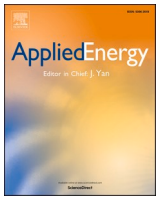
Applied Energy 404 (2026) 127150

Contents lists available at ScienceDirect

Applied Energy

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



Enhancing interpretability and automation in data-driven energy modelling: An analytical approach to change-point regression models

Massimiliano Manfren [a,*], Rundong Liao [b], Benedetto Nastasi [c]

[a] Architecture, Built Environment and Construction Engineering A.B.C., Politecnico di Milano, Via Bonardi 9, 20133 Milano, Italy
[b] College of Electrical Engineering, Sichuan University, Chengdu 610065, China
[c] Department of Industrial Engineering, Tor Vergata University of Rome, Via del Politecnico 1, 00133 Rome, Italy

HIGHLIGHTS

- Automated analytical formulation for 3PH, 3PC and 5P change-point regression.
- Validation against standard 3pH, 3PC and 5P from ASHRAE Guideline 14:2023.
- Open-source framework promoting further development.
- Balancing model sophistication with interpretability for trustworthy ML.
- Analytical formulation for monthly/daily data can be extended to hourly data.

ARTICLE INFO

Keywords:
Measurement and verification
Data-driven energy modelling
Interpretable machine-learning
Regression-based approaches
Energy analytics

ABSTRACT

Reliable, automated Measurement & Verification (M&V) and portfolio-scale energy analytics need models that are both accurate and interpretable. Current practice often relies on change-point regressions whose balance points are found via grid search or optimisation. As an alternative, an analytical formulation for simplified and automated identification of three-parameter heating (3pH), three-parameter cooling (3PC), and five-parameter (5P) models as defined in ASHRAE Guideline 14:2023 is proposed in this paper, with the goal of preserving interpretability also in more sophisticated workflows at the state of the art, which can use this novel formulation at different temporal scales (monthly, daily, and hourly). Standardised test datasets (39 in total) for 3pH, 3PC, and 5P models' testing and the Inverse Modelling Toolkit (IMT) have been used, showing comparable results in the large majority of cases and minor discrepancies in the others. The total batch runtime has been markedly reduced compared to the original implementation. Moreover, datasets from prior studies have been employed to evaluate applicability in real-world scenarios, demonstrating analogous results in this instance as well. While the current formulation is tested with monthly and daily interval data, its incorporation in hourly and sub-hourly resolution modelling workflows can promote further research developments in the area of interpretable data-driven analytics towards the "digital twins" paradigm, where interpretability of machine learning techniques and physical interpretation of underlying parameters is relevant to deliver effective and trusted solutions. Open-source code and datasets are made available to encourage further research on robust, transparent, and scalable data-driven energy modelling methodologies based on M&V principles. In this regard, additional efforts may be pursued to expand the concepts presented for the analytical formulation's development to encompass various automated processes with different objective functions (e.g., lasso, elastic net regression, etc.), model formulations, and constraints (e.g., physics-based interpretation of slopes and change points).

1. Introduction

Measurement and Verification (M&V) techniques [1] provide a solid and empirically grounded basis for improving the performance of facilities and quantifying savings in energy efficiency projects. Yet their potential extends considerably further, encompassing enhanced data-driven operational decision-making, cost and carbon emission savings

* Corresponding author.
E-mail address: massimiliano.manfren@polimi.it (M. Manfren).

https://doi.org/10.1016/j.apenergy.2025.127150
Received 7 May 2025; Received in revised form 13 November 2025; Accepted 23 November 2025
Available online 26 November 2025

Nomenclature

Symbol	Descriptions, Units
b_0	Base load of the 3P model, kW
b_1	Slope of the 3P model, –
b_2	Temperature change point of the 3P model, °C
$b_0^{(c)}$	Base load of the cooling system, kW
$b_0^{(h)}$	Base load of the heating system, kW
$b_0^{(final)}$	Base load of the 5P model, kW
$b_1^{(c)}$	Slope of the cooling model, –
$b_1^{(h)}$	Slope of the heating model, –
$b_2^{(c)}$	Temperature change point of the cooling system, °C
$b_2^{(h)}$	Temperature change point of the heating system, °C
y_i	Observed value, –
$\widehat{y}_i$	Model predicted value, –
$\overline{y}_i$	Mean of observed value, –
df	Error degree of freedom, –
$\overline{\widehat{y}_i - y_i}$	Average of the residuals, –
ε_i	Residual of observed value, –

Abbreviation

$AdjR^2$	Adjusted Coefficient of Determination, –
CV(RMSE)	Coefficient of Variation of the Root Mean Squared Error, %
DW	Durbin-Watson, –
NMBE	Normalised Mean Bias Error, %
R^2	Coefficient of Determination, –
RMSE	Root Mean Squared Error, –
SSE	Sum of Squared Errors, –
3P	Three parameter model, –
5P	Five parameter model, –

quantification, among other scopes [2]. In recent years, the use of smart metering technology enabled the creation of M&V models using high-resolution (hourly/sub-hourly) energy consumption data, capturing transient behaviours overlooked by consolidated M&V approaches, leading to a distinction between “M&V1.0” (traditional) and “advanced M&V" [3] or “M&V2.0” [4]. Concurrently, higher-resolution models can promote further development and incorporation of M&V analytical workflows into new paradigms, such as “Digital Twins” (DT) [5,6], offering additional layers of insight, highlighting operational nuances, and enabling predictive management.

Despite these advances, the extensive use of machine learning (ML) models that, while accurate, lack transparency, remains a fundamental barrier to the development of innovative and fast evolving paradigms such as DT. In fact, while DTs are meant to be the digital counterpart of physical assets, in many cases they make extensive use of data-driven methods that are not interpretable or physics-informed, and these limitations may hinder further developments in this broad research area, with the lack of modelling solutions that can be of general applicability, rather than tailored for single specific cases. Interpretable [7] and physics-informed ML [8], however, can provide transparent solutions in this sense, fostering stakeholders’ trust and facilitating regulatory compliance and practical knowledge transfer, especially when combined with simpler but more consolidated analysis methods such as change-point regression.

In the context of building energy analysis, the change-point identification problem involves determining the optimal point(s) in a continuous variable (typically outdoor air temperature) where the relationship between predictor and response variables undergoes a distinct change in slope. The regression models provided by ASHRAE Guideline 14:2023 [9], leveraging the research developed within ASHRAE Research Project 1050 [10], serve as examples in this context. ASHRAE RP-1050 resulted in the creation of the Inverse Modelling Toolkit (IMT) [11,12], which includes variable-base degree-day models, change-point models and multivariate regression models. As a result, ASHRAE Guideline 14:2023 encompasses seven distinct regression model types for building energy analysis: one-parameter (1P), two-parameter cooling model (2P), three-parameter heating (3pH), three-parameter cooling (3PC), four-parameter heating (4pH), four-parameter cooling (4PC), and five-parameter (5P) models. Each model is meant to represent a typical pattern of energy consumption in buildings.

In this framework, several critical gaps persist:

- Traditional change-point detection methodologies (piecewise linear, change-point modelling) depend on grid search or optimisation. Simplified, analytical solutions compatible with ASHRAE Guideline 14:2023 3-parameter heating (3PH), 3-parameter cooling (3PC) and 5-parameter (5P) models can help streamline analysis without sacrificing accuracy and cover a wide range of applications.
- Automation in current modelling tools and workflows is insufficiently flexible and frequently reduces interpretability in favour of efficiency and accuracy. There is an unmet need for automated solutions applicable across multiple possible workflows while preserving model transparency and leveraging the interpretability of ML methods.
- Extending interpretability beyond simple regression methods to other techniques such as generalised linear models (GLM) and generalised additive models (GAM), as well as robust statistical approaches (e.g., lasso, elastic net, and quantile regression), remains underexplored. Algorithms that produce accurate results but are “black-box” may limit adoption, highlighting the need for methods that are both data-driven and interpretable and physically grounded. Robustness is also crucial when dealing with noisy data and errors in measurements. Techniques like elastic net regression are desirable because they reduce sensitivity to outliers and anomalies in data while maintaining interpretability.

To bridge these gaps, this research proposes an innovative analytical formulation for change-point detection that integrates and automates different modelling steps. The formulation can be used on its own for simple monthly and daily interval regression modelling, as an alternative to other modelling tools, or be incorporated into more sophisticated modelling workflows while maintaining interpretability, fostering trustworthy and scalable applications. Therefore, the main contributions of this paper address the gap identified in the following way:

- Analytical formulation for automated identification of 3-parameter heating (3PH), 3-parameter cooling (3PC) and 5-parameter (5P) models according to ASHRAE-14:2023 classification.
- Validation of the algorithm through standardised ASHRAE-14:2023 test cases for software and empirical building data.
- Open-source code available to encourage further research and practical adoption in different workflows, beyond monthly and daily modelling.

Overall, this research illustrates the development of an analytical formulation capable of performing standard M&V activities with accuracy comparable to tools requiring either grid search, such as IMT in ASHRAE 14:2023, or optimisation [13,14], such as OpenEEmeter 4.0 [15]. This work presents a distinctive approach: rather than substituting

grid search with optimization-based approaches, an analytical formulation is proposed. The formulation preserves ease of use and interpretability, making it easy to audit and integrate in more articulated data-driven workflows. The modelling methodology, with minimal changes, can serve a broad set of objectives. Further, while maintaining the same conceptual framework, the model formulation can be further developed and employed to support more sophisticated workflows to address data up to hourly and sub-hourly resolution, robust model identification and approximated physical interpretation. These aspects, constituting the background and motivation for this research, are discussed in detail in the next section in relation to recent scientific literature.

2. Background and literature review

This section will explain the background of this research and review scientific literature, focusing on measurement and verification methods and their applications (Section 2.1), recent research on interpretable data-driven methods that can leverage hourly and sub-hourly interval data (Section 2.2), and future possibilities related to "digital twins" and physics-informed machine learning techniques (Section 2.3). Finally, the integration of these elements is presented in the summary of findings (Section 2.4).

2.1. Measurement and verification techniques and applications

Measurement and Verification (M&V) is widely acknowledged as a fundamental element of energy management, enabling the validation of savings [16,17] and the continuous evaluation of operational efficiency. Kim and Haberl conducted field tests on the energy performance measurement protocols provided by ASHRAE/CIBSE/USGBC, using three levels of cost/accuracy for the procedures: basic (indicative) [18], intermediate (diagnostic) [19] and advanced (investigative) [20]. The authors highlighted that conducting analysis at an hourly or sub-hourly interval (advanced level) allows for better detection of anomalies in the energy behaviour of buildings and provides a much deeper understanding of their operation, but it requires greater effort in modelling, analysis, and interpretation; conversely, at the intermediate level, using monthly whole-building energy consumption and demand data, which are typically accessible for billing in most existing buildings, is sufficient for performance assessment and benchmarking, provided that the data is gathered over an adequately extended timeframe.

Change-point regression models serve as baseline models for the methodology in Measurement and Verification (M&V) standards such as ASHRAE Guideline 14:2023 [9] and EVO IPMVP [21], among others; typically, they are favoured for their simplicity and ease of interpretation, making them the preferred solution among practitioners, who, in many cases, trade accuracy for practicality [22]. ASHRAE Guideline 14:2023 prescribes seven distinct regression model types for building energy analysis (1P, 2P, 3pH, 3PC, 4pH, 4PC, 5P).

The 3pH, 3PC, and 5P models are extensively used owing to their balance between physical interpretability and simplicity for practical applications. The existence of heating and cooling balance points (i.e., change-point in regression models) creates a close connection with the degree-days methodologies discussed subsequently in this section. The heating balance point denotes the outdoor temperature below which heating of the building is required, whereas the cooling balance point indicates the outdoor temperature above which cooling is required. 4pH and 4PC models also have change-points, but they are used to model consumption patterns where the temperature dependence is not constant, e.g. when latent loads become significant at higher outdoor temperatures in buildings with HVAC.

While simpler regression-based approaches offer the advantage of quick insights at relatively low cost, it is worth noting that more advanced [16,17] data-intensive techniques may capture subtler factors, such as operational details, occupancy, and user behaviour. From a conceptual standpoint interpretability [7], in the context of data-driven methodologies, denotes the ability of humans to understand variations in output resulting from variations in input or algorithmic parameters. When additional techniques must be employed (i.e., "post hoc" techniques model such as SHAP, LIME, etc. [23]) to render interpretable a model that would be otherwise a "black-box", the model is "post hoc" interpretable or explainable [24,25].

For this reason, models that can be directly understood by humans are referred to as having "ante hoc" or "intrinsic" interpretability, a characteristic of methods such as linear multivariate regression and trees or other types of models that represent evolutions of these ones (e.g. GLM, GAM, etc.) [7].

Among the most advanced interpretable machine learning techniques, it is possible to identify symbolic regression [26] and Kolmogorov-Arnold Networks (KAN) [27]. Symbolic regression's primary strength lies in producing explicit mathematical formulas that describe the relationship between variables. The generated equations often correspond to known physical laws or reveal new relationships that scientists can leverage. These two techniques are closely related, as shown by Panczyk et al. [28]. After training KANs transform into symbolic equations, yielding perfectly interpretable models. Creating powerful yet ante hoc interpretable data-driven techniques [29] is an open challenge in the building energy management field [30], because post hoc approaches such as SHAP and LIME provide limited interpretability [23]. The potential to overcome this limitation while maintaining a certain degree of continuity with standardised change-point regression models is discussed later in Section 2.2. Further, the challenge of enhancing transparency in models is acknowledged in both the building energy modelling community [31] and the energy systems modelling community [32]; in general, the problem of interpretability affects not only machine learning (ML) models (which are in many cases "black boxes") [7], but also the entire workflow [33] and its main steps. Another possible solution is that of creating physics-informed machine learning (PIML) models (and workflows) which can meet the dual aims of higher accuracy (inherent to the possibility of employing more sophisticated techniques) and interpretability (due to causality and physical interpretation). This possibility is discussed later in Section 2.3.

Baseline regression models mentioned earlier are used to correlate historical energy use with outdoor air temperature or degree-days, providing a physically meaningful (yet simplistic) interpretation of energy consumption by splitting the weather-dependent component of the energy consumption (i.e., proportional to temperature or degree-days) and the non-weather-dependent component (i.e., the constant part in the piecewise regression). Heating degree-days (HDD) are computed as the cumulative sum of temperature differences below the heating balance-point temperature, whereas cooling degree-days (CDD) are computed as the cumulative sum of temperature differences above the cooling balance-point temperature [34,35]. The ability to differentiate between weather-dependent consumption (temperature-dependent), and weather-independent consumption like lights, plug loads or Domestic Hot Water (DHW) is essential for benchmarking [36,37] and disaggregation [38]. The temperature-dependent component of electricity consumption can become much more relevant with the electrification of heating, for example, by means of air-source heat pumps [39], because of both the heating demand and the variability of the Coefficient of Performance (COP) [40,41].

Additionally, employing outdoor air temperature as a model input allows for regression-based methodologies to evaluate potential changes in building performance under future conditions [42], i.e., accounting for the impact of climate change. At the same time, buildings' physical parameters – building fabric, occupancy, and internal gains – can shift the best-fit heating and cooling balance points quite significantly [43]. Considering the potential variety of architectural designs, climatic conditions, and occupant behaviours, methodologies for establishing base temperatures need to be tailored to enhance model accuracy [44,45], and the evolution of temperatures in time due to climate

change (leading to a decrease of HDD and an increase of CDD in many areas) needs to be assessed [43,46].

More in detail, the energy use in a quasi-steady state calculation considers the balance between losses [47] and gains [48], using standardised formulations [49]; similarly, the components of building energy balance can be reworked to be estimated with regression models [50–52], also using standard formulations [53], based on the energy signature method described in ISO 16346:2013 [54]. When scaled (e.g., by building volume [55]) [49]and grouped (e.g., clustering [56], or building archetypes [57]) such models can be extended to assess an entire building stock, aiding in the analysis of the performance of buildings with different features [49].

Overall, change point regression models, despite their simplicity, can be employed for many purposes connected to the calculation of energy, cost, and carbon emissions savings [58], leveraging either a basic (weather-dependent/weather-independent) or a more articulated (components of energy balance) physical interpretation, previously discussed. Therefore, automating the selection of the best-fitting heating and cooling balance points [59] (i.e., change-points in the regression model) using analytical formulations [60] can avoid the use of grid search or optimisation and serve a broad set of objectives [61] due to their spatial and temporal scalability, as illustrated in the next section.

2.2. Interpretable data-driven models using hourly and sub-hourly data

The ability to develop models that possess good fitting performance with hourly and sub-hourly interval data while preserving ante hoc interpretability (i.e., models that do not necessitate post-hoc techniques like LIME and SHAP [23]) and that maintain a certain degree of continuity with the standardised change-point regression models [9] outlined in Section 2.1 constitutes a significant challenge. This is motivated by the ease of use, extensive range of applications, and scalability already mentioned; yet it conflicts with machine learning models that may offer superior model fitting performance, albeit at the cost of being "black-boxes". A potential solution in this sense is that of developing physics-informed machine learning (PIML) models (and workflows), which will be discussed in Section 2.3 and represents a fast evolving research area.

In this section, the models and workflows discussed are the ones that are more directly related to change-point regression models in ASHRAE Guideline 14:2023. As noted in Section 2.1, while monthly regression-based approaches remain popular and are, in many cases, sufficient for benchmarking, hourly or sub-hourly data can unlock more sophisticated analyses, looking into building energy performance dynamics. Abushakra and Paulus presented in a series of three papers respectively the background, methodology, and results [62] of ASHRAE research project 1404-RP [63] on short-term M&V of whole building energy performance. Regression models were developed employing varying durations of short-term hourly data, spanning from 2 weeks to a full year. The counterintuitive finding was that prolonging the duration of the short-term monitoring period can degrade the predictive ability of the model, while carefully selected shorter periods may yield simpler yet more effective models. Jalori and Reddy have also addressed the need to manipulate correctly hourly building monitoring data in two related publications, one focussing on the detection of diurnal schedules by clustering and the other on regression-based inverse modelling [64], demonstrating the need of proper data segmentation for energy time series.

Recently, Staffel et al. [65] proposed a global model that integrates a five-parameter (5P) change-point structure with multipliers representing diurnal profiles to generate hourly profiles, starting from daily predictions. They introduced a novel input variable, termed building-adjusted internal temperature (BAIT), which is derived from outdoor air temperature, solar radiation, wind speed, and relative humidity, but maintained a similar conceptual framework, where interpretability is preserved, using a variable-base Degree-Days (VB-DD) 5P model on daily interval data. Innovations such split degree-day calculations could enhance these models further by considering intra-day temperature fluctuations [66] and their impact on diurnal profiles. Furthermore, these formulations reflect principles present in the Time of Week and Temperature (TOWT) modelling approach [67,68] modelling approach for load profiles at hourly or sub-hourly intervals. The TOWT algorithm separates energy consumption dependence on building operational strategy (Time Of Week) and weather conditions (outdoor air temperature), providing an intrinsically interpretable framework due to its formulation using multivariate regression. To accomplish this, it employs a workflow with different automated modelling steps, and 5P regression is applied at the beginning to detect occupied/unoccupied (high/low consumption) hours before running separate piecewise regression models, which are then recombined to obtain the complete profile, which considers day-to-day variations in operation.

Along a similar line, Kazmi et al. illustrate how the analysis of outdoor temperature dependence and weekday/weekend differences in operational energy patterns is crucial for the development of data-driven load profiles modelling from individual buildings to urban scale [69]. In relation to outdoor temperature dependence, the electrification of heating and climate change [70] are going to determine significant changes at the system level in energy infrastructure.

The TOWT methodology is particularly useful as it provides both accuracy and interpretability through its framework, hence enhancing simpler M&V methods when data at hourly or sub-hourly resolution is available [71,72]. Open-source software implementations in R and Python languages such as RMV2.0 [73] NMECR [74], and OpenEEmeter 4.0 [15] can be used to improve automation, adaptability, and reproducibility of workflows. Notably, OpenEEmeter 4.0, used also within CalTRACK methods [75], transitioned from exhaustive grid search (OpenEEmeter 3.0) to a global optimization approach [13,14] combined with lasso regression-inspired penalization. This approach reduced computational time significantly while incorporating adaptive loss (objective) functions that can accommodate various regression formulations, including elastic net and other regularisation techniques, to increase the robustness of estimates (e.g., by making it less sensitive to outliers). Automating the identification of 5P models during the execution of the TOWT workflow can serve as an initial step for segmenting the dataset by temperature and it is possible to use it in a forward selection process for model input variables [72]. Further, various authors have recently introduced advancements to the TOWT algorithm, such as Granderson et al. [76], who evaluated different strategies for short-term model retraining to determine flexibility; Mirfin et al. [77], who incorporated solar radiation as an input variable (TOWST); Lopez-Villamor et al. [78], who proposed an Auto-Regressive component (TOW-ARX) together with clustering; and Manfren and Nastasi [71], who modified workflow steps, temperature segmentation and hyper-parameter selection.

An alternative regression modelling technique is used in the software ECAM 7.0 [79], whereby the hourly model relies on day-typing and hour-typing, and many four-parameter models (4P, as defined by ASHRAE-14:2023) are fitted, utilising a tree structure for their integration in the complete model. Overall, it is possible to identify in literature ante hoc interpretable models employing hourly [62–64] and sub-hourly interval data that can maintain continuity with standardised change-point regression models, but due to the necessity of time series data segmentation (according to rules such as day type, hour type, weekday/weekend, occupied/unoccupied hours, etc.), workflow automation becomes more and more relevant for practical applications.

2.3. Digital twins and physics-informed machine learning development

The Digital Twin (DT) paradigm has attracted significant interest recently due to its transformative potential in multiple types of applications. However, the contours of the DT definition are not precisely defined [6], and there are various nuances. Nonetheless, it represents a

concept that can help transform research and applications in multiple phases of the building life cycle [80], from design stage performance simulation [5] to operation [81]. Its application extends beyond individual buildings; due to its scalability, it can be used in urban settings or energy communities for multi-carrier energy systems.

In principle, DTs represent virtual replicas of physical assets which allow for precise monitoring and analysis of energy performance [82], proactive responses to changing demands, and improving operational efficiency [83]. The implementation of digital twins in buildings has been enhanced by the use of sensors, IoT devices and sophisticated data-driven algorithms. These technologies facilitate the creation of reliable energy forecasts [82] that can be used to manage dynamic variations in energy demand and user behaviour. Beyond overall energy consumption prediction, DTs can be applied for specific applications in HVAC performance optimisation, and indoor environmental quality control, among others. A fundamental disadvantage of numerous data-driven methodologies is the lack of a physical interpretation of model parameters, which becomes particularly significant when applied to digital twins, which should mirror the behaviour of the physical asset in a transparent way. A regression layer can function as an interpretable component in a more complex data-driven workflow by constraining the heating and cooling balancing points, the associated slopes, and, where necessary, enforcing convexity.

Physics-informed machine learning (PIML) [8] is a potential answer in this sense because models can be developed using supplementary information acquired by imposing physical laws (acting as "constraints" embedded in the training process) and use training data in arbitrary locations within the space-time domains considered. PIML provides the possibility to integrate noisy data with (physics-informed) mathematical models, overcoming limitation of traditional inverse modelling using Partial Differential Equations (PDE) or Ordinary Differential Equations (ODE) while requiring less training data compared to conventional, non-physics-informed ML. This data efficiency is particularly valuable for practical applications in buildings where extensive training data may not be available.

Ma et al. [84] reviewed methods for PIML for building energy modelling, grouping them by physics-informed input, output, architecture, and ensemble models. In many cases, the physical information (and the underlying "constraints" in modelling) is derived from a detailed "white-box" physics-based model [85] or a "grey-box" (physical-statistical) model, usually in the form of Resistance-Capacitance (RC) model formulation [86]. The principle of embedding physical consistency aligns with the potential incorporation of physical constraints (such as plausible ranges for slopes or change-point temperatures based on building physics knowledge, as well as building energy balance components' estimates [50–52]) directly into the techniques for M&V discussed in Section 2.1. In this approach, an RC model can be developed as a surrogate physical model to generate additional data.

In general, as illustrated by Zanotto and Ferrari [87], building energy balance can be formulated as an Ordinary Differential Equation (ODE) involving heat transfer through the envelope, infiltration and ventilation, solar and internal gains, and heating and cooling systems. This principle can be employed by PIML for model training, but also for data imputation (based on building physics knowledge) [31] and for the identification of lumped physical parameters [88]. In relation to lumped parameter models, RC ones can be formulated to be compatible with current international technical standards for building energy performance, such as ISO 52016:2017 [89] (i.e., detailed "white-box" physics-based models), while leveraging strategies to enable efficient computation, such as matrices pre-inversion [90] and adjustment/scaling methods for lumped physical parameters [91]. In this way, PIML can make effective use of both physical information and a standardised modelling procedure, which can make workflows more transparent and reproducible, connecting it to the methods discussed in Section 2.1, for the reasons expressed earlier in this section. Further, interpretability can be retained by employing formulations such as Kolmogorov-Arnold Networks (KAN) [92] or symbolic regression based on physical principles and constraints.

In this framework, interpretable data-driven hourly/sub-hourly energy models, presented in Section 2.2, may be used in the training process of PIML, as they can provide additional information in the model creation workflow (due to the physical interpretation of the underlying model coefficients and rules for time series data segmentation) and forecasts of building energy operational profiles that can be used in a comparative way (e.g., to detect differences in specific conditions) or by using them together in ensembles of ML techniques [93], which can be trained using archetypes with well-defined physical properties [94], which can be clustered/adjusted/scaled to fit the specific case.

2.4. Summary of literature review findings

Measurement and Verification (M&V) is fundamental for the evaluation of operational efficiency and the robust quantification of energy [16,17], cost, and carbon emission savings [58] in projects. Overall, the literature review highlighted that:

1. Change-point regression approaches, as defined in ASHRAE Guideline 14:2023 [9], remain the interpretable core for M&V. The guideline includes seven specific types of regression models, each one designed to exemplify a standard pattern of energy use found in monitoring.
2. A more in-depth interpretation of regression models can leverage the calculation of the components of the building energy balance in analogy with semi-stationary methods [50,52], with the inclusion of additional input variables [53].
3. Analytical formulations of change-point regression models [60] represent an alternative to grid search and optimisation approaches that can be used on its own for basic monthly and daily interval regression modelling but also incorporated in more complex workflows with hourly and sub-hourly interval data [63,67].
4. Physics-Informed Machine Learning (PIML) [84] can leverage "grey-box" (physical-statistical) model formulations in the Resistance-Capacitance (RC) form [86] that can be compatible with current international technical standards, such as ISO 52016:2017 [90], and become interpretable by means of Kolmogorov-Arnold Networks (KAN) [92] or by employing symbolic regression models based on physical principles and constraints.

These findings can be useful in facilitating the development of the Digital Twin (DT) paradigm. DTs represent virtual counterparts of physical assets, the physical interpretation previously mentioned could be leveraged in the training process of advanced physics-informed machine learning (PIML) models, or work in ensembles of ML models [93]. In this context, analytical formulations of change-point regression models for M&V can be used for a large variety of purposes.

3. Methodology

As described in the introduction and in the literature review, the analytical model formulations developed in this study are meant to simplify the creation of ASHRAE 3pH, 3PC, and 5P models so that they become suitable for implementation in integrated automated workflows without sacrificing interpretability.

As discussed in Section 3.1, this workflow is based on simple piecewise linear regression and the research involved two fundamental elements: revising the algorithm for the explicit solution of 3pH and 3PC models proposed by Paulus [60] and extending its applicability to 5P model introducing additional criteria based on data shape [59] and convexity.

In Section 3.2 a suite of diagnostic metrics for model performance is reported, along with the acceptability criteria prescribed by Measurement and Verification (M&V) guidelines. In Section 4, the models are

then tested against both the cases from ASHRAE Guideline14:2023 and measured data from real buildings.

3.1. Model workflow and formulation

This study primarily focuses on the 3pH, 3PC, and 5P models fitting, for the reasons reported earlier. Fig. 1 illustrates the flowchart for the automatic determination of the temperature change-points for the 3pH and 3PC models. Firstly, the 3pH model utilises the following mathematical expression:

$$y = b_0 + b_1 \cdot \max(b_2 - x, 0) \quad (1)$$

Where, b_0 represents the system's base load, b_1 denotes the slope of the response in the lower temperature range, and b_2 corresponds to the temperature change point in the heating phase.

During the model solution process, the input temperature data *(x)* and the corresponding energy consumption data (*y*) are first sorted in ascending order according to temperature. Next, the candidate change points are traversed. Let the candidate index *m* range from 3 to *n* (where *n* represents the total number of data points), thereby assigning the first $m-1$ data points to the low-temperature segment and the remaining points to the high-temperature segment. For each candidate segmentation, the mean value of the *y* in the high-temperature segment is used to estimate the base load b_0, while the least squares method is applied to the low-temperature segment data to estimate the slope b_1. Specifically, let $\sum x_{low}$ be the sum of the temperatures in the low-temperature segment, $\sum y_{low}$ be the sum of the energy consumption values in that segment, $\sum x_{low}^2$ be the sum of the squares of the temperatures, and $\sum (x_{low} \cdot y_{low})$ be the sum of the products of the temperatures and the energy consumption values. Then, the candidate slope value b_1 is calculated as:

$$b_1 = \frac{(\sum x_{low})(\sum y_{low}) - (m-1) \cdot \sum (x_{low} \cdot y_{low})}{(m-1) \cdot (\sum x_{low}^2) - (\sum x_{low})^2} \quad (2)$$

If the computed result shows that b_1 is close to zero, the model is deemed to have degenerated into a constant model, and b_2 is then set to the median of the temperature data x. Conversely, if b_1 is non-zero, the temperature change point b_2 is further calculated using the low-temperature segment data according to the following formula:

$$b_2 = \frac{\sum y_{low} - (m-1) \cdot b_0 + b_1 \cdot \sum x_{low}}{(m-1) \cdot b_1} \quad (3)$$

Subsequently, based on the current candidate parameter combination, predicted values are constructed and the sum of squared errors (*SSE*) between these predictions and the actual *y* values is calculated. This process is iterated until the optimal parameter combination (b_0, b_1, b_2) that minimises the *SSE* is identified. Even though *SSE* is employed here, the procedure can be changed to employ alternative objective

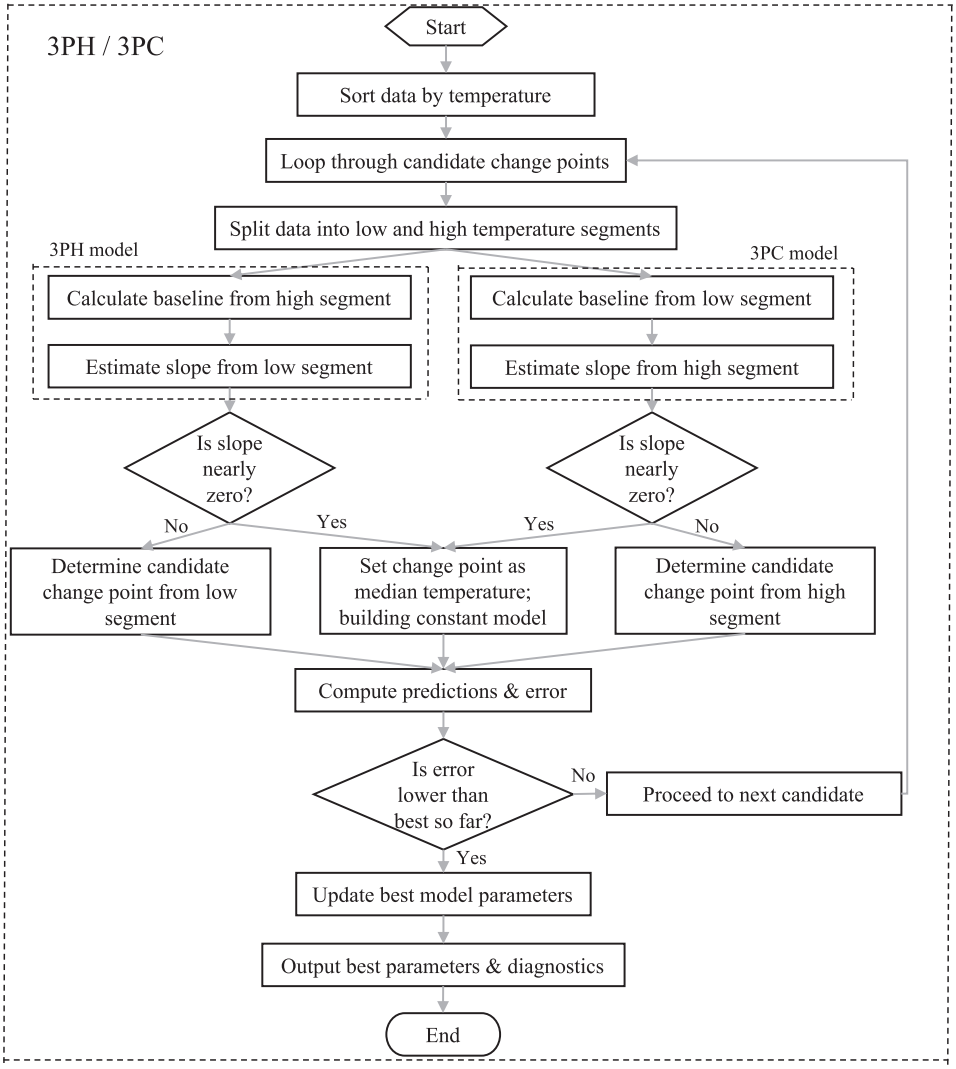


Fig. 1. Flow charts for automatic identification of heating and cooling temperature change points in 3pH and 3PC models respectively.

functions, such as penalties, as in the case of elastic net regression. Clearly, the previous formulas should be adjusted, depending on the objective function choice.

In the 3PC model, the model expression is adjusted as follows:

$$y = b_0 + b_1 \cdot \max(x - b_2, 0) \tag{4}$$

Where, b_0 is estimated as the mean value of the y in the low-temperature segment (i.e. the first $m-1$ data points), b_1 is obtained via the least squares method applied to the high-temperature segment data, and b_2 is defined as the temperature change point. In the high-temperature segment, this is done by setting the various metrics for the high-temperature data: let $\sum x_{high}$ be the sum of the temperatures in the high-temperature segment, $\sum y_{high}$ be the sum of the energy consumption values in that segment, $\sum x_{high}^2$ be the sum of the squares of the temperatures, $\sum \left(x_{high} \cdot y_{high}\right)$ be the sum of the products of the temperatures and the energy consumption values, and n_{high} represents the total number of data points in the high temperature segment. Then, the candidate slope value b_1 is calculated as:

$$b_1 = \frac{n_{high} \cdot \sum \left(x_{high} \cdot y_{high}\right) - \left(\sum x_{high}\right)\left(\sum y_{high}\right)}{n_{high} \cdot \sum x_{high}^2 - \left(\sum x_{high}\right)^2} \tag{5}$$

If b_1 is approximately zero, then the change point b_2 is directly the median of x; Otherwise, use the following formula to calculate b_2:

$$b_2 = \frac{\sum \left(b_0 + b_1 \cdot x_{high} - y_{high}\right)}{n_{high} \cdot b_1} \tag{6}$$

Similarly, by iterating through the candidate partitions and computing the corresponding *SSE*, the parameter combination that minimises the error is ultimately selected as the output of the 3P-cooling model.

Building on these two models, the 5P model further integrates the

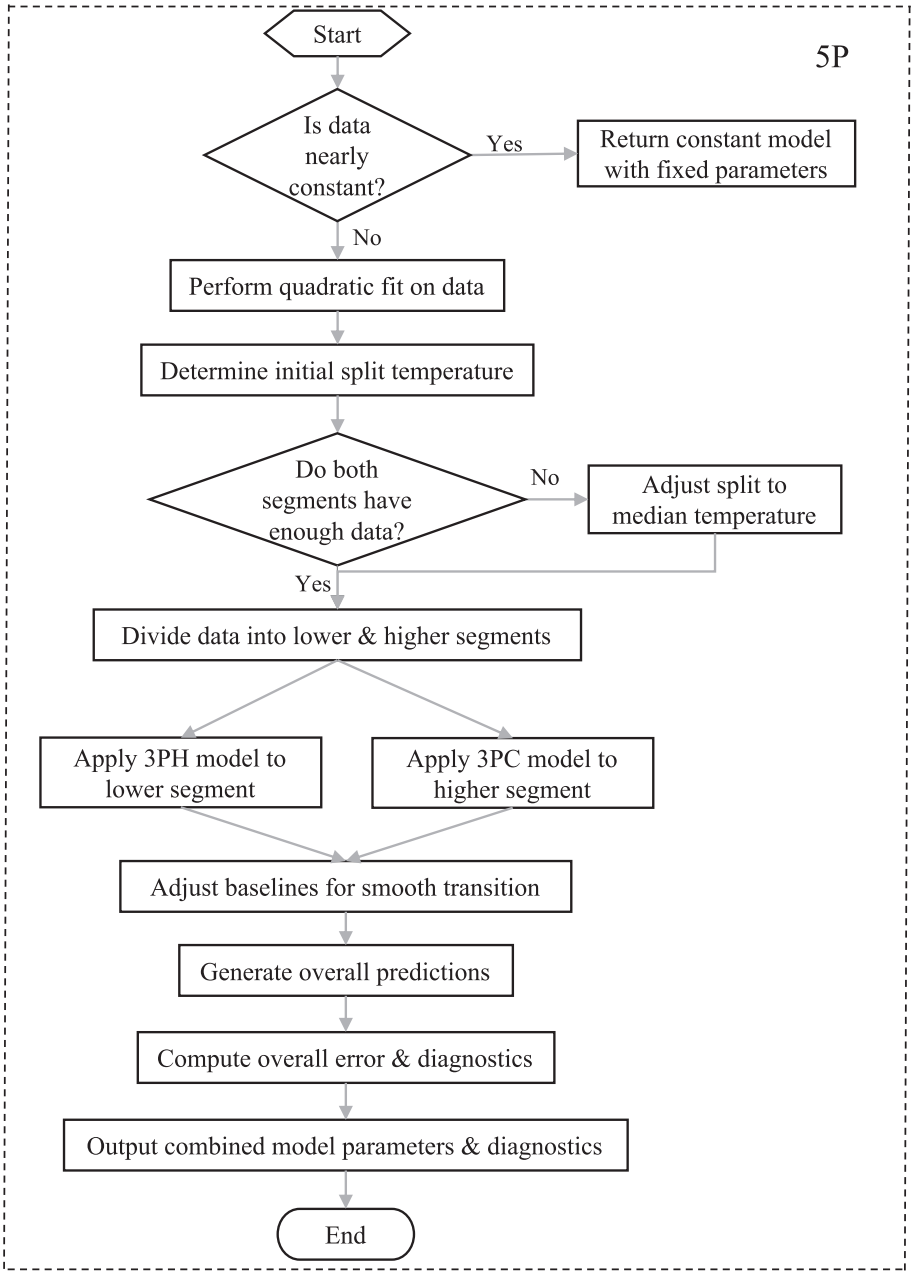


Fig. 2. Flow chart for automatic identification of heating and cooling temperature change points in 5P model.

heating and cooling phases to characterise the system's energy consumption response to temperature changes. Simplifications are made to automatically determine the optimal change point also in this case, avoiding reliance on grid search or optimisation, and instead analysing the data shape, as shown in Fig. 2, and using the approach illustrated earlier and in Fig. 1.

The specific steps are: Firstly, the entire dataset is evaluated to determine whether, in relation to temperature, the data is essentially constant. The temperature data is uniformly divided into four intervals, and the mean and median are computed for these intervals as well as for the entire dataset. Next, the maximum and minimum values among the means and medians of these intervals are identified. If the maximum or minimum of the means and medians deviates by no more than ±5 % relative to the values of the whole dataset, it can be concluded that the data is approximately constant, indicating that the effect of temperature change on energy consumption is negligible and that a constant model is suitable. If this condition is not met, a quadratic polynomial is employed for a global fit, with the expression:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{7}$$

Where β_0, β_1 and β_2 are the fitting parameters. Here, careful attention is given to the curvature of the second-order polynomial regression model to assess data plausibility. The second derivative is determined by the coefficient β_2; when $\beta_2 > 0$, the function is convex, and when $\beta_2 < 0$, it is concave. In practical simulations of heating and cooling processes, the expected physical behaviour typically shows convex curvature. Therefore, if the fitted regression function exhibits concavity, this may be regarded as a data anomaly. If normal, proceed to the next step, in accordance with the properties of quadratic functions, the extremum point is determined using the following formula:

$$x_{min} = -\frac{\beta_1}{2\beta_2} \tag{8}$$

This x_{min} serves as an initial partitioning point for the data, after which the sufficiency of the data in each segment is verified. If this point does not fall within the data range, it indicates that the data exhibit a monotonic heating or cooling pattern. In such cases, linear regression is applied: if the slope is negative, the 3pH model is used exclusively; if the slope is positive, the 3PC model is applied instead. If any segment is found to have an insufficient number of data points, the median of x is employed as an alternative. Subsequently, the data are divided into two parts at x_{min}, one to the left and one to the right, and the previously described 3pH and 3PC models are fitted independently to these segments.

This process yields the heating parameters ($b_0^{(h)}$, $b_1^{(h)}$, $b_2^{(h)}$) and the cooling parameters ($b_0^{(c)}$, $b_1^{(c)}$, $b_2^{(c)}$). To ensure a smooth transition of the overall model at the division point, a unified intercept is defined as the average of the base loads from both segments:

$$b_0^{(final)} = \frac{b_0^{(h)} + b_0^{(c)}}{2} \tag{9}$$

Finally, depending on the temperature interval in which x falls, the 5P model provides the following predictions:

When x is less than or equal to the heating phase's temperature change point $b_2^{(h)}$, the predicted value is calculated using the following formula:

$$y = b_0^{(\text{final})} + \left| b_1^{(h)} \right| \cdot \left(b_2^{(h)} - x \right) \tag{10}$$

When x is greater than or equal to the cooling phase's temperature change point $b_2^{(c)}$, the prediction expression is:

$$y = b_0^{(\text{final})} + b_1^{(c)} \cdot \left(x - b_2^{(c)} \right) \tag{11}$$

If x falls between the two phases, $b_0^{(final)}$ is directly used as the prediction.

In summary, this study determines each candidate temperature change point in turn with a rigorous step-by-step algorithm and achieves automatic optimal parameter selection by minimising the sum of squared errors (*SSE*) between the overall predicted value and the actual data. The search process for the optimal parameters can be further constrained to embed physical plausibility of the estimated quantities.

3.2. Model performance analysis

Following model execution, the overall predictive performance of the model and the characteristics of its residuals were comprehensively evaluated. A range of diagnostic indicators was employed, including the sum of squared errors (*SSE*), the coefficient of determination (R^2), adjusted coefficient of determination ($AdjR^2$), the root mean square error (*RMSE*), the coefficient of variation of *RMSE* (*CV(RMSE)*), the normalised mean bias error (*NMBE*), the *p*-value of the F-test, and the Durbin–Watson (*DW*) statistic. Table 1 provides a brief description of the statistical indicators used for model performance analysis and employed in the comparison of the results obtained by the Inverse Modelling Toolkit (IMT), which is used as a reference in the validation process based on ASHRAE RP-1050 [10] test cases, and the ones obtained with the novel model formulation proposed in this research. Among the statistical indicators considered, *CV(RMSE)* and *NMBE* are the ones used for M&V model acceptance according to ASHRAE Guideline 14:2023 [9]. Table 2 reports, as an example, thresholds for M&V model acceptance [9]using monthly and hourly interval data.

4. Results and discussion

4.1. Basic model shape tests

Ideally, the energy signature curve of a building should be U-shaped. This means that, within the balance point temperature range where no additional heating or cooling is required, the energy consumption remains nearly constant with respect to temperature. As the outdoor air temperature deviates from this balance temperature, the predicted air-conditioning energy consumption is expected to rise due to the

Table 1
Description of the statistical indicators for model diagnostics.

Indicator	Description	Formula	
SSE	Sum of Squared Errors	$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	(12)
R^2	Coefficient of Determination, proportion of variance explained	$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$	(13)
$AdjR^2$	Adjusted R^2, corrects R^2 for number of predictors in linear models	$AdjR^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1}$	(14)
RMSE	Root Mean Squared Error, average magnitude of residuals	$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{df}}$	(15)
CV (RMSE)	Coefficient of Variation of Root Mean Squared Error (RMSE), normalised by mean value	$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{df}}}{\bar{y}_i} \times 100\%$	(16)
NMBE	Normalised Mean Bias Error, measures systematic bias	$NMBE = \frac{\overline{(\hat{y}_i - y_i)}}{\bar{y}_i}$	(17)
p-value	F-statistic, hypothesis testing of overall significance, by	$F = \frac{\frac{R^2}{k}}{\frac{(1-R^2)}{df}}$	(18)
DW	Durbin-Watson statistic, tests for residual autocorrelation	$DW = \frac{\sum_{i=2}^{n} (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n} (\varepsilon_i)^2}$	(19)

Table 2
Examples of acceptance thresholds for M&V models according to ASHRAE Guideline 14:2023.

Diagnostic indicator	Time scale	Acceptable thresholds (%)
CV(RMSE)	Monthly	15
	Hourly	30
NMBE	Monthly	±5
	Hourly	±10

increased demands for heating or cooling. Therefore, it is expected that no alternative mathematical manifestations of the 5P shape will occur in actual building energy consumption patterns. Naturally, the model's form can be verified by utilising the second derivative of an approximating quadratic polynomial function, as described in Section 3.1. Fig. 3 illustrates the normal and abnormal appearances of the 5P, 3pH, and 3PC shapes as determined using this method.

In this study, we initially employed quadratic regression, using a globally positive second derivative as the criterion for normal 5P model shape. However, while this approach captures overall convexity, it lacks sensitivity to local variations. For instance, Fig. 3c and d show that even with global convexity, abnormal local inflection points can occur. To more precisely determine whether the 5P model aligns with theoretical expectations, we recommend further higher-order nonlinear fitting. As demonstrated in Fig. 3e and f, cubic regression reveals one or more inflection points that partition the domain into convex (positive second derivative) and concave (negative second derivative) segments. According to the basic physical interpretation of the models, both the heating and cooling zones should exhibit monotonic convexity (i.e., consistently positive curvature); hence, the presence of local negative curvature indicates an anomaly, and such data should be flagged. However, it should be noted that there are some cases in which a concave shape (i.e., Fig. 3h and j) may correspond to a real physical behaviour, where peak power is capped, for example. However, under

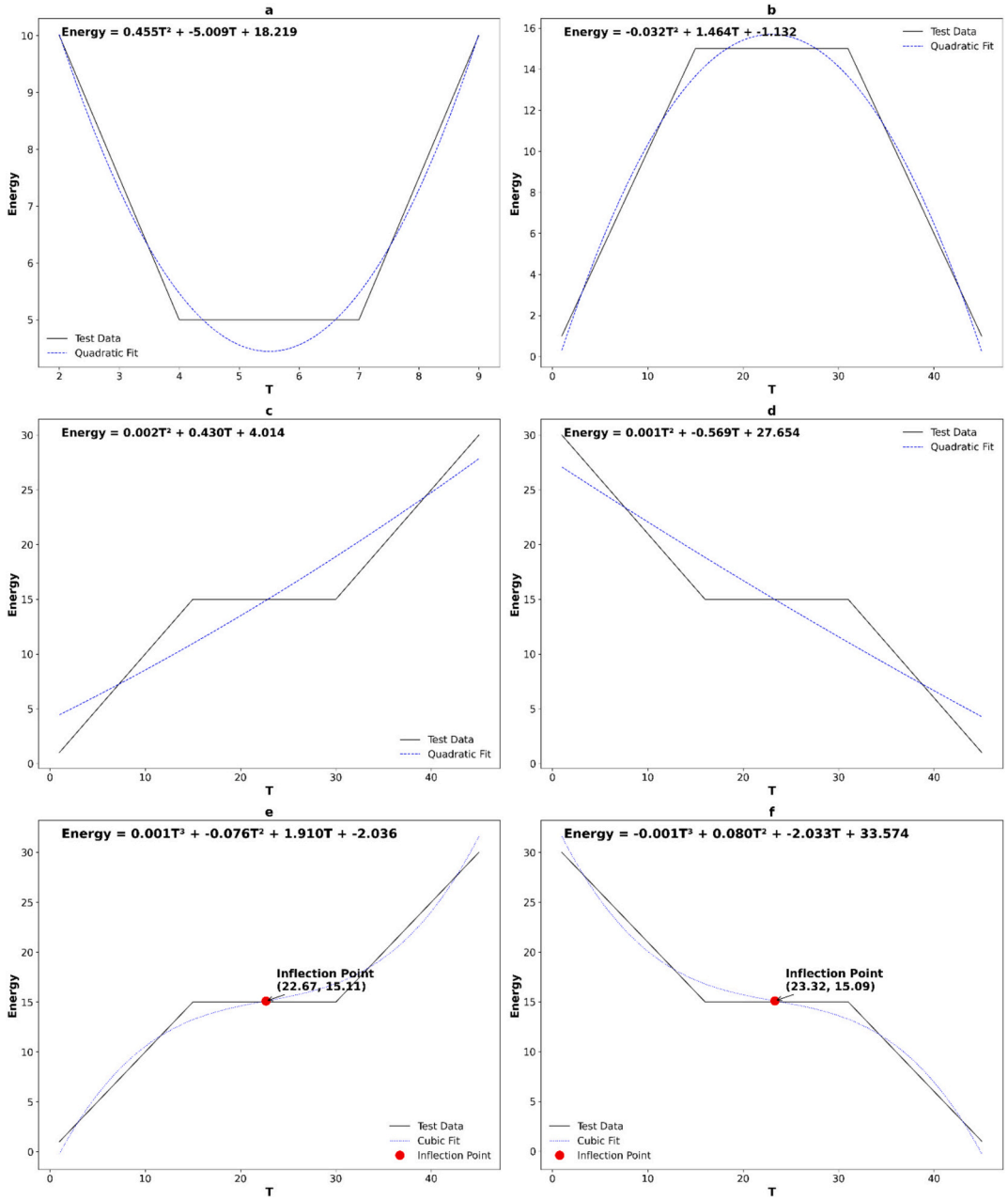


Fig. 3. Normal (**a, g, i**) and abnormal (**b, c, d, e, f, h, j**) shape of 5P, 3pH, and 3PC judgement.

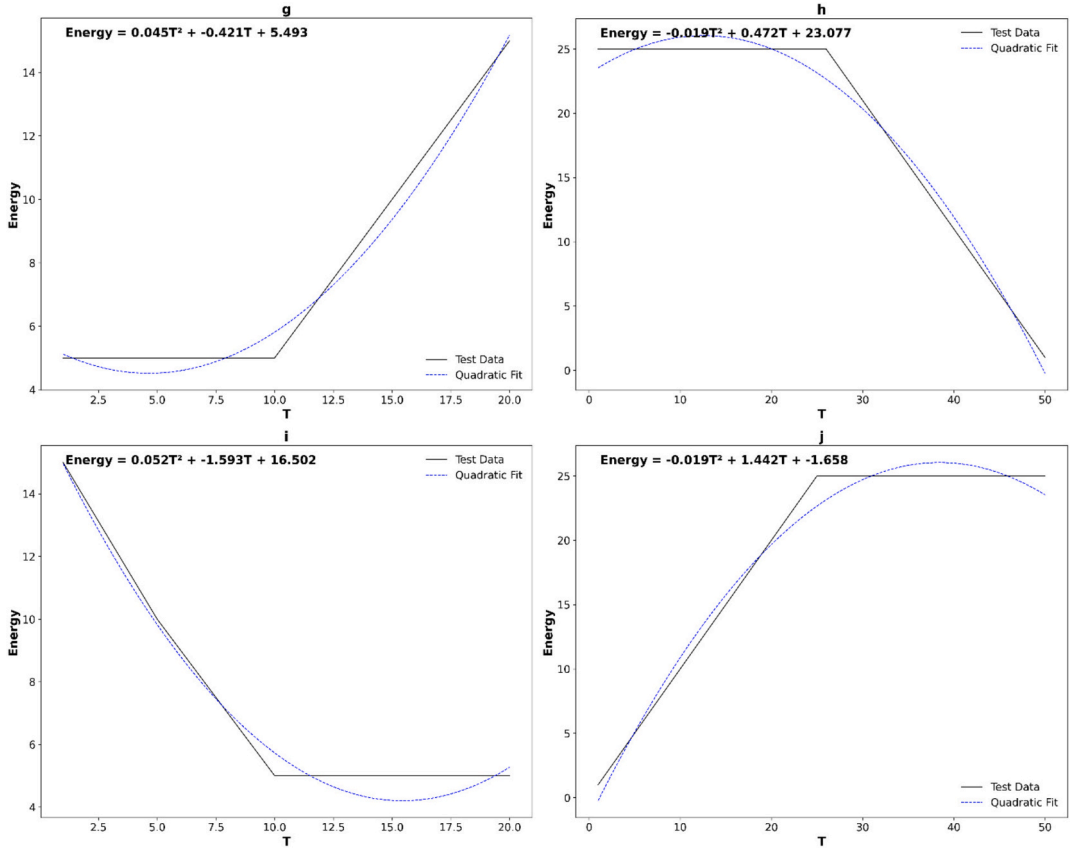


Fig. 3. (*continued*).

normal conditions such shapes should seldom occur. If they do appear, one should first check whether there are errors in measurement or there is effectively a constrained power. If this is the case, the monotonic 3pH or 3PC model described in Section 3.1 may be applied with some changes, but this is not considered in the current implementation of the model. Further, the 2-parameter (2P) cooling model in the current implementation is treated as a limit case of the 3PC model, where there is no change point, but the computed change point coincides with the lower bound of the outdoor temperature range. Finally, 4-parameter heating (4pH) and 4-parameter cooling (4PC) are convex as well and similar in principles to the shapes in Fig. 3g and i, but with a sloped segment instead of a horizontal one; the slope at lower/higher temperatures (below/above the change-point) respectively will be higher. This case is excluded from the current formulation as it is less relevant in the possible automated workflows outlined in Section 2 but may developed in future research. Generally speaking, the purpose of the model shape tests is to determine automatically whether a particular pattern in the energy consumption data is normal or abnormal, according to the physical interpretation, before proceeding with piecewise linear modelling fitting.

4.2. Testing with ASHRAE-14:2023 test cases

The first dataset tested is the ASHRAE RP-1050 test dataset developed by IMT. Among these, there are 13 synthetic datasets each for the 3pH and 3PC models, and 15 synthetic datasets for the 5P model. Detailed test results for each model are presented with tabular and graphical data in Appendices A, B, and C respectively for 3pH, 3PC and 5P models. The results indicate a substantial agreement with IMT's results in most of the cases, while the size of the testing datasets are reported in Appendix E. Here, cases that differ from IMT's test outcomes are reported and discussed. In general, the model successfully passed the ASHRAE shape test, and the test results for the 3pH and 3PC datasets are almost identical to those reported by IMT. The only differences emerged in the tests on the 5P datasets. For most of the 5P datasets, both the model coefficients and statistical indicators deviated only slightly from IMT's findings; however, for the "Large" and "Max_y" datasets, our model shows a smaller spread in residuals pattern, more accurately reflecting the data behaviour compared to the IMT model, as illustrated in Fig. 4.

It can be observed that both our 5P model and IMT's model provide a close fit to the original data for these two datasets. However, the residual plots reveal that, for these specific test sets with very high y-values, the deviation between our 5P model's predictions and the observed values is exceedingly small. This clearly explains the significant difference in the RMSE values reported in Appendix C for these datasets, thereby demonstrating that our 5P model achieves a better fit.

Regarding the computation time, a "wrapper" (i.e., software that encapsulates other software) to run the computationally efficient, Fortran-programmed software tool (IMT) in a Python environment has been created and made available for testing purpose. This choice is because IMT represents the reference software tool in this research, which can be run in a more flexible way (e.g., batch mode, multiple files at a time, etc.) as part of more complex automated workflows enabled by programming languages like Python. The use of a wrapper is an approach frequently used with efficient Fortran/C++ legacy code for numerical methods.

The IMT executable integrated within a Python workflow takes 10.619 s to run all ASHRAE 14:2023 test cases considered (3pH, 3PC, 5P) in a single execution on a PC with an i7-14700KF 3.40 GHz processor, while the Python implementation presented in this research completes the same task in 2.614 s, showing a reasonable computational speed improvement. A direct comparison of the two computational methods would require the complete reimplementation of IMT software

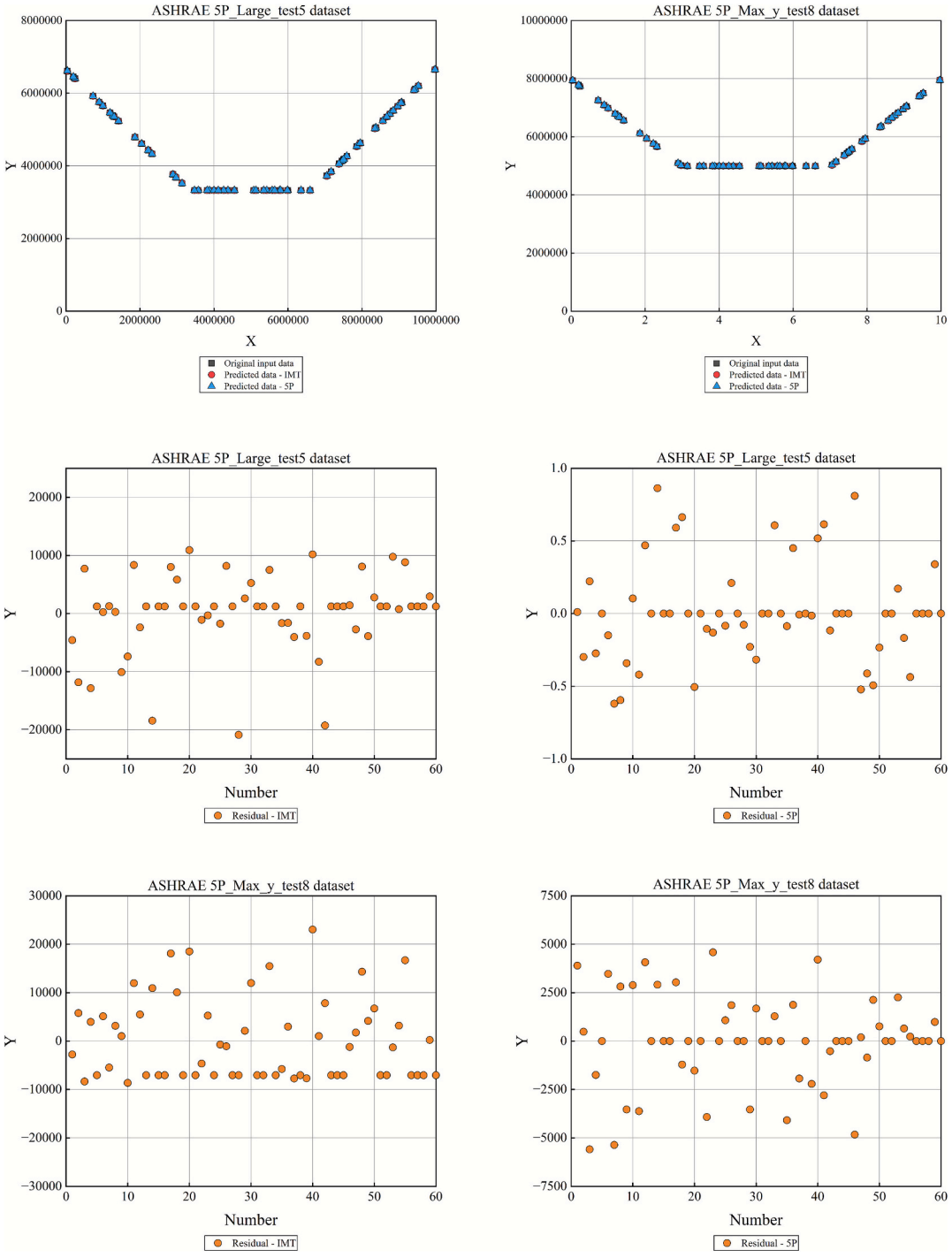


Fig. 4. Comparison of 5P model on the ASHRAE dataset with IMT.

in Python, but this goes beyond the scope of the present research, and it is indicated among the limitations in Section 4.4. In this case, the computational speed advantage would likely be more limited.

However, even a modest computational speed advantage when running a single model (e.g., a single building) could become significant when running multiple models simultaneously (e.g., different model configurations for the same buildings or for multiple buildings). Complex workflows would undoubtedly benefit from a reduction in execution time, even if it were only marginally significant at the level of a single model run.

4.3. Testing with measured data from real buildings

Next, tests were conducted using three years of continuously monitored data from two real-world case studies, both of which had been previously studied and reported. One case involves a passive house located in the Province of Forlì-Cesena in the Emilia-Romagna region of Italy, with monitoring conducted on a monthly basis, referred to as Case 1, previously modelled [95]. The other case pertains to the Hale Court shelter housing development project managed by Portsmouth City Council, which forms part of the EU's Horizon 2020 THERMOSS project [72]; for this project, data were monitored both monthly and daily, and is referred to as Case 2. The test results for the monthly data (Cases 1 and

2) will be presented together in Section 4.3.1, while those for the daily data (only Case 2) will be reported separately in Section 4.3.2. The size of the test datasets is reported in Appendix E.

4.3.1. Testing with monthly data – Cases 1 and 2

In Case 1, various types of energy demand were considered, including:

1. Total electricity demand.
2. Electricity demand for heating, ventilation, and air conditioning (HVAC) and domestic hot water (DHW).
3. Electricity demand for appliances.
4. Energy demand for heating.
5. Energy demand for cooling.

The raw energy data were converted into energy signature by dividing the original energy consumption values by the number of operating hours within the designated time interval. This procedure yields the average power value over the analysis period, in this case, monthly intervals spanning three years (a total of 36 months), rendering the data suitable for regression modelling. Typically, in the regression modelling of raw energy data, it is expressed as a function of energy signature and outdoor air temperature.

This approach is used because outdoor air temperature is the most relevant variable for weather normalisation and is frequently employed as the sole predictor in regression-based methods (i.e. temperature response functions). Fig. 5 presents scatter plots illustrating the relationship between the energy signature and outdoor air temperature for both electricity demand (left panel) and energy demand (right panel). It includes training data (70 %) and test data (30 %), as well as regression prediction data fitted using the 5P model. It can be observed that the fit obtained using the 5P model is consistent with the case data. The segments corresponding to total electricity demand, as well as the HVAC and DHW electricity demand segments, when fitted by the 5P model, successfully pass the shape test and exhibit the anticipated U-shape. For the energy demand segments related to heating and cooling, the algorithm automatically returns the optimal 3pH and 3PC models during the fitting process. The resulting fits likewise pass the shape test, displaying the expected shape.

However, there is a limitation regarding the appliance electricity demand segment. Under normal circumstances, this segment would be expected to conform to a constant model (i.e. displaying no temperature dependency). In this particular case, however, the presence of a solar thermal system that supplies part of the domestic hot water leads to a base load that is influenced by both temperature and solar radiation. Consequently, the fit for the appliance electricity demand segment is also expected to exhibit a U-shape, albeit with a relatively lower magnitude.

Fig. 6 illustrates the monthly time series forecast results of energy demand, based on the regression model fitted to the energy signature. The left panel shows the forecasts for electricity demand, while the right panel depicts those for energy demand. In both instances, the differences between the measured and predicted values are very small.

The error terms for outdoor air temperature dependence and time dependence in Case 1 are plotted in Fig. 7, to highlight potential patterns in the residuals.

In Case 2, the natural gas energy signature was employed as the predictor variable because the electricity demand remained essentially unchanged during the monitoring period. This stability is attributable to the operation of auxiliary equipment, which is not affected by temperature fluctuations and thus operates in an almost constant manner throughout the year. Furthermore, as the case study building underwent retrofitting interventions during the three-year monitoring period, its natural gas demand signature is categorised into three distinct periods. For a detailed description, please refer to Manfren et al. [72]. Fig. 8 displays a scatter plot of natural gas demand versus outdoor air temperature (left panel), derived from fitting the energy signature model using the 5P approach. Since natural gas is only associated with heating and domestic hot water, the algorithm automatically selects the most appropriate 3pH model during the fitting process. The resulting fit also passes the shape test, exhibiting the expected form. On the right panel is the monthly time series forecast of natural gas consumption, derived from the regression model fitted to the energy signature. The model is able to accurately capture the temporal pattern, with only minimal differences between the observed and predicted values. The error terms for outdoor air temperature dependence and time dependence in Case 2 monthly data are plotted in Fig. 9.

Overall, although the 5P model is relatively simple, it has performed very well on the monthly data from both real-world cases, passing the shape test. Moreover, it demonstrates an automated workflow that when the data tend to exhibit a 3pH/3PC shape, the algorithm automatically fits the corresponding model function. In addition, the model's fit results are satisfactory; Table 3 reports the statistical indicators from these two case studies, which can be interpreted in light of the thresholds outlined

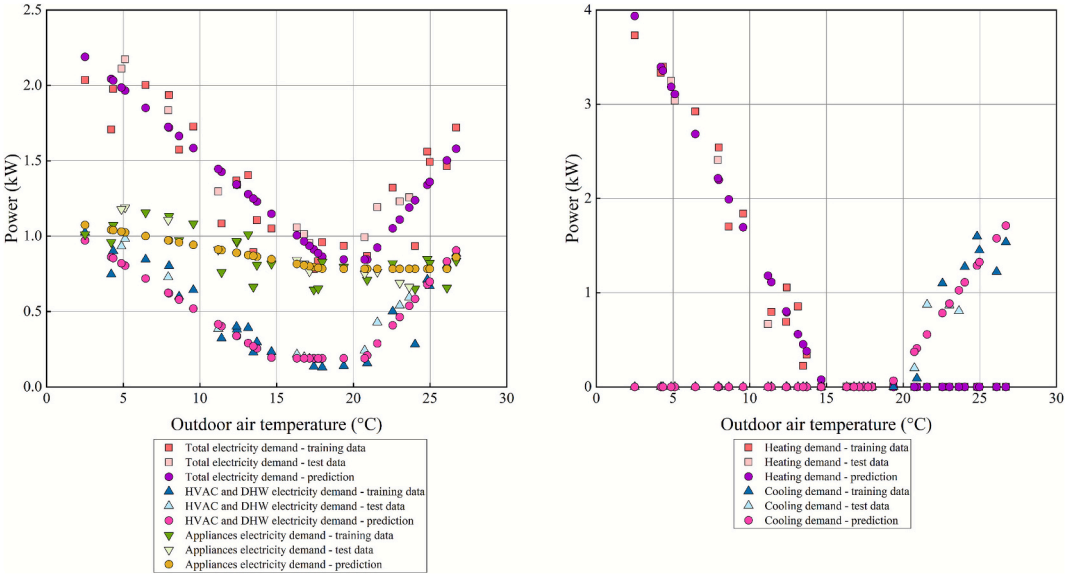


Fig. 5. Monthly electricity and energy demand signature – Case 1.

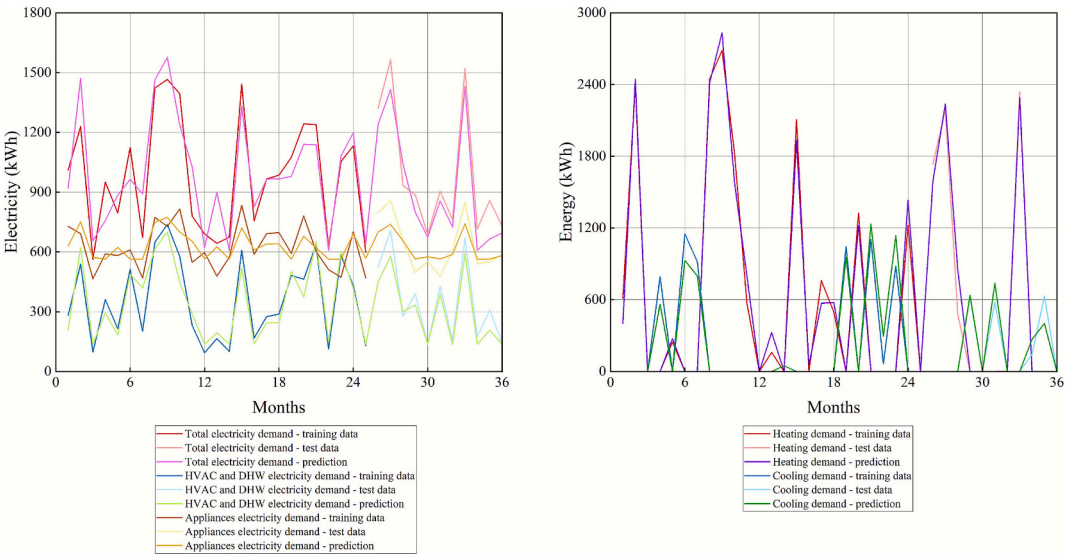


Fig. 6. Monthly electricity and energy consumption time series – Case 1.

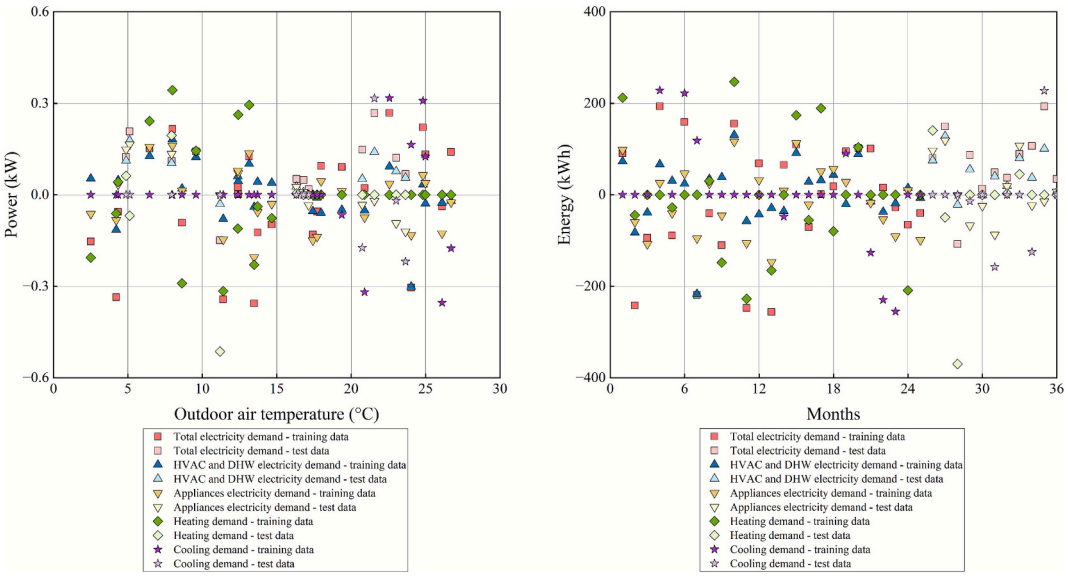


Fig. 7. Error term for the monthly electricity and energy demand signature/consumption time series – Case 1.

in Table 2.

The models developed for Case 1 have ranges of R^2 and *NMBE* similar to the ones found in previous research [95], which were in the range 67.34–91.23 % and – 1.07-0.04 %, respectively, despite using a different regression-based formulation. The main difference identified is in the CV (RMSE) range (10.56–17.88 % in the previous paper), which was computed for the entire dataset originally due to the small number of data points (reported in Appendix E) and has been recomputed using the 70/30 % dataset split. A prolonged monitoring period is necessary for attaining more stable statistical indicators; however, the principal objective of this research is to evaluate the applicability of the modelling methodology because the algorithms were tested with the case reported in Section 4.2 and Appendices A, B, and C. The outcomes of model fitting employing complete datasets, as in prior research, are documented in Appendix D with tabular and graphical data to facilitate better comparison for both Case 1 and Case 2.

Further, the models developed for Case 2 in previous research [72] have ranges of R^2, *NMBE* and *CV(RMSE)* respectively 91.11–97.93 %, −0.02-0.10 % and 6.35–18.25 %, respectively, so similar to the ones found in this research, despite using a different regression-based formulation. Also in this case, the differences are not very large, however, in Case 2 "Period 2 - testing dataset" and "Period 3 – testing dataset", certain statistical indicators in Table 3 are not calculated because, after partitioning the dataset in a 70 %/30 % ratio, a limited number of data points remain (reported in Appendix E), rendering it impossible to obtain stable estimates of the indicators due to the formula's reliance on a correction based upon the model's degrees of freedom. This is a limitation for case 2, as the building has undergone refurbishment in many phases, albeit over relatively brief durations. Consequently, in the original research paper [72], daily and then hourly data models were used. The analysis of daily data for case 2 is detailed later in Section 4.3.2.

4.3.2. *Testing with daily data – Case 2*

As described in the previous Section, Case 2 was divided into Period 1, Period 2, and Period 3 over the entire three-year monitoring period to capture the natural gas demand signature. In this section, the 3pH model is applied to daily natural gas demand data from these three periods to

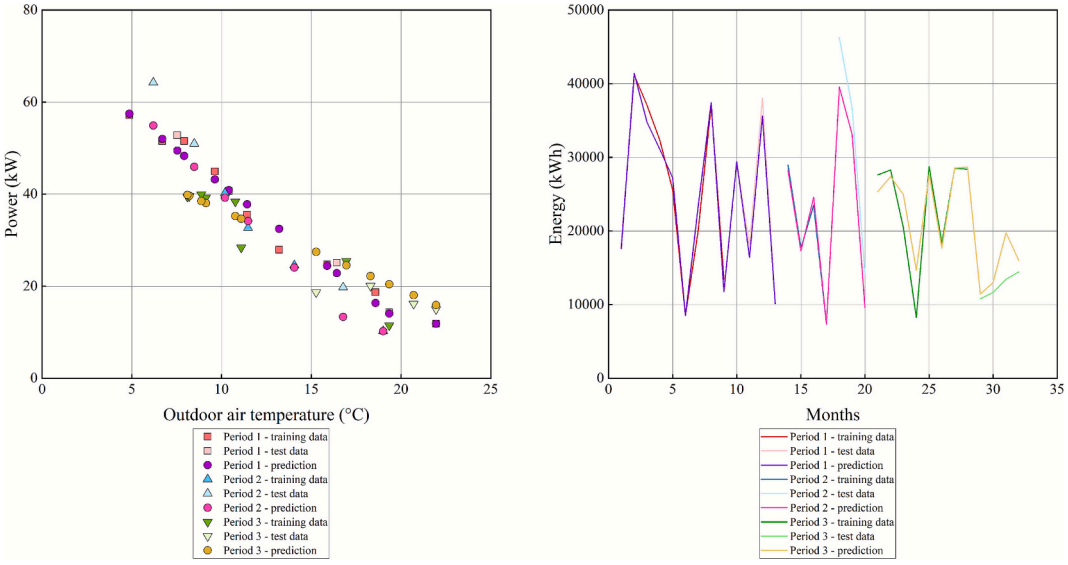


Fig. 8. Monthly natural gas demand signature and consumption time series – Case 2.

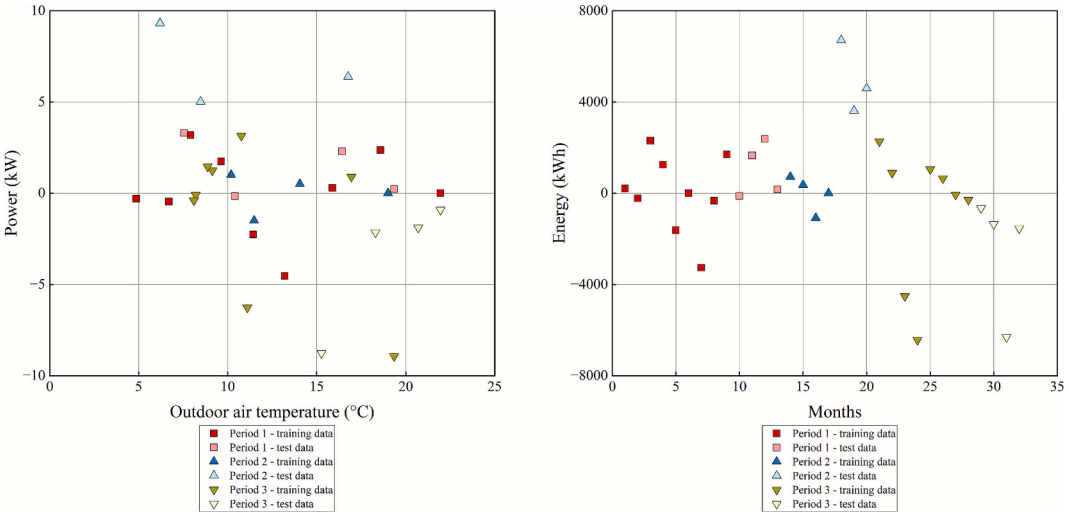


Fig. 9. Error term for the monthly natural gas demand signature and consumption time series – Case 2.

Table 3
Results of the monthly interval data's energy signature analysis.

Case	Type	Statistical indicators					
		SSE	R^2	$AdjR^2$	*RMSE*	*CV(RMSE)*	*NMBE*
			%	%		%	%
Case 1	Total electricity – training dataset	0.81	79.23	75.07	0.20	14.80	0.81
	Total electricity – testing dataset	0.21	89.52	82.53	0.19	13.67	0.21
	HVAC and DHW electricity – training dataset	0.22	88.64	86.36	0.11	21.32	0.22
	HVAC and DHW electricity – testing dataset	0.09	89.48	82.46	0.12	24.50	0.09
	Appliances electricity – training dataset	0.27	55.46	46.55	0.12	13.42	0.27
	Appliances electricity – testing dataset	0.09	74.57	57.62	0.13	14.23	0.09
	Heating – training dataset	0.66	98.32	97.98	0.18	19.34	0.66
	Heating – testing dataset	0.31	98.28	97.85	0.20	23.14	0.31
	Cooling – training dataset	0.50	94.18	93.02	0.16	47.84	0.50
	Cooling – testing dataset	0.18	88.24	85.31	0.15	59.83	0.18
Case 2	Period 1 – training dataset	44.81	97.85	97.13	2.73	7.60	0.00
	Period 1 – testing dataset	16.28	98.11	94.34	4.03	12.15	−4.26
	Period 2 – training dataset	3.51	99.29	97.87	1.87	6.97	0.00
	Period 2 – testing dataset	152.87	–	–	–	–	–
	Period 3 – training dataset	133.73	81.98	74.78	5.17	15.80	3.41
	Period 3 – testing dataset	85.85	–	–	–	–	–

assess the accuracy of the 3pH model on daily data. Fig. 10 shows, on the left panel, a scatter plot of the natural gas energy signature versus outdoor air temperature fitted using the 3pH model, and on the right panel, the daily time series forecast of natural gas consumption based on the corresponding regression. Similarly, since natural gas is solely associated with heating, the algorithm automatically selects the most appropriate 3pH model during the fitting process. The resulting fit passes the shape test, exhibits the anticipated form, and accurately captures the temporal pattern, with a high degree of concordance between the predicted and observed values. The split between training and testing is the same as the one used in the previous section, 70 % for training and 30 % for testing. The error terms for outdoor air temperature dependence and time dependence in Case 2 daily data are plotted in Fig. 11.

In summary, despite its simplicity, the 3pH model successfully passes tests based on daily data. It is capable of automatically fitting to determine the optimal model function, even when confronted with noisy, real-world daily data. Furthermore, the model's fitting results are robust, with the statistical indicators reported in Table 4, which can be interpreted according to the thresholds outlined in Table 2. For Case 2 in the original research [72] statistical indicators R^2 and *CV(RMSE)* were in the range 55.80–91.42 % and 13.89–19.59 %, respectively, so similar to the ones found in this research, despite using a different regression-based formulation. The outcomes of model fitting employing complete datasets, as in prior research, are documented in Appendix D to facilitate better comparison.

4.4. Limitations and further work

The current analytical formulation, while effective for the widely used 3PH, 3PC, and 5P models, represents a subset of the seven regression model types prescribed in ASHRAE Guideline 14:2023. The 2P cooling model is treated as a limit case in the current implementation of the 3PC model, while the 4pH and 4PC models are excluded from it.

The models have been validated using ASHRAE's Inverse Modelling Toolkit (IMT) test cases and then applied to datasets analysed in previous research for further testing. Using the IMT executable in a Python workflow, all the ASHRAE 14:2023 test cases that were taken into consideration throughout the validation process (3pH, 3PC, and 5P) can be run in 10.619 s, whereas the new implementation (pure Python) can do so in 2.614 s, which is a respectable increase. IMT software would need to be completely reimplemented in Python in order to compare the two computing approaches directly in terms of execution time. In this instance, the computational improvement can be smaller but still significant, and this will be considered in future studies. At the same time, including indicators like the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) would aid in better understanding the trade-off between model fitting performance and complexity, particularly in workflows that incorporate additional explanatory variables.

Further, the methodology currently relies on standard least squares fitting and has not yet incorporated advanced objective functions for regularisation (e.g., elastic net) that could enhance the model's robustness when dealing with outliers and anomalies in data without losing interpretability.

Overall, the use of the analytical formulation within hourly and sub-hourly resolution modelling workflows (like Time of Week and Temperature or other regression-based workflows), which are increasingly relevant for advanced M&V applications and digital twin paradigms, seems the more immediate further development for this research, together with the inclusion of different objective/loss functions to enhance robustness.

An alternative solution could be the development of analytical formulations for 4pH and 4PC models, which can then be used for hourly interval modelling in a similar way to ECAM 7.0 [79] software (regression trees), but with a different algorithmic implementation. To automate the data analysis process, every model described above can be employed concurrently, maintaining interpretability as they are all regression-based.

5. Conclusion

The essence of this research lies in simplifying and automating workflows for advanced Measurement and Verification (M&V) practices through an analytical formulation meant to address key knowledge gaps identified. The gaps identified concerned, in particular, the grid search or optimisation required to identify change points (in this specific case, balance points for heating and cooling) in piecewise linear regression models such as the ones proposed by ASHRAE Guideline 14:2023 (namely, 3pH, 3PC, and 5P) and the possibility to streamline and enhance the capabilities of advanced workflows, employing hourly and sub-hourly resolution data and more sophisticated techniques, without sacrificing interpretability.

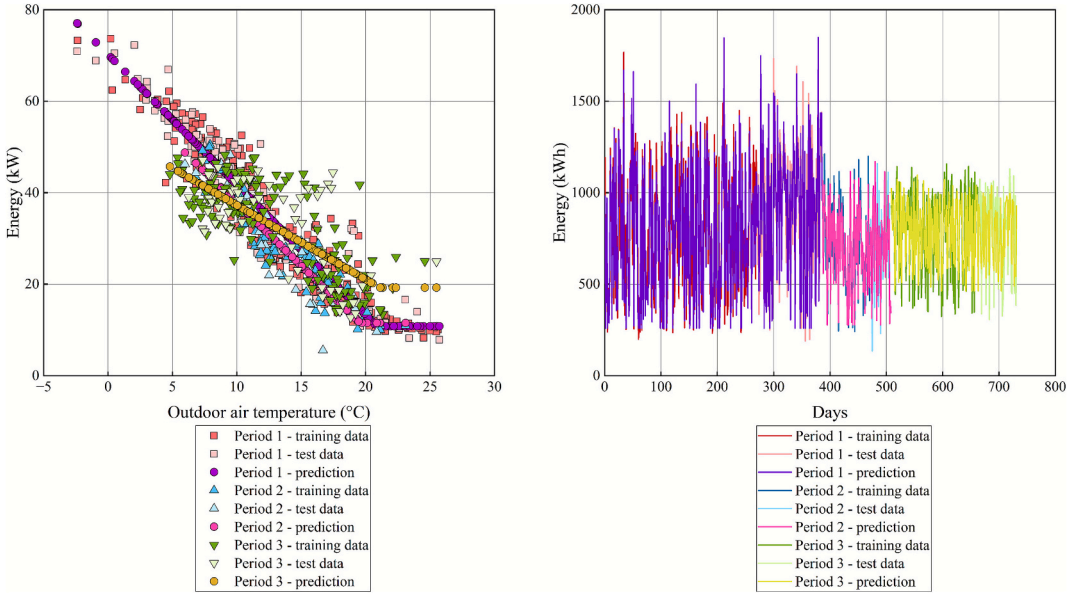


Fig. 10. Daily natural gas demand signature and consumption time series – Case 2.

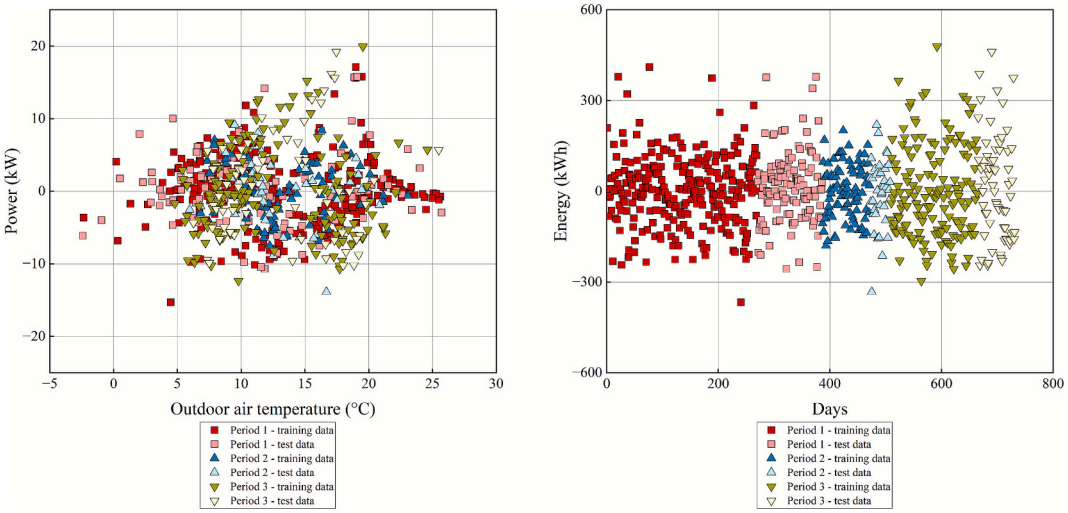


Fig. 11. Error term of the daily natural gas demand signature and consumption time series – Case 2.

Table 4
Results of the daily interval data's energy signature analysis.

Case	Type	Statistical indicators					
		SSE	R^2	$AdjR^2$	*RMSE*	*CV (RMSE)*	*NMBE*
			%	%		%	%
Case 2	Period 1 – training dataset	6083.87	91.12	91.05	4.77	14.58	0.00
	Period 1 – testing dataset	2747.97	91.48	91.33	4.93	12.70	−1.45
	Period 2 – training dataset	1205.01	86.24	85.91	3.83	13.18	0.00
	Period 2 – testing dataset	783.87	83.95	83.00	4.80	16.65	0.95
	Period 3 – training dataset	5985.11	60.07	59.55	6.25	18.90	0.15
	Period 3 – testing dataset	3378.87	45.62	43.92	7.27	21.96	−1.71

Indeed, it is fundamentally important for humans to understand the inner workings of a machine learning workflow for advanced M&V to enable trust by practitioners and stakeholders. Further, interpretability of machine learning models encompasses various potential techniques beyond linear multivariate regression (e.g., lasso, elastic net, quantile regression, symbolic regression, etc.), and the analytical approach proposed can be developed further by incorporating additional characteristics from these techniques.

The analytical approach undergoes validation using both test cases from ASHRAE Guideline 14:2023 and measured data from monitored buildings. For the test cases, the method is dependable and gives results that are very similar to those from the Inverse Modelling Toolkit (IMT) software used by ASHRAE for validation. The ASHRAE test cases considered are 39 in total, divided between 3pH (13), 3PC (13), and 5P (15) model types, showing comparable results in the large majority of cases and minor discrepancies only for the "Large test 5" and "Max y test 8" datasets. The total batch runtime has been sensibly reduced compared to the original implementation, from 10.619 s using the IMT toolkit executable integrated within a Python workflow to 2.614 s using the analytical implementation in Python, in the machine used in this research.

Regarding the use of measured data, the models developed provided similar model fitting capabilities to the ones developed in previous research, but with a higher level of automation, given by the possibility to choose the right model (either 3pH, 3PC or 5P) without compromising interpretability. This workflow can be improved further by using it to create a higher-resolution model with hourly or sub-hourly data intervals (e.g., integrating with real-time data) and exploiting automated workflows already found in literature.

Overall, the test cases and the real case studies indicated the possibility to develop sophisticated analytical workflows while retaining interpretability. By automating the generation processes for interpretable models up to hourly/sub-hourly data intervals, innovative workflows may be created towards the "digital twins" paradigm for building facilities to provide meaningful insights for stakeholders on energy performance and other related issues in a transparent way.

The creation of open-source software aims to promote collaborative research in this area, enabling further exploration of modelling solutions and practical refinement of the ones made available in this research. Future research trajectories include enhancing interpretability across the entire workflow, preventing the emergence of opaque "black-box" stages of the analysis and increasing robustness. Advanced statistical methods such as lasso, elastic net, and quantile regression can be used for this purpose. Finally, the methodology developed and the analytical workflow presented in this research represent advancements in enabling intelligent, interpretable, and automated energy analytics for buildings, where the model identification can be made constraint aware and auditable towards more transparent and effective building management practices.

Software and data

Software and data are made available at: https://github.com/EnergyDEPLOY/MVCP

CRediT authorship contribution statement

Massimiliano Manfren: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Rundong Liao:** Writing – original draft, Visualization, Validation, Software, Resources, Data curation. **Benedetto Nastasi:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declaration of competing interest

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

Appendix A. Appendix A – ASHRAE RP-1050 dataset 3pH test results comparison with the IMT.

Table 5
ASHRAE RP-1050 3pH tests' results.

Dataset	Model	b_0	b_1	b_2	R^2	AdjR^2	RMSE	CV(RMSE)	DW
3_POINT	IMT				Unable to model				
	Proposed	5.0000	1.1111	10.0000	1.000	NaN	NaN	NaN	NaN
5_POINT	IMT	5.0016	1.2535	8.9800	1.000	1.000	0.006	0.078	2.708
	Proposed	5.0000	1.2500	9.0000	1.000	1.000	0.000	0.000	NaN
9000_P	IMT	5.0063	0.9989	4.9984	0.969	0.969	0.289	4.626	1.963
	Proposed	5.0069	0.9996	4.9952	0.969	0.969	0.289	4.626	1.963
Large	IMT	9,370,096.00	1.0000	9,370,100.00	1.000	1.000	1.612	0.000	0.923
	Proposed	9,370,097.40	1.0000	9,370,097.36	1.000	1.000	0.049	0.000	2.887
Max_size_X	IMT	5.0000	0.0000	5,004,678.50	1.000	1.000	0.000	0.003	1.722
	Proposed	5.0000	0.0000	5,004,506.25	1.000	1.000	0.000	0.003	1.788
Max_size_Y	IMT	5,004,020.00	992,540.25	5.0045	1.000	1.000	1328.09	0.022	1.866
	Proposed	5,004,678.78	993,657.06	5.0000	1.000	1.000	196.05	0.003	2.175
PACKED	IMT	30.0132	3.0545	24.5200	0.985	0.985	2.938	6.168	2.235
	Proposed	30.0765	3.0717	24.4098	0.985	0.985	2.967	6.229	2.240
SCATTER	IMT	77.4479	2.9475	25.5000	0.407	0.407	29.301	30.563	2.286
	Proposed	78.5366	2.9427	25.0246	0.405	0.380	29.648	30.925	2.284
Slope_A	IMT				Unable to model				
	Proposed	5.0000	0.0000	25.5000	NaN	NaN	0.000	0.000	NaN
Slope_B	IMT	25.0000	26.0417	25.4800	0.571	0.571	5.520	17.523	0.124
	Proposed	25.0000	12.5000	26.0000	0.571	0.553	5.578	17.709	0.124
Slope_C	IMT	25.0000	−24.0000	25.5000	0.581	0.581	5.204	27.390	0.129
	Proposed	25.0000	−12.0000	26.0000	0.581	0.563	5.259	27.680	0.129
Slope_D	IMT	24.9103	−1.0244	24.5200	1.000	1.000	0.153	0.804	0.207
	Proposed	25.0000	−1.0000	25.0000	1.000	1.000	0.000	0.000	NaN
Small	IMT	0.0005	0.9871	0.0005	1.000	1.000	0.000	0.451	2.099
	Proposed	0.00 05	0.9950	0.0005	1.000	1.000	0.000	0.431	2.103

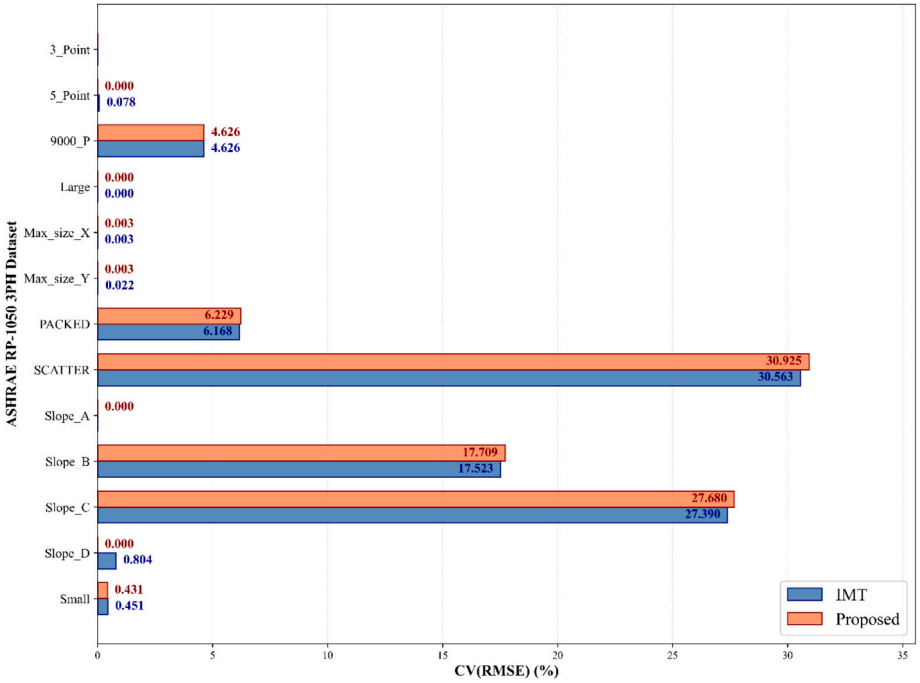


Fig. 12. ASHRAE RP-1050 3PH graphical comparison of *CV(RMSE)* between the IMT and the proposed model.

The *CV(RMSE)* statistical indicator is used for error comparison in tests because it represents *RMSE* normalised with respect to the mean (expressed in percentage), and thus it enables comparison across datasets with different ranges of values, as in the case of the ASHRAE RP-1050 3pH tests. It can

be seen in Fig. 12 how the proposed analytical approach performs slightly better (i.e., lower *CV(RMSE)* value) than IMT in the cases with low *CV (RMSE)*, i.e., less scattered data, or in the limit test cases like "5_point" or "Slope_D", which have an exact solution, represented by a piecewise linear shape. On the other hand, IMT performs slightly better in cases with scattered data, e.g., test cases like "SCATTER", "SLOPE_B" and "SLOPE_C". Overall, the difference between the two approaches in terms of *CV(RMSE)* is very small for all the test cases.

Appendix B. Appendix B – ASHRAE RP-1050 dataset 3PC test results comparison with the IMT.

Table 6
ASHRAE RP-1050 3PC tests' results.

Dataset	Model	b_0	b_1	b_2	R^2	AdjR^2	RMSE	CV(RMSE)	DW
3_POINT	IMT				Unable to model				
	Proposed	5.0000	1.0000	10.0000	1.000	NaN	NaN	NaN	NaN
5_POINT	IMT	5.0077	1.0136	10.1200	1.000	1.000	0.030	0.380	1.187
	Proposed	5.0000	1.0000	10.0000	1.000	1.000	0.000	0.000	NaN
9000_P	IMT	4.9997	0.9996	4.9984	1.000	1.000	0.001	0.008	1.990
	Proposed	5.0000	1.0000	5.0000	1.000	1.000	0.000	0.000	NaN
Large	IMT	443,143.75	1.0000	443,145.47	1.000	1.000	0.478	0.000	0.652
	Proposed	443,143.60	1.0000	443,143.60	1.000	1.000	0.000	0.000	1.500
Max_size_X	IMT	5.0091	0.0000	5,004,678.50	1.000	1.000	0.012	0.190	2.007
	Proposed	5.0000	0.0000	4,969,447.87	1.000	1.000	0.002	0.032	1.829
Max_size_Y	IMT	4,996,527.00	986,543.81	5.0045	1.000	1.000	10,770.09	0.169	1.963
	Proposed	5,004,678.78	993,649.74	5.0364	1.000	1.000	203.934	0.003	2.155
PACKED	IMT	27.4776	1.0379	25.5000	0.999	0.999	0.290	0.854	2.293
	Proposed	27.4136	1.0225	25.1865	0.999	0.999	0.274	0.806	2.613
SCATTER	IMT	44.0362	3.4570	24.5200	0.547	0.547	27.107	40.237	2.601
	Proposed	43.6100	3.3662	23.9264	0.547	0.528	27.388	40.654	2.602
Slope_A	IMT				Unable to model				
	Proposed	5.0000	0.0000	25.5000	NaN	NaN	0.000	0.000	NaN
Slope_B	IMT	25.0000	13.0208	24.0400	0.571	0.571	5.520	17.523	0.124
	Proposed	25.0000	12.5000	24.0000	0.571	0.553	5.578	17.709	0.124
Slope_C	IMT	25.0000	−12.0192	24.0400	0.536	0.536	5.474	28.812	0.109
	Proposed	25.0000	−11.5385	24.0000	0.536	0.516	5.532	29.117	0.109
Slope_D	IMT	24.9103	−1.0244	26.4800	1.000	1.000	0.153	0.804	0.207
	Proposed	25.0000	−1.0000	26.0000	1.000	1.000	0.000	0.000	NaN
Small	IMT	0.0005	1.0039	0.0005	1.000	1.000	0.000	0.084	1.960
	Proposed	0.0005	1.0000	0.0005	1.000	1.000	0.000	0.000	NaN

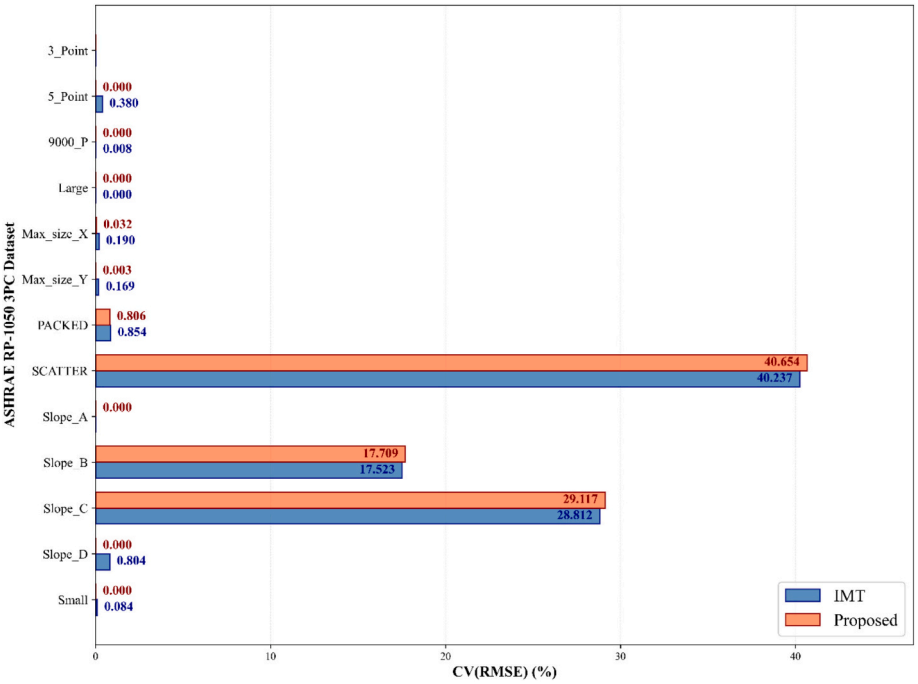


Fig. 13. ASHRAE RP-1050 3PC graphical comparison of *CV(RMSE)* between the IMT and the proposed model.

The *CV(RMSE)* statistical indicator is used for error comparison in tests because it represents *RMSE* normalised with respect to the mean (expressed in percentage), and thus it enables comparison across datasets with different ranges of values, as in the case of the ASHRAE RP-1050 3PC tests. It can be seen in Fig. 13 how the proposed analytical approach performs slightly better (i.e., lower *CV(RMSE)* value) than IMT in the cases with low *CV (RMSE)*, i.e., less scattered data, or in the limit test cases like "5_point" or "Slope_D", which have an exact solution, represented by a piecewise linear shape. On the other hand, IMT performs slightly better in cases with scattered data, e.g., test cases like "SCATTER", "SLOPE_B" and "SLOPE_C". Overall, the difference between the two approaches in terms of *CV(RMSE)* is very small for all the test cases.

Appendix C. Appendix C – ASHRAE RP-1050 dataset 5P test results comparison with the IMT.

Table 7
ASHRAE RP-1050 5P tests' results.

Dataset	Model	b_0	$b_1^{(h)}$	$b2^{(h)}$	$b_1^{(c)}$	$b2^{(c)}$	R^2	$AdjR^2$	RMSE	CV(RMSE)	DW
4_POINT	IMT					Unable to model					
	Proposed					Unable to model					
7_POINT	IMT	4.8723	−2.4657	4.0743	2.4657	6.9257	0.999	0.999	0.077	1.078	3.029
	Proposed	5.0000	−2.5000	4.0000	2.5000	7.0000	1.000	1.000	0.000	0.000	NaN
9000_P	IMT	2.9931	−1.0216	2.9624	0.9710	5.9237	0.946	0.946	0.288	6.798	1.939
	Proposed	3.0028	−0.9995	2.9933	0.9942	5.9943	0.947	0.947	0.287	6.777	1.937
Large	IMT	3,323,121.25	−0.9901	3,348,554.50	1.0063	6,660,803.00	1.000	1.000	6837.156	0.153	2.116
	Proposed	3,324,351.00	−1.0000	3,324,350.03	1.0000	6,648,701.41	1.000	1.000	0.345	0.000	1.666
Max_size_X	IMT	4.9983	0.0000	3,348,554.25	0.0000	6,660,803.00	1.000	1.000	0.010	0.153	2.107
	Proposed	5.0000	0.0000	3,325,350.10	0.0000	6,648,447.16	1.000	1.000	0.002	0.033	2.278
Max_size_Y	IMT	4,993,588.50		2.9826	1,008,166.38	7.0274	1.000	1.000	8566.653	0.145	2.327
	Proposed	4,986,525.97	−998,439.08	2.9981	997,361.22	7.0000	1.000	1.000	2406.329	0.041	2.745
PACKED	IMT	13.6478	−0.8945	17.2947	1.0204	28.7053	0.939	0.939	1.408	7.164	2.714
	Proposed	14.1479	−0.9735	15.5581	1.0888	29.6107	0.939	0.933	1.448	7.369	2.606
SCATTER	IMT	7.1445	−0.8616	23.8116	1.4693	27.0724	0.465	0.465	8.171	45.680	2.749
	Proposed	8.2929	−0.9084	19.5927	1.4293	26.1250	0.440	0.384	8.568	47.898	2.620
Slope_A	IMT					Unable to model					
	Proposed	5.0000	0.0000	25.5000	0.0000	25.5000	1.000	1.000	0.000	0.000	0.000
Slope_B	IMT	5.0000	−8.2500	5.5556	7.8000	9.4444	0.195	0.195	3.388	38.946	0.491
	Proposed	5.0000	−4.5833	6.0000	4.3333	9.0000	0.195	0.124	3.463	39.802	0.491
Slope_C	IMT	15.0000	8.2500	5.5556	−6.6000	9.4444	0.182	0.182	3.260	27.866	0.733
	Proposed	15.0000	4.5833	6.0000	−3.6667	9.0000	0.182	0.109	3.332	28.478	0.733
Slope_D	IMT	15.2201	0.9558	15.6667	−0.9582	30.3333	0.998	0.998	0.216	2.231	0.397
	Proposed	15.0000	1.0000	15.0000	−1.0000	31.0000	1.000	1.000	0.000	0.000	NaN
Slope_E	IMT	15.1639	0.9502	15.6667	1.0159	30.3333	1.000	1.000	0.174	1.055	0.382
	Proposed	15.0000	1.0000	15.0000	1.0000	30.0000	1.000	1.000	0.000	0.000	NaN
Slope_F	IMT	15.1626	−1.0169	15.6667	−0.9501	30.3333	0.999	0.999	0.184	1.199	0.383

(*continued on next page*)

Table 7 (*continued*)

Dataset	Model	b_0	$b_1^{(h)}$	$b2^{(h)}$	$b_1^{(c)}$	$b2^{(c)}$	R^2	$AdjR^2$	RMSE	CV(RMSE)	DW
	Proposed	15.0000	−1.0000	16.0000	−1.0000	31.0000	1.000	1.000	0.000	0.000	NaN
Small	IMT	0.0000	−0.9902	0.0000	1.0064	0.0000	1.000	1.000	0.000	0.000	2.934
	Proposed	0.0000	−1.0000	0.0000	1.0000	0.0000	1.000	1.000	0.000	0.008	1.902

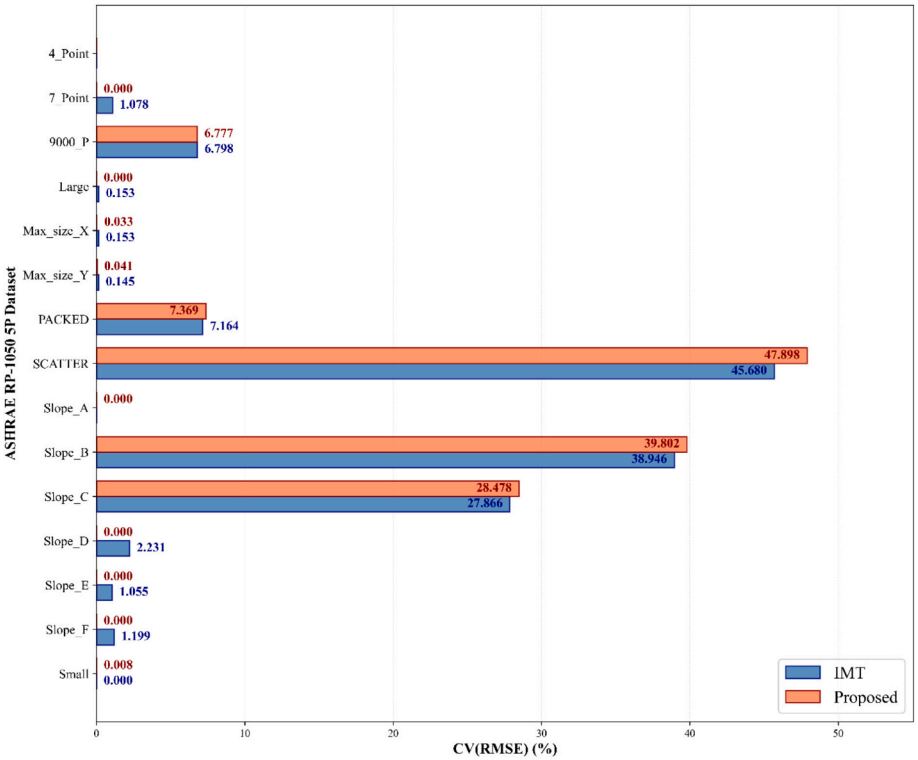


Fig. 14. ASHRAE RP-1050 5P graphical comparison of *CV(RMSE)* between the IMT and the proposed model.

The *CV(RMSE)* statistical indicator is used for error comparison in tests because it represents *RMSE* normalised with respect to the mean (expressed in percentage), and thus it enables comparison across datasets with different ranges of values, as in the case of the ASHRAE RP-1050 5P tests. It can be seen in Fig. 14 how the proposed analytical approach performs slightly better (i.e., lower *CV(RMSE)* value) than IMT in the cases with low *CV(RMSE)*, i.e., less scattered data, or in the limit test cases like "7_point" or "Slope_D" and others, which have an exact solution, represented by a piecewise linear shape. On the other hand, IMT performs slightly better in cases with scattered data, e.g., test cases like "SCATTER", "SLOPE_B" and "SLOPE_C". Overall, the difference between the two approaches in terms of *CV(RMSE)* is very small for all the test cases.

Appendix D. Appendix D – Case 1 and Case 2 results (entire datasets).

Table 8
Case 1 and Case 2 results (entire datasets) – monthly data.

Case	Dataset type	Statistical indicators					
		SSE	R^2	$AdjR^2$	*RMSE*	*CV(RMSE)*	*NMBE*
			%	%		%	%
Case 1	Total electricity	0.90	84.88	82.93	0.17	12.46	0.38
	HVAC and DHW electricity	0.27	90.56	89.34	0.09	18.65	−0.31
	Appliances electricity	0.30	69.25	68.35	0.09	10.50	0.00
	Heating	0.96	98.33	98.11	0.22	12.99	0.01
	Cooling	0.68	93.34	92.48	0.24	23.72	0.00
Case 2	Period 1	58.41	98.03	97.64	2.42	6.88	0.00
	Period 2	64.34	96.93	95.40	4.01	11.58	0.00
	Period 3	127.68	90.74	88.69	3.77	13.62	0.60

Table 9
Case 2 results (entire datasets) – daily data.

Case	Dataset type	Statistical indicators					
		SSE	R^2	$AdjR^2$	*RMSE*	*CV(RMSE)*	*NMBE*
			%	%		%	%
Case 2	Period 1	8820.40	91.50	91.45	4.80	13.89	0.00
	Period 2	1980.12	85.49	85.24	4.08	14.06	0.00
	Period 3	9359.30	55.86	55.46	6.52	19.71	0.00

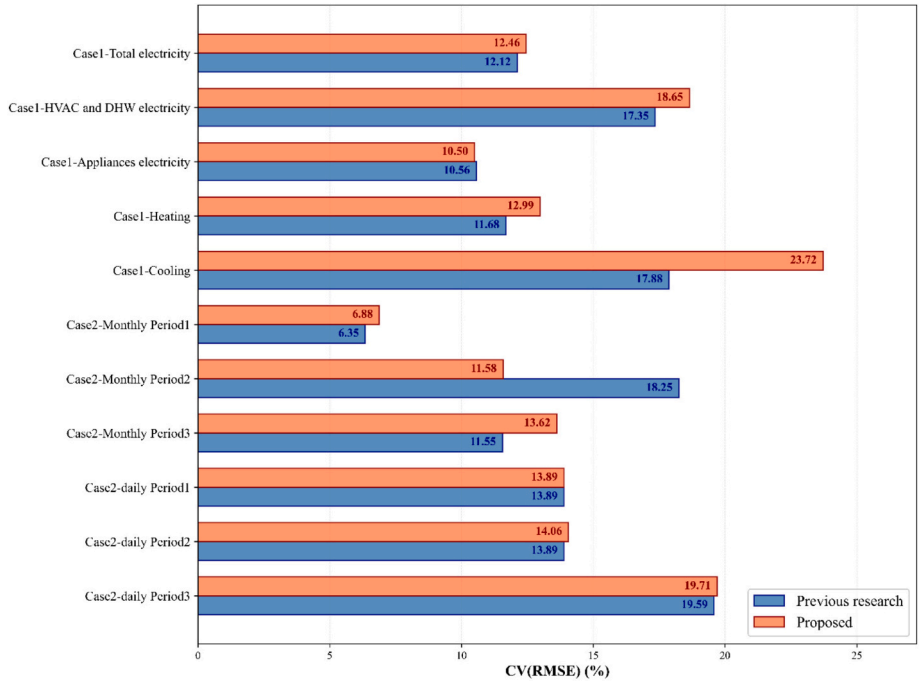


Fig. 15. Case study 1 and 2 graphical comparison of *CV(RMSE)* obtained using grid search and the proposed model.

The *CV(RMSE)* statistical indicator is used for error comparison in a similar way to Appendices A, B, and C. The results in Fig. 15 indicate that in most of the cases the difference between the grid search approach and the new proposed one is small, even though there is a notable difference for "Case 1-cooling", where grid search performs better, while for "Case 2-Monthly Period2", the new approach performs better than grid search.

Appendix E. Appendix E – Size of research datasets employed for testing.

Table 10
Size of ASHRAE RP-1050 3PC, 3pH, and P test datasets.

Case	Dataset name	Size
3pH model	3_POINT	3
	5_POINT	5
	9000_P	9000
	Large	13
	Max_size_X	60
	Max_size_Y	60
	PACKED	50
	SCATTER	50
	Slope_A	50
	Slope_B	50
	Slope_C	50
	Slope_D	50
	Small	61
3PC model	3_POINT	3
	5_POINT	5
	9000_P	9000
	Large	13
	Max_size_X	60
	Max_size_Y	60

(*continued on next page*)

Table 10 (*continued*)

Case	Dataset name	Size
	PACKED	50
	SCATTER	50
	Slope_A	50
	Slope_B	50
	Slope_C	50
	Slope_D	50
5P model	4_POINT	4
	7_POINT	7
	9000_P	9000
	Large	60
	Max_size_X	60
	Max_size_Y	60
	PACKED	45
	SCATTER	45
	Slope_A	50
	Slope_B	50
	Slope_C	50
	Slope_D	50
	Slope_E	50
	Slope_F	45
	Small	60

Table 11
Size of Case1 and Case 2 datasets.

Case	Dataset type	Size
Case 1	Total electricity – training dataset	25
	Total electricity – test dataset	11
	HVAC and DHW electricity – training dataset	25
	HVAC and DHW electricity – test dataset	11
	Appliances electricity – training dataset	25
	Appliances electricity – test dataset	11
	Heating – training dataset	25
	Heating – test dataset	11
	Cooling – training dataset	25
	Cooling – test dataset	11
Case 2 – monthly data	Period 1 – training dataset	9
	Period 1 – test dataset	4
	Period 2 – training dataset	4
	Period 2 – test dataset	3
	Period 3 – training dataset	8
	Period 3 – test dataset	4
Case 2 – daily data	Period 1 – training dataset	270
	Period 1 – test dataset	116
	Period 2 – training dataset	85
	Period 2 – test dataset	37
	Period 3 – training dataset	156
	Period 3 – test dataset	67

Data availability

The link to the open-source code is reported at the end of the paper.

References

[1] Haberl JS, Culp CC. Measurement and verification of energy savings. Energy Management Handbook: River Publishers; 2020. p. 733–95.
[2] Kromer S. The role of the measurement and verification professional: Judgment and decision-making in the application of M&V. River Publishers; 2024.
[3] Granderson J, Fernandes S. The state of advanced measurement and verification technology and industry application. The Electricity Journal 2017;30:8–16. https://doi.org/10.1016/j.tej.2017.08.005.
[4] Franconi E, Gee M, Goldberg M, Granderson J, Guiterman T, Li M, et al. The status and promise of advanced M&V: An overview of "M&V 2.0" methods, tools, and applications. 2017.
[5] de Wilde P. Building performance simulation in the brave new world of artificial intelligence and digital twins: a systematic review. Energ Buildings 2023;292:113171. https://doi.org/10.1016/j.enbuild.2023.113171.
[6] Abdelrahman M, Macatulad E, Lei B, Quintana M, Miller C, Biljecki F. What is a digital twin anyway? Deriving the definition for the built environment from over 15,000 scientific publications. Build Environ 2025;274:112748. https://doi.org/10.1016/j.buildenv.2025.112748.
[7] Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: fundamental principles and 10 grand challenges. Statistic Surveys 2022;16:1–85.
[8] Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. Nature reviews. Physics 2021;3:422–40. https://doi.org/10.1038/s42254-021-00314-5.
[9] ASHRAE guideline 14–2023: Measurement of energy, demand, and water savings; American Society of Heating, Refrigerating and air-Conditioning Engineers: Atlanta, GA, USA. 2023.
[10] Kissock JK, Haberl JS, Claridge DE. Development of a toolkit for calculating linear, change-point linear and multiple-linear inverse building energy analysis models, ashrae research project 1050-rp, final report. 2002.
[11] Kissock JK, Haberl JS, Claridge DE. Inverse modeling toolkit: numerical algorithms. ASHRAE Trans 2003;109:425.
[12] Haberl JS, Sreshthaputra A, Claridge DE, Kissock JK. Inverse model toolkit: application and testing. ASHRAE Trans 2003;109:435.
[13] Sikes T, Aligholian A, Chulock J, Scheer A. OpenEEmeter 4.0: Final Model Specification and Results. Linux Foundation (LF) Energy 2023;1:1–33.
[14] LF Energy OpenEEmeter 4.0 Release (https://lfenergy.org/the-openeemeter-working-group-is-pleased-to-release-lf-energy-openeemeter-4-0/)
[15] OpenEEmeter 4.0 (https://github.com/openeemeter/eemeter)

[16] Grillone B, Danov S, Sumper A, Cipriano J, Mor G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renew Sustain Energy Rev 2020;131:110027. https://doi.org/10.1016/j.rser.2020.110027.

[17] Alrobaie A, Krarti M. A review of data-driven approaches for measurement and verification analysis of building energy retrofits. Energies (Basel) 2022:15. https://doi.org/10.3390/en15217824.

[18] Kim H, Haberl JS. Field-test of the new ASHRAE/CIBSE/USGBC performance measurement protocols for commercial buildings: basic level. ASHRAE Trans 2012; 118:135.

[19] Kim H, Haberl J. Field-test of the ASHRAE/CIBSE/USGBC performance measurement protocols: part I intermediate level energy protocols. Sci Technol Built Environ 2018;24:281–97. https://doi.org/10.1080/23744731.2017.1368836.

[20] Kim H, Haberl J. Field-test of the ASHRAE/CIBSE/USGBC performance measurement protocols: part II advanced level energy protocols. Sci Technol Built Environ 2018;24:298–315. https://doi.org/10.1080/23744731.2017.1368837.

[21] EVO. IPMVP New Construction Subcommittee. International Performance Measurement & Verification Protocol: Concepts and option for determining energy Savings in New Construction, Volume III; Efficiency Valuation Organization (EVO): Washington, DC, USA, 2003.

[22] Afroz Z, Burak Gunay H, O'Brien W, Newsham G, Wilton I. An inquiry into the capabilities of baseline building energy modelling approaches to estimate energy savings. Energ Buildings 2021;244:111054. https://doi.org/10.1016/j.enbuild.2021.111054.

[23] Chen Z, Xiao F, Guo F, Yan J. Interpretable machine learning for building energy management: a state-of-the-art review. Advances in Applied Energy 2023;9: 100123. https://doi.org/10.1016/j.adapen.2023.100123.

[24] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15. https://doi.org/10.1038/s42256-019-0048-x.

[25] Rudin C, Radin J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. Harv Data Sci Rev 2019:1.

[26] Bahmani B, Suh HS, Sun W. Discovering interpretable elastoplasticity models via the neural polynomial method enabled symbolic regressions. Comput Methods Appl Mech Eng 2024;422:116827. https://doi.org/10.1016/j.cma.2024.116827.

[27] Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, et al. Kan: Kolmogorov-arnold networks. ArXiv Preprint ArXiv:240419756. 2024.

[28] Panczyk NR, Erdem OF, Radaideh MI. Opening the Black-Box: Symbolic Regression with Kolmogorov-Arnold Networks for Energy Applications. ArXiv Preprint ArXiv: 250403913. 2025.

[29] Grillone B, Mor G, Danov S, Cipriano J, Sumper A. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. Appl Energy 2021;301:117502. https://doi.org/10.1016/j.apenergy.2021.117502.

[30] Alrobaie AS, Krarti M. Measurement and verification building energy prediction (MVBEP): an interpretable data-driven model development and analysis framework. Energ Buildings 2023;295:113321. https://doi.org/10.1016/j.enbuild.2023.113321.

[31] Liguori A, Quintana M, Fu C, Miller C, Frisch J, van Treeck C. Opening the black box: towards inherently interpretable energy data imputation models using building physics insight. Energ Buildings 2024;310:114071. https://doi.org/10.1016/j.enbuild.2024.114071.

[32] Pfenninger S, Hirth L, Schlecht I, Schmid E, Wiese F, Brown T, et al. Opening the black box of energy modelling: strategies and lessons learned. Energ Strat Rev 2018;19:63–71. https://doi.org/10.1016/j.esr.2017.12.002.

[33] Pfenninger S. Open code and data are not enough: understandability as design goal for energy system models. Progress in Energy 2024;6:33002. https://doi.org/10.1088/2516-1083/ad371e.

[34] Meng Q, Xi Y, Zhang X, Mourshed M, Hui Y. Evaluating multiple parameters dependency of base temperature for heating degree-days in building energy prediction. Build Simul 2021;14:969–85. https://doi.org/10.1007/s12273-020-0752-9.

[35] Park S, Shim J, Song D. Issues in calculation of balance-point temperatures for heating degree-days for the development of building-energy policy. Renew Sustain Energy Rev 2021;135:110211. https://doi.org/10.1016/j.rser.2020.110211.

[36] Hong S-M, Paterson G, Burman E, Steadman P, Mumovic D. A comparative study of benchmarking approaches for non-domestic buildings: part 1 – top-down approach. Int J Sustain Built Environ 2013;2:119–30. https://doi.org/10.1016/j.ijsbe.2014.04.001.

[37] Burman E, Hong S-M, Paterson G, Kimpian J, Mumovic D. A comparative study of benchmarking approaches for non-domestic buildings: part 2 – bottom-up approach. Int J Sustain Built Environ 2014;3:247–61. https://doi.org/10.1016/j.ijsbe.2014.12.001.

[38] Borgato N, Bordignon S, Prataviera E, Garay-Martinez R, Zarrella A. Enhanced methodology for disaggregating space heating and domestic hot water heat loads of buildings in district heating networks. Appl Therm Eng 2025;263:125296. https://doi.org/10.1016/j.applthermaleng.2024.125296.

[39] Wang Y, Wang J, He W. Development of efficient, flexible and affordable heat pumps for supporting heat and power decarbonisation in the UK and beyond: review and perspectives. Renew Sustain Energy Rev 2022;154:111747. https://doi.org/10.1016/j.rser.2021.111747.

[40] Lazzarin R, Busato F, Noro M. Heat pumps in refurbishment of existing buildings. J Fed Eur Heat Vent Air Cond Assoc REHVA 2012;6:45–9.

[41] Busato F, Lazzarin RM, Noro M. Two years of recorded data for a multisource heat pump system: a performance analysis. Appl Therm Eng 2013;57:39–47. https://doi.org/10.1016/j.applthermaleng.2013.03.053.

[42] Yan D, Wu Y, Malik J, Hong T. Ten questions on future and extreme weather data for building simulation and analysis in a changing climate. Build Environ 2025; 269:112461. https://doi.org/10.1016/j.buildenv.2024.112461.

[43] Li Y, Wang W, Wang Y, Xin Y, He T, Zhao G. A review of studies involving the effects of climate change on the energy consumption for building heating and cooling. Int J Environ Res Public Health 2021:18. https://doi.org/10.3390/ijerph18010040.

[44] Meng Q, Mourshed M. Degree-day based non-domestic building energy analytics and modelling should use building and type specific base temperatures. Energ Buildings 2017;155:260–8. https://doi.org/10.1016/j.enbuild.2017.09.034.

[45] Meng Q, Xiong C, Mourshed M, Wu M, Ren X, Wang W, et al. Change-point multivariable quantile regression to explore effect of weather variables on building energy consumption and estimate base temperature range. Sustain Cities Soc 2020; 53:101900. https://doi.org/10.1016/j.scs.2019.101900.

[46] Spinoni J, Vogt JV, Barbosa P, Dosio A, McCormick N, Bigano A, et al. Changes of heating and cooling degree-days in Europe from 1981 to 2100. Int J Climatol 2018; 38:e191–208. https://doi.org/10.1002/joc.5362.

[47] Pernigotto G, Gasparella A. Quasi-steady state and dynamic simulation approaches for the calculation of building energy needs: Part 1 thermal losses. Building Simulation Applications BSA 2013. In: 1st IBPSA Italy Conference, Bozen-Bolzano, 30th January-1st February 2013; 2013. p. 293–302.

[48] Pernigotto G, Gasparella A. Quasi-steady state and dynamic simulation approaches for the calculation of building energy needs: Part 2 thermal gains. In: Building Simulation Applications BSA 2013: 1st IBPSA Italy Conference, Bozen-Bolzano, 30th January-1st February 2013; 2013. p. 303–12.

[49] Prades-Gil C, Viana-Fons JD, Masip X, Cazorla-Marín A, Gómez-Navarro T. An agile heating and cooling energy demand model for residential buildings. Case study in a mediterranean city residential sector. Renew Sustain Energy Rev 2023;175: 113166. https://doi.org/10.1016/j.rser.2023.113166.

[50] Vesterberg J, Andersson S, Olofsson T. Robustness of a regression approach, aimed for calibration of whole building energy simulation tools. Energ Buildings 2014;81: 430–4. https://doi.org/10.1016/j.enbuild.2014.06.035.

[51] Vesterberg J, Andersson S, Olofsson T. A single-variate building energy signature approach for periods with substantial solar gain. Energ Buildings 2016;122: 185–91. https://doi.org/10.1016/j.enbuild.2016.04.040.

[52] Vesterberg J, Andersson S, Olofsson T. Calibration of low-rise multifamily residential simulation models using regressed estimations of transmission losses. J Build Perform Simul 2016;9:304–15. https://doi.org/10.1080/19401493.2015.1067257.

[53] Tronchin L, Manfren M, Nastasi B. Energy analytics for supporting built environment decarbonisation. Energy Procedia 2019;157:1486–93. https://doi.org/10.1016/j.egypro.2018.11.313.

[54] ISO 16346:2013, Energy performance of buildings — Assessment of overall energy performance. 2013.

[55] Pistore L, Pernigotto G, Cappelletti F, Gasparella A, Romagnoni P. A stepwise approach integrating feature selection, regression techniques and cluster analysis to identify primary retrofit interventions on large stocks of buildings. Sustain Cities Soc 2019;47:101438. https://doi.org/10.1016/j.scs.2019.101438.

[56] Westermann P, Deb C, Schlueter A, Evins R. Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. Appl Energy 2020;264:114715. https://doi.org/10.1016/j.apenergy.2020.114715.

[57] Pasichnyi O, Wallin J, Kordas O. Data-driven building archetypes for urban building energy modelling. Energy 2019;181:360–77. https://doi.org/10.1016/j.energy.2019.04.197.

[58] Woo J, Fatima R, Kibert CJ, Newman RE, Tian Y, Srinivasan RS. Applying blockchain technology for building energy performance measurement, reporting, and verification (MRV) and the carbon credit market: a review of the literature. Build Environ 2021;205:108199. https://doi.org/10.1016/j.buildenv.2021.108199.

[59] Paulus MT, Claridge DE, Culp C. Algorithm for automating the selection of a temperature dependent change point model. Energ Buildings 2015;87:95–104. https://doi.org/10.1016/j.enbuild.2014.11.033.

[60] Paulus MT. Algorithm for explicit solution to the three parameter linear change-point regression model. Sci Technol Built Environ 2017;23:1026–35.

[61] Manfren M, Sibilla M, Tronchin L. Energy modelling and analytics in the built environment—a review of their role for energy transitions in the Construction sector. Energies 2021:14. https://doi.org/10.3390/en14030679.

[62] Abushakra B, Paulus MT. An hourly hybrid multi-variate change-point inverse model using short-term monitored data for annual prediction of building energy performance, part III: results and analysis (1404-RP). Sci Technol Built Environ 2016;22:984–95. https://doi.org/10.1080/23744731.2016.1215659.

[63] Abushakra B, Reddy A, Vipul Singh. ASHRAE research project report 1404-RP, measurement, modeling, analysis and reporting protocols for Short-term M&V of whole building energy performance, Arizona State University, USA. 2012.

[64] Jalori S, Reddy TA. A unified inverse modeling framework for whole-building energy interval data: daily and hourly baseline modeling and short-term load forecasting. ASHRAE Trans 2015;121:156.

[65] Staffell I, Pfenninger S, Johnson N. A global model of hourly space heating and cooling demand at multiple spatial scales. Nat Energy 2023. https://doi.org/10.1038/s41560-023-01341-5.

[66] Kheiri F, Haberl JS, Baltazar J-C. Split-degree day method: a novel degree day method for improving building energy performance estimation. Energ Buildings 2023;289:113034. https://doi.org/10.1016/j.enbuild.2023.113034.

[67] Price P. Methods for analyzing electric load shape and its variability. Lawrence Berkeley National Laboratory Report LBNL-3713E; 2010.
[68] Mathieu JL, Price PN, Kiliccote S, Piette MA. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2011; 2:507–18.
[69] Kazmi H, Fu C, Miller C. Ten questions concerning data-driven modelling and forecasting of operational energy demand at building and urban scale. Build Environ 2023;239:110407. https://doi.org/10.1016/j.buildenv.2023.110407.
[70] Harang I, Heymann F, Stoop LP. Incorporating climate change effects into the European power system adequacy assessment using a post-processing method. Sustainable Energy, Grids and Networks 2020;24:100403. https://doi.org/10.1016/j.segan.2020.100403.
[71] Manfren M, Nastasi B. Interpretable data-driven building load profiles modelling for measurement and verification 2.0. Energy 2023:128490. https://doi.org/10.1016/j.energy.2023.128490.
[72] Manfren M, James PAB, Aragon V, Tronchin L. Lean and interpretable digital twins for building energy monitoring – a case study with smart thermostatic radiator valves and gas absorption heat pumps. Energy AI 2023;14:100304. https://doi.org/10.1016/j.egyai.2023.100304.
[73] RMV2.0 - LBNL M&V2.0 Tool (https://lbnl-eta.github.io/RMV2.0/)
[74] NMECR (https://kw-labs.github.io/nmecr/)
[75] CalTRACK. CalTRACK Methods (http://docs.caltrack.org/en/latest/methods.html)
[76] Granderson J, Fernandes S, Crowe E, Sharma M, Jump D, Johnson D. Accuracy of hourly energy predictions for demand flexibility applications. Energ Buildings 2023;295:113297. https://doi.org/10.1016/j.enbuild.2023.113297.
[77] Mirfin A, Xiao X, Jack MW. TOWST: a physics-informed statistical model for building energy consumption with solar gain. Appl Energy 2024;369:123488. https://doi.org/10.1016/j.apenergy.2024.123488.
[78] Lopez-Villamor I, Eguiarte O, Arregi B, Garay-Martinez R, Garrido-Marijuan A. Time of the week AutoRegressive eXogenous (TOW-ARX) model to predict thermal consumption in a large commercial mall. Energy Conversion and Management: X 2024;24:100777. https://doi.org/10.1016/j.ecmx.2024.100777.
[79] ECAM 7.0 (https://sbwconsulting.com/ecam/).
[80] Arowoiya VA, Moehler RC, Fang Y. Digital twin technology for thermal comfort and energy efficiency in buildings: a state-of-the-art and future directions. Energy and Built Environment 2024;5:641–56. https://doi.org/10.1016/j.enbenv.2023.05.004.
[81] Cespedes-Cubides AS, Jradi M. A review of building digital twins to improve energy efficiency in the building operational stage. Energy Informatics 2024;7:11. https://doi.org/10.1186/s42162-024-00313-7.
[82] Boukaf M, Fadli F, Meskin N. A comprehensive review of digital twin technology in building energy consumption forecasting. IEEE Access 2024;12:187778–99. https://doi.org/10.1109/ACCESS.2024.3498107.
[83] Hwang J, Kim J, Yoon S. DT-BEMS: digital twin-enabled building energy management system for information fusion and energy efficiency. Energy 2025; 326:136162. https://doi.org/10.1016/j.energy.2025.136162.
[84] Ma Z, Jiang G, Hu Y, Chen J. A review of physics-informed machine learning for building energy modeling. Appl Energy 2025;381:125169. https://doi.org/10.1016/j.apenergy.2024.125169.
[85] Von Krannichfeldt L, Orehounig K, Fink O. Combining physics-based and data-driven modeling for building energy systems. Appl Energy 2025;391:125853. https://doi.org/10.1016/j.apenergy.2025.125853.
[86] Li Y, O'Neill Z, Zhang L, Chen J, Im P, DeGraw J. Grey-box modeling and application for building energy simulations - a critical review. Renew Sustain Energy Rev 2021;146:111174. https://doi.org/10.1016/j.rser.2021.111174.
[87] Ferrari S, Zanotto V, others. Building energy performance assessment in southern Europe. Springer; 2016.
[88] Baasch G, Westermann P, Evins R. Identifying whole-building heat loss coefficient from heterogeneous sensor data: an empirical survey of gray and black box approaches. Energ Buildings 2021;241:110889. https://doi.org/10.1016/j.enbuild.2021.110889.
[89] ISO. ISO 52016-1:2017(en) Energy performance of buildings — Energy needs for heating and cooling, internal temperatures and sensible and latent heat loads — Part 1: Calculation procedures. 2017.
[90] Lundström L, Akander J, Zambrano J. Development of a space heating model suitable for the automated model generation of existing multifamily buildings—a case study in Nordic climate. Energies 2019:12. https://doi.org/10.3390/en12030485.
[91] Lecocq EM, Pascual J, Salom J. Development of physically coherent grey-box models for residential buildings using a simplified adjustment method. Energ Buildings 2025;328:115215. https://doi.org/10.1016/j.enbuild.2024.115215.
[92] Chen X, Lv G, Zhuang X, Duarte Roa C, Schiavon S, Geyer P. Integrating symbolic neural networks with building physics: a study and proposal. Journal of Building Engineering 2025;111:113033. https://doi.org/10.1016/j.jobe.2025.113033.
[93] Sarmas E, Forouli A, Marinakis V, Doukas H. Baseline energy modeling for improved measurement and verification through the use of ensemble artificial intelligence models. Inf Sci (N Y) 2024;654:119879. https://doi.org/10.1016/j.ins.2023.119879.
[94] Ali U, Bano S, Shamsi MH, Sood D, Hoare C, Zuo W, et al. Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. Energ Buildings 2024;303:113768. https://doi.org/10.1016/j.enbuild.2023.113768.
[95] Manfren M, James PAB, Tronchin L. Data-driven building energy modelling – an analysis of the potential for generalisation through interpretable machine learning. Renew Sustain Energy Rev 2022;167:112686. https://doi.org/10.1016/j.rser.2022.112686.