

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

Data-driven building energy modelling - An analysis of the potential for generalisation through interpretable machine learning

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Massimiliano Manfren, Patrick AB James, Lamberto Tronchin (2022). Data-driven building energy modelling - An analysis of the potential for generalisation through interpretable machine learning. RENEWABLE & SUSTAINABLE ENERGY REVIEWS, 167, 1-13 [10.1016/j.rser.2022.112691].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/890245> since: 2022-07-09

*Published:*

DOI: <http://doi.org/10.1016/j.rser.2022.112691>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Manfren M., James P.A., Tronchin L., Data-driven building energy modelling – An analysis of the potential for generalisation through interpretable machine learning, (2022) Renewable and Sustainable Energy Reviews, 167, art. no. 112686

The final published version is available online at: <https://doi.org/10.1016/j.rser.2022.112691>

### Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# Data-driven building energy modelling – An analysis of the potential for generalisation through interpretable machine learning

Massimiliano Manfren<sup>a</sup>, Patrick AB James<sup>a</sup>, Lamberto Tronchin<sup>\*b</sup>

<sup>a</sup> Faculty of Engineering and Physical Sciences, University of Southampton,  
Boldrewood Campus – SO16 7QF Southampton, United Kingdom

<sup>b</sup> Department of Architecture (DA), University of Bologna, Via Cavalcavia 61 47521,  
Cesena, Italy

## Abstract

Data-driven building energy modelling techniques have proven to be effective in multiple applications. However, the debate around the possibility of generalisation is open. Generalisation involves the ability of a machine-learning model to adapt to previously unseen data and perform in a satisfactory way. Besides that, while machine-learning techniques are extremely powerful, interpretability, i.e. the ability for humans to predict how the model output will change in response to a change in input data or algorithmic parameters, is essential to attain a "human-in-the-loop" approach and creating feedback loops aimed at continuous improvement of efficiency measures in buildings.

A flexible regression-based approach is developed and tested on a Passive House building in this study. The formulation employs dummy (binary) variables as a piecewise linearization method, and the rules for creating them are explicitly stated to ensure interpretability. Furthermore, the possibility of automating the model selection process using statistical indicators is described, including specific indicators used in Measurement and Verification (M&V) for the acceptance of calibrated energy models.

The valuable insights that can be found using data-driven methods are reported and discussed, emphasising limitations and constraints, as well as the potential for future research focused on systems of (interpretable data-driven) models that can exploit the techniques' spatial and temporal scalability. Finally, the physical interpretation of model coefficients and the analytical formulations for energy model decomposition can be used to supplement the scalability of data-driven techniques and create more sophisticated systems of interconnected models.

**Keywords:** Data-driven energy modelling, Interpretable machine-learning, Regression-based approaches, Generalisation, Building energy modelling, Measurement and Verification, Energy Analytics.

## Highlights:

- Data-driven techniques can be used effectively in building energy modelling.
- Generalisation and interpretability are crucial to improve their applicability.
- An interpretable data-driven modelling approach is tested on a Passive House building.
- Rules for modelling workflow automation are presented.
- Indications for further developments of systems of models are reported.

---

\* Corresponding author:

Tel: +39 0512093384; fax: +39 0512093384

e-mail: lamberto.tronchin@unibo.it

## 1 Nomenclature

Table 1: Nomenclature

Variables and parameters		
Symbol	Quantity	Unit
$a_0, b_0, c_0$	regression coefficients, intercept	kW
$a_1, b_1, c_1$	regression coefficients, temperature dependence term	kW/K
$a_2, b_2, c_2$	regression coefficients, solar radiation dependence term	m <sup>2</sup>
$Cv(RMSE)$	coefficient of variation of RMSE	-
$I_{sol}$	total solar radiation on horizontal surface (direct and diffuse) average hourly value on monthly base	kW/m <sup>2</sup>
$MAPE$	mean absolute percentage error	-
$NMBE$	normalized mean bias error (expressed in percentage)	-
$q_h$	energy signature heating	kW
$q_b$	energy signature base load	kW
$q_c$	energy signature cooling	kW
$R^2$	determination coefficient (expressed in percentage)	-
$X_h$	dummy variable (binary 0-1) heating	-
$X_b$	dummy variable (binary 0-1) base load	-
$X_c$	dummy variable (binary 0-1) cooling	-
$\theta_e$	outdoor air temperature	°C
$\varepsilon_h$	error term heating	kW-kWh
$\varepsilon_b$	error term base load	kW-kWh
$\varepsilon_c$	error term cooling	kW-kWh
$\sigma$	standard deviation	kWh

## 2 Introduction

Data-driven building energy modelling approaches, using machine-learning techniques, have proven to be effective in multiple applications across building life cycle phases [1], from design [2] to operation [3]. Their use for energy performance benchmarking can provide highly valuable insights to help reducing energy consumption and related carbon emissions, both for newly constructed buildings and for existing or retrofitted ones. However, the debate around the possibility of generalisation of data-driven approaches is still open [4]. The term generalisation indicates the ability of a machine-learning model to adapt to a new dataset (i.e. previously unseen data) and perform reasonably well, considering statistical performance indicators.

In fact, while a large number of energy prediction techniques have been proposed in research in the last decades, yet there is still no general consensus on which techniques perform better in which specific problems in the energy field (e.g. load prediction, measurement and verification procedures, anomaly detection, data-driven predictive control, etc.). Furthermore, the presence of a wide range of building construction typologies, technologies, and end-uses (e.g., residential or non-residential) in the building stock results in a highly heterogeneous set of design and operational conditions.

In general, multiple “performance gaps” (i.e. differences between expected and measured performance) can be present in buildings and adequate analysis frameworks should be set up [5]. At the same time, there is a problem of literacy around this topic [6] and the need for robust enquiry and repeatable scientific methods when addressing building performance [7]. Essentially, “performance gaps” can involve all the phases of the building life cycle [8] and building energy performance analysis requires a careful consideration of both human and technical factors [9]; the use of standardized assumptions in modelling is not sufficient.

For this reason, building energy model calibration techniques has been reviewed extensively for example by Coakley et al. [10] and, more recently, Chong et al. [11]. In the latter paper, the input-output analysis process performed for building model calibration is discussed in detail, clarifying the difference between variables (i.e. model states that evolves during simulation) and parameters (i.e. quantities that describe the property of objects in the model).

Going back to parameters and variables in building energy modelling, the two most common input variables found in the systematic review by Chong et al. [11] are outdoor air temperature and solar radiation. The most important is outdoor air temperature, as evidenced also by the widespread adoption of methods based on temperature response functions [12].

In this sense, a typical example of generalisation can be the use of a model trained with a certain weather dataset and then tested with another weather dataset, to understand the impact of weather variability (e.g. normalization of energy demand with respect to weather data). This is, for example, the counterfactual approach used in Measurement and Verification (M&V) protocols, which would be recalled later. According to this approach energy savings are calculated as the difference between the energy used after a certain point in time (e.g. an intervention of the building, a change in operation, etc.) and a baseline estimate. This type of generalisation has limited impact because the model only applies to the specific building, but its value can be derived from the ability to compare the (case specific) model parameters to those of buildings with similar characteristics.

While the use of case studies in building performance research is a common practice, the creation of machine learning models trained on datasets that are too small and too case specific has questionable value when looked from the point of view of applications at large scale in the built environment. What appears more interesting is the potential to extend the applicability of models at larger scales (e.g. neighbourhood, cities, building stock) and to extract additional insights regarding the performance of building technologies and people behaviour.

Aside from generalisation, the interpretability and explainability of machine learning models are also to be considered. We can find definitions in recent standardization focused on Artificial Intelligence (AI) software testing [13], but they are concept important and “slippery” at the same time [14].

Interpretability refers to the ability to predict what will be the model output given a change in input data or in algorithmic parameters. In other words, human can understand the rationale behind model output and the algorithmic logic can be inspected. Further, in some cases it is even possible to discern a potential causality in the relationship among variables (normally only correlation and not causation can be proved) when more information is available. Interpretable models are, for example, linear multivariate regression or regression and decision trees, where the impact of regression coefficients (in the case of linear regression) or rules (in the case of trees) can be easily understood by humans.

On the other hand, explainability refers to the extent to which the internal mechanics of a machine-learning algorithm can be explained in human terms. High performance machine learning models such as neural network and random forests can be explained, but they are not generally interpretable and, for this reason, they are frequently indicated with the term “black-box”. Therefore, interpretability, while representing a constraint from the point of view of the modelling technique selection (in the sense that it reduces the potential spectrum of techniques that can be used), can provide advantages in terms of simplicity and transparency. These advantages have been discussed in recent research [15]. Further, there is a need to open the “black-box” of energy modelling [16] to make it easily applicable.

Returning to the generalisation issue, large scale data collection and analysis using both top-down [17] and bottom-up approaches [18] in an integrated way is crucial to obtain robust evidence about building performance at large scale. From this point of view, important initiatives have been conducted in recent years such as the Building Data Genome Project [19] and the Building Data Genome Project 2 [20] (with energy meter data from the ASHRAE Great Energy Predictor III competition).

In general, the use of crowd-sourced building data [21] is an important element to promote advances in building performance research based on open data [22]. However, open-data on measured performance are not generally available on a large scale base and large scale data collection requires relevant investments. Furthermore, there is a practical requirement to integrate various types of data when analysing building performance [23] in order to enable effective benchmarking at the level of single technologies, entire buildings, and building stock.

As a result, harmonised and (temporally and spatially) scalable techniques are critical when generating estimates that can be used not only for single cases but also for large-scale analysis. Together, these two features can help overcoming at least partly the limitations inherent to the use of small datasets and/or individual case studies previously mentioned, by enhancing the comparability and transparency of the modelling approach. These aspects are discussed in this research, where a flexible data-driven energy model formulation is presented and tested on data from a Passive House residential building case study. The model formulation is interpretable and conceived to be used in a general way for monthly energy data analysis. Further, it can be extended to other time intervals (daily, hourly, sub-hourly) and spatial scales [15], linked to standardized analytical formulations [24] and applied across different life cycle phases [3], as shown in previous research.

The models are trained with different datasets, using electric and thermal energy data measured during a building monitoring period of 3 years, and their goodness of fit is discussed by means of statistical indicators. Finally, indications of future research directions in the broad area of data-driven energy modelling are given in relation to research outcomes.

### **3 Background on interpretable data-driven methods**

In this research the use of interpretable techniques and harmonized procedures is the starting point. The concept of interpretability was discussed previously in the introduction, while the term harmonized indicates procedures that are codified, like the ones proposed by Measurement and Verification (M&V) protocols such as ASHRAE 14:2014 [25], Efficiency Value Organization (EVO) [26], Federal Energy Management Program (FEMP) [27], as well as other technical standards reported later in this Section. The procedures and techniques proposed in these protocols have been adopted, for example, in projects such as the Uniform Methods Project (UMP) [28], aimed at providing robust and empirically grounded techniques to benchmark energy efficiency measures with a uniform approach, and the Investor Confidence Project (ICP) [29], aimed at de-risking investment in energy efficiency and increasing trust by private and public investors. Further, the applicability of these techniques can be extended to electric load profiles analysis and prediction at large scale [30], with the possibility to analyse Demand Response events [31] as well.

As reported in the introduction, weather variables such as outdoor air temperature and solar radiation are essential inputs for model calibration [11] and an example of largely diffuse interpretable approach is the variable-based degree-days regression, originally proposed by Kissock et al. in the Inverse Modeling Toolkit (IMT) [32], which has been included in

ASHRAE 14:2014 [25] and has been evolving steadily with different algorithmic formulations for the automatic the selection of the change-point model [33] and explicit solution of the three-parameter (heating and base load) linear change-point model [34]. The definition of degree-days is itself part of international standardization [35]. For these reasons, interpretable machine-learning techniques are compatible with the energy performance analysis approach proposed by ISO 50006:2014 [36] and with energy signatures model formulation, defined in ISO 16346:2013 [37]. The analysis of model outputs can be used as a feedback in a continuous improvement logic, as indicated in ISO 50001:2018 [38].

Energy signature is obtained by dividing the energy demand with respect to the amount of operating hours in the time interval considered in the analysis, i.e. determining an average power over it. This data transformation presents some advantages related to the approximated physical interpretation of regression coefficient [23] and its scalability for building stock analysis [24].

Additionally, the use of outdoor air temperature and solar radiation as inputs can be found in interpretable regression-based approaches such as co-heating test [39] as well. Co-heating test is based on building zone thermal balance approximation and can be used to provide reliable evidence on the actual performance of building fabric [40]. Clearly, the uncertainty impact determined by unmonitored energy sources and sinks [41] must be considered in order to enable the correct quantification of thermo-physical properties (due to the use of energy balance as the underlying principle), within the general problem of energy balance decomposition.

In this respect, it is possible to find examples of building energy balance decomposition using regression-based approaches in research studies focused on multi-scale analysis of thermo-physical properties of buildings [42], robust energy model calibration [43] and energy analytics for building decarbonisation [24]. All these regression-based approaches can be complemented by the use of Statistical Process Control (SPC) techniques [44] for the graphical identification of performance anomalies [45], coherently with the indications of ISO 50006:2014 [46].

Essentially, interpretable regression-based methods can be based on general piecewise linearization methods [47] and use dummy (binary variables) to handle non-linearities (as will be illustrated in Section 4.1 for this research). In this way, they can become highly versatile [48] and be used for a variety of purposes, including the dynamic hourly and sub-hourly modelling of electric load profiles, as shown by the Time Of Week and Temperature (TOWT) model formulation aimed at analysing electric load shapes and variability [30], quantifying changes in electricity use in Demand Side Management (DSM) events [31] and improve efficiency program outcomes [49].

The combination of simple yet powerful interpretable data-driven methods makes it possible to envision systems of interconnected models [50] that can act as “digital twins”, sharing essential open data and standards [22].

A "digital twin" is a digital representation of a process that can overcome the limitations of traditional simulation-based engineering approaches (such as those used in building performance simulation) by demonstrating how a process performs in real time and under real-world conditions. Interpretability is useful (from the standpoint of the user) because it allows the user to easily predict what will happen when the model input is changed. The more "transparent" the model is to the user, the more it can encourage a "human-in-the-loop approach" that "black-box" (non-interpretable) techniques cannot.

## 4 Methods

In previous research work conducted on the case study building [51], which will be illustrated in more detail in Section 5, it was demonstrated that 2 years of monthly monitoring data were required to achieve calibration according to M&V procedures (the calibration process was performed incrementally) and that the solar radiation variable played an important role in model performance improvement. At that point, separate models for heating, cooling and base load were fitted, as in research focused on energy model decomposition [43] and calibration via regression [52].

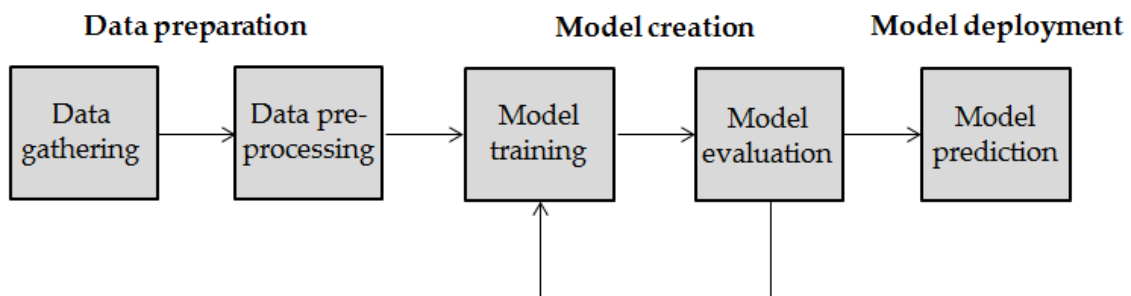
Instead, in this research the separate sub-models are combined into a single model (which is the sum of the individual sub-models) by introducing additional variables (dummy, 0-1 binary variables) to the original datasets using rules that are explained in Section 4.1. The rules themselves have a physical meaning and are meant to retain the ability to connect the two approaches in the future (i.e. piecewise linear regression and analytical formulations of building energy balance). In this way, energy balance decomposition and identification of thermo-physical properties (depending on the variables monitored) may be enabled by a unique integrated workflow.

The performance of models is tested by fitting them to multiple datasets of measured electric and thermal demand in the building. The details of the model formulation and calibration criteria are reported in Sections 4.1 and 4.2, respectively; the characteristics of the case study and datasets used in this research are then described in Section 5.

### 4.1 Model formulation

While the main goal of this research is to model the building's energy behaviour using data-driven techniques, energy signatures (i.e. energy demand divided by the number of operating hours in the time interval considered, expressed as an average power) are used in the workflow because of their advantages in terms of approximate physical interpretation, as discussed previously in Section 3. As a result, measured energy data (data gathering) are pre-processed to obtain energy signatures and dummy variables are included (data pre-processing), then models are trained on energy signatures (model training) and evaluated against statistical indicators of goodness of fit (model evaluation); finally, data are post-processed for graphical visualization purpose and both energy and energy signatures are plotted in time and with respect to outdoor air temperature (i.e. the fundamental input variable). The overall workflow is summarized in Figure 1 and explained hereafter in more detail.

Figure 1: Diagram of the modelling workflow



The goal of the formulation presented is to obtain a multi-output regression model that can then evolve and incorporate other analytical formulations to enable a more in-depth (physics-informed) analysis of the model coefficients, following the arguments presented in



Section 3 and recalled at the beginning of this Section. To this end, the different individual regression models are combined using dummy variables (0-1, binary variables) [47] in order to obtain a single model that can be used in a flexible way for all the datasets of monitored energy data, reported in Section 5.

The rules to create dummy variables (and including them as additional variables in the datasets) are summarized in Table 2 and illustrated more in detail later in this Section with respect to model formulation.

*Table 2: Rules for dummy (binary) variables creation*

<b>Rule</b>	<b>Description</b>	<b>Variables</b>
1	If the energy demand is greater than 0 for the corresponding sub-model (e.g. heating, cooling or base load), then dummy variable is equal to 1.	$X_h, X_b, X_c$
2	If the outdoor air temperature is lower than balance point temperature for heating (i.e. heating base temperature), the dummy variable for heating is equal to 1.	$X_h$
3	If the outdoor air temperature is greater than balance point temperature for cooling (i.e. cooling base temperature), the dummy variable for cooling is equal to 1.	$X_c$
4	All the dummy variables (that partition heating, cooling and base load demands) should be coherent with the schedules of operation for building services (i.e. months of heating and cooling system operation).	$X_h, X_b, X_c$
5	The dummy variables for base load are assumed to be 1 in all the months (i.e. electricity and hot water demand are always present).	$X_b$

Because of the sub-metering of energy demand (i.e. the subdivision by type of end-use), the information to derive the dummy variables, using the rules reported in Table 2, was available in this research, but rules 2 and 3 can be applied even when this subdivision is unknown, due to the physical meaning of balance point temperatures for heating and/or cooling (i.e. base temperatures [53] for variable-base heating [54] and cooling degree-days [55] calculations). Further, rule 4 can be inferred as well due to typical operation strategies of buildings.

In other words, it may not be possible to attribute precisely from the very beginning dummy variables (to partition the dataset for the heating, cooling and base load demand) if only total electricity demand is present (and sub-metering data are not available); nonetheless, the search for the best fitting model can be performed iteratively (as shown in Figure 1 for model creation), using ranges of balance point temperatures for heating and cooling (dependent on building characteristics [54]) and using contextual information to determine the schedules of building operation. After that, the performance of multiple models can be ranked with respect to the statistical indicators that are used later to evaluate models' acceptability (thresholds for model calibration, proposed by protocols and standards). Finally, the models' outputs are constrained to be positive (energy metering data are positive quantities) as in some cases the output of the regression model can assume a small negative value, using change-point models.

Indeed, the fundamental reason behind the creation of this model formulation is the possibility to use a single and flexible model for the analysis of the energy demand at different levels in the building (corresponding to the different datasets reported in Section 5) in a multi-input/multi-output fashion and then proceed by stepwise regression with backward elimination of variables (i.e. reducing the amount of variables) to obtain the final model version, i.e. automating the model selection process using rules and statistical indicators to rank their goodness of fit. Indeed, the form of the final model can be constrained in such a way that an approximated physical interpretation of the model coefficients is retained (to enhance transparency and enable further decomposition) [24], coherently with the arguments presented in Section 3.

As specified before, the regression model formulation used in this research corresponds to the extension of the one previously tested in relation to multi-scale analysis of building performance [42] and energy analytics that can be used to support the decarbonisation of built environment [24].

The overall model is the sum of three sub-models for heating, base load and cooling, and dummy variables are used to exclude the part of the model which are not pertinent (e.g. thermal demand in the case of only heating or only cooling component). One dummy variable is used for each sub-model, namely  $X_h$  for heating,  $X_b$  for base load and  $X_c$  for cooling. Finally, the intercept of the complete model is set to 0 (i.e. regression through the origin) and the output constrained to be positive. The dependent variables are the energy demand data reported in Section 5 (electric and thermal) and the two independent variables are outdoor air temperature  $\theta_e$  and total solar radiation on horizontal surface  $I_{sol}$ , which represent the two most common variables considered in model calibration [11]. Tables 3 and 4 report the formulas for Type 1 model (only outdoor air temperature  $\theta_e$  as input variable) and Type 2 model (outdoor air temperature  $\theta_e$  and solar radiation on horizontal surface  $I_{sol}$  as input variables), respectively.

*Table 3: Regression sub-models for heating, cooling and baseline demand analysis, combined by means of dummy variables – model Type 1*

Demand	Sub-models
Heating	$q_h = X_h(a_0 + a_1\theta_e) + \varepsilon_h$ (1)
	$q_h = a_0(X_h) + a_1(X_h\theta_e) + \varepsilon_h$ (2)
Base load	$q_b = X_b(b_0 + b_1\theta_e) + \varepsilon_b$ (3)
	$q_b = b_0(X_b) + b_1(X_b\theta_e) + \varepsilon_b$ (4)
Cooling	$q_c = X_c(c_0 + c_1\theta_e) + \varepsilon_c$ (5)
	$q_h = c_0(X_c) + c_1(X_c\theta_e) + \varepsilon_c$ (6)

*Table 4: Regression sub-models for heating, cooling and baseline demand analysis, combined by means of dummy variables – model Type 2*

Demand	Sub-models
Heating	$q_h = X_h(a_0 + a_1\theta_e + a_2I_{sol}) + \varepsilon_h$ (7)
	$q_h = a_0(X_h) + a_1(X_h\theta_e) + a_2(X_hI_{sol}) + \varepsilon_h$ (8)
Base load	$q_b = X_b(b_0 + b_1\theta_e + b_2I_{sol}) + \varepsilon_b$ (9)
	$q_b = b_0(X_b) + b_1(X_b\theta_e) + b_2(X_bI_{sol}) + \varepsilon_b$ (10)
Cooling	$q_c = X_c(c_0 + c_1\theta_e + c_2I_{sol}) + \varepsilon_c$ (11)
	$q_h = c_0(X_c) + c_1(X_c\theta_e) + c_2(X_cI_{sol}) + \varepsilon_c$ (12)

Model type 1 is the sum of the three sub-models in formulas 2, 4 and 6, while model type 2 is the sum of the three sub-models in formulas 8, 10 and 12. The formulas 2, 4, 6 and 8,10, 12 are chosen (instead of 1, 3, 5 and 7, 9, 11 respectively) because they represent the regression models created using the dummy variables where the new variables are  $X_h$ ,  $X_h\theta_e$ ,  $X_hI_{sol}$ , etc. In other words, the dummy (binary) variables are used to create interactions among variables, instead of relying just on the original variables  $\theta_e$  and  $I_{sol}$ .

#### 4.2 Model calibration criteria

In this Section the calibration criteria for model acceptability are introduced. Following the indications proposed by Measurement and Verification (M&V) protocols at the state-of-the-art such as ASHRAE 14:2014 [25], Efficiency Value Organization (EVO) IPMVP [26], and Federal Energy Management Program (FEMP) [27] (cited previously in Section 3), the thresholds of acceptability for regression models as calibrated with monthly data and hourly data (for the sake of comparison) are reported in Table 5.

Table 5: Thresholds of acceptability for M&V models as calibrated with monthly and hourly data

Interval	Metric	ASHRAE Guidelines 14	IPMVP	FEMP
Monthly	NMBE	±5	±20	±5
	Cv(RMSE)	15	-	15
Hourly	NMBE	±10	±5	±10
	Cv(RMSE)	30	20	30

The statistical indicators chosen to enable the comparison between measured and predicted data (obtained by regression) for the monitoring period are  $R^2$  (Coefficient of determination),  $MAPE$  (Mean Absolute Percentage Error),  $NMBE$  (Normalized Mean Bias Error) and  $Cv(RMSE)$  (Coefficient of Variation of Root Mean Square Error). While the first two indicators ( $R^2$  and  $MAPE$ ) are general statistical indicators for the evaluation of goodness of fit of models,  $NMBE$  and  $Cv(RMSE)$  are statistical indicators considered for the specific purpose of model calibration and, for this reason, proposed by the M&V protocols ASHRAE 14:2014, FEMP and EVO-IPMVP reported above in Table 5.

Nonetheless, the coefficient of determination  $R^2$  (defined in the range of 0-100 percent, or 0-1), is a suitable indicator as reported in ISO 50006:2014 [36], even though the limitation that stems from the fact that  $R^2$  is inherently related to the model's slope (i.e. the dependence on input variables) have to be acknowledged. Higher slope values will correspond to higher  $R^2$  values even with the same predicted variable variance. An  $R^2$  value greater than 80 % indicates a good model fit and 75 % is indicated in IPMVP Guidelines for Assessing Uncertainty as a reference value [56].  $MAPE$  is not indicated specifically in model calibration procedures but it is included in this research to give an idea of what is the percentage of variation (in absolute value) of the error term with respect to the measured data.

Overall, the statistical indicators reported are used to provide thresholds for the acceptability of models, but they can be complemented by the visual analysis of modelling results, up to the creation of control charts [45], used in Statistical Process Control (SPC) [44], as shown at the end of Section 6.1.

## 5 Case study description

The case study chosen is a Passive House residential building located in the Province of Forli-Cesena, in the Emilia Romagna Region of Italy. For three years, the case study building was monitored and data from energy meters and weather stations were collected. The case study building is characterized by highly insulated envelope components built in accordance with Passive House standards. Further, the building's technical systems include mechanical ventilation with heat recovery (MVHR), a ground-source heat pump system (GSHP) for heating, cooling and DHW services, a photovoltaic system for on-site electricity generation, and a solar thermal system with storage, to integrate DHW production. The number of occupants of the building is 5. Table 6 summarizes the most important building data.

Table 6: Building design data assumptions

Group	Type	Unit	Design
Geometry	Gross volume	m <sup>3</sup>	1557
	Net volume	m <sup>3</sup>	1231
	Heat loss surface area	m <sup>2</sup>	847
	Net floor area	m <sup>2</sup>	444
	Glazed area/total wall area ratio percentage	%	22,5
	Surface/volume ratio	1/m	0,54
Envelope	U value external walls	W/(m <sup>2</sup> K)	0,18

	U value roof	W/(m <sup>2</sup> K)	0,17
	U value transparent components	W/(m <sup>2</sup> K)	0,83
HVAC and DHW	Ground-source heat pump (GSHP) - Brine/Water Heat Pump (B0/W35)*	kW	8,4
	Borehole heat exchanger (2 double U boreholes)	m	100
On-site energy production	Building Integrated Photo-Voltaic (BIPV) - Polycrystalline silicon	kW <sub>p</sub>	9,2
	Solar thermal - Glazed flat plate collector	m <sup>2</sup>	4,32
	Domestic hot water storage	m <sup>3</sup>	0,74

\* EN14511 test condition in heating mode, brine at 0 °C and water 35 °C with supply-return temperature difference  $\Delta t = 10$  °C.

Previously, research was conducted on this case study, with the goal of testing the use of regression methods to aid in the progressive calibration of a Resistance-Capacitance (RC) energy model [51]. The deviations between measured and simulated energy demand were used to control the progressive model calibration process with the aid of regression models. In that case, models were meant to define a data envelopment of possible operational energy performance, approximating simulation output. It was shown how 2 years of data were necessary to reach calibration and how the design assumption were to be revised in order to match model prediction with measured data.

Unlike previous research on this case study, a novel model formulation is tested in this case, which has been presented in Section 4.1. The monitoring period chosen for model training and calibration is 3 years and the measured energy demand dataset are reported in Table 7.

*Table 7: Datasets used in research*

Dataset	End-use	Interval	Monitoring period
Electric energy	Total	Monthly	3 years
	HVAC, DHW	Monthly	3 years
	Appliances and lighting	Monthly	3 years
Thermal energy	Heating	Monthly	3 years
	Cooling	Monthly	3 years

## 6 Results and discussion

This study aims the test, as introduced in Section 4.1, a novel formulation of the regression-based approach used in previous research. This formulation is proposed with the goal of incorporating it into an interpretable data-driven building energy modelling workflow that can supplement other state-of-the-art techniques and tools and evolve potentially into a system of interconnected models, leveraging the fundamental features presented in Section 3. Section 6.1 discusses the results of model training from both a visual and numerical standpoint, emphasizing the importance of the "human-in-the-loop" approach for machine learning methods. After that, in Section 6.2, the limitations of the techniques at this stage are discussed and finally future research possibilities are presented in Section 7.

### 6.1 Model training and calibration

The data-driven energy modelling process is presented in this Section, beginning with regression models fitted to energy signature data, for the various types of energy demands considered in this research, namely:

1. total electricity demand;
2. electricity demand for HVAC and DHW;
3. electricity demand for appliances and lighting;

4. thermal demand for heating;
5. thermal demand for cooling.

As introduced in Section 3, energy signature is calculated by dividing the original energy data by the amount of operating hours in the interval considered, thereby determining an average power value for the interval of analysis (monthly in the case, for 36 months, 3 years). Then, measured and predicted energy demand is compared by multiplying the energy signature by the number of operating hours in the specific time interval of analysis. The comparison is performed both numerically and visually, considering thresholds of statistical indicators representing the goodness of fit of models (i.e. indicating practical limits for their acceptability), reported in Table 5 in Section 4.2. The regression models developed are independent on the specific weather data used, as weather data are the independent variables (air temperature and solar radiation in this case, for the reasons described in the introduction and in Section 3), while the energy signature (i.e. average power over the time interval) is the dependent variable.

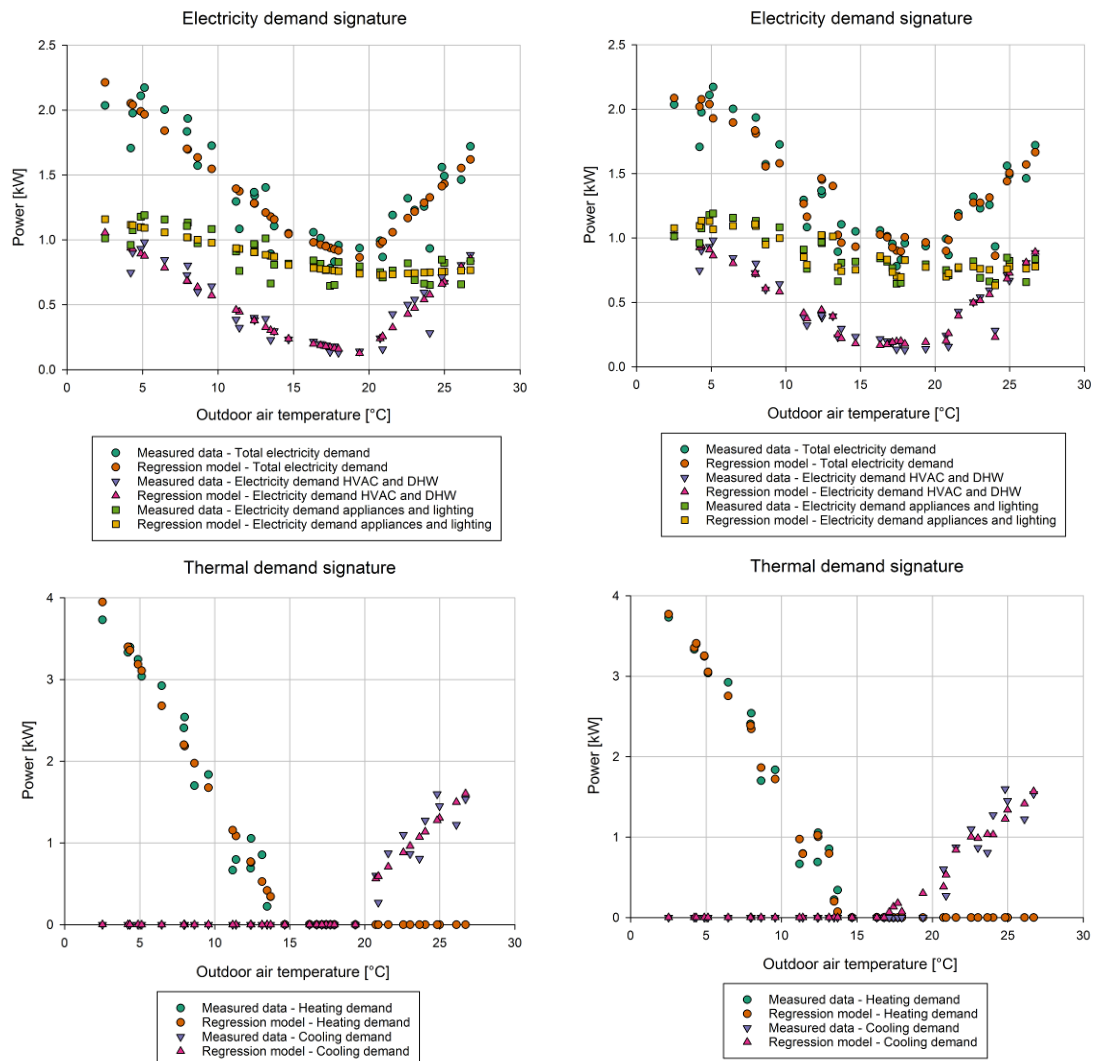


Figure 2: Electricity and thermal demand signature – monthly values, model type 1 on the left, model type 2 on the right

In Figure 2 energy signatures of electricity demands (above) and thermal demands (below) are represented as a function of outdoor air temperature because it represents the

most relevant variable for weather normalization and is frequently used as the only variable in regression-based approaches (i.e. temperature response functions [12]), as explained in the introduction and Section 3. The results of regression models type 1 (illustrated in Table 3, Section 4.1) are reported on the left, while the ones of regression models type 2 (illustrated in Table 4, Section 4.1) on the right. Model type 1 is clearly a simple piecewise linear model (points are on straight lines, with change-points) as a function of 1 input variable, outdoor air temperature, while model type 2 points are more scattered and able to approximate the measured data even better.

Furthermore, considering the change-points (i.e. the balance point temperatures for heating and cooling respectively) of the piecewise linear model type 1, an alternative visualization of the thermal demand for heating and cooling can be provided, using variable-based degree-days [35], computed on a monthly base, in Figure 3. This visualization also confirms in general the goodness of fit of the models presented.

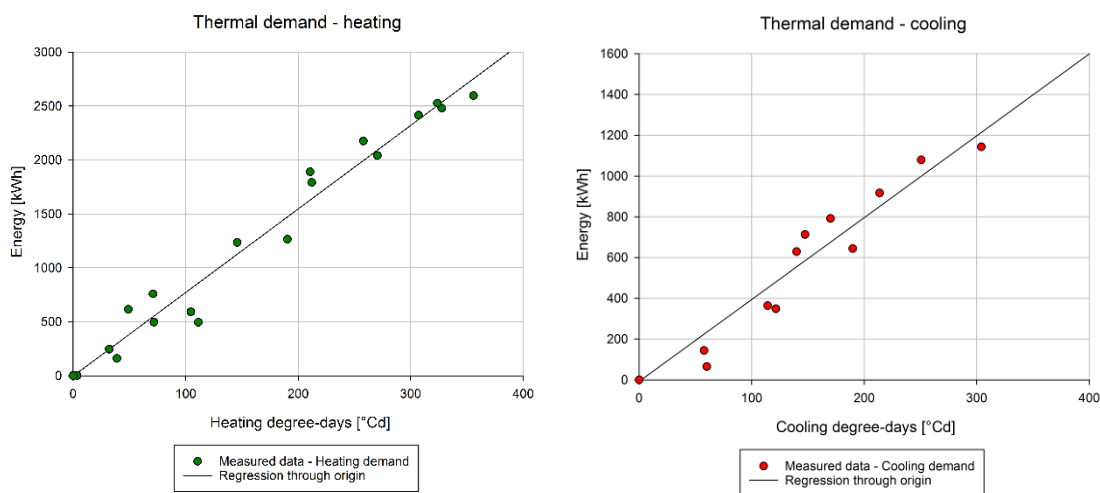


Figure 3: Thermal energy demand – Variable-base degree-days, heating demand on the left, cooling demand on the right

The model formulations (obtained as the sum of the sub-models, using dummy variables) presented in Table 3 and 4 respectively for model type 1 and type 2, are able to fit all the five different datasets (total electricity, electricity for HVAC and DHW, electricity for appliances and lighting, thermal demand for heating, thermal demand for cooling) reasonably well, as indicated in Tables 8 and 9, again respectively for model type 1 and type 2. In general,  $R^2$  values are high in all cases ( $> 80\%$ ), with the notable exception of the electricity for lighting and appliances for model type 1 (67.34%).

Electricity for appliances and lighting is highly dependent on user behaviour and less dependent on weather, essentially only on the amount of daylight hours (from sunrise to sunset) for lighting. The thresholds of acceptability for model calibration (according to M&V protocols reported in Section 4.2), specified in Table 5 are satisfied in all the cases, except for model type 1 trained on HVAC and DHW electricity demand, where the  $Cv(RMSE)$  is greater than 15% (17.35%).  $NMBE$  is very small in all cases (the maximum value found is 1.51%, for thermal demand for cooling, using model type 1) and the maximum value of  $MAPE$  is 13.46%, showing how the models are overall able to fit all the dataset well.

Table 8: Results of analysis of measured and predicted data for the monitoring period –  
Model type 1

Dataset	End-use	Energy indicators		Statistical indicators			
		Energy measured	Energy predicted	$R^2$	$MAPE$	$NMBE$	$Cv(RMSE)$
		kWh	kWh	%	%	%	%
Electric energy	Total	35130	34819	84.41	10.33	-0.88	12.12
	HVAC, DHW	12270	12139	91.23	12.59	-1.07	17.35
	Appliances and lighting	22860	22868	67.34	9.50	0.04	10.56
Thermal energy	Heating	23790	23795	98.29	12.83	0.02	11.68
	Cooling	6838	6735	97.01	5.48	-1.51	13.52

Table 9: Results of analysis of measured and predicted data for the monitoring period –  
Model type 2

Dataset	End-use	Energy indicators		Statistical indicators			
		Energy measured	Energy predicted	$R^2$	$MAPE$	$NMBE$	$Cv(RMSE)$
		kWh	kWh	%	%	%	%
Electric energy	Total	35130	34849	92.24	6.73	-0.80	8.88
	HVAC, DHW	12270	12195	95.53	13.46	-0.61	12.75
	Appliances and lighting	22860	22872	85.26	6.21	0.05	7.31
Thermal energy	Heating	23790	23798	99.30	9.26	0.03	6.20
	Cooling	6838	6827	96.47	3.65	-0.16	11.93

However, due to the fact that the results of model formulations reported in Tables 8 and 9 use dummy variables (0-1), which are used essentially to include/exclude (turn on-off) part of the models, there are conditions where the output is 0, in particular for the thermal model for heating in summer (no-heating) and, viceversa, the thermal model for cooling in winter (no-cooling).

This effect determines a series of points (with output equal to 0) that have to be excluded from the computation of statistical indicators. For this reason, statistical indicators are re-computed by excluding zeros and reported in Table 10 hereafter for the thermal demand (heating and cooling) sub-models. The statistical indicators present some differences but all the models are still calibrated according to thresholds in Table 5, except for type 1 cooling model.

Table 10: Results of analysis of measured and predicted data for the monitoring period –  
Statistical indicators recomputed for thermal demand sub-models

Model	End-use	Statistical indicators			
		$R^2$	$MAPE$	$NMBE$	$Cv(RMSE)$
		%	%	%	%
Type 1	Heating	98.32	13.47	0.03	7.96
	Cooling	92.48	10.85	2.78	17.88
Type 2	Heating	95.91	20.28	0.02	12.42
	Cooling	90.91	19.19	-1.51	17.77

After that, the regression models fitted to energy signatures are used as a basis to compute a monthly time series of energy demand data for the monitoring period (36 months, 3 years) as shown in Figures 4. In this case also, the results of regression models type 1 are reported on the left, while the ones of regression models type 2 are reported on the right, with the electricity demand above and the thermal demand below. The difference between measured and predicted energy demand is very small in all the cases.

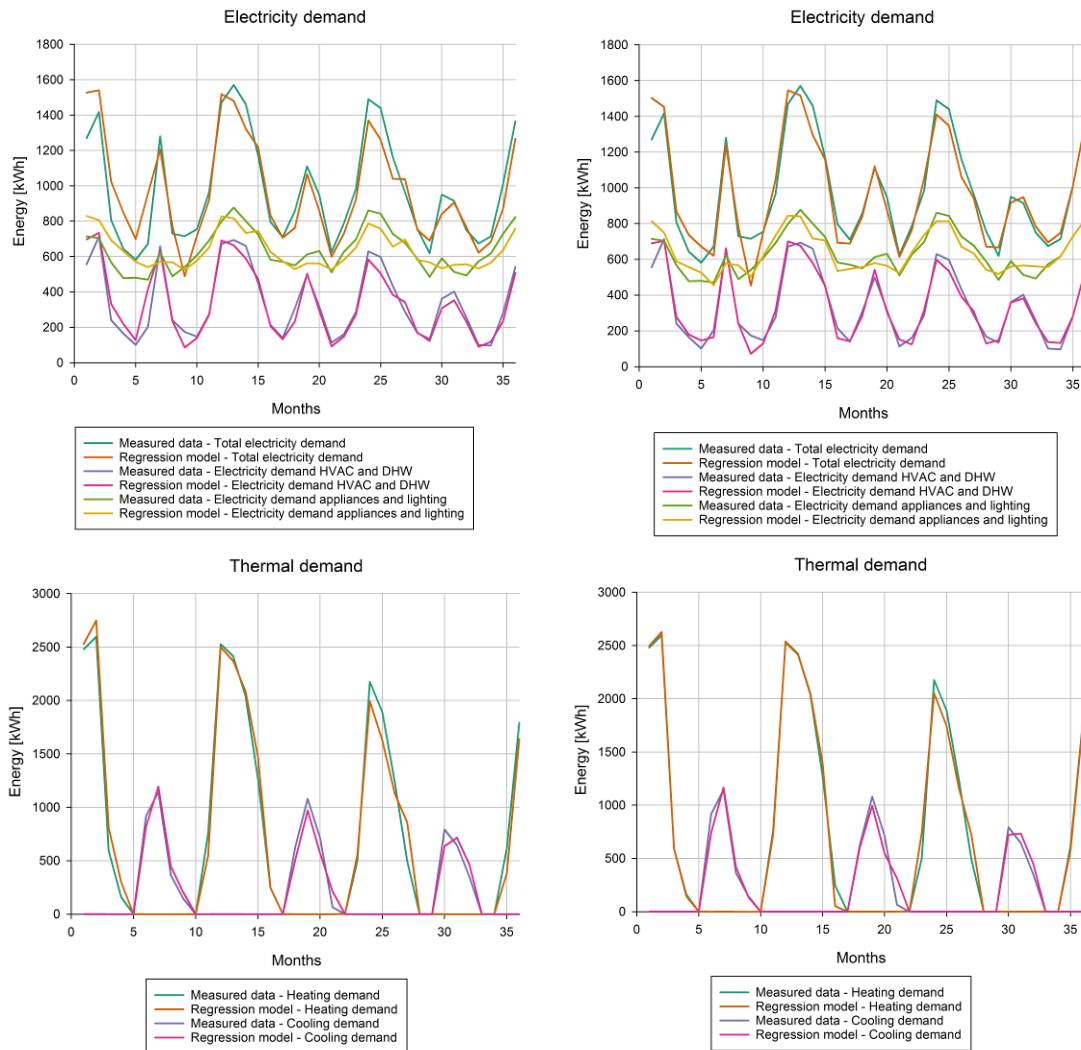


Figure 4: Electricity and thermal demand time series – monthly values, model type 1 on the left, model type 2 on the right

The next step is the visual analysis which involves plotting the differences between measured and predicted data (i.e. residuals, error term) as a function of both outdoor air temperature (the most important input variable) and time (monthly intervals of the monitoring period), to highlight potential patterns in residuals. The error term is plotted in Figures 5 and 6, respectively for outdoor air temperature dependence and time dependence. Again, model type 1 results are reported on the left, while model type 2 results are reported on the right.



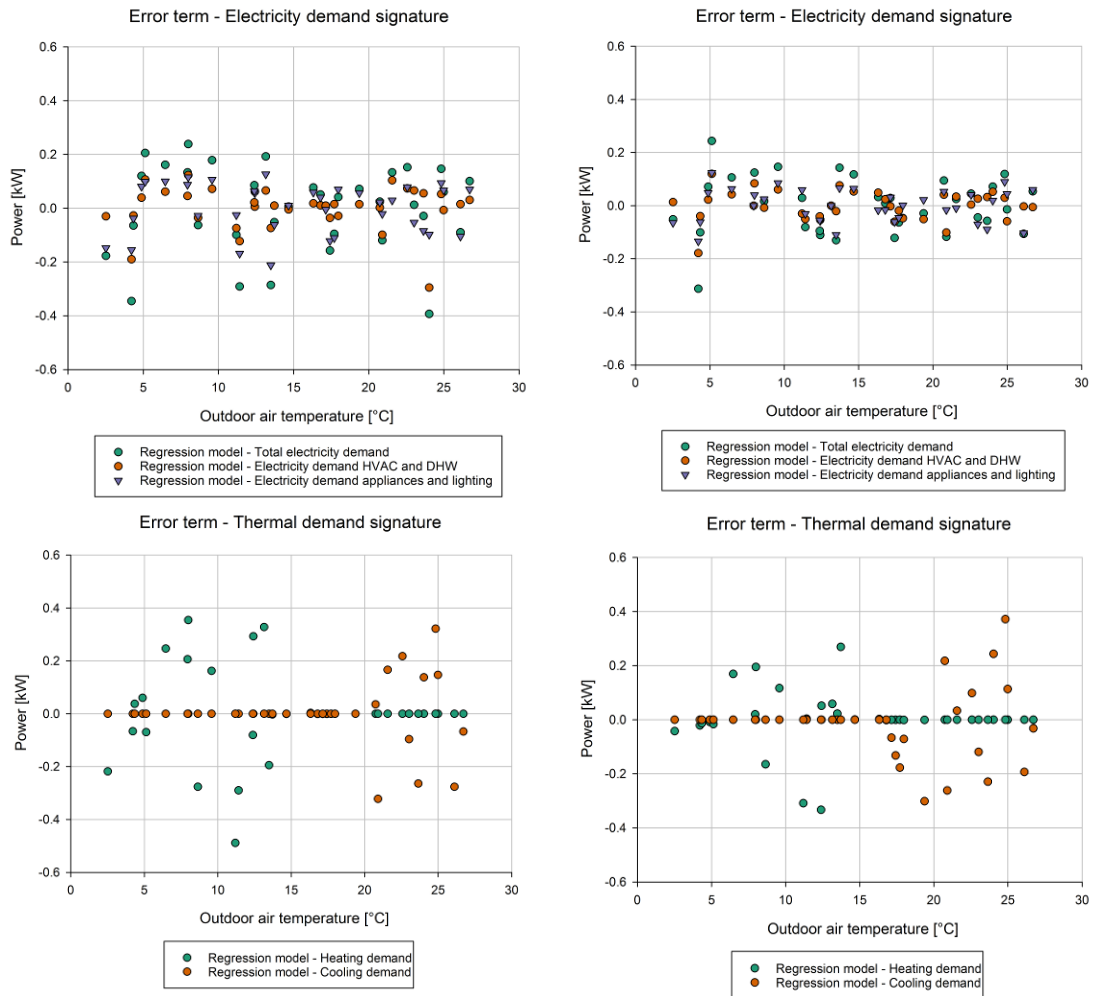


Figure 5: Error term of energy signature models of electric and thermal demand – monthly values, model type 1 on the left, model type 2 on the right

The residuals with respect to temperature (Figure 5) seem relatively uniform with some exception at lower temperature, around 4°C (model type 1 and 2), and higher temperature, around 24 °C (model type 1). As expected, the spread of data around the mean (equal to zero) is larger in model type 1 compared to model type 2. The error term for the thermal models (heating and cooling) is equal to zero in many point because of the model formulation using dummy variables, as explained before in relation to the necessity of re-computing the statistical indicators, reported in Table 10. The residuals (error term) with respect to time (Figure 6), highlight how the electricity demand was lower than the one predicted by the models in the first months of the monitoring period (1-5 in particular), for both type 1 and 2 models, while the distribution in the case of thermal demands seems more uniform.

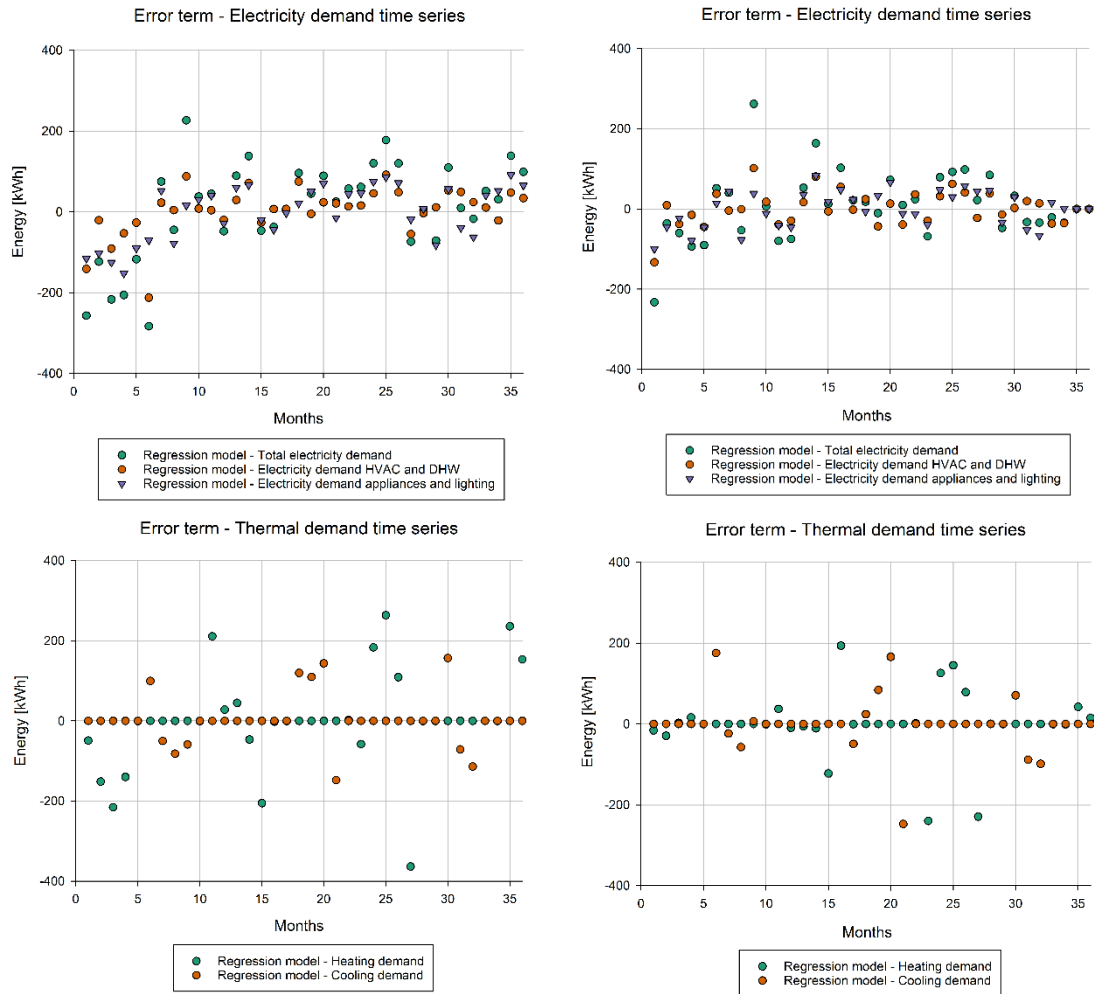


Figure 6: Error term of electricity and thermal demand time series – monthly values, model type 1 on the left, model type 2 on the right

The final step of the visual analysis of residuals (error term), can be performed using a control chart [45] as in Statistical Process Control (SPC) [44]. In this case, for the sake of simplicity, only the control chart for the residuals of total electricity demand is presented in Figure 7. However, the technique can be used for all the other datasets to analyse performance anomalies at multiple levels in the building (when sub-metering data are available). Further, it can be combined with the visualization of the cumulative sum of differences between measured and predicted values (CUSUM) [46] and other innovative related techniques when additional explanatory variables (covariates) are present [57].

In a control chart, the error term is plotted with respect to the number (1, 2 and 3) of standard deviations  $\sigma$  from the mean (equal to zero). In general, the limit of 3 standard deviations from the mean is considered as an indicator of anomaly in the series of residuals. As can be seen in Figure 7, both models type 1 and 2 indicate that months 1 and 9 have larger deviations from the mean, and month 9 is flagged as an anomaly ( $> 3 \sigma$ ) by model type 2. The spread of residuals is much smaller for model type 2 (as expected) and this implies that this model can indicate tighter boundaries for the statistical control of energy performance in time.

As a conclusion, both visual and numerical data analysis is considered in the workflow aimed at model testing; the models were calibrated in most of the conditions but, beyond

calibration, one of the essential goals was to test how simple, intuitive and interpretable (in a "human in the loop" approach) the workflow could be.

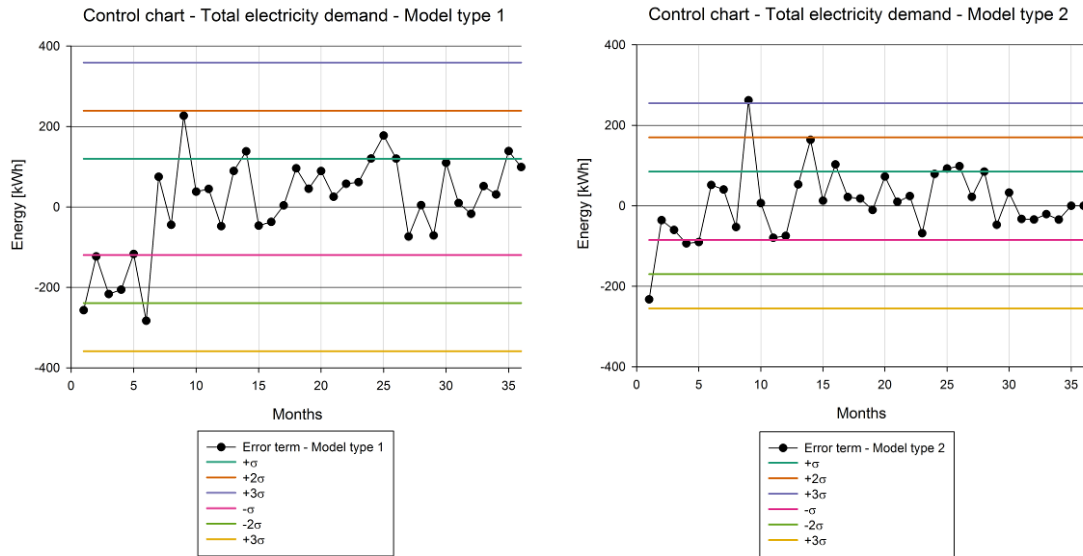


Figure 7: Statistical process control of total electricity demand residuals – monthly values, model type 1 on the left, model type 2 on the right

## 6.2 Limitations

The fundamental limitations of the approach proposed in this study are based on the requirement for rules to define dummy variables. As such, the approach is semi-automated. In any case, the rules are stated transparently and are intrinsically linked to the approximated physical interpretation of the regression coefficients [24].

Furthermore, solar radiation data are generally less widely available than temperature data. However, solar radiation variable is critical to improving model performance for very efficient buildings, in which passive solar gains play a critical role. To overcome this limitation, it is possible to consider the relationship between temperature and time of year with solar radiation [58] and use additional models to perform a reconstruction of solar data, in case they are not directly available.

Additionally, sub-metering data are used in this study to test the approach at multiple levels within the building. This level of detail may not be available on a large scale basis, but disaggregation techniques [41], supplemented by analytical formulations of the building energy balance [43], can be used instead and can be part of future research.

Finally, monthly interval data were used in this study rather than daily, hourly, or sub-hourly data. Even though building performance simulation tools can provide hourly and sub-hourly outputs, the calibration of building energy models using monthly data remains widespread, as evidenced by systematic reviews by Coakley et al. [10] and, more recently, Chong et al. [11]. In fact, data collected on a monthly base are simpler to find (e.g. by digitalizing utility billing data). While monthly-based models do not fully exploit the possibilities provided by higher data granularity, they are computationally inexpensive and can still provide useful insights on energy performance if calibrated.

Following the discussion of current limitations and the underlying potential for further development, the following Section provides indications for future research.

## 7 Further research potential

Further research can be developed in a variety of directions based on the results presented and discussed in the preceding Section. The first development path can be focused on scalability (both temporal and spatial) and integration across life cycle phases (from early design to operation). In fact, regression-based approaches can be applied to energy signature data with monthly, daily, and hourly/sub-hourly time intervals (temporal scalability), but also at multiple levels in energy models (spatial scalability), as shown in previous research [15], thereby adapting them based on the scope and on the granularity of data available.

In general, using data at higher resolution (daily, hourly, sub-hourly) could aid in increasing the complexity and capabilities of models (e.g. load profiles modelling, DSM events modelling, etc.). However, maintaining a certain degree of comparability with models built using monthly data is critical to enabling the effective integration of short-term and high-frequency data with long-term and low-frequency data (i.e. alternating different monitoring strategies), as demonstrated in ASHRAE 1404-RP [59].

Nonetheless, more sophisticated models can perform a variety of functions, potentially being integrated as a system. In particular, they may evolve into “digital twins”, which can supplement building performance simulation tools and track the evolution of building performance across life cycle phases, starting from parametric simulation in design phase, to initial commissioning, and up continuous monitoring/commissioning. Overall, the possibility to use flexible and interpretable models calibrated on measured data (at different intervals) could enable multiple feedback loops in design and operation practices in the built environment, in a continuous improvement logic [38].

A second path of development can be the one making use of the approximated physical interpretation of regression-models’ coefficients. Indeed, identifying lumped thermo-physical properties for buildings is critical to understanding the performance of building technologies.

This was the case, for example, of previous research aimed at incrementally calibrating a Resistance Capacitance (RC) [51] building energy model on the same case study building; similar examples can be found in recent literature [52]. Using the rigorous and harmonised rules, which serve as the foundation of M&V protocols, in conjunction with other standards, focused on building performance simulation, can help improve the energy modelling process through empirically grounded and tested methods.

The third direction is related to heat pump performance analytics and electric grid interaction. The regression model results can be used to implement a calculation of the part load ratio (PLR) of a heat pump based on the actual balance-point temperature of the building (which may vary significantly depending on building characteristics) rather than a fixed reference point (i.e. 16 °C) as in current technical standards dealing with heat pump test conditions [60] and performance assessment at part-load and on a seasonal basis [61].

In the latter, part load ratio is defined as the outdoor temperature in the interval of calculation minus 16 °C, divided by the reference design temperature minus 16 °C, assuming the same value both for heating and cooling, when the heat pump is reversible (like in this case, providing both heating and cooling services). This is not the case in real building operation (as shown also in this research) and regression models can improve the accuracy of analytical methods that are present in technical standardization at the state-of-the-art. By exploiting the scalability of the modelling approach, it could be possible to evaluate and forecast much better the performance of heat pumps and their interaction with the electric grid at large scale.

Overall, these examples of research development paths are clearly not exhaustive, but the evolution of interpretable data-driven modelling approaches into "digital twins" characterised by systems of interacting models is a promising research direction that could help to accelerate building stock decarbonisation through innovative services and technologies [62].

## 8 Conclusion

Data-driven building energy modelling approaches based on machine learning have proven to be effective in a variety of applications. However, there are numerous issues to consider in today's research. To begin with, multiple "performance gaps" (differences between expected and measured performance) are commonly found in buildings, and appropriate analysis techniques should be used. Among those using machine learning, the issue of generalisation (i.e. a model's ability to perform adequately on previously unseen data) is a major topic.

Other critical issues, in addition to generalisation, are interpretability and explainability; interpretability (i.e., the ability for humans to understand the rationale behind model output and inspect the algorithmic logic) in particular is critical to fostering a "human-in-the-loop" approach.

In response to these issues, in this study we examined a Passive House building with two primary objectives. The first one was creating a simple and flexible regression model formulation (i.e. able to fit multiple energy datasets, maintaining the same underlying structure), using dummy variables, and indicating a way to automate the process of model performance comparison and selection.

Statistical indicators were employed to this end, considering general ones and M&V specific ones (i.e. energy model calibration criteria). Dummy variables were included in the datasets by means of interpretable rules that can satisfy the fundamental constraints considered as the motivation of this research (i.e. using interpretable models as a way to enhance the "human-in-the-loop" approach and referring to harmonized procedures to improve model reproducibility and standardization).

Despite their simplicity, the model formulations proposed are able to fit the data reasonably well and can be considered calibrated, following M&V procedures, in most of the cases. Indeed, even with limited information, such as monthly interval data (utility bills and temperatures), the proposed model formulations are appropriate for quick and low-cost (but robust being based on M&V) performance evaluation, which can support the design of energy efficiency measures to meet decarbonisation targets and to reduce reliance on fossil fuels. Further, it can provide a low cost "entry-level" "digital twin" for the building, which can then evolve including more information and a more sophisticated modeling strategy (i.e. employing daily, hourly, sub-hourly interval data, as indicated in Section 7).

The second objective was to identify future research paths based on the first objective's research outcomes and to evaluate them in light of current research developments in the broad area of data-driven building energy modelling. The first path identified is based on temporal and spatial scalability of techniques, as well as on model integration across building life cycle phases. The second path relies on the approximated physical interpretation of regression-model coefficients and on the related analytical formulations (e.g. on the thermo-physical parameters of the building). The third path is concerned with heat pump performance analytics and electric grid interaction, both of which are critical to meeting decarbonisation targets (by means of electrification of heating but considering electric grid supply constraints as well).

Overall, the future research paths indicated are clearly not exhaustive, but they aim to highlight the potential of systems of interacting (interpretable data-driven) models that can evolve into "digital twins" which can help accelerate the transition to a more efficient and decarbonized building stock. Simultaneously, systems of interacting models can assist in addressing important issues such as generalisation, interpretability, and explainability, overcoming limitations inherent in the use of machine learning models and promoting a "human-in-the-loop" approach to a wider audience, not just the research community.

## References

- [1] Hong T, Wang Z, Luo X, Zhang W. State-of-the-art on research and applications of machine learning in the building life cycle. *Energy Build* 2020;212:109831. <https://doi.org/https://doi.org/10.1016/j.enbuild.2020.109831>.
- [2] Westermann P, Evins R. Surrogate modelling for sustainable building design – A review. *Energy Build* 2019;198:170–86. <https://doi.org/https://doi.org/10.1016/j.enbuild.2019.05.057>.
- [3] 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>.
- [4] Miller C. More Buildings Make More Generalizable Models—Benchmarking Prediction Methods on Open Electrical Meter Data. *Mach Learn Knowl Extr* 2019;1. <https://doi.org/10.3390/make1030056>.
- [5] de Wilde P. The gap between predicted and measured energy performance of buildings: A framework for investigation. *Autom Constr* 2014;41:40–9. <https://doi.org/http://doi.org/10.1016/j.autcon.2014.02.009>.
- [6] Imam S, Coley DA, Walker I. The building performance gap: Are modellers literate? *Build Serv Eng Res Technol* 2017;38:351–75. <https://doi.org/10.1177/0143624416684641>.
- [7] de Wilde P. “The building performance gap: Are modellers literate?” *Build Serv Eng Res Technol* 2017;38:757–9. <https://doi.org/10.1177/0143624417728431>.
- [8] van Dronkelaar C, Dowson M, Burman E, Spataru C, Mumovic D. A Review of the Energy Performance Gap and Its Underlying Causes in Non-Domestic Buildings. *Front Mech Eng* 2016;1:17. <https://doi.org/10.3389/fmech.2015.00017>.
- [9] Yoshino H, Hong T, Nord N. IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy Build* 2017;152:124–36. <https://doi.org/https://doi.org/10.1016/j.enbuild.2017.07.038>.
- [10] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. *Renew Sustain Energy Rev* 2014;37:123–41. <https://doi.org/http://dx.doi.org/10.1016/j.rser.2014.05.007>.
- [11] 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/https://doi.org/10.1016/j.enbuild.2021.111533>.
- [12] Fazeli R, Ruth M, Davidsdottir B. Temperature response functions for residential energy demand – A review of models. *Urban Clim* 2016;15:45–59. <https://doi.org/https://doi.org/10.1016/j.uclim.2016.01.001>.
- [13] ISO/IEC TR 29119-11:2020(en) Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems.
- [14] Lipton ZC. The Mythos of Model Interpretability: In machine learning, the

- concept of interpretability is both important and slippery. *Queue* 2018;16:31–57.
- [15] 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>.
  - [16] Pfenninger S, Hirth L, Schlecht I, Schmid E, Wiese F, Brown T, et al. Opening the black box of energy modelling: Strategies and lessons learned. *Energy Strateg Rev* 2018;19:63–71. <https://doi.org/https://doi.org/10.1016/j.esr.2017.12.002>.
  - [17] Hong S-M, Paterson G, Burman E, Steadman P, Mumovic D. A comparative study of benchmarking approaches for non-domestic buildings: Part 1 – Top-down approach. *Int J Sustain Built Environ* 2013;2:119–30. <https://doi.org/https://doi.org/10.1016/j.ijsbe.2014.04.001>.
  - [18] Burman E, Hong S-M, Paterson G, Kimpian J, Mumovic D. A comparative study of benchmarking approaches for non-domestic buildings: Part 2 – Bottom-up approach. *Int J Sustain Built Environ* 2014;3:247–61. <https://doi.org/https://doi.org/10.1016/j.ijsbe.2014.12.001>.
  - [19] Miller C, Meggers F. The Building Data Genome Project: An open, public data set from non-residential building electrical meters. *Energy Procedia* 2017;122:439–44. <https://doi.org/https://doi.org/10.1016/j.egypro.2017.07.400>.
  - [20] Miller C, Kathirgamanathan A, Picchetti B, Arjunan P, Park JY, Nagy Z, et al. The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. *Sci Data* 2020;7:368. <https://doi.org/10.1038/s41597-020-00712-x>.
  - [21] Robertson C, Mumovic D, Hong SM. Crowd-sourced building intelligence: the potential to go beyond existing benchmarks for effective insight, feedback and targeting. *Intell Build Int* 2015;7:147–60. <https://doi.org/10.1080/17508975.2014.987639>.
  - [22] Manfren M, Nastasi B, Groppi D, Astiaso Garcia D. Open data and energy analytics - An analysis of essential information for energy system planning, design and operation. *Energy* 2020;213. <https://doi.org/10.1016/j.energy.2020.118803>.
  - [23] Manfren M, Nastasi B. From in-situ measurement to regression and time series models: An overview of trends and prospects for building performance modelling. *AIP Conf Proc* 2019;2123:20100. <https://doi.org/10.1063/1.5117027>.
  - [24] Tronchin L, Manfren M, Nastasi B. Energy analytics for supporting built environment decarbonisation. *Energy Procedia* 2019;157:1486–93. <https://doi.org/https://doi.org/10.1016/j.egypro.2018.11.313>.
  - [25] ASHRAE. ASHRAE Guideline 14-2014: Measurement of Energy, Demand, and Water Savings; American Society of Heating, Refrigerating and Air-Conditioning Engineers: Atlanta, GA, USA, 2014.
  - [26] EVO. IPMVP New Construction Subcommittee. International Performance Measurement & Verification Protocol: Concepts and Option for Determining Energy Savings in New Construction, Volume III; Efficiency Valuation Organization (EVO): Washington, DC, USA, 2003.
  - [27] FEMP. FEMP. Federal Energy Management Program, M&V Guidelines: Measurement and Verification for Federal Energy Projects Version 3.0, U.S. Department of Energy Federal Energy Management Program 2008.
  - [28] Jayaweera T, Haeri H, Gowans D. The Uniform Methods Project: Methods for Determining Energy Efficiency Savings for Specific Measures. Contract 2013;303:275–3000.
  - [29] Investor Confidence Project (<https://europe.eepperformance.org/>), accessed

- 31/08/2020 n.d.
- [30] Price P. Methods for analyzing electric load shape and its variability, Lawrence Berkeley National Laboratory Report LBNL-3713E. 2010.
  - [31] 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.
  - [32] Kissock JK, Haberl JS, Claridge DE. Inverse modeling toolkit: numerical algorithms. *ASHRAE Trans* 2003;109:425.
  - [33] Paulus MT, Claridge DE, Culp C. Algorithm for automating the selection of a temperature dependent change point model. *Energy Build* 2015;87:95–104. <https://doi.org/https://doi.org/10.1016/j.enbuild.2014.11.033>.
  - [34] Paulus MT. Algorithm for explicit solution to the three parameter linear change-point regression model. *Sci Technol Built Environ* 2017;23:1026–35.
  - [35] ISO 15927-6:2007 Hygrothermal performance of buildings — Calculation and presentation of climatic data — Part 6: Accumulated temperature differences (degree-days).
  - [36] ISO 50006:2014, Energy management systems — Measuring energy performance using energy baselines (EnB) and energy performance indicators (EnPI) — General principles and guidance.
  - [37] ISO 16346:2013, Energy performance of buildings — Assessment of overall energy performance.
  - [38] ISO 50001:2018, Energy management systems - Requirements with guidance for use.
  - [39] Bauwens G, Roels S. Co-heating test: A state-of-the-art. *Energy Build* 2014;82:163–72. <https://doi.org/https://doi.org/10.1016/j.enbuild.2014.04.039>.
  - [40] 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>.
  - [41] Li M, Allinson D, Lomas K. Estimation of building heat transfer coefficients from in-use data. *Int J Build Pathol Adapt* 2020;38:38–50. <https://doi.org/10.1108/IJBPA-02-2019-0022>.
  - [42] Tronchin L, Manfren M, Tagliabue LC. Optimization of building energy performance by means of multi-scale analysis – Lessons learned from case studies. *Sustain Cities Soc* 2016;27:296–306. <https://doi.org/https://doi.org/10.1016/j.scs.2015.11.003>.
  - [43] Vesterberg J, Andersson S, Olofsson T. Robustness of a regression approach, aimed for calibration of whole building energy simulation tools. *Energy Build* 2014;81:430–4. <https://doi.org/https://doi.org/10.1016/j.enbuild.2014.06.035>.
  - [44] Braga LC, Braga AR, Braga CMP. On the characterization and monitoring of building energy demand using statistical process control methodologies. *Energy Build* 2013;65:205–19. <https://doi.org/https://doi.org/10.1016/j.enbuild.2013.05.002>.
  - [45] Flores M, Naya S, Fernández-Casal R, Zaragoza S, Raña P, Tarrío-Saavedra J. Constructing a Control Chart Using Functional Data. *Math* 2020;8. <https://doi.org/10.3390/math8010058>.
  - [46] Fichera A, Volpe R, Cutore E. Energy performance measurement, monitoring and control for buildings of public organizations: Standardized practises compliant with the ISO 50001 and ISO 50006. *Dev Built Environ* 2020;4:100024. <https://doi.org/https://doi.org/10.1016/j.dibe.2020.100024>.
  - [47] Lin M-H, Carlsson JG, Ge D, Shi J, Tsai J-F. A Review of Piecewise



- Linearization Methods. *Math Probl Eng* 2013;2013:101376.  
<https://doi.org/10.1155/2013/101376>.
- [48] Bemporad A. Piecewise linear regression and classification 2021.  
<https://doi.org/10.48550/ARXIV.2103.06189>.
- [49] Borgeson SD. Targeted efficiency: Using customer meter data to improve efficiency program outcomes. University of California, Berkeley; 2013.
- [50] Bollinger LA, Davis CB, Evins R, Chappin EJJ, Nikolic I. Multi-model ecologies for shaping future energy systems: Design patterns and development paths. *Renew Sustain Energy Rev* 2018;82:3441–51.  
<https://doi.org/https://doi.org/10.1016/j.rser.2017.10.047>.
- [51] Tronchin L, Manfren M, James PAB. Linking design and operation performance analysis through model calibration: Parametric assessment on a Passive House building. *Energy* 2018;165:26–40.  
<https://doi.org/https://doi.org/10.1016/j.energy.2018.09.037>.
- [52] Fumo N, Torres MJ, Broomfield K. A multiple regression approach for calibration of residential building energy models. *J Build Eng* 2021;43:102874.  
<https://doi.org/https://doi.org/10.1016/j.jobe.2021.102874>.
- [53] Lee K, Baek H-J, Cho C. The Estimation of Base Temperature for Heating and Cooling Degree-Days for South Korea. *J Appl Meteorol Climatol* 2014;53:300–9.
- [54] 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/https://doi.org/10.1016/j.enbuild.2017.09.034>.
- [55] Peña Suárez JN, del Campo Díaz VJ. Degree-days in a Caribbean and tropical country: the Dominican Republic’s case. *Int J Ambient Energy* 2021;42:795–800.  
<https://doi.org/10.1080/01430750.2019.1566175>.
- [56] Uncertainty Assessment for IPMVP, International Performance Measurement & Verification Protocol, Efficiency Valuation Organization (EVO) n.d.
- [57] Centofanti F, Lepore A, Menafoglio A, Palumbo B, Vantini S. Functional Regression Control Chart. *Technometrics* 2021;63:281–94.  
<https://doi.org/10.1080/00401706.2020.1753581>.
- [58] Vesterberg J, Andersson S, Olofsson T. A single-variate building energy signature approach for periods with substantial solar gain. *Energy Build* 2016;122:185–91. <https://doi.org/https://doi.org/10.1016/j.enbuild.2016.04.040>.
- [59] Abushakra B, Reddy A, Singh V. ASHRAE Research Project Report 1404-RP, Measurement, Modeling, Analysis and Reporting Protocols for Short-term M&V of Whole Building Energy Performance, Arizona State University, USA. 2012.
- [60] EN 14511-2:2018 Air conditioners, liquid chilling packages and heat pumps for space heating and cooling and process chillers, with electrically driven compressors. Test conditions.
- [61] 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.
- [62] Manfren M, Nastasi B, Tronchin L, Groppi D, Garcia DA. Techno-economic analysis and energy modelling as a key enablers for smart energy services and technologies in buildings. *Renew Sustain Energy Rev* 2021;150:111490.  
<https://doi.org/https://doi.org/10.1016/j.rser.2021.111490>.