

**A methodology to estimate space heating and domestic hot water energy demand profile in residential buildings from low-resolution heat meter data**

Leiria, Daniel; Johra, Hicham; Marszal-Pomianowska, Anna; Pomianowski, Michal Zbigniew

*Published in:*  
Energy

*DOI (link to publication from Publisher):*  
[10.1016/j.energy.2022.125705](https://doi.org/10.1016/j.energy.2022.125705)

*Creative Commons License*  
CC BY 4.0

*Publication date:*  
2023

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*

Leiria, D., Johra, H., Marszal-Pomianowska, A., & Pomianowski, M. Z. (2023). A methodology to estimate space heating and domestic hot water energy demand profile in residential buildings from low-resolution heat meter data. *Energy*, 263(Part B), Article 125705. <https://doi.org/10.1016/j.energy.2022.125705>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

**Take down policy**

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.





# A methodology to estimate space heating and domestic hot water energy demand profile in residential buildings from low-resolution heat meter data

Daniel Leiria<sup>\*</sup>, Hicham Johra, Anna Marszal-Pomianowska, Michal Zbigniew Pomianowski

Aalborg University, Department of the Built Environment, Thomas Manns Vej 23, 9220, Aalborg Øst, Denmark

## ARTICLE INFO

### Keywords:

Smart energy meter  
Data disaggregation  
Building energy usage  
Load profiles  
Time series analysis  
District heating

## ABSTRACT

This article presents a new methodology to disaggregate the energy demand for space heating (SH) and domestic hot water (DHW) production from single hourly smart heat meters installed in Denmark. The new approach is idealized to be easily applied to several building typologies without the necessity of in-depth knowledge regarding the dwellings and their occupants. This paper introduces, tests, and compares several algorithms to separate and estimate the SH and DHW demand. To validate the presented methodology, a dataset of 28 Danish apartments with detailed energy monitoring (separated SH and DHW usage) is used. The comparison shows that the best method to identify energy demand data points corresponding to DHW production events is the so-called “maximum peaks” approach. Furthermore, the best algorithm to estimate the SH and DHW separately is a combination of two methods: the Kalman filter and the Support Vector Regression (SVR). This new methodology outperforms the current Danish compliances typically used to estimate the annual DHW usage in residential buildings.

## Credit author statement

**Daniel Leiria:** Conceptualization; Methodology; Software; Validation; Formal analysis; Investigation; Data curation; Writing - original draft; Writing - review & editing; Visualization. **Hicham Johra:** Conceptualization; Methodology; Resources; Writing - review & editing; Supervision. **Anna Marszal-Pomianowska:** Conceptualization; Resources; Writing - review & editing; Supervision. **Michal Zbigniew Pomianowski:** Conceptualization; Resources; Writing - review & editing; Supervision; Project administration; Funding acquisition. All authors have read and agreed to the published version of the manuscript.

## 1. Introduction

With the growing global concern regarding climate changes and the sustainability of our technologies, the different sectors of our society are challenged and urged to take a sharp turn to alleviate their impact on the environment. This is especially the case for energy production, distribution, and usage activities. Among them, the building sector has a major role in this sustainability transition. According to [1], the European Union (EU) building sector has an estimated share of 40% of the total energy end-usage, where 79% of it is for space heating (SH) and

domestic hot water (DHW) production alone [2]. Specifically, in Denmark, 81.8% of the annual energy is used for heating (SH and DHW) in a typical house, while the other appliances (electrical consumers, lighting, etc.) have an annual share of 18.2% [3]. Regarding the Danish heating demand, 64% of the housing stock is connected to the district heating (DH) network. Furthermore, around 50% of the building stock in Iceland, Lithuania, Estonia, Sweden, Finland, Russia, Poland, and Northern China have their energy demand for space heating, cooling, and domestic hot water provided by district heating and cooling (DHC) networks [4]. The DHC systems and their potential for cost-effective, flexible, and sustainable heating and cooling supply are considered a strategic component of the roadmap toward a low-carbon future and gas-free neighborhoods in Europe, the USA, Canada, and Asia [5,6].

Research in the field of DH system improvement and integration of renewable energy sources leads to new DH concepts or configurations called “generations”. Currently, the newly-installed and refurbished DH networks are transitioning from the 3rd to the 4th generation [7]. The 4th generation of district heating (4GDH) systems is mainly characterized by low-temperature heat-carrier fluid supply (40–70°C). The articles [7–10] outline several uprising advantages of implementing the 4GDH systems. Some of these advantages are the increase of energy efficiency in the network distribution due to the lowering of heat losses, a higher output capacity from different low-temperature sources

<sup>\*</sup> Corresponding author.

E-mail address: [dle@build.aau.dk](mailto:dle@build.aau.dk) (D. Leiria).

<https://doi.org/10.1016/j.energy.2022.125705>

Received 12 April 2022; Received in revised form 28 August 2022; Accepted 8 October 2022

Available online 12 October 2022

0360-5442/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature			
<i>Acronyms</i>			
4GDH	4th generation district heating	C	Cost (SVR parameter)
CPT	Change point temperature	$C_{p,water}$	Specific heat capacity of water
DHW	Domestic hot water	$E_{DHW, estim}$	Estimated domestic hot water energy usage
DH	District heating	$E_{DHW}$	Measured domestic hot water energy usage
DHC	District heating and cooling	$E_{DHW, compl}$	Yearly estimated domestic hot water energy usage by the Danish compliance calculation
EU	European Union	$E_{SH, estim}$	Estimated space heating energy usage
HVAC	Heating, ventilation, and air conditioning	$E_{SH, max}$	Maximum measured space heating in the dataset
NA	Not available – missing data point	$E_{SH}$	Measured space heating energy usage
nZEB	Nearly zero-energy building	$E_{SH, min}$	Minimum measured space heating in the dataset
SH	Space heating	$\overline{E_{SH}}$	Mean measured space heating in the dataset
SMA	Simple moving average	$E_{Total}$	Measured total heat demand for space heating and domestic hot water production
SVR	Support vector regression	$T_{cold}$	Cold water supply temperature from Danish standards
NMBE	Normalized mean bias error	$T_{DHW}$	Domestic hot water supply temperature from Danish standards
CVRMSE	Coefficient of variation of the root mean square error	$\alpha$	Moving average weight
<i>Symbols and variables</i>		$\gamma$	Gamma (SVR parameter)
A	Heated area	$\rho_{water}$	Water density

integrated with DH systems, a smaller risk of pipe leakages caused by thermal stress, a better relation with the new building requirements regarding thermal usage, to name a few. Nevertheless, the 4GDH transition also faces particular challenges in decreasing the supply temperature needed for the building's SH and DHW demands. The challenges are the proper coordination in integrating the multiple low-temperature heat and waste-heat sources (renewable and recycled), the coupling to other energy grids (e.g., electricity, gas), the smart monitoring and control of such thermal grids and all its sub-components (including accurate prediction of production and demand, and demand-side management of the heat end users), the cost-effectiveness and the achievement of high reliability of heat supply at all time within a given (and often inflexible) legislative framework, and the operation of oversized or faulty systems on the building side.

It is clear from the barriers stated above that it is necessary to understand the DH network in detail. Therefore [7,11], outline the importance of smart meter data in the future of district heating. This metering initiative makes it possible: to efficiently manage the energy production, distribution grid, and the end-consumers; to optimize the DH system and its interconnection with other energy sources; to detect and fix the different faults occurring in the system; and to provide more information to the end-users regarding their energy usage, instigating them to change their consumption behavior.

As a front-runner, Denmark has made a great effort to install smart heat meters in buildings connected to the heating grid, and from 2027 it will be obligatory to collect dynamic heating data by using smart meters for every building connected to the DH grid [12]. These meters have up to 1-hour resolution measurements, and their collected data is easily accessible by utility companies. This metering initiative aims to obtain a detailed insight into the heat load patterns in each building and, when coupled with other sources of information, to unravel the reasons behind them. Even though this initiative is a significant step toward reaching the energetic goals set by Denmark [13,14] and the EU [15], it has a major drawback with respect to its data collection. In most buildings, only one smart energy meter is being installed per household. Each meter thus collects the total heat usage without distinguishing the energy used for SH or DHW production. Regarding SH, it depends on the outdoor conditions, building characteristics, occupants' preferences, and installed space heating systems [16]. In contrast, DHW production is correlated with people's consumption habits and the installed hot water production system. Because these two types of energy usage are associated with different variables, it is essential to estimate them separately

to have a deeper insight into the building itself and its occupants [17].

Another aspect to consider on the importance of knowing these energy shares is regarding refurbishment initiatives. In [18], the authors argue that global building regulations have stricter SH efficiency rules while overlooking DHW consumption. Therefore, these new buildings, also known as low-energy buildings, have a much higher DHW share due to the continuous decrease of SH usage over the years and the higher levels of comfort concerning heating practices demanded by the residents.

Thus, a better assessment of the thermal appliances can be achieved by disaggregating the energy used in buildings. This contributes to a more detailed understanding and control on the user side and promotes better decision-making strategies regarding heat production and distribution.

### 1.1. Literature review

As mentioned above, most installed smart energy meters only measure the building's total heat usage. These total measurements often equal the sum of SH and DHW in a household. Even though this is already a great source of information, a clear distinction between SH and DHW production must be made. To tackle this problem, several research studies have developed different methods to estimate both utilities from total heating measurements. The present research focuses on instantaneous DHW production systems without thermal storage tanks, due to being a typical installation in Danish households, and all apartments in the dataset had this type of system. Hence most of the reviewed articles are regarding disaggregating methods applied in these systems.

One of the first studies to explore this problem is [19], which presents a statistical time-series approach to estimate the SH from the total heat usage measurements. The method assumes that the space heating demand varies smoother due to small outdoor temperature changes than the DHW usage, which, conversely, is more sporadic with higher peaks due to the very short time length of the different hot water draw-off events. This method estimates the SH by applying a kernel smoother to the total data points, where all measurements above a defined smoothed threshold are due to DHW usage. This method seems promising, and the authors formulated several kernel functions to increase the estimation accuracy. Nevertheless, it still lacks validation with separated space heating and DHW usage measurements, which the authors did not have at the time. Another drawback of this method is the necessity of high-resolution data (10-min measurements) to detect the

sporadic peaks from DHW usage. Unfortunately, most of the installed smart meters in Denmark do not provide these high-frequency measurements.

Differently in [20], a simpler methodology is proposed to disaggregate the smart meters data by considering that the total measurements are equal to the DHW usage during summer, i.e., no SH demand. Based on this assumption, their approach does not estimate the different household heating utilities during the whole year but estimates the household average DHW load profile. If defined correctly, this type of profile provides valuable information concerning the customers' DHW habits. Regarding the method's accuracy, it is shown that it performs better for newly-built households with a large DHW usage share. However, the authors also concluded that several houses use space heating during summer, invalidating their initial assumption and significantly decreasing the profile accuracy. Similarly, in [21], a method is proposed to decompose SH and DHW usage in total measurements. The proposed method is called *hybrid summer signature*. It is based on discovering the DHW profiles when the total heating is equal to the DHW usage (no SH demand), taking into account the outdoor temperature. When the DHW profiles are discovered, the space heating demand equals the subtraction of the total values and the DHW daily profiles. The method was validated with several Norwegian buildings (apartments and hotels) and compared with other existing methods.

In [22], another approach is proposed to separate the different measurements in a Norwegian hotel. Two methods were presented and compared. Both approaches began by estimating the SH demand through its linear dependency on the outdoor temperature. The main difference between the methods is that the first calculates the DHW needs by subtracting the estimated SH from the total measured heat demand. And the second method, before calculating the DHW usage, the SH (already calculated by its outdoor temperature dependency) is adjusted by applying a singular spectrum analysis algorithm. The second methodology had the highest accuracy in predicting both heating utilities. With a different approach [23], estimates the SH and DHW usage weekly profiles using grey-box models. Their study concluded that the calculated values were slightly overestimated compared to the actual measurements. However, the method is accurate, and the authors argue that the models can be improved to increase even further its accuracy. The methodology developed in [24] is also worth mentioning. A pattern recognition algorithm was applied to disaggregate SH from other appliances in two households in the UK. Nevertheless, the household's heating source is a natural gas boiler instead of DH to provide thermal energy to SH, DHW, and cooking utilities (e.g., oven).

## 1.2. Contributions

Some of the methods developed to disaggregate the heating measurements are present in the section above. However, they have some drawbacks that this methodology attempts to solve. Firstly, this novel method aims to separate these energy shares using 1-h resolution measurements, which was proven by [19,23] to be extremely difficult and susceptible to inaccurate estimations. Another problem that the present methodology seeks to address is its non-dependence on other sources of information. Some of the reviewed methods require more information regarding the building (e.g., thermal envelope properties) and people (e.g., consumption habits) to proceed with SH and DHW estimation. This information is usually difficult to retrieve. Therefore, the proposed technique requires only the hourly total recorded heating values from the heat meters and the associated local weather data (outdoor temperature and global radiation). Lastly, the methodology algorithms were made simple and easy to implement and do not require any grey-box models' calibration.

Moreover, the contributions of this paper are:

1. The development of a new methodology to disaggregate SH and DHW from 1-h resolution total heating measurements. Besides, the

method's algorithm is created to be easily implemented and only requires weather data as input.

2. The validation of the present methodology with a dataset of separated measurements of the different heating appliances from 28 Danish single-family apartments. All the apartments have an instantaneous DHW production system without a storage tank.
3. The comparison between DHW demand estimated through our disaggregation method and the current Danish annual DHW compliances. In order to assess the method's performance compared with the current calculation used in Denmark for the energy labeling in buildings.

## 1.3. Outline

Following this section, the developed methodology is described. The results from the method's validation are presented and discussed in section 3. The article closes with the main conclusions and suggestions for further work in sections 4 and 5. All the algorithms developed in this work are coded with the software *Rstudio* [25].

## 2. Methodology

### 2.1. Research roadmap

The method assumes that the SH system continuously operates during the heating season. At the same time, the DHW usage is expected to be produced sporadically throughout the day. Thus, during a day (which has 24 recorded data points – hourly resolution), only a few of these points will consist of collective SH and DHW production, whereas the other measurements will be SH usage alone. Every measurement identified with DHW production is converted to a missing point (NA point). Hence, each NA value is constituted by two energy shares, one for SH and another for DHW usage. Conversely, the non-NA values only have the SH share. Because the non-NA points in the dataset are the ones with SH usage alone, they are used to estimate the SH component of the NA points. The DHW usage in each NA point is calculated a posteriori through the difference between the total heat measurement from the smart meter and the estimated SH.

Based on these assumptions, several approaches have been developed to find the best procedure to separate and estimate the utilities' heating usage. In Fig. 1, one can see the research roadmap with the different studied approaches.

After the datasets are retrieved and pre-processed in step 1, different approaches to separate the data points are investigated in step 2. The energy separation stage identifies and labels all hours when the dwellers use DHW. In step 3, the points labeled as "not having DHW" (SH only) will be used to estimate the SH share of the points labeled with DHW and SH usage happening simultaneously. In step 4, the estimated values are compared to the actual separated measurements and the Danish DHW compliance calculations to test the methods' accuracy.

### 2.2. Dataset description and pre-processing

The dataset used in this study for validation is constituted of 28 apartments. All apartments are located in a social housing complex in Aalborg, Denmark. The complex was gradually renovated to the nearly Zero Energy Building (nZEB) standard from 2012 to 2020. The apartments included in this block were modernized in 2015. The concrete sandwich elements in the façade were replaced with insulated wooden cassettes with different façade cladding (i.e., brick wall, wood, or zinc). The roof construction was supported with new insulation. The heating, ventilation, and air conditioning (HVAC) installations were replaced with new ones. The interior of the apartments was fully renovated, and the new space heating installation includes radiators in all rooms and kitchens and underfloor heating in the bathrooms and hallways. The heat for SH and DHW is produced at the building block level and

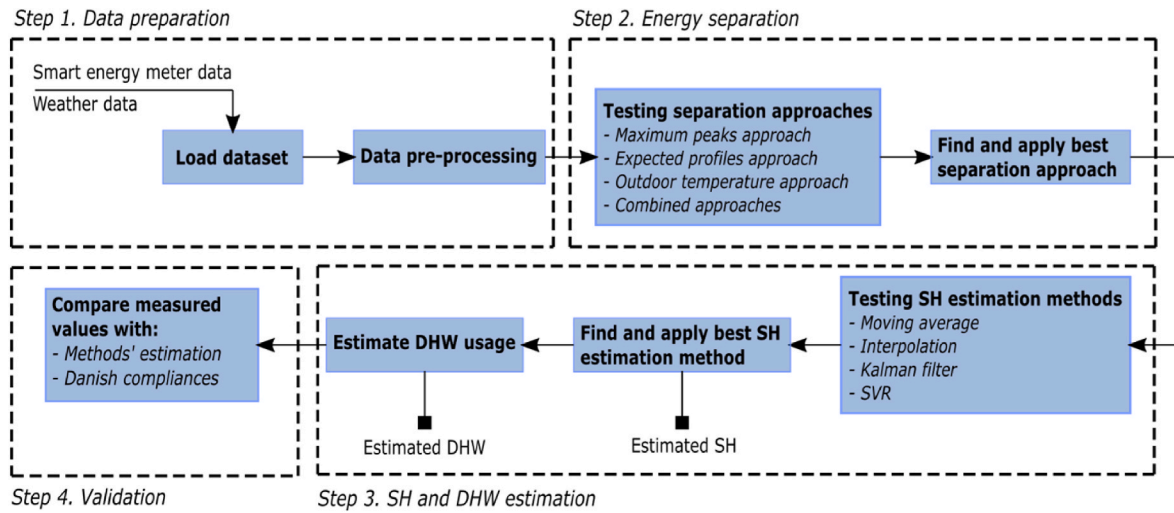


Fig. 1. Research roadmap.

distributed to each apartment. Apartments are equipped with individual SH and total heat demand meters (measuring SH and DHW without other appliances, e.g., electricity). The DHW is calculated through the difference between measurements from the meters. The floor area of the apartments is between 97 and 112 m<sup>2</sup>.

The local weather data is extracted from the Danish Meteorologic Institute (DMI) website. The outdoor temperature and the global radiation were the only variables extracted with 1-h time resolution. The selected weather station is Tylstrup, as it is the closest station from Aalborg available in the DMI database.

In this work, the data pre-processing consisted in detecting the number of missing and negative measurements and removing them. In the 28 apartments dataset (187 123 data points), with approximately nine months of monitoring for each dwelling, there are 46 661 missing hours (~25% of the dataset). The apartment with the lowest missing data has approximately 3% missing data. On the other hand, some apartments have up to 43% of missing data. Regarding negative energy usage measurements (erroneous values), there are few dwellings with those. In total, these values only represent 0.013% of the dataset. Therefore, all missing measurements and erroneous values were removed from the dataset before its analysis.

## 2.3. Energy separation

In Denmark, the SH system generally operates continuously during the heating season, while the DHW is only produced sporadically throughout the day. Thus, only a few hours of the day correspond to the majority of the DHW usage, whereas the other data points are SH usage alone. To estimate the SH and DHW usage, it is thus necessary to identify which hourly measurement data points correspond to DHW and SH use from those only comprising SH demand. To that matter, five new approaches to identify these points are developed and investigated in this paper. All these methods are tested against ground truth data from the 28-apartment dataset in section 3.

### 2.3.1. Maximum peaks approach

This method starts from the premise that the outdoor temperature has small fluctuations during the day, contributing to smooth SH demand variations throughout its continuous daily operation. Considering this assumption, all meters' significant peaks in the measured heat can be accounted for DHW usage. Therefore, the "maximum peaks" algorithm detects all daily highest data points ( $E_{Total}$ ) and considers them as comprising DHW production and SH ( $E_{Total} = E_{SH} + E_{DHW}$ ). If a data point is not one of the maximum values, it is considered only SH usage

( $E_{Total} = E_{SH}$ ). For each day, the method assumes the seven-highest measurements as DHW production, while the other 17 hourly data points are considered SH alone. It is also assumed a daily sleeping period from 1:00–4:00 h. Therefore, only SH operates during this period, and the high values are due to the low outdoor temperatures. In Fig. 2, one can see the algorithm's data flow diagram (a) and the representation of the method during a day for a single household (b).

After detecting all data points with DHW usage, they are converted into NA-values, and the household's dataset is updated with only SH measurements and the NA-values.

### 2.3.2. Expected profiles approach

This new method follows the same reasoning as the one used behind the "Maximum peaks approach". However, it is based on the hypothesis that weekdays have a certain regularity (i.e., routine) regarding the hot water usage pattern, as opposed to weekends. In this study, weekdays are considered from Monday to Thursday, while weekends are considered from Friday to Sunday. The reason for this division is that it is expected that a larger variation of hot domestic water usage occurs on Fridays afternoon and evening. This reasoning is also corroborated in [23], where Fridays were considered a different profile from the other weekdays. Therefore, from Monday to Thursday, the daily profile was separated into three groups – morning (5:00–11:00 h), afternoon (12:00–16:00 h), and evening (17:00–00:00 h). The highest value is found and considered as being "SH + DHW" in each time range. During the morning and evening periods, the adjacent hours (−1 and +1 h) of the peak heating usage are also identified as "SH + DHW". Concerning periods spanning from Fridays to Sundays, the "Maximum peaks approach" is used to detect the "SH + DHW" points because it is not likely to follow a routine. In Fig. 3, one can see the algorithm's data flow diagram (a) and the representation of the method during a day for a single household (b).

After detecting all data points with DHW usage, they are converted into NA-values, and the household's dataset is updated with only SH measurements and the NA-values.

### 2.3.3. Outdoor temperature approach

It is known that building SH needs have a strong negative linear correlation with outdoor temperature during the heating season [26]. However, if this trend with the outdoor temperature is not observable with the total energy measurements, it is due to significant DHW production events.

As illustrated in Fig. 4, this method starts by subsetting each household's dataset with only the total measured values from 1:00–4:00



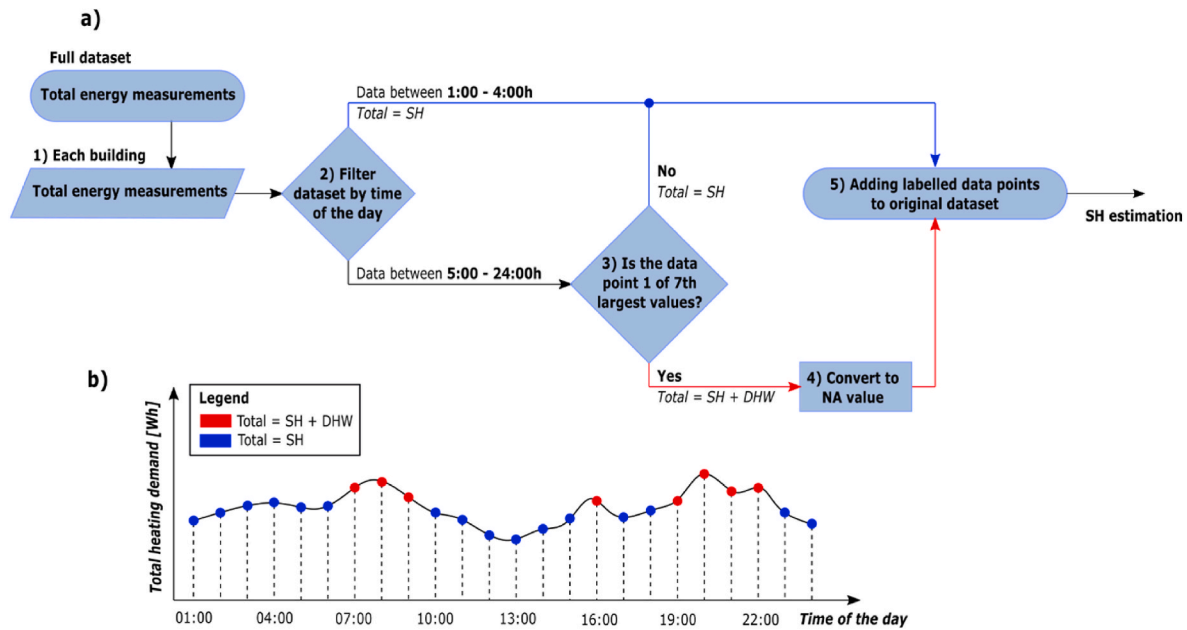


Fig. 2. a) Data flow diagram: Maximum peaks approach; b) Method representation.

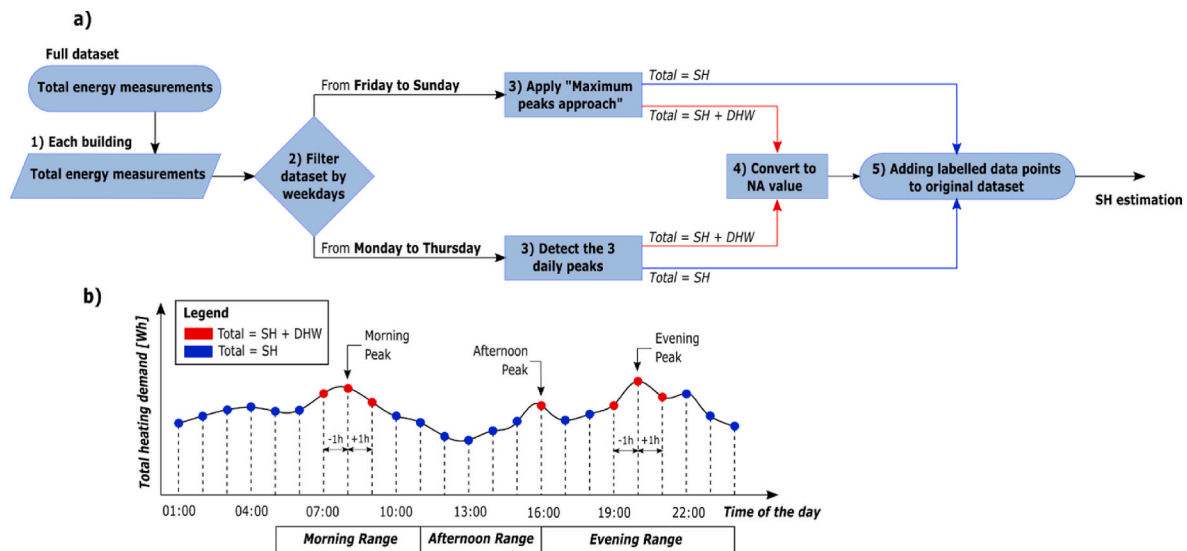


Fig. 3. a) Data flow diagram: Expected profiles approach; b) Method representation.

h (step 2). This time-conditioned subset is used because it is assumed that during this period, people are asleep. Thus, all total measurements must be SH usage only. With the subset, it is generated a piecewise linear regression [27] that estimates the SH demand as a function of the outdoor temperature for both the heating and no-heating seasons (commonly known as the “heat demand signature curve”). The junction between the negative linear trend of the heating season and the linear constant (horizontal) trend of the no-heating season forms the “change point temperature” (CPT) [28] (step 3).

In step 4, a prediction interval is developed for the two seasons. For the heating season, the interval’s tolerance is iteratively defined as 0.90, and for the no-heating period, a narrower tolerance of 0.60. By establishing the prediction intervals, the building’s dataset is divided into data points that are positioned above the intervals (step 5). If the measurement is below the interval, it follows the SH trend and therefore is SH usage ( $E_{\text{Total}} = E_{\text{SH}}$ ). If the value is outside the interval, the total energy equals SH and DHW simultaneously ( $E_{\text{Total}} = E_{\text{SH}} + E_{\text{DHW}}$ ). All

points with DHW usage are converted into NA-values in step 6. The last step is updating the building’s dataset with only SH measurements and the NA-values. In Fig. 4, one can see the approach’s representation.

#### 2.3.4. Combined approaches

The combined method merges the separation techniques described in subsections 2.3.1 and 2.3.3. Two different combined approaches were developed. “Combined method 1” only categorizes a data point as “SH + DHW” if both approaches, “Maximum peaks” and “Outdoor temperature”, together label the same point as “SH + DHW”. This method’s data flow diagram can be seen in Fig. 5a. The “Combined method 2” categorizes a data point according to its measured total heating usage. If the total energy of the datapoint is lower than 250 Wh or higher than 3,000 Wh, then the “Outdoor temperature approach” is used. If not, the “Maximum peaks approach” is used. These threshold values are established due to a preliminary investigation of the performance of the “Outdoor temperature” approach for one of the apartment’s data. This

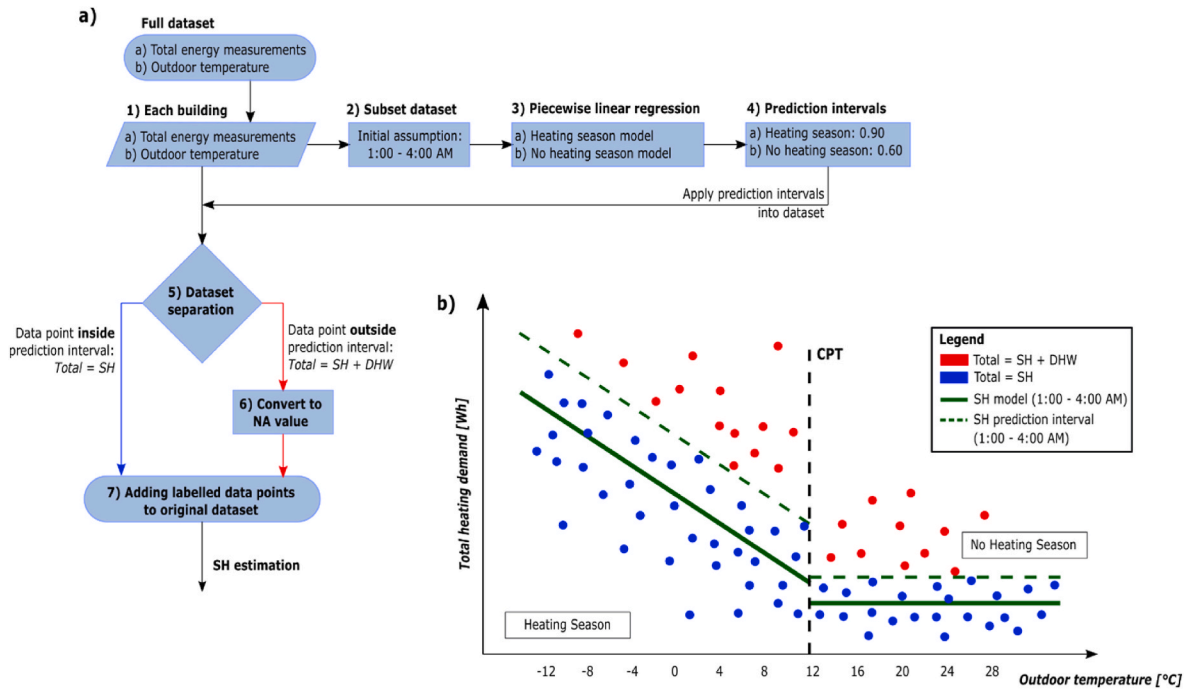


Fig. 4. a) Data flow diagram: Outdoor temperature approach; b) Method representation.

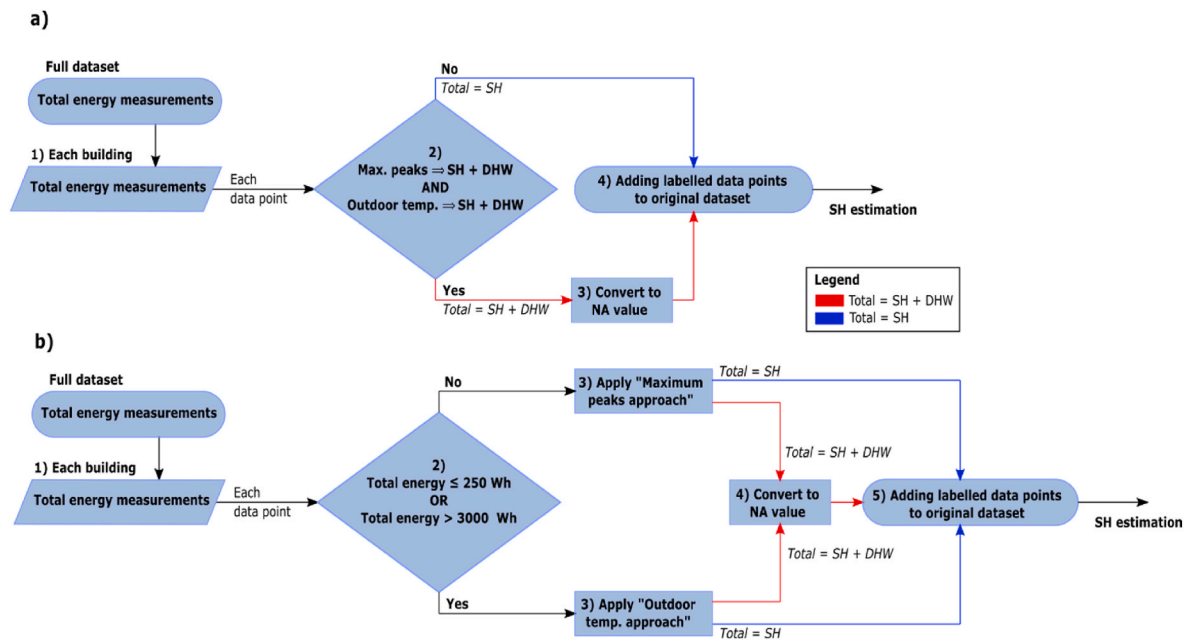


Fig. 5. a) Data flow diagram: Combined 1 approach; b) Data flow diagram: Combined 2 approach.

preliminary test showed that the “Outdoor temperature” method performed better for total heat data points below 250 Wh and above 3,000 Wh. It is advised for each building case to perform a preliminary calculation to establish these heating thresholds accurately because they might differ for each building. The second merged method is presented in Fig. 5b.

Like other approaches, after detecting all data points with DHW usage, they are converted into NA-values, and the household’s dataset is updated with only SH measurements and the NA-values.

## 2.4. Space heating and DHW estimation

At this stage, the DH dataset consists of NA-values and measurements that only quantify SH ( $E_{\text{Total}} = E_{\text{SH}}$ ). The next step is to estimate the SH usage ( $E_{\text{SH,estim}}$ ) in the NA-values by considering the known SH data points ( $E_{\text{SH}}$ ). After obtaining the  $E_{\text{SH,estim}}$  of the NA-points, the DHW usage is calculated as  $E_{\text{DHW,estim}} = E_{\text{Total}} - E_{\text{SH,estim}}$ . To calculate the SH demand in the missing points, several methods have been implemented and benchmarked hereafter.

### 2.4.1. Interpolation – Univariable estimator

Firstly, linear interpolation, cubic spline interpolation, and Stineman



interpolation are tested. The linear method calculates the NA-value(s) by assuming a linear relationship between its known neighboring points. To estimate the missing values, the cubic spline method fits a third-order polynomial between the known SH data points. The Stineman interpolation also applies a third-order polynomial into the time series; however, it preserves its monotonicity. These estimation algorithms are derived from the R-package *imputeTS* [29].

#### 2.4.2. Moving average – Univariable estimator

This method is one of the most commonly used in data analysis for smoothing time series. It consists in averaging the values with their neighboring points. The width of neighboring points used to calculate this average is designated as the “window”. This window-size variable, or range, must be set beforehand. A range equal to 2 ( $k = 2$ ) has been selected in this study, which means that any NA-value is estimated by averaging its two previous and two succeeding points. Different weight-averaging techniques are also tested. These techniques are the simple moving average, linear weighted moving average, and exponential weighted moving average.

When estimating the missing points, the simple moving average has equal-weighted neighboring points ( $\alpha = 1/4$ ). The linear weighted moving average follows an arithmetic progression by decreasing the weight of the furthest neighboring points. For a window-size of 2, the previous and succeeding points to the NA-value weigh 0.5 ( $\alpha = 1/2$ ), and the second previous (and succeeding) points weigh 0.33 ( $\alpha = 1/3$ ). For an exponential weighted moving average, the nearest points weigh 0.5 ( $\alpha = 1/2^1$ ), followed by a weight of 0.25 ( $\alpha = 1/2^2 = 1/4$ ). These estimation algorithms are derived from the R-package *imputeTS* [29].

#### 2.4.3. Kalman filtering – Univariable estimator

A Kalman filter is tested with four different model implementations: a structural time series model with and without smoothing and an ARIMA (Autoregressive Integrated Moving Average) model with and without smoothing.

The structural time series model is based on the function “StructTS”, which consists of a linear Gaussian state-space model for univariate time series. The ARIMA model is from the function “auto.arima”, which finds the best ARIMA model for each building’s time series. Both models are tested with and without smoothing. These estimation algorithms are derived from the R-package *imputeTS* [29].

#### 2.4.4. Support vector regression (SVR) – multivariable estimator

Contrary to the aforementioned methods, this estimation technique considers other inputs to calculate SH missing data points. The support vector regression (SVR) is a machine learning method that trains a model with the values labeled as “SH only”. The input data to estimate a given SH point is the outdoor temperature, the global solar radiation measured two and one hours prior, and the SH + DHW points (smart meter measurements) before and after the missing point. The SVR model uses a radial kernel function with the parameters C (cost) and  $\gamma$  (gamma) equal to 7 and 0.01, respectively. This estimation algorithm is derived from the R-package *e1071* [30,31].

#### 2.4.5. Combined Kalman filtering and SVR – Univariable/multivariable estimator

From preliminary results, the Kalman smoothing techniques are the best methods to predict space heating from the total heat use. However, as explained, these methods depend on the neighboring data points, which can also be missing in some cases (missing data gap larger than 1 h). To tackle the problem, this algorithm is refined to use the smoothed Kalman filter with the model “StructTS” only when the number of hours missing consecutively is equal to or below 2 ( $\text{Gap} \leq 2$ ). If the data gap is larger, the SVR is applied instead with the same parameters described above. These estimation algorithms are derived from the R-packages *imputeTS* and *e1071* [29–31]. One can see in Table 1 all the tested estimation methods and their parameters.

**Table 1**  
Estimation methods.

Method	Parameters	Type	Input	Reference
Interpolation	Linear Cubic spline Stineman	Univariable	$E_{Total}$	[29]
Moving average	Simple, $k = 2$ Linear, $k = 2$ Exponential, $k = 2$	Univariable	$E_{Total}$	[29]
Kalman filtering	Model: <i>StructTS</i> Smoothing: <i>True</i> Model: <i>StructTS</i> Smoothing: <i>False</i> Model: <i>auto.arima</i> Smoothing: <i>True</i> Model: <i>auto.arima</i> Smoothing: <i>False</i>	Univariable	$E_{Total}$	[29]
Support vector regression (SVR)	Kernel: <i>Radial</i> Cost: 7 Gamma: 0.01	Multivariable	$T_{out} [i-1, i-2]$ $Rad [i-1]$ $E_{Total} [i-1, i+1]$	[30,31]
Kalman filtering & SVR	<b>Gap <math>\leq 2</math> h</b> Model: <i>StructTS</i> Smoothing: <i>True</i> <b>Gap &gt; 2 h</b> Kernel: <i>Radial</i> Cost: 7 Gamma: 0.01	Multivariable	$T_{out} [i-1, i-2]$ $Rad [i-1]$ $E_{Total} [i-1, i+1]$	[29–31]

With the application of some of these methods, the estimated space heating ( $E_{SH,estim}$ ) can be negative or higher than the total energy measurements. Therefore, if  $E_{SH,estim}$  is negative, it is set to zero; and if  $E_{SH,estim}$  is larger than  $E_{Total}$  (SH + DHW – Smart meter’s measurements), it is set to  $E_{Total}$ .

#### 2.5. Methodology validation

To benchmark the accuracy of these different estimation methods, two different comparison metrics are computed: the normalized mean bias error (NMBE) and the coefficient of variation of the root mean square error (CVRMSE). These metrics are commonly used to assess numerical models’ performance (accuracy) in the energy and building systems field. They can evaluate the distance between the output time series of a numerical simulation and a reference time series [32,33]. The NMBE is given as a percentage (see Equation (1)) and measures the global bias of the estimation methods. If the value is negative, the method is globally underpredicting and overpredicting if positive.

$$NMBE = \frac{\sum_{i=1}^n (E_{SH,estim}[i] - E_{SH}[i])}{n} \times \frac{100}{E_{SH,max} - E_{SH,min}} \quad (1)$$

The CVRMSE is also given as a percentage and estimates the point-to-point difference between the measurements (ground truth) and estimated values (see Equation (2)).

$$CVRMSE = \sqrt{\frac{\sum_{i=1}^n (E_{SH,estim}[i] - E_{SH}[i])^2}{n}} \times \frac{100}{E_{SH}} \quad (2)$$

Where:

NMBE: Normalized mean bias error [%]

CVRMSE: Coefficient of variation of the root mean square error [%]

$E_{SH,estim}[i]$ : Estimated space heating [kWh]

$E_{SH}[i]$ : Measured space heating [kWh]

$E_{SH,max}$ : Maximum measured space heating in the dataset [kWh]

$E_{SH,min}$ : Minimum measured space heating in the dataset [kWh]

$\overline{E_{SH}}$ : Mean measured space heating in the dataset [kWh]

$n$ : Number of measurements in the dataset [–]

After selecting the best estimation method to obtain the SH demand using the above metrics (Equations (1) and (2)), the method is applied to calculate the SH in all apartments and predict the DHW need ( $E_{DHW,estim} = E_{Total} - E_{SH,estim}$ ). The DHW estimated demand is finally compared with the actual DHW measurements and the Danish compliance calculations to investigate if the developed methodology outperforms the current Danish calculations in predicting the DHW household needs.

In Denmark, the DHW consumption in households is currently predicted using the compliance calculation of 250 L/m<sup>2</sup> per year [34]. Similarly, the inlet water (cold) and outlet water (DHW) temperatures are considered to be 10 °C and 55 °C, respectively [35]. By knowing the area of the different apartments, the yearly energy usage for DHW production is calculated through Equation (3):

$$E_{DHW,compl} = \frac{1}{3600} \times 0.25A \times \rho_{water}c_{p,water} \times (T_{DHW} - T_{cold}) \quad (3)$$

Where:

$E_{DHW,compl}$ : Estimated DHW energy usage from Danish compliances [kWh/year]

0.25A: 0.25 m<sup>3</sup> water volume per m<sup>2</sup> of heated area per year [m<sup>3</sup>/year]

$\rho_{water}c_{p,water}$ : Water density per water-specific heat capacity – Constant value: 4177 [kJ/m<sup>3</sup>°C]

$T_{DHW}$ : DHW supply temperature from Danish standards – Constant value: 55 [°C]

$T_{cold}$ : Cold water supply temperature from Danish standards – Constant value: 10 [°C]

### 3. Results and discussion

This section presents the results of applying the different methods and their validation to find the best methodology. Moreover, the estimated DHW usage from the best methodology is compared with the current Danish compliance, which estimates the yearly DHW production.

#### 3.1. Energy demand separation

The five DHW separation methods presented herebefore are tested against measurements (ground truth) from 28 apartments in Denmark that have separated metering of SH and DHW energy usage. The validation consists in assessing the identification accuracy of the different approaches.

In Fig. 6, one can see the total percentage of incorrectly identified points in all apartments. This percentage is divided into total heating intervals (measured by the smart meters) to see if the methods perform better at different energy demand levels.

In Fig. 6, one can observe that “maximum peaks” and “combined 2” approaches are the best for categorization, with 20% incorrectly identified in all apartments. The method with the highest inaccuracy is the “outdoor temperature” approach, with a value of 27%. It is also seen that for different heating intervals, some approaches performed better than others. However, such differences are too small to conclude that the measured heating intensity affects the approach’s performance.

The quantity of correctly and incorrectly labeled (identified) points per approach was also analyzed without dividing by measured energy levels. The explanation of these attributed labels and how they affect the methodology are in Table 2, and its results are in Fig. 7.

One can conclude from the results presented in Fig. 7 that the methods “combined 2” and “maximum peaks” have a similar identification performance. However, the “maximum peaks” approach is preferred as the best separation algorithm from these results because the “combined 2” method is rooted in the “outdoor temperature” approach, which has the largest percentage of incorrectly identified points.

To conclude, the incorrect identified points percentage of each separation approach is calculated for each apartment. This analysis, in

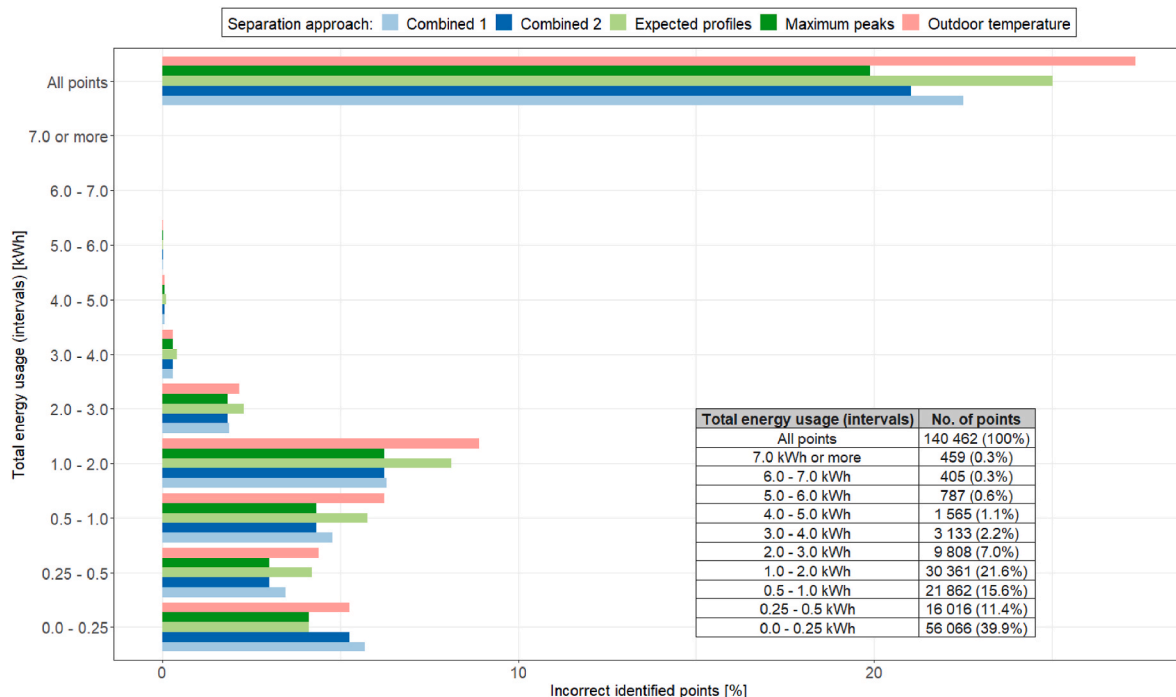


Fig. 6. Incorrectly identified points percentage in the overall dataset for each separation approach.

**Table 2**  
Labeling types.

Case	Correct label	Attributed label	Type	Explanation
$E_{DHW} \neq 0$	“SH + DHW”	“SH + DHW”	Correct	Points correctly identified and converted into NA-values to be estimated by the SH models
		“Only SH”	Incorrect	Incorrectly identified points that will be used to train the SH models – Will negatively affect the model's training
$E_{DHW} = 0$	“Only SH”	“SH + DHW”	Incorrect	Incorrectly identified points that will be converted into NAs and estimated by the models – Will affect the models' accuracy negatively
		“Only SH”	Correct	Correctly identified points that will be used to train the SH models

Fig. 8, is made to understand if the different apartments influence the overall performance of the different methods.

In Fig. 8, one can see the overall incorrect percentage of identified points (x-axis) for each apartment per separation method (y-axis/colored legend). From the figure, it is possible to observe the percentage distribution and extreme cases.

The results show that the different methods have their inaccuracy distributions between 10% and 40%, see Fig. 8. The “outdoor temperature” approach underperforms the most. One can observe that the best approaches are the “combined 2” and “maximum peaks”, with a slightly smaller difference in the mean value in the latter. Based on the analysis and application of the methods on the 28 apartments dataset, the preferred method to disaggregate the DH dataset is the “maximum peaks” approach.

### 3.2. Space heating and DHW estimation

After separating the data points, the following step estimates the SH based on data assumed to be only SH usage. Several methods are tested

to determine the most accurate one for this specific application. One can see in Table 3 the NMBE and CVRMSE calculated for each estimation method for the whole dataset (28 apartments).

The results show that most methods have similar values in both metrics, which means that methods differ slightly from each other. The worst-performing method is the cubic spline interpolation, indicating that cubic polynomial is not the best mathematical function to estimate space heating. The best method is the combined Kalman filtering and SVR according to both metrics.

In Fig. 9, one can see the overall error between the estimation and the measurements of SH and DHW of the different apartments. The overall error is calculated by comparing the difference between the aggregated measurements and estimated values during the measurement period.

As one can see from Fig. 9, the overall SH error (green color) is primarily negative (underestimated), with 18 apartments between -10% and 0%. Furthermore, the households with the extreme error values are one apartment with less than -15% error and another with almost +50% error (overpredicted).

Regarding the DHW prediction (blue color), the error distribution is wider than the space heating. In this case, five apartments have an overestimated DHW demand above +25%. The extreme DHW prediction is one household with an overestimation of +85% and four apartments with an underestimation slightly higher than -10%.

Several factors influence the method's estimations and the overall error. Foremost, the separation approach inaccurately identifies some of the points, influencing from the beginning, the estimation accuracy. Another factor is the presence of missing values in the initial dataset. As one can see in section 2.1, the dataset comprises about 25% of missing measurement points. Because the estimation relies on determining the SH demand based on its neighboring points, several missing measurements negatively impact the overall method's performance. Moreover, in section 2.1, it is shown that the retrieved weather data is not at the exact location where the dwellings are located. Besides this, the possibility of different heating systems, a large SH share, the unique dwellers' routines, or the DHW share being equal to zero (no occupancy) may influence the method's performance, and they might be the reasons behind the extreme cases.

The present research also compares the estimated DHW values with

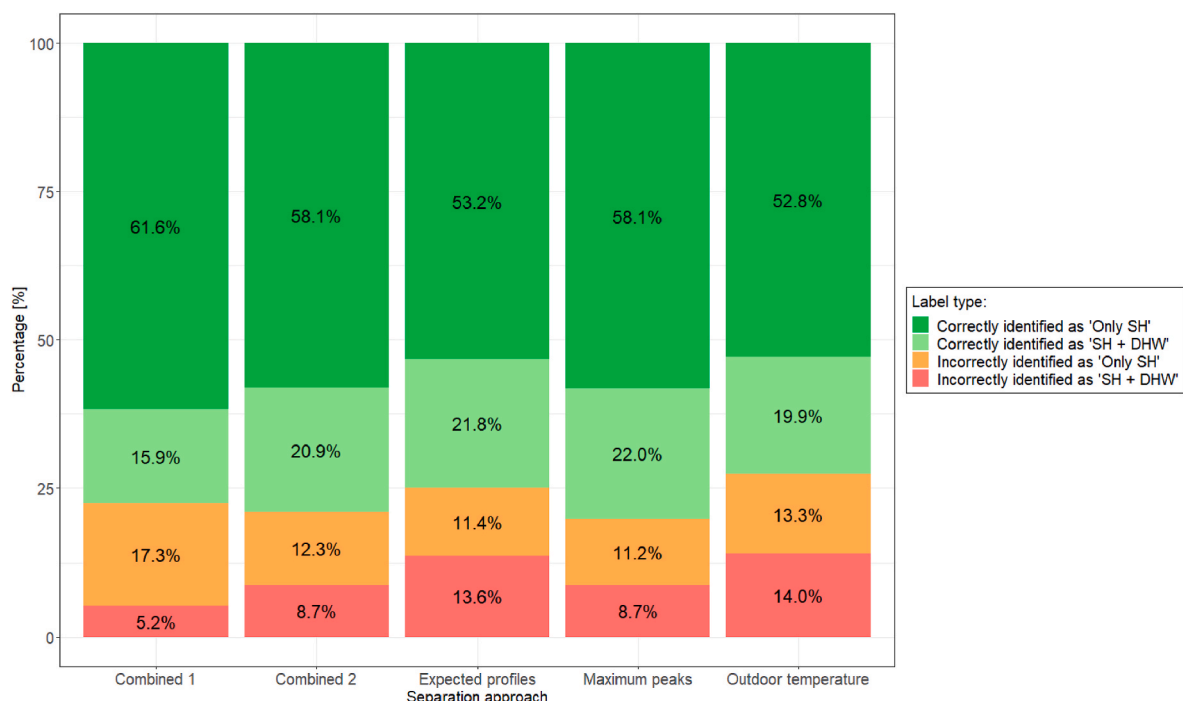


Fig. 7. Attributed labels percentage in the overall dataset for each separation approach.

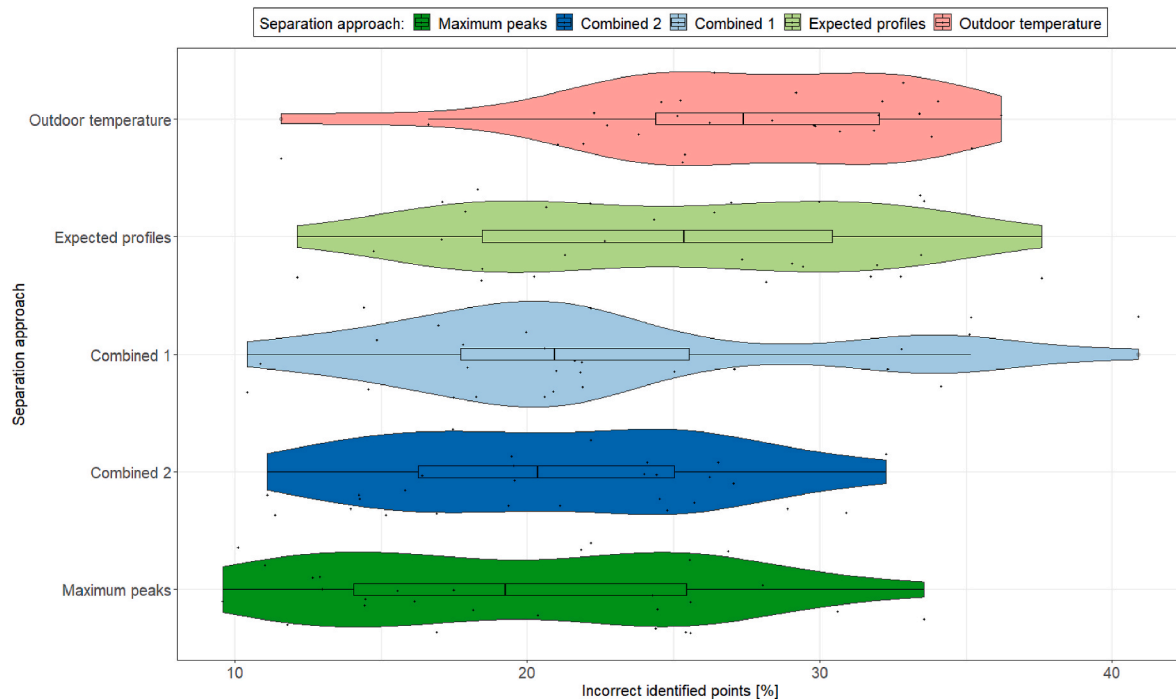


Fig. 8. Incorrectly identified percentage of separation approaches for each apartment (each point is one apartment).

Table 3

Each SH estimation method's NMBE and CVRMSE for all apartments.

Method	Type	Method specifications	NMBE	CVRMSE
Interpolation	Univariable	Linear	−0.25%	54.67%
		Cubic spline	−0.24%	59.60%
		Stineman	−0.25%	54.96%
Moving average	Univariable	Simple	−0.26%	54.62%
		Linear	−0.26%	54.60%
		Exponential	−0.26%	54.67%
Kalman filtering	Univariable	StructTS – Smoothed	−0.27%	53.86%
		StructTS – No smoothed	−0.28%	55.29%
		ARIMA – Smoothed	−0.25%	54.16%
		ARIMA – No smoothed	−0.24%	56.26%
		-	-	-
Support vector regression (SVR)	Multivariable	-	−0.18%	54.46%
Kalman filtering & SVR	Multivariable	StructTS – Smoothed	−0.10%	52.49%
		-	-	-

the Danish compliance calculation used to predict the annual DHW demand in households. The results of this comparison are in Table 4.

As shown in Table 4, there are three types of values per DHW usage. The actual DHW demand ( $E_{DHW}$ ), the compliance calculation of DHW demand used in Denmark ( $E_{DHW, comp}$ ), and the estimated DHW from the developed methodology ( $E_{DHW, estim}$ ). The “average” values are the aggregated DHW usage divided by the number of data points (hours). For the case of the DHW measurements and estimation, the number of data points is the number of measurement hours in each apartment. For the compliance case, the number of data points is the number of hours in a year. The “average” values are determined to be able to compare all three DHW usage types and calculate the error between the actual measurements and the compliance/estimation values. In most apartments, the developed methodology outperforms (bold values) the current Danish compliance calculations. Even though the disaggregation

method has a good performance in estimating the DHW usage for most apartments, there are few cases where the error is significant. The reason behind it might be due to numerous measurement hours missing in the initial dataset or due to the lack of dwellers in the households during the measurement period. However, from the results, it is argued that the method can be applied to predict the household's DHW energy use instead of what has been used to make the dwelling's energy assessment in Denmark. Also, it is clear that basing the Danish DHW compliance calculations only on the building area is imprecise; hence the research must shift towards the occupancy number and its behavior.

Even though some apartments have large SH and DHW estimation errors, this data-driven methodology is quite appreciable when considering its simplicity and the fact that no detailed building information, often unknown, is required (e.g., people habits description, system identification, building envelope characteristics, etc.). Another advantage of this method over some of the existing ones reported in the literature is the possibility of using hourly measurement data. Finally, the method outperforms the floor area-based compliance method currently used in Denmark for estimating DHW production.

#### 4. Conclusion

This article introduces a new data-driven methodology to estimate the SH and DHW from low-resolution heat meter data. The method's novelty is the possibility of applying it to hourly heating measurements without in-depth knowledge of the building and its occupants. The developed method is the combination of two algorithms to i) identify from the total heat measurements the points with DHW production and ii) estimate from the identified DHW usage points, the SH and DHW usage. This research tested several alternative methods for both algorithms to find the best point separation and energy estimation techniques. The different methods can be seen in Table 5:

The validation process shows that the best-performing method to detect when the DHW is being used is the “maximum peaks” approach, with a successful identification rate of approximately 80%. The best algorithm to estimate the SH demand in the identified points is the combination of SVR and Kalman filters (smoothed “StructTS” model). This estimation method has an NMBE of −0.10% and CVRMSE of

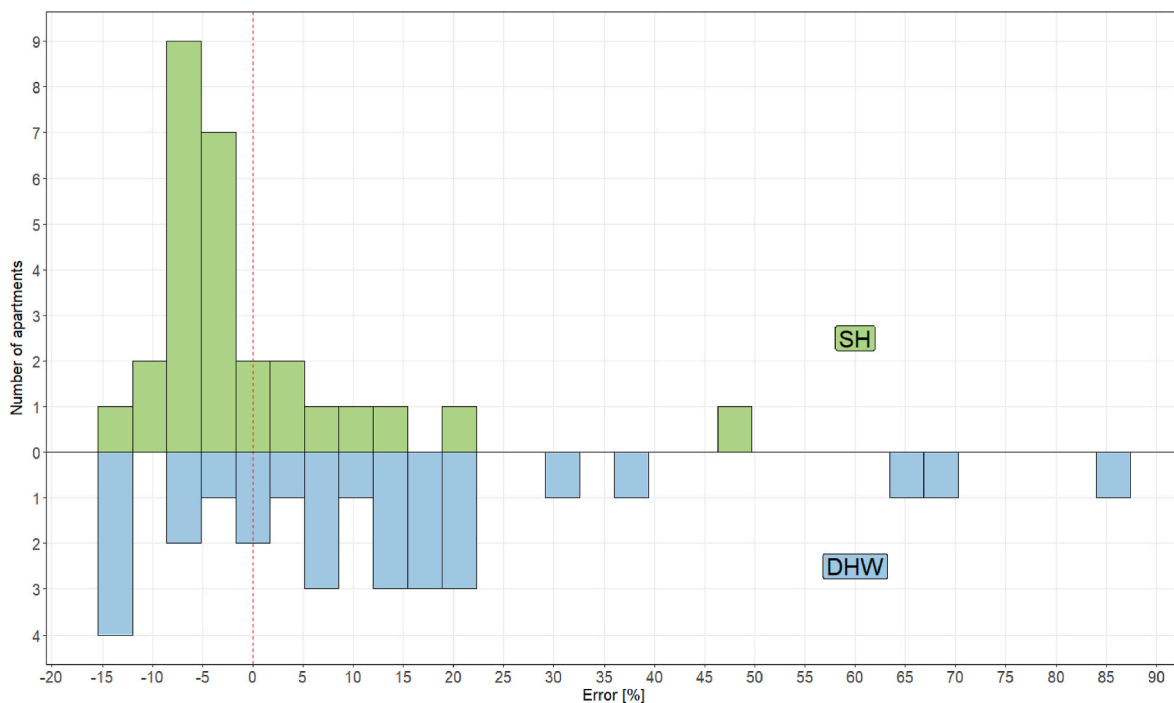


Fig. 9. Overall error of each apartment's SH and DHW estimation.

Table 4

Comparison between the Danish compliance values and the estimation results. The bold error values indicate the best performing method between the novel approach developed in this study and the Danish compliance calculations.

Apartment ID	Area [m <sup>2</sup> ]	$E_{DHW}$ [kWh/h]	$E_{DHW, compl}$ [kWh/h]	$E_{DHW, estim}$ [kWh/h]	Error between $E_{DHW}$ and $E_{DHW, compl}$	Error between $E_{DHW}$ and $E_{DHW, estim}$
666	112	0.314	0.167	0.315	−47%	<b>0%</b>
668	111	0.286	0.165	0.346	−42%	<b>21%</b>
669	110	0.184	0.164	0.224	−11%	22%
670	111	0.588	0.165	0.555	−72%	−6%
671	110	0.247	0.164	0.295	−34%	<b>20%</b>
697	111	0.692	0.165	0.606	−76%	−12%
698	111	0.674	0.165	0.627	−75%	−7%
699	110	0.678	0.164	0.588	−76%	−13%
700	111	0.074	0.165	0.137	123%	<b>85%</b>
701	111	0.167	0.165	0.196	−1%	18%
702	110	0.088	0.164	0.115	87%	<b>32%</b>
724	110	0.229	0.164	0.255	−28%	<b>11%</b>
726	111	0.116	0.165	0.132	43%	<b>14%</b>
727	111	0.103	0.165	0.121	61%	<b>18%</b>
728	110	0.148	0.164	0.203	<b>11%</b>	37%
729	110	0.144	0.164	0.161	14%	<b>12%</b>
730	111	0.388	0.165	0.406	−57%	5%
731	111	0.087	0.165	0.142	90%	<b>63%</b>
732	110	0.406	0.164	0.347	−60%	−15%
734	97	0.091	0.145	0.106	59%	<b>17%</b>
735	111	0.328	0.165	0.347	−50%	<b>6%</b>
736	111	0.336	0.165	0.34	−51%	<b>1%</b>
739	111	0.524	0.165	0.561	−68%	<b>7%</b>
740	111	0.164	0.165	0.159	<b>1%</b>	−3%
741	111	0.237	0.165	0.253	−30%	<b>7%</b>
742	97	0.145	0.145	0.167	−1%	15%
743	111	0.461	0.165	0.403	−64%	−13%
745	111	0.093	0.165	0.157	78%	<b>69%</b>

52.49%, being the lowest metric values of all tested SH estimation algorithms. Therefore the chosen overall method to disaggregate SH and DHW demand from the total heat measurements is the “maximum peaks” approach for identification purposes and the combined methods of SVR and Kalman filter to estimate SH needs.

Table 5

List of tested methods in this study.

i) Identification/separation methods	ii) SH estimation methods
Maximum peaks	Interpolation
Expected profiles	Moving average
Outdoor temperature	Kalman filter
Combined 1	Support vector regression (SVR)
Combined 2	Kalman filter & SVR (combined methods)

The overall methodology predicts the SH demand with an error between −10% and 10% for most dwellings. Concerning DHW estimation, the error is slightly wider, with most apartments falling between −15% and 15%. Moreover, this study compared the estimated DHW demand from the method with the actual measurements and the current Danish DHW compliance calculations. This comparison concludes that the developed methodology outperforms the Danish compliance calculations in most cases. Furthermore, it is argued that this disaggregation method can be applied to predict the household's SH and DHW energy shares. The authors also argue that estimating the DHW energy usage by relying solely on the building's area is erroneous (currently being done in Denmark and other European countries). Thus, future research efforts must move toward estimating the heating usage in buildings considering the dwellers' number and more specific building typology regarding DHW use (currently, in Denmark, only two are present: residential and other).

Finally, this data-driven method is simple to compute and understand, and if validated with more building cases and proved to be robust, it can be applied in the future by DH companies and energy auditors. Also, this methodology can be used without having additional detailed information about the building and its dwellers and can be used with 1-h resolution data, which is often the status of the buildings and their metering installations. The authors argue that this method is relevant to the energy and buildings field when considering these advantages, more specifically for the analysis of the energy performance gap, the DHW usage assessment (which has been overlooked until recent years), clustering of different SH usage patterns according to their systems and user's practices, and energy-efficiency decision-making.



## 5. Further work

A suggestion for further work is the application of this methodology with other datasets for further validation and robustness analysis. Preferably, datasets should come from various countries to ensure the methodology's robustness and applicability in different cases. This study used several algorithms to estimate space heating (e.g., SVR, moving average, etc.). However, this work can be further developed by investigating other estimation methodologies that can be found in the literature (e.g., neural networks, random forest regression, etc.).

It is also suggested to benchmark this novel methodology with other existing disaggregation methods on a common dataset. Furthermore, a more extensive effort must be made to collect good quality datasets – with hourly resolution (or higher) – of separated energy usage for space heating and domestic hot water in buildings with instantaneous hot water production systems.

## Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 893945 (E-DYCE). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 958345 (PRELUDE).

## Declaration of competing interest

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

## Data availability

The authors are unable or have chosen not to specify which data has been used.

## Acknowledgments

The authors would like to acknowledge the Himmerland housing association and Varmekontrol A/S for their support in data collection and to all partners involved in the E-DYCE and PRELUDE projects. Also, the first author would like to express his gratitude to his colleagues (Kamilla H. Andersen, Markus Schaffer, Martin Frandsen, Martin Veit, and Simon P. Melgaard) in the Ph.D. "Corner Office" for the valuable discussions and support.

## References

- [1] European Commission. *Energy performance of buildings directive* [Online]. 2022. Available: [https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive\\_en](https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en).
- [2] European Commission. *Heating and cooling* [Online]. 2022. Available: [https://energy.ec.europa.eu/topics/energy-efficiency/heating-and-cooling\\_en](https://energy.ec.europa.eu/topics/energy-efficiency/heating-and-cooling_en).
- [3] Danish Energy Agency. "Energy Statistics 2020," 2020. [Online]. Available: [https://ens.dk/sites/ens.dk/files/Statistik/energy\\_statistics\\_2020.pdf](https://ens.dk/sites/ens.dk/files/Statistik/energy_statistics_2020.pdf).
- [4] Werner S. International review of district heating and cooling. *Energy* 2017;137: 617–31. <https://doi.org/10.1016/j.energy.2017.04.045>. Oct. 15.
- [5] IRENA. *Renewable Energy Policies in a Time of Transition*. 953; 2018.
- [6] Paardekooper S, et al. *Heat roadmap Europe 4: quantifying the impact of low-carbon heating and cooling roadmaps*. Aalborg: Aalborg Universitetsforlag; 2018.
- [7] Lund H, et al. "4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems.," *Energy* 2014;68:1–11. <https://doi.org/10.1016/j.energy.2014.02.089>. Apr. 15.
- [8] Lund H, et al. Perspectives on fourth and fifth generation district heating. *energy* 2021;227. <https://doi.org/10.1016/j.energy.2021.120520>. 120520, Jul.
- [9] Dalla Rosa A, et al. General rights IEA DHC annex X report: toward 4th generation district heating experience and potential of low-temperature district heating. 2014.
- [10] Li H, Nord N. Transition to the 4th generation district heating - Possibilities, bottlenecks, and challenges. *Energy Proc Sep.* 2018;149:483–98. <https://doi.org/10.1016/j.egypro.2018.08.213>.
- [11] Volkova A, Mašatin V, Siirde A. Methodology for evaluating the transition process dynamics towards 4th generation district heating networks. *energy* May 2018;150: 253–61. <https://doi.org/10.1016/j.energy.2018.02.123>.
- [12] European Parliament, the Council. Directive (EU) 2018/2002 of the European parliament and of the council of 11 December 2018 amending directive 2012/27/EU on energy efficiency. *Off J Eur Union Dec.* 2018;328:210–30.
- [13] The Danish Government. *A Green and Sustainable World - the Danish Government's long-term strategy for global climate action*. Oct. 2020.
- [14] Energistyrelsen. *Danish climate policies* [Online]. 2022. Available: <https://ens.dk/en/our-responsibilities/energy-climate-politics/danish-climate-policies>.
- [15] European Commission. *Communication from the commission to the European parliament. Brussels: the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions — The European Green Deal*; 2019.
- [16] Do Carmo CMR, Christensen TH. Cluster analysis of residential heat load profiles and the role of technical and household characteristics. *Energy Build Aug.* 2016; 125:171–80. <https://doi.org/10.1016/j.enbuild.2016.04.079>.
- [17] Gram-Hanssen K. "New needs for better understanding of household's energy consumption - behaviour, lifestyle or practices?,". *Architect Eng Des Manag* 2014; 10(1–2):91–107. <https://doi.org/10.1080/17452007.2013.837251>. Apr.
- [18] Pomianowski MZ, Johra H, Marszal-Pomianowska A, Zhang C. Sustainable and energy-efficient domestic hot water systems: a review. *Renew Sustain Energy Rev* 2020;128(1):109900. <https://doi.org/10.1016/j.rser.2020.109900>.
- [19] Bacher P, de Saint-Aubain PA, Christiansen LE, Madsen H. Non-parametric method for separating domestic hot water heating spikes and space heating. *Energy Build Oct.* 2016;130:107–12. <https://doi.org/10.1016/j.enbuild.2016.08.037>.
- [20] Marszal-Pomianowska A, Zhang C, Pomianowski M, Heiselberg P, Gram-Hanssen K, Hansen AR. Simple methodology to estimate the mean hourly and the daily profiles of domestic hot water demand from hourly total heating readings. *Energy Build Feb.* 2019;184:53–64. <https://doi.org/10.1016/j.enbuild.2018.11.035>.
- [21] Lien S, Ivanko D, Sartori I. *Domestic hot water decomposition from measured total heat load in Norwegian buildings*. In: *BuildSIM-nordic 2020*. Oslo: Norway; 2020. p. 244–51.
- [22] Ivanko D, Sørensen ÅL, Nord N. "Splitting measurements of the total heat demand in a hotel into domestic hot water and space heating heat use,". *Energy* 2021;219: 119685. <https://doi.org/10.1016/j.energy.2020.119685>. Mar.
- [23] Hedegaard RE, Kristensen MH, Petersen S. "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* 2020;172:12001. <https://doi.org/10.1051/e3sconf/202017212001>.
- [24] Alzaatreh A, Mahdjoubi L, Gething B, Sierra F. Disaggregating high-resolution gas metering data using pattern recognition. *Energy Build Oct.* 2018;176:17–32. <https://doi.org/10.1016/j.enbuild.2018.07.011>.
- [25] RStudio. (2022, Mar. 25). RStudio [Online]. Available: <https://www.rstudio.com/products/rstudio/>.
- [26] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Appl Energy Feb.* 2019; 236:1078–88. <https://doi.org/10.1016/j.apenergy.2018.12.042>.
- [27] Crawley MJ. *The R book*. first ed. John Wiley and Sons Ltd; 2007.
- [28] Ulseth R, Lindberg KB, Georges L, Alonso MJ, Utne Å. "Measured load profiles and heat use for 'low energy buildings' with heat supply from district heating,". *Energy Proc* 2017;116:180–90. <https://doi.org/10.1016/j.egypro.2017.05.066>.
- [29] Moritz S, Bartz-Beielstein T. imputeTS: time series missing value imputation in R. *R J* 2017;9(1):207–18. <https://doi.org/10.32614/rj-2017-009>.
- [30] Meyer D. Support Vector Machines: The interface to libsvm in package e1071. 2022 [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/ijcnn.ps.gz>.
- [31] D. Meyer et al. (2022, Feb. 07). CRAN - Package e1071," CRAN - Package [Online]. Available: <https://cran.r-project.org/web/packages/e1071/index.html>.
- [32] Johra H, Mans M, Filonenko K, De Jaeger I, Saelens D, Tvedebrink T. Evaluating different metrics for inter-model comparison of urban-scale building energy simulation time series. In: *Proceedings of building simulation 2021: 17th conference of international building performance simulation association*. Belgium: Bruges; 2021.
- [33] O. Perpinan Lamigueiro. (2022, Feb. 21). tdr: Target Diagram [Online]. Available: <https://cran.r-project.org/web/packages/tdr/tdr.pdf>.
- [34] Aggerholm S, Skovgaard M. *SBI-ANVISNING 213: bygningers energibehov*. Copenhagen: SBI Forlag; 2018.
- [35] *Norm for vandinstallationer*. Dansk Standard – DS, 439; 2009. 2009.