ORIGINAL RESEARCH **Open access**

Regression-Based Digital Twins for Multi-Temporal Building Energy Analytics and Performance Verification

Kristian Olsen[1*], Ingrid Berg[1], Magnus Dahl[2], Sigrid Haugen[3]

Abstract

Shifting toward low-carbon energy frameworks introduces significant energy efficiency challenges. Within building renovation initiatives, advanced solutions like intelligent control systems and heat pumps are progressively replacing traditional equipment to decrease carbon footprints and realize operational energy and financial savings, alongside secondary advantages. Nonetheless, empirical evaluations of building performance frequently expose a substantial discrepancy between anticipated energy consumption during the design phase and the actual usage during operation. Consequently, streamlined and comprehensible digital twins are essential for monitoring building energy, ensuring the longevity of savings, and fostering ongoing performance enhancements. In the present study, easily interpretable regression models are developed using multi-scale temporal data (monthly, daily, and hourly). These models are cohesively combined to validate the efficiency gains achieved through smart thermostatic radiator valves (TRVs) and gas absorption heat pumps (GAHPs), while also offering a deeper understanding of the overall building functionality. Additionally, as a component of the analytical investigation, the time of week and temperature (TOWT) methodology is modified and evaluated against its initial framework. The selected case study focuses on the Hale Court sheltered housing complex in Portsmouth (UK). This facility served as a real-world testing ground for novel systems like TRVs and GAHPs under the European Union Horizon 2020 THERMOSS project. The findings are leveraged to demonstrate potential broader applications for energy signature modeling, emphasizing the consequences for both energy administration and the advancement of cutting-edge building technologies.

Keywords Energy management, Energy signature, Data-driven methods, Digital twins, Thermostatic radiator valves, Gas absorption heat pumps

*Correspondence:
Kristian Olsen
kristian.olsen@ntnu.no
[1] Department of Artificial Intelligence for Hydroelectric Systems, Faculty of Engineering, Norwegian University of Science and Technology, Trondheim, Norway
[2] Department of Smart Energy Networks and Machine Learning, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway
[3] Department of Intelligent Power Systems, Faculty of Engineering, University of Bergen, Bergen, Norway

Introduction

Shifting towards low-carbon energy infrastructures presents notable difficulties regarding energy efficiency. While electrifying end-use applications is broadly acknowledged as a critical step toward total decarbonization [1], the resulting displacement of environmental impacts [2] must not be overlooked. Concurrently, using hydrogen to replace natural gas and other fossil fuels to meet decarbonization targets introduces multiple obstacles, with financial costs representing one of the primary hurdles [3].

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Broadly speaking, assessing the effectiveness of energy efficiency interventions requires robust methodologies, particularly given the substantial capital investments involved [4, 5]. Furthermore, such evaluations can serve as a foundational step for subsequent, highly detailed analyses that rely on precisely calibrated simulations [6].

This investigation incorporates cutting-edge heating technologies, specifically highlighting gas absorption heat pumps (GAHPs) as viable substitutes for traditional and condensing boilers. A GAHP operates as a thermal apparatus that relies on the combustion of natural gas to power an absorption cycle. During this process, thermal energy is extracted from the external environment (the heat source) and transferred to the water circulating within the heating network (the heat sink), operating at a superior efficiency level than standard condensing boilers. While GAHPs deliver functional benefits comparable to air source heat pumps (ASHPs), they are powered by natural gas rather than electrical energy. Consequently, GAHP systems present a practical alternative for upgrading condensing boilers. However, this substitution may induce a long-term technological lock-in effect, given its continued reliance on fossil fuels and operation at elevated heating temperatures.

Operationally, the efficiency of a GAHP is dictated by the temperatures of both the heat source (specifically, the external air temperature) and the heat sink (the water temperature within the heating circuit). Although standardized testing protocols exist [7] for such equipment, empirical field evaluations are required to assess any decline in performance—relative to baseline data gathered under controlled conditions—caused by operating under partial-load scenarios. Contemporary research has specifically focused on evaluating the partial load efficiency of GAHPs [8] and contrasting GAHP systems [9] against alternative heating solutions. Findings from these previous investigations are evaluated against the outcomes derived from our specific case study. Furthermore, intelligent heating management devices, such as smart thermostatic radiator valves (TRVs), are frequently highlighted as highly prospective technologies. Previous literature suggests they possess considerable potential [10, 11]; however, this potential requires rigorous verification as it is contingent upon various elements, notably operational habits and occupant behavior. Therefore, grasping real-world operational tactics is essential for bridging the current knowledge deficit [12] concerning actual building efficiency.

Pinpointing both functional and design-related complications through digital counterparts of physical items, procedures, or services is an active area of investigation across numerous domains. A wide array of definitions for “digital twins” can be found throughout the literature [13], frequently blurring the distinction between a standard model and a true digital twin [14]. Moreover, this subject is garnering significant interest within the construction and energy industries [15], driven by the escalating integration of Artificial Intelligence (AI) and Machine Learning (ML) instruments. Nevertheless, although advanced ML methodologies like deep learning have proven highly successful across various use cases [16], they predominantly suffer from a lack of interpretability [17], remaining a persistent obstacle today [18]. As demonstrated by Chen *et al*. [19], complex machine learning algorithms deployed for building energy supervision generally rely on post-hoc interpretation methods, such as LIME and SHAP [20], to explain their outputs. However, straightforward and transparent examination of the underlying algorithmic reasoning remains highly problematic. This limitation creates significant challenges across various dimensions, encompassing AI and ML ethical considerations [21] as well as the confidence of industry professionals within the energy field.

The focal point of this investigation is the Hale Court sheltered housing complex, managed by the Portsmouth City Council. This facility underwent refurbishment under the EU Horizon 2020 THERMOSS initiative [22]. The Hale Court property consists of 80 apartments separated into a North and a South block, with each block operating its own distinct thermal network. This evaluation focuses exclusively on the North block, tracking its performance both before and after the renovation efforts. The upgrade was executed in two distinct phases: initially, smart TRVs were implemented, followed subsequently by the integration of GAHPs. The recorded data is categorized into three distinct timeframes: the pre-renovation phase, the post-TRV installation phase, and the period following the GAHP integration and mechanical room modernization. The operational data gathered before any upgrades serve as the foundational baseline for assessing the energy conservation realized by the newly introduced systems. A series of models were progressively developed within a unified analytical pipeline. Within this framework, innovative methodologies are evaluated using monthly, daily, and hourly data to ensure scalability across timeframes. Simultaneously, these methods preserve the inherent (ante-hoc [20])

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

transparency characteristic of regression techniques, thereby supplying valuable perspectives on the operational success of the newly implemented equipment and the holistic efficiency of the structure.

Nomenclature

Table 1 shows the nomenclature.

Table 1. Nomenclature.

Symbol	Description	Unit
a_0	Intercept term of the heating-component regression model	kW
a_1	Temperature-dependent coefficient in the heating-component regression model	kW/K
a_2,a_3	Coefficients representing solar-radiation influence and differenced-temperature dependence in the heating component	m², kW/K
a_j	Regression coefficient corresponding to hour *j* of the week	kW
b_k	Regression coefficient associated with temperature segment *k*	kW/K
c_o	Intercept coefficient of the TOWT model	kW
d_0,d_1,d_2	Regression parameters of the Qth model	kW, kW/K, kW/K
e_0,e_1,e_2	Regression parameters of the GUEmax model	–, 1/K, 1/K
Cv(RMSE)	Coefficient of variation of the root mean square error	–
GUE	Gas utilization efficiency	–
GUE_{max}	Peak gas utilization efficiency	–
gue	Partial-load fraction of GUE	–
I_{sol}	Monthly averaged hourly total solar irradiance on a horizontal plane (direct and diffuse components)	kW/m²
L_{occ}	Load demand during occupied building periods	kW
L_{unocc}	Load demand during unoccupied building periods	kW
MAPE	Mean absolute percentage error	–
NMBE	Normalized mean bias error (reported as a percentage)	–
Q_{NG}	Energy supplied by natural gas	kWh
Q_{th}	Thermal energy output	kWh
q_h	Heating energy signature based on natural gas consumption	kW
R^2	Coefficient of determination (expressed as a percentage)	–
$T(i)_k$	Outdoor air temperature corresponding to segment *k* at time interval *i* in the TOWT model	–

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

$t_{ow,j}$	Binary time-of-week variable in the TOWT model	–
SE	Standard error	kW, –
X_h	Binary heating indicator variable (0–1)	–
θ_e	Outdoor ambient air temperature	°C
$\theta_{e,diff}$	Differenced outdoor air temperature	°C
$\theta_{e,sup}$	Supply temperature of the water-heating system	°C
ε	Error component	kW

Literature Review

Current and future perspectives on data-driven approaches for energy monitoring and management need to resolve two primary challenges: delivering a dependable evaluation of the true efficiency impact of various technologies, and facilitating enduring energy savings alongside the ongoing refinement of management strategies. This aligns with the concepts introduced earlier concerning digital twin applications [13, 15].

Consequently, the foundation of this study relies on regression-oriented methodologies outlined in cutting-edge Measurement and Verification (M&V) frameworks. These include standards like ASHRAE 14:2014 [23], the Efficiency Valuation Organization (EVO) [24], and the Federal Energy Management Program (FEMP) [25]. The advancement of these frameworks is commonly referred to as Advanced M&V or M&V 2.0 [26].

This progression incorporates the application of complex algorithms, higher-resolution data (such as readings from smart meters), enhanced precision, and more rapid feedback loops. Within this context, the concept of “digital twins” is applied due to their capacity to dynamically simulate real-world behavior and supply actionable analytics for pinpointing and resolving operational problems.

In the current investigation, the regression frameworks suggested by ASHRAE 14:2014 [23] and ISO 50,006:2014 [27] serve as a foundational baseline. These represent the most established methodologies, featuring typical recurring energy consumption profiles suitable for diverse building categories and climatic conditions. Establishing this baseline allows for a robust comparison of efficiency interventions monitored over varied timeframes and evaluated via distinct modeling approaches. These models were selected based on specific traits detailed later in this section. Additionally, every model introduced in this study utilizes energy signatures in accordance with ISO 16,346:2013 [28]. Specifically, they estimate the average power by dividing the total energy by the number of operational hours within the given analytical period—whether monthly, daily, or hourly. This methodology was adopted to ensure the models possess a required level of temporal scalability. Such scalability is also beneficial for visual analytics, enabling a coherent comparison between long-term (low-frequency) and short-term (high-frequency) energy measures. Indeed, regression methodologies are extensively acknowledged within the domain of building energy management, spanning applications from fundamental and intermediate tiers [29] to highly sophisticated ones [30].

Primarily, M&V frameworks applied to building energy oversight depend on interval energy data, which acts as the dependent variable, and meteorological data, serving as independent variables. These are often supplemented by other independent factors, such as dummy variables used to represent distinct occupancy patterns and operational states derived from contextual data. For weather-normalizing energy usage, outdoor air temperature is the most critical independent variable [23, 31, 32]. Furthermore, professionals in the field frequently favor univariate regression models [33] for statistical energy modeling at the whole-building level. Although this methodology might yield models with lower

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

accuracy than those produced by more complex algorithms, it remains highly valuable because of its straightforward interpretability [34].

The utility of regression-driven methods extends beyond the energy monitoring of entire buildings; it can be broadened by incorporating supplementary variables beyond merely the outdoor air temperature required for weather normalization. For instance, in regression analyses applied to co-heating tests [35, 36]—which assess the performance of a building's thermal envelope—horizontal solar radiation is integrated to factor in solar heat gains within the thermal energy equilibrium of specific building zones. These tests typically span an average of 15 days [37]. Moreover, calculating the difference in outdoor air temperatures across a daily time series (i.e., finding the variation between consecutive days) streamlines the assessment of the building envelope's thermal inertia [38, 39]. Concurrently, the supply temperature of the heating system is analyzed alongside the outdoor air temperature to evaluate the efficiency of air source heat pumps (ASHPs) for optimization and control purposes [40, 41].

Regarding regression methods that utilize hourly data intervals, the time of week and temperature (TOWT) algorithm was initially introduced by Price [42] to examine the profile and variability of electrical loads. Subsequently, it has been applied to measure shifts in electricity consumption resulting from demand response initiatives [43] and within large-scale utility efficiency schemes [44]. The TOWT approach has also undergone successful validation on the Advanced M&V testing platform established by the Efficiency Valuation Organization (EVO) [45]. This platform serves as an open, independent testing environment, utilizing electricity usage data from 367 facilities distributed across various North American regions. The TOWT framework is available as an R software package known as RMV2.0 [46], which concurrently incorporates a gradient boosting machine technique [47]. Nevertheless, the TOWT model remains notably more user-friendly, requiring the calibration of only a single hyper-parameter. Currently, multiple open-source software libraries deploy TOWT alongside other regression methodologies. Notable examples include the R package NMECR [48], the Python-based OpenEEmeter [49] (which integrates Caltrack procedures [50]), and the recently developed EENSIGHT [51]. Furthermore, comprehensive systematic reviews have explored the challenges associated with energy baseline formulation [52], statistical models for whole-building energy consumption [33], the deployment of physics-based and data-driven methods for measurement and verification in building renovations [53, 54], and broadly, the need for interpretable energy analytics within the built environment [55, 56].

Lastly, as demonstrated by Chen *et al*. [19], advanced machine learning techniques utilized in building energy management predominantly rely on post-hoc interpretation methods, such as LIME and SHAP [20], to clarify their outputs. In contrast, this research specifically opted for regression-based strategies because of their inherent (ante-hoc) transparency and interpretability.

Materials and Methods

The predictive model structures adopted in this study were originally conceived for various distinct applications. However, they exhibit numerous shared characteristics that allow them to be seamlessly combined into a cohesive workflow. This integrated process begins with baseline regression models that use the energy signature as the dependent variable and outdoor air temperature as the independent variable. These models are evaluated across various monitoring timeframes. Subsequently, these initial regression models are refined by incorporating supplementary variables to enhance their overall goodness of fit. The foundational models also serve to evaluate continuous model retraining over a rolling 15-day data horizon and to analyze the outdoor temperature sensitivity of the time-of-week and temperature (TOWT) algorithm. This algorithm is subsequently restructured and compared against its standard configuration. Additionally, a specific regression framework incorporating an extra variable is formulated to evaluate the operational efficiency of gas absorption heat pumps (GAHPs). Finally, the requisite standards for validating the models, based on established measurement and verification (M&V) protocols.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Regression-based approaches for M&V and their extensions with additional variables

During the preliminary phase, before model training, the dataset underwent processing to address missing entries via interpolation, as the empirical data exhibited only negligible gaps. The hourly records for outdoor air temperature and solar radiation were deemed sufficiently precise for these analyses. Conversely, the half-hourly natural gas consumption data required sanitization; a moving average filter was applied to reduce noise (i.e., eliminating minor fluctuations caused by measurement inaccuracies), after which the data was consolidated into hourly intervals. The foundational regression models are based on ASHRAE 14:2014 [23] guidelines, treating outdoor air temperature θe as the primary input parameter. These models are subsequently adapted to incorporate further variables.

These supplementary variables include the aggregate solar radiation measured on a horizontal plane, Isol, along with the differenced daily mean temperature, θe, diff (representing the day-to-day variance in outdoor air temperature), justified by the rationale. Moreover, to accurately represent piecewise linear trajectories, binary dummy variables (0–1), designated as Xh, function as interaction terms. These are multiplied by the core variables (θe, Isol, θe, diff) in accordance with the established conventions and methodologies detailed in prior literature [57].

These principles dictate that change-point temperatures must be selected from within plausible boundaries. Since these change-points correspond to the thermal balance-point temperatures of the facility [58, 59], they are governed by both the physical attributes of the building and its operational parameters, imposing physical limits on their values. Additionally, these rules require enforcing constraints on the model's output. Because energy signatures must remain strictly positive, applying a piecewise linear structure might inadvertently generate slight negative outputs when the signature approaches zero; such anomalies must be eliminated. The specific mathematical expressions governing the monthly and daily energy signature models are detailed in **Table 2.**

Table 2. Formulation of monthly and daily energy signature models tested.

Model name	Input variables	Interval	Formulation	Empty cell
θ_e	θ_e	Monthy/daily	$q_h = X_h(a_0+a_1\theta_e) + b_0 + \varepsilon$	(1)
$\theta_e I_{sol}$	θ_e, I_{sol}	Monthy/daily	$q_h = X_h (a_0 + a_1\theta_e + a_2 I_{sol}) + b_0 + \varepsilon$	(2)
$\theta_e \theta_{e,diff}$	$\theta_e, \theta_{e,diff}$	Daily	$q_h = X_h (a_0 + a_1\theta_e + a_2\theta_{e,diff}) + b_0 + \varepsilon$	(3)
$\theta_e I_{sol} \theta_{e,diff}$	$\theta_e, I_{sol}, \theta_{e,diff}$	Daily	$q_h = X_h (a_0+ a_1\theta_e + a_2 I_{sol} + a_3\theta_{e,diff}) + b_0 + \varepsilon$	(4)

To evaluate the potential uncertainty arising from the reliance on short-term data and to mitigate the risk of overfitting, baseline models undergo incremental retraining using a 15-day rolling horizon. Conceptually, this procedure resembles model cross-validation: the comprehensive dataset is partitioned into smaller 15-day subsets, and the resulting individual models are subsequently benchmarked against the primary model trained on the complete dataset.

Furthermore, supplementary dummy variables are integrated into the time-of-week and temperature (TOWT) framework. The authors have previously utilized TOWT to simulate electrical load profiles under conditions of high variability [60]. In the current study, however, an innovative formulation is introduced and evaluated against the implementation found in the RMV2.0 R package [46]. The fundamental structure of the TOWT model is relatively straightforward, consisting of two piecewise linear sub-models: one representing occupied periods, Locc (characterized by high demand), and another for unoccupied periods, Lunocc (characterized by low demand), as illustrated below:

$$L_{occ} = \sum_{j=1}^{n-1} b_k I_{occ,j} + \sum_{k=1}^{m} b_k \left(T(i)_k\right) + c_0 \quad (5)$$

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

$$L_{unocc} = \sum_{j=1}^{n-1} b_k t_{ow,j} + \sum_{k=1}^{m} b_k \left(T(i)_k\right) + c_0 \quad (6)$$

Nevertheless, owing to this architectural design, two distinct strategies are required to execute both the temporal (Time of Week) and temperature segmentations. The temporal division inherently relies on the temperature segmentation, given that occupied and unoccupied intervals are determined automatically by evaluating the electrical load in relation to the outdoor air temperature. As previously detailed, this investigation prioritizes the energy signature associated with natural gas usage, qh, rather than the electrical load.

Temperature segmentation is executed to maintain continuity in the model's calculated temperature response, employing the algorithm described in Mathieu *et al*. [43]; this approach "indirectly" enforces constraints to preserve continuity, even while applying a standard regression algorithm that inherently lacks such constraints. Furthermore, the temperature response serves as the intersection between the TOWT framework and the regression models outlined in **Table 2**, a relationship that will be graphically demonstrated through energy signatures.

Table 3 delineates the core attributes of the original RMV2.0 implementation alongside the novel reformulation advanced in this study, providing a summary of the modifications and additions incorporated at each phase of the modeling process.

Table 3. RMV2.0 implementation of TOWT compared to the reformulation proposed.

Workflow phase	Element	TOWT–RMV2.0	TOWT–Revised Formulation
Data processing	Temporal segmentation	The time-of-week indicator (tow,j) is represented as a binary (dummy) variable. The model employs n−1 hourly variables for a week (168−1 = 167), while the 168th hour is absorbed into the intercept coefficient c_0.	Holiday-related time-of-week variables are treated identically to Sundays.
	Temperature segmentation	The temperature variable $T(i)_k$ is continuous and can be expressed in either Celsius or Fahrenheit. Here, *m* denotes the number of temperature bins, corresponding to *m−1* breakpoints in the piecewise-linear temperature-response function. Default breakpoints are set at (40, 55, 65, 80, 90 °F / 4.4, 12.8, 18.3, 26.7, 32.2 °C). A minimum of 1 breakpoint is required, with at least 20 observations above the highest and below the lowest breakpoint. Segmentation follows the procedure described in Mathieu *et al*. [43]. Identical temperature intervals are applied to both occupied and unoccupied operating modes.	The variable $T(i)_k$ is generated using the same methodology as the original approach [43]. However, breakpoint locations are determined from the outdoor-temperature distribution, producing an even number of segments with the central breakpoint positioned at the median (50th percentile). Segment widths may differ above and below the median, depending on the data distribution. A minimum of 6 segments is recommended. The same segmentation scheme is used for occupied and unoccupied periods.
Model development	Identification of occupied/unoccupied	Occupancy status is determined through a regression model incorporating Heating Degree-Days (HDD), Cooling	Occupied and unoccupied periods are identified using a temperature-based regression model built on the

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

	periods (high/low demand)	Degree-Days (CDD), and an intercept term. Weekly periods that remain underpredicted for a specified proportion of observations are classified as occupied. The default threshold is 65% (equivalent to 0.65 in the model settings). HDD and CDD are calculated using base temperatures of 50°F and 65°F, respectively.	segmentation method described above. The default classification threshold remains 0.65, consistent with the original implementation.
	Overall model structure	The final model consists of a weighted aggregation of multiple regression models generated across several runs, with weighting influenced by hyperparameter selection.	The complete model combines a regression model for occupied conditions, another for unoccupied conditions, and a residual model that captures only Time-Of-Week (TOW) effects.
	Hyperparameters	A single hyperparameter is employed, representing the time scale (in days) of the weighting function used to merge the various regression models.	One hyperparameter is used, expressed in weeks or months, to define additional temporal segmentation required for modeling residuals as a TOW-dependent function without temperature effects.
Result visualization and analysis	Temperature response	Load-temperature relationships are displayed using scatter plots.	Energy signatures are plotted against outdoor air temperature over different temporal resolutions (monthly, daily, and hourly). These plots facilitate interpretation of energy behavior and comparison of the piecewise-linear temperature response derived from the TOWT reformulation against alternative regression-based methods.
	Time-series representation	Measured and predicted values are displayed together in a time-series plot.	No modifications.
	Weekly operational profiles	Weekly operating patterns are visualized through a two-dimensional heatmap.	No modifications.
	Residual analysis	Residuals are visualized over time and against outdoor air temperature.	No modifications.
	Measured versus predicted comparison	A scatter plot comparing measured and predicted values is used to identify deviations and model performance.	No modifications.

Alternative software versions of TOWT exist, such as the Python-based Caltrack/OpenEEmeter [50] and the NMECR R package [48], as previously noted. The NMECR package permits the arbitrary definition of temperature change-points and allows for temperature segments containing uniform quantities of data points (i.e., quantile-based segmentation). It

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

retains the RMV2.0 methodology of utilizing days as hyperparameters for weighting, whereas the Caltrack/OpenEEmeter framework utilizes a monthly weighting strategy.

In the final phase of the analytical workflow, the performance of the GAHP is evaluated utilizing the Gas Utilization Efficiency (GUE), the primary metric for its thermal conversion efficiency. GUE is determined by the ratio of the supplied thermal output to the consumed fuel energy—a concept closely related to the Coefficient of Performance for electrical heat pumps. The variable gue represents its Part Load Fraction (PLF), expressed as follows:

$$GUE = \frac{Q_{th}}{Q_{NG}} \quad (7)$$

$$gue = \frac{GUE}{GUE_{max}} \quad (8)$$

Where Qth and QNG represent the thermal output (kWh) and the natural gas energy input (kWh) during the analyzed timeframe, respectively. The term gue denotes the ratio of the empirical GUE observed under partial-load operation to the peak measurable value obtained under steady-state, standardized testing conditions. Both Qth and QNG can similarly be quantified as power (kW) by dividing the energy totals by the relevant time duration (in this instance, hours). The total energy derived from the gas is computed by multiplying the metered gas volume (m3) by its lower heating value (kWh/m3).

For this application, the objective is to formulate a straightforward methodology for forecasting Qth and GUE under fluctuating, real-world operational scenarios, yielding a practical model for energy management applications. Initially, the GAHP performance metrics (specifically Qth and GUE) provided in the manufacturer's technical specifications are interpolated in accordance with the methodology outlined in EN 15,316–4–2:2017 [61]. Manufacturer datasheets present outcomes derived from steady-state, full-load testing scenarios [7]; the discrepancy between the regression-derived GUEmax and the empirically recorded GUE is analyzed subsequently in the context of metered energy consumption. The independent variables for both multivariate regression models—one predicting Qth and the other GUEmax—are the outdoor air temperature, θe (°C), and the heating circuit's supply temperature, θsup (°C). The dependent variables are the thermal power output, Qth (kW), and the previously defined GUE. These models are structured mathematically as follows:

$$Q_{th} = d_0 + d_1\theta_e + d_2\theta_{sup} \quad (9)$$

$$GUE_{max} = e_0 + e_1\theta_e + e_2\theta_{sup} \quad (10)$$

Statistical criteria for model acceptability according to M&V principles

Contemporary Measurement and Verification (M&V) frameworks [23-25] establish acceptable performance thresholds for predictive models relying on statistical metrics, specifically the Normalized Mean Bias Error (NMBE) and the Coefficient of Variation of the Root Mean Square Error (CV(RMSE)). These thresholds are detailed in **Table 4** below. Because this study utilizes monthly, daily, and hourly datasets, the acceptability criteria for the daily intervals are posited to fall between the monthly and hourly benchmarks. This is calculated as a straightforward arithmetic mean of the two corresponding values specified in ASHRAE Guideline 14:2014 [23], following the approach suggested by Meng and Mourshed [62]. The coefficient of determination, R2 (ranging from 0 to 1, or 0 to 100%, indicating a robust correlation when exceeding 75%), is also evaluated, given its inclusion in standards like ISO 50,006:2014 [27].

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Nevertheless, its inherent limitations must be recognized. Specifically, a primary limitation is that R2 is fundamentally linked to the model's regression slope (for instance, its sensitivity to θe, Isol, and θe, diff in the models presented in **Table 2**). Steeper slopes generally yield inflated R2 values, even if the variance of the forecasted variable remains constant. Consequently, while R2 is included for comprehensive reporting, it is imperative to utilize supplementary statistical metrics when rigorously assessing model accuracy.

Table 4. Thresholds of acceptability for M&V models as calibrated with monthly, daily and hourly data.

Interval	Metric	ASHRAE Guidelines 14/ Meng and Mourshed [62]
Monthly	*NMBE*	±5
	Cv(RMSE)	15
Daily	*NMBE*	±7.5
	Cv(RMSE)	22.5
Hourly	*NMBE*	±10
	Cv(RMSE)	30

Case Study Description

As detailed in the introduction, the Hale Court complex is a sheltered housing facility constructed in 1984 and managed by the Portsmouth City Council. The property comprises 80 individual units, featuring a combination of studio apartments alongside one-, two-, and three-bedroom dwellings. Structurally, the development is divided into two primary sections: the North Block and the South Block. Key architectural details are summarized in **Table 5**, and the foundational floor plan is illustrated in **Figure 1.**

Table 5. Hale Court building blocks data.

Building characteristic	Description
Gross floor area	**North Block:** approximately **1300 m²; South Block:** approximately **2100 m²**
Residential units	**North Block: 32** flats; **South Block: 48** flats
Apartment categories	Studio, **1-bedroom**, **2-bedroom**, and **3-bedroom** units
Construction year	**1984**
Occupant profile	Sheltered accommodation designed for senior residents and elderly occupants

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

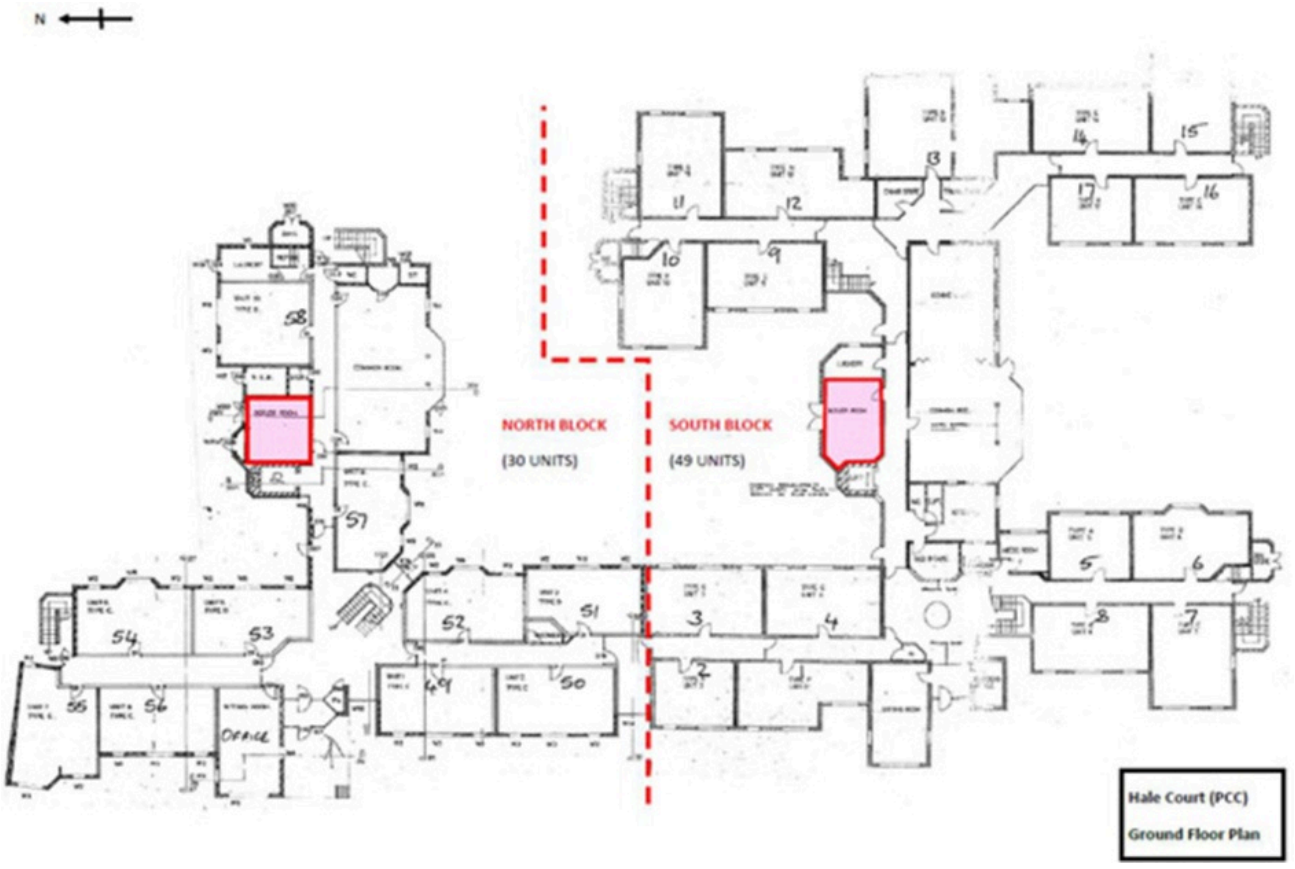


Figure 1. Hale Court – North Block/South Block - Floorplan Ground Floor.

The Hale Court facility underwent extensive renovations as part of the European Union Horizon 2020 THERMOSS initiative [22]. This retrofitting project primarily targeted two specific energy conservation strategies: the integration of Smart Thermostatic Radiator Valves (TRVs) and Gas Absorption Heat Pumps (GAHPs). These additions were executed concurrently with a comprehensive modernization of the mechanical plant room, which included the deployment of an advanced energy monitoring infrastructure. The chronological phasing of the renovation activities and the corresponding monitoring protocols are outlined in **Table 6.** The current investigation focuses exclusively on the North Block, enabling a precise evaluation of the energy reductions achieved through the combined use of Smart TRVs and GAHPs. Specifically, the thermal system upgrades within the North Block's mechanical room included the following actions:

- Retaining the pre-existing conventional gas boilers;
- Eliminating the legacy domestic hot water generation units;
- Installing a cascaded array of 3 GAHPs;
- Integrating a new thermal energy storage system;
- Restructuring the overall hydraulic piping network.

Table 6. Hale Court retrofit intervention periodization.

Monitoring stage	Intervention status	Building section(s)	Monitoring dates/Description

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

1	Pre-intervention period	North Block / South Block	December 2017 – December 2018
2	After the first intervention and before the second intervention	North Block / South Block	January 2019 – July 2019
	Smart Thermostatic Valves (STVs) installation	North Block / South Block	STVs were installed in January 2019. Energy consumption anomalies were observed during January and February 2019.
3	Following the second intervention (divided into commissioning and operational stages)	North Block	Phase 1: Plant-room refurbishment and GAHP deployment (July–September 2019)
		North Block	Phase 2: GAHP commissioning (September 2019)
		North Block	Phase 3: System operated using both GAHP and boilers (October 2019 – January 2020)
		North Block	Phase 4: Exclusive GAHP operation for testing purposes (February 2020)

The technical capacities of the newly implemented heat generation equipment within the plant room are documented in **Table 7.**

Table 7. Heat system generators data.

System technology	Parameter	Unit	Specification
Gas boilers	Equipment model	–	REMEHA GAS 110 Eco
	Rated thermal output	kW	115 × 2
	Maximum design operating temperature	°C	80
	Boiler classification	–	Condensing
Gas absorption heat pump (GAHP)	Equipment model	–	GHP AWO 38
	Thermal capacity at A7W50 conditions	kW	114.9 (cascade arrangement of **3** units)
	Gas Utilization Efficiency (A7W50)	–	1.52
	Natural gas input power at A7W50	kW	75.6
	Maximum heating-water supply temperature	°C	65
	Allowable ambient operating temperature range	°C	−20 to +40

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Regarding operational control, the system utilizes a fixed-point strategy with internal compensation (excluding external weather compensation). Consequently, during active operation, the supply water temperature within the primary circuit fluctuates between 60 °C and 65 °C. Maintaining this elevated temperature is essential because, as depicted in the schematic diagram in **Figure 2**, the primary loop is responsible for fulfilling both the space heating requirements and the domestic hot water needs of the residential units. Standard radiators serve as the terminal heat emission units within the dwellings. The inability to lower this supply temperature represents a significant barrier to maximizing the theoretical efficiency limits of both the overarching heating network and the specific GAHP equipment, a limitation that will be analyzed in greater depth.

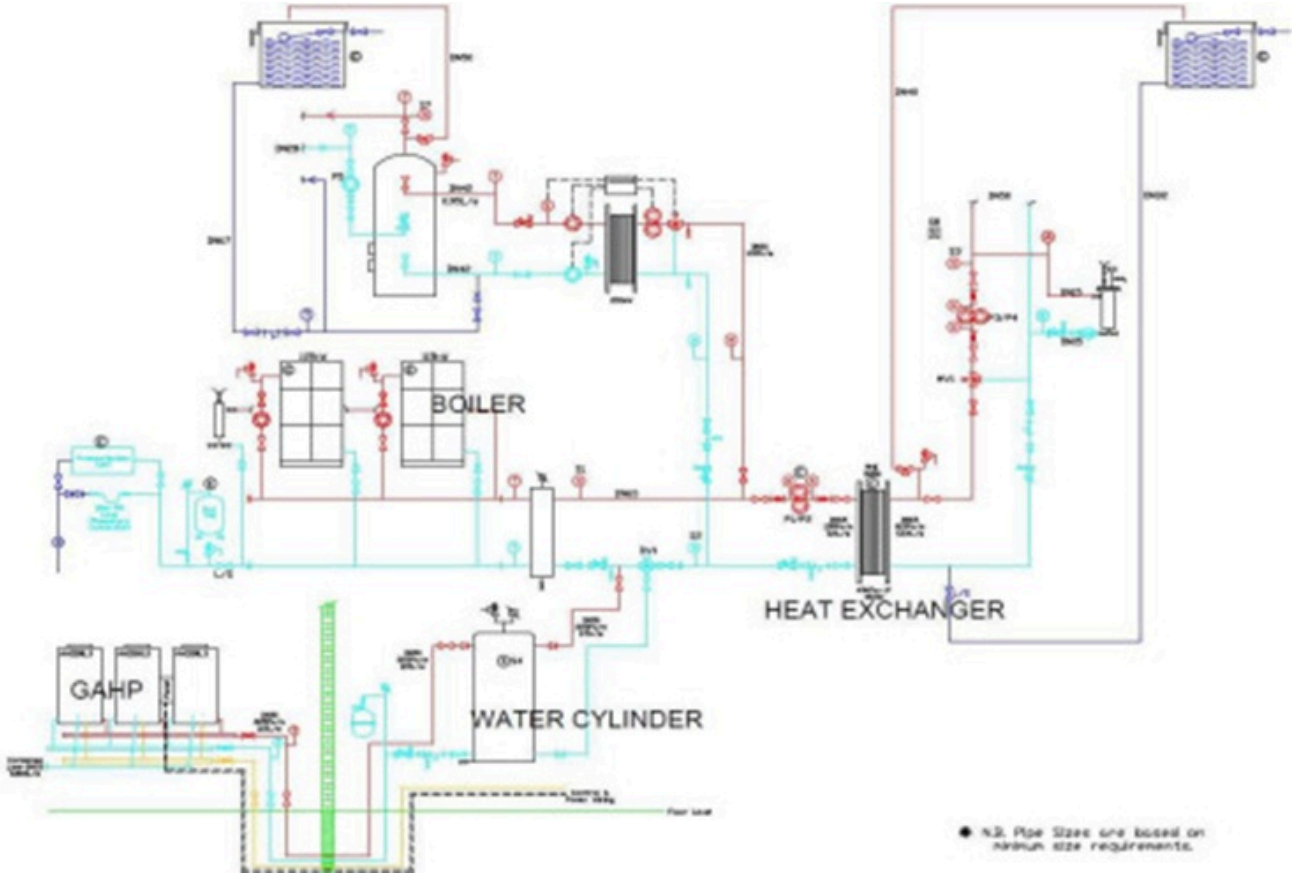


Figure 2. Schematic Layout of Heating System in Hale Court North.

Results and Discussion

Testing regression models θe, θeIsol, θeθe,diff, and θeIsolθe,diff

The preliminary stages of this study are outlined below, demonstrating how foundational models—initially detailed in ASHRAE 14:2014 [23] and subsequently adapted in the literature [57]—serve as a fundamental baseline for valid comparisons. Subsequently, extra independent variables are incorporated into these models, adhering to the rationale established.

The initial evaluation focuses on model θe **(Table 2)**, utilizing the natural gas demand signature as the dependent variable. This approach was chosen because, as clearly illustrated in **Figure 3**, electricity consumption remains largely unchanged both before and after the retrofit.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

This figure reveals that auxiliary systems do not exhibit temperature-dependent operation; their average annual functioning remains steady and thus warrants no further analysis here. Nevertheless, earlier studies have demonstrated the application of regression methodologies on substantially fluctuating electrical load profiles [60].

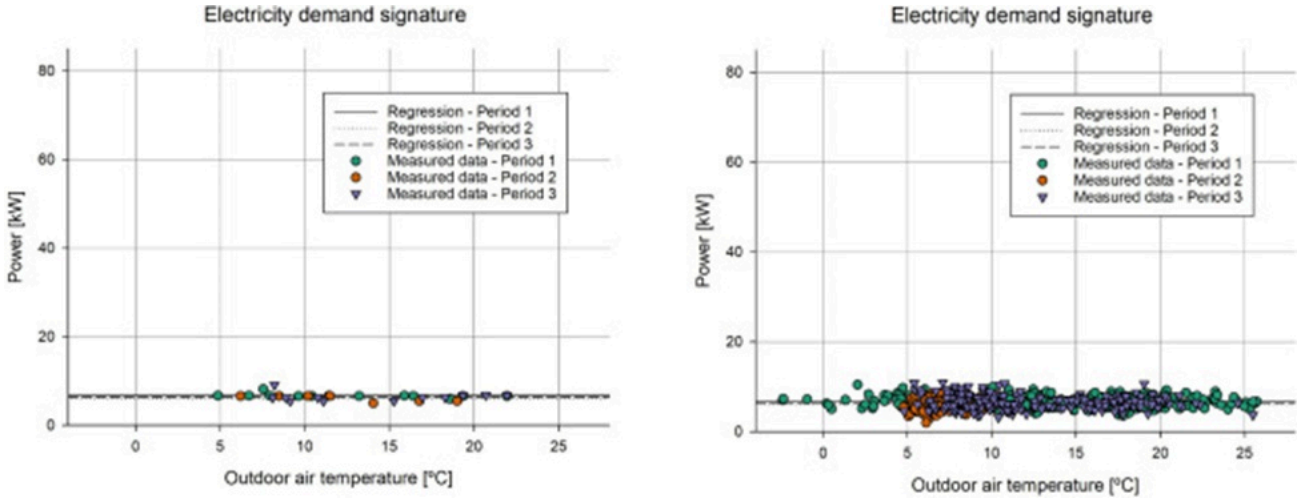


Figure 3. Electricity demand signature – monthly (left) and daily (right) data.

Conversely, examining the slope derived from the energy signature model fitted to the natural gas demand in **Figure 4** reveals that the installed heating system is significantly oversized for the building's true requirements, even when projections are extended to extreme cold temperatures.

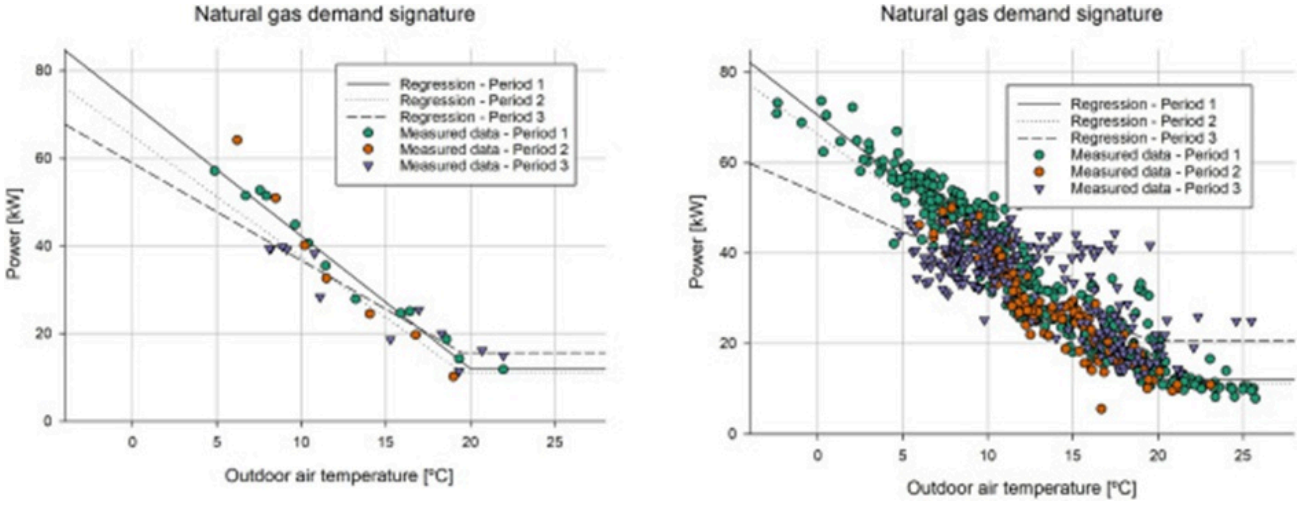


Figure 4. Natural gas demand signature – monthly (left) and daily (right) data.

Across the various monitoring phases, the piecewise linear regression model exhibits change-points between 19.5–20.0 °C. While these figures significantly exceed those typically observed in standard buildings [62], they are driven by the exceptionally high mean temperatures maintained in the bedrooms and living areas (22.9–23.5 °C). Notably, the base temperature used to normalize energy usage in Display Energy Certificates (DEC) is 15.5 °C, and technical standards for determining heat pump part-load ratios specify 16 °C [63]. Because both standard benchmarks deviate drastically from the empirical measurements in this study, the need to formulate customized regression models to capture true building performance is evident.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Above the change-point temperature, the flat segment of the natural gas signature reflects the baseline domestic hot water (DHW) requirement. This aligns with the summer season, which features zero space heating needs and exclusively DHW usage. **Table 8** presents the findings of the natural gas signature evaluation, demonstrating an overall strong model fit across all three periods when assessed against the acceptability thresholds for R2 and Mean Absolute Percentage Error (MAPE), detailed in **Table 4**. For daily data instances, the Normalized Mean Bias Error NMBE sits at 0. This outcome arises inherently from the mathematical formulation of regression techniques, which aim to minimize the sum of squared differences between observed values and model forecasts. Consequently, CV(RMSE) serves as the most pertinent statistical indicator here, particularly given the constraints of R2 previously discussed.

Table 8. Results of analysis for natural gas demand signature – Model θe.

Time interval	Period	CV(RMSE) (%)	NMBE (%)	MAPE (%)	R^2 (%)	Predicted energy (P)	Measured energy (M)
Monthly	1	6.35	−0.02	4.89	97.93	332,672	332,722
	2	18.25	0.02	20.44	97.54	93,474	93,466
	3	11.55	0.10	11.60	91.11	242,829	242,586
Daily	1	13.89	0.00	13.04	91.42	320,182	320,182
	2	13.89	0.00	12.17	85.48	84,955	84,955
	3	19.59	0.00	18.33	55.80	177,111	177,111

Figures 5 and 6 illustrate the time-series plots of regression forecasts (at both monthly and daily resolutions, alongside their respective deviations) covering the complete 33-month monitoring window for both facilities. The graphs indicate that, notwithstanding their straightforward structure, the models closely track the temporal trends, displaying substantial deviations between predicted and actual consumption at only a handful of data points.

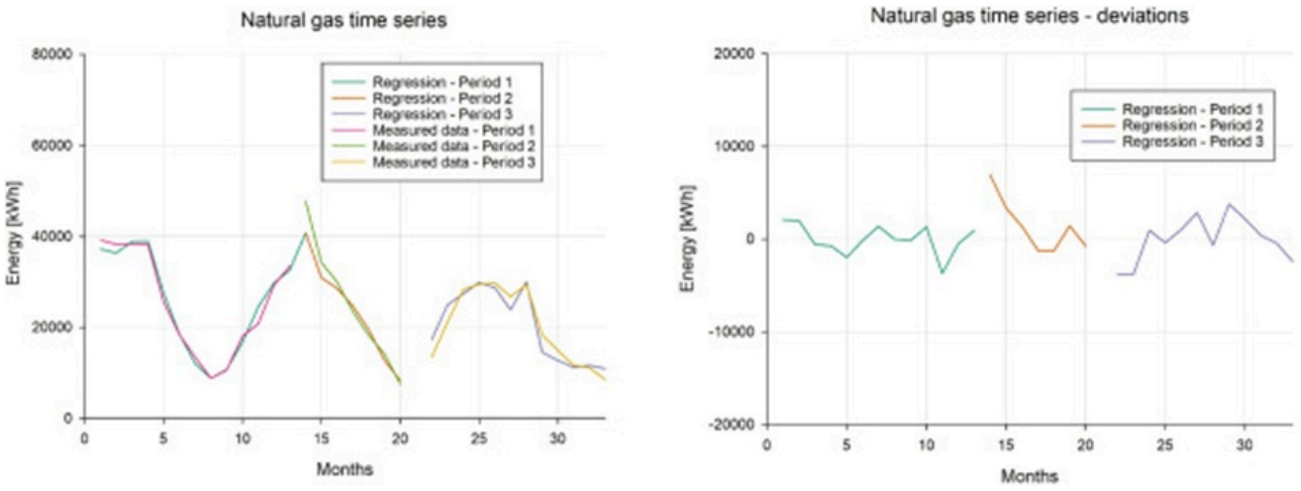


Figure 5. Time series of natural gas demand (left) and deviations (right)–monthly data.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

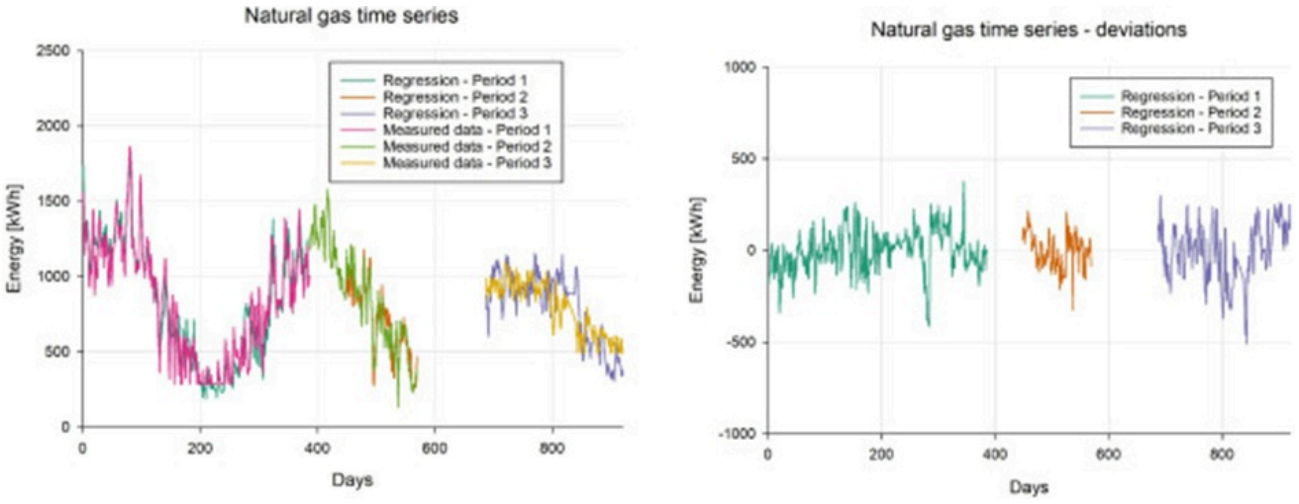


Figure 6. Time series of natural gas demand (left) and deviations (right) – daily data.

The data gap corresponds to the timeframe during which the heating infrastructure underwent retrofitting with the integration of GAHPs. Additionally, incomplete documentation led to the exclusion of several more days during the daily time-series data cleaning phase. Subsequently, the discrepancies between empirical measurements and model forecasts were scrutinized to identify any systematic patterns. For this analysis, no discernible structural patterns emerged within the deviations.

Lastly, following the methodology outlined in ISO 50,006:2014 [27], a time series was generated to track the variance between actual natural gas usage and the consumption predicted by models calibrated on historical data. This standard dictates that a “baseline” model must initially be trained and subsequently applied to assess performance during the “reporting” period. These figures are then aggregated via a cumulative sum approach to determine the overall energy conserved (representing the tangible savings realized relative to the operations in period 1 and period 2).

Examining the time series of absolute and cumulative savings presented in **Figure 7** reveals a surge in domestic hot water consumption during the pandemic era (the final segment of the monitoring timeline). This spike is likely attributable to intensified cleaning and sanitation activities. **Table 9** displays the computed performance enhancements, framed as a percentage decrease in energy usage against the baseline metrics across the distinct monitoring phases, which are further elaborated upon in **Table 6.**

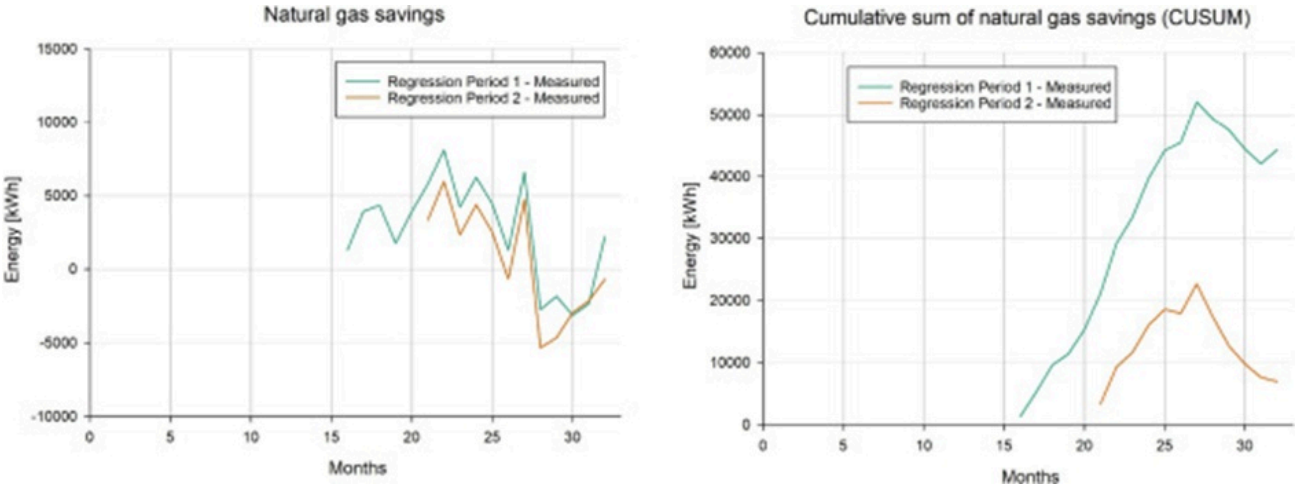


Figure 7. Time series of natural gas savings and cumulative sum of savings.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 9. Energy demand reduction using model prediction–Weather and behaviour normalized.

Period	Description	Overall energy reduction	Relative reduction
1	Pre-retrofit baseline condition	**0 (baseline)**	–
2	Installation of Thermostatic Radiator Valves (TRVs)	**7.2%**	**0 (baseline)**
3	Gas Absorption Heat Pumps (GAHPs) deployment and plant room modernization	**25.8%**	**20.0%**

To ensure a valid and standard comparison [27] under normal climatic circumstances, annual energy consumption was estimated using the formulated piecewise linear regression models alongside a representative local weather file (weather normalization). Furthermore, this calculation assumed a domestic hot water requirement consistent with pre-pandemic levels to account for behavioral normalization.

The deployment of smart TRVs yielded a moderate 7.2% decline in natural gas usage; however, this closely aligns with existing literature. Previous studies document potential reductions between 5%–10% via simulations in the UK residential context [10], and empirical long-term field assessments in Poland show actual savings of 7.1%–23.3% [11]. Benchmarking against these alternate case studies suggests that the conservation potential of the TRVs was not fully realized in this deployment, leaving room for optimization of their operational configurations. An exhaustive examination of operational behaviors within this overarching project confirmed this hypothesis, though those specific findings are excluded here to maintain the paper's scope. Ultimately, these outcomes reinforce the necessity for empirical evidence regarding the impact of smart controls in the residential sector [9], alongside careful evaluations of their cost-benefit ratio. The combined integration of Smart TRVs and GAHPs led to a 25.8% drop in energy demand, with the relative improvement determined by GAHPs sitting at 20.0%. This dynamic will be investigated in greater depth , where regression-based techniques are used to characterise GAHP performance and validate energy savings.

The evaluation sequence proceeds by incorporating supplementary variables to construct models θeIsol, θeθe,diff and θeIsolθe,diff, as specified in **Table 2**. The findings presented in **Tables 10-12** for models θeIsol, θeθe,diff, and θeIsolθe,diff, respectively, demonstrate that integrating these extra variables yields a generally uniform improvement across statistical metrics (as theoretically anticipated). Nevertheless, the gains observed in R2, MAPE, and CV(RMSE) remain negligible in this particular scenario, justifying the adoption of model θe as a practical reference framework for the digital twin application due to its simplicity. Naturally, for structures characterized by substantial thermal mass and/or intense solar heat gains, deploying models with extended variable sets could guarantee a marked upgrade in predictive capability.

Table 10. Results of analysis for natural gas demand signature–Model θeIsol.

Time interval	Period	CV(RMSE) (%)	NMBE (%)	MAPE (%)	R^2 (%)	Energy predicted (P)	Energy measured (M)
Monthly	1	6.33	−0.01	4.96	97.95	332,692	332,722
	2	5.54	−0.02	8.19	99.79	93,457	93,466
	3	10.07	0.08	10.13	93.25	242,772	242,586
Daily	1	13.58	0.00	13.13	91.81	320,182	320,182

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

	2	13.25	0.00	12.00	86.79	84,955	84,955
	3	19.55	0.00	18.35	55.99	177,111	177,111

Table 11. Results of analysis for natural gas demand signature – Model θeθe, diff.

Time interval	Period	CV(RMSE) (%)	NMBE (%)	MAPE (%)	R^2 (%)	Energy predicted (P)	Energy measured (M)
Daily	1	13.01	0.00	12.50	92.48	320,182	320,182
	2	13.46	0.00	11.89	86.37	84,955	84,955
	3	19.21	0.00	17.87	57.51	177,111	177,111

Table 12. Results of analysis for natural gas demand signature–Model θeIsolθe, diff.

Time interval	Period	CV(RMSE) (%)	NMBE (%)	MAPE (%)	R^2 (%)	Energy predicted (P)	Energy measured (M)
Daily	1	12.67	0.00	12.60	92.87	320,182	320,182
	2	12.80	0.00	11.60	87.66	84,955	84,955
	3	19.16	0.00	17.88	57.70	177,111	177,111

Testing regression model θe with rolling horizon retraining

This segment details the outcomes derived from training model θe on a daily data interval; however, deviating from the approach , a 15-day rolling horizon is applied in accordance with the theoretical foundation. The outputs from this sequence of calibrated θe models are then benchmarked against the overarching θe model. Fundamentally, this specific modeling paradigm serves two core objectives.

Primarily, it aims to emulate regression applications within a “digital twin” framework, wherein the system undergoes frequent recalibration, and its model coefficients are routinely updated. Secondarily, it facilitates cross-validation across the varied monitoring blocks by fragmenting the comprehensive dataset into discrete 15-day data sets. By doing so, fluctuations in coefficient estimates can be analyzed, and the potential risk of overfitting can be rigorously assessed.

To reliably gauge the heating-dependent segment of the regression, daily intervals featuring a peak temperature of 16 °C were selected, even though the building’s true balance point lies between 19.5–20.0 °C (data points situated too close to this change point tend to produce more unstable estimates). Given its potential for broad physical interpretation (similar to principles utilized in co-heating tests, for example), the heating demand regression slope (designated as parameter a1 for the models outlined in **Table 2**) is analyzed. As shown in **Table 13**, the average coefficient estimates derived from the daily θe model applied to the total dataset closely mirror those generated by the rolling-horizon model retraining (subsets), thereby validating the effectiveness of the approach.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 13. Results of regression model θe with digital twin approach–Estimates of the slope of the regression model when θe >16 °C.

Period	Digital twin θe (15-day rolling horizon)–Max	Digital twin θe (15-day rolling horizon)–Average	Digital Twin θe (15-day rolling horizon)–Min	Daily θe	Monthly θe
1	−2.26 kW/K	−3.03 kW/K	−3.37 kW/K	−2.88 ± 0.05 kW/K	−3.01 ± 0.15 kW/K
2	−2.56 kW/K	−2.86 kW/K	−3.23 kW/K	−2.81 ± 0.11 kW/K	−3.13 ± 0.29 kW/K
3	−1.47 kW/K	−1.73 kW/K	−2.18 kW/K	−1.61 ± 0.11 kW/K	−2.27 ± 0.28 kW/K

Testing towt model reformulation and benchmarking with implementations at the state-of-the-art

As outlined in **Table 3**, a primary distinction between the current state-of-the-art TOWT implementation in RMV2.0 and the newly proposed adaptation is the fact that the temperature is subsetted around the median (i.e. 50 percentile of data) and that the breadth of these temperature bins hinges on the absolute range (min-max) of the available data (which may result in different widths on the left and right sides of the temperature distribution). The alternative TOWT implementation introduced in this study deploys a piecewise linear function to distinguish between periods of high and low consumption, categorized as occupied and unoccupied hours in RMV2.0, which instead relies on a regression regarding Heating Degree-Days (HDD) and Cooling Degree-Days (CDD). This formulation, employing energy signatures, enables direct comparison between the piecewise linear models in **Table 2** (focusing solely on model θe here for simplicity) and the specific TOWT module that models the outdoor temperature dependence of the load/signature.

Distinct signatures are independently fitted for periods of high consumption (flagged as occupied in the original formulation) and low consumption (flagged as unoccupied in the original formulation). Evaluating the left-hand side of **Figures 8-10** across the three distinct monitoring periods reveals substantial alignment between the temperature-dependent piecewise linear component of the TOWT and model θe, with minor exceptions during summer when the TOWT more accurately mirrors empirical conditions, and during severe winter conditions when the temperature is low. A slope reduction occurs compared to the simplest regression model. Ultimately, this newly adapted model reformulation fits the data well and operates comfortably inside the acceptable model tolerance bounds established by ASHRAE 14:2014 and reported in **Table 4.**

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

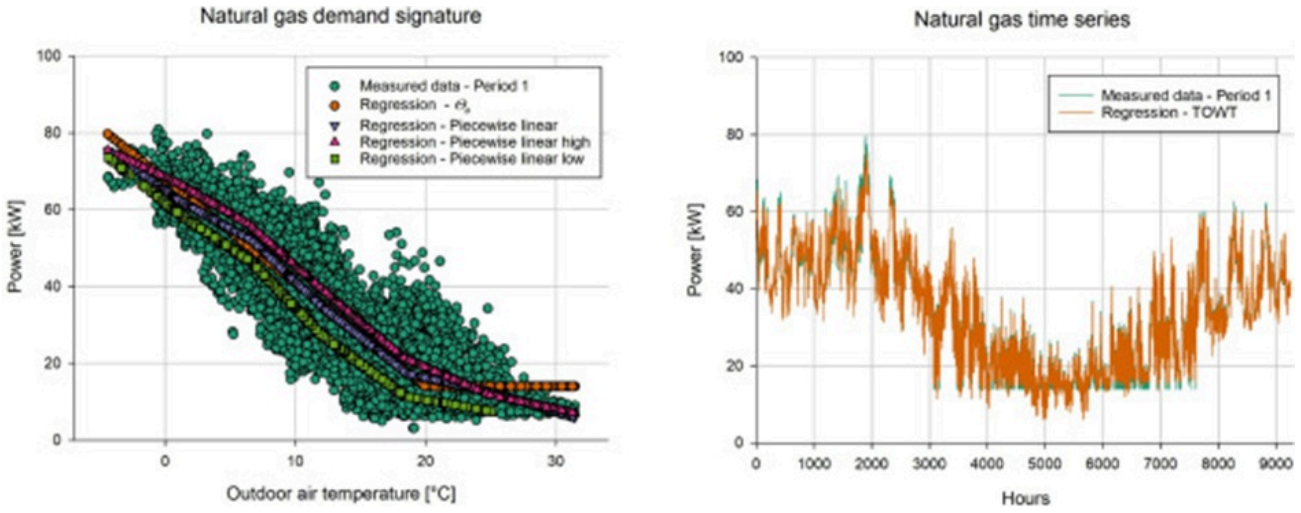


Figure 8. Natural gas demand signature (left) and time series (right) with hourly data, TOWT model – Period 1.

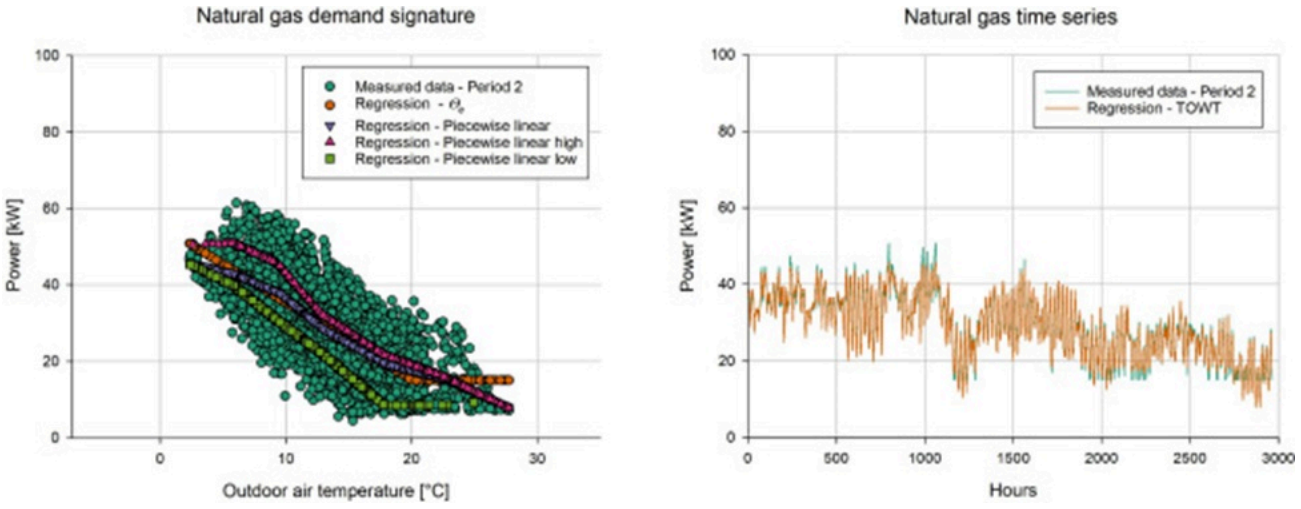


Figure 9. Natural gas demand signature (left) and time series (right) with hourly data, TOWT model – Period 2.

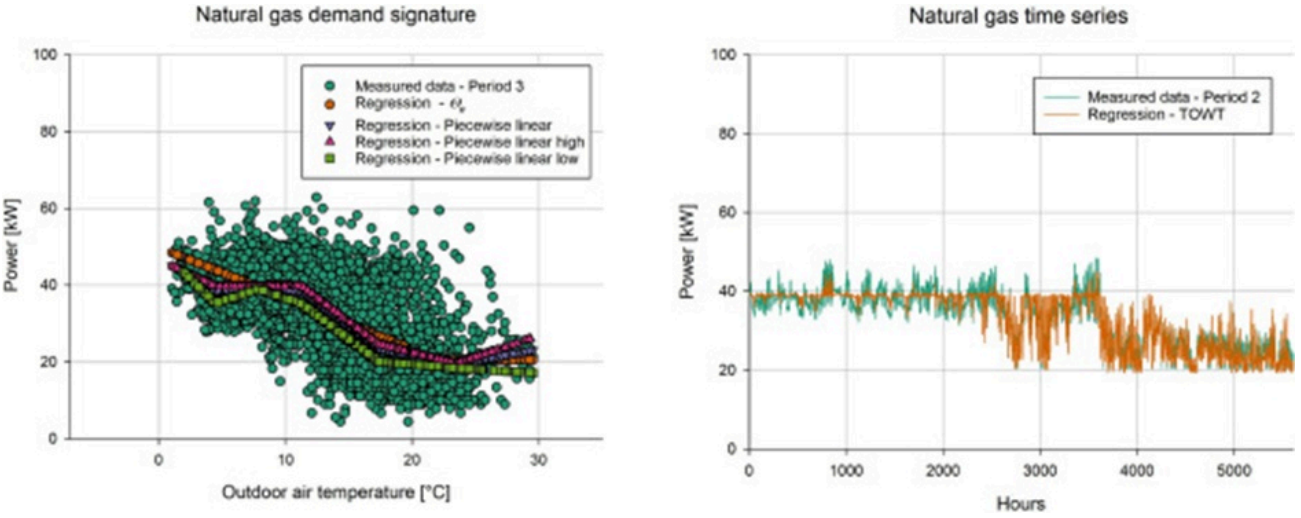


Figure 10. Natural gas demand signature (left) and time series (right) with hourly data, TOWT model – Period 3.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

The inherent scalability of visualizing energy signatures as a function of outdoor air temperature facilitates the straightforward cross-examination of model estimates from monthly to daily and hourly data formats, effectively demonstrated in **Figure 4** (monthly and daily interval) and **Figures 8-10** (hourly interval). This feature is particularly relevant when conducting coherent comparisons between low-frequency/long-term measures and high-frequency/short-term logs in a meaningful way. Moreover, evaluating the change-points and slopes of these energy signatures unlocks a multitude of analytical insights.

Finally, in numerical terms, the adapted TOWT framework fits the data well across the designated periods, as evidenced by **Table 14**. For this specific evaluation, the Mean Absolute Percentage Error (MAPE) indicator is deliberately excluded. Because hourly time series frequently capture instances where data approaches or equals 0 (i.e., if the system powers down for a segment of the day), MAPE ceases to be an appropriate measurement indicator.

Table 14. Results of analysis for natural gas demand signature, hourly data–TOWT model reformulation–1 month hyper-parameter.

Period	CV(RMSE) (%)	NMBE (%)	R^2 (%)	Energy Predicted (P)	Energy Measured (M)
1	14.70	0.00	91.65	319,408	319,408
2	13.01	0.00	88.72	86,772	86,772
3	17.04	0.00	66.56	186,636	186,636

The analytical outcomes achieved closely parallel those generated by the RMV2.0 TOWT configuration utilizing a 15-day hyper-parameter, detailed in **Table 15**, even though the weighting protocol implemented in RMV2.0 triggers a significantly higher volume of model runs, quantified in **Table 16**, particularly if the hyper-parameter is set to 1 day. Mechanically, the RMV2.0 setup requires an initial processing cycle to distinguish between occupied and unoccupied hours utilizing a degree-days model, after which the temporal sequence is segmented according to the hyper-parameter to generate a coefficient matrix designed to weight numerous regressions.

Table 15. Results of analysis for natural gas demand signature, hourly data – TOWT model implemented in RMV2.0 and different hyper-parameters.

Period	Hyper-parameter (days)	CV (RMSE) (%)	NMBE (%)	R^2 (%)	Energy predicted (P)	Energy measured (M)
1	30	15.25	0.24	90.43	318,951	319,408
	15	13.71	0.11	92.26	319,321	319,408
	1	3.69	0.00	99.44	319,751	319,408
2	30	16.83	0.22	82.24	85,995	86,772
	15	15.42	0.12	85.10	86,081	86,772
	1	4.03	0.00	98.98	86,499	86,772
3	30	19.39	−0.43	61.85	187,399	186,636
	15	17.22	−0.29	69.90	187,170	186,636
	1	4.57	−0.02	97.88	186,650	186,636

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 16. Comparison of the number of model runs between TOWT in RMV2.0 and the model reformulation.

Period	Days	TOWT–RMV2.0 (Hyper-parameter)	TOWT–Reformulation (Hyper-parameter)
		30 days	15 days
1	386	14	27
2	124	6	10
3	234	9	17

Conversely, the proposed adaptation also initiates with a single foundational model run to fit the piecewise linear temperature component required to segregate high (i.e., occupied) from low (i.e., unoccupied) consumption intervals. From there, however, merely two model runs are needed to determine the distinct high/low consumption piecewise components, as illustrated on the left side of **Figure 8-10**. Finally, a solitary model run constructs the Time of Week (TOW) model governing the residuals. In this specific design, residuals are differentiated by month, aligning with the hyper-parameter utilized in the newly proposed methodology. By drastically cutting down the number of models fitted, this strategy aims to distinctly reduce processing overhead while also mitigating the risk of model overfitting and limiting parameter bloat. A structurally comparable month-based weighting scheme is currently utilized within the Caltrack/OpenEEmeter [50] TOWT implementation, which nonetheless retains a temperature segmentation modeling approach heavily akin to RMV2.0 [46].

Testing regression model for GAHP performance characterization in variable conditions

This section provides a more detailed examination of the energy demand reduction achieved by the GAHPs (20.0%, as detailed in **Table 9**), taking into account the authentic physical dynamics of the heat pump technology. **Table 17** presents the outcomes derived from fitting the dual models described in **Eqs. 9 and 10** , corresponding to Qth and GUEmax, respectively.

Table 17. Regression model R2 and standard error (SE).

Metric	GUEmax (Value)	GUEmax (Unit)	Qth (Value)	Qth (Unit)
R^2	97.4	%	97.3	%
SE	0.03	–	2.21	kW

It is evident that, despite their simplicity, the models demonstrate a strong fit to the data, characterized by a high coefficient of determination (R2) and a low standard error for both Qth and GUEmax.

Subsequently, the model forecasts for Qth and GUE are illustrated in **Figure 11** (left and right panels, respectively) alongside the computed GAHP metrics, which were derived via interpolation from technical datasheet information. Furthermore, the graphs display individual regression lines corresponding to distinct supply temperatures (50, 55, 60, 65 °C). The observation that these slopes vary only slightly supports the formulation of a unified regression model utilizing

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

two input variables: θe and θsup. The scope of the model could be broadened by integrating dummy variables (e.g., to address non-linearity and change points), similar to the frameworks outlined in **Table 2**. However, such an expansion is unnecessary for the present scenario.

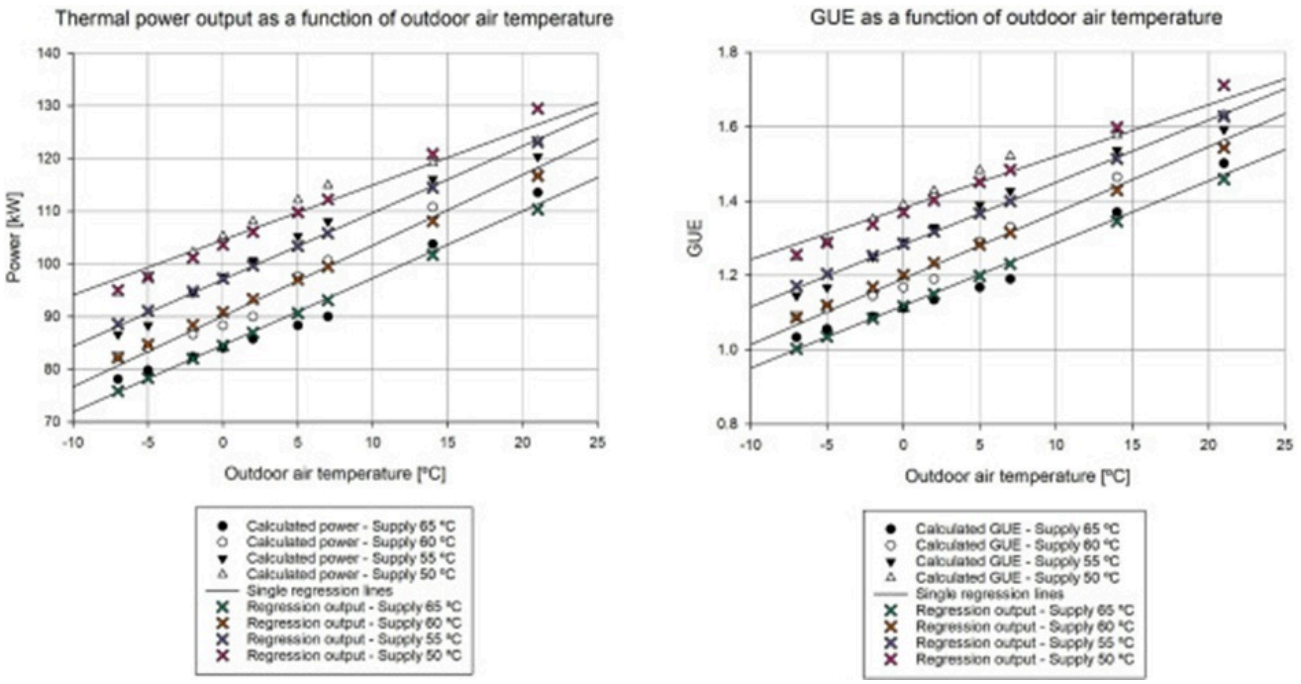


Figure 11. Regression model for the calculation of thermal power output Qth and GUEmax for the GAHP in variable conditions.

A study authored by Schmitt-Gehrke *et al*. [8] evaluated various GAHP models under partial load states, computing the GUE, which generally falls between 0.80–0.90. As detailed previously at the conclusion, this metric quantifies part-load performance deterioration and mirrors the Part Load Fraction (PLF) applied to electrical heat pumps. The findings from the current investigation align with those reported in the aforementioned literature. For a rudimentary verification, assuming an average outdoor air temperature of 14 °C (applicable to monitoring period 3) alongside a supply temperature between 60–65 °C (due to fixed-point control and a lack of weather compensation), **Figure 4** suggests a GUEmax of approximately 1.4. Applying a GUE of 0.85 (which sits within the previously noted literature range for partial load GAHP evaluations), the adjusted final GUE reaches roughly 1.2. This figure essentially corroborates the empirical performance enhancement realized by the GAHP (a 20.0% gain over condensing boilers). Under these specific operating parameters, the legacy condensing boilers operate with an efficiency closely approaching 1 (0.97–0.98); thus, a GUE near 1.2 aligns well with a 20% efficiency gain. The condensing boilers in this specific building fail to operate in condensing mode because of the elevated supply temperatures (approximately 60–65 °C). Naturally, a more rigorous assessment would necessitate the precise weighting of the GUE against varying load demands and outdoor air temperatures. Such a procedure would closely resemble the technical standardization protocols [63, 64] used to calculate the Seasonal Coefficient of Performance (SCOP) in electrical heat pumps, incorporating specific adjustments for part-load operation across diverse states. Nevertheless, the approach introduced in this study offers a rapid, straightforward method for performance verification, yielding outcomes comparable to alternative investigations that documented a 10% to 15% boost in annual heating efficiency when utilizing GAHPs instead of conventional alternatives [65].

The degree of performance enhancement relative to natural gas boilers is inherently limited by the heating system's supply temperature, as illustrated by the GUE value ranges shown in **Figure 11**. Consequently, facilities requiring elevated supply temperatures will inevitably experience smaller relative performance gains compared to structures operating with lower supply temperatures. This principle is critical and must be factored into retrofit planning; overly

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

optimistic projections regarding GAHP efficiency could jeopardize the return on investment, which is strictly tied to the tangible energy conservation realized.

Further research potential

The findings presented and evaluated, open several avenues for future exploration. A primary enhancement focuses on synthesizing these diverse methodologies into a cohesive workflow that efficiently leverages temporal scalability. To illustrate, analyzing daily data allows for the identification of specific weekly trends (such as variations in occupancy and operational modes across different weekdays, or pronounced contrasts between weekends and weekdays). Furthermore, incorporating hourly and sub-hourly records facilitates the characterization of operational schedules and the building's dynamic thermal response, thereby aiding in the optimization of system parameters. Broadly speaking, the concurrent availability of long-term/low-frequency data (e.g., monthly records spanning several years) and short-term/high-frequency logs (sub-hourly, hourly, daily) substantially strengthens performance evaluations while streamlining analytical efforts. **Table 18** provides a comprehensive summary of these prospective research trajectories, categorized by specific applications and domains. As highlighted in the table, employing an approximate physical interpretation serves as a crucial mechanism for establishing sensible boundaries and parameters within the modeling framework, ultimately fostering greater confidence in the generated predictions. Moreover, this distinctive trait of interpretable approaches has the potential to connect with cutting-edge research regarding whole-building energy model calibration [6] through a multi-tiered methodology [66]. Within this paradigm, aggregate physical characteristics are isolated [67] and applied to streamlined physical models, facilitating dynamic operational improvements and boosting overall building flexibility [68].

Table 18. Further potential research developments with respect to areas and applications.

Research area	Applications	Potential development
Modelling techniques	Automated modelling workflow	Further development of data pre-processing and post-processing pipelines. An automated forward model selection approach (e.g., incremental inclusion of additional variables) can be integrated with cross-validation techniques.
	Regression-based formulation using approximated physical models	Building balance-point temperatures correspond to change-points in piecewise regression, while slope parameters are strongly linked to the overall heat transfer coefficient.
Building technologies	Whole-building performance analysis and model calibration	Approximate physical models and analytical formulations can define feasibility bounds and constraints for model parameters, utilizing energy balance relationships at multiple levels of the building system.
	Construction technologies, air-tightness, and ventilation	The overall heat transfer coefficient of building zones can be estimated through regression-based methods.
	Technical systems	Building balance-point temperature (i.e., change-points in piecewise regression) affects weather-compensated control of heat pumps and influences performance metrics such as COP and GUE.

Conclusion

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

This study introduced several designs for streamlined, interpretable (regression-oriented) “digital twins” intended for tracking building energy dynamics. Initially, incorporating supplemental variables demonstrated prospective advancements over current state-of-the-art model architectures. Subsequently, model resilience was evaluated through a retraining protocol employing a 15-day rolling horizon, which revealed an acceptable degree of stability within the coefficient estimations. Additionally, the study put forth an adapted version of the time of week and temperature (TOWT) model, delivering strong predictive performance alongside improved interpretability via energy signature visualizations. Ultimately, regression techniques were also utilized to characterize GAHP performance.

In summary, the research illustrated that transparent, regression-centric methodologies function efficiently across diverse temporal analytical scales. Because they operate adequately even when constrained to limited inputs like monthly records, they prove highly valuable for rapid, cost-effective performance appraisals (utilizing basic outdoor air temperatures and utility invoices). Even so, model accuracy can be elevated by refining data resolution from monthly down to daily, and eventually to hourly or sub-hourly intervals, alongside the integration of further variables. Specifically, adopting hourly or sub-hourly intervals permits the precise mapping of operating routines (e.g., the time-of-week model module) and the building’s dynamic thermal attributes (such as its reaction to external temperatures), information that is pivotal for tuning operational configurations. Lastly, harnessing the models’ temporal scalability allows for the strategic combination of low-frequency/long-term tracking (monthly data spanning years) with high-frequency/short-term readings (sub-hourly, hourly, or daily metrics).

During the tracking phase, recorded energy reductions were verified by normalizing against operational parameters and climatic factors. The implementation of Smart TRVs yielded a 7.2% decline in energy consumption, whereas the joint deployment of GAHPs and TRVs resulted in a 25.0% total reduction. While the savings attributed to TRVs fell toward the bottom of the anticipated spectrum (5%–20%), optimized system calibrations could likely enhance this outcome. The 20.0% relative drop in energy demand facilitated by GAHPs is practically attainable under the operating conditions observed during the study. Although GAHP systems present a credible substitute for both condensing and traditional boilers, achieving stringent low-carbon objectives would require them to operate on synthetic fuels or hydrogen, introducing an element of unpredictability regarding their future application.

Prospective scientific advancements could encompass refining the modeling protocols outlined in this manuscript, as well as applying them to specialized practical scenarios. The former involves seamlessly embedding these modeling strategies into an analytical pipeline designed to deliver visual and quantitative insights, using cross-validation frameworks and forward model selection (thereby mitigating the risk of overfitting and strengthening estimate reliability). The latter focuses on constructing regression architectures grounded in an approximated physical reality (effectively a “grey-box” methodology). This approach yields a deeper understanding of the facility’s overall performance (including enabling comprehensive building energy model calibration), alongside insights into operational tactics, technical infrastructure, and building materials. Ultimately, while such modeling paradigms might lack the raw predictive power and precision of cutting-edge machine learning algorithms, their approximate physical grounding remains crucial for interpretability—a central pillar of this study. By ensuring that model-derived analytics are comprehensible from a human perspective, this transparency fundamentally cultivates heightened trust in the resulting outputs.

Acknowledgements

None

Conflict of interest

None

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Financial support

This study and project are financially supported by the EU Research and Innovation programme Horizon 2020 through number 723562–THERMOSS. The authors would like to thank the European Commission for enabling the funding of this project.

Ethics statement

None

Received: 18 Sep 2023 Revised: 24 Nov 2023 Accepted: 11 Dec 2023
Published online: 15 June 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Rosenow J, Eyre N. Reinventing energy efficiency for net zero. Energy Res Soc Sci. 2022;90:102602. https://doi.org/10.1016/j.erss.2022.102602.

2. Markard J, Geels FW, Raven R. Challenges in the acceleration of sustainability transitions. Environ Res Lett. 2020;15:81001. https://doi.org/10.1088/1748-9326/ab9468.

3. Sunny N, Mac Dowell N, Shah N. What is needed to deliver carbon-neutral heat using hydrogen and CCS? Energy Environ Sci. 2020;13:4204-24. https://doi.org/10.1039/D0EE02016H.

4. Jayaweera T, Stern F, Violette DM, Baumgartner R. The uniform methods project: methods for determining energy efficiency savings for specific measures. Contract. 2013;303:275-3000.

5. Investor Confidence Project [Internet]. [cited 2020 Aug 31]. Available from: https://europe.eeperformance.org/

6. Chong A, Gu Y, Jia H. Calibrating building energy simulation models: a review of the basics to guide future work. Energy Build. 2021;253:111533. https://doi.org/10.1016/j.enbuild.2021.111533.

7. EN 12309-3:2014. Gas-fired sorption appliances for heating and/or cooling with a net heat input not exceeding 70 kW. Test conditions. Brussels: European Committee for Standardization; 2014.

8. Schmitt-Gehrkea P, Buchina O, Cigandaa JLC, Grafa R, Kühnb A, Zieglera F. Part load performance of gas fired absorption heat pumps. In: Proc th IEA Heat Pump Conf; 2017.

9. Lomas KJ, Oliveira S, Warren P, Haines VJ, Chatterton T, Beizaee A, et al. Do domestic heating controls save energy? A review of the evidence. Renew Sustain Energy Rev. 2018;93:52-75.

https://doi.org/10.1016/j.rser.2018.05.002.

10. Marshall E, Steinberger JK, Dupont V, Foxon TJ. Combining energy efficiency measure approaches and occupancy patterns in building modelling in the UK residential context. Energy Build. 2016;111:98-108. https://doi.org/10.1016/j.enbuild.2015.11.039.

11. Cholewa T, Siuta-Olcha A, Balaras CA. Actual energy savings from the use of thermostatic radiator valves in residential buildings – long term field evaluation. Energy Build. 2017;151:487-93. https://doi.org/10.1016/j.enbuild.2017.06.070.

12. Hahn J, Heiler S, Kane MB, Park S, Jensch W. The information gap in occupant-centric building operations: lessons learned from interviews with building operators in Germany. Front Built Environ. 2022;8. https://doi.org/10.3389/fbuil.2022.838859.

13. Dalibor M, Jansen N, Rumpe B, Schmalzing D, Wachtmeister L, Wimmer M, et al. A cross-domain systematic mapping study on software engineering for digital twins. J Syst Softw. 2022;193:111361. https://doi.org/10.1016/j.jss.2022.111361.

14. Wright L, Davidson S. How to tell the difference between a model and a digital twin. Adv Model Simul Eng Sci. 2020;7:13. https://doi.org/10.1186/s40323-020-00147-4.

15. de Wilde P. Building performance simulation in the brave new world of artificial intelligence and digital twins: a systematic review. Energy Build. 2023;292:113171. https://doi.org/10.1016/j.enbuild.2023.113171.

16. Tran M, Amer M, Dababat A, Abdelaziz AY, Dai H-J, Liu M-K, et al. Robust fault recognition and correction scheme for induction motors using an effective IoT with deep learning approach. Measurement. 2023;207:112398. https://doi.org/10.1016/j.measurement.2022.112398.

17. ISO/IEC. ISO/IEC TR 29119-11:2020(en) software and systems engineering — software testing — part 11: guidelines on the testing of AI-based systems. Geneva: International Organization for Standardization; 2020.

18. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: fundamental principles and 10 grand challenges. Stat Surv. 2022;16:1-85.

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

20. Molnar C. Interpretable machine learning [Internet]. Section 3.2 Taxonomy of Interpretability Methods. [cited 2023 May 24]. Available from: https://christophm.github.io/interpretable-ml-book/taxonomy-of-interpretability-methods.html

21. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harvard Data Sci Rev. 2019;1. https://doi.org/10.1162/99608f92.8cd550d1.

22. THERMOSS project [Internet]. [cited 2021 Apr 22]. Available from: https://energy.soton.ac.uk/project/thermoss

23. ASHRAE. ASHRAE Guideline 14-2014, Measurement of energy, demand, and water savings. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2014.

24. International Performance Measurement and Verification Protocol (IPMVP): Volume I: Concepts and Options for Determining Energy and Water Savings. EVO 10000–1:2012. Washington, DC: Efficiency Valuation Organization (EVO); 2012.

25. FEMP. M&V guidelines: measurement and verification for performance-based contracts, version 4.0. 2015.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

26. Gallagher CV, Leahy K, O'Donovan P, Bruton K, O'Sullivan DTJ. Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. Energy Build. 2018;167:8-22. https://doi.org/10.1016/j.enbuild.2018.02.023.

27. ISO 50006:2014, Energy management systems — measuring energy performance using energy baselines (EnB) and energy performance indicators (EnPI) — general principles and guidance. Geneva: International Organization for Standardization; 2014.

28. ISO 16346:2013, Energy performance of buildings — assessment of overall energy performance. Geneva: International Organization for Standardization; 2013.

29. 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.

30. 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.

31. Abels B, Sever F, Kissock K, Ayele D. Understanding industrial energy use through lean energy analysis. SAE Int J Mater Manuf. 2011;4:495-504. https://doi.org/10.4271/2011-01-0319.

32. Lin G, Claridge DE. A temperature-based approach to detect abnormal building energy consumption. Energy Build. 2015;93:110-8. https://doi.org/10.1016/j.enbuild.2015.02.013.

33. Fu H, Baltazar J-C, Claridge DE. Review of developments in whole-building statistical energy consumption models for commercial buildings. Renew Sustain Energy Rev. 2021;147:111248. https://doi.org/10.1016/j.rser.2021.111248.

34. 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. Energy Build. 2021;244:111054. https://doi.org/10.1016/j.enbuild.2021.111054.

35. Bauwens G, Roels S. Co-heating test: a state-of-the-art. Energy Build. 2014;82:163-72. https://doi.org/10.1016/j.enbuild.2014.04.039.

36. Jack R, Loveday D, Allinson D, Lomas K. First evidence for the reliability of building co-heating tests. Build Res Inf. 2018;46:383-401. https://doi.org/10.1080/09613218.2017.1299523.

37. Alzetto F, Farmer D, Fitton R, Hughes T, Swan W. Comparison of whole house heat loss test methods under controlled conditions in six distinct retrofit scenarios. Energy Build. 2018;168:35-41. https://doi.org/10.1016/j.enbuild.2018.03.024.

38. Masuda H, Claridge DE. Inclusion of building envelope thermal lag effects in linear regression models of daily basis building energy use data. 2012.

39. Danov S, Carbonell J, Cipriano J, Martí-Herrero J. Approaches to evaluate building energy performance from daily consumption data considering dynamic and solar gain effects. Energy Build. 2013;57:110-8. https://doi.org/10.1016/j.enbuild.2012.10.050.

40. Verhelst C, Degrauwe D, Logist F, Van Impe J, Helsen L. Multi-objective optimal control of an air-to-water heat pump for residential heating. Build Simul. 2012;5:281-91.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

https://doi.org/10.1007/s12273-012-0061-z.

41. Péan T, Salom J, Costa-Castelló R. Configurations of model predictive control to exploit energy flexibility in building thermal loads. In: 2018 IEEE Conf Decis Control. 2018. p. 3177-82. https://doi.org/10.1109/CDC.2018.8619452.

42. Price P. Methods for analyzing electric load shape and its variability. Berkeley, CA: Lawrence Berkeley National Laboratory; 2010. Report LBNL-3713E.

43. 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. https://doi.org/10.1109/TSG.2011.2149545.

44. Borgeson SD. Targeted efficiency: using customer meter data to improve efficiency program outcomes [dissertation]. Berkeley, CA: University of California; 2013.

45. EVO's Advanced M&V Testing Portal [Internet]. [cited 2026 Jul 6]. Available from: https://mvportal.evo-world.org/

46. RMV2.0 - LBNL M&V2.0 Tool [Internet]. [cited 2026 Jul 6]. Available from: https://lbnl-eta.github.io/RMV2.0/

47. Touzani S, Granderson J, Fernandes S. Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy Build. 2018;158:1533-43. https://doi.org/10.1016/j.enbuild.2017.11.039.

48. NMECR [Internet]. [cited 2026 Jul 6]. Available from: https://kw-labs.github.io/nmecr/

49. OpenEEmeter [Internet]. [cited 2026 Jul 6]. Available from: https://github.com/openeemeter/eemeter

50. CalTRACK Methods [Internet]. [cited 2026 Jul 6]. Available from: http://docs.caltrack.org/en/latest/methods.html

51. EENSIGHT [Internet]. [cited 2026 Jul 6]. Available from: https://github.com/hebes-io/eensight

52. Qaisar I, Zhao Q. Energy baseline prediction for buildings: a review. Results Control Optim. 2022;7:100129. https://doi.org/10.1016/j.rico.2022.100129.

53. Alrobaie A, Krarti M. A review of data-driven approaches for measurement and verification analysis of building energy retrofits. Energies. 2022;15. https://doi.org/10.3390/en15217824.

54. 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.

55. Manfren M, Nastasi B, Tronchin L. Linking design and operation phase energy performance analysis through regression-based approaches. Front Energy Res. 2020;8:288. https://doi.org/10.3389/fenrg.2020.557649.

56. 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.

57. 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.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

58. Krese G, Lampret Ž, Butala V, Prek M. Determination of a building's balance point temperature as an energy characteristic. Energy. 2018;165:1034-49. https://doi.org/10.1016/j.energy.2018.10.025.

59. Hao Z, Zhang X, Xie J, Yin K, Liu J. Balance point temperature and heating degree-days in different climate conditions for building energy efficiency applications. Build Environ. 2022;216:109013. https://doi.org/10.1016/j.buildenv.2022.109013.

60. Nastasi B, Manfren M, Groppi D, Lamagna M, Mancini F, Astiaso Garcia D. Data-driven load profile modelling for advanced measurement and verification (M&V) in a fully electrified building. Build Environ. 2022;221:109279. https://doi.org/10.1016/j.buildenv.2022.109279.

61. EN 15316-4-2:2017. Energy performance of buildings. Method for calculation of system energy requirements and system efficiencies. Space heating generation systems, heat pump systems, Module M3-8-2, M8-8-2. Brussels: European Committee for Standardization; 2017.

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

63. EN 14825:2018. Air conditioners, liquid chilling packages and heat pumps, with electrically driven compressors, for space heating and cooling. Testing and rating at part load conditions and calculation of seasonal performance. Brussels: European Committee for Standardization; 2018.

64. EN 12309-6:2014. Gas-fired sorption appliances for heating and/or cooling with a net heat input not exceeding 70 kW. Calculation of seasonal performances. Brussels: European Committee for Standardization; 2014.

65. Kühn A, Graf R, Ciganda JC, Schmitt-Gehrke P, Gassenmeyer F, Blum N, et al. Betriebsstrategien für Gasabsorptionswärmepumpen. 2016.

66. Yang Z, Becerik-Gerber B. A model calibration framework for simultaneous multi-level building energy simulation. Appl Energy. 2015;149:415-31. https://doi.org/10.1016/j.apenergy.2015.03.048.

67. 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. Energy Build. 2021;241:110889. https://doi.org/10.1016/j.enbuild.2021.110889.

68. Li H, Hong T. On data-driven energy flexibility quantification: a framework and case study. Energy Build. 2023;296:113381. https://doi.org/10.1016/j.enbuild.2023.113381.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press