The time-series comparison, reported in Fig. 14b, further confirms

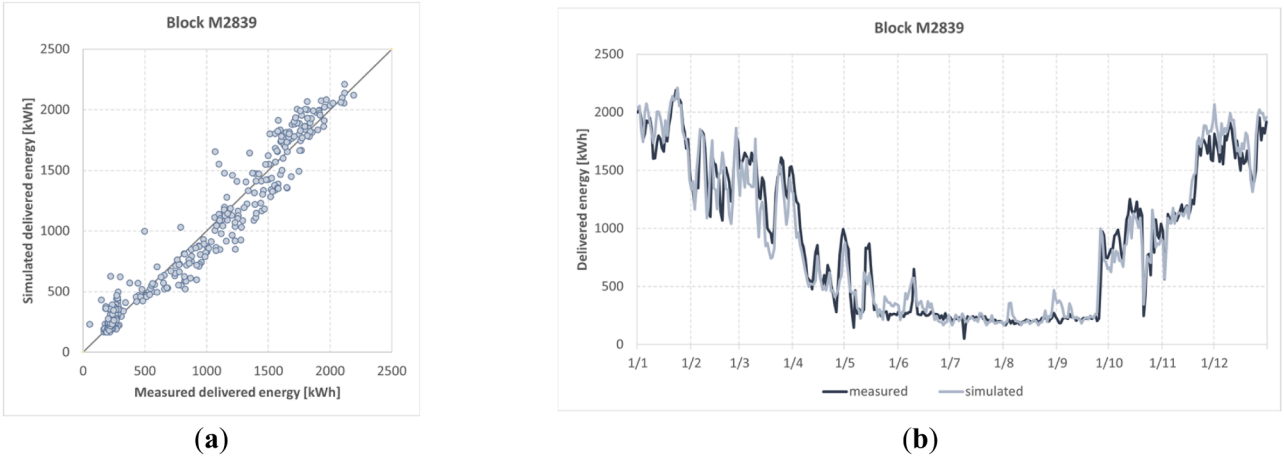


Fig. 14. Daily measured vs simulated delivered energy for urban block M2839: measured-simulated (a) and time-series comparison (b).

that the simulation reproduces the seasonal pattern and the timing of the main demand variations. Residual discrepancies are partly attributable to the absence of an explicit modelling of intermittent system operation and occupant-driven usage patterns, which were assumed to be continuous over the year and can therefore limit the model's ability to reproduce short-term fluctuations.

4.2.2. Five building case study

For the residential urban block M2814, composed of five buildings, the optimiser achieved a *CV(RMSE)* of 21.6 % after the generation and evaluation of approximately 6000 individuals; the corresponding results are reported in Table 8. In this case, the input vector minimising the objective function maintained constant value across all five buildings for the daily DHW consumption per occupant $V_W = 35.0$ l/(d·pers.) and for the space heating set-point temperature $\theta_{H;set} = 19.5$ °C. Consequently, the automatic calibration process mainly adjusted the occupied area per person, with values ranging between 29 and 31 m^2/pers. In this context, lower occupant densities are typically associated with smaller buildings, while higher densities tend to be assigned to larger buildings. These results remain physically acceptable; however, the optimiser predominantly adjusts a limited subset of parameters while keeping others constant, indicating limited identifiability at the urban-block scale. This behaviour, already observed in the single-building calibration, becomes more pronounced at larger spatial resolution, where the number of unknown terms in the mathematical system increases and compensation effects among parameters are more likely to occur.

Fig. 15 reports the monthly delivered energy intensity for the urban block M2814, normalised by the total conditioned floor area. The stacked columns represent the simulated energy demand of the individual buildings (M2814, M2810, M2815, M4410, and M2813), while the adjacent grey bars indicate the observed monthly energy consumption. The results show that the overall energy demand of the urban block is largely driven by a limited subset of buildings. In particular, buildings M2810 and M2815 provide the dominant contribution throughout the year, especially during the heating season (January-March and October-December), where they largely determine the magnitude of the winter peaks. Together, these two buildings account for >60 % of the urban block's total conditioned floor area (Table 5).

The figure further emphasises the complexity of the calibration problem, as the dominant role of specific buildings increases the sensitivity of the optimisation to a reduced number of parameters. In addition, a systematic underestimation of the simulated combined space-heating and domestic hot water demand is observed during the coldest months, namely January and December. This discrepancy may be attributable to the assumption of non-intermittent operation of the space heating and DHW technical building systems. It is further exacerbated by the representation of each building as a single thermal zone, which prevents the model from capturing heterogeneous demand profiles across different units within the same building.

At a daily temporal resolution, Fig. 16a presents the measured-simulated comparison, while Fig. 16b depicts illustrates the temporal distribution of delivered energy. The comparison in Fig. 16a shows a strong correlation between simulated and observed values ($R^2 = 0.94$). Both figures confirm the presence of a mismatch between simulated and observed energy demand during winter months, consistent with the trends identified at the monthly scale.

Table 8
Calibration results for urban block M2814.

Block ID	Bldg. code	*CV(RMSE)* [–]	A_{pers} [m^2/pers.]	V_W [l/(d·pers.)]	$\theta_{H;set}$ [°C]
	M2814		30		
	M2810		29		
M2814	M2815	21.6 %	29	35.0	19.5
	M4410		31		
	M2813		31		

4.2.3. Discussion

The calibration results highlight both methodological and modelling limitations inherent to multi-building calibration at the urban-block scale. When simultaneously calibrating multiple buildings, the optimisation process exhibits a clear compensation effect, whereby discrepancies in one building are offset by adjustments in others. This effect is observed in both single-building and large-scale calibrations, where compensatory parameter adjustments can occur; however, it is particularly pronounced at the aggregated (urban-block) scale. In this case, larger buildings tend to dominate the total demand—compared to single-building calibrations—due to their greater conditioned volume and envelope surface. As a consequence, the optimiser preferentially acts on parameters associated with these buildings, such as the space-heating set-point temperature, to minimise the overall error metric.

In this context, discrepancies between the parameters obtained through single-building and group-level calibration are expected, as the two approaches target fundamentally different optimisation objectives. While single-building calibration identifies building-specific optima that best reproduce the measured consumption of each individual building, urban-block calibration seeks a single (shared) parameter set that minimises an aggregated error metric across multiple buildings. The resulting solution therefore represents a compromise across buildings with different thermal characteristics and demand profiles, which can lead to deviations from the individual-building optima.

This behaviour underscores the need to constrain the variability range of the calibrated parameters. Broad parameter bounds increase the likelihood of compensatory solutions that achieve good statistical agreement with measured data while converging towards values that may be weakly identifiable or only marginally physically plausible. Restricting the admissible parameter space is therefore essential to steer the optimisation towards solutions that are not only statistically robust but also physically meaningful.

Beyond calibration-specific aspects, the results reflect intrinsic limitations of the modelling framework. The UBEM adopted in this study relies on a typical single thermal zone representation per building, which restricts its ability to capture intra-building thermal heterogeneity and zone-level operational differences. Moreover, the model does not explicitly represent intermittent system operation, either in time or intensity. Heating schedules and occupant-driven variability are therefore assumed to be continuous (Fig. 4), limiting the model's capability to reproduce short-term fluctuations and peak demand dynamics.

Finally, the interpretation of calibration performance is further complicated by the use of aggregated measured energy data, combining space heating and domestic hot water consumption, where a relevant example is reported in Fig. 15. This aggregation obscures the contribution of individual end uses and reduces the model's ability to attribute mismatches to specific demand components.

4.3. General considerations

The Urban Building Energy Models are inspired by the experience developed through building energy simulation tools, incorporating assumptions and simplifications to represent heat exchanges at the urban scale. As the scale of analysis increases, the level of geometric-informative details in UBEMs tends to decrease, amplifying the uncertainty associated with input data coverage in large-scale energy assessments. This challenge is exacerbated by the lack of city-scale energy consumption data with sufficient spatial and temporal resolution, which hinders effective model calibration and validation. Additionally, due to the high computational demands of models, grey-box and black-box approaches are typically preferred. Combined with limited availability of reliable consumption data, these factors reduce the feasibility of iterative validation processes required to improve the accuracy and reliability of UBEM outputs.

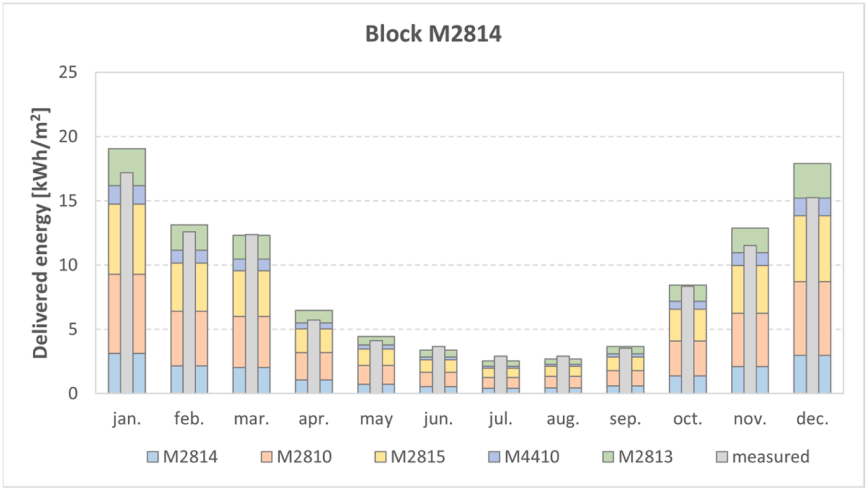


Fig. 15. Monthly delivered energy intensity for the urban block M2814, expressed in kWh/m^2. The stacked columns represent the simulated contributions of the individual buildings (M2814, M2810, M2815, M4410, and M2813), while the adjacent grey column indicates the measured energy consumption for each month.

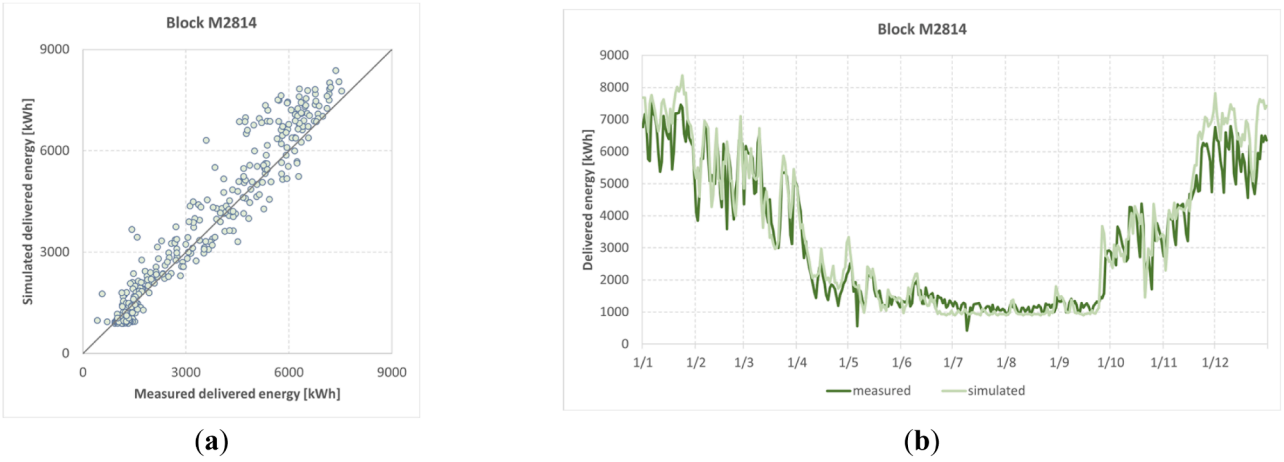


Fig. 16. Daily measured vs simulated delivered energy for urban block M2814: measured-simulated (a) and time-series comparison (b).

The UBEM field still lacks in standardisation. As a result, modelling assumptions are often derived from technical standards, reports, and recognised references developed for building energy modelling. For instance, ASHRAE Guideline-14 (2023) defines reference thresholds for *MBE* and *CV(RMSE)* to assess calibration in BEM tools. However, given the greater level of assumptions, simplifications, and uncertainties, it is questionable whether the same statistical acceptability ranges can be applied. It is likely that less severe statistical thresholds would be more appropriate for urban-scale applications.

UBEM heavily relies on local or regional geospatial databases that, in the best-case scenario, provide information on building use, construction period, and volumes. Any inaccuracies in these datasets—such as misclassification of construction period, incorrect building use, or geometric errors—directly propagate into the UBEM, affecting the assignment of representative non-geometrical properties (building archetypes), the modelling of occupant-related loads, and the assessment of energy performance.

The ideal approach for calibrating an UBEM would include all thermally simulated buildings within a single unified scenario, computing the total energy demand of the building stock as the sum of the individual building's consumption, and progressively adjusting the parameters to meet the calibration criteria. However, this methodology presents two major challenges.

First, the calculation time required to simulate hundreds or thousands of buildings simultaneously, along with the multiple iterations typical of calibration procedure, is extremely high. Second, the dimensionality of the input parameter space increases exponentially with the number of buildings, significantly slowing convergence to an optimal solution.

Furthermore, performing automatic calibration at the urban block level can introduce imbalances, as it may lead to an unrealistic and disproportionately overestimation of energy demand in buildings with larger conditioned volumes. To mitigate this compensation effect, calibration at the city-block scale may be more appropriate than broader district-level approaches.

5. Conclusions

The newly approved Energy Performance of Buildings Directive (EU 2024/1275) sets out the national building renovation plan, which can be effectively supported through the integration of UBEM and building archetypes. However, the lack of accessible, comprehensive, and reliable city-wide building energy consumption data—combined with the high computational cost of repeatedly running large-scale energy models—has left most UBEM applications unvalidated and uncalibrated.

This study presents a novel methodology that employs the CMA-ES/HDE evolutionary algorithm to calibrate large-scale urban energy models at both the individual and urban block levels. Unlike

conventional practices in the UBEM field—where optimisation techniques are often avoided due to the intensive computational demands of iterative procedures—this approach demonstrates their practical feasibility, scalability, and transparency. Compared to surrogate- or inverse-model approaches, the workflow is implemented directly within the simulation engine, avoiding additional modelling layers and allowing explicit inspection of convergence behaviour and calibrated parameter distributions. The proposed framework should therefore be seen as complementary to Bayesian approaches, prioritising scalability and transparency over full uncertainty quantification in data-limited urban-scale applications. A key strength of the proposed method lies in its ability to capture occupant behaviour and activity patterns, as well as its adaptability to energy consumption data with varying spatial resolutions, offering a valuable guidance for future large-scale calibration applications. Furthermore, this work addresses a common gap in existing models by enhancing the representation of occupant-driven energy use within building archetypes.

The automatic calibration methodology based on the CMA-ES/HDE was applied to a case study located in Monthey, Switzerland, involving approximately 200 residential buildings. To address uncertainties in large-scale input data, an archetype-based UBEM approach was adopted, assigning building archetypes by construction period and using simplified shoebox geometries, corresponding to LoD1. CitySim was used to simulate energy exchanges in the urban environment. The objective function minimised the coefficient of variation of the root-mean-square error, *CV(RMSE)* by comparing simulated daily space heating and domestic hot water demands with measured district heating data from 2020. Calibration was performed at both single-building and multi-building scales, depending on the spatial resolution of available consumption data.

Four key input variables were optimised: air change rate, occupied area per person, daily volume of DHW per person, and heating set-point temperature. Overall, the calibration at the individual building scale was successful in 43 % of cases. Building portfolio calibrations were affected by a compensation effect, where energy demand was disproportionately assigned to specific buildings to meet the *CV(RMSE)* criterion. Although the methodology lacks the nuance typically associated with engineering judgment, it remains the only practical option for calibrating large-scale urban energy models. Furthermore, the results indicate that calibration is more effective when performed at the individual building level, rather than by aggregating multiple buildings within a single large-scale model.

The main limitations to the applicability of the proposed methodology are as follows.

- The availability of large-scale building energy consumption data with sufficient spatial and temporal resolution remains limited.
- The calibration process is computationally intensive. On average, approximately two hours are required to run 2000 individuals (i.e., single-building calibration). All simulations were carried out using CitySim on a workstation equipped with an Intel Core i9–14,900 K processor (24 cores, 32 threads, 3.2 GHz) and 32 GB of RAM. Trained surrogate models offer a more efficient solution to address this challenge.
- The absence of engineering judgment in automated procedures may lead to a bias towards convergence by adjusting the most sensitive parameter(s), potentially reducing the variability of others. Nevertheless, manual calibration at the city-block scale is unfeasible, making automated approaches a necessary trade-off.

Future research lines to enhance the proposed calibration procedure may focus on:

- Developing and integrating surrogate models to reduce computational demands while preserving accuracy and scalability, in line with the current approach.
- Extending the validation of the proposed calibration procedure by testing the model on additional datasets and case studies, as soon as new high-resolution energy consumption data become available, in order to assess its robustness. This will allow assessing the transferability of the calibrated parameters and reducing potential overfitting effects associated with the use of a single dataset.
- Improving the quality of input data, particularly by incorporating high-resolution energy consumption data to support broader applicability across different municipalities.
- Collaborating with data managers at local, regional, and national levels to ensure consistent, accurate, and up-to-date datasets.

CRediT authorship contribution statement

Matteo Piro: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Jérôme H. Kämpf:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Ilaria Ballarini:** Writing – review & editing, Supervision. **Vincenzo Corrado:** Writing – review & editing, Supervision.

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.

Acknowledgements

The authors would like to thank Satom SA for providing the monitoring data of the district heating network for the municipality of Monthey.

Data availability

The data that has been used is confidential.

References

ASHRAE. (2023). Measurement of energy, demand, and water savings guideline 14. www.ashrae.org/technology.

Ballarini, I., Corgnati, S. P., & Corrado, V. (2014). Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project. *Energy Policy, 68*, 273–284. https://doi.org/10.1016/j.enpol.2014.01.027

Biljecki, F., Ledoux, H., & Stoter, J. (2016). An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems,, 59*, 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005

Borges, P., Travesset-Baro, O., & Pages-Ramon, A. (2022). Hybrid approach to representative building archetypes development for urban models – A case study in Andorra. *Building and Environment*, 215. https://doi.org/10.1016/j.buildenv.2022.108958

BUILD_ME. (2025). BUILD_ME web-site. Retrieved from https://www.buildings-mena.com/typologies. Accessed April 30, 2025.

Carnieletto, L., Ferrando, M., Teso, L., Sun, K., Zhang, W., Causone, F., Romagnoni, P., Zarrella, A., & Hong, T. (2021). Italian prototype building models for urban scale building performance simulation. *Building and Environment, 192*. https://doi.org/10.1016/j.buildenv.2021.107590

Cerezo, C., Sokol, J., AlKhaled, S., Reinhart, C., Al-Mumin, A., & Hajiah, A. (2017). Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City. *Energy and Buildings, 154*, 321–334. https://doi.org/10.1016/j.enbuild.2017.08.029

Chen, Y., Deng, Z., & Hong, T. (2020). Automatic and rapid calibration of urban building energy models by learning from energy performance database. *Applied Energy*, 277. https://doi.org/10.1016/j.apenergy.2020.115584

Choi, S., & Yoon, S. (2024). GPT-based data-driven urban building energy modeling (GPT-UBEM): Concept, methodology, and case studies. *Energy and Buildings*, 325. https://doi.org/10.1016/j.enbuild.2024.115042

Choi, S., Yi, D. H., Kim, D. W., & Yoon, S. (2025). Multi-source data fusion-driven urban building energy modeling. *Sustainable Cities and Society, 123*. https://doi.org/10.1016/j.scs.2025.106283

Chong, A., & Menberg, K. (2018). Guidelines for the bayesian calibration of building energy models. *Energy and Buildings, 174*, 527–547. https://doi.org/10.1016/j.enbuild.2018.06.028

CitySim Solver. (2025). CitySim solver web-site. Retrieved from https://github.com/kaemco/CitySim-Solver. Accessed April 30, 2025.

Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews, 37*, 123–141. https://doi.org/10.1016/j.rser.2014.05.007

Davila, Cerezo, C, Reinhart, F, C., & Bemis, J. L (2016). Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets. *Energy, 117*, 237–250. https://doi.org/10.1016/j.energy.2016.10.057

Dilsiz, A. D., Nweye, K. E., Wu, A. J., Kämpf, J. H., Biljecki, F., & Nagy, Z. (2023). How spatio-temporal resolution impacts urban energy calibration. *Energy and Buildings*, 292. https://doi.org/10.1016/j.enbuild.2023.113175

EA CMA-ES/HDE. (2025). *EA CMA-ES/HDE github web-site*. Retrieved from https://github.com/jkaempf/EA Accessed April 30, 2025.

Emmanuel, W., & Kämpf, J. (2015). A verification of CitySim results using the BESTEST and monitored consumption values. In *Proceedings of BSA Conference 2015: Second Conference of IBPSA-Italy* (pp. 215–222).

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). (2007). Thermal performance of buildings - Transmission and ventilation heat transfer coefficient - Calculation method (ISO 13789:2007).

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).

Fabrizio, E., & Monetti, V. (2015). Methodologies and advancements in the calibration of building energy models. *Energies, 8*(4), 2548–2574. https://doi.org/10.3390/en8042548

Faure, X., Lebrun, R., & Pasichnyi, O. (2024). Impact of time resolution on estimation of energy savings using a copula-based calibration in UBEM. *Energy and Buildings*, 311. https://doi.org/10.1016/j.enbuild.2024.114134

Ferrando, M., Causone, F., Banfi, A., Corrado, V., Ballarini, I., Piro, M., Zarrella, A., Carnieletto, L., Borgato, N., Evola, G., Detommaso, M., Nicoletti, F., Vallati, A., & Fiorini, C. V. (2025). Developing an Italian library of reference buildings for urban building energy modeling (UBEM): Lessons learnt from the URBEM project. *Energies, 18*(22), 6026. https://doi.org/10.3390/en18226026

Gholami, M., Torreggiani, D., Tassinari, P., & Barbaresi, A. (2021). Narrowing uncertainties in forecasting urban building energy demand through an optimal archetyping method. *Renewable and Sustainable Energy Reviews*, 148. https://doi.org/10.1016/j.rser.2021.111312

Goy, S., Maréchal, F., & Finn, D. (2020). Data for urban scale building energy modelling: Assessing impacts and overcoming availability challenges. *Energies, 13*(6). https://doi.org/10.3390/en13164244

Gu, Y., Tian, W., Song, C., & Chong, A. (2023). Quantifying the effects of different data streams on the calibration of building energy simulation. *Energy and Buildings*, 296. https://doi.org/10.1016/j.enbuild.2023.113352

Haneef, F., Pernigotto, G., Gasparella, A., & Kämpf, J. H. (2021). Application of urban scale energy modelling and multi-objective optimization techniques for building energy renovation at district scale. *Sustainability, 13*(20). https://doi.org/10.3390/su132011554

Happle, G., Fonseca, J. A., & Schlueter, A. (2018). A review on occupant behavior in urban building energy models. *Energy and Buildings, 174*, 276–292. https://doi.org/10.1016/j.enbuild.2018.06.030

Heo, Y., Choudhary, R., & Augenbroe, G. A. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings, 47*, 550–560. https://doi.org/10.1016/j.enbuild.2011.12.029

Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, 168. https://doi.org/10.1016/j.buildenv.2019.106508

Hwang, J., Lim, H., & Lim, J. (2024). Reducing uncertainty of building shape information in urban building energy modeling using bayesian calibration. *Sustainable Cities and Society, 116*. https://doi.org/10.1016/j.scs.2024.105895

Johari, F., Munkhammar, J., Shadram, F., & Widén, J. (2022). Evaluation of simplified building energy models for urban-scale energy analysis of buildings. *Building and Environment*, 211. https://doi.org/10.1016/j.buildenv.2021.108684

Johari, F., Shadram, F., & Widén, J. (2023). Urban building energy modeling from geo-referenced energy performance certificate data: Development, calibration, and validation. *Sustainable Cities and Society*, 96. https://doi.org/10.1016/j.scs.2023.104664

Judkoff, R., & Neymark, J. (1995). *International energy agency building energy simulation test (BESTEST) and diagnostic method*. www.nrel.gov/docs/legosti/old/6231.pdf.

KAEMCO. (2025). *KAEMCO company web-site*. Retrieved from https://www.kaemco.ch/download.php Accessed April 30, 2025.

Kämpf, J. H., & Robinson, D. (2007). A simplified thermal model to support analysis of urban resource flows. *Energy and Buildings, 39*(4), 445–453. https://doi.org/10.1016/j.enbuild.2006.09.002

Kämpf, J. H., & Robinson, D. (2009). A hybrid CMA-ES and HDE optimisation algorithm with application to solar energy potential. *Applied Soft Computing Journal, 9*(2), 738–745. https://doi.org/10.1016/j.asoc.2008.09.009

Kämpf, J. H., Wetter, M., & Robinson, D. (2010). A comparison of global optimization algorithms with standard benchmark functions and real-world applications using EnergyPlus. *Journal of Building Performance Simulation, 3*(2), 103–120. https://doi.org/10.1080/19401490903494597

Kim, H. G., & Kim, S. S. (2024). UBEM calibration method by energy consumption disaggregation using a change-point model. *Journal of Building Performance Simulation*, 1–28. https://doi.org/10.1080/19401493.2024.2412135

Kolbe, T. H., Kutzner, T., Smyth, C. S., Nagel, C., Roensdorf, C., & Heazel, C. (2021). *OGC document: 20-010 OGC city geography markup language (CityGML) part 1: Conceptual model standard*. http://portal.opengeospatial.org/public_ogc/change_request.php.

Kolbe, T. H. (2009). Representing and exchanging 3D City models with CityGML. In S. Lee Jiyeong, & Zlatanova (Eds.), *3D Geo-Information sciences* (pp. 15–31). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87395-2_2.

Langevin, J., Reyna, J. L., Ebrahimigharehbaghi, S., Sandberg, N., Fennell, P., Nägeli, C., Laverge, J., Delghust, M., Mata, Van Hove, M., Webster, J., Federico, F., Jakob, M., & Camarasa, C. (2020). Developing a common approach for classifying building stock energy models. *Renewable and Sustainable Energy Reviews*, 133. https://doi.org/10.1016/j.rser.2020.110276

Li, Y., & Feng, H. (2025). Integrating urban building energy modeling (UBEM) and urban-building environmental impact assessment (UB-EIA) for sustainable urban development: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 213. https://doi.org/10.1016/j.rser.2025.115471

Lundström, L., & Akander, J. (2019). Bayesian calibration with augmented stochastic state-space models of district-heated multifamily buildings. *Energies, 13*(1). https://doi.org/10.3390/en13010076

Manso-Burgos, Ribó-Pérez, D, As, Van, J., Montagud-Montalvá, C., & Royo-Pastor, R. (2023). Diagnosis of the building stock using energy performance certificates for urban energy planning in mediterranean compact cities. Case of study: The city of València in Spain. *Energy Conversion and Management, 20*. https://doi.org/10.1016/j.ecmx.2023.100450

Mattsson, M., Danielski, I., Olofsson, T., & Nair, G. (2025). Archetypes-based calibration for urban building energy modelling. *Energy and Buildings*, 343. https://doi.org/10.1016/j.enbuild.2025.115843

Nageler, P., Zahrer, G., Heimrath, R., Mach, T., Mauthner, F., Leusbrock, I., Schranzhofer, H., & Hochenauer, C. (2017). Novel validated method for GIS based automated dynamic urban building energy simulations. *Energy, 139*, 142–154. https://doi.org/10.1016/j.energy.2017.07.151

Nagpal, S., Mueller, C., Aijazi, A., & Reinhart, C. F. (2019). A methodology for auto-calibrating urban building energy models using surrogate modeling techniques. *Journal of Building Performance Simulation, 12*(1), 1–16. https://doi.org/10.1080/19401493.2018.1457722

Nouvel, R., Zirak, M., Coors, V., & Eicker, U. (2017). The influence of data quality on urban heating demand modeling using 3D city models. *Computers, Environment and Urban Systems,, 64*, 68–80. https://doi.org/10.1016/j.compenvurbsys.2016.12.005

Oraiopoulos, A., & Howard, B. (2022). On the accuracy of urban building energy modelling. Renewable and sustainable energy reviews, 158. https://doi.org/10.1016/j.rser.2021.111976.

Pasichnyi, O., Wallin, J., & Kordas, O. (2019). Data-driven building archetypes for urban building energy modelling. *Energy, 181*, 360–377. https://doi.org/10.1016/j.energy.2019.04.197

Perez, D. (2014). *A framework to model and simulate the disaggregated energy flows supplying buildings in urban areas* [Doctoral dissertation, Ecole Polytechnique Fédérale de Lausanne (EPFL)] https://infoscience.epfl.ch/handle/20.500.14299/101151.

Perwez, U., Shono, K., Yamaguchi, Y., & Shimoda, Y. (2023). Multi-scale UBEM-BIPV coupled approach for the assessment of carbon neutrality of commercial building stock. *Energy and Buildings*, 291. https://doi.org/10.1016/j.enbuild.2023.113086

Piro, M., Ballarini, I., & Corrado, V. (2023). From building Energy models (BEM) to Urban building Energy models (UBEM): Input data and modelling approaches. Building Simulation Conference proceedings, 18, 3475–3482. https://doi.org/10.26868/25222708.2023.1662.

Piro, M., Bianco Mauthe Degerfeld, F., Ballarini, I., & Corrado, V. (2024). The challenges for a holistic, flexible and through-life updated energy performance certificate. *Sustainable Energy Technologies and Assessments*, 69. https://doi.org/10.1016/j.seta.2024.103922

Prataviera, E., Vivian, J., Lombardo, G., & Zarrella, A. (2022). Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis. *Applied Energy*, 311. https://doi.org/10.1016/j.apenergy.2022.118691

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling - A review of a nascent field. *Building and Environment, 97*, 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001

Robinson, D., & Stone, A. (2005). A simplified radiosity algorithm for general urban radiation exchange. *Building Services Engineering Research and Technology, 26*(4), 271–284. https://doi.org/10.1191/0143624405bt133oa

Robinson, D., Haldi, F., Leroux, P., Perez, D., Rasheed, A., & Wilke, U. (2009). CitySim: Comprehensive micro-simulation of resource flows for sustainable urban planning. *IBPSA 2009 - International Building Performance Simulation Association, 2009*, 1083–1090. https://doi.org/10.26868/25222708.2009.1083-1090

Salvalai, G., Zhu, Y., & Sesana, M. M. (2024). From building energy modeling to urban building energy modeling: A review of recent research trend and simulation tools. *Energy and Buildings*, 319. https://doi.org/10.1016/j.enbuild.2024.114500

Samareh Abolhassani, S., Zandifar, A., Ghourchian, N., Amayri, M., Bouguila, N., & Eicker, U. (2024). Occupant counting model development for urban building energy modeling using commercial off-the-shelf wi-fi sensing technology. *Building and Environment*, 258. https://doi.org/10.1016/j.buildenv.2024.111548

Sokol, J., Cerezo Davila, C., & 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. https://doi.org/10.1016/j.enbuild.2016.10.050

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews, 13*(8), 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033

Swiss Society of Engineers and Architects (SIA). (2006). *Conditions d'utilisation standard pour l'energie et les installations du bâtiment (SIA 2024)*. In French.

swissBUILDINGS3D. (2025). *swissBUILDINGS3D web-site*. Retrieved from https://www.swisstopo.admin.ch/en/landscape-model-swissbuildings3d-2-0 Accessed April 30, 2025.

Tognoli, M., Peronato, G., & Kämpf, J. H. (2023). Prediction of building hourly heating demand from CityGML files: Training workflow and deployment as an API. *Building Simulation Conference Proceedings, 18*, 2932–2939. https://doi.org/10.26868/25222708.2023.1570

URBEM. (2025). *URBEM web-site*. Retrieved from https://www.urbem.polimi.it/ Accessed April 30, 2025.

Wang, C. K., Tindemans, S., Miller, C., Agugiaro, G., & 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. https://doi.org/10.1080/19401493.2020.1729862

Wang, C., Ferrando, M., Causone, F., Jin, X., Zhou, X., & Shi, X. (2022). Data acquisition for urban building energy modeling: A review. *Building and Environment*, 217. https://doi.org/10.1016/j.buildenv.2022.109056

Wang, C., Wang, X., Causone, F., Yang, Y., Li, P., & Shi, X. (2026). Balancing effects of uncertain parameters on the accuracy and stability of urban building energy modeling. *Building and Environment*, 288. https://doi.org/10.1016/j.buildenv.2025.113886

Xia, C., & Hu, Y. (2024). Profiling residential energy vulnerability: Bayesian-based spatial mapping of occupancy and building characteristics. *Sustainable Cities and Society, 114*. https://doi.org/10.1016/j.scs.2024.105667

Yang, Y., Tan, Z., & Schläpfer, M. (2025). Assessing the space-use efficiency of French cities by coupling city volumes with mobile data traffic. *Sustainable Cities and Society, 124*. https://doi.org/10.1016/j.scs.2025.106292

Yi, D. H., & Park, C. S. (2021). Model selection for parameter identifiability problem in bayesian inference of building energy model. *Energy and Buildings*, 245. https://doi.org/10.1016/j.enbuild.2021.111059