Energy 283 (2023) 128490

Contents lists available at ScienceDirect

Energy

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

Interpretable data-driven building load profiles modelling for Measurement and Verification 2.0

Massimiliano Manfren [a], Benedetto Nastasi [b,*]

[a] Faculty of Engineering and Physical Sciences, University of Southampton, Boldrewood Campus, SO16 7QF, Southampton, United Kingdom
[b] Department of Planning, Design & Technology of Architecture, Sapienza University of Rome, Via Flaminia, 72 00196, Rome, Italy

ARTICLE INFO

Handling editor: Neven Duic

Keywords:
Data-driven methods
Interpretability
Regression-based approaches
Measurement and verification
M&V 2.0
Energy analytics
Energy management
TOWT

ABSTRACT

Accelerating the decarbonisation of the built environment necessitates increasing electrification of end-uses, which in turn poses the issue of rethinking the role of energy efficiency in conjunction with flexibility in grid interaction. This requires a better understanding of the electricity load profiles at hourly or sub-hourly intervals using techniques that are simple, reliable, and interpretable. To this extent, this study proposes a reformulation of the Time Of Week and Temperature modelling approach. This approach is able to separate the energy consumption dependence on building operational characteristics (Time Of Week) and on weather (outdoor air temperature), through a highly automated modelling workflow, necessitating minimal effort for model tuning. These features, along with its intrinsic interpretability due to its formulation using multivariate regression and the availability of open-source software, makes it an ideal starting point for applied research. The case study selected for the research is a fully electrified public building in Southern Italy. The building has been monitored for 5 years, before, during and after the COVID-19 lockdown. The novel model formulation is calibrated using hourly interval data with a Coefficient of Variation of Root Mean Square Error in the range of 20.0–28.5% throughout the various monitoring periods. The counterfactual analysis of electricity consumption indicates a 10.7–26.7% decrease in electricity consumption due to operational adjustments following COVID-19 lockdown, highlighting the impact of behavioural change. Finally, the possibility of additional workflow automation and enhanced interpretability is discussed.

Nomenclature

Variables and parameters

Symbol	Quantity	Unit
A	average measured energy consumption	kWh
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, intercept	
$CV(RMSE)$	coefficient of variation of $RMSE$ (expressed in percentage)	-
E_i	difference between measured and predicted energy consumption for each data point (error term)	kWh
L_{occ}	electric load when the building is occupied	kW
L_{unocc}	electric load when the building is not occupied	kW
n	number of data points, number of hours in a week	-
$NMBE$	normalised mean bias error (expressed in percentage)	-
p	degrees of freedom of the model	-
$RMSE$	root mean square error	kW
R^2	determination coefficient (expressed in percentage)	-
$T(i)_k$	outdoor air temperature value for the segment k at time interval i	°C
$t_{ow,j}$	time of week binary variable	-

1. Introduction

The necessity to accelerate decarbonisation of the building stock [1] involves a substantial switch in consumption from fuels to electricity, which carries the risk of a shifting impact [2], despite the ever increasing quota of energy from renewables [3]. In other words, the growing electrification of building stock [4] necessitates a more thorough examination of electricity use in order to meet heating and cooling demands, as well as those for lighting, appliances, and electric cars [5].

* Corresponding author.
E-mail address: benedetto.nastasi@outlook.com (B. Nastasi).

https://doi.org/10.1016/j.energy.2023.128490
Received 24 January 2023; Received in revised form 26 May 2023; Accepted 19 July 2023
Available online 25 July 2023
0360-5442/

This presents the challenge of reinventing energy efficiency [6] in the context of achieving decarbonisation targets [1] while resolving the inherent flexibility concerns associated with electrification [7]. In the process of increasing the energy efficiency of the building stock, multiple issues must be considered, including refurbishment options [8], appropriate design and sizing of technical systems [9], and neighbourhood [10] or district [11] and community-scale solutions that require adequate policies [12], which are part of the current scientific debate at global scale [13]. It is crucial to monitor [14] and control the dynamics of energy usage and grid interaction, especially when on-site renewable power generation is present and load-matching and grid interaction [15] capabilities can be enhanced. In addition, a sufficient length for the monitoring period [16] and a suitable data granularity are essential for evaluating the actual dynamic building energy behaviour and deriving relevant insights.

Indeed, energy bills with monthly data resolution are the simplest form of datasets that can be obtained and analysed [17], but this is insufficient if the objective is promoting a better integration of renewables [18] and new business models [19] on the energy market, where hourly to minute-level data resolution is crucial [20]. Additionally, regarding the dynamic interaction between end-users and buildings, research projects such as IEA EBC Annex 79 [21] have emphasised the use of dynamic data to increase the understanding of user behaviour and preferences. Despite the fact that user behaviour is not often the single most important factor in energy consumption, there is frequently a considerable lack of information about its actual impact on energy consumption, even among building managers, resulting in a significant information gap [22]. This is clearly an issue in light of the fact that the exploitation of potential energy savings and flexibility reserves can have both a major local (e.g., individual users, energy communities, etc.) but also national-level implications.

In essence, the aforementioned issues, namely electrification, efficiency, and flexibility, provide the framework for this study, the objective of which is to propose a novel formulation of the energy time series modelling approach known as Time Of Week and Temperature (TOWT) [23] and to test it under conditions identified as critical for the original TOWT software implementation, as discussed in prior research [24]. The chosen case study is the Procida Town Hall in Southern Italy, which was monitored for approximately five years before, during, and after the COVID-19 pandemic lockdown period. The original goal of the research was achieving a better knowledge of the dynamic building behaviour (i.e. electric load profiles, the building is fully electrified) by deploying methodologies that may be employed in a relatively straightforward manner and automated further while retaining an "intrinsic interpretability" [25], discussed more in detail in Section 2. At the state-of-the-art, in the context of Artificial Intelligence and Machine Learning (AI/ML) software testing [26], "interpretability" is defined as the "level of understanding how the underlying (AI) technology works" which entails how the model output and algorithmic logic can be understood in human terms. On the one hand, techniques such as linear multivariate regression or regression and decision trees, for instance, are classified as interpretable since they are easily comprehensible in human terms. On the other hand, sophisticated data-driven methods used in a variety of applications in the energy sector are often "black boxes", for example load prediction for buildings with the presence of electric vehicles (EVs) [27], automated rejection of unreliable predictions when online building energy data are used [28], treatment of the fairness/accuracy trade-off problem for data-driven methods in indoor environment modelling [29] and many others. Despite having a lower accuracy in many cases with respect to "black-box" methods, (intrinsically) interpretable methods can be insightful [30] and be also preferred by practitioners [31]. Promoting a "human in the loop" approach to data-driven methods and increasing literacy both regarding energy and AI/ML tools by leveraging simpler techniques, which can provide adequate performance and automation potential but are easier to understand in human terms, is part of the research conducted. In this regard, the choice of the TOWT approach depends on its ability to provide an algorithmic formulation that is at the same time interpretable (regression-based), largely automated (minimal effort needed for model tuning) and able to separate the energy consumption dependence on building operational characteristics (Time Of Week) and on weather (outdoor air temperature). The ability of the TOWT algorithm to identify peculiar operational characteristics, including the automated identification of "occupied" and "unoccupied" hours, can contribute to fill the information gap [22] discussed before in relation to user behaviour. Section 2 elaborates more in detail on why the TOWT regression-based technique was selected over other possibilities.

2. Literature review

The acceleration of the decarbonisation process through electrification of end-uses and greater penetration of renewable power generation into the electric systems, as stated in the introduction, necessitates innovative approaches to efficiency [6] and flexibility [7].

In this regard, data-driven load profile modelling methodologies should be capable of handling increasingly detailed data, on the order of hours and minutes, with an increased level of automation. Advanced Measurement and Verification (M&V) or M&V 2.0 [32] typically refers to applications of AI/ML techniques on granular measurements of energy use, which should deliver greater model precision, statistical reliability, and real-time feedback on performance and operational issues [33], ideally employing sophisticated visualization techniques. Among regression-based techniques, the Time Of Week and Temperature (TOWT) algorithm, proposed initially by Price [23] to analyse electric load shape and its variability and then used for the quantification of changes in electricity use due to demand response implementation [34] and in the context of utility scale efficiency programs [35], has been chosen as a starting point for this study. The reasons for the choice are as follows. First, the ability to separate the energy consumption dependence on operational characteristic (Time Of Week) and weather (outdoor air temperature). After that, intrisic interpretability due to its regression-based formulation. In addition, its automated workflow, necessitating minimal effort for model tuning (just one hyper-parameter, as will be explained later). Then, its availability as open-source software in the R packages "RMV2.0" [36], "NMECR" [37], and the Python package OpenEEmeter, which implements "Caltrack" methods [38]. EVO's Advanced M&V Testing Portal [39], provides electricity consumption data from 367 buildings in various regions of North America for the purpose of open and independent testing of M&V algorithms and TOWT was successfully tested, showing a good level of accuracy.

The model input is a time series of hourly or sub-hourly outdoor air temperatures and energy data (load profile). The time stamp of the series identifies the day of the week, and extra variables are automatically added in the construction of the regression model (that is, differentiating each day and hour of the week to capture specific recurring weekly patterns of operation).

In addition, the temperature dependence of electric load is accounted for by separating load data into temperature bins. It can handle both Demand Response (DR) and Non Routine Events (NRE), i.e. unexpected changes in occupancy and operation strategies, employing methods established by Killick to deal with the automated detection of change-points [40], implemented in the R package "Change-point" [41], as well as in "RMV2.0" [36].

TOWT uses variable-base degree-days regression as part of its automated workflow, which makes it potentially comparable with variable-base degree-days (VB-DD) regression algorithms, which are closely relared to change-point methods, originally proposed by Kissock et al. in the Inverse Modeling Tool (IMT) [42], which have been included in ASHRAE 14:2014 [43] and have been steadily evolving over time with the introduction of algorithmic techniques for the selection of base temperature [44], up to the explicit solution for the three parameters case [45].

At a very basic level, as evidenced by recent research, variable-base degree-days regression algorithms methods are versatile enough to be used for a variety of purposes throughout the building life cycle [46] and the analytics generated can be used for multiple applications in the built environment [47], where spatial and temporal scalability of modelling techniques is crucial to overcome practical problems. Further, VB-DD have a “transversal” role that spans from long-term projections of climate change impact [48], to climate classification in relation to building energy performance [49] (where the choice of the base temperature is important in relation to building characteristics [50] and can be defined through additional parameters [51]), and to methods to assess climate uncertainty in energy system models [52], including the problem of heat demand electrification at scale [53]. While degree-days definition is standardised [54], newer formulations are aimed at adapting it to problems where building dynamic behaviour has to be studied [55].

At the same time, as illustrated by Kim and Haberl, change-point regression is part of the normalisation of measured energy performance in both intermediate level [56] and advanced level applications [57]. As explained before, the objective of the research is developing a reformulation of TOWT algorithm, whose peculiar characteristics are reported in Section 3. Among the novel features introduced there is an alternative way to perform the segmentation of outdoor air temperature data and a different way to identify the “occupied” and “unoccupied” periods, based on piecewise linear regression instead of variable-base degree-days. This alternative formulation enables the comparison of the (piecewise linear) temperature response model fitted in the automated workflow with change-point models leveraging energy signature [58] (i.e. energy divided by the number of operating hours in the time interval of the analysis, corresponding to an average power [59]), instead of energy, enhancing its interpretability, determined by the possibility to compare data at different temporal intervals such as monthly, daily, and hourly/sub-hourly (energy signature is an average power, which can be compared at different temporal scales, as function of outdoor air temperature). These new features can help overcome some of the limitations mentioned in the introduction, such as supporting the improvement of building management strategies, which is not always possible due to limitations in the quantity and quality [60] of available data, resulting in an information gap [22], as pointed out in the introduction. Further, the reconstruction of missing parts of energy time series to evaluate the dynamic performance of building assets when data scarcity occurs [61] is a topic that has been explored by Lamagna et al. [62] in relation to electric load profiles. The possibility to use both short-term/high-frequency measurements (i.e. daily, hourly, sub-hourly) and long-term/low-frequency measurements (i.e. monthly, over multiple years ideally) in an integrated way has been tested previously in a successful way in ASHRAE project RP-1404 [63]. Even when the building is equipped with smart meters that measure aggregated data (potentially enabling sophisticated data analysis workflows), there is the need for data normalisation and disaggregation to understand the impact of efficiency measures within a single building or across sets of comparable buildings [64]. The simple comparison with existing datasets or similar building stock data is insufficient because of the need to “normalise” performance with respect to weather and operational characteristics (including user behaviour), which is one of the key characteristics of M&V2.0 techniques.

With respect to the evaluation of model performance (i.e. judging the model’s ability to reproduce measured energy consumption), the statistical indicators commonly employed by state-of-the-art protocols for Measurement and Verification (M&V), such as ASHRAE 14:2014 [43], Efficiency Value Organization (EVO) IPMVP [65], and Federal Energy Management Program (FEMP) [66] are Normalised Mean Bias Error, *NMBE* and Coefficient of Variation of Root Mean Square Error, *CV (RMSE)*. In addition, standards such as ISO 50006:2014 [67] consider the coefficient of determination R^2 as well. Being based on measured performance, these standard and protocols are empirically grounded and represent the results of multiple years of field testing, with intermediate [56] or advanced [57] levels of application, as explained before.

In relation to the choice of data-driven modelling techniques, it is important at present to reflect on concepts such as "interpretability" and "explainability" as a result of the growing use of AI/ML techniques in the energy sector and to be able to distinguish between “ante-hoc” or “intrinsic” interpretability, and “post-hoc” interpretability [25]. The ISO/IEC TR 29119–11:2020 standard [26] for Artificial Intelligence (AI) software testing defines "interpretability" as the "level of understanding how the underlying (AI) technology works”. The definition provided does not apply to the extent to which the internal mechanics of the machine learning algorithm can be explained in human terms (i.e the “explainability” representing the “level of understanding how the AI-based system came up with a given result” according to the above mentioned standard) but rather to the possibility (for a human) to predict the model’s output given a change in input data or algorithmic parameters. Interpretability is an important but problematic attribute for ML techniques [68], necessitating particular methodologies when it is used to identify causality [69], which is part of the “grand challenges” related to interpretability in ML [70].

In simpler terms, the more "transparent" the model is to the user, the more it can support a "human-in-the-loop approach" that "black-box" solutions cannot. Techniques based on linear multivariate regression and regression trees are classified as interpretable, whereas others, such as random forests or neural networks (or other ML methods), are classified as "black-box" (non-interpretable), even if their computation is explainable. While providing high performance, explainable techniques, need the introduction of additional approaches such as SHAP (SHapley Additive exPlanations) [71] or LIME (Local Interpretable Model-agnostic Explanations) which makes them less intuitive, compared to “intrinsic” (“ante-hoc”) interpretable techniques. For example, SHAP technique is used by Zhang et al. [71] to provide insights into building performance and identify variables influencing energy consumption and greenhouse gas emissions. However, in a recent state-of-the-art review on interpretability of ML methods used in building energy management, Chen et al. [72] indicated (among other challenges) that “the prevalent techniques such as SHAP and LIME can only provide limited interpretability”. Therefore, the use of ML techniques to address energy performance estimation [73] in buildings and benchmarking needs to confront with the problem of lack of interpretability [74], and to the need for feature engineering to make them applicable to building assets in a scalable way [75].

In conclusion, the presence of multiple interpretable AI/ML models that can work in synergy (being designed to be temporally and spatially scalable [47], leveraging similar rules and standards, combined into systems [76] and sharing crucial information [77]) can significantly enhance the robustness of the energy performance assessment of building at scale and create multiple feedback loops in a continuous improvement logic [78], following a Deming cycle or PDCA (Plan-Do--Check-Act reported in technical standardization of energy management [79]), which is crucial for the achievement of efficiency targets, as well as the closely connected flexibility targets, within a market-enabled mechanism [80], in which techno-economic optimization is essential [81]. Section 3 describes the reformulation of the TOWT algorithmic approach using RMV2.0 as a reference and the previous monitoring activity conducted for the selected case study [24].

3. Methods

The methods used in this study are described hereafter in Sections 3.1 and 3.2. Section 3.1 describes the model formulation and workflow, whereas Section 3.2 describes the criteria used to evaluate the model’s acceptability as calibrated for the specific case.

3.1. Model formulation and workflow

In this study, a novel formulation of the Time Of Week and Temperature (TOWT) [23] algorithm, originally implemented in RMV2.0 software [36], is proposed. This modelling approach was selected due to its potential interpretability, which is inherently related to its formulation based on piecewise linear regression and this aspect is stressed in the results and discussion section. An hourly or sub-hourly dataset with a timestamp, outdoor air temperature, and energy is all that is required as model input, and this make it simple to set-up. The timestamp of the series determines the day of the week, and the temperature data is binned based on the algorithm implementation. Only one hyper-parameter is present in the novel formulation proposed, similarly to its original implementation, but with a different meaning. The model is formulated as follows:

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

$$L_{unocc} = \sum_{j=1}^{n-1} a_j t_{ow,j} + \sum_{k=1}^{m} b_k \left(T(i)_k\right) + c_0 \tag{2}$$

Two types of variables are considered in the model formulation.

- $t_{ow,j}$ is a time of the week binary (0–1) variable;
- $T(i)_k$ is temperature variable, representing a temperature segment (binned data), identified algorithmically, as illustrated below.

The time of the week ($t_{ow,j}$) variable is a binary variable, *n-1* reflects the number of hours in a week (168-1 = 167), and the last term (168th) is included in the intercept term c_0, this is a typical strategy when modelling with dummy variables [82]. The temperature variable $T(i)_k$ is a continuous variable with an arbitrary temperature scale, and *m* represents the number of segments employed while binning the temperature data. To ensure continuity in the (piecewise linear) temperature response component of the function, the algorithm described in Ref. [34] is used to generate temperature segments. The number of temperature segments in the initial version was 6; however, more recent implementations have the ability to automate the selection and to subset temperature intervals with equal width or equal number of data points. The user may also specify the change-points for the piecewise linear function explicitly as an input to the method. However, a potential advantage of TOWT over change-point approaches (i.e. VB-DD regression) is that the change-points of the piecewise linear function do not need to be specified by the user, but can be computed in an automated way. In the reformulation proposed the change-points can be chosen based on quantiles of outdoor air temperature data, considering 1 change-point in the median of the outdoor air temperature distribution and a symmetric subdivision into intervals. This subdivision can be refined further by comparing the change-point initially selected with the ones identified by change-point methods (ideally plotting energy signature as a function of outdoor air temperature) and this makes it more "flexible" with respect to the original implementation in RMV2.0 and, at the same time, it provides a sort of continuity between the two modelling approaches.

As shown in formulas 1 and 2, the model is split into a part computing the load profile for "occupied" hours and the other one for "unoccupied" hours. The criterion for partitioning the model into "occupied" and "unoccupied" hours is based on an occupancy threshold factor, which distinguishes between hours with higher consumption, which are considered as "occupied", and hours with lower consumption, which are considered as "unoccupied". Therefore, the criterion is used to discriminate between "high" and "low" energy consumption interval in the series, not necessarily the actual occupancy value (i.e. number of people in the building), even though "high" or "low" consumption conditions can be considered effectively a "proxy" of occupancy level, at least in relation to the use of electric appliances. "High" and "low" demand periods are classified based on a regression of load with respect to outdoor air temperature, using VB-DD regression in the original formulation and a piecewise linear function in the reformulation proposed, whose details are reported later in Table 1. The time intervals of the week when the regression usually underpredicts the load are labelled as "high" demand, while the rest of the intervals are labelled as "low" demand. The above mentioned threshold is used in this way, if the regression underpredicts for a certain fraction (0–100%) of the time, the corresponding interval is assumed to be in the "high" demand mode. In its original implementation RMV2.0 uses a threshold of 65% (0.65 fraction), while other implementations such as R package NMECR [37] give the possibility to set it arbitrarily, making the model more flexible. This possibility is considered as well in the new implementation proposed in this research, but the default parameter value (65%) is used as it seems to work appropriately also in this istance.

Finally, in terms of model tuning, in RMV2.0 implementation there is just one hyper-parameter, represented by the time scale of the weighting function (multiple regressions are trained and weighted), i.e. the number of days nearby the predicted day that are used for weighting. In the novel formulation proposed one hyper-parameter is used as well, but representing the temporal segmentation (expressed in months or fraction of months) needed to model residuals as a Time Of Week (TOW) function without temperature dependence, which is eliminated in the initial part of the modelling workflow, using the regression against temperature. Another alternative formulation of the hyper-parameter for TOWT is the one implemented in Caltrack [38], where the number of nearby months used for weighting is considered. For the sake of simplicity, the weighting approach, which intrinsically depends on the choice of hyper-parameter, is omitted from formulas 1 and 2. The characteristics of the novel algorithmic implementation proposed and its workflow are summarized in Table 1.

3.2. Model calibration criteria

As previously discussed in Section 2, the use statistical indicators such as the Normalised Mean Bias Error (*NMBE*) and Coefficient of Variation of Root Mean Square Error (*CV(RMSE)*) to establish acceptable thresholds for models is common to standards and protocols such as ASHRAE 14:2014 [43], Efficiency Value Organization (EVO) IPMVP [65] and Federal Energy Management Program (FEMP) [66]. Table 2 provides a summary of the acceptability thresholds used in this research (for hourly interval data, due to the nature of the technique employed) and taken from ASHRAE 14:2014. As per ISO 50006:2014 [67], the coefficient of determination R^2 (which spans from 0 to 100%, or 0 to 1) is also considered, but its limitations must be understood. These limitations arise from the strong relationship between R^2 and the model's slope (i.e. its dependence on input variables). Even when the variance of the predicted variable is the same, models with higher slope values will have higher R^2 values. Nonetheless, the IPMVP Guidelines for Assessing Uncertainty [83] consider an R^2 value of 75% as a reference. Below are the formulas for calculating the statistical indicators.

The first indicator is the Normalised Mean Bias Error (*NMBE*), reported in Equation (4). It is calculated by taking the sum of the error terms E_i in each time interval (difference between the measured and predicted value), and dividing it by the average energy consumption *A* and by the difference between number of data points and degrees of freedom of the model *(n-p),* and expressed as a percentage.

$$A = \frac{\sum_{i}^{n} M_i}{n} \tag{3}$$

Table 1
TOWT modelling approach reformulation proposed.

Workflow step	Component	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). Default change points in RMV2.0 implementation are present (40, 55, 65, 80, 90 F/4.4, 12.8, 18.3, 26.7, 32.2 °C), while in the reformulation 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). Temperature segmentation is performed as described in Ref. [34]. Depending on the distribution of outdoor air temperature data, 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 (high, low demand).
Model training	Detection of occupied/unoccupied hours (high/low demand)	Differently from the original implementation (using Heating Degree-Days (HDD) and Cooling Degree-Days (CDD) and an intercept term regression) the occupied/unoccupied periods (high/low demand) 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 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, using only 1 hyper-parameter.
	Hyper-parameters	One hyper-parameter is present, expressed in months or fractions of 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 plot	Energy signature is plotted with respect to outdoor air temperature. 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 (change-point methods) at different time intervals (monthly, daily, hourly).
	Time series plot	Time series of measured data are plotted with respect to predicted ones.
	Load profiles	Typical load profile for working days and weekend conditions.
	Other potential visualizations	Weekly patterns of operations are shown with a 2D heatmap (as in RMV2.0), residuals are plotted in time and with respect to outdoor air temperature (as in RMV2.0). A scatterplot of measured vs predicted data is used to highlight possible deviations (as in RMV2.0).

Table 2
Thresholds of acceptability for M&V models as calibrated with hourly data.

Interval	Metric	ASHRAE Guidelines 14
		%
Hourly	***NMBE***	±10
	Cv(RMSE)	30

$$NMBE = \frac{\sum_{i}^{n} E_i}{(n-p) * A} * 100 = \frac{\sum_{i}^{n} E_i * n}{(n-p) * \sum_{i}^{n} M_i} * 100 \tag{4}$$

The second indicator is the Coefficient of Variation of Root Mean Squared Error (*CV(RMSE)*), as shown in Equation (6). It is calculated as the ratio of the Root Mean Squared Error (*RMSE*) computed in Equation (5), which represents the sample deviation of the differences between measured and predicted values, divided by the average measured electricity consumption (*A*) computed in Equation (3), and expressed in percentage. The lower the *CV(RMSE)* value, the better the model fit.

$$RMSE = \sqrt{\frac{\sum_{1}^{n} E_i^2}{n-p}} \tag{5}$$

$$CV(RMSE) = \frac{RMSE}{A} * 100 \tag{6}$$

4. Case study description

The building chosen for this study is the Procida City Hall, a completely electrified building. It was monitored for nearly 5 years, from February 2018 to December 2022. This period, as shown in Table 3, includes (i) the building's typical energy demand during normal operations before COVID-19 pandemic (period 1), followed by (ii) a significant decrease in energy demand during the COVID-19 pandemic lockdown (period 2), (iii) a new pattern of energy usage influenced by the implementation of smart working procedures and reduced occupancy (period 3), and (iv) a structural implementation of smart-working involving at least 1 day per week remotely and a (period 4), representing the "new normal" condition.

The electricity consumption data gathered for this study has a 15-min resolution, which is then aggregated to provide an hourly profile that corresponds to the time series of outdoor air temperature measurements. The features of the building and equipment were surveyed initially to quantify their rated power and typical operational schedules. Two types of schedules for weekdays have been identified; working hours are from 8:00 to 14:35 on all weekdays, except on Thursday, when the working hours are from 8:00 to 17:30. Moreover, during the weekend (Saturday and Sunday) the office is running from 8:00 to 18:00, because of tourism related services. Lift and electric water heaters are activated on request (i.e. they don't have a precise operational schedule) while appliances such as lighting, computers and printers are continuously operating for the whole office working time (clearly not necessarily simultaneously). In terms of heating and cooling services, the building is equipped with individual electric heaters for the winter season, which are activated by each employee in their room as needed, and a few reversible split air-conditioners for the summer season, with the same control strategy (i.e. no centralised control). The climate in the area is Mediterranean, with mild winters and hot summers. However, due to the holiday season and the building's location in front of the bay, which allows an effective natural ventilation, the impact of high temperatures on building energy consumption is mitigated. The key characteristics of the electricity use in the building are summarized in Table 4. As previously mentioned, the schedules reported indicate the period when appliances are typically on, not necessarily simultaneously at the rated power, as it depends on the coincidence factor in their use, which is not reported here.

Table 3
Monitoring period subdivision.

Period		Dates	Notes
1	Before COVID-19 lockdown	From February 01, 2018 to March 07, 2020	All employees work in presence according to the usual schedule
2	During COVID-19 lockdown	From March 08, 2020 to October 07, 2020	No presence in the first 6 weeks of the period due to hard lock-down while few officers were allowed to enter to the building after that.
3	After COVID-19 lockdown	From October 08, 2020 to November 10, 2021	General rule of remote working applies to this period with an average ratio of 50% of the employees, reduced to 25% in the last 10 weeks.
4	Back to normality with structural smart working option	From November 11, 2021 to December 31, 2022	Most sectors and usage of space are back to normality pre-pandemic plus the structural implementation for Public Administration of 1-day remote work.

5. Results and discussion

In this Section, the results of numerical and visual data analysis are provided and discussed in order to evaluate the performance of the novel TOWT algorithm formulation proposed, highlighting limitations and showing the possibility for additional research inherent in the usage of this technique. First, the outcomes of the process of model fitting are presented in Section 5.1, which compares measured and estimated energy consumption for the various monitoring periods listed in Tables 3 and in Section 4. The impact of the hyper-parameter selection (time interval of the segmentation of the time series, as described in Section 4.1) on model goodness of fit is explored in the same section, employing the indications mentioned in Table 2 to determine if the model may be considered as calibrated or not. The outcomes of model fitting are given in the same section as a graphical interpretation of the results that could aid comprehension of the case study's key characteristics. The calibrated energy models are then deployed in Section 5.2 to compare the performance of the building in different time periods using a counterfactual method. This enables the comparison of building operating performance under same conditions, in this example a typical meteorological year, which would otherwise be impossible due to the necessity to normalise data in relation to weather and occupancy (measured data are highly dependent on the specific weather and operational conditions of the monitored period). Again, numerical and graphical analysis tools are combined in this instance. Then, in Section 5.3, the expected daily load profiles for typical workday and weekend conditions for the winter/summer (January and July) and intermediate (April and October) seasons are compared to emphasise seasonal variability and the impact of building operation choices. Finally, Section 5.4 outlines limitations and prospects for future research.

5.1. Results of models' fitting in the four monitoring periods

In this Section Table 5 presents the results of model testing for the various time periods, considering energy indicators (measured and predicted energy consumption) and statistical indicators (R^2, *NMBE*, and *CV(RMSE)*, introduced in Section 3.2). The models fit the measured data well, within the model calibration limits established by ASHRAE 14:2014 [43] in most of the cases. There is a negligible difference between the predicted and measured electric energy consumption in each period, and the *NMBE* is practically equal to zero. Indeed, this is a consequence of the proposed TOWT model reformulation, which is based on regression (which minimises the distance between model prediction and measurement data) and includes a model for the residuals as a function of the time of week dependence. In addition, R^2 decreases significantly in period 2 compared to period 1; after the lockdown, in periods 3 and 4, R^2 is still lower than in period 1, indicating a lower predictability of electricity demand as a result of less regular operation of the building. The most pertinent statistical indicator to evaluate model performance is *CV(RMSE)*, which is greater than 20% in all cases and greater than 30% (threshold for acceptability of the model as calibrated as shown in Section 3.2) in two cases, period 2 and 3 with a hyper-parameter of 1 month. The impact of the hyper-parameter, representing the period of the temporal segmentation of the time series for the time-of-week model of residuals, is relevant in the proposed reformulation, although less pronounced compared to the previous research conducted on this case study [24], which used TOWT in the RMV2.0

Table 5
Results of TOWT model training in the different monitoring periods (1–4).

	Hyper-parameter	Energy indicators		Statistical indicators		
Monitoring Period	Months (entire/fraction)	Energy measured	Energy predicted	R^2	*NMBE*	*CV (RMSE)*
		kWh	kWh	%	%	%
1	1	126,882	127,516	87.6	0.0	25.2
	0.5	126,882	126,374	91.7	0.0	20.8
2	1	23,363	23,550	65.4	0.0	33.5
	0.5	23,363	23,527	76.8	0.0	28.4
3	1	53,587	53,212	82.5	0.0	31.5
	0.5	53,587	53,909	88.6	0.0	26.0
4	1	74,002	74,520	66.9	0.0	32.6
	0.5	74,002	73,632	76.4	0.0	28.5

Table 4
Building electricity end-uses and operational schedules.

End-use	Type of appliance	Power	Period of operation	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
		kW	–	–	–	–	–	–	–	–
Office work	**Printers**	12.3	Year	08:00–14:35	08:00–14:35	08:00–14:35	08:00–17:30	08:00–14:35	08:00–18:00	08:00–18:00
	Computers	14.7	Year	08:00–14:35	08:00–14:35	08:00–14:35	08:00–17:30	08:00–14:35	08:00–18:00	08:00–18:00
Lighting	**Lighting**	6.8	Year	08:00–14:35	08:00–14:35	08:00–14:35	08:00–17:30	08:00–14:35	08:00–18:00	08:00–18:00
Heating/ cooling	**Electric heaters**	49.5	Winter	08:00–14:35	08:00–14:35	08:00–14:35	08:00–17:30	08:00–14:35	08:00–18:00	08:00–18:00
	Split air-conditioners	8.8	Year	08:00–14:35	08:00–14:35	08:00–14:35	08:00–17:30	08:00–14:35	08:00–18:00	08:00–18:00
	Fans	0.7	Summer	08:00–14:35	08:00–14:35	08:00–14:35	08:00–17:30	08:00–14:35	08:00–18:00	08:00–18:00
Water heating	**Electric water heaters**	3.6	Year	On request	On request	On request	On request	On request	On request	On request
Other uses	**Lift**	13	Year	On request	On request	On request	On request	On request	On request	On request
	Others	2.5	Year	On request	On request	On request	On request	On request	On request	On request

implementation, where the hyper-parameter was the time scale of the weighting function, i.e. the number of days nearby the day of the week predicted. In the paper presenting the initial implementation of the TOWT algorithm, Mathieau et al. [34] highlighted the potential problem of model overfitting, stating that the model can easily overfit the data, reducing its ability to generalise. This aspect will be discussed in the context of the interpretation of model results later on in this section.

Following the analysis of energy and statistical indicators reported in Table 5 for the various monitoring periods and fitted models, the visual interpretation of results is discussed below, taking into account two peculiar aspects: the temperature-dependent model component and the time-of- week-dependent model component. For all monitoring periods, a scatterplot identifying the temperature dependence of load is presented alongside a time series of measured electric load profiles plotted against predicted data. The visualization of data for the first monitoring period (from February 01, 2018 to March 07, 2020) is limited to one year (from February 2018 to February 2019) to keep the amount of data in the visualization comparable to periods 2, 3 and 4. It can be clearly seen in Fig. 1 on the left, how the temperature regression in high demand hours, determined using the approach described in Section 3.1, clearly depends on outdoor air temperature (practically a straight line below 16 °C and with a moderate slope also above 22 °C), while the regression in low demand hours remains substantially horizontal (i.e. non dependent on temperature). As described in Section 2 regarding change-point models, when a significant portion of electric load is related to heating and/or cooling services, outdoor air temperature is frequently the most influential variable. On the right hand side of Fig. 1 it is possible to see that the model fits the electricity demand time series (in this case using 0.5 months hyper-parameter) well, with the exception of some peak conditions, which are more difficult to predict compared to the typical weekly behaviour.

Moving to period 2 corresponding to the COVID-19 pandemic lockdown (from March 08, 2020 to October 07, 2020), it can be seen in Fig. 2 that the range of electric load is much smaller (as expected due to the end-use of the building). As a result, the scale of the y-axis on the left side of Fig. 2 has been reduced from 60 to 30 kW. The data distribution appears less predictable and uniform. The regression with respect to temperature demonstrates that the hours with low demand differ significantly from the overall temperature regression and from the hours with high demand. Lastly, the model is able to fit the electricity demand time series shown in Fig. 2 on the right within the calibration ranges when using 0.5 months as hyper-parameter, but it cannot reproduce correctly the irregular spikes in the series.

Figs. 3 and 4 depict the monitoring periods 3 (from October 08, 2020 to November 10, 2021) and 4 (from November 11, 2021 to December 31, 2022) described in Table 5, following the COVID-19 lockdown. Period 4 exhibits a more pronounced temperature dependence above 22 °C but the temperature response component in periods 3 and 4 remains relatively similar. The temperature dependence of load in winter appears to be very different from period 1; while in periods 3 and 4 it is possible to observe an increase in the slope of the regression for hours of high demand below 16 °C, the slope is close to 0 or even negative at lower temperatures, whereas it was positive in Fig. 1, for period 1. This indicates a significant change in the operational regime of the building between periods 3 and 4 compared to period 1, whereas between periods 1 and 4 the summer behaviour appears to be quite similar. Also in this instance, the model fits the time series data slightly better than in period 2, but the inability to reproduce spikes that are irregular with regard to the time of week component of the regression remains.

The TOWT algorithm reformulation was able to successfully fit models in challenging conditions (as indicated also by the results of previous work on this case study) creating calibrated models before (period 1), during (period 2), and after covid (periods 3 and 4), with the post-covid situation being more dynamic than the pre-covid situation, even for the period 4 ("back to normality"). Interestingly, the examination of the temperature response highlighted operational changes that would have been difficult to uncover otherwise and this aspects is particularly relevant for the research which is aimed at integrating the use of TOWT algorithm with piecewise linear change-point models, for the reasons discussed in Section 2, weather normalisation in particular. Nonetheless, the monitored data correspond to distinct dsltime periods with distinct meteorological and operational conditions. For this reason, the trained and calibrated models are used to generate forecasts using a typical meteorological year (TMY) weather data file as input, enabling as such a comparison of the pre-covid and post-covid situations under idealised typical conditions, using a conterfactual approach. This is illustrated in Section 5.2.

5.2. Model predictions comparison for a typical meteorological year

The models fitted for periods 1, 3, and 4 are compared using a counterfactual method in this section. Period 2, during the COVID-19 lockdown, is excluded since its behaviour is not reflective of the building's operations, but rather an outlier. The use of a counterfactual aims to predict what the building operation would have been, based on the characteristics of the periods during which the models are trained, but applied to a typical meteorological year (TMY), in order to enable comparability in reference weather conditions, which would otherwise be impossible due to the specificity of the monitored weather conditions and operating schedules of every single period. The electrical demand signature shown in Fig. 5 indicates that the period 1 (pre-COVID-19) behaviour has higher load peaks throughout the winter, whereas these

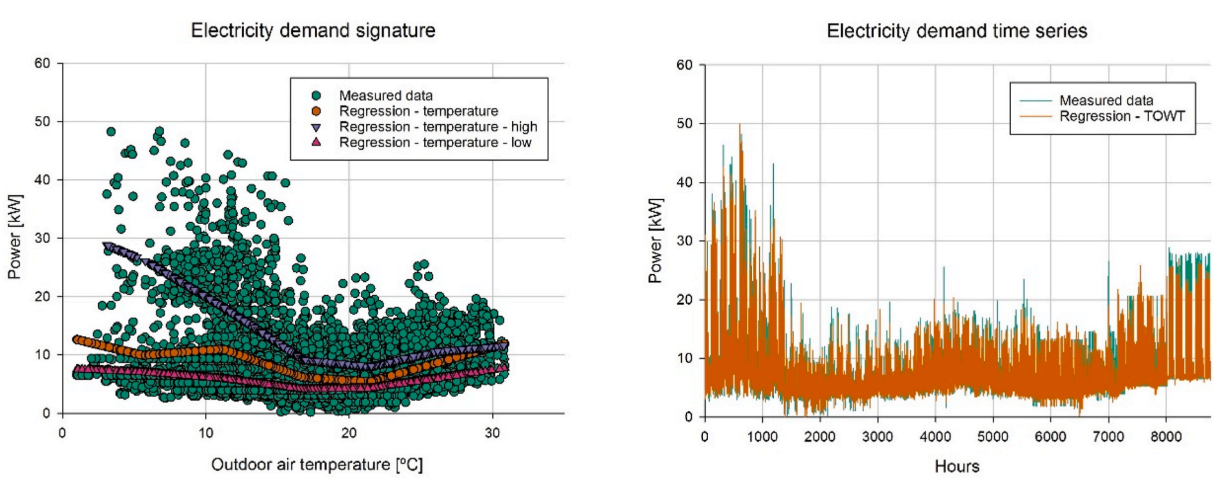

Fig. 1. Electricity demand signature (left) and time series (right) with hourly data, TOWT model – Period 1.

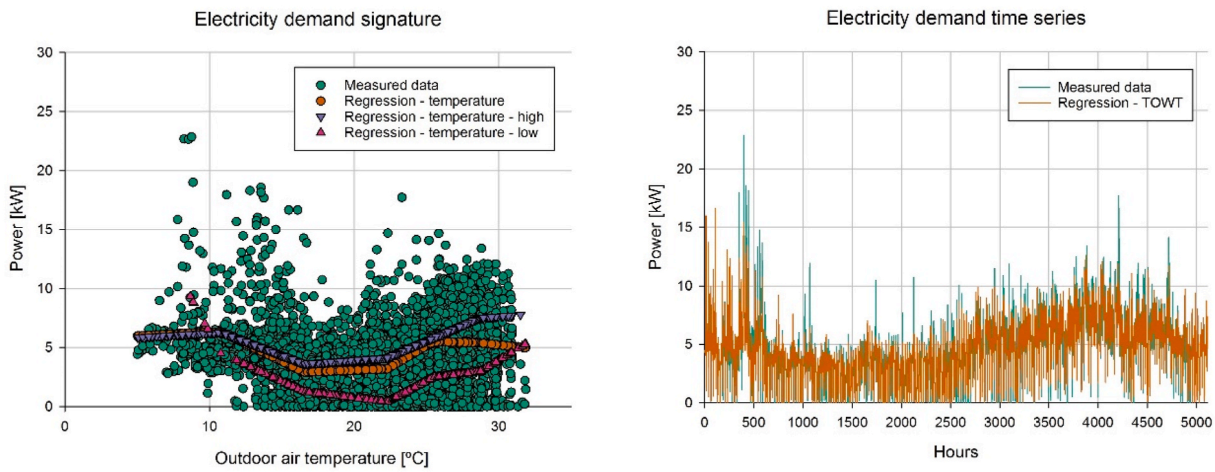

Fig. 2. Electricity demand signature (left) and time series (right) with hourly data, TOWT model – Period 2.

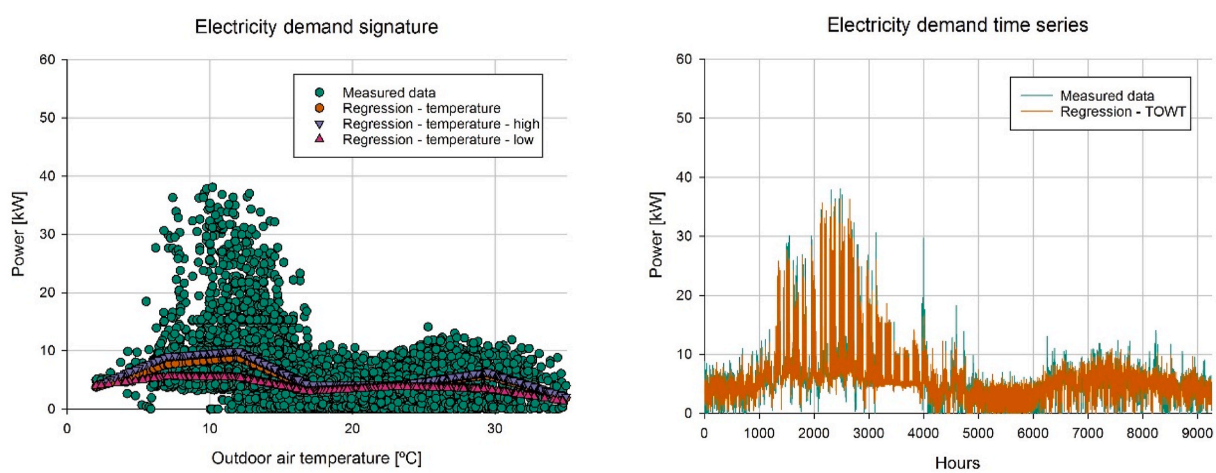

Fig. 3. Electricity demand signature (left) and time series (right) with hourly data, TOWT model – Period 3.

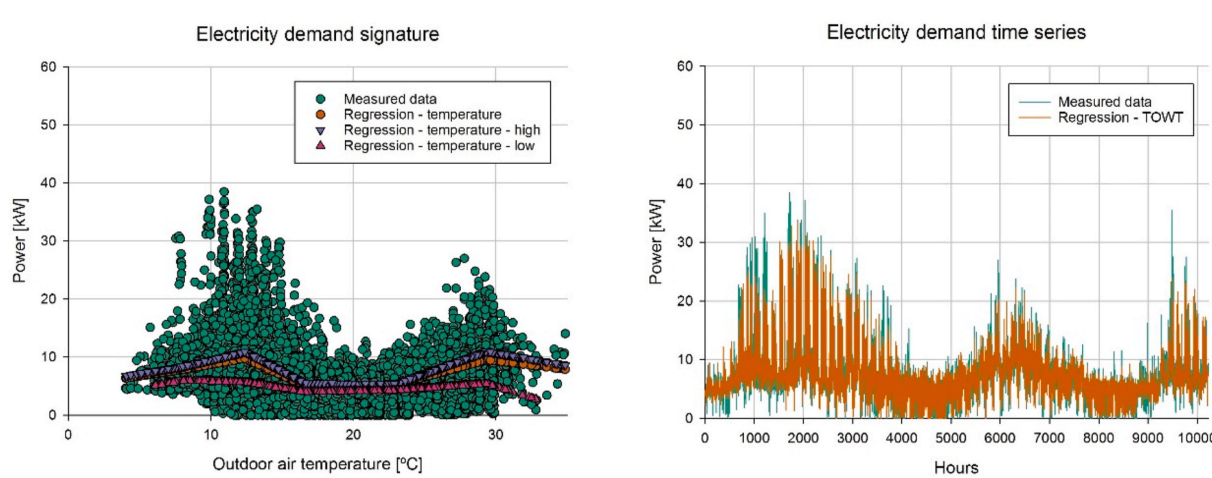

Fig. 4. Electricity demand signature (left) and time series (right) with hourly data, TOWT model – Period 4.

peaks are reduced in periods 3 and 4 (post-COVID-19) due to smart working. While the peaks vary significantly between times, the regression with respect to outdoor air temperature, which captures the "average" temperature response and is used to distinguish between high and low demand conditions as explained in Section 3.1, remains relatively stable. In other words, the weekday component of the model, which reflects the dynamic operation schedule of the building, is considerably more variable than the building's predominant energy signature as a function of outdoor air temperature. Indeed, the energy signature regression is the part of the TOWT modelling workflow which makes it comparable with piecewise-linear change-points method recalled in Section 3 and makes the results more easily "interpretable".

The predictions for the different monitoring time periods, reported in Table 6, suggest that the pre-COVID-19 operational strategy consumed more energy than the post-COVID-19 strategy (- 26.7% in period 3 and - 10.07% in period 4) compared to period 1. However, in period 4, energy consumption is 21.8% more than in period 3. Overall, this emphasises the significance of gaining an accurate understanding of building operation strategies and behaviour (i.e. overcoming the "information gap" described in Section 2), which could aid in energy conservation (together with other more pervasive efficiency measures, of course).

The predicted load profiles for a typical meteorological year are then sorted to create load duration curves, which are useful for understanding peak conditions (occurring only a few hours per year) and prevailing operational conditions, for the reasons outlined in Section 2. Fig. 6 demonstrates that the peak load for period 1 is greater than for periods 3 and 4, and that post-COVID-19 conditions tend to flatten the load duration curve. Despite the fact that the curve for period 4 is higher than the one for period 3, they have a comparable behaviour.

5.3. Load profiles comparison for a typical meteorological year

Following the counterfactual approach described in Section 5.2, the models fitted for periods 1, 3, and 4 are compared based on the typical daily profiles for an average working day (weekday) and weekend day (weekend) during specific months. The month's temperature conditions are those of a typical meteorological year (TMY) weather data file, to enable a meaningful comparison. The months of January and July (winter/summer conditions) and April and October (spring/autumn conditions) are considered to model both extreme and intermediate cases, with the purpose of demonstrating both the impact of seasonal variability (typical workday and weekend conditions for the winter/summer seasons) and the impact of building operation choices. Immediately evident from Fig. 7 is that the use of electric heaters in the winter results in a higher average demand for January compared to the summer, despite the behaviour appearing to be quite different in the different periods, with a flat demand during the average working day for the pre-COVID-19 conditions, while a more pronounced peak for post-COVID-9 conditions when it is concentrated in the morning in period 3 and during midday in period 4. There are variations in demand throughout the year, but weekend demand is relatively stable. Specifically, the summer weekday behaviour of period 4 appears remarkably comparable to that of period 1.

Moving on to the analysis of intermediate seasons, Fig. 8 reveals that the difference between working days and weekends is present but far less pronounced than in Fig. 7. This provides a basis for discussion regarding the amount of energy that could be saved by better scheduling

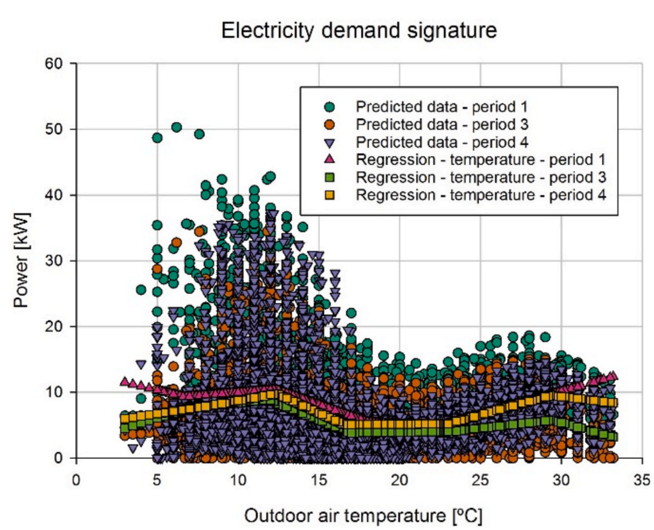

Fig. 5. Electricity demand signatures prediction for a typical meteorological year using the model trained in period 1, 3 and 4.

Table 6
Energy demand prediction for a typical meteorological year using the model trained in period 1, 3 and 4.

Model training period	Description	Overall reduction	Relative reduction/ increase
1	Before COVID-19 lockdown	0 (baseline)	–
3	After COVID-19 lockdown (initial period)	- 26.7%	0 (baseline)
4	Back to normality with structural smart working option	- 10.7%	+21.8%

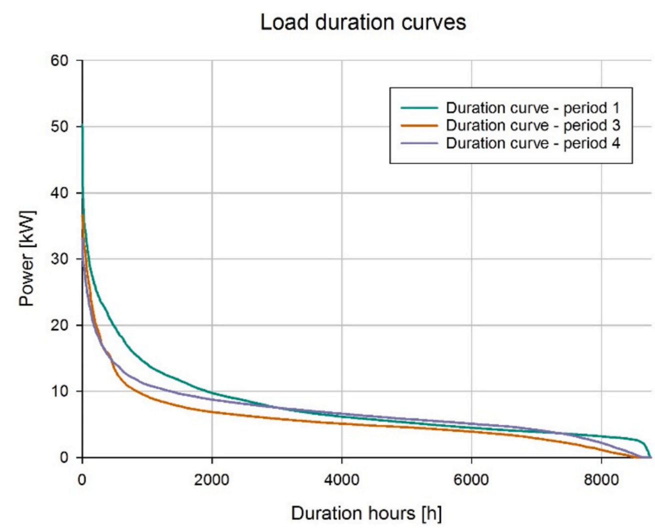

Fig. 6. Load duration curves for a typical meteorological year using the model trained in period 1, 3 and 4.

of appliances, i.e. making them operate only when needed, during peak demand periods in winter and summer. In general, however, the normal operational profiles, both throughout the week and on weekends, are lower after COVID-19 than they were before, coherently with the calculated savings which were reported in Table 5.

5.4. Limitations and further research

Due to variable operational regimes, the pre-COVID-19 (period 1) load profiles were more predictable than the post-COVID-19 (periods 3 and 4) ones, as indicated by the results of the numerical and visual analysis of electricity demand time series on which TWOT models were fitted, separating the different time periods. The COVID-19 lockdown (period 2) was an exception in itself, and less predictable conditions were expected in combination with a decrease in energy demand as a result of remote working and reduced operation hours. Regarding the data-driven technique chosen, the Time Of Week and Temperature (TOWT) algorithm, chosen for its simplicity and ease of use in the initial phase of experimentation [24], has been reformulated so that, for the temperature response component, it is comparable to the consolidated piecewise linear modelling approaches employed for M&V. The thermal response can be viewed via the lens of the approximated physical interpretation of coefficients, which improves its ability to be understood in human terms and to create insights. Considering then the issues of hyper-parameter tuning (e.g., indicating range and step of hyper-parameters) and subsetting of the original time series, the procedure can be automated further in the event of changes to the building's operating schedules (e.g. using contextual information, in this case the periods before, during and after COVID, characterised by different operational regimes). As illustrated in Section 5.1, the outputs of an automated model fitting process can be ranked according to their

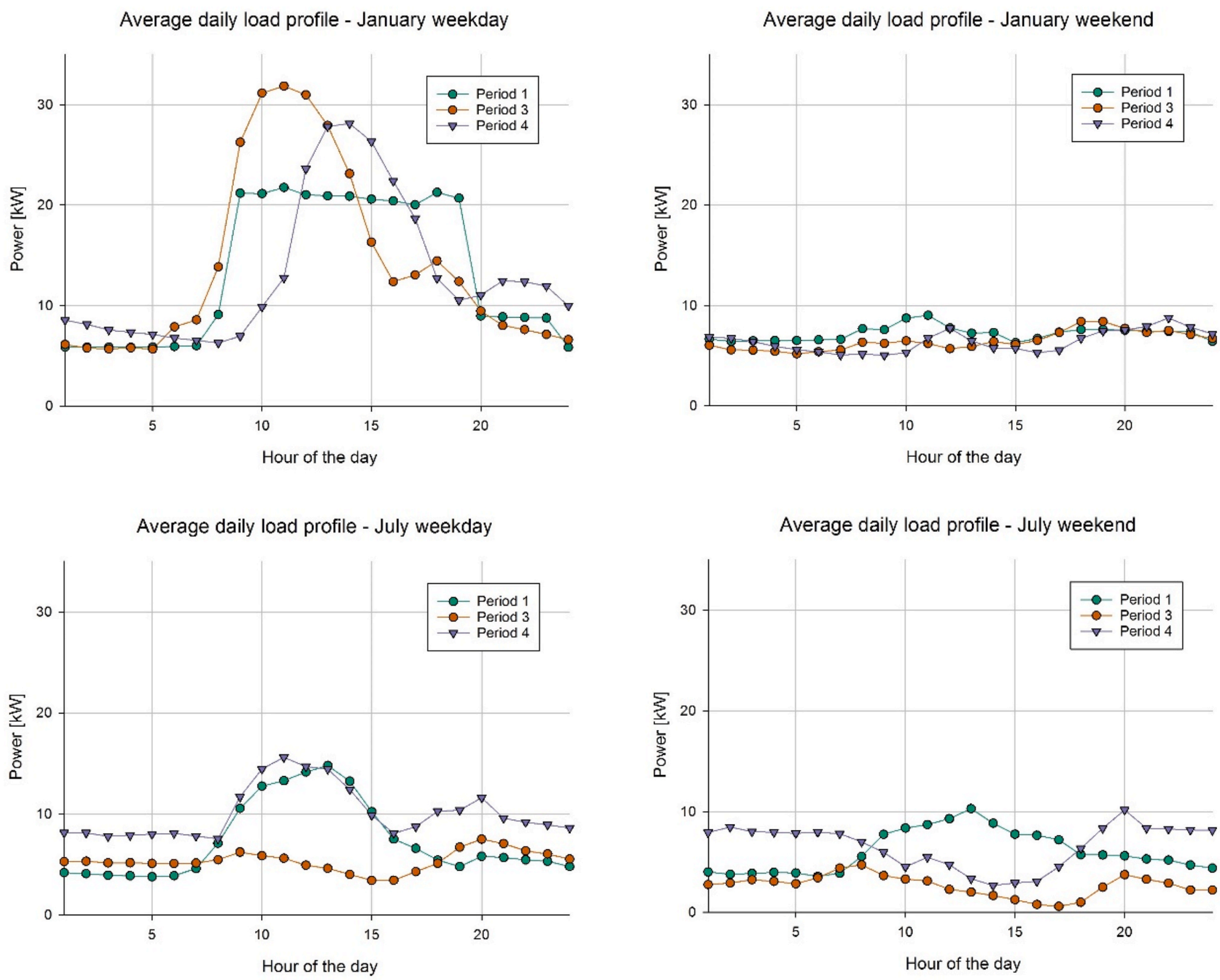

Fig. 7. Load profiles comparison for typical weekday and weekend conditions in January and July for period 1, 3 and 4.

goodness of fit with respect to statistical indicators (R^2, *NMBE*, and *CV (RMSE)* presented in Section 3.2) while simultaneously being presented graphically. This would enable the method's interpretability to be preserved even in a highly automated process whose primary objective is extracting the two main components of building behaviour: the time of week (schedule of operation) and the temperature response (piecewise linear function, with approximated physical interpretation).

This research indicates that the operational regime of the building (i. e. time of week component) is far more variable than the temperature response, which represents the "average" operating conditions at different times as a function of outdoor air temperature. Retaining the same formulation, the temperature response can be developed further with respect to hours with high demand and low demand. In fact, the algorithmic reformulation proposed to enable the seamless integration of TOWT with piecewise linear change-point methods opens up the possibility to include additional variables (other than the time series stamp and temperature, as in its original formulation), which will be the focus of future research. Additionally, automated cross-validation could be applied to test the robustness of model estimates. Lastly, the use of these regression-based approaches in conjunction with statistical process control techniques (SPC) may enable real-time identification of operational issues or drifting of building performance relative to expected behaviour (i.e. anomaly detection), leveraging numerical and graphical techniques once more. In other words, these methods are potentially capable of operating as "digital twins" that work as virtual representations of energy-related processes within buildings.

6. Conclusions

The need to accelerate the decarbonisation of the building stock environment necessitates a substantial switch in consumption from fossil fuels to electricity and the electrification process poses the problem of rethinking energy efficiency while addressing the inherent flexibility concerns. This, in turn, requires a better understanding of the electricity load profiles and, for this reason, this study proposed a reformulation of the Time Of Week and Temperature (TOWT) algorithm, which has been tested on a case study, the Procida Town Hall that is completely electrified and has been monitored for nearly five years, from February 2018 to December 2022. The novel model formulation is calibrated using hourly interval data with a Coefficient of Variation of Root Mean Square Error in the range of 20.0–28.5% throughout the 4 monitoring periods considered. The counterfactual analysis of electricity consumption, conducted using a Typical Meteorological Year weather data file, indicates a 10.7–26.7% decrease in electricity consumption due to operational adjustments following COVID-19 lockdown, showing the potential impact of behavioural change. Overall, the goal of the research was to attain a performance comparable to the original implementation of TOWT algorithm (described in Section 5.1) while enhancing its interpretability and automation potential (both discussed in Section 5.4). Particularly, interpretability is determined by the ability to represent the electric load based on a weekly schedule (i.e., differentiating each day and hour of the week to capture specific recurring weekly patterns of operation) and on outdoor air temperature response,

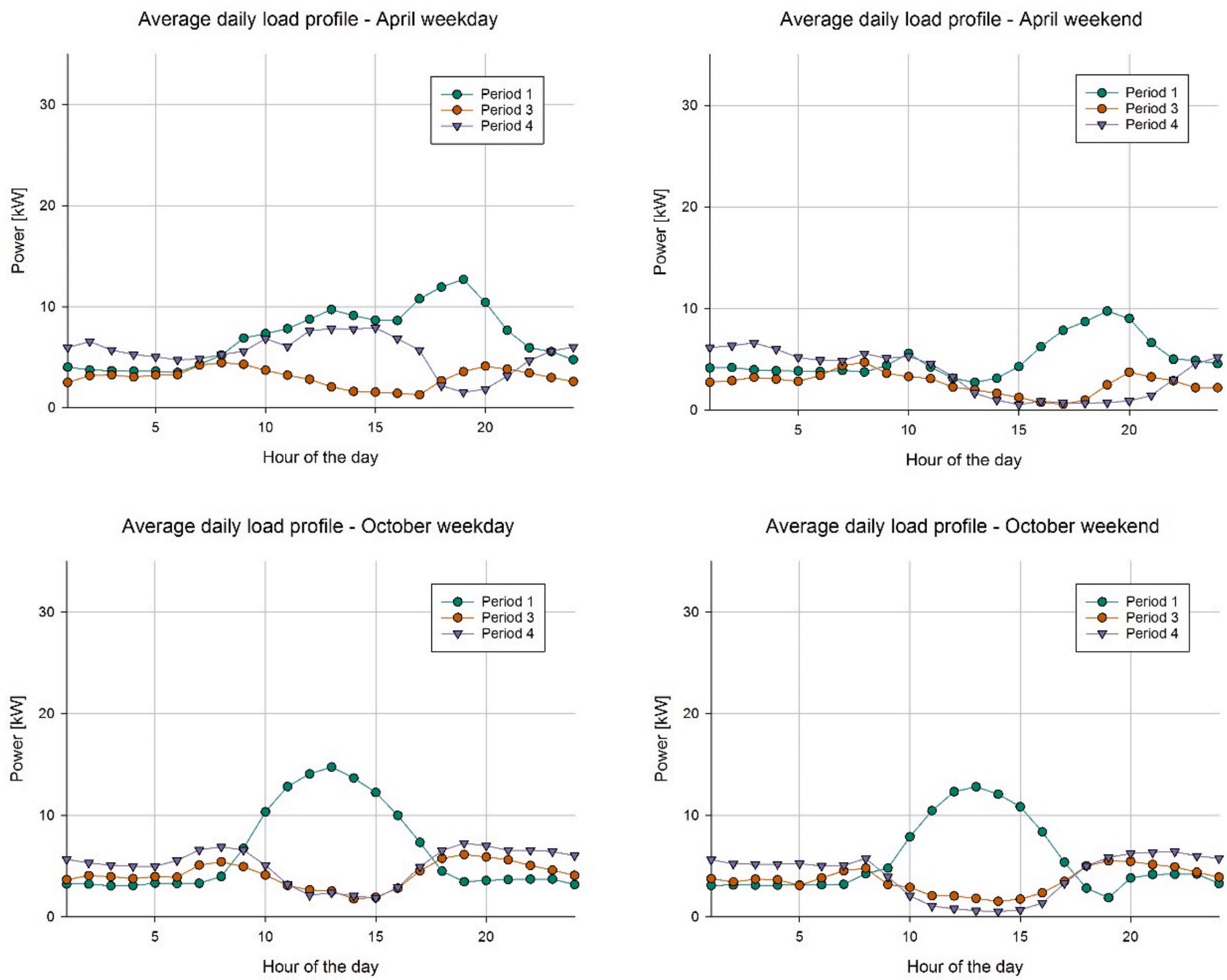

Fig. 8. Load profiles comparison for typical weekday and weekend conditions in April and October for period 1, 3 and 4.

which accounts for electric loads determined by heating and cooling services (i.e. temperature dependent component of energy consumption). Thus, the (relatively) more stable component of building performance (i.e. the temperature response component) is separated from the most dynamic component (i.e. the time of week schedule of operation).

Finally, the use of statistical process control (SPC) is a potential further development for the proposed technique, which may enable real-time identification of operational anomalies by combining numerical and graphical techniques to obtain an interpretable "digital twin", i.e. a dynamic virtual representation of energy-related processes within the building, easily understandable in human terms.

Author contributions

Benedetto Nastasi: Concept, Investigation, Methodology, Data curation, Writing-reviewing and editing, Project administration; **Massimiliano Manfren**: Concept, Investigation, Methodology, Visualization, Software, Writing-reviewing and editing, Project administration.

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 data that has been used is confidential.

References

[1] Norton B, Gillett WB, Koninx F. Decarbonising buildings in europe: a briefing paper. Proc Inst Civ Eng - Energy 2021;174:147–55. https://doi.org/10.1680/jener.21.00088.

[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] Haas R, Duic N, Auer H, Ajanovic A, Ramsebner J, Knapek J, et al. The photovoltaic revolution is on: how it will change the electricity system in a lasting way. Energy 2023;265:126351. 10.1016/j.energy.2022.126351.

[4] Slorach PC, Stamford L. Net zero in the heating sector: technological options and environmental sustainability from now to 2050. Energy Convers Manag 2021;230:113838. 10.1016/j.enconman.2021.113838.

[5] Calise F, Cappiello FL, Dentice d'Accadia M, Vicidomini M. Smart grid energy district based on the integration of electric vehicles and combined heat and power generation. Energy Convers Manag 2021;234:113932. 10.1016/j.enconman.2021.113932.

[6] Rosenow J, Eyre N. Reinventing energy efficiency for net zero. Energy Res Soc Sci 2022;90:102602. 10.1016/j.erss.2022.102602.

[7] D'Ettorre F, Banaei M, Ebrahimy R, Pourmousavi SA, Blomgren EMV, Kowalski J, et al. Exploiting demand-side flexibility: state-of-the-art, open issues and social perspective. Renew Sustain Energy Rev 2022;165:112605. j.rser.2022.112605.

[8] Manfren M, Aste N, Leonforte F, Del Pero C, Buzzetti M, Adhikari RS, et al. Parametric energy performance analysis and monitoring of buildings-HEART project platform case study. Sustain Cities Soc 2020:102296.

[9] Dongellini M, Naldi C, Morini GL. Influence of sizing strategy and control rules on the energy saving potential of heat pump hybrid systems in a residential building. Energy Convers Manag 2021;235:114022. j.enconman.2021.114022.
[10] Ala-Juusela M, Crosbie T, Hukkalainen M. Defining and operationalising the concept of an energy positive neighbourhood. Energy Convers Manag 2016;125:133–40. j.enconman.2016.05.052.
[11] Adhikari RS, Aste N, Manfren M. Optimization concepts in district energy design and management – a case study. Energy Proc 2012;14:1386–91. j.egypro.2011.12.1106.
[12] Kona A, Bertoldi P, Kılkış Ş. Covenant of mayors: local energy generation, methodology, policies and good practice examples. Energies 2019;12. https://doi.org/10.3390/en12060985.
[13] Kılkış Ş, Krajačić G, Duić N, Rosen MA, Ahmad Al-Nimr M. Accelerating mitigation of climate change with sustainable development of energy, water and environment systems. Energy Convers Manag 2021;245:114606. 10.1016/j.enconman.2021.114606.
[14] Rehman H ur, Korvola T, Abdurafikov R, Laakko T, Hasan A, Reda F. Data analysis of a monitored building using machine learning and optimization of integrated photovoltaic panel, battery and electric vehicles in a Central European climatic condition. Energy Convers Manag 2020;221:113206. 10.1016/j.enconman.2020.113206.
[15] Salom J, Marszal AJ, Widén J, Candanedo J, Lindberg KB. Analysis of load match and grid interaction indicators in net zero energy buildings with simulated and monitored data. Appl Energy 2014;136:119–31. 10.1016/j.apenergy.2014.09.018.
[16] National Institute of Building Sciences (NIBS). New buildings Institute, national environmental balancing bureau. Data needs for achieving high performance buildings: high-performance building data collection initiative. 2011. Washington, DC.
[17] Mancini F, Lo Basso G. How climate change affects the building energy consumptions due to cooling, heating, and electricity demands of Italian residential sector. Energies 2020;13. https://doi.org/10.3390/en13020410.
[18] Fabrizio E, Filippi M, Virgone J. An hourly modelling framework for the assessment of energy sources exploitation and energy converters selection and sizing in buildings. Energy Build 2009;41:1037–50. 10.1016/j.enbuild.2009.05.005.
[19] Wang S, Zang Y, Ge W, Wang A, Li D, Tang J. Data-driven real-time pricing strategy and coordinated optimization of economic load dispatch in electricity market. Front Energy Res 2021;9:434.
[20] Zhang J, Zhang H, Ding S, Zhang X. Power consumption predicting and anomaly detection based on transformer and K-means. Front Energy Res 2021;9:681.
[21] IEA. Annex 79 - occupant-centric building design and operation n.d.
[22] 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.
[23] Price P. Methods for analyzing electric load shape and its variability. Lawrence Berkeley National Laboratory Report LBNL-3713E; 2010.
[24] 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. 10.1016/j.buildenv.2022.109279.
[25] Interpretable machine learning, Section 3.2 Taxonomy of Interpretability Methods, Christopher Molnar (https://christophm.github.io/interpretable-ml-book/taxonomy-of-interpretability-methods.html).
[26] ISO/IEC. ISO/IEC TR 29119-11:2020(en) Software and systems engineering. Software testing. Part 11: Guidelines on the test AI-base syst 2020.
[27] Zhang X, Kong X, Yan R, Liu Y, Xia P, Sun X, et al. Data-driven cooling, heating and electrical load prediction for building integrated with electric vehicles considering occupant travel behavior. Energy 2023;264:126274. 10.1016/j.energy.2022.126274.
[28] Liang X, Zhu X, Chen K, Chen S, Jin X, Du Z. Endowing data-driven models with rejection ability: out-of-distribution detection and confidence estimation for black-box models of building energy systems. Energy 2023;263:125858. j.energy.2022.125858.
[29] Sun Y, Haghighat F, Fung BCM. Trade-off between accuracy and fairness of data-driven building and indoor environment models: a comparative study of pre-processing methods. Energy 2022;239:122273. 10.1016/j.energy.2021.122273.
[30] 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. 10.1016/j.enbuild.2021.111054.
[31] Fu H, Baltazar J-C, Claridge DE. Review of developments in whole-building statistical energy consumption models for commercial buildings. Renew Sustain Energy Rev 2021;147:111248. 10.1016/j.rser.2021.111248.
[32] Franconi E, Gee M, Goldberg M, Granderson J, Guiterman T, Li M, et al. The status and promise of advanced M&V: an overview of "M&V 2.0" methods, tools, and applications. 2017.
[33] 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. 10.1016/j.enbuild.2018.02.023.
[34] 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.
[35] Borgeson SD. Targeted efficiency: using customer meter data to improve efficiency program outcomes. Berkeley: University of California; 2013.
[36] RMV2.0 - LBNL M&V2.0 Tool (https://lbnl-eta.github.io/RMV2.0/).
[37] NMECR (https://kw-labs.github.io/nmecr/).
[38] CalTRACK. CalTRACK Methods (http://docs.caltrack.org/en/latest/methods.html).
[39] EVO's Advanced M&V Testing Portal. https://mvportal.evo-world.org/.
[40] Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc 2012;107:1590–8. https://doi.org/10.1080/01621459.2012.737745.
[41] Killick R, Haynes K, Eckley IA. {changepoint}: an {R} package for changepoint analysis. J Stat Softw 2014;58:1–19.
[42] Kissock JK, Haberl JS, Claridge DE. Inverse modeling toolkit: numerical algorithms. ASHRAE Trans 2003;109:425.
[43] ASHRAE. ASHRAE guideline 14-2014: measurement of energy, demand, and water savings. Atlanta, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2014. 2014.
[44] Paulus MT, Claridge DE, Culp C. Algorithm for automating the selection of a temperature dependent change point model. Energy Build 2015;87:95–104. 10.1016/j.enbuild.2014.11.033.
[45] Paulus MT. Algorithm for explicit solution to the three parameter linear change-point regression model. Sci Technol Built Environ 2017;23:1026–35.
[46] 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.
[47] 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.
[48] Spinoni J, Vogt JV, Barbosa P, Dosio A, McCormick N, Bigano A, et al. Changes of heating and cooling degree-days in Europe from 1981 to 2100. Int J Climatol 2018;38:e191–208. 10.1002/joc.5362.
[49] Omarov B, Memon SA, Kim J. A novel approach to develop climate classification based on degree days and building energy performance. Energy 2023;267:126514. 10.1016/j.energy.2022.126514.
[50] Park S, Shim J, Song D. Issues in calculation of balance-point temperatures for heating degree-days for the development of building-energy policy. Renew Sustain Energy Rev 2021;135:110211. 10.1016/j.rser.2020.110211.
[51] Meng Q, Xi Y, Zhang X, Mourshed M, Hui Y. Evaluating multiple parameters dependency of base temperature for heating degree-days in building energy prediction. Build Simul 2021;14:969–85. https://doi.org/10.1007/s12273-020-0752-9.
[52] Plaga LS, Bertsch V. Methods for assessing climate uncertainty in energy system models — a systematic literature review. Appl Energy 2023;331:120384. 10.1016/j.apenergy.2022.120384.
[53] Lizana J, Halloran CE, Wheeler S, Amghar N, Renaldi R, Killendahl M, et al. A national data-based energy modelling to identify optimal heat storage capacity to support heating electrification. Energy 2023;262:125298. 10.1016/j.energy.2022.125298.
[54] ISO 15927-6:2007 Hygrothermal performance of buildings — Calculation and presentation of climatic data — Part 6: Accumulated temp differ (degree-days).
[55] Kheiri F, Haberl JS, Baltazar J-C. Split-degree day method: a novel degree day method for improving building energy performance estimation. Energy Build 2023;289:113034. 10.1016/j.enbuild.2023.113034.
[56] 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.
[57] 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.
[58] 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. 10.1016/j.rser.2022.112686.
[59] ISO 16346:2013, Energy performance of buildings. Assess overall energy perform. 2013.
[60] Lei L, Wu B, Fang X, Chen L, Wu H, Liu W. A dynamic anomaly detection method of building energy consumption based on data mining technology. Energy 2023;263:125575. 10.1016/j.energy.2022.125575.
[61] Fan C, Lei Y, Sun Y, Piscitelli MS, Chiosa R, Capozzoli A. Data-centric or algorithm-centric: exploiting the performance of transfer learning for improving building energy predictions in data-scarce context. Energy 2022;240:122775. 10.1016/j.energy.2021.122775.
[62] Lamagna M, Nastasi B, Groppi D, Nezhad MM, Garcia DA. Hourly energy profile determination technique from monthly energy bills. Build Simul 2020;13:1235–48. https://doi.org/10.1007/s12273-020-0698-y.
[63] Abushakra B, Reddy A, Singh Vipul. ASHRAE research project report 1404-RP, measurement, modeling, analysis and reporting protocols for short-term M&V of whole building energy performance. USA: Arizona State University; 2012.
[64] Leiria D, Johra H, Marszal-Pomianowska A, Pomianowski MZ. A methodology to estimate space heating and domestic hot water energy demand profile in residential buildings from low-resolution heat meter data. Energy 2023;263:125705. 10.1016/j.energy.2022.125705.
[65] International Performance Measurement and Verification Protocol (IPMVP). Volume I: concepts and options for determining energy and water savings. EVO 10000–1. Washington, DC: Efficiency Valuation Organization; 2012 [EVO)].
[66] Femp. M&V guidelines: measurement and verification for performance-based contracts. 2015., version 4.0.
[67] ISO 50006:2014, Energy management systems. Measuring energy performance using energy baselines (EnB) and energy performance indicators (EnPI). Gen princ guid 2014.

[68] Lipton ZC. The Mythos of Model Interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 2018;16:31–57.
[69] Chen X, Abualdenien J, Singh MM, Borrmann A, Geyer P. Introducing causal inference in the energy-efficient building design process. Energy Build 2022;277:112583. 10.1016/j.enbuild.2022.112583.
[70] 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.
[71] Zhang Y, Teoh BK, Wu M, Chen J, Zhang L. Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. Energy 2023;262:125468. 10.1016/j.energy.2022.125468.
[72] 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. 10.1016/j.adapen.2023.100123.
[73] Naji S, Keivani A, Shamshirband S, Alengaram UJ, Jumaat MZ, Mansor Z, et al. Estimating building energy consumption using extreme learning machine method. Energy 2016;97:506–16. 10.1016/j.energy.2015.11.037.
[74] Galli A, Piscitelli MS, Moscato V, Capozzoli A. Bridging the gap between complexity and interpretability of a data analytics-based process for benchmarking energy performance of buildings. Expert Syst Appl 2022;206:117649. 10.1016/j.eswa.2022.117649.
[75] Miller C. What's in the box?! Towards explainable machine learning applied to non-residential building smart meter classification. Energy Build 2019;199:523–36. 10.1016/j.enbuild.2019.07.019.
[76] Bollinger LA, Davis CB, Evins R, Chappin EJL, Nikolic I. Multi-model ecologies for shaping future energy systems: design patterns and development paths. Renew Sustain Energy Rev 2018;82:3441–51. 10.1016/j.rser.2017.10.047.
[77] 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.
[78] 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. 10.1016/j.dibe.2020.100024.
[79] ISO 50001. Energy manag syst - Requir guid use 2018.
[80] Report on European Energy Efficiency Services markets and quality. Horizon, project QualitEE 2018, https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bf6a077e&appId=PPGMS; 2020.
[81] 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. 10.1016/j.rser.2021.111490.
[82] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. OTexts; 2018.
[83] Uncertainty assessment for IPMVP, international performance measurement & verification protocol. Efficiency Valuation Organization (EVO); 2019.