Energy and AI 14 (2023) 100304

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.sciencedirect.com/journal/energy-and-ai

Lean and interpretable digital twins for building energy monitoring – A case study with smart thermostatic radiator valves and gas absorption heat pumps

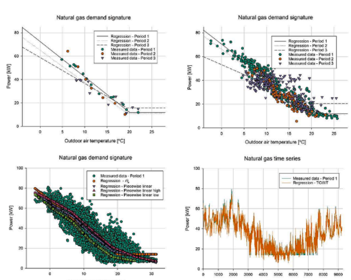
Massimiliano Manfren [a,*], Patrick AB James [a], Victoria Aragon [a], Lamberto Tronchin [b]

[a] Faculty of Engineering and Physical Sciences, University of Southampton, Boldrewood Campus – SO16 7QF Southampton, United Kingdom
[b] Department of Architecture (DA), University of Bologna, Via Cavalcavia 61 47521 Cesena, Italy

HIGHLIGHTS

- Energy signature regression is used for M&V with monthly,daily and hourly data.
- Energy signature models are retrained as "digital twins" with rolling horizon.
- TOWT regression model is reformulated to enhance interpretability.
- A regression model is developed to characterize GAHP performance.
- Energy savings due to smart heating controllers (TRVs) and GAHPs are verified.

GRAPHICAL ABSTRACT



ARTICLE INFO

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

ABSTRACT

The transition to low carbon energy systems poses challenges in terms of energy efficiency. In building refurbishment projects, efficient technologies such as smart controls and heat pumps are increasingly being used as a substitute for conventional technologies with the aim of reducing carbon emissions and determining operational energy and cost savings, together with other benefits. Measured building performance, however, often reveals a significant gap between the predicted energy use (design stage) and actual energy use (operation stage). For this reason, lean and interpretable digital twins are needed for building energy monitoring aimed at persistence of savings and continuous performance improvement. In this research, interpretable regression models are built with data at multiple temporal resolutions (monthly, daily and hourly) and seamlessly integrated with the goal of verifying the performance improvements due to Smart thermostatic radiator valves (TRVs) and gas absorption heat pumps (GAHPs) as well as giving insights on the performance of the building as a whole. Further, as part of modelling research, time of week and temperature (TOWT) approach is reformulated and benchmarked against its original implementation. The case study chosen is Hale Court sheltered housing, located in the city of Portsmouth (UK). This building has been used for the field-testing of innovative technologies such as TRVs and GAHPs within the EU Horizon 2020 project THERMOSS. The results obtained are used to illustrate possible extensions of the use of energy signature modelling, highlighting implications for energy management and innovative building technologies development.

* Corresponding author.
E-mail address: M.Manfren@soton.ac.uk (M. Manfren).

https://doi.org/10.1016/j.egyai.2023.100304

Available online 26 September 2023

1. Introduction

The transition to low carbon energy systems poses challenges in terms of energy efficiency. The electrification of end-uses is widely recognised as a key component for achieving complete decarbonisation [1], even though the shift of impacts [2] should not be underestimated. At the same time, the use of hydrogen as a substitute for natural gas (and other fuels) in order to achieve decarbonization goals presents a number of challenges, cost being amongst the most significant ones [3].

In general, the evaluation of the impact of energy efficiency measures has to be conducted using reliable methods, considering the relevant investments needed [4,5], and can constitute the starting point for other more in depth evaluations, using detailed calibrated simulations [6].

Innovative heating technologies are part of this research and gas absorption heat pumps (GAHPs) can be used as an alternative to conventional and condensing boilers. A GAHP is a thermal machine that uses natural gas combustion process as heat source to feed an absorption cycle. In the cycle, heat is subtracted from outdoor air (heat source) and released to water in the heating system (heat sink), with higher efficiency compared to a condensing boiler. GAHPs can provide similar services compared to air source heat pumps (ASHPs) but are fed by natural gas instead of electricity. Therefore, GAHP technology can be considered as an option for the replacement of condensing boilers, even though it creates a potential lock-in condition in the long-term, due to the use of fossil fuels and higher temperature heating systems.

In operation terms, GAHP performance depends on the temperature of the heat source (i.e. outdoor air temperature) and of the heat sink (i.e. water temperature in the heating system) and, while testing standards exists [7] for this technology, field testing is necessary to evaluate performance degradation (compared to data measured in standard test conditions) determined by operation at part load conditions. Recent studies have considered, in particular, the part load performance testing of GAHPs [8] and the comparison of GAHP technology [9] with other options in heating systems . The results of these studies are compared with the ones obtained in our case study. Additionally, smart heating controller, such as Smart Thermostatic Radiator Valves (TRVs) are often reported as a promising technology, and studies indicate a certain potential [10,11], which has to be verified because it depends on multiple factors and, in particular, on operational patterns and user behaviour. In this sense, it is crucial to understand actual operation strategies to overcome the information gap [12] regarding building performance.

The identification of operational (as well as design) issues using digital replicas of objects/processes/services is part of current research in several fields and a multiplicity of “digital twin” definitions are present across different studies [13], in which sometimes the difference between model and digital twin is blurred [14]. Further, this topic is attracting a lot of attention also in building and energy sectors [15], due to the increasing use of Artificial Intelligence (AI) and Machine Learning (ML) tools; however, while sophisticated ML techniques such as deep learning have been effective in multiple applications [16], they lack interpretability [17], which is today an open challenge [18]. As show by Chen et al. [19] sophisticated machine learning approaches used in building energy management employ, in most of the cases, post-hoc techniques such as LIME and SHAP [20] to interpret their results, but it is very difficult to inspect the model algorithmic logic in a simple and transparent way. This represents an issue from multiple points of view, including AI/ML ethics [21] and trust by practictioners in the energy sector.

The case study chosen for this research is Hale Court sheltered housing development, run by Portsmouth City Council, which has been retrofitted as part of the EU Horizon 2020 project THERMOSS [22]. Hale Court development comprises 80 flats divided in two blocks, North and South, each with independent heat networks. The analysis concentrates on the North block which has been monitored before and after retrofit interventions and has been retrofitted in two steps, first with the installation of Smart TRVs and then with the installation of GAHPs. Monitored data are split into three periods, namely before retrofit, after Smart TRVs installation and after GAHPs installation and plant room upgrade. The performance before retrofit represents the baseline for the evaluation of the energy savings achieved by new technologies. Multiple models are created incrementally in an integrated workflow, where novel formulations are tested with data at monthly, daily and hourly resolution, to provide temporal scalability, while retaining (intrinsic/ante-hoc [20]) interpretability of regression-based approaches and providing insights on the performance of new technologies installed and on the building as a whole. The origin of the integrated workflow used in this research is described in Section 3, while the technical details of its implementation are specified in Section 4.

2. Nomenclature

3. Literature review

Data-driven methods for energy monitoring and management must address two important issues in present and future perspectives: providing a reliable assessment of the actual impact of different technologies in terms of efficiency and enabling persistence of saving and continuous improvement of management practices, following the considerations expressed in the introduction regarding the applications of digital twins [13,15].

For this reason, the starting point in this research are the regression-based approaches proposed by Measurement and Verification (M&V) protocols at the state-of-the-art such as ASHRAE 14:2014 [23], efficiency value organization (EVO) [24], federal energy management program (FEMP) [25], whose evolution is indicated with the term Advanced M&V or M&V 2.0 [26]. This evolution involves the use of more sophisticated algorithms, more granular data (e.g. smart metre readings), greater accuracy and faster feedback. In this sense, the term "digital twins" is used because of their ability to reproduce actual behaviour on a dynamic basis and to provide useful analytics for the identification and solution of operation issues.

In this research, the regression-based approaches proposed by ASHRAE 14:2014 [23] and by ISO 50,006:2014 [27] are used as a baseline (most consolidated approaches, with typical recurring energy demand characteristics for different building types and climates), to enable a meaningful comparison of efficiency measures monitored in different periods and estimated using different types of models, which are chosen because of their characteristics, which will be illustrated later in this section. Further, all models presented in this research employ energy signatures as defined in ISO 16,346:2013 [28], i.e. they model the average power obtained by dividing the energy with respect to the amount of operating hours in the time interval considered, namely monthly, daily, hourly. This choice is based on the necessity to achieve a certain degree of temporal scalability for the models, which can be exploited also in visual analytics to compare long-term measures (low-frequency data) and short-term measures (high frequency) in a meaningful way. Indeed, regression-based approaches are widely recognized in the building energy management sector, with a variety of applications from simple and intermediate level [29] to advanced ones [30].

First of all, M&V approaches used in building energy management are based on energy interval data (dependant variable) and on weather data (independent variables), along with other independent variables (e.g. dummy variables to model different occupancy and operational regimes, extracted from contextual information). The most important independent variable for weather normalization of energy consumption is outdoor air temperature [23,31,32] and univariate regression formulations are preferred by practitioners [33] for whole-building statistical modelling. While this approach can lead to less accurate models compared to more sophisticated techniques, it is nonetheless insightful due to its interpretability [34].

The applicability of regression-based approaches is not limited to whole building energy monitoring, but it can be extended further with the use of additional variables, beyond outdoor air temperature which is needed for weather normalization of energy demand. In regression for co-heating tests [35,36] (building fabric performance characterisation), for example, solar radiation on horizontal surfaces is used to account for the solar gains component in the thermal energy balance of building zones, with tests lasting around 15 days on average [37]. Furthermore, differencing outdoor air temperatures in a time series with daily interval (i.e., taking differences between one day and the previous day) simplifies the quantification of the influence of the building fabric's inertia [38,39]. Additionally, heating system supply temperature is used in combination with outdoor air temperature to characterise the performance of air source heat pumps (ASHPs) for control purpose [40,41].

Amongst the regression-based techniques leveraging hourly interval data, time of week and temperature (TOWT) algorithm was proposed initially by Price [42] to analyse electric load shape and its variability. It has been used then for the quantification of changes in electricity use due to demand response [43] and in the context of utility scale efficiency programs [44]. TOWT was successfully tested as well in the advanced M&V testing portal of efficiency valuation organization (EVO) [45], a portal for open and independent testing, with electricity consumption data from 367 buildings located in various regions of North America. TOWT is a software R package named RMV2.0 [46]; in this package, a gradient boosting machine method is implemented as well [47]. However, TOWT is much easier to use, with only one hyper-parameter to be tuned. At present, various open software packages exists that implement TOWT and other regression-based techniques, namely R package NMECR [48], Python package OpenEEmeter [49] that implements Caltrack methods [50], or the recent EENSIGHT [51]. Further, extensive reviews have been developed around the problem of energy baseline modelling [52], whole-building statistical energy consumption models [33], use of physics-driven and data-driven approaches for measurement and verification in building retrofits [53,54], and more, in general, interpretable energy analytics in the construction sector [55,56]. Finally, as show by Chen et al. [19] sophisticated machine learning approaches used in building energy management employ, in most of the cases, post-hoc techniques to interpret their results such as LIME and SHAP [20]. Instead, regression-based approaches are selected for this study due to their intrinsic (ante-hoc) interpretability.

4. Methods

The model formulations used in this research have been developed for different applications, as indicated in Section 3, but they share a series of similarities that make them suitable for implementation in an integrated workflow, illustrated in Section 4.1. The integrated workflow starts with simple baseline regression models using energy signature as dependant variable and outdoor air temperature as independent variable, tested in the different monitoring periods reported later in Section 5. After that, the regression models are modified by means of additional variables to improve their goodness of fit. The simple baseline models are then used to test model retraining with a rolling horizon of 15 days of data and to check the outdoor air temperature response of time of week and temperature (TOWT) algorithm, which is reformulated and benchmarked against its original implementation. A regression model is used as well to characterise the performance of gas absorption heat pumps (GAHP), using an additional variable as well. Finally, criteria for model acceptability according to measurement and verification (M&V) guidelines are reported in Section 4.2.

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

In the preparation phase, before training models, data have been processed to remove missing values using interpolation, due to the minimal gaps found in measured data. Hourly outdoor air temperature and solar radiation data collected were already sufficiently accurate for this type of applications, while natural gas half-hourly data were cleaned by using a moving average filter aimed at de-noising (e.g. removing

Table 1
Nomenclature.

Variables and parameters Symbol	Quantity	Unit
a_0	regression coefficients heating component, intercept	kW
a_1	regression coefficients heating component, temperature dependence term	kW/K
a_2,a_3	regression coefficients heating component, solar radiation dependence and differenced temperature depence term	m^2,kW/K
a_j	regression coefficients, j is the hour of the week	kW
b_k	regression coefficients, k is the temperature segment	kW/K
c_o	regression coefficient TOWT, intercept	kW
d_0,d_1,d_2	regression coefficients Q_{th} model	kW,kW/K,kW/K
e_0,e_1,e_2	regression coefficients GUE_{max} model	-,1/K,1/K
Cv(RMSE)	coefficient of variation of RMSE	-
GUE	gas utilization efficiency	-
GUE_{max}	maximum gas utilization efficiency	-
gue	part load fraction of GUE	-
I_{sol}	total solar radiation on horizontal surface (direct and diffuse) average hourly value on monthly base	kW/m^2
L_{occ}	load when the building is occupied	kW
L_{unocc}	load when the building is not occupied	kW
MAPE	mean absolute percentage error	-
NMBE	normalized mean bias error (expressed in percentage)	-
Q_{NG}	natural gas energy	kWh
Q_{th}	thermal energy	kWh
q_h	energy signature heating (natural gas)	kW
R^2	determination coefficient (expressed in percentage)	-
$T(i)_k$	outdoor air temperature value for the segment k at time interval i, TOWT model	
$t_{ow,j}$	time of week binary variable, TOWT model	-
SE	standard error	kW,-
X_h	dummy variable (binary 0–1) heating	-
θ_e	outdoor air temperature	°C
$\theta_{e,diff}$	outdoor air temperature differenced	°C
$\theta_{e,sup}$	water heating system supply temperature	°C
ε	error term	kW

small oscillations around the actual value that depend on measurement errors) and aggregated at hourly interval. As introduced in Section 3, the simple regression models used as baseline are derived from ASHRAE 14:2014 [23] and consider outdoor air temperature θ_e as fundamental input but are then reformulated with the inclusion of additional variables.

Amongst the additional variables there is total solar radiation on horizontal surface I_{sol} and differenced daily average temperature $\theta_{e,\ diff}$ (i.e. the difference between outdoor air temperature in one day and the day before) for the reasons expressed in Section 3. Further, to be able to model piecewise linear shapes, dummy variables (0–1) such as X_h are used as an interaction term (i.e. multiplied by other variables θ_e, I_{sol}, $\theta_{e,\ diff}$) following the same rules and conventions illustrated in previous research [57].

These rules involve the selection of change-point temperatures within credible ranges (change-points represent the balance-point temperatures for the building [58,59] and are a function of building characteristics and operation strategies, so they are physically constrained) and the inclusion of a constraint in model output. In fact, energy signatures are positive quantities and using a piecewise linear formulation can lead to small negative values when signature is near 0, which have to be removed. The formulations of monthly and daily energy signature models are reported in Table 2.

In order to assess the degree of uncertainty that may be introduced when only short-term measures are available and the risk of overfitting, baseline models are tested with incremental re-training with a 15-days rolling horizon, for the reasons discussed in Section 3. This process is similar, in principles, to a cross-validation of models, where the entire dataset is subsetted into smaller datasets with 15 days of data each, and the individual models are then compared to the original model trained on the entire set.

Moreover, additional dummy variables are used also in time of week and temperature (TOWT) approach. TOWT has been used before by the authors to model electric load profiles in highly variable conditions [60]. However, in this paper a novel formulation is proposed and benchmarked with R package RMV2.0 [46] implementation. TOWT model formulation is simple in principle, as it is composed by two sub-models (piecewise linear), one for occupied L_{occ} (high demand) and one for unoccupied L_{unocc} (low demand) periods, as shown hereafter:

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

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

However, due to its structure, two strategies are needed to perform the temporal (Time Of Week) and temperature segmentation. The temporal segmentation is intrinsically dependant on the temperature segmentation as occupied/unoccupied periods are identified in an automated way, considering the electric load with respect to outdoor air temperature. In this research, as explained before, we focus on energy signature of natural gas consumption q_h instead of electric load.

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

Model name	Input variables	Interval	Formulation	
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)

Temperature segmentation is conducted in such a way that continuity in the temperature response computed by the model is preserved, using the algorithm reported in [43]; in this way constraints are imposed "indirectly" to retain continuity, while using a conventional regression algorithm, which doesn't include constraints. Additionally, temperature response is the point of contact between the regression models in Table 2 and TOWT, as will be shown graphically in Section 6.3 by means of energy signatures.

In Table 3 the fundamental characteristics of the approach originally implemented in RMV2.0 and the reformulation proposed in this research are indicated, summarizing the elements added or changed in each step of the modelling workflow.

Other software implementations of TOWT are available, for example the Python package Caltrack/OpenEEmeter [50] and the R package NMECR [48], reported before in Section 3. The latter provides the possibility to define temperature change-points arbitrarily as well as temperature segments with an equal amount of data points (i.e. based on quantiles of data), but keeps the same weighting approach using days as hyper-parameters, as in RMV2.0, while Caltrack/OpenEEmeter employs a monthly weighting approach.

In the last step of the research workflow, the GAHP performance is characterised using *GUE* (Gas Utilization Efficiency), which is the parameter describing its thermal conversion efficiency. *GUE* is defined as the ratio between the delivered thermal energy and the fuel energy (and substantially similar to the Coefficient of Performance for an electric heat pump) and *gue* correspond to its Part Load Fraction (*PLF*) as follows:

$$GUE = \frac{Q_{th}}{Q_{NG}} \tag{7}$$

$$gue = \frac{GUE}{GUE_{max}} \tag{8}$$

where Q_{th} and Q_{NG} are respectively the heat output (kWh) and gas energy input (kWh) of the period under consideration, *gue* is the ratio between the actual GUE value in part-load conditions and the maximum value measurable in steady-state conditions (i.e. in standardized test conditions). Q_{th} and Q_{NG} can be expressed also in terms of power (kW), by dividing the quantity by the time interval (hours in this case). The energy content of the gas is calculated with the measured gas volume (m^3) and lower heating value (kWh/m^3). In this case, our goal is to create a simple approach to predict Q_{th} and *GUE* in variable operational conditions (i.e. realistic operation), thereby obtaining a model useful for energy management purpose. First, GAHPs performance data (i.e. Q_{th} and *GUE*) available from technical datasheets are interpolated using the approach explained in EN 15,316–4–2:2017 [61]. Technical data sheets report the results of steady-state test conditions at full load [7]; the gap between GUE_{max} (determined by the regression) and the actual measured *GUE* is discuss later in relation to metered energy demand. The input data for the two multivariate regression models (one for Q_{th} and one for GUE_{max}) are θ_e outdoor air temperature (°C) and θ_{sup} supply temperature of the heating loop (°C), while the output data are Q_{th} expressed as thermal power output (kW) and *GUE*, defined before. The models are formulated as follows:

$$Q_{th} = d_0 + d_1\theta_e + d_2\theta_{sup} \tag{9}$$

$$GUE_{max} = e_0 + e_1\theta_e + e_2\theta_{sup} \tag{10}$$

4.2. Statistical criteria for model acceptability according to M&V principles

As introduced in Section 3, Measurement and Verification (M&V) approaches at the state-of-the-art [23–25] provide thresholds of acceptability for the models using statistical indicators such as Normalized Mean Bias Error, *NMBE*, and Coefficient of Variation of Root

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

Workflow step	Component	TOWT – RMV2.0	TOWT – reformulation
Data preparation	Temporal segmentation	Time of week variable ($t_{ow,j}$) is a binary variable (or dummy variable), *n-1* is the number of hours of the week (i.e. 168–1 = 167), the last term (168th) is included in the intercept term c_0.	Time of week variables for Holidays are modelled as Sundays.
	Temperature segmentation	Temperature variable $T(i)_k$ is a continuous variables with arbitrary temperature scale (Celsius or Fahrenheit), *m* is the number of segments chosen when binning the temperature data (i.e. *m-1* change points for the piecewise linear function used to represent the temperature response component). The default change points are (40, 55, 65, 80, 90 F/ 4.4, 12.8, 18.3, 26.7, 32.2 °C). At least 1 change point should be present and at least 20 data point above the highest and below the lowest change-point need to be included. Temperature segmentation is performed as described in [43]. The same temperature segments are used for occupied/unoccupied modes.	Temperature variable $T(i)_k$ is computed with the same algorithm as in the original implementation [43], but the change points are derived from the analysis of the distribution of outdoor air temperature, creating an even number of segments, with the central change-point corresponding to the median of data (50 percentile). Depending on the distribution of data, temperature segments may have different widths below/above the median value. At least 6 segments are suggested. The same temperature segments are used for occupied/unoccupied modes.
Model training	Detection of occupied/ unoccupied hours (high/ low demand)	The occupied/unoccupied periods are detected by running a regression model with two variables Heating Degree-Days (HDD) and Cooling Degree-Days (CDD) and an intercept term. The time of the week period which are underpredicted for a certain percentage of time are assumed to be occupied. The default percentage is 65 (threshold equal to 0.65 in model settings). Degree-days are computed with base temperatures of 50 and 65, respectively for heating and cooling.	The occupied/unoccupied periods are detected by running a regression model with respect to temperature using the temperature segmentation criterion proposed above. The threshold considered is equal to 0.65 as the default in the original implementation.
	Overall model	The overall model is the weighted sum of models generated in multiple runs, which are dependant on the hyper-parameter choice.	The overall model is created as the sum of one regression model for occupied periods, one model for unoccupied periods and a model for residuals, considering only Time-Of-Week (TOW) dependence.
	Hyper-parameters	One hyper-parameter is present, representing the time scale, expressed in days, of the weighting function used to combine different regression models created.	One hyper-parameter is present, expressed in weeks or months, to determine the additional temporal segmentation needed to model residuals as a TOW function, without temperature response.
Visualization and interpretation of results	Temperature dependence	Temperature dependence of load is shown as a scatterplot.	Energy signature is plotted with respect to outdoor air temperature at different time intervals (monthly, daily, hourly). The energy signature interpretation is used to derive additional insights and to compare the piecewise linear temperature response obtained by TOWT reformulation with the ones obtained using other regression-based approaches.
	Time series	Time series of measured data are plotted with respect to predicted ones.	No changes.
	Weekly patterns	Weekly patterns of operations are shown with a 2D heatmap.	No changes.
	Residuals	Residuals are plotted in time and with respect to outdoor air temperature.	No changes.
	Measured vs predicted	A scatterplot of measured vs predicted data is used to highlight possible deviations.	No changes.

Mean Square Error, *CV(RMSE)*, summarized hereafter in Table 4. In this research monthly, daily and hourly data are used and thresholds of acceptability for daily data is assumed to be in the range between monthly and hourly, with a simple average between the two values provided by ASHRAE Guidelines 14:2014 [23], as proposed by Meng et al. [62]. The coefficient of determination R^2 (defined in the range between 0 and 100%, or 0–1, with a strong correlation when greater than 75%), is considered as well, because it is reported for example in ISO 50,006:2014 [27], but its limitations have to be acknowledged. In fact, its limitation reside in the fact that R^2 is inherently related to the slope of the model (e.g. the dependence on θ_e, I_{sol}, $\theta_{e,\ diff}$ for the models in Table 2). Higher slopes have a tendency to show a higher value of R^2 even though the variance of the predicted variable is the same. As a result, when evaluating a model, it is critical to use other metrics in addition to R^2, but it is reported for the sake of completeness.

5. Case study description

As explained in the introduction, Hale Court is a sheltered housing development run by Portsmouth City Council, built in 1984. Hale Court consists of 80 flats with a mix of studio, 1 bedroom, 2 bedroom and 3-bedroom flats. The building complex is composed of two blocks, North and South, and essential data are summarized in Table 5, while the plan is reported in Fig. 1.

Hale Court has been refurbished during the EU Horizon 2020 project THERMOSS [22]. The retrofit intervention has been concentrated on two energy efficiency measures in particular Smart Thermostatic Radiator Valves (TRVs) and Gas Absorption Heat Pumps (GAHPs) that have been installed, contextually to a general upgrade of the plant room, with the installation of a monitoring system. The periodization of the retrofit intervention and monitoring strategy is reported in Table 6. In this research, the analysis refers to the North block, where the savings due to both Smart TRVs and GAHPs can be assessed. More in detail, the

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

Interval	Metric	ASHRAE Guidelines 14/Meng et al.
Monthly	*NMBE*	±5
	Cv(RMSE)	15
Daily	*NMBE*	±7.5
	Cv(RMSE)	22.5
Hourly	*NMBE*	±10
	Cv(RMSE)	30

refurbishment of heating system for the North Block plant room has involved:

- keeping the existing gas boilers;
- removal of the original domestic hot water heaters;
- addition of a cascade of 3 GAHP;
- incorporation of thermal storage;
- reconfiguration of piping layout.

The sizes of the heat generators installed in the plant room are reported in Table 7.

From the point of view of control, the strategy adopted is fixed point with compensation (but no-weather compensation) so the hot water temperature in the primary loop oscillates between 60 and 65 °C in operation. Such a high temperature is necessary because, as it can be seen in the scheme in Fig. 2, it has to serve both the space heating and the domestic hot water demands for the dwellings. The terminal units of heating system are radiators. The impossibility to reduce supply temperature is a major limitation to the achievement of the full efficiency potential for the heating system and GAHP technology, as will be discussed later in Section 6.4.

6. Results and discussion

Following the application of the methods introduced in Section 4 to the case study described in Section 5, in this section multiple model implementations are discussed. In Section 6.1, regression models at the state of the art (which represent the most consolidated techniques, used for measument and verification) are used as baseline and then modified with the inclusion of additional independent variables, assessing their performance improvement. After that, in Section 6.2, the robustness of baseline models for the different periods of monitoring (reported in Table 6) is tested by re-training them with a rolling horizon of 15 days. Then, in Section 6.3 a reformulation of the Time Of Week and Temperature (TOWT) approach is proposed and benchmarked against the original formulation. Finally, the GAHP performance is characterised using regression as well in Section 6.4, providing an additional validation to the savings estimated. Overall, the results indicate how it is possible to provide an integrated workflow, starting from a simple baseline model and further research potential is described in Section 6.5.

6.1. Testing regression models θ_e, $\theta_e I_{sol}$, $\theta_e \theta_{e,diff}$, and $\theta_e I_{sol} \theta_{e,diff}$

The initial steps of the research are presented hereafter, indicating how baseline models, originally included in ASHRAE 14:2014 [23] and modified as shown in literature [57], can be used as a starting point to enable a meaningful comparison. Then additional independent variables are added to the models, following the logic described in Section 3 and 4.

The first model tested is model θ_e from Table 2 and the test is developed using natural gas demand signature as predicted variable, because the electricity demand is practically constant before and after retrofit, as can be clearly seen in Fig. 3. In this figure, there is no temperature dependence for the operation of auxiliaries, i.e. they are running practically in a constant way (on average) during the year and they are not analysed further. Nonetheless, the application of regression-based techniques to highly variable electric load profiles has been presented in previous research [60].

Table 5
Hale Court building blocks data.

Total floor area	North block: approximately 1300 m^2 South block: approximately 2100 m^2
Number of flats	North block: 32 South block: 48
Type of flats	Studio, 1 Bed, 2 Bed and 3 Bed
Year of construction	1984
Resident demographics	Sheltered housing, residence for elderly citizens

Instead, if the slope obtained by the energy signature model fitting in Fig. 4 for natural gas demand is considered, it is possible to see how the heating system capacity installed is oversized compared to the actual needs, even when extrapolating for very low temperatures.

The change-points of the piecewise linear regression model in the different monitoring periods are in the range 19.5–20.0 °C.

These values are much larger than the one normally encountered in buildings [62], but there reasons behind this high value is the very high average temperatures in living rooms and bedrooms, in the range 22.9–23.5 °C. It is worth noting that the base temperatures used for the normalization of energy demand in Display Energy Certificates (DEC) is 15.5 °C and the one reported in technical standardization for the calculation of the part load ratio of heat pump operation [63] is 16 °C; both temperatures are very far from the measured one in this case and this confirms the importance of developing tailored regression models that reflect the actual behaviour of a building.

The constant part of the natural gas signature (above the change-point temperature) represents the average demand for domestic hot water (DHW), as it corresponds to summer months when there is no heating demand and just DHW demand. In Table 8 the results of natural gas demand signature analysis are reported, indicating in general a good fit of the models in all the three periods, interpreting them in light of the thresholds of acceptability reported in Table 4, R^2 and Mean Absolute Percentage Error, *MAPE*. The Normalized Mean Bias Error *NMBE* is 0 in the cases with daily data. This is due to the fact that regression techniques tend to have a value equal to 0 because of their formulation (minimization of sum of the squares of the differences between measures and model predictions). The more relevant statistical parameter in this case is *CV(RMSE)*, considering also the limitations of R^2 reported before in Section 4.2.

Figs. 5 and 6 depict the results of the regression predictions plotted in time (with monthly and daily intervals, together with the deviations) for the entire monitoring period (33 months) for both buildings. It can be seen that, despite their simplicity, the models can fairly accurately approximate the temporal pattern, with only a few points where the difference between measured and predicted consumption is significant.

The period when the heating plant was retrofitted with the installation of GAHPs represents the data gap. Furthermore, more days were excluded from the data cleaning process in the time series with daily data due to incomplete records. The next step involved the analysis of deviations between measured data and model prediction, to verify if there is a recurring pattern in them. In this case there is no apparent pattern in deviations.

Finally, the time series of differences between the measured and predicted (by the models fitted in the previous period) natural gas consumption was computed using the method described in ISO 50,006:2014 [27]. According to this standard a "baseline" model is trained first and the used for performance verification in the "reporting" period. The same data are then aggregated using cumulative sum to obtain the total amount saved (i.e. the actual savings achieved with respect to period 1 and period 2 operation).

From the analysis of the time series of savings and cumulative savings reported in Fig. 7, it can be seen how during the pandemic period (the last part of the monitoring period) domestic hot water usage increased, probably due to an higher demand of water for cleaning and sanification purposes. In Table 9 the calculation of the performance improvement is reported, expressed in terms of percentage of energy demand reduction compared to a baseline value in the different monitoring periods, described in more detail in Table 6.

To obtain an appropriate comparison [27] in typical climate conditions, the energy demand as been calculated on a yearly basis, using the piecewise linear regression models developed and a standard weather data file for the location (weather normalization), considering the same domestic hot water demand of the pre-pandemic period (normalization of user behaviour).

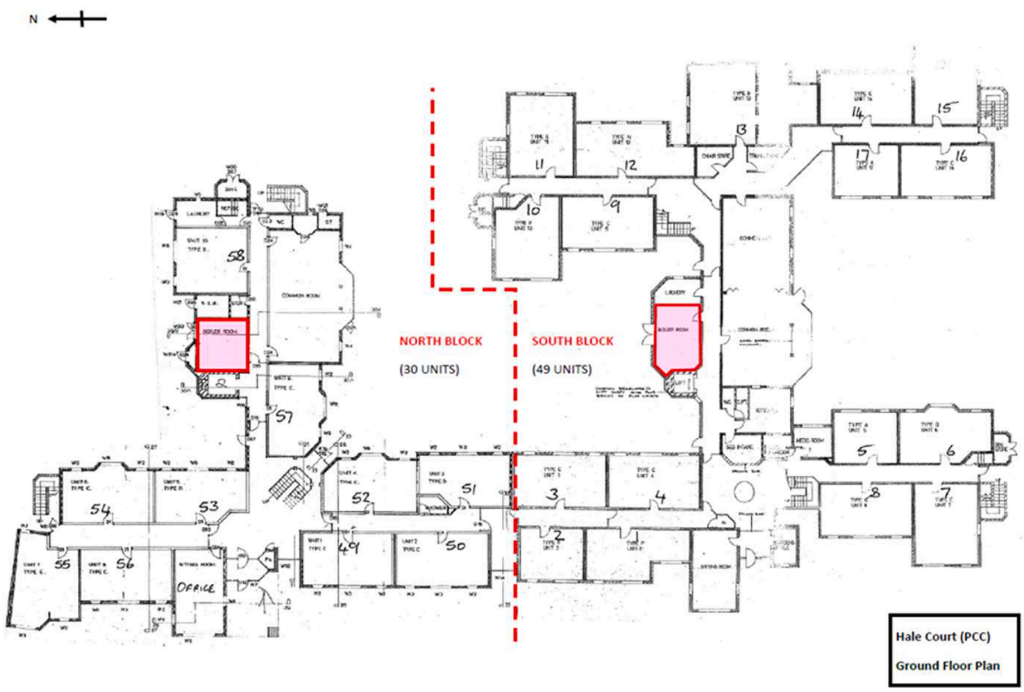


Fig. 1. Hale Court – North Block/South Block - Floorplan Ground Floor.

Table 6
Hale Court retrofit intervention periodization.

Period		Building blocks	Dates (monitoring phases)
1	Before any intervention	North block/ South Block	From Dec 2017 to Dec 2018
2	Post first intervention & Pre second intervention	North block/ South Block	Jan 2019 – July 2019 Smart Thermostatic Valves installation in January 2019. January and February 2019 present anomalies in energy consumption
3	Post second intervention (split into two parts, commissioning and actual operation)	North block	Phase 1: Plant room upgrade and GAHP installation, July to Sep 2019. Phase 2: Commissioning of GAHP, Sep 2019 Phase 3: October 2019/January 2020 the system is running with GAHP and boilers Phase 4: February 2020 just GAHP for testing purpose

Table 7
Heat system generators data.

Technology	Variable	Unit	Value
Gas Boilers	Model type	–	REMEHA GAS 110 Eco
	Thermal power output	kW	115 × 2
	B oiler max. design temperature	°C	80
	Typology	–	Condensing
Gas Absorption Heat Pump	Model type	–	GHP AWO 38
	Thermal power output A7W50	kW	114.9 (3 appliances cascade configuration)
	Gas Utilization Efficiency A7W50	–	1.52
	Natural gas power input A7W50	kW	75.6
	M ax. heating water flow temperature	°C	65
	P ermissible Ambient Temperature	°C	−20 to +40

On the one hand, the impact of smart TRVs (7.2% reduction of natural gas demand) is modest but essentially in line with values found in literature, which indicates respectively a range 5–10%, determined by simulation in the UK residential context [10], and 7.1–23.3%, measured empirically in Poland with a long-term field evaluation [11]. Following the comparison with other case studies, it appears evident that the saving potential of TRVs has not been exploited completely in this case and their operational settings can be improved. This fact has been clarified by a more in depth data analysis of operational patterns conducted within the same project, which is not presented in this case as it goes beyond the scope of this paper. In general, these results confirm the need of evidence on the impact of smart controls in the residential sector [9] and an adequate consideration of their cost-benefit ratio. The energy demand reduction determined by Smart TRVs and GAHPs together is 25.8% and the relative improvement determined by GAHPs is 20.0%. This aspect will be analysed more in detail in Section 6.4, where regression-based technique is used for GAHP performance characterisation and validation of energy savings.

The model testing procedure is continued by including additional variables which lead to models $\theta_e I_{sol}$, $\theta_e \theta_{e,diff}$, and $\theta_e I_{sol} \theta_{e,diff}$, defined in Table 2 in Section 4.1. As shown by the results in Tables 10, 11, and 12 for model $\theta_e I_{sol}$, $\theta_e \theta_{e,diff}$, and $\theta_e I_{sol} \theta_{e,diff}$ respectively, the addition of variables improves the statistical indicators more or less uniformly (as it would be normally expected). However, the improvements in R^2, *MAPE*, and *CV(RMSE)* in this specific case is insignificant, justifying the use of model θ_e as a reference for the digital twin implementation due to its simplicity. Clearly, in cases of buildings with high solar gains and/or high thermal inertia (following the arguments reported in Sections 3 and 4), models with additional variables may guarantee a significant improvement of the predictive ability.

6.2. Testing regression model θ_e with rolling horizon retraining

In this section, the results obtained by fitting model θ_e using daily data interval are presented but, differently from Section 6.1, a rolling

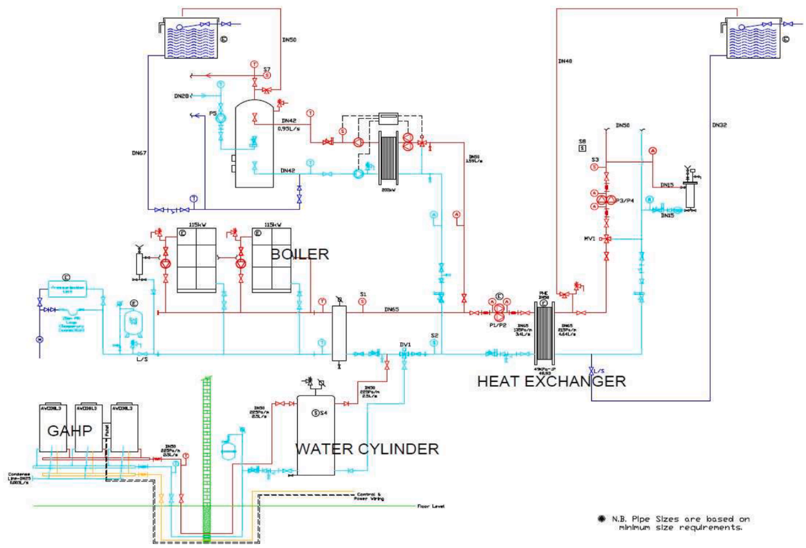


Fig. 2. Schematic Layout of Heating System in Hale Court North.

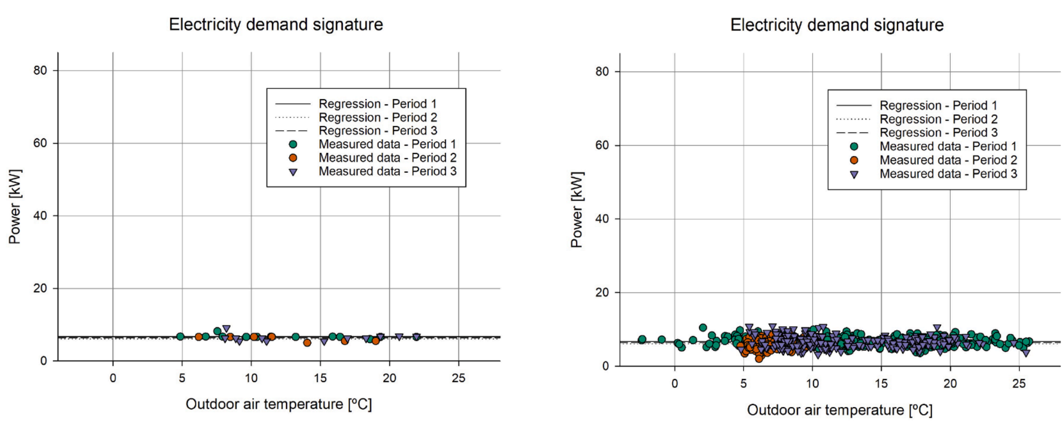


Fig. 3. Electricity demand signature – monthly (left) and daily (right) data.

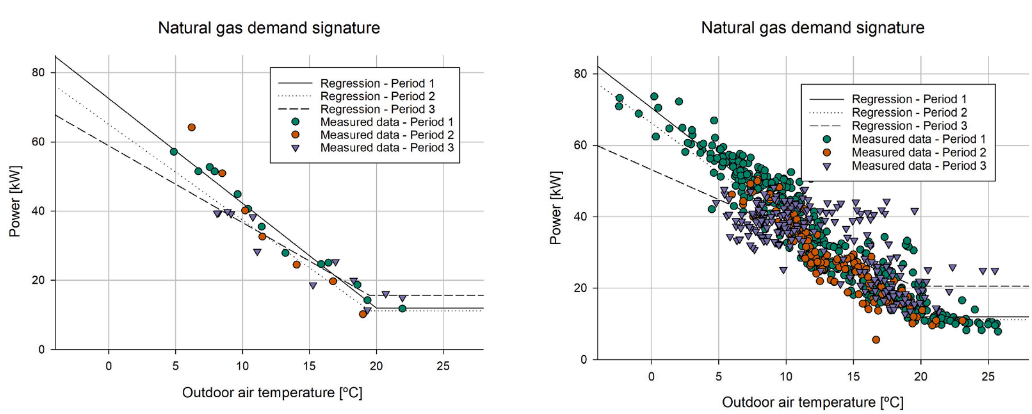


Fig. 4. Natural gas demand signature – monthly (left) and daily (right) data.

horizon of 15 days is chosen, following the considerations described in Section 3. The results of the series of θ_e models fitted is compared to the θ_e model fitted to the entire series of data and reported in Section 6.1. This modelling approach has, in principles, two main objectives.

The first one is reproducing the use of regression in a "digital twin" fashion where model coefficients are periodically recomputed and model is recalibrated. The second one is performing a cross-validation of models' results for the different monitoring periods, subsetting the entire dataset into individual sets of 15 days of data, for which the variability of coefficients' estimates is analysed and the potential over-fitting risk is evaluated.

The daily ranges were selected with a maximum temperature of 16 °C to accurately estimate the heating dependant part of the model, despite the fact that the balance point for the building is in the range

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

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

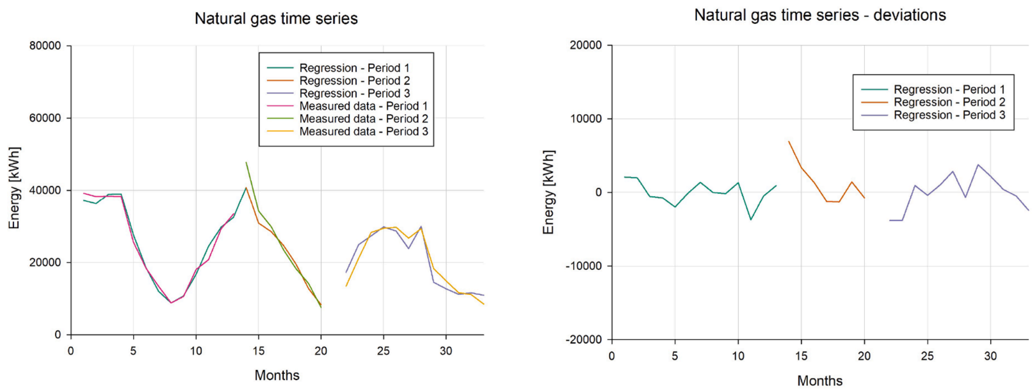


Fig. 5. Time series of natural gas demand (left) and deviations (right) – monthly data.

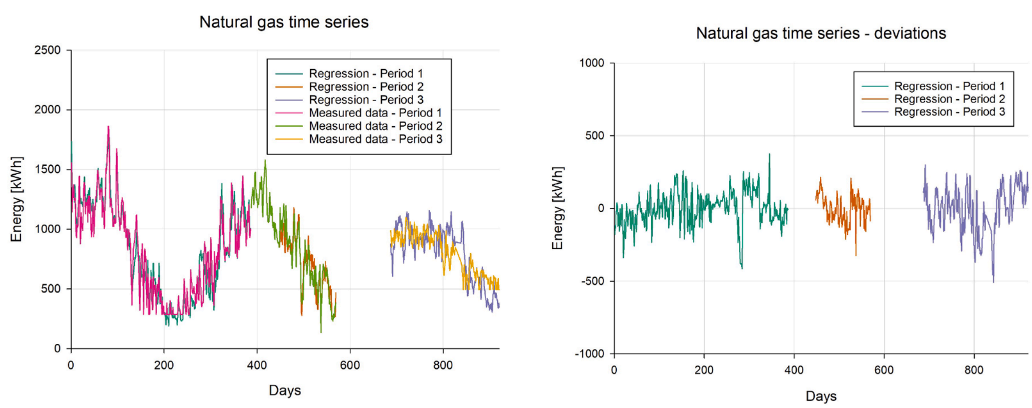


Fig. 6. Time series of natural gas demand (left) and deviations (right) – daily data.

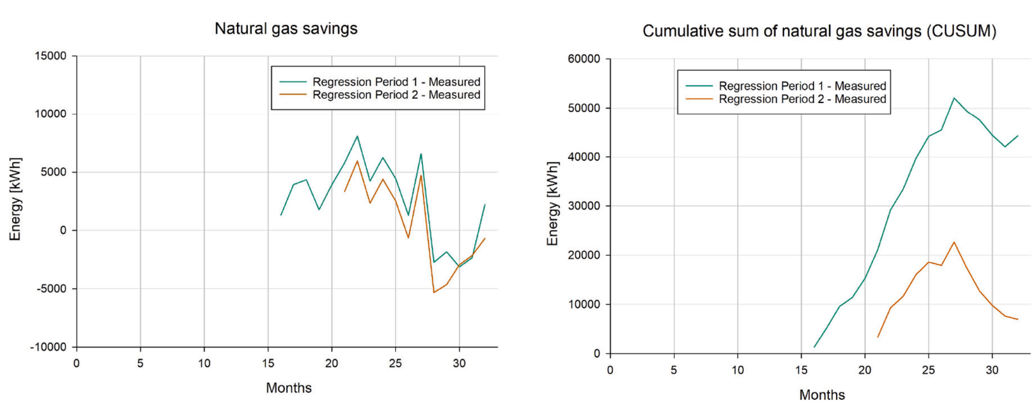


Fig. 7. Time series of natural gas savings and cumulative sum of savings.

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

Period	Description	Overall reduction	Relative reduction
1	Before retrofit	0 (baseline)	–
2	TRVs installation	7.2%	0 (baseline)
3	GAHPs and plant room upgrade	25.8%	20.0%

19.5–20.0 °C (points which are near to the change-point would give more unstable estimates). Because of the possibility of exploiting its approximated physical interpretation (as in co-heating tests, for example), the slope of the regression for heating demand is analysed (parameter a_1 for the models reported in Table 2 in Section 4). The results in Table 13 show that, on average, the coefficients' estimates obtained with the daily θ_e model fitted to the entire dataset are not very different from the ones obtained with rolling horizon model retraining (subsets), indicating the effectiveness of the approach.

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

As indicated in Table 3, a major difference between the state-of-the-art implementation of TOWT in RMV2.0 and the reformulation proposed resides in the fact that temperature is subsetted with respect to the median (i.e. 50 percentile of data) and that the width of the temperature segments is dependant on the range (min-max) of the available data (may be different in the left and right side of the temperature distribution). The alternative implementation of TOWT proposed in this research makes use of a piecewise linear function to distinguish between periods of high and low consumption, denoted as occupied and unoccupied hours in RMV2.0, which uses instead a regression with respect to Heating Degree-Days (HDD) and Cooling Degree-Days (CDD). This formulation, employing energy signature, enables the comparison between piecewise linear models in Table 2 (in this section only model θ_e is considered for simplicity) and the component of TOWT model which represents the outdoor temperature dependence of load/signature.

Separated signatures are fitted for periods of high consumption (flagged as occupied in the original formulation) and low consumption (flagged as unoccupied in the original formulation). As can be seen in Figs. 8, 9 and 10 on the left-hand side, for the three periods of monitoring, there is a good agreement between the piecewise linear temperature-dependant component of the TOWT and the model θ_e, with some exceptions during summer when the TOWT fits better the actual conditions and in winter when temperature is low and there is reduction of the slope, compared to the simplest regression model. Overall, the model reformulation proposed fits data well and within the limits for model acceptability given by ASHRAE14:2014 and reported in Table 4.

Indeed, the scalability of energy signature visualisation as a function of outdoor air temperature enables the comparison of model estimates from monthly, to daily and hourly data, for example in Figs. 4 (monthly and daily interval) and Fig. 8-10 (hourly interval). This feature is particularly relevant if there is the necessity to compare long-term/low-

Table 13
Results of regression model θ_e with digital twin approach – Estimates of the slope of regression model, when θ_e >16 °C, parameter a_1.

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

Table 10
Results of analysis for natural gas demand signature – Model $\theta_e I_{sol}$.

Interval	Period	Energy indicators		Statistical indicators			
		Energy measured (M) kWh	Energy predicted (P) kWh	R^2 %	*MAPE* %	*NMBE* %	*CV(RMSE)* %
Monthly	1	332,722	332,692	97.95	4.96	−0.01	6.33
	2	93,466	93,457	99.79	8.19	−0.02	5.54
	3	242,586	242,772	93.25	10.13	0.08	10.07
Daily	1	320,182	320,182	91.81	13.13	0.00	13.58
	2	84,955	84,955	86.79	12.00	0.00	13.25
	3	177,111	177,111	55.99	18.35	0.00	19.55

Table 11
Results of analysis for natural gas demand signature – Model $\theta_e \theta_{e,diff}$.

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

Table 12
Results of analysis for natural gas demand signature – Model $\theta_e I_{sol} \theta_{e,diff}$.

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

frequency measures and short-term/high frequency measures in a meaningful way. Further, there are multiple insights that can be derived by interpreting energy signature slopes and change-points, as discussed in Section 3.

Finally, in numerical terms, the reformulation of TOWT fits data well for the selected periods, as indicated in Table 14. In this case the Mean Absolute Percentage Error (*MAPE*) indicator is excluded because hourly time series may have data that are near or equal to 0 (i.e. if the system is turn-off for a period of the day) and *MAPE* is not an appropriate indicator.

The results achieved are very similar to the ones which can be achieved with TOWT implementation in RMV2.0 with 15 days hyper-parameter, as reported in Table 15, even though the weighting approach implemented in RMV2.0 determines a much larger number of model runs, reported in Table 16, in particular if the hyper-parameter is set to 1 day. In fact, the RMV2.0 implementation requires an initial run to distinguish between occupied and unoccupied hours using a degree-days model, after which the time series is segmented based on the hyper-parameter to produce a matrix of coefficients used to weight multiple regressions.

Instead, the proposed reformulation requires as well one initial model run to fit the piecewise linear temperature component required to distinguish between high (i.e., occupied) and low (i.e., unoccupied) consumption hours. After that however, two model runs are required to determine the piecewise linear component for high/low consumption hours, as depicted in Figs. 8-10 on the left, and finally a single model run is required to develop a Time Of Week (TOW) model for the residuals. The residuals are differentiated by month in this case, as this is the hyper-parameter used in the new formulation proposed. The lower amount of models fitted aims to reduce clearly the computational effort but also the risk of model overfitting and limiting the number of model parameters. A weighting approach with hyper-parameters expressed in months is present in the Caltrack/OpenEEmeter [50] implementation of TOWT, which keeps however a modelling approach for temperature segmentation similar to RMV2.0 [46].

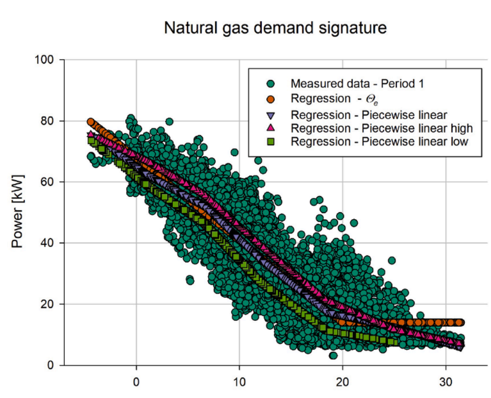
6.4. *Testing regression model for GAHP performance characterization in variable conditions*

In this section the energy demand reduction determined by the GAHPs (20.0% as indicated in Table 9 in Section 6.1) is analysed more in depth, considering the actual physical behaviour of the heat pump technology. In Table 17 the results obtained by fitting the two models reported in formulas 9 and 10 in Section 4.1 are reported, respectively for Q_{th} and GUE_{max}.

It can be observed that, while being simple, models fit data well with a high value of the coefficient of determination R^2 and a small standard error for both Q_{th} and GUE_{max}.

The results of model predictions for Q_{th} and GUE are then plotted in Fig. 11 (on left and right side respectively) against the calculated GAHP data, determined by means of interpolation, using data from technical datasheets. The charts indicate also the single regression lines for each supply temperature (50, 55, 60, 65 °C) and it can be seen how the slopes are just moderately different each other; this justifies the creation of a single regression model with two input variables θ_e and θ_{sup}. The model applicability can be extended further by means of dummy variables (i.e. to handle change points and non-linearity), in analogy with the models presented in Table 2. This is not required, however, in this case.

In a research published by Schmitt-Gehrke et al. [8] multiple types of GAHPs were tested in partial load conditions, calculating the *gue*, whose values range typically between 0.80–0.90 This parameter accounts for part load performance degradation and is defined similarly to Part Load Fraction (*PLF*) for electrical heat pumps, as explained before at the end of Section 4.1. The values found in this study are coherent with the ones found in that research. If, for the sake of verification and in a very simplistic way, the average outdoor air temperature conditions of 14 °C is assumed (for monitoring period 3) with a supply temperature in the range 60–65 °C (given the fixed point control and the absence of weather compensation) a GUE_{max} of around 1.4 is obtained, looking at Fig. 4. Then, if a *gue* equal to 0.85 is considered (within the range specified above from literature focused on part load GAHP testing), the final corrected *GUE* is around 1.2, which is essentially compatible with the measured performance improvement achieved by the GAHP (20.0% compared to condensing boilers) because the existing condensing boilers, in these operating conditions, have an efficiency that is near to 1 (0.97–0.98, thereby a *GUE* of around 1.2 is compatible with a 20% savings). In this case study building in fact, condensing boilers are not working in condensing mode due to the high temperature of supply (around 60–65 °C) as reported in Section 4. Clearly, a more in depth analysis requires the correct weighting of *GUE* with respect to outdoor air temperature and load conditions, with a method substantially similar to the one specified in technical standardization [63,64] for the calculation of Seasonal Coefficient of Performance (*SCOP*) for electrical heat pumps and with the correction for part load performance in the different conditions. Nonetheless, the method presented in this research provides a quick and simple way to verify performance, that gives results similar to the ones obtained in other studies where a 10 to 15% improvement of the annual heating efficiency for GAHP with respect to other options was found [65]. This percentage of performance improvement with respect to natural gas boilers is fundamentally constrained by the supply temperature of the heating system, as can be seen from the ranges of *GUE* values presented in Fig. 11. In other words, in buildings with higher supply temperatures the relative performance improvement is necessarily smaller than in buildings with lower supply temperatures. This fact is important and needs to be considered when planning a retrofit intervention because optimistic assumption on GAHP performance may



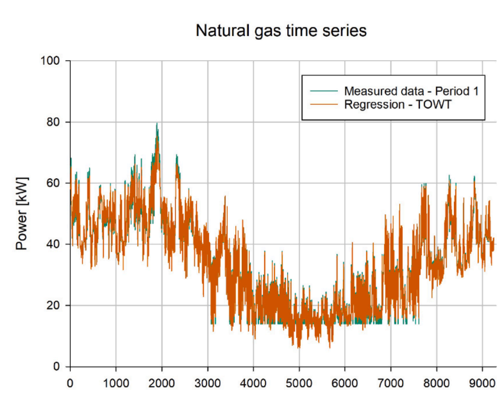
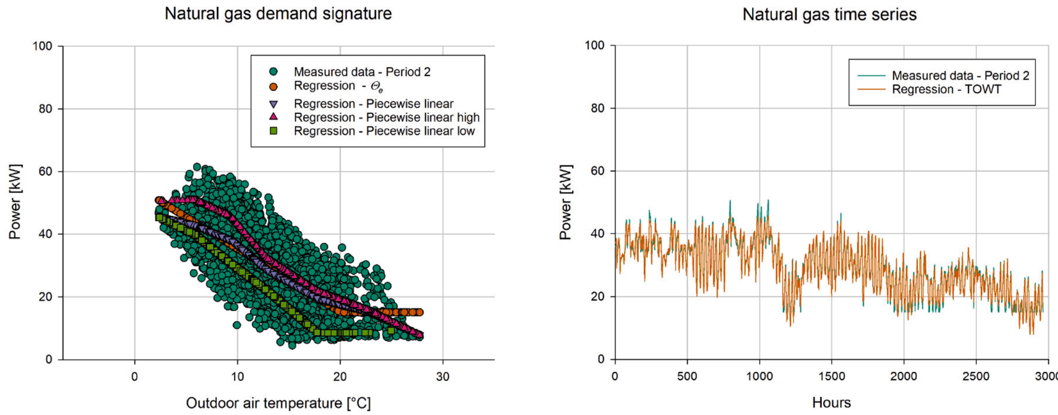


Fig. 8. Natural gas demand signature (left) and time series (right) with hourly data, TOWT model – Period 1.



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

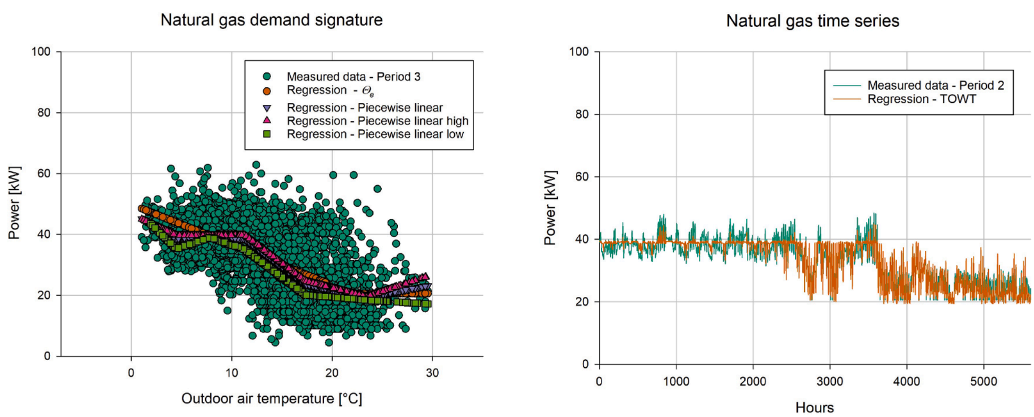


Fig. 10. Natural gas demand signature (left) and time series (right) with hourly data, TOWT model – Period 3.

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

Period	Energy indicators		Statistical indicators		
	Energy measured (M) kWh	Energy predicted (P) kWh	R^2 %	*NMBE* %	*CV (RMSE)* %
1	319,408	319,408	91.65	0.00	14.70
2	86,772	86,772	88.72	0.00	13.01
3	186,636	186,636	66.56	0.00	17.04

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

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

Period	Days	TOWT – RMV2.0			TOWT - reformulation
		Hyper-parameter 30 days	15 days	1 day	Hyper-parameter 1 Month
1	386	14	27	387	4
2	124	6	10	125	4
3	234	9	17	235	4

Table 17
Regression model R2 and Standard Error (SE).

Metric	Q_{th}		GUE_{max}	
	Unit	Value	Unit	Value
R^2	%	97.3	%	97.4
SE	kW	2.21	–	0.03

compromise the return of the investment, which depends on the actual energy savings achieved.

6.5. Further research potential

Multiple areas of development are possible starting from the results presented and discussed in Section 6. The first area of improvement involves the integration of the different techniques in a seamless workflow where temporal scalability is exploited effectively. For example, using daily data it becomes possible to understand if there are

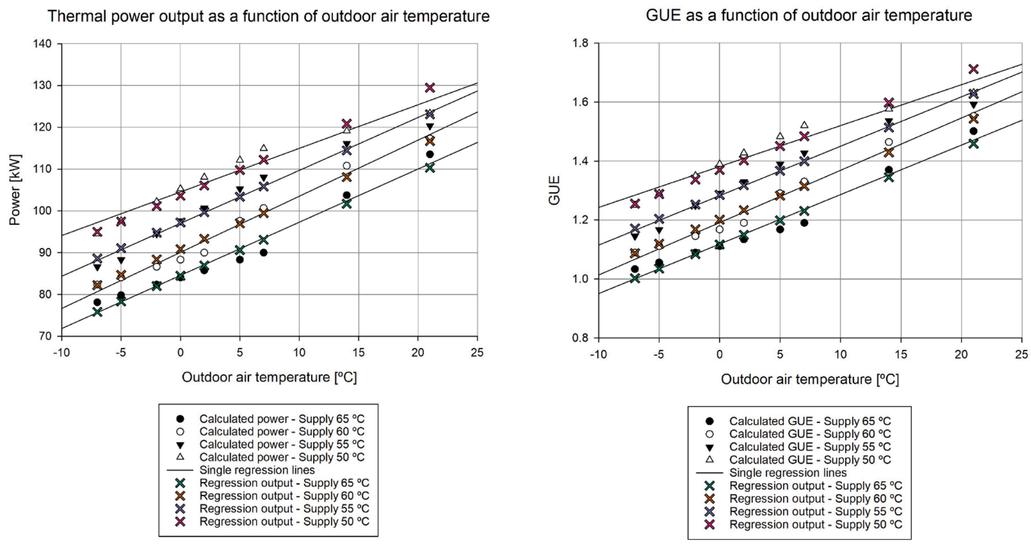


Fig. 11. Regression model for the calculation of thermal power output Qth and GUEmax for the GAHP in variable conditions.

particular weekly patterns (i.e. difference between operational regimes and occupancy in different days of the week, or if there is a significant difference between weekdays and weekend). Then, using hourly and sub-hourly data it become possible to characterize the building dynamic thermal behaviour and operating schedules; this can help optimise operational settings. In general, the presence of both short-term/high-frequency measurements (daily, hourly, sub-hourly) and long-term/low-frequency ones (monthly, across multiple years) can greatly improve the robustness of the performance assessment and optimise the effort spent in the analysis process. Overall, further potential research developments are summarised in Table 18, considering the areas and specific applications. As can be seen in the table, the use of approximate physical interpretation is a relevant element which helps in setting appropriate boundaries and constraints in the modelling process, while providing a greater sense of trust in models' outputs. Further, this peculiar characteristics of interpretable methods can help bridge the current state of the art research on whole-building energy model calibration [6] by means of a multi-level process [66], where lumped physical properties are identified [67] and used in simplified physical models to optimize operation and enhance building flexibility [68] dynamically.

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

Research area	Applications	Potential development
Modelling techniques	Automated modelling workflow	Data pre- and post-processing can be developed further. An automated forward model selection (i.e. including additional variables incrementally) can be combined with cross-validation.
	Regression-based formulation leveraging approximated physical models	Building balance-point temperatures correspond to change-points in piecewise regression and slopes are closely related to overall heat transfer coefficient.
Building technologies	Whole-building performance analysis/model calibration	Approximate physical models and analytical formulations can be used to define boundaries of acceptability and constraints for model parameters, leveraging energy balance at multiple levels in the building system.
	Construction technologies, air-tightness and ventilation	Overall heat transfer coefficient of building zones can be identified via regression.
	Technical systems	Building balance-point temperature (change-points of piecewise regression) influences heat pump weather compensated control and efficiency (COP, GUE).

7. Conclusion

In this research, multiple formulations of lean and interpretable (regression-based) "digital twins" for building energy performance monitoring have been proposed. First of all, potential improvements with respect to model formulations at the state-of-the-art have been shown with the use of additional variables. Then, the robustness of models has been tested by retraining them with a 15 days rolling horizon, indicating a reasonable level of stability of models' coefficients estimates. Further, a reformulation of the time of week and temperature (TOWT) model has been proposed, which achieves good performance with an enhanced interpretability, using energy signature visualisation. Finally, GAHPs performance has been characterised using regression as well.

Overall, it was shown how interpretable regression-based approaches can be used effectively at multiple temporal scales of analysis and, therefore, how they can work reasonably well even with limited information such as monthly data, making them suitable for quick and inexpensive performance assessment (i.e. using utility bills and outdoor air temperatures). Nonetheless, models' performance can be improved by increasing the data granularity from monthly to daily and then to hourly/sub-hourly and by including additional variables. In particular, the use of hourly/sub-hourly data makes it possible to characterize the building dynamic thermal behaviour (e.g. outdoor air temperature response) and operating schedule (e.g. time of week model component) and this, in turn, can be used to optimize operational settings. Finally, by exploiting model temporal scalability it becomes possible to alternate short-term/high-frequency measurements (daily, hourly, sub-hourly) and long-term/low-frequency ones (monthly, across multiple years) in an optimal way.

In the monitoring process, energy savings were normalised by weather and operational conditions to verify them. Smart TRVs reduced energy demand by 7.2% and TRVs and GAHPs combined reduced it by 25.0%. TRVs savings were in the lower part of the expected range (5–20% see Section 3), but better system tuning could increase this value. As indicated in Section 6.4, the GAHPs' relative energy demand reduction of 20.0% is realistic given the monitoring period's operating conditions. GAHP technology may be a viable alternative to conventional and condensing boilers, but it would need to be run on hydrogen or synthetic fuels to meet low-carbon emission targets and this represents a factor of uncertainty in a future perspective.

Possible research developments include both the enhancement of the modelling procedure described in this paper and its implementation in a particular set of applications. The first aspect involves the seamless incorporation of modelling techniques into a workflow that aims to provide both numerical and visual analytics through the use of forward model selection and cross-validation procedures (to enhance robustness of estimates and reduce overfitting risk). The second aspect involves the formulation of regression models based on an approximated physical interpretation (i.e. a "grey-box" approach), which can provide additional insight into the performance of the building as a whole (including supporting whole building energy model calibration), as well as construction technologies, technical systems, and operational strategies. In fact, despite the fact that these modelling strategies may be less accurate and powerful than other machine learning approaches at the state of the art, an approximate physical interpretation is relevant in light of the interpretability objective, which is also at the core of this research, because the analytics derived from models can be more easily understood in human terms, ideally promoting a greater sense of trust in the models' output.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

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

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 the uniform methods project : methods for determining energy efficiency savings for specific measures. Contract 2013;303:275–3000.
[5] Investor Confidence Project (https://europe.eeperformance.org/), accessed 31/08/2020.
[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.
[8] Schmitt-Gehrkea P., Buchina O., Cigandaa J.L.C., Grafa R., Kühnb A., Zieglera F. Part load performance of gas fired absorption heat pumps. 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 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] Interpretable machine learning, Section 3.2 Taxonomy of Interpretability Methods, Christopher molnar (https://christophm.github.io/interpretable-ml-book/taxonomy-of-interpretability-methods.html), accessed on 24/05/82023. n.d.
[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 (https://energy.soton.ac.uk/project/thermoss), accessed 22/04/2021.
[23] ASHRAE, ASHRAE Guideline 14-2014. Measurement of energy, demand, and water savings; american society of heating. Atlanta, GA, USA: Refrigerating and Air-Conditioning Engineers; 2014. 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). n.d.
[25] FEMP, M&V guidelines: measurement and verification for performance-based contracts, version 4.0 2015.
[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 2014.
[28] ISO 16346:2013, Energy performance of buildings — assessment of overall energy performance 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.
[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, DE Claridge. 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 D.E. 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. 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. 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.
[44] Borgeson SD. Targeted efficiency: using customer meter data to improve efficiency program outcomes. Berkeley: University of California; 2013.
[45] EVO's Advanced M&V Testing Portal (https://mvportal.evo-world.org/).
[46] RMV2.0 - LBNL M&V2.0 Tool (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 (https://kw-labs.github.io/nmecr/).
[49] OpenEEmeter (https://github.com/openeemeter/eemeter).
[50] CalTRACK. CalTRACK Methods (http://docs.caltrack.org/en/latest/methods.html).
[51] EENSIGHT (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.
[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.
[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.
[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.
[65] Kühn A, Graf R, Ciganda JC, Schmitt-Gehrke P, Gassenmeyer F, Blum N, et al. Betriebsstrategien f{ü}r 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.