Smart Energy 3 (2021) 100035

Contents lists available at ScienceDirect

Smart Energy

journal homepage: www.journals.elsevier.com/smart-energy

Using data from smart energy meters to gain knowledge about households connected to the district heating network: A Danish case

Daniel Leiria*, Hicham Johra, Anna Marszal-Pomianowska, Michal Zbigniew Pomianowski, Per Kvols Heiselberg

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

ARTICLE INFO

Article history:
Received 12 December 2020
Received in revised form
21 May 2021
Accepted 9 June 2021
Available online 12 June 2021

Keywords:
District heating
Smart energy meters
Big data mining
Linear regression analysis
Energy performance certificates
Building characterization

ABSTRACT

In Europe, one of the most sustainable solutions to supply heat to buildings is district heating. It has good acceptance in the Northern countries, a low-carbon footprint, and can easily integrate intermittent renewable energy sources when coupled to the electrical grid. Even though district heating is seen as a vital element for a sustainable future, it requires extensive planning and long-term investments. To increase the understanding of the district heating network performance and the demand-side dynamics of the connected buildings, several countries, including Denmark, have installed smart heat meters in different cities. In that context, this paper presents several methodologies to analyze the datasets from the smart heat meters installed in a small Danish town. The first method is concerning data curation to remove the anomalies and missing data points. The second method analyses measured variables (heat consumption, outdoor temperature, wind speed, and global radiation) to acquire new knowledge on the building characteristics. These results were compared with the values given by the energy performance certificates of a smaller sample of 41 households. Finally, to communicate and visualize the analysis outputs in a user-friendly way, an interactive web interface tool has been created.

1. Introduction

Denmark has very ambitious goals in terms of the sustainability of its energy sector. In 2030, the share of renewable energy sources (RES) shall be 55% in final energy usage, 100% in electricity use and 90% in the district heating (DH) sector [1]. The transition to a low-carbon future is built on two pillars: lowering energy demand and increasing intermittent RES usage, such as wind and solar power.

Buildings account for roughly 40% of all energy use in the European Union [2] and have, therefore, a prominent role in this transition towards sustainability. In addition to a general decrease of their energy demand, buildings should be operated in a smart way, i.e., modulate their energy demand according to the availability of local RES.

District Heating systems also play a vital role in shifting to a low carbon society as they provide most of the energy used for space heating and domestic hot water in Danish buildings (>60%). In other countries like Iceland, Poland, Lithuania, Estonia, Sweden, Finland, and Northern China, more than half of their building stock is connected to the DH grid [3]. In addition, DH systems can integrate a wide range of RES [4] and excess heat from local industrial processes. However, this requires a change in the DH network's operation by lowering supply and return temperatures and intelligent control at the building/customer level with heat demand-side management.

The digitalization of energy use in buildings brings new opportunities. However, smart technologies in residential buildings are expected to be adopted by 27% of EU households by 2025 [5], leaving 73% of building stock as conventional buildings equipped solely with energy meters lacking IoT (Internet of Things) enabled sensors and devices. Therefore, the lack of available technological infrastructure is a significant barrier that will prevent a substantial EU building stock share from being involved in the proactive and integral part of the evolving energy system. Moreover, current methods for assessing and optimizing building performance and evaluating smart readiness are based on simplified calculations [6,7]. Therefore, the difference between predicted and actual energy use of a building (the performance gap) can be up to a factor of 2.5 [8,9].

Currently, the real-time consumption data from DH smart heat

* Corresponding author.
E-mail address: dle@build.aau.dk (D. Leiria).

https://doi.org/10.1016/j.segy.2021.100035

meters is primarily used for billing the customers. In Denmark, it will be obligatory to collect dynamic heating data by using smart meters for every building connected to the DH grid from 2027 [10]. Therefore, this paper aims to present how to exploit these energy meters' large potential and how they can deliver new information on building connected to the DH grid: faulty operation of the ventilation/infiltration system, poor performance of the building envelope, etc. This information is crucial for CO_2-saving actions in energy-optimized buildings that smartly interact with the local DH system and facilitate its green transition.

These data will also provide a better knowledge of the actual building's energy use, which enables an accurate estimation and planning of the demand-side management for heating grids. By increasing the accuracy, it pushes forward the DH systems into the 4th Generation District Heating (4GDH). The 4GDH systems are the new evolutionary step of the DH networks, where the fluid-supply temperature is lowered (50–60 °C) to decrease the overall heat losses, reduce the distribution pipe diameters, lower the fluid flow rates and increase, consequently, the system's efficiency [11]. Even though the 4GDH systems bring several advantages, the system's implementation must be precise to assure the user's comfort. For this reason, it is argued that smart energy data will benefit the transition to the 4GDH systems because all data collected and its analysis will provide a better understanding of the grid, allowing a more reliable design and implementation of the 4GDH networks.

This paper's contributions are a methodology used to perform the cleansing and preparation of the DH dataset to be used for further analysis. Moreover, the treated dataset can be applied to more precisely assess the building's transmission losses, the ventilation and infiltration dependency on the energy consumption, and the impact of the solar radiation has on the building to decrease the heating demand. This study also proposes incorporating the methodology results to compare with the Energy Performance Certificates (EPC) calculations to evaluate their similarity and detect the key variables that might contribute to the energy performance gap.

After reviewing the current publications about data analysis of smart meters in DH systems in section 2, the current study's methodology is detailed in section 3. Section 4 presents and discusses the results from the application of the methods developed in this study. The article closes with conclusions and suggestions for future work.

2. State of the art

In a recent follow-up study of heat load profiles using smart heat data [12], Calikus et al. [13] presented an automated method to analyze heat load profiles for non-residential buildings. Since the latter have clear occupation schedules and heating management systems, their heat demand profiles are easier to predict compared to residential houses, in which occupants' behavior and heating practices are more diverse [14]. With the application of smart heat meters in the residential sector, the recent research work has focused on identifying typical customer segments according to heat consumption [15–17]. Johra et al. [18] have shown that clustering of buildings according to metered parameters, i.e., the temperature of return water to DH network (T_{return}) and the temperature difference (ΔT) between the supply and return fluid, can help to identify buildings with efficient heating systems.

Gianniou et al. [19] presented a simple methodology with uniform and steady-state assumptions about the occupants in all houses, i.e., heating practices, occupancy schedule, and single-zone modeling of the building. The smart heat data are used to derive the temperature setpoint and U-value of the building envelope. However, the number of occupants and their heating habits [14] and the number and use of appliances [20,21] are different and dynamic. Thus the proposed methodology needs to adjust for such dynamic boundary conditions.

Recently, studies have also shown that real-time heat data could facilitate the field of urban building energy modeling (UBEM) in the calibration of archetype building energy models [22], for modeling of demand-respond [23], and load forecasting [24]. The smart electricity meter data have proven to provide great environmental, social and economic benefits [25]. However, the research work on smart heat data is still in its infancy, and we have yet to discover the knowledge gain captured in real-time heat data to speed up the green transition of building stock and energy systems. Therefore, there is a great need for research on identifying novel methodologies to convert the promising potential of smart meters to transform conventional buildings into energy-optimized and smart buildings.

3. Methodology

The dataset provided by the DH company Aalborg Forsyning [26] consists of heat meter recordings from 1665 buildings (mostly single-family dwellings) located in a small town in the North of Jutland, Denmark. All the smart energy meters installed in the buildings measure the cumulative energy and fluid usage, the hourly-averaged supply and return fluid temperatures and the instantaneous measurements of fluid flowrate, supply and return temperatures. The devices also store the faults detected during operation. The space heating and heating for domestic hot water production are accounted together. The measurement time period spans from the October 1, 2018 until the October 7, 2019. The measurements are recorded every hour. Because of the scope of the research project supporting the current study, only the single-family residential buildings connected to the DH network were studied.

3.1. Data pre-processing

In this study, the methodology to pre-process the DH dataset is similar to the one described by Johra et al. [18]. This algorithm performs the cleansing and treatment of the data for further analysis. The flowchart in Fig. 1 details the different steps of the data treatment.

The first part of the algorithm organizes the dataset, performs resampling, and quantifies the number of anomalies detected by the smart energy meters and the possible data outliers. The second part of the algorithm determines the number of missing data points in the DH dataset as well as the gap sizes of the consecutive missing values for each building. Buildings with data gaps larger than nine consecutive hours or buildings with more than 10% of data missing are excluded from the rest of the analysis. Different imputation techniques are then applied to generate the remaining missing data points. In this case study, the most suitable imputation for cumulative values was found to be linear interpolation. For instantaneous values, it is the exponential weighted moving average with a window-size of 8 data points before and after the missing value. In the last part of the algorithm, a well-known statistical technique of outlier detection is applied. It considers values higher or lower than 1.5 of the variable's interquartile range as an outlier. This technique evaluates the maximum cumulative energy and fluid-volume consumption parameters and the maximum mean calculated power to identify the non-residential buildings. The latter can then be excluded from further analysis to fit the scope of the current study. As an output of this pre-processing algorithm, when applied to the dataset of 1665 buildings, only 969 buildings (58.2% of the original dataset) fulfill all the requirements to be used for further analysis.

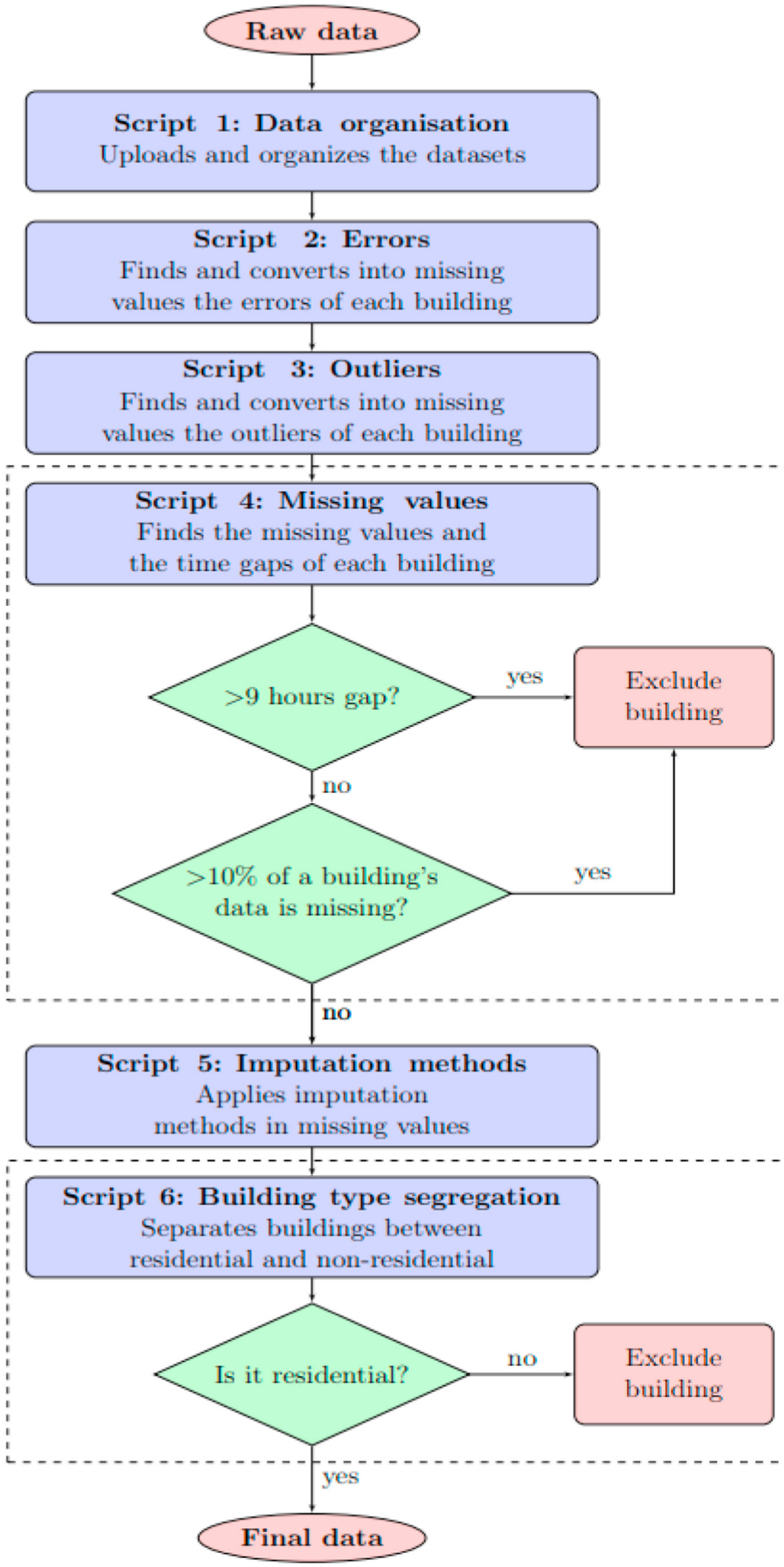

Fig. 1. Pre-processing algorithm.

The weather data (outdoor temperature, wind speed and solar radiation) from the local weather station is also integrated into the dataset.

In addition, information extracted from the EPC of each building is integrated into the dataset. This EPC reports the building energy use and production from which an energy label is issued. Label A represents the best energy performance, while the letter G is the worst energy performance grade that can be obtained [27]. Because EPC is only mandatory for buildings that are for sale or rent, EPC information could not be extracted for all buildings of the case study. For the available EPC reports, the extracted parameters are the construction year, the year of major renovation (if any), the heated surface area (m^2), the total specific heat losses from the opaque and glazed envelope (W/K), the solar exposure of the glazed elements, the volumetric flow rates of natural and mechanical ventilation for Winter and Summer seasons (litres/s.m^2), the description of the ventilation and the space heating systems, the estimated energy usage (kWh/m^2 year) and its associated energy label. From the dataset of 969 buildings, a subset of 41 buildings was selected as they present good quality EPC information.

3.2. District heating variables and coefficients

The recorded variables from the smart energy meters data are as follows: the cumulative energy usage (E_{cum}), cumulative fluid use (V_{cum}), the product between the cumulative water use and the hourly-averaged supply and return temperatures ($V_{cum}\overline{T}_s$ and $V_{cum}\overline{T}_r$). From these measurements, other variables are calculated, as described hereafter.

3.2.1. Hourly-averaged temperature and temperature difference

The hourly-averaged temperature recorded by the sensors is calculated back from (1):

$$\overline{T}_x = \frac{V_{cum}\overline{T}_x}{V_{cum}} \tag{1}$$

Where the variable $\overline{T}_x$ can be either the supply or the return fluid temperature. With the estimated hourly-averaged temperatures, the temperature difference is calculated as:

$$\overline{\Delta T} = \overline{T}_s - \overline{T}_r \tag{2}$$

3.2.2. Hourly energy usage and fluid volume use

The smart energy meters measure the building energy usage and fluid use by summing it up with the previous measurements. The current hourly values of energy usage and fluid flow are thus obtained by subtracting the previous data point from the current one:

$$E_i = E_{cum,\,i} - E_{cum,\,i-1} \tag{3}$$

$$V_i = V_{cum,i} - V_{cum,i-1} \tag{4}$$

3.2.3. Building thermal characteristics: heat transmittance, ventilation/infiltration and solar gains

The heating demand of a building is calculated from the steady-state energy balance between the heat losses through the envelope, ventilation and infiltration, and the heat gains from solar radiation and other internal loads (occupants, equipment, etc.):

$$E_{demand} = E_{trans} + E_{vent} - E_{solar} - E_{int} \tag{5}$$

The sum of heat losses by transmission, ventilation and infiltration can be expressed as a function of the temperature difference between the indoor and outdoor environments:

$$E_{trans} + E_{vent} = (UA + nc_p\rho)(T_{int} - T_{out}) \tag{6}$$

Where the U-value is the overall thermal transmittance of the building (W/m^2 K), A is the overall envelope area (m^2), n is the volumetric flow rate of the ventilation and infiltration (m^3/s), C_p is the specific heat capacity of the air at a constant-pressure system (J/kg K), and ρ is the air density (kg/m^3). By using the values calculated in equation (3), a correlation between heating demand and outdoor temperature is deduced for each building. Fig. 2 shows a scatter point plot from one of the buildings in the dataset where the heating season points (Autumn, Winter and Spring) are marked in red and the no-heating season (Summer) in blue.

However, one can see that there is no clear correlation between energy consumption and outdoor temperature beyond the high levels of heat demanded when the outdoor temperature is low. This unclear relationship between the variables might be due to the low resolution of the energy meter and the DHW energy share mixed with the space heating needs. To tackle that, the hourly energy data points were summed up for each day, and the outdoor conditions were averaged over the same period of time. By using this methodology, similar to the one used in [19], a more explicit relationship between the outdoor temperature and the energy demand can be observed (see Fig. 3).

The shape of the cloud of points in Fig. 3 is commonly called "hockey-stick" by utility companies. It represents the piecewise linear correlation between the heat demand and the outdoor temperature during the heating season and the constant energy values during the season without space heating needs and only domestic hot water production.

The linear relationship $y = mx + b$ between these two variables during the heating season can be linked to the steady-state energy balance of the building by combining equations (5) and (6):

$$E_{demand} = E(T_{out}) = -(UA + nc_p\rho)T_{out} + (UA + nc_p\rho)T_{int} - E_{solar} - E_{int} \tag{7}$$

Where the m- (slope of the line) is the term $-(UA + nc_p\rho)$ and the b- is the term $(UA + nc_p\rho)T_{int} - E_{solar} - E_{int}$:

Heating season linear regression:

$$E(T_{out}) = m_{Heating} \times T_{out} + E(T_{out} = 0) \tag{8}$$

No heating season linear regression:

$$E(T_{out}) = E\left(T_{out} = T_{No\ Influence}\right) \tag{9}$$

As seen, the m- is a descriptive value of the building characteristics, which is dependent on the ventilation/infiltration levels and the transmission losses. The value $T_{No\ Influence}$ is an outdoor temperature threshold under which the heating demand is no longer dependent on the outside temperatures. This may represent the outside temperature when the heating system is turned off. In order to quantify the parameters of (7), it is necessary to isolate the outdoor factors. The outdoor cause that influences the solar gains is solar radiation, and the infiltration/ventilation losses are mainly dependent on the wind speed. Therefore, solar radiation and wind speed are used as filtering conditions to isolate data points that are more conditioned on one energy component than the others. In order to reduce the impact of the internal gains and the potential

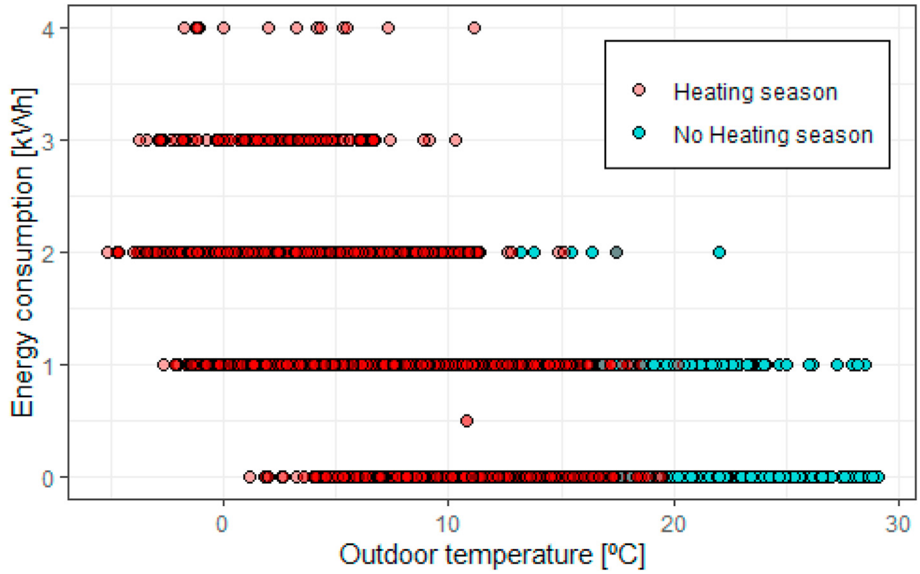

Fig. 2. Scatterplot between energy usage and outdoor temperature for a particular building.

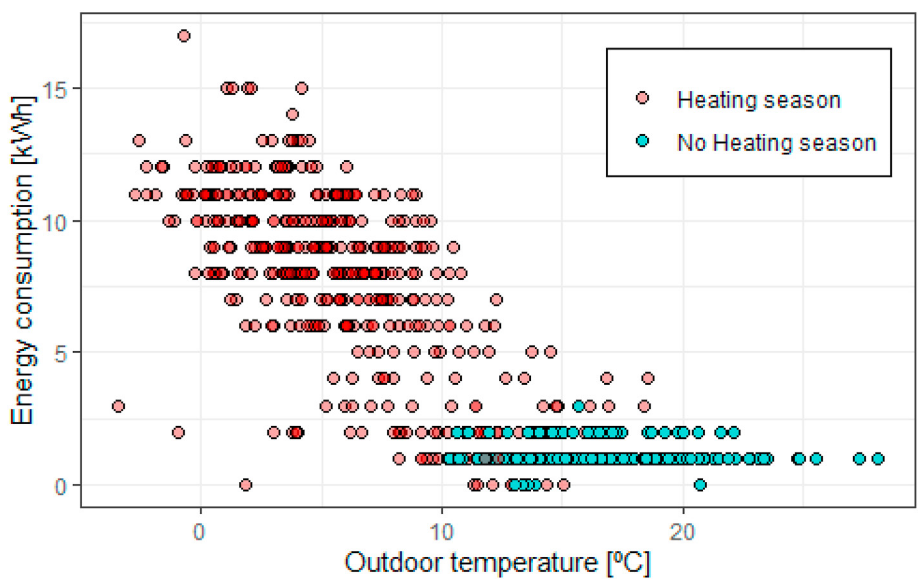

Fig. 3. Scatterplot between energy usage and outdoor temperature of a particular building with daily aggregated values.

natural ventilation from the opening of windows, the data points in the time intervals 11:00–16:00 and 23:00–04:00 were selected, as people are usually not at home or are sleeping. Fig. 4 presents the distribution of different weather parameters. One can see that the outdoor temperature and the wind speed have a quasi-normal distribution with a mean value of 6.4 °C and 3.6 m/s, respectively. In terms of solar radiation, the mean value is 103.5 W/m^2, with most of the values below 100 W/m^2, which is expected for Denmark.

The filtering conditions to be applied in each building of the dataset were selected by taking into account the weather variable's distributions (see Table 1).

The temperature/transmission component is isolated in the dataset when considering zero solar radiation (night period) and the wind velocity lower than 2 m/s. The same reasoning is followed for the ventilation/infiltration factor, but considering the data points where the wind speed is higher than 3 m/s. The solar gain component is filtered when the radiation is higher than 30 W/m^2 (daytime), and the wind velocity is lower than 2 m/s. The same methodology of linear regression is then reapplied on these new filtered subsets to obtain the parameters that are dependent on a specific energy component for the heating season:

Transmission losses condition:

$$E(T_{out}) = m_1(UA) \times T_{out} + b_1(UA) \tag{10}$$

Solar gains condition:

$$E(T_{out}) = m_2(UA) \times T_{out} + b_2(E_{solar}) \tag{11}$$

Ventilation and infiltration losses condition:

$$E(T_{out}) = m_3(n) \times T_{out} + b_3(n) \tag{12}$$

Equation (10) is the linear regression made from the data points

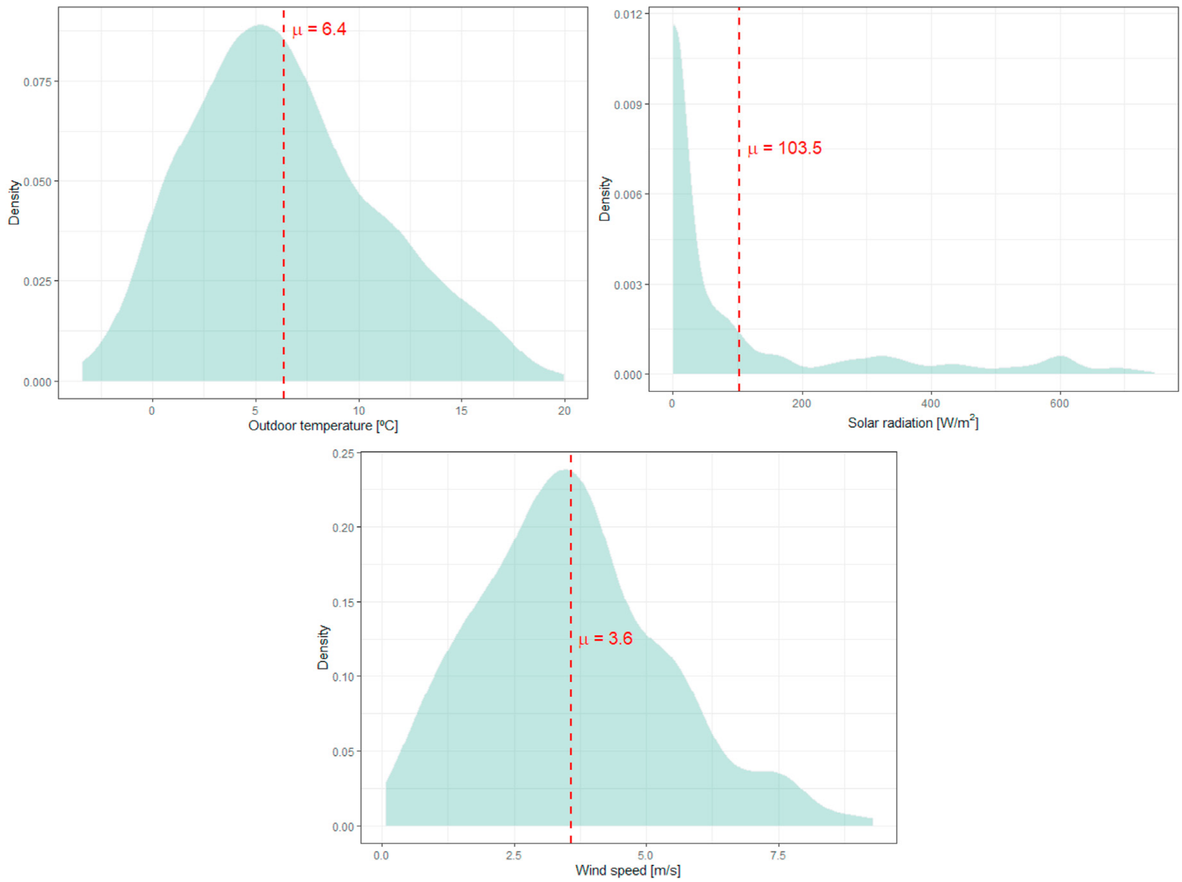

Fig. 4. Outdoor variables distribution.

Table 1
Subset filtering conditions.

Isolated energy component	Solar radiation ($\boldsymbol{R_d}$) [W/m^2]	Wind speed ($\boldsymbol{v_{wind}}$) [m/s]
Transmission losses (Outdoor temperature)	0	≤2
Solar gains (Solar radiation)	>30	≤2
Ventilation and infiltration losses (Wind speed)	0	>3

subset that is highly correlated with the outdoor temperature. Therefore the value m_1 is much more dependent on the constant *UA* than the air rate from the ventilation/infiltration component. In equation (11), b_2 is much more correlated with the solar gains than with the other energy parameters. In equation (12), by following the same reasoning, the slope m_3 characterizes the ventilation and infiltration impact on the building's heating demand. It is still important to highlight that these values are still influenced by the outdoor temperature, the building's air change rate, indoor temperature setpoint and internal gains. Regarding the values extracted from the equations, it is concluded that the higher the absolute value of m_1, the higher the heat transmission losses will be through the building's envelope. The b_2 has the solar gain as negative term; therefore, the lower this coefficient is, the higher the solar gains in the building will be. This happens in buildings with large glazing areas, low use of solar shading, or most glazing surfaces facing south. The coefficient m_3 is much more dependent on the energy losses due to wind speed. For this value, the higher its absolute value is, the higher the impact of wind on a building. This is seen in buildings with high infiltration/ventilation rates due to their ventilation system, or where the windows are opened frequently with the heating system in operation or due to high air leakage.

3.3. Comparison of the district heating coefficients with the EPC results

After extracting all DH variables from the dataset and calculating the linear regression coefficients, the derived building characteristics are compared with the building information reported in the EPC. These inputs are the information collected by the EPC certifier during the assessment and are considered reference values to evaluate the accuracy of the coefficients. The inputs are valuable to the utility companies when analyzing the building heating data because they are more detailed than what is available in the Danish building and housing register.

At this research stage, all EPC information is extracted manually to be analyzed. Therefore, a subset of 41 buildings was used. Among the subset, five buildings had significant renovations. All the buildings have a similar heated surface area, with a mean and standard deviation of 137.1 ± 32.1 m^2. The main ventilation system is natural ventilation. Few of them also have mechanical ventilation systems installed. Apart from two cases with indirect connection, the space heating system of the buildings is directly connected to the thermal grid without any intermediate heat exchanger. The domestic hot water is produced in a heat exchanger, often located

Table 2
Number of buildings with the same EPC label.

EPC label	Number of buildings
A2010	2
B	16
C	8
D	10
E	3
F	1
G	1
Total	41

in the utility room. Each building is assigned an energy label, which represents its energy performance level. One can see in Table 2 the number of buildings for each energy label.

The first variable that will be compared is the difference between the total measured energy usage by the smart meters and the annual estimated energy by the EPCs (kWh/m^2 year):

$$\Delta E = E_{EPC} - E_{DH} \tag{13}$$

The EPC annual energy, E_{EPC}, is the predicted household's energy usage concerning space heating, DHW and the electricity consumption by the buildings' installed systems. Because it is only being studied single-family houses, the majority of the E_{EPC}-value is the sum between space heating and DHW, as it is in the smart meter's measurements.

In equation (13), if the value ΔE is negative, then the EPC underestimated the building's energy usage and vice-versa. If the EPCs are in good agreement with the smart meter recordings, it might be argued that the former can be used by the utility companies when designing the expansion of their network.

The main parameters that are extracted and calculated from the linear regressions are m_1, b_2 and m_3. The estimate of these values can be compared against the input parameters from the EPC. For the case of m_1, its dependency is with the transmission losses of a building. From the EPC, the value used to test m_1 is the total specific heat loss from the opaque and glazed elements on the building. The parameter b_2 quantifies the dependency of solar radiation on the overall energy usage. Therefore the coefficient was compared with the heat gain share from the heat balance of all windows in a building. The heat gain share in the energy balance is dependent on the window's area, orientation, inclination and total solar energy transmittance (g_w), which are all described in the EPC. The ventilation and infiltration losses (m_3) are compared with the ventilation heat transfer coefficient provided by the TABULA WebTool [28].

4. Results

4.1. Comparison between district heating smart meter results and building EPCs

4.1.1. Smart energy meters measurements and EPC energy predictions

In Fig. 5, one can see the relationship between the EPC estimated energy consumption and the energy usage measured by the smart energy meters. The data points in the red upper zone indicate an

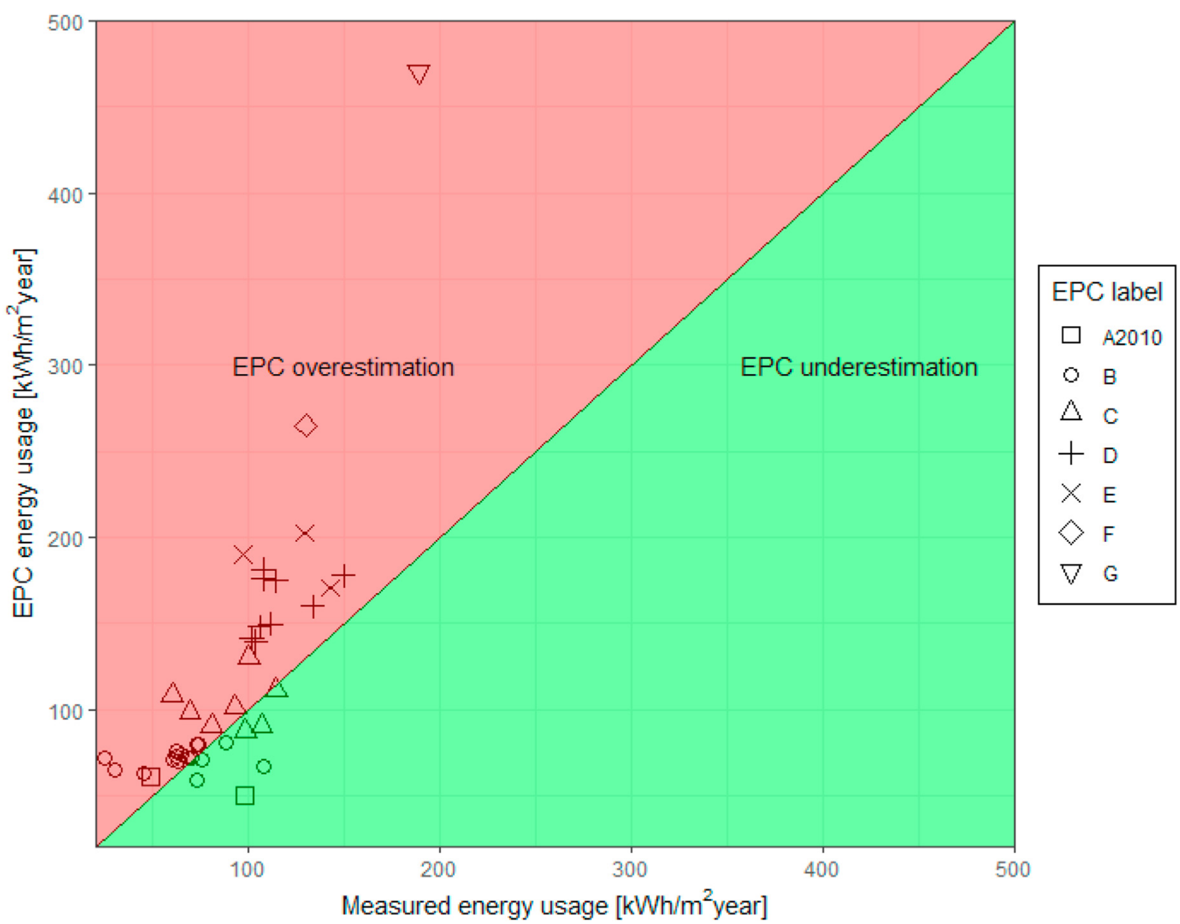

Fig. 5. Relationship between EPC energy prediction and the measured building's consumption.

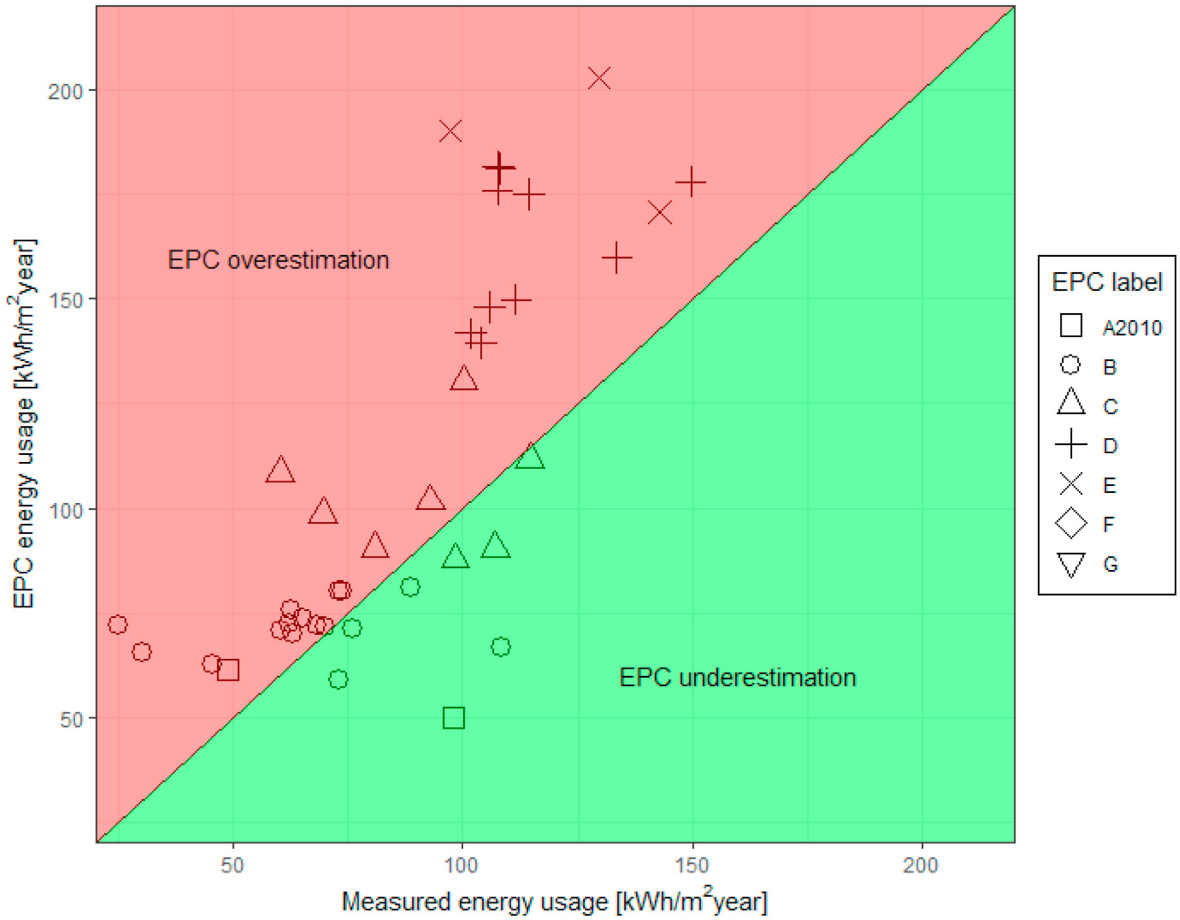

Fig. 6. Relationship between EPC energy prediction and the measured building's consumption (Zoom).

overestimation of the energy use by the EPC. Conversely, the green zone's data points indicate an underestimation of the energy use by the EPC.

Fig. 6 shows the same plot as in Fig. 5 but focuses on the lower range of energy use.

The figure above shows a significant mismatch between the EPC estimation and the actual heating need of many buildings throughout all energy label categories. Building energy performance from the EPC can thus be a problem if used by the DH utility companies when planning the extension of their thermal grid. Estimating the heat demand from the EPC of existing buildings connected to the DH network could lead to a large oversizing of the latter. Unfortunately, the size of the analyzed sample is too small to make definite conclusions.

4.1.2. Smart energy meters measurements and EPC building thermal characteristics

Regarding the building's thermal characteristics, a methodology was developed to quantify the envelope and ventilation/infiltration heat losses and heat gains from solar radiation. These characteristics can be assessed in a more accurate manner than what is stated in EPCs. It is thus possible to identify the main reason behind the high energy demand of certain buildings. Building system faults could thus be detected. A renovation scheme could also be suggested to the households.

In Fig. 7, the correlation between the specific transmission heat losses from the EPC and the m_1 coefficient is presented.

The specific transmission losses from the EPC are the product between the total area of the opaque and glazed envelope elements and their thermal transmittance. There is a good correlation between the EPC values and the m_1 coefficients obtained by linear regression on the smart meter data. This implies that this methodology is suitable to understand the building's transmission heat losses from the smart energy meters. Hence, it can be used as an indicator to the utility companies to identify the buildings where their space heating usage is highly dependent on the envelope heat losses.

Regarding the solar gains, the b_2-coefficient was calculated. This value is connected with the impact that solar radiation has on the building's heat demand. And the lower this variable is, the higher the solar gains are in a building. In Fig. 8, three of the EPC label categories with more buildings in the subset are represented, with each building's b_2-coefficient and their associated overall solar exposure.

In the plot, each point represents a building, and it is expected that the buildings with more prominent south solar exposure have lower b_2 (green data points). However, this relation is not observed. In label B, there are some buildings with low b_2 coefficients exposed mainly to the north. As for the label C and D buildings, some south solar-exposed buildings have large b_2-values. The reason behind it might be that the building's linear regressions are not adequate and that other unknowns significantly impact the building's energy performance.

Concerning the ventilation and infiltration losses, they are estimated by using the m_3 coefficient. To evaluate this variable's relevance, the correlation between the m_3 value and the expected heat transfer coefficient by ventilation in the buildings was estimated. Moreover, the assessors have sometimes described the

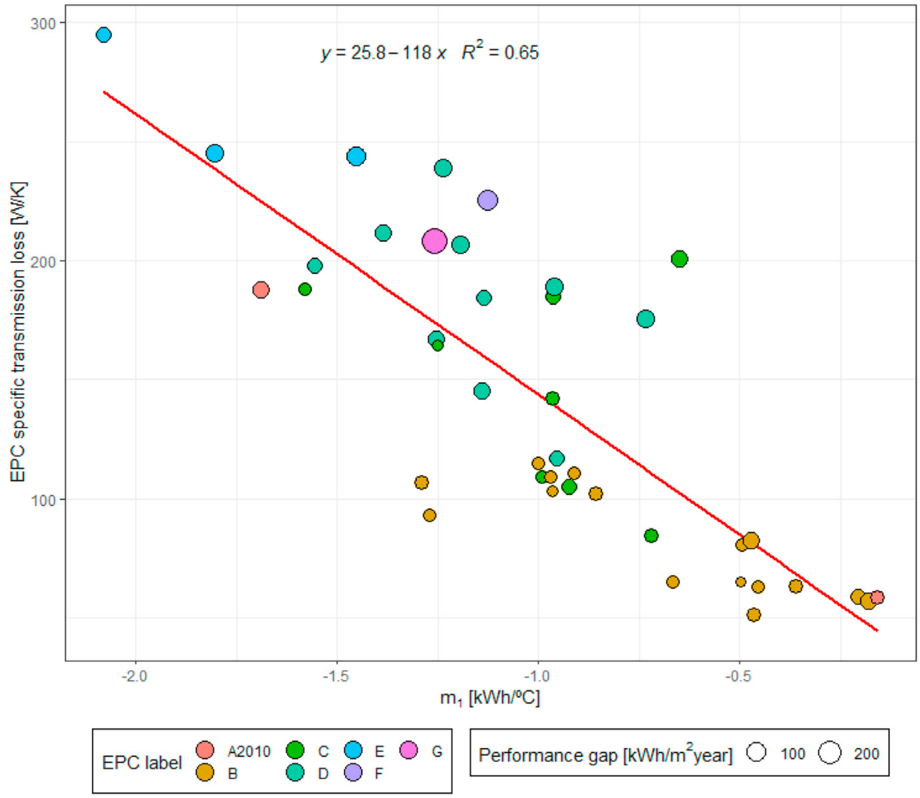

Fig. 7. Correlation between EPC specific transmission losses and the m_1-coefficient.

condition of the building joints (construction joints, joints at windows and door openings) to assess the building's leakage. Therefore the condition of the building joints was also compared to the m_3 coefficient (see Fig. 9).

In the scatterplot, one can observe a reasonably good correlation between what are the expected losses and the m_3 values. Therefore this method might be adequate to identify buildings with high ventilation and infiltration losses. Another point worth mentioning is regarding the building (Label A2010) with the lowest m_3 coefficient (-2.19 kWh/°C). Even though it is newly built and with a good EPC label, it presents high ventilation and infiltration losses. In the EPC assessment, this is one of the few buildings with mechanical ventilation, which might be the reason for high ventilation/infiltration losses and, therefore, a low m_3 coefficient. This also shows that the utility companies can also detect the households with large heat demand due to ventilation and advice their customers to take actions to reduce their consumption. Concerning the joints condition, buildings with low m_3 coefficients are expected to be less air tight. In the boxplot, a large portion of the building did not have a description (Not defined). The "Reasonable" conditions have low m_3 values, which is expected. The description "Not good" is only two buildings where its m_3 should be much lower than it is. Therefore, by observing the plot, the sample is too small to draw meaningful conclusions regarding the buildings' envelope airtightness.

Even though this methodology is quite promising due to its simplicity, it might not perform very well, in some cases, when compared with the EPC values. Several reasons might explain the poor performance of this method. The energy consumption measured by the devices is for space heating and DHW production. Therefore, it is hard to isolate the energy required only for space heating, which depends on the outdoor conditions. Also, the method used in this study tried to evaluate the different energy components on the building's heat balance by filtering the data points that have adequate outdoor conditions to eliminate certain terms of the heat balance. However, this filtering might not be perfect or not minimize enough certain gain terms in the energy balance. Additional unknowns might also have a significant impact that cannot be captured by the energy meters, EPC or weather station, e.g., people behavior, natural ventilation through windows openings, internal gains. Furthermore, in the filtering conditions, another problem that might cause the methodology not to work correctly is the studied location's meteorological characteristics. This study was performed in Aalborg, which is known for being a windy city with low sun exposure; therefore, the data points will not be equally distributed, creating unrealistic linear regressions without any physical meaning, as seen in Fig. 10.

Because there are not that many daylight hours during the heating season in Aalborg, it might be the main reason for the b_2 coefficient to contradict the EPC results. So, this coefficient might be more accurate for countries where the daylight hours are much higher than Denmark. The methodology accuracy was also tested when compared with the values from the EPCs. The EPCs are highly dependent on the assessor's knowledge and inputs as well as the building standard values provided in the national building standards. However, the standards and EPC inputs may not be the actual building's values, contradicting the calculated coefficients consequently. Another reason worthy of mentioning is the small EPC sample size (only 41 buildings).

4.2. Interactive web-based interface for data visualization - shiny

The methodology presented above is of great interest for households, urban planners, municipalities and utility companies

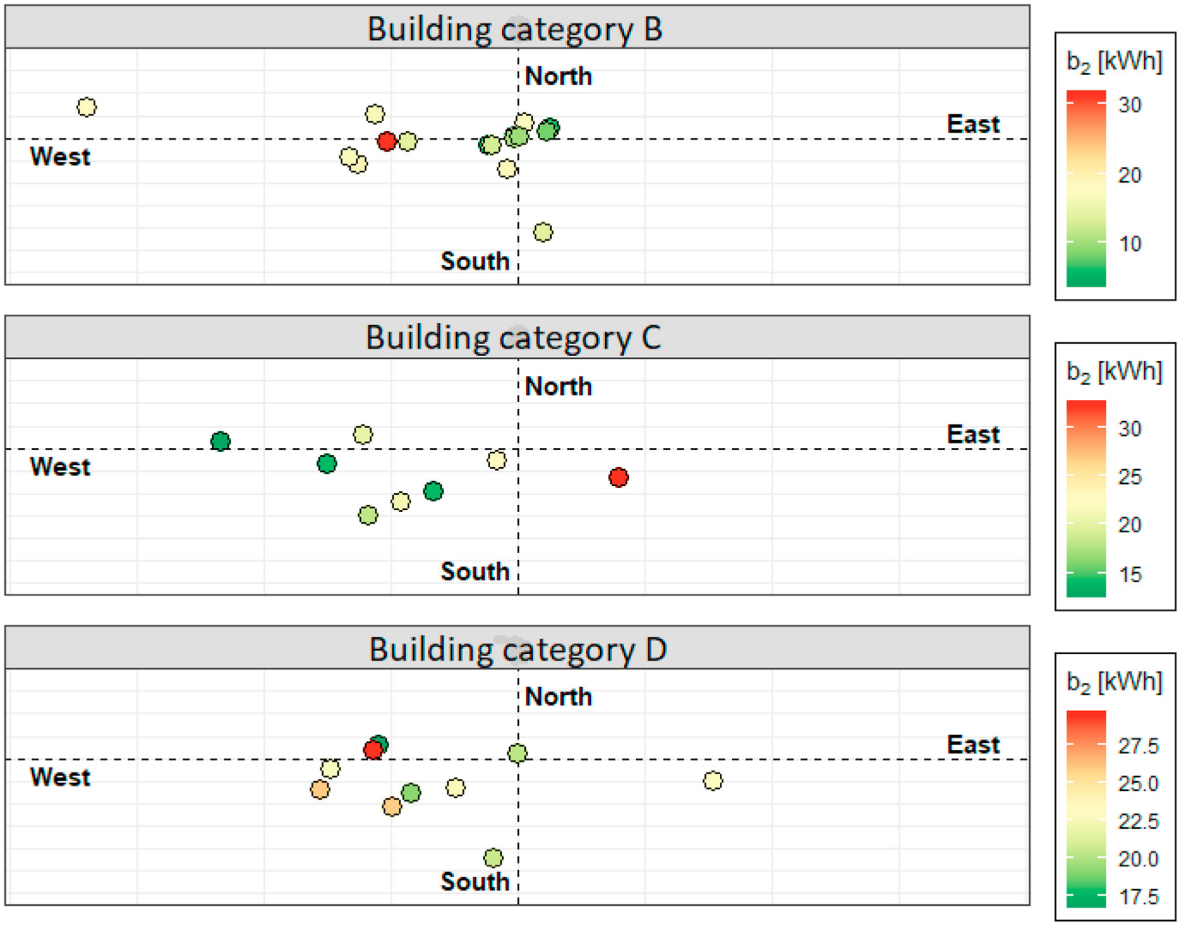

Fig. 8. Relation between EPC overall solar exposure and the b_2 coefficient.

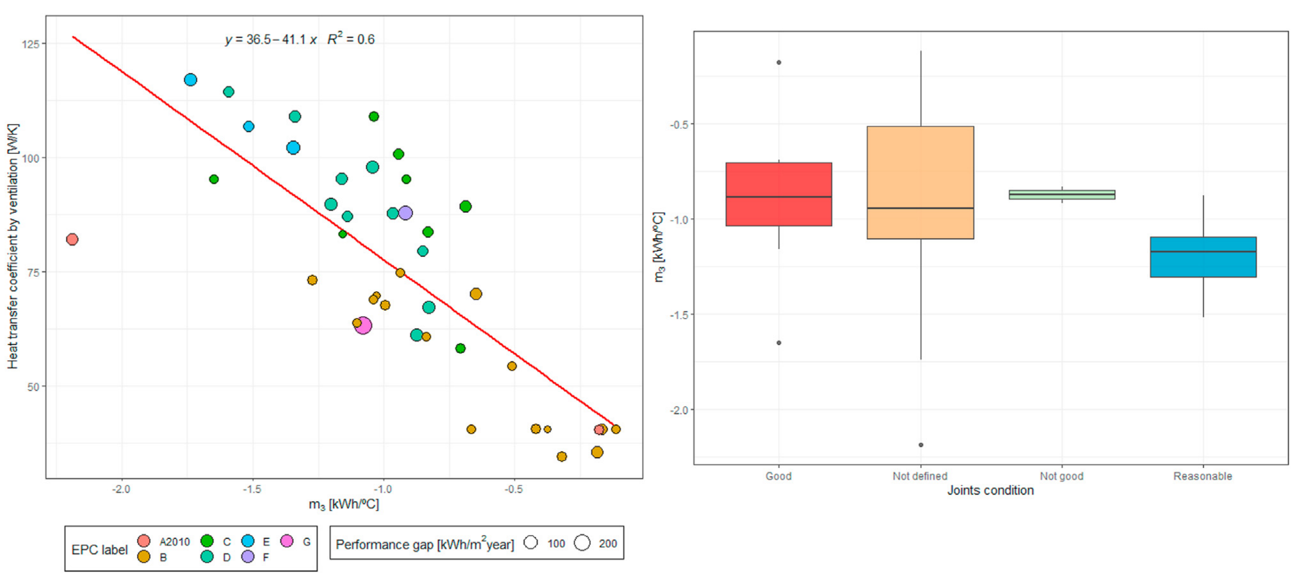

Fig. 9. Correlation between EPC specific ventilation loss and the m_3 coefficient. Relation between the building's joints condition and the m_3 coefficient.

managing energy distribution grids like district heating networks. However, sharing so much information with building professionals in a concise yet meaningful and flexible way can be challenging. Given this, it was chosen to bundle all the data analysis results of this study into a web-based interactive map. Maps are intuitive tools to rapidly grasp an overview of the state and characteristics of a given cluster of buildings or an entire city.

The "Shiny" package [29] is a free library for the R programming environment that enables the simple development of web-based graphical interfaces to display data plots and navigable satellite

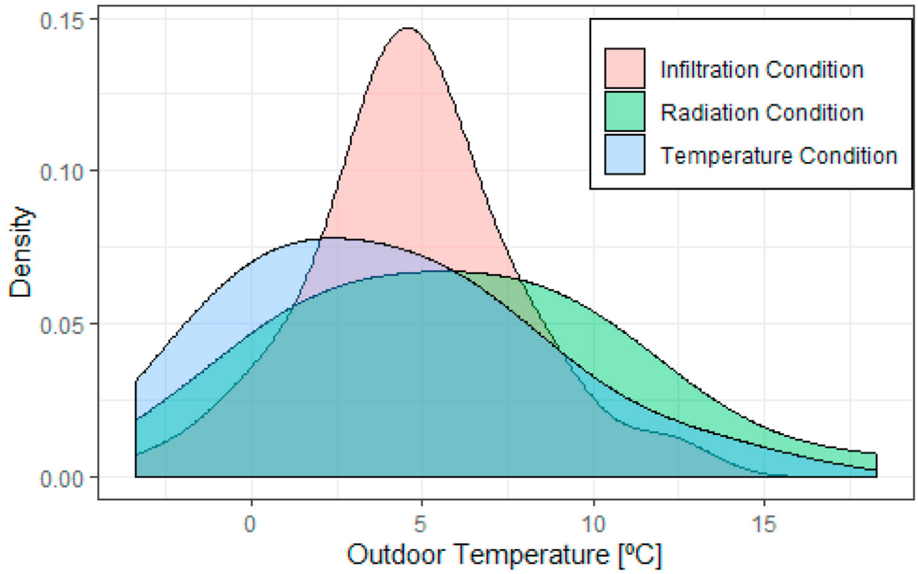

Fig. 10. Filtered data points distribution.

maps with super-imposed graphics, symbols, and data points (bubble maps). A dedicated data visualization interface has thus been created with the "Shiny" package to present the processed smart energy meter data to a larger audience.

In Fig. 11, one can see an overview of the web-based interface. The user can navigate on the map and display overlaying colored points (corresponding to a color scale) for the following building characteristics: district heating fluid supply temperature, the temperature difference between supply and return fluid of district heating, yearly volume of fluid passing through the substation of the building, yearly heating demand per m^2, clustering categories, amount of erroneous data from the smart meter, and amount of missing data from the smart meter. The user can filter the visible building data points on the map by selecting the filtering range corresponding to the aforementioned parameters.

It is also possible to include all the other building characteristics calculated during the data processing or extracted from the national building register. However, for clarity, these have not been included here.

In the interface, the user can select a specific point on the map and display a summary of the building characteristics (yearly values) together with the address, year of construction and energy label. The user can also open the data time series of the selected case and browse through it by selecting a specific period of time. Furthermore, display correlation plots between different measurement parameters in a selected building pop-up window.

Finally, a parallel plot (see Fig. 12) can be generated to give an overview of all analysis results and building parameters of the different households. Each line passing through the different parameter columns is a unique building case. The user can narrow

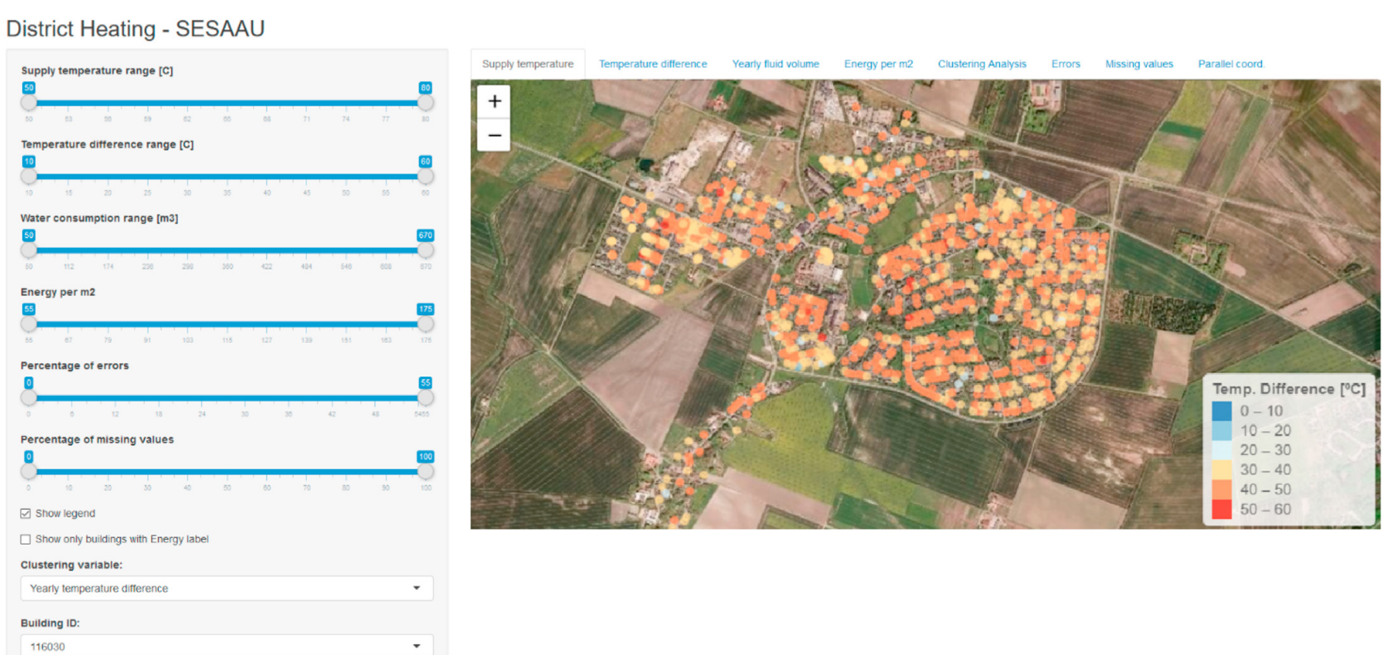

Fig. 11. Overview of the graphical user interface to visualize this study's processed data: district heating temperature difference to buildings in a specific region.

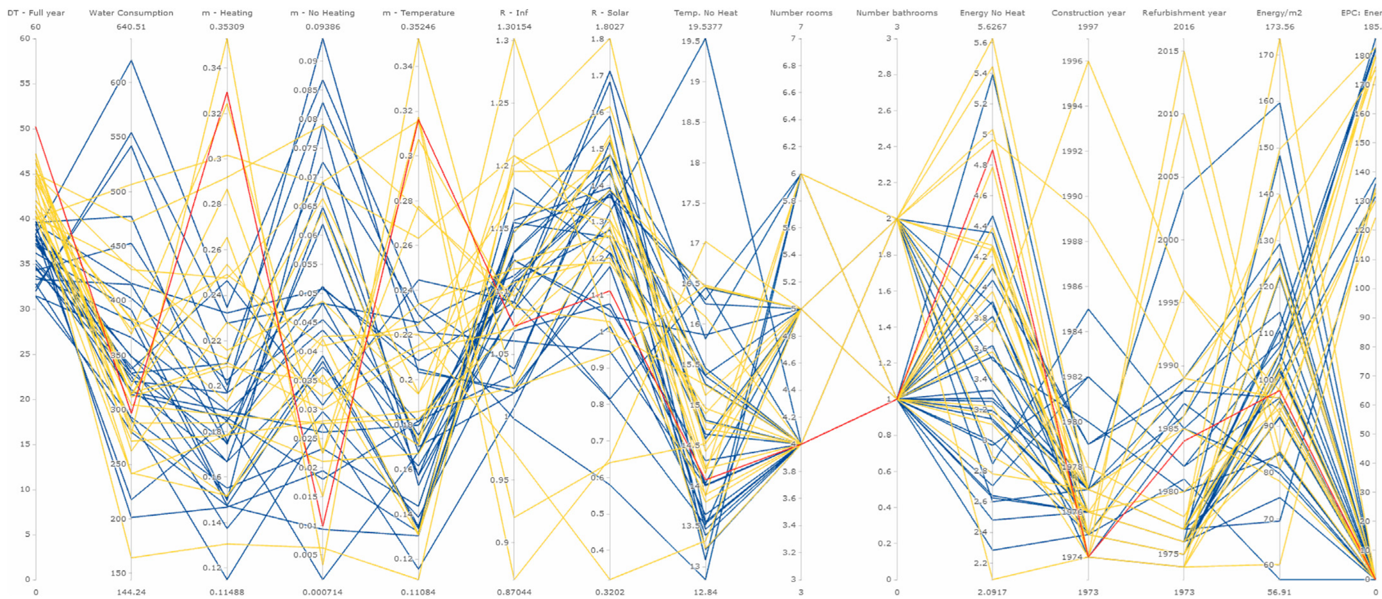

Fig. 12. Parallel plot of the building characteristics for some of the buildings in the case study.

down each parameter's range to identify and compare the other characteristics of the remaining building cases. The parameter and characteristic columns can be manually swapped around by the user.

5. Discussion

The systematic implementation of smart meters in the district heating network opened up for the utilities to learn more about their customers. The potential of this action is significant and should be beneficial for both the production and demand-side (i.e., utilities and customers). The digitalization of the demand side is also foreseen as the key component of the district heating transition towards 4th generation systems (i.e., 4GDH).

The presented methodology shows how utilities can identify customers with high energy use and determine the reasons for their performance. DH utilities can thus provide their customers with tailored-suited energy-efficiency actions, thereby lowering the network energy demand more efficiently, allowing a successful transition towards 4GDH.

Another pillar of the 4GDH concept is the high share of fluctuating renewable energy sources (e.g., wind and sun) on the production side. The methodology provides utilities with knowledge on which users are expected to use more or less heating depending on weather conditions like wind speed and solar radiation. Commonly the energy use at the demand side is correlated only with the outdoor temperature. Neither solar radiation nor wind conditions are taken into consideration when sizing the production mix. With a higher share of intermitted RES in 4GDH, the knowledge on the expected energy use must be better foreseen.

The expansion of hourly data from smart heat meters delivers great potential to learn more about buildings. In Figs. 11 and 12, one can see the easiness of identifying the building characteristics, the data outliers, and the corresponding clients. For example, a district heating company can map the clients with a high percentage of missing data or erroneous data from their smart meters and send a technician to verify and repair the latter. The utilities would also be very interested in identifying clients with a very low fluid-temperature difference because they lessen the district heating system's energy efficiency.

6. Conclusion

In this study, a simple methodology was used to treat and analyze the data recorded by smart energy meters installed in 1665 buildings connected to a district heating network. Regarding the different variables measured by the devices, all of them can be used to evaluate the building's energy performance concerning the different building characteristics, systems and user's behavior. To evaluate the validity of the methodology, its results were compared with the information from energy performance certificates of a smaller sample (41 buildings).

The methodology developed in this study aimed to assess several household characteristics from the smart meter data analysis: the actual heating season in the building and the influence of the outdoor temperature, wind speed, and solar radiation on the energy usage. When compared to EPC information (considered as reference), the simple linear regression method gives positive results for the outdoor temperature and wind speed influence. However, it is not conclusive for the sensibility of solar radiation. For the Danish case, the outdoor temperature and the wind speed influence can be assessed by the utility companies to understand the source of significant heat losses in the buildings connected to the grid. Even in the present research, it was seen that a highly efficient building had significant ventilation and infiltration losses, and therefore a large performance gap, most likely due to their ventilation system operation. To assess the solar gain, it is expected that the methodology might work in countries with more considerable daylight hours.

In the paper, it was also compared the smart energy meters measurements with the EPC estimations. The small sample used shows that the difference between energy estimations and measurements increases for buildings labeled as low-energy efficient. In terms of District Heating systems, the EPC inputs, even though used to test the accuracy of the linear regression method, cannot be used as guidelines for design and planning the creation or expansion of the DH networks.

In this study, a simple data visualization interface created with the "Shiny" package (R) is presented as a starting point for creating a real tool that can be used for the utility and consulting companies to analyze the energy meter data and detect possible problems occurring at their customers.

In the present paper, several parameters increased the uncertainty associated with this simple methodology. It is required to adjust this methodology with more DH data points, more buildings with EPC information, and higher resolution DH data with indoor measurements that will clarify which values are more accurate, the linear regression coefficients or the EPC inputs. As suggestions for further work, it is idealized to use other linear regression algorithms that are less susceptible to the different outdoor conditions distribution and outliers and apply grey-box models with system identification techniques.

Author contributions

Daniel Leiria: Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing - original draft; Writing - review & editing. **Hicham Johra:** Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing - original draft; Writing - review & editing. **Anna Marszal-Pomianowska:** Methodology; Project administration; Resources; Supervision; Validation; Writing - original draft; Writing - review & editing. **Michal Zbigniew Pomianowski:** Project administration; Supervision; Writing - review & editing. **Per Kvols Heiselberg:** Funding acquisition; Supervision; Writing - review & editing.

All authors have read and agreed to the published version of the manuscript.

Funding

This work was part of the E-DYCE project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 983945. This work was also supported by the InterHUB project funded by Aalborg University's strategic funding for interdisciplinary research.

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.

Acknowledgments

The authors would like to thank Aalborg Forsyning that contributed by providing the data from smart energy meters. And to Flávio do Vale Wichman for contributing with the data from the building's EPCs.

References

[1] Energi- Forsynings- og Klimaministeriet. “Denmark's integrated national energy and climate plan under the regulation OF the EUROPEAN parliament and OF the council on the governance of the energy union and climate action. 2019.
[2] Energy performance of buildings directive. https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en. [Accessed 25 March 2021]. accessed.
[3] Top district heating countries – euroheat & power 2015 survey analysis | euroheat & power. https://www.euroheat.org/news/district-energy-in-the-news/top-district-heating-countries-euroheat-power-2015-survey-analysis/. [Accessed 25 March 2021]. accessed.
[4] Lund H, Möller B, Mathiesen BV, Dyrelund A. The role of district heating in future renewable energy systems. Energy Mar. 2010;35(3):1381–90. https://doi.org/10.1016/j.energy.2009.11.023.
[5] Statista.com. Smart Home penetration rate per segment in Europe_ 2023 _ Statista. https://www-statista-com.zorac.aub.aau.dk/forecasts/887721/smart-home-penetration-rate-per-segment-in-europe. [Accessed 25 March 2021]. accessed.
[6] Smart readiness indicator for buildings | smart readiness indicator for buildings. https://smartreadinessindicator.eu/. [Accessed 25 March 2021]. accessed.
[7] Bsi and Cen, “Bs En 15603. Energy performance of buildings. Overall energy use and definition of energy ratings. Accessed European Standard 2008:1–45 [Online]. Available: https://standards.iteh.ai/catalog/standards/cen/7a0df579-c84c-4223-a270-23785e7e3f9f/en-15603-2008. [Accessed 19 May 2021].
[8] Molin A, Rohdin P, Moshfegh B. Investigation of energy performance of newly built low-energy buildings in Sweden. Energy Build Oct. 2011;43(10): 2822–31. https://doi.org/10.1016/j.enbuild.2011.06.041.
[9] Gram-Hanssen K, Georg S, Christiansen ET, Heiselberg PK. How building regulations ignore the use of buildings, what that means for energy consumption and what to do about it. Jun. 2017. p. 2095–104. Accessed: Mar. 25, 2021. [Online]. Available: https://vbn.aau.dk/en/publications/how-building-regulations-ignore-the-use-of-buildings-what-that-me.
[10] Council of the European Union. “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 2018;61(210–328). Accessed: Mar. 25, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2018.328.01.0210.01.ENG&toc=OJ%3AL%3A2018%3A328%3ATOC.
[11] Lund H, et al. 4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems. Energy Apr. 15, 2014;68:1–11. https://doi.org/10.1016/j.energy.2014.02.089. Elsevier Ltd.
[12] Gadd H, Werner S. Heat load patterns in district heating substations. Appl Energy Aug. 2013;108:176–83. https://doi.org/10.1016/j.apenergy.2013.02.062.
[13] Calikus E, Nowaczyk S, Sant'Anna A, Gadd H, Werner S. A data-driven approach for discovering heat load patterns in district heating. Appl Energy Oct. 2019;252:113409. https://doi.org/10.1016/j.apenergy.2019.113409.
[14] Gram-Hanssen K. Residential heat comfort practices: understanding users. Build Res Inf Mar. 2010;38(2):175–86. https://doi.org/10.1080/09613210903541527.
[15] Gianniou P, Liu X, Heller A, Nielsen PS, Rode C. Clustering-based analysis for residential district heating data. Energy Convers Manag Jun. 2018;165: 840–50. https://doi.org/10.1016/j.enconman.2018.03.015.
[16] Tureczek AM, Nielsen PS, Madsen H, Brun A. Clustering district heat exchange stations using smart meter consumption data. Energy Build Jan. 2019;182: 144–58. https://doi.org/10.1016/j.enbuild.2018.10.009.
[17] Le Ray G, Pinson P. Online adaptive clustering algorithm for load profiling. Sustain. Energy, Grids Networks Mar. 2019;17:100181. https://doi.org/10.1016/j.segan.2018.100181.
[18] Johra H, Leiria D, Heiselberg P, Marszal-Pomianowska A, Tvedebrink T. Treatment and analysis of smart energy meter data from a cluster of buildings connected to district heating: a Danish case. In: In *E3S Web of conferences*, 172; Jun. 2020. p. 12004. https://doi.org/10.1051/e3sconf/202017212004.
[19] Gianniou P, Reinhart C, Hsu D, Heller A, Rode C. Estimation of temperature setpoints and heat transfer coefficients among residential buildings in Denmark based on smart meter data. Build Environ Jul. 2018;139:125–33. https://doi.org/10.1016/j.buildenv.2018.05.016.
[20] Widén J, Lundh M, Vassileva I, Dahlquist E, Ellegård K, Wäckelgård E. Constructing load profiles for household electricity and hot water from time-use data-Modelling approach and validation. Energy Build Jul. 2009;41(7): 753–68. https://doi.org/10.1016/j.enbuild.2009.02.013.
[21] Marszal-Pomianowska A, Heiselberg P, Kalyanova Larsen O. Household electricity demand profiles - a high-resolution load model to facilitate modelling of energy flexible buildings. Energy May 2016;103:487–501. https://doi.org/10.1016/j.energy.2016.02.159.
[22] M. H. Kristensen, R. E. Hedegaard, and S. Petersen, “Hierarchical calibration of archetypes for urban building energy modeling,” Energy Build, vol. 175, pp. 219–234, Sep. 2018, doi: 10.1016/j.enbuild.2018.07.030.
[23] Hedegaard RE, Kristensen MH, Pedersen TH, Brun A, Petersen S. Bottom-up modelling methodology for urban-scale analysis of residential space heating demand response. Appl Energy May 2019;242:181–204. https://doi.org/10.1016/j.apenergy.2019.03.063.
[24] M. H. Kristensen, R. E. Hedegaard, and S. Petersen, “Long-term forecasting of hourly district heating loads in urban areas using hierarchical archetype modeling,” Energy, vol. 201, p. 117687, Jun. 2020, doi: 10.1016/j.energy.2020.117687.
[25] Alahakoon D, Yu X. Smart electricity meter data intelligence for future energy systems: a survey. IEEE Trans. Ind. Informatics Feb. 2016;12(1):425–36. https://doi.org/10.1109/TII.2015.2414355.
[26] Aalborg forsyning. https://www.aalborgforsyning.dk/. [Accessed 25 March 2021]. accessed.
[27] Energimærkning boliger | energistyrelsen. https://sparenergi.dk/forbruger/boligen/energimaerkning-boliger. [Accessed 25 March 2021]. accessed.
[28] Wohnen Institut, GmbH Umwelt. Tabula WebTool. https://webtool.building-typology.eu/#bm. [Accessed 25 March 2021]. accessed.
[29] RStudio. Shiny from RStudio. 2017. https://shiny.rstudio.com/. [Accessed 25 March 2021]. accessed.