Applied Thermal Engineering 263 (2025) 125296

Contents lists available at ScienceDirect

Applied Thermal Engineering

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

Research Paper

Enhanced methodology for disaggregating space heating and domestic hot water heat loads of buildings in district heating networks

Nicola Borgato [a,*], Sara Bordignon [a], Enrico Prataviera [a], Roberto Garay-Martinez [b], Angelo Zarrella [a]

[a] Department of Industrial Engineering - Applied Physics Section, University of Padova, Via Venezia 1, 35131 Padova, Italy
[b] Instituto Tecnológico Deusto, DeustoTech - Universidad de Deusto, Avda. Universidades 24, 48007, Bilbao, Spain

ARTICLE INFO

Keywords:
Linear regression models
Energy signature curve
Heat load disaggregation
Season threshold identification
Heat load clusterization

ABSTRACT

This paper presents an innovative approach to disaggregate a building's global heat consumption into space heating and domestic hot water heat load components using Energy Signature Curve models. The study addresses the challenges associated with these models, which often fail to represent daily trends accurately and do not account for dynamic changes in building usage. Four approaches based on linear regression models are compared to determine the most accurate method for space heating and domestic hot water disaggregation. The state-of-the-art Energy Signature Curve is compared with three improved alternatives. A new algorithm for automatic season threshold identification is proposed. The comparison with consumption data indicates that the proposed methodology significantly improves the accuracy in heat load disaggregation, with the superior performance provided by the model based on a 24-hour energy threshold. This advancement can potentially optimize district heating network management and support retrofit interventions by providing detailed consumption profiles.

1. Introduction

In recent years, the focus on sustainability in the energy sector has grown, leading to significant research aimed at improving energy efficiency. The building sector accounts for 40 % of the EU's total energy consumption and over 33 % of its energy-related greenhouse gas emissions [1,2]. Space heating (SH) and domestic hot water (DHW) systems collectively account for over 27 % of final energy usage and 20 % of greenhouse gas emissions in the EU, with SH contributing approximately 85 % and DHW around 15 % to heat demand [3]. Improving energy efficiency in buildings' energy systems is a challenge for new constructions and primarily for existing buildings [4]. Energy Roadmap 2050 [5] sets ambitious goals for improving energy efficiency in buildings and reducing emissions. Effective management of building energy consumption is essential for sustainability and lowering greenhouse gas emissions. [6]. However, researchers and policymakers often lack complete data on building envelopes, usage, occupant behavior, weather, and energy systems, making it challenging to predict energy consumption and retrofit impacts. The increasing availability of smart meters worldwide enhances understanding of consumption patterns and provides valuable data for energy load planning, enabling more accurate energy assessments and retrofit proposals [7–9].

A key challenge in this effort lies in accurately disaggregating the energy consumption in buildings, particularly distinguishing between DHW and SH loads [10]. Disaggregating DHW and SH heat load components can be crucial in retrofitting interventions. By identifying the specific heat consumption patterns of DHW and SH, retrofitting can be adapted to address the unique needs of each building, leading to more effective insulation, heating system upgrades, and other energy-saving measures [11,12]. Disaggregation can also be relevant in optimizing energy efficiency and operational strategies in District Heating Networks (DHNs). The load components distinction can improve system performance by supporting better control and management of the heat supply, thus reducing energy waste and enhancing the network's overall efficiency [13–15]. Accurate heat load forecasts, made possible through disaggregation, can also facilitate the use of intermittent renewable energy sources, such as solar and wind power, by aligning the heat supply with their availability [16,17].

1.1. Heat loads disaggregation methods

Heat load disaggregation methods focus on estimating SH and DHW heat loads separately. Historically, SH systems have been the primary

* Corresponding author.
E-mail address: nicola.borgato.1@studenti.unipd.it (N. Borgato).

https://doi.org/10.1016/j.applthermaleng.2024.125296
Received 9 September 2024; Received in revised form 19 December 2024; Accepted 19 December 2024
Available online 25 December 2024

Nomenclature

Acronyms

CSD	Change Season Day
DHN	District Heating Network
DHW	Domestic Hot Water
DTW	Dynamic Time Warping
DTWD	Dynamic Time Warping Distance
ESC	Energy Signature Curve
KPI	Key Performance Indicator
MAE	Mean Absolute Error
PML	Percentage of Matching Labels
PMLHR	Percentage of Matching Labels with Hour Range
RMSE	Root Mean Square Error
SEC	Summer Energy Consumption Deviation
SH	Space Heating
WEC	Winter Energy Consumption Deviation
YEC	Yearly Energy Consumption Deviation

Symbols

CPT	Change Point Temperature (°C)
E	Global Heat Consumption (kWh)
$E_{DHW\ balanced,\ i}$	DHW heat load component after balancing operation at hour i (kWh)
$E_{DHW\ boundary\ max,\ i}$	Maximum DHW boundary at hour i (kWh)
$E_{DHW\ boundary\ min,\ i}$	Minimum DHW boundary at hour i (kWh)
$E_{SH\ balanced,\ i}$	SH heat load component after balancing operation at hour i (kWh)
E_{DHW}	Modelled DHW heat load component (kWh)
E_i	Global Heat Consumption at Hour i (kWh)
E_{loss}	Modelled heat losses in the water circuit (kWh)
E_{max}	Maximum value of Global Heat Consumption (kWh)
$E_{no\ CSDs}$	Global Heat Consumption without Change Season Days (kWh)
E_{ref}	Heat consumption threshold (kWh)
$E_{ref,\ n}$	Heat consumption threshold at iteration step n (kWh)
E_{SH}	Modelled SH heat load component (kWh)
T_{out}	Outdoor Air Temperature (°C)
$T_{out,\ i}$	Outdoor Air Temperature at Hour i (°C)
X_i	Measured input values
Y_i	Estimated Heat Consumption at Hour i (kWh)
Y_{summer}	Estimated Heat Consumption in the summer season (kWh)
Y_{winter}	Estimated Heat Consumption in the winter season (kWh)
α	Intercept of the linear regression model (kWh)
β	Slope coefficient of the linear regression model (kWh/°C)

subject of performance investigations, while DHW production systems have been overlooked due to their perceived minor impact on overall heating energy demand [10]. However, as building performance improves and passive house technologies are adopted, the significance of DHW heat load increases. [18,19]. Additionally, in warm climates, lower SH demands due to higher outdoor temperatures increase the relative importance of DHW heat loads.[20]. From these perspectives, identifying the different heat uses in buildings is extremely useful for establishing the best retrofit interventions.

The disaggregation approaches are based on building energy modelling methods: white-box or engineering approach, black-box or data-driven approach, and grey-box or hybrid approach [21–23]. White box models are physics-based models that rely on detailed knowledge of the system's internal processes [22]. White box models simulate SH and DHW heat loads separately, starting from separate inputs. One of the main limitations of the white box approach is the difficulty of finding accurate information [24] on the building geometry and schedules, especially in existing buildings. Additionally, these models demand expertise, proper assumptions, and significant time for both development and simulation [24–26]. In addition, domestic hot water models are usually based on standard consumption profiles and technical standards methods [27–30]. Standards' limitations include the risk of highly overestimating design flow rates compared to actual measured flow rates, even more than plausible tolerance for accounting for system operation in suboptimal conditions [10] and not accounting for influential factors like occupancy, leading to potential inaccuracies in DHW consumption predictions [31].

Black box models are data-driven approaches [32] that learn patterns from historical energy consumption data to predict future demand [33]. They are usually divided into two categories [32]: statistical models [34,35], and machine learning models [36–38]. Once trained, the model can predict future energy consumption using new input data [23]. They are advantageous when detailed system knowledge is lacking but depend heavily on the quality and quantity of available data [39].

Black box models have been widely used in the literature to disaggregate SH and DHW. Bacher et al. [40] proposed a non-parametric method using kernel smoothing to identify DHW heating peaks in total heat load time series data for single-family homes, but this requires high-resolution data (at least 15-minute intervals) and relies on instantaneous DHW production. It is ineffective when DHW relies on thermal storage or when DHW load is less than SH. Leiria et al. [41] utilized local weather data and total heat measurements through Kalman filtering and Support Vector Regression to estimate SH and DHW loads separately. In [42], Another approach used singular spectrum analysis in a Norwegian hotel to adjust SH load based on outdoor temperature. It demonstrates the versatility of disaggregation methods across different building types [43] and scales [44]. Among the weaknesses, they depend on the quality and quantity of input data [11,45], and it cannot simulate consumption changes due to new boundary conditions like retrofits or modifications to heat generation systems.

Grey box models combine physical laws with statistical methods to create semi-empirical models that can capture system behavior with less data than white box models but with more interpretability than black box models [33]. Various studies have applied these models for SH and DHW disaggregation. For instance, Schaffer et al. [45] developed a computationally fast algorithm utilizing data from smart heat meters. Hedegaard et al. [46] applied a grey-box approach (5R2C) for separating SH and DHW in district heating systems. It is advantageous as it doesn't rely on high-resolution measurements. Grey box models can adapt to new data and boundary conditions without the need for complete reconfiguration [47,48].

The choice of model for SH and DHW disaggregation depends on factors like data availability, accuracy needs, and application. White box models are best for detailed analysis in existing buildings with accurate information and design phases, while black box models suit scenarios with stable boundary conditions and abundant data. Grey box models balance accuracy with data needs. A common limitation for all approaches is the availability and quality of data on building properties, user behavior, and energy consumption.

1.2. Energy signature curve (ESC) models

In this paper, the Energy Signature Curve (ESC) is used to analyze energy consumption in buildings, focusing on DHW and SH heat load disaggregation. ESC models provide valuable insights into the impact of temperature on building energy consumption and enable utility managers and planners to estimate energy savings accurately [49]. They serve a dual purpose: they allow for analyzing a building's energy

consumption [50], and forecasting the building's energy demand [51–53]. They belong to data-driven models because they are based on linear regressions but also have some characteristics of hybrid models, such as the direct dependence of the model's parameters on the heat transfer phenomena of the building [54]. The ESC represents energy consumption relative to outdoor temperatures, with two segments with different slopes related to heating needs during different seasons. [55]. Day et al. [56] indicate that slope coefficients from the ESC model can accurately estimate a building's U-value. The Change Point Temperature (CPT) is the outdoor temperature where SH is no longer required, marking the transition from winter, when both SH and DHW are needed, to summer, when typically only DHW is required. This distinction may not align perfectly with the actual changes in seasons. CPTs generally range from 10 to 15 °C for the European building stock and are often based on prior knowledge of location or building behavior [51,57]. Pedersen's work [58] contributed to understanding ESC construction methodologies by developing a mathematical procedure for identifying CPT.

Heat load in summer is often treated as a constant [55]. However, several studies show a slight correlation between summer heat consumption and outdoor air temperature, [45,59,60]. Some authors proposed a piecewise linear model with more than two segments to better capture building heat consumption during middle-season days and to have more precise information for retrofit interventions [61,62].

In addition, researchers have found a strong correlation between energy consumption, outdoor air temperature, and the hour and type of day (weekday or weekend). This holds for both winter and summer seasons. Ding et al. [51] illustrated the importance of considering the hour of the day as a parameter for creating an ESC with reasonable accuracy. Sharifi et al. [14] showed that neglecting hour and day types resulted in a low ESC accuracy ($R^2 = 0.59$) when using a single model for the whole year. Generally, the ESC is built by considering a continuous energy consumption for the points before and after T_{out} = CPT [42,63–65]. However, Anjomshoaa et al. [66] showed that, in some cases, it is possible to have a gap between summer and winter energy consumption.

Concerning SH and DHW heat load disaggregation, Ivanko et al. [42] proposed a method where DHW heat consumption was obtained by subtracting the SH modelled using the ESC method from the measured energy values, creating high inaccuracies in this load estimation. Milić et al. [67] used an ESC model to analyze thermal characteristics of 70 multi-family buildings, aiming to identify those in need of retrofitting, but did not provide specific models for DHW, relying instead on July consumption as an indicator. Ivanko et al. [10] developed a clustering algorithm for nursing homes to analyze DHW consumption patterns, emphasizing the need for models based on actual data rather than standard profiles for accurate representation. Some studies use a consumption threshold to differentiate between winter and summer seasons, which can enhance accuracy, in models involving multiple weather parameters [34].

ESC models are simple to build and require minimal data, mainly historical energy consumption and outdoor temperatures [54]. ESC models have a wide range of usage. They can predict future energy consumption and effectively disaggregate DHW and SH heat load. However, their accuracy is highly dependent on the quality and quantity of the input data. High-frequency time series data (hourly resolution is preferable) are necessary to create accurate curves [68]. Moreover, ESC models are static and do not account for dynamic changes in building occupancy, usage, or energy efficiency improvements over time, even if these interventions impact ESC parameters [69]. ESC models assume a strong correlation between energy consumption and outdoor temperature, but they are usually built using data on energy consumption instead of the energy demand of the building. So, high inaccuracies may occur in buildings with significant heat gains (internal, solar radiation) [53,54].

1.3. Objectives and novelty

The paper aims to determine the best ESC-based model for analyzing a building's heat consumption and for disaggregating SH and DHW heat loads. The ESC models combine an acceptable accuracy with low computational effort in analyzing the overall building energy consumption and estimating building parameters [70]. Existing works on this topic show three main problems with these models.

First, using a single ESC to describe the energy consumption of an entire year cannot represent daily trends in the energy consumption [14]. The accuracy of the ESC model is higher when multiple ESCs are used based on the day and hour types [51].

The second issue is related to the non-accurate DHW estimation. The common practice is to assume the DHW heat consumption of the building is constant with the outdoor temperature and equal to the average heat load in the summer season [54], or as the subtraction between measured heat consumption and modeled SH heat load [42].

Third, the method for the identification of the summer and winter seasons: the most used method is the CPT approach, but mathematical algorithms to derive the CPT value are few, and its determination is generally based on previous knowledge and experience [71,72].

This paper compares four different linear regression models to find the best SH and DHW disaggregation methods to tackle these research issues. The work proposes solutions for each of the three main issues. Each solution is designed to maintain the model's simplicity and minimize the required inputs. The goal is to keep the final model free from assumptions regarding building geometry, properties, usage, internal gains, and mechanical systems or appliances. The representation of hourly trends of heat consumption was achieved by proposing the usage of 24 independently calibrated ESCs, one per each hour of the day. To address the lack of a dedicated DHW model, a reference DHW heat consumption profile was obtained thanks to the detailed summer heat consumption analysis. Then, an algorithm for automatically identifying season thresholds (CPT or heat consumption threshold, depending on the model) is proposed. The goal of this study is to simplify the process of identifying seasons by proposing a tool that can be applied across various case studies, regardless of prior knowledge. Models that utilize the new method for season identification, which is based on heat consumption thresholds, demonstrate greater accuracy and better reflect the heat consumption of buildings.

This work aims to show that incorporating these three new features can achieve higher accuracy compared to the current state-of-the-art ESC methods while maintaining nearly the same computational time and avoiding additional assumptions regarding input variables.

The manuscript is structured as follows: Section 2 provides an overview of the methodology. Section 3 describes the case study. Sections 5 and 6 summarize KPIs and results, followed by conclusions.

2. Methodology

In this work, four regression models were applied to a case study of 27 buildings connected to the district heating network of Tartu, Estonia. The first model (CPT-1) uses the CPT to identify the threshold between the winter and summer seasons and builds a single ESC using the global heat consumption measurements of an entire year. The second model uses the CPT, but 24 independent ESC models are created, one per each hour of the day (CPT-24). The third model uses a heat consumption threshold (E_{ref}) to separate the winter and summer seasons and builds a single model using the global heat consumption measurements of an entire year (Eref-1). The fourth model uses a heat consumption threshold and creates 24 independent models, one per hour of the day (Eref–24). Table 1 presents a summary of the main characteristics of each model. A detailed analysis of the DHW heat load was applied to every model and used to estimate the DHW heat load.

This section presents the methodology applied to process the dataset containing the hourly total heating demand of the analyzed buildings, as

Table 1
Summary table of the main characteristics of implemented models.

Model	Seasons threshold	Type of linear regression model
CPT-1	CPT	Whole-year linear model
CPT-24	CPT	24 independent linear models
Eref-1	E_{ref}	Whole-year linear model
Eref-24	E_{ref}	24 independent linear models

follows:

- Dataset segmentation
- ESC calculation:
  - Using CPT split
  - Using E_{ref} split
- SH and DHW disaggregation:
  - For CPT models
  - For E_{ref} models
- DHW profiles clusterization

2.1. Hourly data segmentation

Ensuring data quality and pertinence for modeling is crucial [48]. Therefore, the initial phase involves data pre-processing, including data cleaning and selecting the inputs for models' calibration, which are the outdoor air temperature, T_{out}, and global heat consumption, E, i.e., the building total SH and DHW consumption). Hourly resolution measurements for an entire year are used for T_{out} and E. Data are grouped into 24 vectors per building, each representing one hour of the day. This subdivision aims to calibrate 24 models individually, one for each hour of the day. In this way, unusual energy consumption patterns or exceptional loads during specific hours of the day can be easily detected [34,73].

2.2. Energy signature curve

The ESC is a linear data-driven model that represents the linear dependency between the heating demand of a building and the outdoor air temperature. Usually, ESC is split into two segments by a critical outdoor temperature, the CPT, see Fig. 1. The points below the CPT represent the winter season, with both DHW and SH heat consumptions. In contrast, the points above the CPT correspond to the summer season, with only the DHW heat load. Typically, winter energy consumption is characterized by a strong dependence on outdoor temperature, while this dependence is weak during the summer season. The ESC can be represented using a piecewise function, as shown in Eq. (1) [65].

$$\begin{cases} Y_i = \alpha_1 + \beta_1 \times T_{out,i} & \text{if } T_{out,i} < CPT \\ Y_i = \alpha_2 + \beta_2 \times T_{out,i} & \text{if } T_{out,i} \geq CPT \end{cases} \quad (1)$$

In the previous equation, α_1, α_2 (kWh) are the intercepts with the vertical axis, β_1, β_2 (kWh/°C) are the slope coefficients (α and β are calculated using the least-square linear regression method), T_{out} is the outdoor air temperature, and Y is the predicted energy demand obtained from the linear regression model. Therefore, this work notation refers to the measured consumption when E is used, while Y indicates the model estimation. As emerged in paragraph 1.2, despite a significant part of ESC models using a piecewise continuous linear function at CPT, the discontinuity is acceptable, depending on the use of the building. Switching on or off the heating system in the middle seasons could significantly change the overall energy consumption of the building, which a continuous ESC cannot represent. In the case of the 24 hourly-based models, for each model, the subdivision into the two seasons and the equations' calibration are independently applied to the 24 datasets previously described. As mentioned, two approaches were used to divide the summer and winter seasons. The first is based on a temperature threshold (CPT), while the second uses a heat consumption threshold (E_{ref}) [34].

2.2.1. CPT calculation algorithm

A mathematical methodology was applied to automatically determine CPT [74]. A first regression uses the two heat consumption points corresponding to the highest $T_{out,}$ and the slope coefficient β is obtained. The calculation is repeated by considering step-by-step heat consumption values corresponding to progressively decreasing T_{out}, obtaining the slope coefficients. The number of heat consumption points used for the linear regression increases step by step, with a corresponding decrease in the T_{out} values considered. The calculation ends when all heat consumption values are considered for the regression. The CPT is identified as the T_{out} value in correspondence with a significant β decrease. In fact, when winter season points are included in the iteration, β suddenly starts decreasing due to the increased temperature dependency. So, CPT is the T_{out} at which β begins deviating from 0 more than a certain tolerance. A scheme that represents the algorithm for CPT calculation is reported in Fig. 2.

2.2.2. Eref calculation algorithm

This method uses a heat consumption threshold (E_{ref}) to identify the summer and winter seasons as an alternative to the CPT. The heat consumption values above the threshold belong to the winter season (SH plus DHW season), while the values below are considered summer season points (only DHW season). A season splitting based on E_{ref} attributes the measured E values to the specific season depending on the heat consumption value instead of the T_{out} value. In fact, due to higher consumption, points with $E > E_{ref}$ are supposed to refer to an aggregated load of SH and DHW heat load and are assigned to the winter season. Vice versa, points below the CPT with a very low E are considered only DHW, not total consumption. Therefore, this method is introduced because it allows a more realistic estimation of mid-season points.

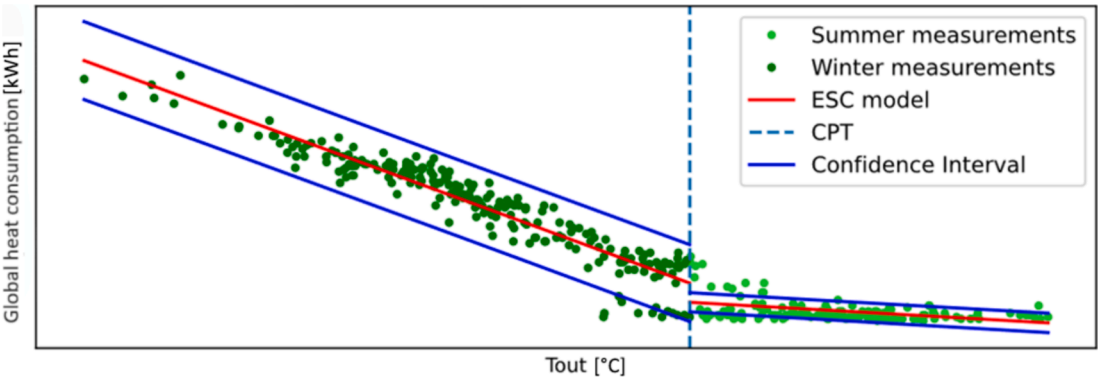

Fig. 1. Example of energy signature curve.

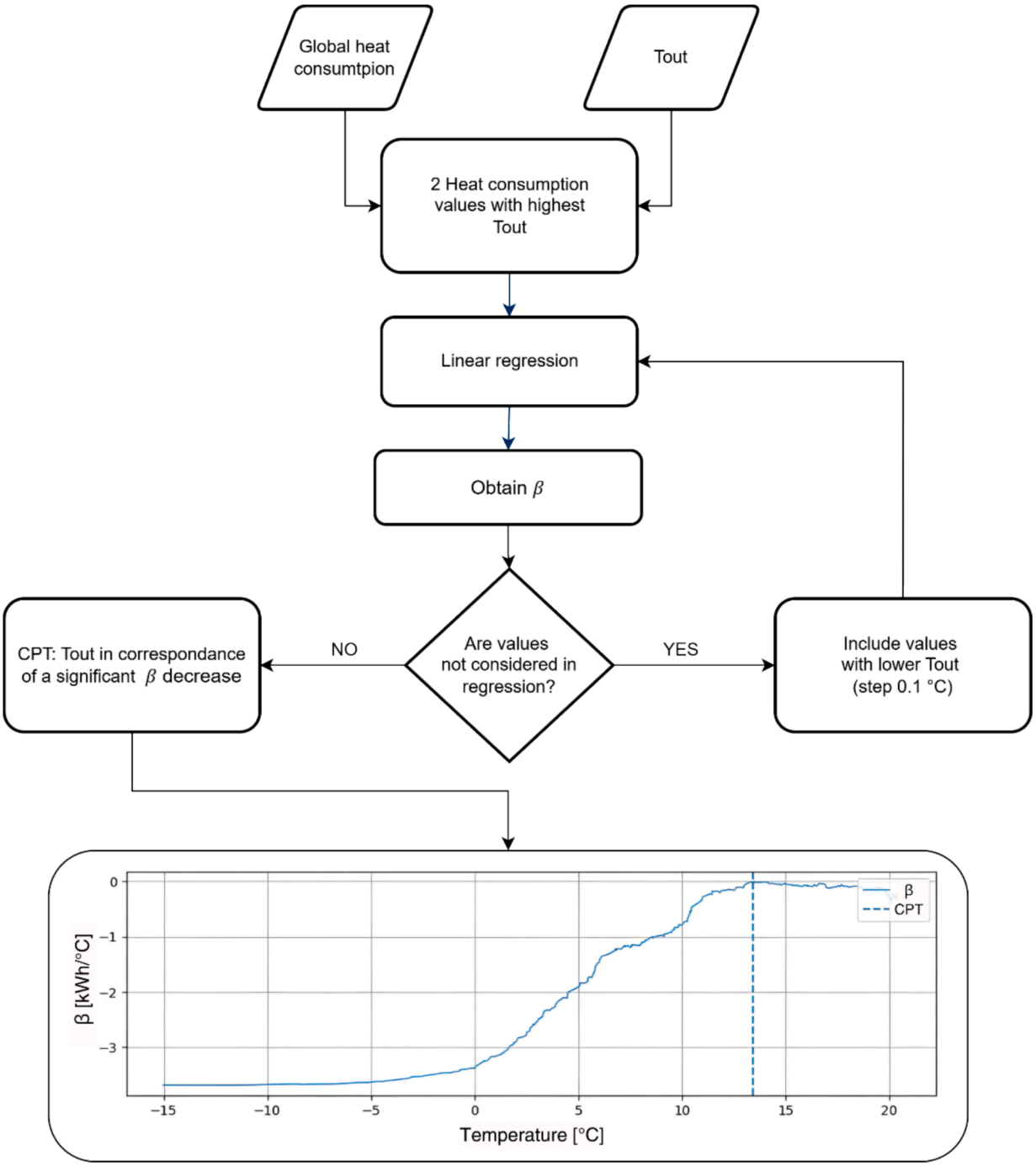

Fig. 2. CPT calculation algorithm flowchart.

An algorithm was implemented to identify the E_{ref} automatically.

1. The first step is determining the change season days (CSDs), characterized by an intermediate behavior between summer and winter. Some days only have DHW demand even if T_{out} is low, and others have DHW and low SH demand. CSDs are excluded from the values used in the E_{ref} calculation to reduce the possible inaccuracies caused by intermediate heat consumption values. CSDs are identified through a peak detection algorithm [75]. Assuming that heat consumption during summer is almost constant and has lower fluctuation than in winter, the detected peaks in heat consumption are efficiently used to identify the change-season period. The implemented peak detection algorithm is calibrated to detect only relevant peaks, i.e., sudden evident increases in heat consumption. Days between the end of the summer season and the first detected peak are considered the change between summer and winter seasons (the first part of CSDs). Similarly, the CSDs corresponding to the change between winter and summer seasons were identified at the beginning of the summer season. A new dataset called $E_{no\ CSDs}$ is created with the exclusion of the points included in the CSDs.
2. The second step is to calculate the first approximation of E_{ref}. Previous studies have highlighted that the E_{ref} variation range is between 0 and $E_{max}/2$, where E_{max} is the maximum measured global heat consumption [34]. The procedure starts by splitting the $E_{no\ CSDs}$ dataset into two parts using a variable E_{ref} threshold. This value changes step by step ($E_{ref,\ n}$), beginning from $E_{ref} = E_{min}$ and being

increased by a tolerance per every step (in this study, fixed at 0,1 kWh). Linear regression is conducted in each step for both winter (values with $E_i > E_{ref,\ n}$) and summer (values with $E_i < E_{ref,\ n}$) seasons. The mean absolute error (MAE) and the root mean squared error (RMSE) of the regressions are calculated in each step. The heat load that minimizes the errors in both winter and summer season regressions determines the final chosen E_{ref} value.

Despite the calculation result, the algorithm guarantees a minimum number of points above and below E_{ref} to maintain a reasonable number of hours for the heating and summer season. The minimum number of points above E_{ref} corresponds to 50 % of the $E_{no\ CSDs}$; instead, the minimum number of points below E_{ref} corresponds to the points with $T_{out} >$ CPT.

2.2.3. Implementation of confidence intervals

Despite the season detection method, final ESCs must be cleaned by outliers, which deviates from the linear regression. Outliers in heat consumption predominantly occur during transitional seasons, when outdoor air temperature values are around the change point temperature (CPT). These anomalies are often due to the space heating plant operating under atypical conditions. An example is when the outdoor temperature is higher than the CPT, but the plant is still operational, or vice versa. Removing these outliers is crucial to avoid incorporating these extraordinary heating plant behaviors into the dataset used for linear regression. Their effect is particularly relevant in the summer season, where the trend is almost horizontal. Even a few outliers can slope the trend line significantly, leading to unrealistically high-temperature dependence. By excluding them from the regression input dataset, a model that more accurately represents the typical behavior of the building is achieved.

From this perspective, the method from [65] involving Confidence Intervals (CI) is applied. CI calculation defines a confidence band outside which points are considered outliers. The estimated range is based on a multiple of the statistic's standard error (SE), determined based on a theoretical distribution. The t-Student distribution was used, applying a CI of 95 %.

Confidence intervals are then applied iteratively to both winter and summer seasons, thus identifying and removing outliers and recalibrating the ESC. The linear regression is performed separately for the winter and summer seasons. The iterative process continues until all experimental values are included in the confidence interval zone or the number of iterations reaches a limit of 4 iterations. It is worth mentioning that, for the analyzed cases the number of deleted outliers was never above 2 % of the yearly measurements.

2.3. SH and DHW heat consumption disaggregation

Two different algorithms were used to disaggregate the global heat consumption into SH and DHW heat loads, depending on whether the CPT or E_{ref} split is applied.

2.3.1. Load disaggregation method for CPT models

This methodology is similar to the method proposed by [42]. It is based on identifying a minimum load during the summer season, considered the DHW base load. The SH fraction of the ESC is then derived by shifting downward the original ESC model by the base load value (Eq. (2).

$$E_{SH} = Y\left(T_{out,i}\right) - \min\left(Y\left(T_{out,i}\right)\right) \tag{2}$$

where $T_{out,\ i}$ represents the outdoor air temperature at a particular hour of the year, Y the heat consumption points obtained through the linear regression model, and E_{SH} the values for space heating consumption. The difference between global heat demand (E_i) and the calculated SH demand (E_{SH}) is assumed to be the DHW demand [40].

Energy losses in the DHW circuit (E_{loss}) must also be considered part of the DHW heat load because, in a DHW system with continuous circulation, the DHW system operates continuously to deliver hot water. First, an average DHW heat consumption profile for each month in the summer season is calculated. The SH period depends on the specific location, and national standards often regulate it. However, it is acceptable for most buildings in mild and cold mid-latitude climates to consider July and August as part of the summer season. Therefore, August and July's average daily energy consumption profile is calculated. Assuming negligible DHW use during nighttime [76], the mean thermal energy demand between 0:00 am and 4:00 am is computed to account for the energy losses in the water circuit. The SH and DHW thermal load separation was conducted using Eq. (3).

$$E_{DHW} = \begin{cases} E_i - E_{SH} + E_{loss}, & if\ E_i > E_{SH} \\ E_{loss}, & if\ E_i \le E_{SH} \end{cases} \tag{3}$$

Ultimately, E_{SH} is recalculated as a difference between E and E_{DHW}. If any of the DHW or the SH energy load is lower than zero, it is set to zero, and the other component load is recalculated to match the measured heat consumption [42]. SH and DHW heat consumption values are calculated for each hour of the year.

2.3.2. Load disaggregation method for Eref models

The DHW heat load calculated with CPT models presents significant fluctuations. This is because, while the SH load is calculated through the ESC model, the DHW load is derived by subtracting the calculated space heating from the measured data. This leads to the inevitable association of the load fluctuations to the DHW load, while the SH profile results in smoother and linear results with the outdoor temperature variation.

To overcome this issue, in the E_{ref}-1 and E_{ref}-24 models, two different regression models are used to calculate the SH and DHW heat loads along the year, expressed in Fig. 3. After E_{ref} is identified, the points with an overall thermal load lower than E_{ref} are used to build the DHW heat load regression line (purple), and the points with an overall thermal load higher than E_{ref} are used to construct the total heat load regression line (SH plus DHW). Generally, the DHW heat load presents a weaker dependence on T_{out}; therefore, its regression line is characterized by a lower slope coefficient. Concerning the heat load of the winter season, the SH load regression line, represented with a green line, is obtained by subtracting the DHW heat load regression line from the total heat load regression line. Negative values are set equal to 0. In addition to the SH and DHW lines, upper and lower boundaries for the DHW heat consumption are set (purple dashed lines). These limits are parallel to the DHW model, interpolating the minimum and maximum summer season load, respectively.

Only days with a daily average $T_{out} > 18\ °C$ were considered to select the maximum summer heat load. This operation excluded from the possible summer maximum heat load any possible SH heat load demand caused by low T_{out} even during the summer season.

Once the SH heat load model and the DHW heat load model are calibrated, each hour's measured energy consumption (E_i) is compared with the sum of the models' outputs ($E_{SH,\ i} + E_{DHW,\ i}$) for that T_{out}. The difference between E_i and $E_{SH,\ i} + E_{DHW,\ i}$ corresponds to the deviation between the model and the measured data. This deviation is attributed to the DHW heat load if it is lower than the minimum or maximum DHW boundary lines. Conversely, if this deviation exceeds the boundary lines regarding heat load, it is attributed to both the DHW heat load (up to the maximum and minimum limit) and the SH heat load. For the Eref-24 model, this heat load disaggregation method is applied independently to each ESC.

2.4. DHW heat load clusterization

After disaggregating the DHW and SH heat consumption data, a dedicated methodology to cluster summer DHW heat demand profiles is

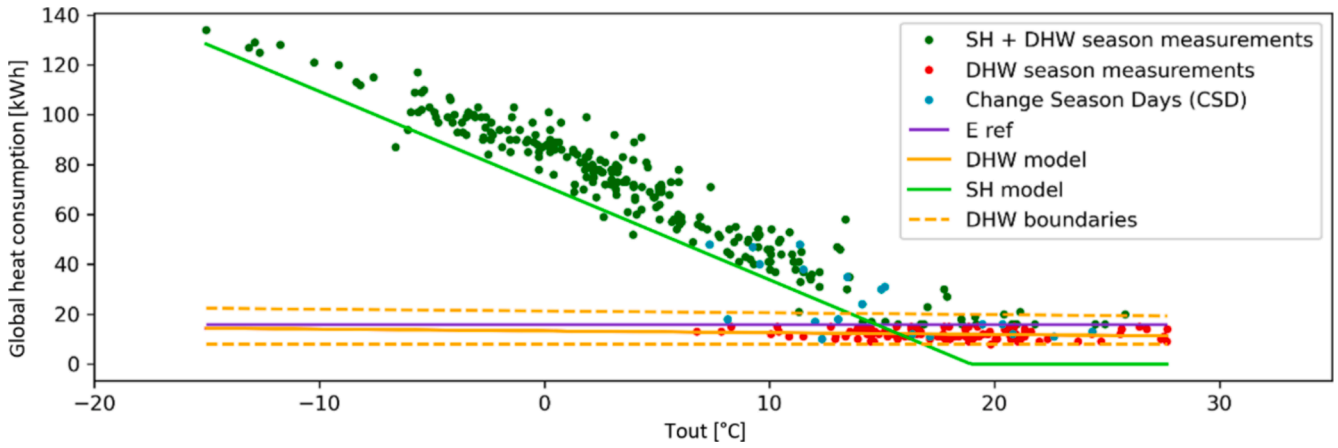

Fig. 3. Example of modelled SH and DHW in Eref model.

implemented [10]. DHW heat load clusterization can serve two purposes: improving the DHW profile in the winter season obtained using the disaggregation algorithm and providing a benchmark to test the performance of disaggregation models. The first purpose is discussed in the current subsection, while the second is described in subsection 2.5.

The implemented approach comprehensively examines DHW heat consumption profiles during summer to identify daily similarities and delineate standardized consumption patterns. The purpose is to derive a reference pattern to improve the DHW heat consumption profile quality obtained from the disaggregation algorithms. The process employs statistical methods for identifying similar consumption patterns on different days of the week, followed by statistical grouping to recognize minimum, medium, and peak consumption hours within daily profiles. The following paragraphs briefly overview the procedure, described in detail in [10].

The method starts by selecting the summer season through the start and end days identified in Section 2.2.2. The daily DHW heat load consumption profiles are clustered into two statistically similar groups. First, DHW heat consumption data is organized into daily profiles. Then, the Student's *t*-test is used to assess if the mean values of DHW heat consumption between two days are equal. Then, Fisher's exact test evaluates the similarity of variances. Days with a high matching rate are considered to have similar DHW heat consumption profiles and are clustered together. The algorithm was implemented to recognize a maximum of two groups. In case differences among days are low, a single group is created.

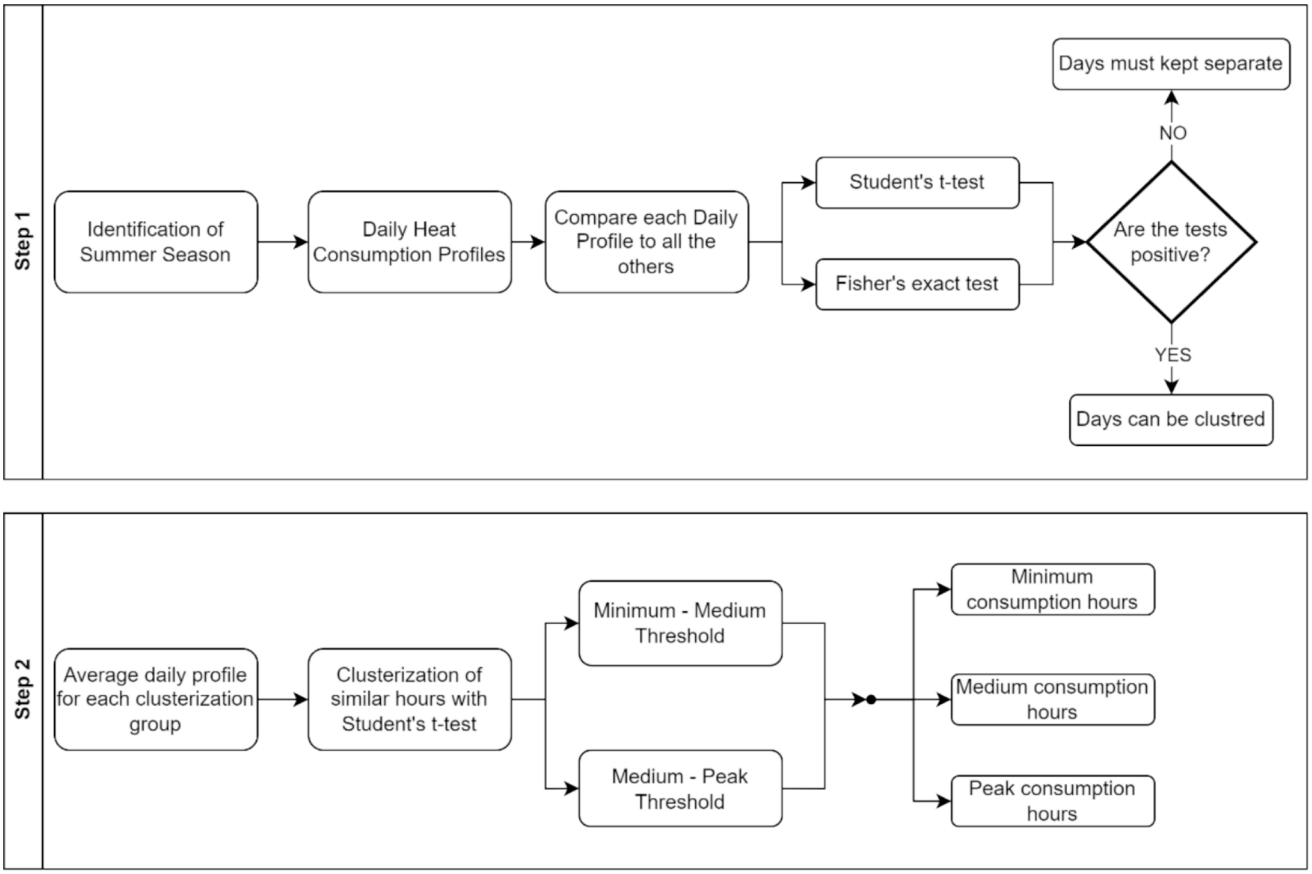

Fig. 4. Example of modelled SH and DHW in Eref models.

The second step of the clusterization procedure consists of identifying hours with minimum, medium, and peak consumption. Similar hourly heat loads are clustered within each group's average consumption profiles using the Student *t*-test. As a result, the method identifies critical thresholds that separate the DHW heat use profile into peak, average, and minimum heat consumption zones. The scheme of this procedure is reported in Fig. 4.

2.5. Key performance Indicators (KPIs)

In this work, the accuracy of the models is evaluated depending on the modeling phase. First, KPIs for assessing the ESC's performances in interpolating the input data are applied to each model. For KPI calculations, the input data is the initial dataset of hourly heat consumption, without removing outliers. Second, qualitative KPIs for evaluating the performances of the disaggregation algorithms are used, with a specific focus on DHW heat consumption profiles. The latter is needed as no specific dataset with a split of DHW and SH data is available. Table 2 presents a summary of the KPIs used in this work. It includes their names and the application category (linear regression performance or disaggregation performance).

Considering ESCs, the accuracy in predicting energy demand is evaluated using R^2, the Yearly Energy consumption Deviation, YEC, and the summer and winter season energy consumption deviation, WEC and SEC, respectively. R^2, Eq. (4), tests the linear regression model's ability to approximate input data. This KPI can be used only in winter when the linear model slope differs widely from zero, as R^2 tends to 0 with horizontal lines, even if the interpolation line fit is good. YEC, WEC, and SEC, eq. (5), 6, and 7, are introduced to study the model's accuracy in terms of energy [34]. Winter and summer seasons are defined in the previous Sections, depending on the model type.

$$R^2 = 1 - \frac{\sum (X_i - Y_i)^2}{\sum (X_i - X_{mean})^2} \tag{4}$$

$$YEC = 100 \times \frac{|\sum X_i - \sum Y_i|}{\sum X_i} \tag{5}$$

$$WEC = 100 \times \frac{|\sum X_{i,winter} - \sum Y_{i,winter}|}{\sum X_{i,winter}} \tag{6}$$

$$SEC = 100 \times \frac{|\sum X_{i,summer} - \sum Y_{i,summer}|}{\sum X_{i,summer}} \tag{7}$$

Y_i is the value obtained from the prediction model, X_i is the measured data (E_i), and X_{mean} is the mean value of the measured data. The subscripts "winter" and "summer" indicate that only values from the corresponding seasons are considered for that KPI calculation.

Concerning the disaggregation model performance, more qualitative parameters are used to evaluate the resulting SH and DHW loads. In this case, since no separate DHW and SH benchmarks are available, an alternative benchmark is set to assess the reliability of the calculated results. In particular, the benchmark is defined as the global heat consumption in summer, which reasonably consists only of DHW, as described in Section 2.3.

Table 2
Summary table of the KPIs and their characteristics.

KPI	Application category
R^2	Linear Regression
YEC	Linear Regression
WEC	Linear Regression
SEC	Linear Regression
PML	Disaggregation Algorithm
PMLHR	Disaggregation Algorithm
DTWD	Disaggregation Algorithm

The first indicator introduced in this paper is the Percentage of Matching Labels (PML). The DHW profiles derived from the disaggregation algorithms are scrutinized for each model using the procedure outlined in Section 2.4. This involves assigning a "minimum," "medium," or "peak" load label to each hour. The PML indicator is a crucial tool in this process, representing the percentage of the label correspondence between the benchmark DHW heat consumption profile hours and the DHW heat consumption profile hours during the summer months. In simpler terms, it compares DHW peak/medium/minimum load hours to the corresponding benchmark, resulting in the percentage of correctly labeled hours. The expression of PML is reported in Eq. (8), where $n_{correctly\ labeled}$ represents the total number of hours with the correct label identified in the analyzed period and n_{hours} is the total number of analyzed hours.

In addition to the PML, a variation was proposed to consider potential time shifts in the demand, known as the PML Hourly Range, or PMLHR. Like the PML, this indicator no longer evaluates matching between corresponding hours but within a broader time range. For instance, if the analyzed profile is labeled as "medium" at 11:00 am, a "medium" hour between 10:00–12:00 am would result in a positive match. The introduction of the PMLHR serves a specific purpose, to provide an indicator less influenced by the local fluctuation of the DHW profile obtained from disaggregation, thereby offering a more comprehensive evaluation. The expression of PMLHR is reported in Eq. (9), where $n'_{correctly\ labeled}$ represents the total number of hours with the correct label identified in the analyzed period. It was considered a correct label for correlating the analyzed hour with either the label of the previous hour or the following hour. n_{hours} is the total number of analyzed hours.

The third qualitative indicator is Dynamic Time Warping Distance (DTWD). Dynamic Time Warping (DTW) is an algorithm used to measure the similarity between two temporal sequences [77,78]. The corresponding DTWD parameter is a crucial algorithm component, providing the similarity between two temporal sequences. It finds an optimal match between the corresponding points of two sequences, and the output of the DTW algorithm is the distance between the two analyzed time series, the DTWD. In this research, the DTWD calculation was applied as exposed in [79], benchmarking the four models' disaggregated DHW heat profiles. The expression of DTWD is reported in Eq. (10), where $d_\varphi(X, Y)$ is the average accumulated distortion between the warped time series X (the analyzed DHW profile) and Y (the reference DHW profile). A more detailed explanation of the dynamic time-warping algorithm is presented in Appendix A.

$$PML = 100 \times \frac{n_{correctlylabeled}}{n_{hours}} \tag{8}$$

$$PMLHR = 100 \times \frac{n'_{correctlylabeled}}{n_{hours}} \tag{9}$$

$$DTWD(X, Y) = \min_{\varphi} d_\varphi(X, Y) \tag{10}$$

3. Case study

The case study focuses on the DHN in Tartu, Estonia. The data was sourced from two primary inputs: the heat consumption data from 42 smart energy meters located in various substations of the DHN and climatic data from a weather station managed by the University of Tartu [34]. Each substation corresponds to a single building. All these substations are in the Tarkon sub-network, and data about hourly heat consumption are collected remotely by the DHN operator [80]. The network covers 5.34 km and delivers up to 4.3 MW of heat, with a total annual energy consumption of around 8.2 GWh for both SH and DHW heat production. A list of intended uses of buildings is provided in [34]; however, due to privacy concerns, no association is available between the end-user and the substation ID (Table 3). For the same reason, no

Table 3
Summary of the intended use of buildings analyzed from [34].

Intended use of buildings	Number of DH substations
Residential Apartments	25
Private House	7
Commercial Buildings	1
Educational Buildings	8
Offices	1

information about building geometry, envelope properties, occupancy, year of construction or appliances was available.

The heat consumption data, recorded hourly, included total energy consumption and operational variables such as supply and return temperatures in the heat distribution loop of the primary and secondary circuits of the network (Fig. 5). The position of the smart meters was defined by the operator of the District Heating Network for measuring global heat consumption of every building and water temperature set-points. The energy meter accuracy respects the European directives, keeping the error below 5 %. The climatic data comprised outdoor temperature, global solar irradiance, wind speed, and direction, sampled at 15-minute intervals and resampled to match the hourly frequency of the energy data.

A further selection of buildings is conducted in this work. Only buildings with summer heat consumption are used for the analysis. Buildings with null demand for the entire summer period (when T_{out} is higher) are excluded because no DHW exists in these buildings. So, the dataset used to apply disaggregation algorithms includes 27 buildings.

4. Results

The results are divided into two parts: an evaluation of the performance of the four linear regression approaches and a presentation of the performance of the disaggregation algorithms.

4.1. Linear regression approaches

Concerning the performance of linear regression approaches, Fig. 6 displays the four resulting models for a sample building (Bd 08). All plots represent the heat consumption on the y-axis and the outdoor air temperature on the x-axis. SH and DHW season measurements are highlighted with green points, while only DHW season measurements are represented with red ones. Fig. 6 (a) and (b) represent the CPT-1 model and the 5:00 pm hour of the CPT-24 model, respectively. In both Subfigures, the continuous light-green line represents the SH plus DHW model, while the continuous orange line represents the DHW model in the DHW season. The continuous blue lines represent the confidence intervals. The light-blue dotted line is drawn in correspondence with CPT.

Comparing the CPT-24 and CPT-1 approaches reveals that the CPT-24 approach can better represent the peaks of the building's heat consumption profile. The main reason is that the CPT-24 approach includes the energy consumption peaks in the interpolation, whereas the CPT-1 approach excludes them from the confidence intervals. Indeed, by creating a single model for every hour, confidence intervals are created according to the heat consumption of the analyzed hour. So, peaks that occur in specific hours are inside the CI, and the model represents them. On the contrary, they are treated as outliers in the CPT-1 approach.

Fig. 6 (b) shows that continuity between the SH + DHW model (light-green line) and DHW model (orange line) in correspondence with CPT is not guaranteed. Most of the existing state-of-art models based on ESC assume continuity in correspondence with CPT. However, this cannot be imposed by hypothesis, and this assumption could lead to significant inaccuracies in heat consumption estimation. As in the proposed example, there could be a step in heat consumption when the SH power plant is turned on.

Moving to Eref models, changing the method to distinguish the summer and winter seasons affects the model accuracy, especially during the change season days period, shown in red in Fig. 6 (c), are considered. Middle-season days could have different behaviors depending on the heat supply system settings. It is possible to have points with only DHW heat consumption even if T_{out} is lower than CPT or to have SH and DHW heat consumption if T_{out} is higher than CPT (Fig. 6 (c) in blue). This phenomenon is relevant, especially in cold climates, with high T_{out} variability during the days of the middle season. The CPT approach cannot correctly identify these consumption values due to the subdivision of DHW (summer) and DHW plus SH (winter) seasons based on temperature threshold. In the linear SH plus DHW model in Fig. 6 (a), these values are considered outliers to be removed.

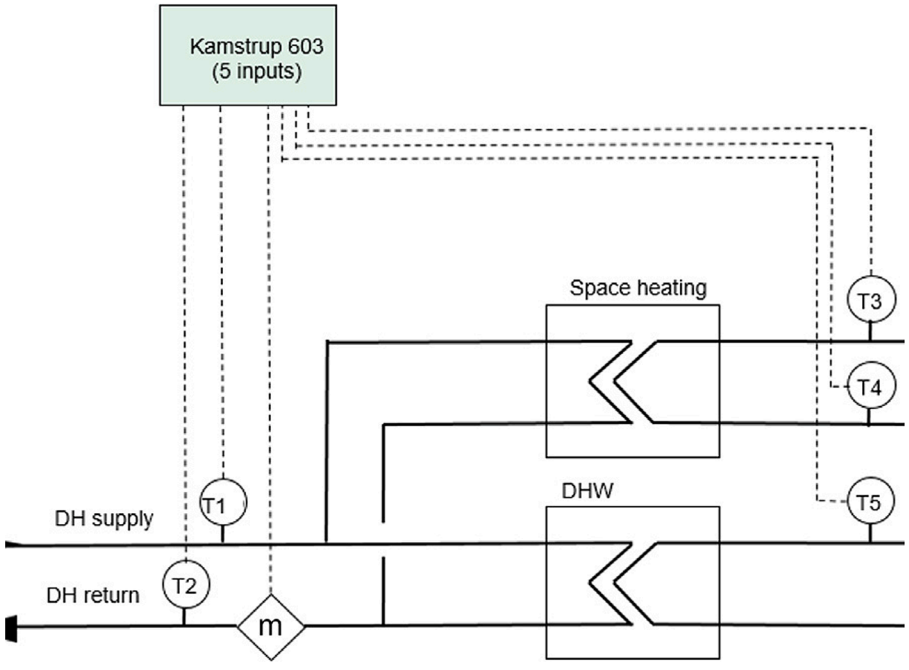

Fig. 5. Location and layout of the smart energy meters in the DH in Tartu [10].

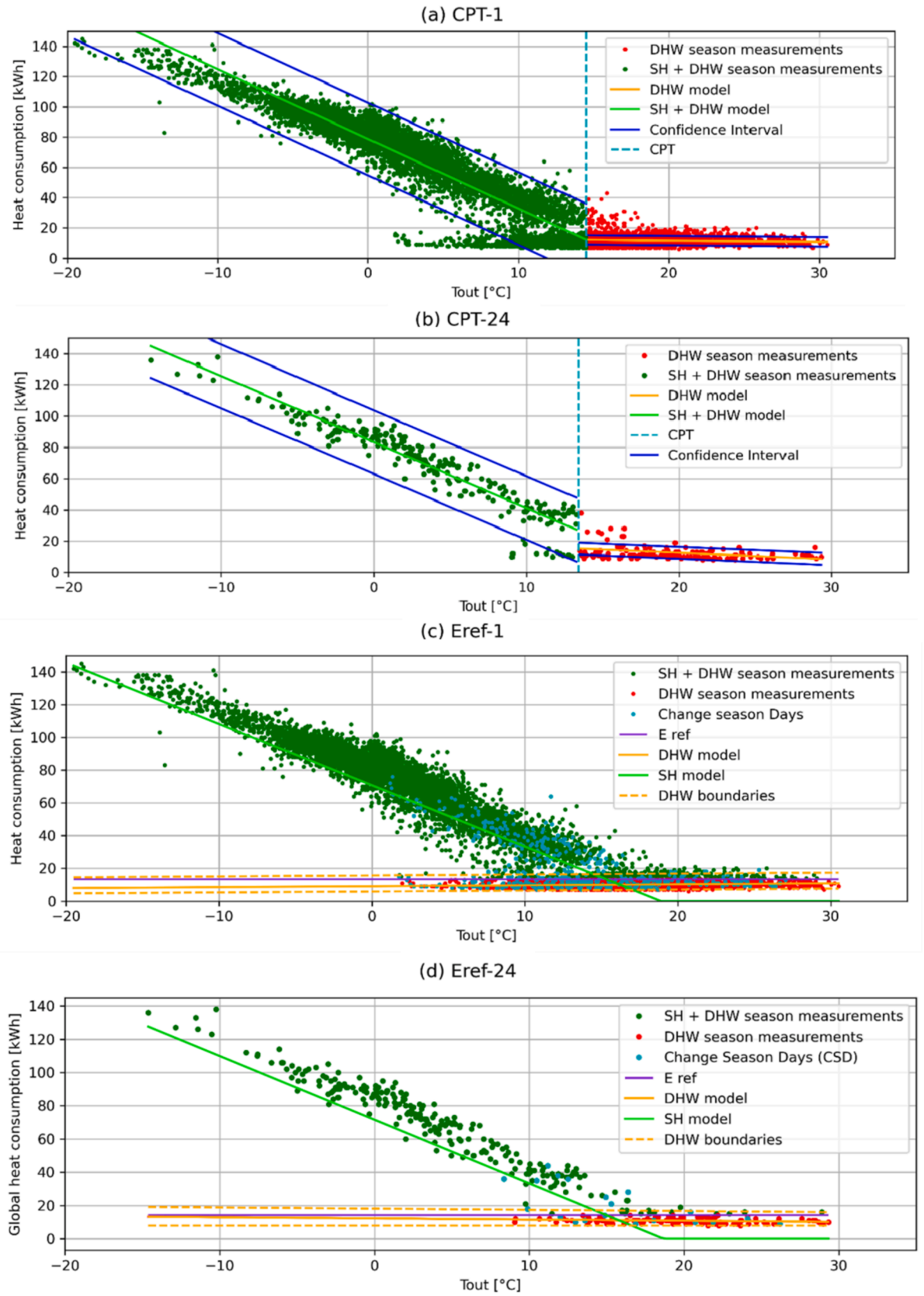

Fig. 6. Linear regression model results of four approaches for a single building (Bd 08); (a) CPT-1 approach; (b) an hour of the CPT-24 approach (5:00 pm – 6:00 pm); (c) Eref-1 approach; (d) an hour of the Eref-24 approach (5:00 pm – 6:00 pm).

Fig. 6 (c) and (d) represent the Eref-1 model and the 5:00 pm hour of the Eref-24 model, respectively. In these plots, the light-green continuous line represents the SH model instead of the total DHW + SH model for clarity of representation. Change-season days (CSD) are highlighted in light blue. The horizontal purple line represents the E_{ref} threshold, and the dotted orange lines are the boundaries of minimum and maximum possible values for the DHW heat load component.

Eref-based approaches can better interpolate middle-season days. It is evident in Fig. 6 (c) and (d) that the combination of the SH model (light-green line) and DHW model (orange line) can accurately describe low heat consumption values in correspondence of low T_{out} and high heat consumption correspondence of high T_{out}. The model better interpolates these points, but they are also attributed to the right component of heat load. As explained above, low heat consumption values in the low T_{out} zone are due to the DHW heat load's unusual decrease in outdoor air temperature on summer days. Similarly, high heat consumption in the high T_{out} zone is due to SH heat demand in case of unusual peaks of outdoor air temperature on winter days. The peculiar cold climate of the analyzed case study explains the high variability in T_{out} within the hours of the same day.

Table 4 summarizes the KPIs used to describe the performance of the four linear regression models, as presented in Section 2.5, applied to all 27 buildings from the case study. The values shown in the table are the average values for the dataset. By analyzing average KPIs, hourly-based models (CPT-24 and Eref-24) present higher R^2 and lower YEC, SEC, and WEC, highlighting their higher capacity to accurately describe the building's heat load. As already said, the reason is their ability to follow the setpoint scheduling of SH and the peaks in certain hours of the DHW heat load component. Despite the better performances of hourly-based models (CPT-24 and Eref-24), the average R^2 is relatively low, suggesting that these models could not accurately represent the actual behavior of the building. However, average values are obtained by averaging the R^2 obtained in every analyzed building in the dataset. So, low values play an essential role in decreasing the average R^2. Although this effect, low R^2 values are reported in the results to prove two aspects. The first one is that these simplified models struggle to represent the behavior of every building. This is primarily due to their simplicity. This issue is particularly notable in buildings with irregular heat consumption profiles, such as occasionally heated buildings or buildings where DHW heat consumption is in the same order of magnitude or even higher than the SH one. So, other types of models must be chosen to represent heat consumption profiles with high fluctuations accurately.

The second point is that even in the worst cases, hourly-based models (CPT-24 and Eref-24) show better performance than the other approaches.

To conclude the analysis of the R^2 indicator, while hourly-based models (CPT-24 and Eref-24) demonstrate better performance compared to other approaches, their simplicity limits their ability to accurately represent buildings with irregular heat consumption profiles, necessitating the use of more complex models for such cases.

By looking at the comparison between SEC and WEC, summer is the most critical season overall. This confirms that approximating summer heat load (just DHW) with an almost constant value produces high inaccuracies in heat load estimation and may bring a high deviation from the actual heat consumption of the season. However, deviations are lower in the Eref-24 because the 24-hour ESCs mitigate them. DHW heat consumption depends primarily on the users' behavior, and the 24-hour approach allows the account of recursive heat consumption peaks caused by users' habits (i.e., meals, cooking, showers).

The YEC value is close to WEC for all models, which suggests that summer heat consumption inaccuracies slightly affect the accuracy of yearly heat consumption estimation, as summer heat consumption is significantly lower than winter's. This is a consequence of the cold climate of the case study that makes SH the significant component.

YEC values can also be compared with results published in Lumbreras et al. [34], where a different model was used to evaluate the same case study. The mentioned study developed a data-driven model based on multiple linear regression combined with decision trees. Their YEC values go from a minimum of 0 % to a maximum of 18 %, with an average of around 5 %. The best-performing model proposed in the current work (Eref-24) has a mean YEC value of 7 %, indicating that the performance of the models is similar despite the proposed Eref-24 approach being simpler to implement and requiring lower computational resources.

Fig. 7 shows the yearly global heat consumption calculated with the four approaches (colored columns) and measured (dotted column) for every building analyzed in the case study. The Eref-1 approach tends to overestimate yearly global heat consumption for most buildings. The influence of summer peaks in DHW heat consumption of the year causes this. These peaks contribute to an increase in the value of the DHW heat consumption model, although they appear only at specific hours of the day. So, the presence of DHW heat consumption peaks in particular hours of the day, increasing the estimated DHW heat consumption for every hour of the year.

For all approaches, the similarity between the estimated heat consumption and the measured heat consumption is independent of the magnitude of the building's heat consumption.

Table 4
Summary table of KPIs. For every approach, the average value for the dataset and the minimum and maximum values are in brackets [min; max]. The best values are highlighted in bold.

KPI	CPT-24	CPT-1	Eref-24	Eref-1
R^2	0.6	0.55	**0.6**	0.5
	[0.35; 0.85]	[0.15; 0.85]	[0.36; 0.86]	[0.20; 0.80]
YEC	7.2 %	10 %	**7 %**	9.1 %
	[0 %; 16.3 %]	[0 %; 31.2 %]	[0 %; 15.8 %]	[5.9 %; 13.5 %]
WEC	4.9 %	7.7 %	**4.7 %**	6.6 %
	[0 %; 15.6 %]	[0 %; 20.8 %]	[0 %; 15 %]	[3.3 %; 9.5 %]
SEC	32 %	36 %	**26 %**	42 %
	[0 %; 50 %]	[0 %; 68 %]	[0 %; 48 %]	[4 %; 74 %]

4.2. Disaggregation algorithms

The current subsection provides an overview of the performance of disaggregation methods. Fig. 8 shows the results of the four disaggregation methods. All plots represent the heat consumption of a sample building, Bd 08, versus the outdoor air temperature. The SH component is represented in green, the DHW component is represented in red, and the global measured heat consumption is represented in blue. Each plot describes an entire year of values (8760 points − hours).

From the analysis of the plots, many factors suggest the Eref-24 as the best model. First, in all approaches, the linear trend of the SH is evident. However, in the case of CPT-1 and CPT-24, SH heat load stops for T_{out} higher than CPT (13 °C), while in Eref-1 and Eref-24, some SH consumption is still present in correspondence with higher T_{out}. As mentioned, this SH heat load consumption is caused by a sudden increase of T_{out} during daylight hours of change-season days. For Tartu's cold climate, a model that can identify the SH heat component in correspondence with high T_{out} can provide a more realistic representation of the power plant operation. On the contrary, CPT approaches associate such heat load only to DHW, causing unrealistic peaks in DHW heat load consumption, notable in red values between 10 °C and 16 °C in Fig. 8 (a) and (b).

A second factor concerns the variability of the SH heat load component. The variability of SH values increases in CPT-24 and Eref-1 approaches, and it is the highest in the Eref-24 one. Even if SH heat load has a linear dependence on the T_{out}, a higher dispersion of SH points is caused by the better capability of representing the actual behavior of the building, which presents some fluctuations due to differences in users' behavior, days of the week and hour of the day.

The third factor is related to the DHW maximum heat load. In CPT

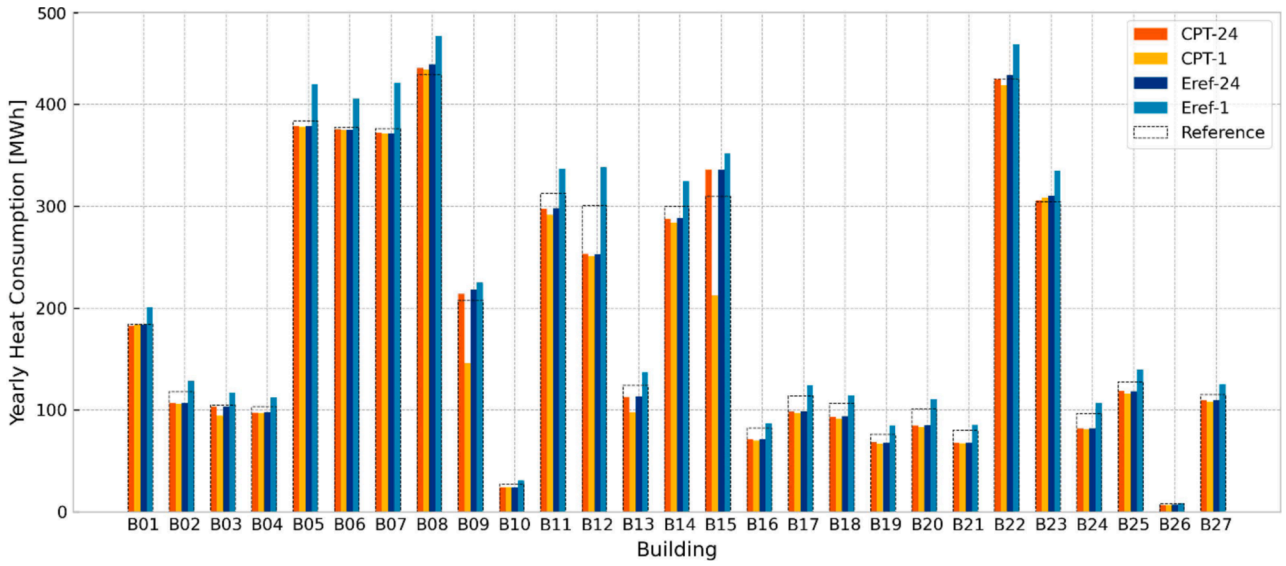

Fig. 7. Yearly heat consumption measured (Reference) and estimated with the four approaches.

approaches, the peak DHW heat load occurs in middle season days (between 10 °C and 16 °C). As mentioned, this unrealistic behavior is caused by the absence of the SH component error imposed by the disaggregation models. In Eref-1, the maximum DHW heat load is limited by the boundary conditions described in Section 2.3.2. In the case of the buildings presenting some heat consumption peaks in summer load (high T_{out}), the upper DHW boundary could be high, producing unrealistic peaks in DHW heat consumption all over the year. In addition, the variability of DHW peaks in the Eref-1 approach is null because all DHW heat load peaks are limited by the same boundary value for the entire year without considering the presence of peaks just in particular hours. The most realistic profile regarding the value and variability of DHW heat load peaks is obtained with the Eref-24 approach.

Fig. 9 shows the DTWD values of the four models for every analyzed building (the colored columns) and the global heat consumption in one year per building (the dotted column). DTWD, expressed in kWh, is the distance between the DHW profile obtained from disaggregation models and the DHW profiles in the summer season for every building, as described in Section 2.5.

The 24 models (CPT-24 and Eref-24) present lower DTWD than the corresponding single-year models (CPT-1 and Eref-1) for every analyzed building. In particular, the DTWD of the CPT-24 approach is around 2.5 times lower than the CPT-24 approach, 8.3 times lower than the CPT-1 approach, and 10 times lower than Eref-1 on average. This confirms the better capability of 24-hour approaches and of the CPT-24 approach in describing the DHW heat load profiles.

Eref-24 shows the lowest DTWD in every analyzed building. Thanks to the combination of the 24-hour models and the boundary limit for DHW heat load, the DHW heat load profile obtained from the disaggregation operation is much more similar to the DHW heat load profile in only the DHW season (summer). The magnitude of the parameters does not seem to correlate with the magnitude of the building's yearly heat consumption.

PML shows acceptable percentages for all the approaches. The PML of minimum consumption hours is around 65 %, the PML of medium consumption hours is 70 %, and the PML of peak consumption hours is 40 % on average for all approaches. While no relevant differences among the approaches emerge in the PML of minimum consumption hours, CPT-24 and Eref-24 show higher percentages in peak consumption hours. The Eref-24 approach has almost 50 % PML in peak consumption.

PMLHR shows higher values than PML due to the definition of this indicator. The average PMLHR of this approach is 78 % for minimum consumption hours, 92 % for medium consumption hours, and 60 % for peak consumption hours. Eref-24 shows higher PMLHR values in medium and peak consumption hours than the other approaches. The same indicator for CPT-1 (the state-of-the-art approach) shows a value of 65 % for minimum consumption hours, 90 % for medium consumption hours, and 50 % for peak consumption hours. DTWD, PML, and PMLHR confirm that DHW profiles obtained with the Eref-24 approach are the most accurate regarding the magnitude and distribution of minimum, medium, and peak consumption hours.

Despite the results showing an increase in performance with the 24 models (CPT-24 and Eref-24) compared to the state-of-the-art approach (CPT-1), there could be some drawbacks.

The first concerns computational time. Instead of creating a single model, these approaches create 24 independent ESC models for each building. However, these simple models require low computational effort, so the increase in computational time is practically not noticeable.

The second drawback is the amount of data needed for model training. When a single model is created for the whole year, the entire dataset is used for training, enhancing the model's accuracy. With the proposed approaches, only 1/24 of the available dataset is used for each ESC. This problem is relevant in cases where a small amount of measures is available, for example, due to a short period of sampling of a low resolution. However, thanks to the model's simplicity, good results can also be obtained with a low number of training points. Further development of the work involves the effect of low-resolution data on the models' performance.

5. Conclusions

This work proposed a simple tool for disaggregating SH and DHW heat load components from a building's global heat consumption. The intent was to use a data-driven model with parameters correlated with the building's characteristics. The necessity was to create a model based only on temperature and heat consumption data without any other information on the building.

The state-of-the-art in this field consists of using the ESC to model the SH heat load and obtaining the DHW component by subtracting the modeled SH to measure global heat consumption. The main issues with this approach are the lack of a dedicated model to DHW, with a consequent imputation of the model uncertainty entirely on the DHW component; the low accuracy in the analysis of building with variable heat load among hours of the day (i.e., setpoint changes); the high

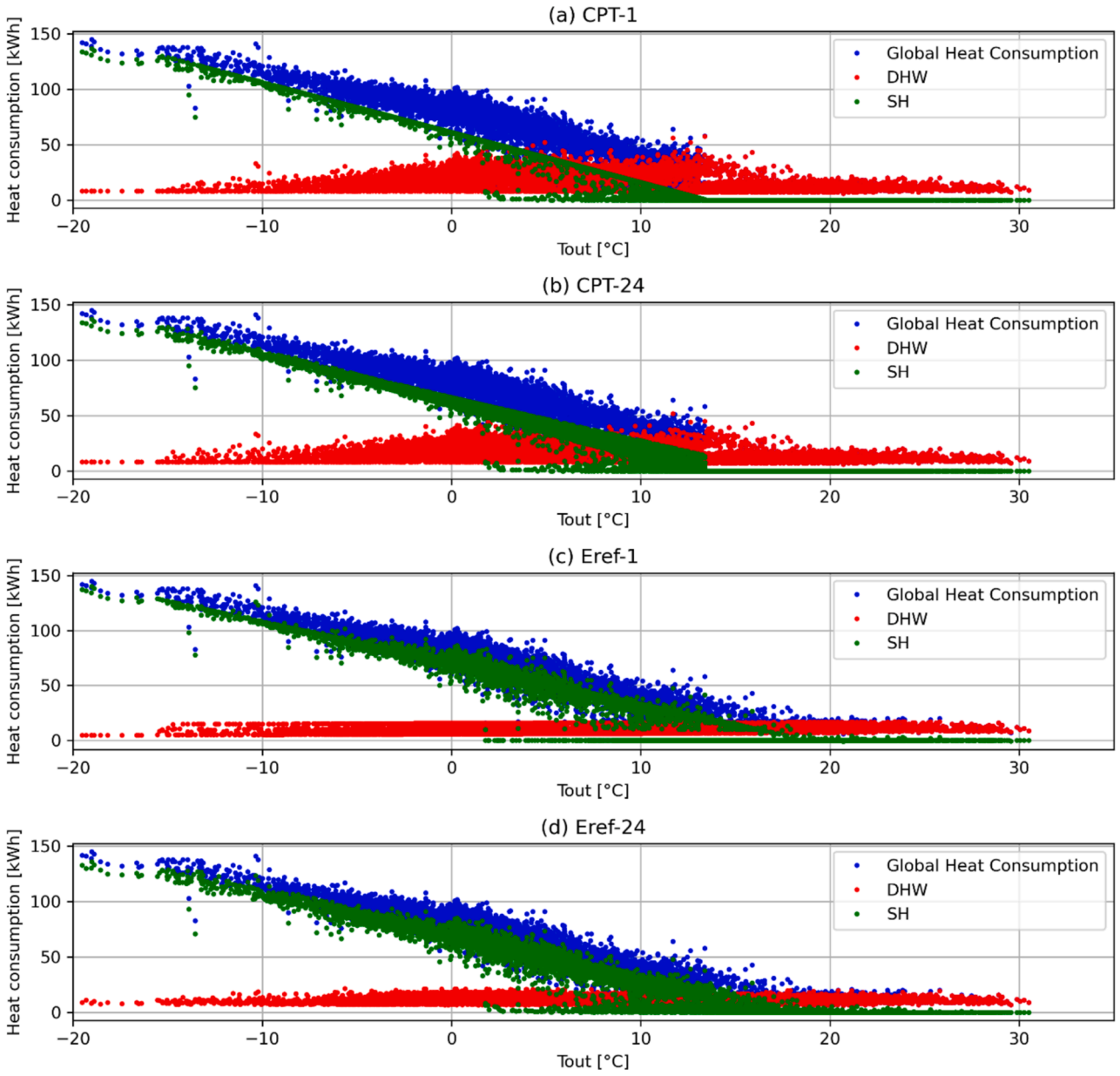

Fig. 8. DHW and SH disaggregated profiles of four approaches for a single building (Bd 08). (a) CPT-1 approach (b) CPT-24 approach; (c) an hour of the Eref-24 approach (5:00 pm – 6:00 pm); (d) Eref-1 approach.

uncertainty in the middle-season days analysis caused by the CPT season identification method.

Four approaches were proposed to solve these problems, and a performance comparison was conducted to find the best approach. An automatic procedure for identifying the heat consumption threshold between the summer and winter (E_{ref}) is proposed.

The four approaches have been applied to 27 buildings connected to the DHN of Tartu, Estonia. Quantitative and qualitative KPIs evaluated the four approaches' performances. R^2, YEC, WEC, and SEC were used to quantify the performances of the linear regression models of the approaches. Models developing independent curves for each hour (CPT-24 and Eref-24) show higher values of R^2 (around 0.6 on average, with a maximum of 0.86) compared to single model approaches (CPT-1 and Eref-1). YEC, WEC, and SEC values confirm that the Eref-24 approach has the highest yearly heat consumption estimation accuracy.

New qualitative KPIs were proposed to overcome the lack of SH and DHW separately measured to use as benchmarks. DTWD, PML, and PMLHR compare the DHW profiles obtained through disaggregation algorithms with the measured summer heat consumption profiles when only DHW heat load is presented.

The Eref-24 approach shows the lowest DTWD in every analyzed building, indicating a higher correspondence between estimated and measured DHW profiles. PML and PMLHR confirm the accuracy of this disaggregation approach. PMLHR indicates an hour identification correspondence of 78 % for minimum consumption hours, 92 % for medium consumption hours, and 60 % for peak consumption hours on average. In addition, the Eref-24 approach creates plausible DHW and SH profiles in buildings with setpoint changes during the day. The results make it possible to conclude that Eref-24 is much better than the state-of-the-art approach (CPT-1) in terms of accuracy, reliability, and range of application.

This disaggregation approach could enhance the management of district heating networks. Using the proposed disaggregation algorithm, the DHW and SH energy consumptions for each building connected to the network can be derived. With this information, the DHN operator can adjust heat production based on outdoor air temperature and the

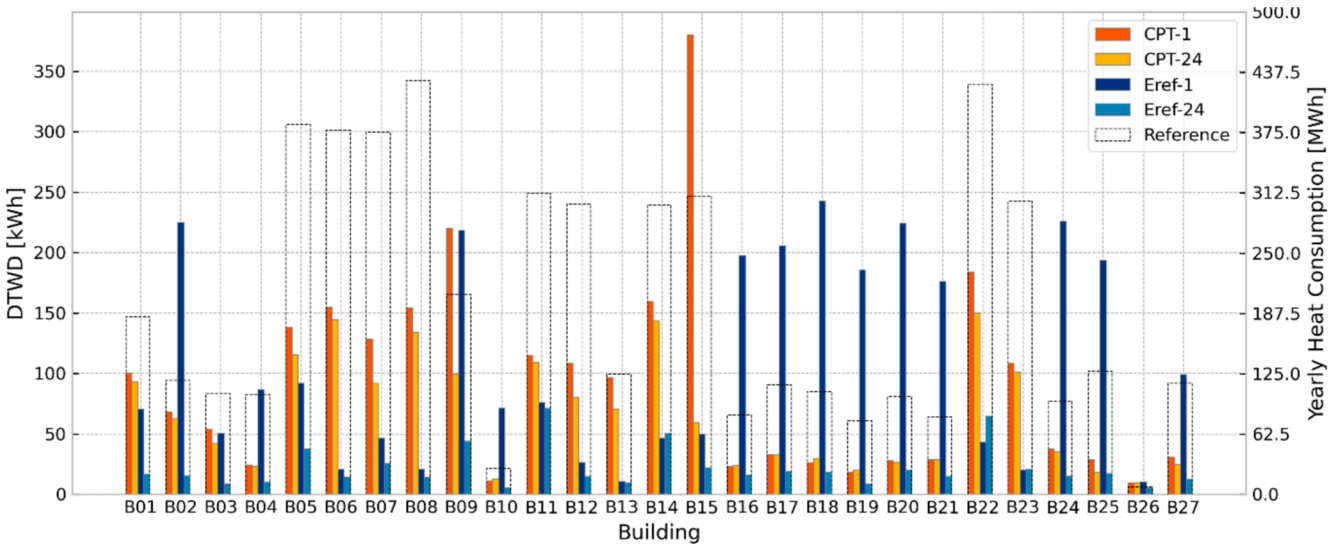

Fig. 9. DTWD and yearly heat consumption measured (Reference) for the four approaches.

consumption patterns of the buildings to minimize heat losses. Additionally, the parameters derived from the Eref-24 approach can be effectively utilized in retrofit interventions.

Declaration of competing interest

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

Appendix A

The study by Toni Giorgino [78] provides a detailed explanation of the Dynamic Time Warping algorithm. DTW compares two time series by stretching or compressing them locally to minimize the distance between them. This appendix summarizes the main points and equations used to implement it in the analysis. Assuming that the goal is comparing two time series: a test or query, $X = (x_1, \ldots, x_N)$, and a reference $Y = (y_1, \ldots, y_M)$. To clarify, the index $i = 1 \ldots N$ will be used to refer to the elements in set X and $j = 1 \ldots M$ to refer to the elements in set Y. Assume there is a non-negative local dissimilarity function f defined for every pair of elements x_i and y_i. The notation is reported in Eq. A(1).

$$d(i,j) = f(x_i, y_i) \geq 0 \tag{A1}$$

In the Eq. A(1) d represents the cross-distance matrix between vectors X and Y. It is the only input to the DTW algorithm. Elements of time series to analyze, x_i and y_i, only enter the computation through the arguments of f. The most common choice is to use the Euclidean distance to compute the cross-distance matrix. However, different definitions could also be helpful, as reported in [78]. The core part of the algorithm consists of the warping curve $\varphi(k)$, $k = 1 \ldots T$:

$$\begin{aligned} &\varphi(k) = \left(\varphi_x(k), \varphi_y(k)\right) \ with \\ &\varphi_x(k) \in \{1 \ldots N\}, \\ &\varphi_y(k) \ \in \{1 \ldots M\} \end{aligned} \tag{A2}$$

The warping functions φ_x and φ_y remap X and Y time indices, respectively. It is a function that approximates a mapping from the time axis of pattern X onto that of pattern Y. To ensure acceptable warps, constraints are usually applied to φ. For example, monotonicity is enforced to preserve their chronological order and avoid nonsensical loops. Their expression is reported in Eq. A(3).

$$\begin{aligned} &\varphi_x(k+1) \geq \varphi_x(k) \\ &\varphi_y(k+1) \geq \varphi_y(k) \end{aligned} \tag{A3}$$

Given φ, the average accumulated distortion between the warped time series X and Y is computed following Eq. A(4).

$$d_\varphi(X,Y) = \sum_{k=1}^{T} \frac{d\left(\varphi_x(k), \varphi_y(k)\right) \times m_\varphi(k)}{M_\varphi} \tag{A4}$$

In Eq. A(4), $d_\varphi(X, Y)$ is the time-normalized distance between two series X and Y, or the average accumulated distortion between the warped time series X and Y, as explained above. $m_\varphi(k)$ is a per-step nonnegative weighting coefficient, $M\varphi$ is the corresponding normalization constant, which ensures that the accumulated distortions are comparable along different paths. The numerator is a reasonable measure for the goodness of the warping function φ. The minimum value of $d_\varphi(X, Y)$ is achieved when the warping function φ is optimally determined to adjust the timing differences. This minimum residual distance can be viewed as the distance between patterns X and Y after the timing differences have been eliminated. It is expected to

remain stable despite fluctuations along the time axis. So, the expression of the dynamic time warping distance (DTWD) between the two analyzed time series is expressed in Eq. A(5).

$$DTWD(X,Y) = \min_{\varphi} d_{\varphi}(X,Y) \tag{A5}$$

The DTWD value represents the stretch-insensitive measure of the inherent difference between two time series. With this index it is possible to check the similarity between two time series even if they are not aligned in the temporal axis.

Data availability

Data will be made available on request.

References

[1] EEA - European Environment Agency, "EEA greenhouse gases - data viewer," EEA greenhouse gases. Accessed: Dec. 22, 2023. [Online]. Available: https://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer.

[2] European Environment Agency, Annual European Union greenhouse gas inventory 1990–2021 and inventory report 2023, 2023. Accessed: Mar. 26, 2024. [Online]. Available: https://www.eea.europa.eu/publications/annual-european-union-greenhouse-gas-2.

[3] A. Thonipara, P. Runst, C. Ochsner, K. Bizer, Energy efficiency of residential buildings in the European Union-An exploratory analysis of cross-country consumption patterns ☆, 2019, doi: 10.1016/j.enpol.2019.03.003.

[4] EU, "Energy efficiency in buildings," 2020.

[5] EU, Energy efficiency in buildings, 2020.

[6] M. in Choi et al., Empirical study on optimization methods of building energy operation for the sustainability of buildings with integrated renewable energy, Energy Build, vol. 305, p. 113908, Feb. 2024, doi: 10.1016/J.ENBUILD.2024.113908.

[7] K.S. Cetin, Z. O'Neill, Smart meters and smart devices in buildings: a review of recent progress and influence on electricity use and peak demand, Curr. Sustain./Renew. Energy Reports 4 (1) (Mar. 2017) 1–7, https://doi.org/10.1007/S40518-017-0063-7/METRICS.

[8] P.S.G. De Mattos Neto et al., Energy consumption forecasting for smart meters using extreme learning machine ensemble, Sensors 21 (2021) 8096, doi: 10.3390/S21238096.

[9] Y. Wang, Q. Chen, T. Hong, C. Kang, Review of smart meter data analytics: applications, methodologies, and challenges, IEEE Trans. Smart Grid 10 (3) (May 2019) 3125–3148, https://doi.org/10.1109/TSG.2018.2818167.

[10] D. Ivanko, H.T. Walnum, N. Nord, Development and analysis of hourly DHW heat use profiles in nursing homes in Norway, 110070, *Energ. Buildings* 222 (2020), https://doi.org/10.1016/j.enbuild.2020.110070.

[11] D.H.W. Li, et al., Estimating residential space heating and domestic hot water from truncated smart heat data, 022017, *J. Phys. Conf. Ser.* 2600 (2) (Nov. 2023), https://doi.org/10.1088/1742-6596/2600/2/022017.

[12] A. Jahanbin, G. Semprini, M. Goni, Techno-economic analysis of a novel retrofit solution for the domestic hot water system: a comparative study, 117363, *Energy Convers Manag* 292 (Sep. 2023), https://doi.org/10.1016/J.ENCONMAN.2023.117363.

[13] K. Lygnerud, et al., A study on how efficient measures for secondary district heating system performance can be encouraged by motivational tariffs, *Energy Sustain Soc* 13 (1) (2023) 1–15, https://doi.org/10.1186/S13705-023-00417-0/FIGURES/10.

[14] M. Sharifi, A. Kouti, E. Lambie, Y. Ma, M.F. Boneta, M.H. Shamsi, A comprehensive framework for data-driven building end-use assessment utilizing monitored operational parameters, *Energies (Basel)* 16 (20) (2023) Oct, https://doi.org/10.3390/en16207132.

[15] B. Vand, R. Ruusu, A. Hasan, and B. Manrique Delgado, Optimal management of energy sharing in a community of buildings using a model predictive control, Energy Convers Manag 239 (2021) 114178, doi: 10.1016/J.ENCONMAN.2021.114178.

[16] A. Golla, J. Geis, T. Loy, P. Staudt, C. Weinhardt, An operational strategy for district heating networks: application of data-driven heat load forecasts, *Energy Informatics* 3 (1) (Oct. 2020) 1–11, https://doi.org/10.1186/S42162-020-00125-5/TABLES/6.

[17] A. Pena-Bello, P. Schuetz, M. Berger, J. Worlitschek, M.K. Patel, D. Parra, Decarbonizing heat with PV-coupled heat pumps supported by electricity and heat storage: Impacts and trade-offs for prosumers and the grid, 114220, *Energy Convers Manag* 240 (Jul. 2021), https://doi.org/10.1016/J.ENCONMAN.2021.114220.

[18] T. Peters, J. Weyer, "7. Passivhus Norden | Sustainable Cities and Buildings Brings practitioners and researchers together Architectural Design for Low Energy Housing-Experiences From Two Recent Affordable Housing Projects in Denmark, 2015. [Online]. Available: www.7phn.org.

[19] V. Földváry, D. Petráš, Energy use and thermal comfort of two apartment buildings before and after refurbishment in Slovakia, 2016, Accessed: Mar. 26, 2024. [Online]. Available: https://www.researchgate.net/publication/293488143.

[20] M. Karami, F. Javanmardi, Performance assessment of a solar thermal combisystem in different climate zones, Asian J. Civil Eng. 21 (5) (Jul. 2020) 751–762, https://doi.org/10.1007/S42107-020-00236-0/TABLES/3.

[21] X. Li, J. Wen, Review of building energy modeling for control and operation, Renew. Sustain. Energy Rev. 37 (Sep. 2014) 517–537, https://doi.org/10.1016/J.RSER.2014.05.056.

[22] V.S.K.V. Harish, A. Kumar, A review on modeling and simulation of building energy systems, *Renew. Sustain. Energy Rev.* 56 (Apr. 2016) 1272–1292, https://doi.org/10.1016/J.RSER.2015.12.040.

[23] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, P. Chatellier, "Modeling and forecasting building energy consumption: a review of data-driven techniques," Sustain. Cities Soc. 48 (2019) 101533, doi: 10.1016/J.SCS.2019.101533.

[24] Y. Wei, et al., A review of data-driven approaches for prediction and classification of building energy consumption, *Renew. Sustain. Energy Rev.* 82 (Feb. 2018) 1027–1047, https://doi.org/10.1016/J.RSER.2017.09.108.

[25] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: a review, *Renew. Sustain. Energy Rev.* 23 (Jul. 2013) 272–288, https://doi.org/10.1016/J.RSER.2013.03.004.

[26] M.S. Al-Homoud, Computer-aided building energy analysis techniques, *Build. Environ.* 36 (4) (May 2001) 421–433, https://doi.org/10.1016/S0360-1323(00)00026-3.

[27] E. Fuentes, L. Arce, J. Salom, A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis, *Renew. Sustain. Energy Rev.* 81 (Jan. 2018) 1530–1547, https://doi.org/10.1016/J.RSER.2017.05.229.

[28] ASHRAE, ASHRAE 90.2 Energy-Efficient Design of Low-Rise Residential Buildings, 1994.

[29] UNE and EN, UNE EN12976. Thermal solar systems components. Factory made systems. Test methods, 2006.

[30] EN, EN 15316-3-1:2006. Heating systems in buildings. Method for calculation of system energy requirements and system efficiencies. Domestic hot water systems, characterisation of needs., 2007.

[31] L. Tronchin, K. Fabbri, Energy performance building evaluation in Mediterranean countries: comparison between software simulations and operating rating simulation, Energ. Build. 40 (7) (Jan. 2008) 1176–1187, https://doi.org/10.1016/J.ENBUILD.2007.10.012.

[32] Y. Sun, F. Haghighat, B.C.M. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, 110022, *Energ. Build.* 221 (Aug. 2020), https://doi.org/10.1016/J.ENBUILD.2020.110022.

[33] K. Arendt, M. Jradi, H. R. Shaker, C.T. Veje, Comparative analysis of white-, gray- and black-box models for thermal simulation of indoor environment: teaching building case study, in: 2018 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA, 2018, Accessed: Mar. 26, 2024. [Online]. Available: www.ibpsa.us.

[34] M. Lumbreras, et al., Data driven model for heat load prediction in buildings connected to District Heating by using smart heat meters, 122318, *Energy* 239 (Jan. 2022), https://doi.org/10.1016/J.ENERGY.2021.122318.

[35] R.J. Hyndman, G. Athanasopoulos, Chapter 8 ARIMA models, Forecasting: principles and practice, 2nd edition. Accessed: Mar. 26, 2024. [Online]. Available: https://otexts.com/fpp2/arima.html.

[36] N. Pachauri, C.W. Ahn, Regression tree ensemble learning-based prediction of the heating and cooling loads of residential buildings, *Build. Simul.* 15 (11) (Nov. 2022) 2003–2017, https://doi.org/10.1007/S12273-022-0908-X/METRICS.

[37] C. Deb, F. Zhang, J. Yang, S.E. Lee, K.W. Shah, A review on time series forecasting techniques for building energy consumption, *Renew. Sustain. Energy Rev.* 74 (2017) 902–924, https://doi.org/10.1016/J.RSER.2017.02.085.

[38] Z. Wang, R.S. Srinivasan, A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models, in: Proceedings - Winter Simulation Conference, vol. 2016-February, pp. 3438–3448, Feb. 2016, doi: 10.1109/WSC.2015.7408504.

[39] D. Mariano-Hernández, L. Hernández-Callejo, F. S. García, O. Duque-Perez, A.L. Zorita-Lamadrid, A Review of energy consumption forecasting in smart buildings: methods, input variables, forecasting horizon and metrics, Applied Sciences 10 (2020) 8323, doi: 10.3390/APP10238323.

[40] P. Bacher, P. Anton de Saint-Aubain, L. Engbo Christiansen, H. Madsen, Non-parametric method for separating domestic hot water heating spikes and space heating, *Energ. Buildings* 130 (2016) 107–112, https://doi.org/10.1016/j.enbuild.2016.08.037.

[41] D. Leiria, H. Johra, A. Marszal-Pomianowska, M.Z. Pomianowski, A methodology to estimate space heating and domestic hot water energy demand profile in residential buildings from low-resolution heat meter data, 125705, *Energy* 263 (Jan. 2023), https://doi.org/10.1016/J.ENERGY.2022.125705.

[42] D. Ivanko, Å. Lekang Sørensen, N. Nord, "Splitting measurements of the total heat demand in a hotel into domestic hot water and space heating heat use," 2020, doi: 10.1016/j.energy.2020.119685.

[43] L. Georges, M. Haase, V. Novakovic, P.G. Schild, "Domestic hot water decomposition from measured total heat load in Norwegian buildings, 31-38, pp. 13–14, 2020, Accessed: Mar. 27, 2024. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2684373.
[44] Y. Hu, et al., A data-driven approach for the disaggregation of building-sector heating and cooling loads from hourly utility load data, 101175, *Energ. Strat. Rev.* 49 (Sep. 2023), https://doi.org/10.1016/J.ESR.2023.101175.
[45] M. Schaffer, J. Widén, J.E. Vera-Valdés, A. Marszal-Pomianowska, T.S. Larsen, Disaggregation of total energy use into space heating and domestic hot water: a city-scale suited approach, 130351, *Energy* 291 (Mar. 2024), https://doi.org/10.1016/J.ENERGY.2024.130351.
[46] R.E. Hedegaard, M.H. Kristensen, S. Petersen, "Experimental validation of a model-based method for separating the space heating and domestic hot water components from smart-meter consumption data, E3S Web of Conferences, vol. 172, p. 12001, Jun. 2020, doi: 10.1051/E3SCONF/202017212001.
[47] Y. Li, Z. O'Neill, L. Zhang, J. Chen, P. Im, J. DeGraw, Grey-box modeling and application for building energy simulations – a critical review, 111174, *Renew. Sustain. Energy Rev.* 146 (Aug. 2021), https://doi.org/10.1016/J.RSER.2021.111174.
[48] C. Deb, A. Schlueter, Review of data-driven energy modelling techniques for building retrofit, 110990, *Renew. Sustain. Energy Rev.* 144 (Jul. 2021), https://doi.org/10.1016/J.RSER.2021.110990.
[49] S. Oh, J.F. Gardner, Large scale energy signature analysis: tools for utility managers and planners, Sustainability 14 (2022) 8649, doi: 10.3390/SU14148649.
[50] M. Bhatnagar, J. Mathur, V. Garg, Determining base temperature for heating and cooling degree-days for India, J. Build. Eng. 18 (Jul. 2018) 270–280, https://doi.org/10.1016/J.JOBE.2018.03.020.
[51] Y. Ding, H. Brattebø, N. Nord, A systematic approach for data analysis and prediction methods for annual energy profiles: an example for school buildings in Norway, 111160, Energ. Build. 247 (Sep. 2021), https://doi.org/10.1016/J.ENBUILD.2021.111160.
[52] C. Ghiaus, Experimental estimation of building energy performance by robust regression, *Energ. Buildings* 38 (6) (Jun. 2006) 582–587, https://doi.org/10.1016/J.ENBUILD.2005.08.014.
[53] R. Hitchin, I. Knight, Daily energy consumption signatures and control charts for air-conditioned buildings, *Energ. Buildings* 112 (Jan. 2016) 101–109, https://doi.org/10.1016/J.ENBUILD.2015.11.059.
[54] M. Eriksson, J. Akander, B. Moshfegh, Development and validation of energy signature method – Case study on a multi-family building in Sweden before and after deep renovation, 109756, *Energ. Buildings* 210 (2020), https://doi.org/10.1016/j.enbuild.2020.109756.
[55] S. Schneider, P. Hollmuller, J. Chambers, M. Patel, A heat demand load curve model of the swiss national territory, IOP Conf Ser Earth Environ Sci 290(1) (2019), doi: 10.1088/1755-1315/290/1/012107.
[56] T. Day, K. Morris, I. Chaer, A. Gillich, An energy analysis methodology for residential heat pump retrofits, Build. Serv. Eng. Res. Technol. (2024), https://doi.org/10.1177/01436244241289521.
[57] S. Werner, District heating and cooling, Reference module in earth systems and environmental sciences, 2013, doi: 10.1016/B978-0-12-409548-9.01094-0.
[58] L. Pedersen, Load modelling of buildings in mixed energy distribution systems, no. February. 2007. [Online]. Available: http://ntnu.diva-portal.org/smash/record.jsf?pid=diva2:122458.
[59] I. Meireles, V. Sousa, B. Bleys, B. Poncelet, Domestic hot water consumption pattern: Relation with total water consumption and air temperature, 112035, *Renew. Sustain. Energy Rev.* 157 (Apr. 2022), https://doi.org/10.1016/J.RSER.2021.112035.
[60] J.Y. Lee, T. Yim, Energy and flow demand analysis of domestic hot water in an apartment complex using a smart meter, 120678, *Energy* 229 (Aug. 2021), https://doi.org/10.1016/J.ENERGY.2021.120678.
[61] Y. Ding, T.O. Timoudas, Q. Wang, S. Chen, H. Brattebø, N. Nord, A study on data-driven hybrid heating load prediction methods in low-temperature district heating: An example for nursing homes in Nordic countries, 116163, *Energy Convers Manag* 269 (Oct. 2022), https://doi.org/10.1016/J.ENCONMAN.2022.116163.
[62] R. Zmeureanu, Assessment of the energy savings due to the building retrofit, *Build. Environ.* 25 (2) (Jan. 1990) 95–103, https://doi.org/10.1016/0360-1323(90)90020-R.
[63] J.D. Balcomb, Energy signatures: a proposed new design tool, 1986.
[64] P. Rohdin, V. Milic, M. Wahlqvist, B. Moshfegh, On the use of Change-point models to describe the energy performance of historic buildings, in: The 3rd International Conference on Energy Efficiency in Historic Buildings, 2018, pp. 512–520.
[65] T. Tereshchenko, D. Ivanko, N. Nord, I. Sartori, Analysis of energy signatures and planning of heating and domestic hot water energy use in buildings in Norway, E3S Web of Conferences 111 (2019) 06009, doi: 10.1051/E3SCONF/201911106009.
[66] A. Anjomshoaa, M. Salmanzadeh, Estimation of the changeover times and degree-days balance point temperatures of a city using energy signatures, *Sustain. Cities Soc.* 35 (Nov. 2017) 538–543, https://doi.org/10.1016/J.SCS.2017.08.028.
[67] V. Milić, P. Rohdin, Screening of thermal characteristics and assessment of comparative energy efficiency potential in a residential district, *Adv. Build. Energy Res.* 17 (3) (May 2023) 255–276, https://doi.org/10.1080/17512549.2023.2183522.
[68] A. Acquaviva et al., Energy signature analysis: knowledge at your fingertips, in: Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015, pp. 543–550, Aug. 2015, doi: 10.1109/BIGDATACONGRESS.2015.85.
[69] G. Nordström, H. Johnsson, S. Lidelöw, Using the energy signature method to estimate the effective U-value of buildings, Smart Innovat. Syst. Technol. 22 (2013) 35–44, https://doi.org/10.1007/978-3-642-36645-1_4/COVER.
[70] Y. Zhang, Z. O'Neill, B. Dong, G. Augenbroe, Comparisons of inverse modeling approaches for predicting building energy performance, *Build. Environ.* 86 (Apr. 2015) 177–190, https://doi.org/10.1016/J.BUILDENV.2014.12.023.
[71] Q. Meng, et al., Change-point multivariable quantile regression to explore effect of weather variables on building energy consumption and estimate base temperature range, 101900, *Sustain. Cities Soc.* 53 (Feb. 2020), https://doi.org/10.1016/J.SCS.2019.101900.
[72] D. Lindelöf, Bayesian estimation of a building's base temperature for the calculation of heating degree-days, *Energ. Build.* 134 (Jan. 2017) 154–161, https://doi.org/10.1016/J.ENBUILD.2016.10.038.
[73] M. Lumbreras, G. Diarce, K. Martin, R. Garay-Martinez, B. Arregi, Unsupervised recognition and prediction of daily patterns in heating loads in buildings, 105732, *J. Build. Eng.* 65 (Apr. 2023), https://doi.org/10.1016/J.JOBE.2022.105732.
[74] L. Pedersen, R. Ulseth, Method for load modelling of heat and electricity demand method for load modelling of heat and electricity demand, in: 10th International Symposium on District Heating and Cooling, vol. Sektion 5 b, no. Heat/cold generation, 2006.
[75] Brakel JPGV, "Robust peak detection algorithm using z-scores." Accessed: Apr. 04, 2024. [Online]. Available: https://stackoverflow.com/questions/22583391/peak-signal-detectionin-realtime-timeseries-data/22640362#22640362.
[76] Eurostat, How women and men spend their time : 12/2003, 2003.
[77] P. Senin, Dynamic Time Warping Algorithm Review, 2008.
[78] D.J. Berndt, J. Clifford, Using dynamic time warping to find patterns in time series, in: Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, in AAAIWS'94. AAAI Press, 1994, pp. 359–370.
[79] T. Giorgino, Journal of statistical software computing and visualizing dynamic time warping alignments in R: the dtw package, *J. Stat. Soft.* 31 (2009), https://doi.org/10.18637/jss.v031.i07.
[80] GREN Eesti. Accessed: Apr. 12, 2024. [Online]. Available: https://gren.com/ee/.