Energy 239 (2022) 122318

Contents lists available at ScienceDirect

Energy

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

Data driven model for heat load prediction in buildings connected to District Heating by using smart heat meters

Mikel Lumbreras [a, *], Roberto Garay-Martinez [b], Beñat Arregi [b], Koldobika Martin-Escudero [a], Gonzalo Diarce [a], Margus Raud [c], Indrek Hagu [c]

[a] ENEDI Research Group, Department of Energy Engineering, Faculty of Engineering of Bilbao, University of the Basque Country UPV/EHU, Pza, Ingeniero Torres Quevedo 1, Bilbao, 48013, Spain
[b] TECNALIA, Basque Research and Technology Alliance (BRTA), Bizkaia Science and Technology Park, Astondo Bidea 700, Derio, Spain
[c] GREN Eesti, Turu 18, Tartu, Estonia

ARTICLE INFO

Article history:
Received 14 May 2021
Received in revised form
6 October 2021
Accepted 9 October 2021
Available online 12 October 2021

Keywords:
Load forecasting
Heat meters
Data-driven model
Building
District Heating

ABSTRACT

An accurate characterization and prediction of heat loads in buildings connected to a District Heating (DH) network is crucial for the effective operation of these systems. The high variability of the heat production process of DH networks with low supply temperatures and derived from the incorporation of different heat sources increases the need for heat demand prediction models. This paper presents a novel data-driven model for the characterization and prediction of heating demand in buildings connected to a DH network.

This model is built on the so-called Q-algorithm and fed with real data from 42 smart energy meters located in 42 buildings connected to the DH in Tartu (Estonia). These meters deliver heat consumption data with a 1-h frequency. Heat load profiles are analysed, and a model based on supervised clustering methods in combination with multiple variable regression is proposed. The model makes use of four climatic variables, including outdoor ambient temperature, global solar radiation and wind speed and direction, combined with time factors and data from smart meters. The model is designed for deployment over large sets of the building stock, and thus aims to forecast heat load regardless of the construction characteristics or final use of the building. The low computational cost required by this algorithm enables its integration into machines with no special requirements due to the equations governing the model.

The data-driven model is evaluated both statistically and from an engineering or energetic point of view. R^2 values from 0.70 to 0.99 are obtained for daily data resolution and R^2 values up to 0.95 for hourly data resolution. Hourly results are very promising for more than 90% of the buildings under study.

1. Introduction

Energy consumption in buildings accounts for up to 40% of the total energy consumption in the European Union (EU) [1]. Considering this, increasing energy efficiency in buildings is one of the key targets of the EU strategy for the de-carbonisation of the economy ([2,3]).

Current District Heating (DH) networks are responsible for covering around 13% of the total thermal energy demand in the EU [4]. The evolution of DH networks over the years has been reducing supply temperatures, originally in the range of 80 °C and over, with the progressive implementation of the so-called 4th Generation District Heating (4GDH) ([5,6]) or Ultra Low Temperature (ULT) DH networks, which supply heat at temperatures around 45 °C. This has enabled an increased integration of low grade energy sources such as solar thermal (ST) systems [7] or waste heat (WH) streams ([8–10]) in the heat network.

The increasingly important role of renewable energy sources in 4GDH increases the variability of the heat generation profile in the heat production facilities. This requires the introduction of energy generation flexibility techniques to adapt heat production and demand in the network. To do so, accurate characterization methods for heat loads are required, so the available energy sources can be correctly managed with respect to such external variables as

* Corresponding author.
E-mail address: mikel.lumbreras@ehu.eus (M. Lumbreras).

https://doi.org/10.1016/j.energy.2021.122318

Acronyms

EU	European Union
DH	District Heating
ULT	Ultra-Low Temperature
ST	Solar Thermal
WH	Waste Heat
EC	European Commission
NN	Neural Network
SARIMA	Seasonal Autoregressive Integrated Moving Average models
DT	Decision Trees
DHW	Domestic Hot Water
SH	Space Heating
MPC	Model Predictive Control

Nomenclature

T_{OUT}	Outdoor Temperature [°C]
G_T	Global solar irradiance [W/m^2]
W_S	Wind speed [m/s]
W_D	Wind Direction [−]
Q	Heat Load [kWh]
α_0	Climate-independent demand [kWh]
α_1	Temperature parameter in [kWh/°C]
α_2	Solar radiation parameter in [kWh·m^2/W]
α_3	Wind-speed parameter in [kWh·s/m]
α_4	Wind Direction parameter in [kWh]
Q_{REF}	Reference demand dividing climatic dependent heat-load [kWh]
R^2	R-squared Value [−]
SSE	Sum Squared Regression Error [−]
SSYY	Sum Squared Total Error [−]
YEC	Yearly energy consumption deviation [%]

weather [11].

Heat meters are increasingly common in buildings, allowing the thermal energy consumption of each consumer from the heat network to be measured ([12,13]). Modern devices allow the hourly or sub-hourly gathering of energy and additional operational variables, including continuous communication with the DH utility. These devices are being widely implemented across the EU, mandated by Directive 2018/2002 [2]. This directive deals with the disaggregation of the final energy use by customers and the obligation to implement remote reading functionalities. Therefore, all meters will be remotely readable by January 2027. The remote access of such data leads to different energy management systems of heat production in DH networks, such as [14,15], based on frequent readings of smart heat meters at consumer level. These systems usually perform short-term forecasting in the range of some hours or days.

With regard to heat load forecasting alternatives for buildings, there has been recent research on white-box model forecasting based on such tools as EnergyPlus [16] or TRNSYS [17], including their calibration against meter data. However, these methods are not valid at DH scale, as the DH utility does not have the required information to develop such models (architectural data, use patterns, etc.) and the model development and calibration process is considered to be time and resource intensive. Thus, this approach is not considered to be reasonable on a district or city scale.

A more suitable alternative is provided by data-driven demand forecasting models that are partially or fully based on heat meter data. A wide variety of data-driven models exists, ranging from black-box models in which no prior knowledge from the building is required, such as the simplest energy signatures ([18–21]), up to more complex grey-box models formulated through differential equations that combine metered data with prior physical knowledge in building scale ([22–24]). Regarding grey-box models, Madsen et al. [22] developed a model based on discrete-time building performance, whereas Andersen et al. [23] described the time modelling of the heat dynamics using stochastic differential equations. Similarly, Bacher et al. [24] applied grey-box modelling for different applications regarding heat dynamics of a building, such as, control of indoor climate and energy consumption forecasting. In the following paragraphs a more extensive review of the studies related to data-driven models for demand characterization and prediction is provided.

Data driven models, based on different machine learning methods focused on electricity consumption, have been widely used in recent years. Many of the conclusions and knowledge acquired throughout the studies based on electricity consumption data are also applicable to space heating consumption. Tureczek et al. studied the conclusions from more than 30 papers about the applicability of clustering techniques to electricity consumption profiles in Ref. [25]. McLoughlin et al. [26] presented a study about electricity use patterns within the residential sector in Ireland, based on different clustering processes. This study characterized diurnal, intra-daily, seasonal and between customer electricity use. In Ref. [27], it is concluded that climatic conditions highly affect final electricity consumption in dwellings, and in Ref. [28] data from more than 4500 smart meters were used to conclude that individual electricity loads should be differentiated by use categories (residential, industrial, etc.), weekday and weekend, and summer and winter.

In contrast to the advanced situation of electricity consumption analysis, forecasting methods applied to heat loads are relatively new, and this research field is yet to be consolidated. To the authors' knowledge, initial works in this field ([29,30]) were developed in the early 2000s. In Ref. [29], a simple model was developed for forecasting demand in a DH network using outdoor temperature and social behaviour. Besides [30], presented an energy signature model for modelling different variables for the operation of a DH network.

Different types of data-driven models can be applied, which can be clustered in two main groups: grey-box models and black-box models. Grey-box models integrate prior physical knowledge and are typically formulated as state-space models through a set of stochastic linear differential equations, either in discrete or continuous time. Grey box models require a deep understanding of all relevant phenomena in a building that impact instantaneous or cumulated values of the load. Due to the complexity of these models, many grey-box models in the literature have been formulated for individual components of the building, such as walls or windows [31]. Thus, it is challenging to fit suitable grey-box models for multi-element systems such as buildings, because the interaction between the different elements and parameters is frequently unknown or too complex to be explicitly formulated.

In contrast, black-box data-driven models do not require the differential equations that govern building physics to be understood and implemented. Such models are purely based on data and can be trained to infer relations between inputs and outputs using statistical techniques with no physical interpretation.

Energy signature models are one of the simplest types of black-

box models, but these can provide successful results for monthly or seasonal data. Energy signature models are widely applied data-driven models that express the heating energy use as a function of weather variables. The first references to this type of studies were registered towards the end of the 1980s. In Ref. [18], a completely static energy signature was successfully presented for daily or lower frequency data; whereas [19] studied the statistical dependence between weather variables and the heating demand in buildings. In energy signature literature, outdoor temperature is considered to be the most dominant weather variable ([19,20]). The usual choice of outdoor temperature as the unique predictor variable can be partially explained by the difficulty to access good historical data of other climatic variables. In Ref. [21], outdoor temperature, global solar radiation and wind speed were used as the weather parameters for the models. In other studies, such as [32], relative humidity was also included. Relative humidity is not included in the model proposed in this work, since the climate in Tartu is very cold and dry, its impact on the heating energy use thus being low. Results from all these studies concluded that outdoor temperature is the most influential parameter and it is highly commendable to consider solar radiation [33]. Moreover, considering that the system is only supplying heat to buildings, it does not need to account for latent refrigeration loads. However, energy signatures are only valid for low-resolution predictions, such as weekly or monthly accumulated energy forecasting.

For daily or hourly heat load forecasting, more sophisticated models can be found. Grosswindhager et al. in Ref. [34], presented an approach for on-line short-term load forecasting using seasonal autoregressive integrated moving average models (SARIMA) in state space representation, resulting a mean absolute percentage error (MAPE) of 4.4% for the accumulated demand in the DH network. In Ref. [35], different SARIMA models are compared for prediction of heat demand of the total district, with minimum MAPE above 5% and R^2 values around 0.7. Other models based purely on machine learning techniques, such as neural networks (NN), have recently been applied for the calculation of heat loads in buildings. In Ref. [36], a study of several NN architectures is presented to forecast heat loads in a residential building in Canada in the very short term (hourly) and short term (daily). The obtained prediction results vary from MAPE values around 3–4%. In turn [37], presents an NN-based model with 13 input variables including weather, energy and social behaviour parameters in order to predict the hourly heat demand of a commercial building. In this final case, an error of 3.2% is obtained. In general, the greatest inconvenience of this type of sophisticated models is that the phenomena that really define the heating demand in a building are usually unknown.

Due to the greater number of years that electricity meters have been installed, most of the studies regarding energy forecasting are applied for electricity demand, despite some studies for heating demand can also be found ([38,39]). Besides, most of the studies related to the data driven models for heating demand prediction are usually limited to applying them to a specific building [40] and more general models are applied to low time-resolution (weekly or monthly) predictions ([41,42]). In both [41,42], simplified models are developed using monthly data for the forecasting of monthly energy use in buildings. In the checked literature, models with higher prediction accuracy are based on sophisticated machine-learning techniques (e.g. Ref. [43]) in which it is not fully clear the phenomena that determines the exact value of the heat consumption in a building.

Therefore, the following novelties can be outlined from this study.

a) Definition of the model: a multistep model is proposed based on supervised clustering learning and multivariable regression. The presented methodology enables the characterization and the short and mid-term heat load forecast for different buildings relying on weather data and calendar information. Dependencies between different variables are built and the parameter identification is run for relevant subsets of the calendar. In Ref. [44], it is shown that user behaviour and its impact on the heating demand correlate with time or calendar variables in commercial buildings. This effect, albeit with different correlation coefficients, is also replicable for residential and other types of buildings.
b) Wide range of applicability: the multi variable model presented in this study aims to be valid for any type of building, regardless of the heating profile or final use, since the building stock connected to a DH network is usually made up of all kinds of building types. Thus, the model is applied to 42 buildings located in Tartu (Estonia) and connected to the same sub-network of a DH system.
c) High temporal resolution predictions: the present model is applied to hourly and daily data, thus meeting the necessity of high temporal resolution models.
d) Simplicity and accuracy: the proposed model is based on relatively simple equations and the low calculation/processing cost, and the consideration of any type of final use of the buildings modelled, makes it suitable for deployment on such large scales as full DH networks.

2. Methodology

This section outlines the general methodology applied to the data. Firstly, Section 2.1 introduces the sources of the used data within the model and how this data is pre-processed before starting to develop the model. Section 2.2 provides an insight into the initial analysis of the available data and the identification of the time-correlation of the demand by means of Decision-Trees (DT). In Section 2.3, the data-driven model is defined, showing the equations and the training process in different steps. Finally, in Section 2.4, an approach to the analysis of the results is presented. The rest of the paper consists of the presentation and discussion of the results, analysing their implications for the real application. The detailed methodology is illustrated in Fig. 1.

The steps are detailed in the following subsections.

2.1. Data sources & data pre-processing

The starting point for this work is the data collection from two different sources: Data from DH substations (heating demand, among others) and data from a weather station, both for the year 2019.

The heating load profile consists of data from 42 substations of the DH network in Tartu (Estonia). All these substations are located in the sub-network of Tarkon and each contains a smart energy meter that is constantly measuring different variables in the system and sending it remotely to GREN [45], the DH operator, every hour. Table 1 summarizes the type of buildings monitored in the sub-network.

The location scheme of the smart energy meters in the heat network is shown in Fig. 2. Each building is identified by a code (ID number), completely independent from the real address of the location to avoid any type of identification problems and to preserve the privacy of the users.

The energy meters read and send data of different variables, such as the total energy consumption and different supply and

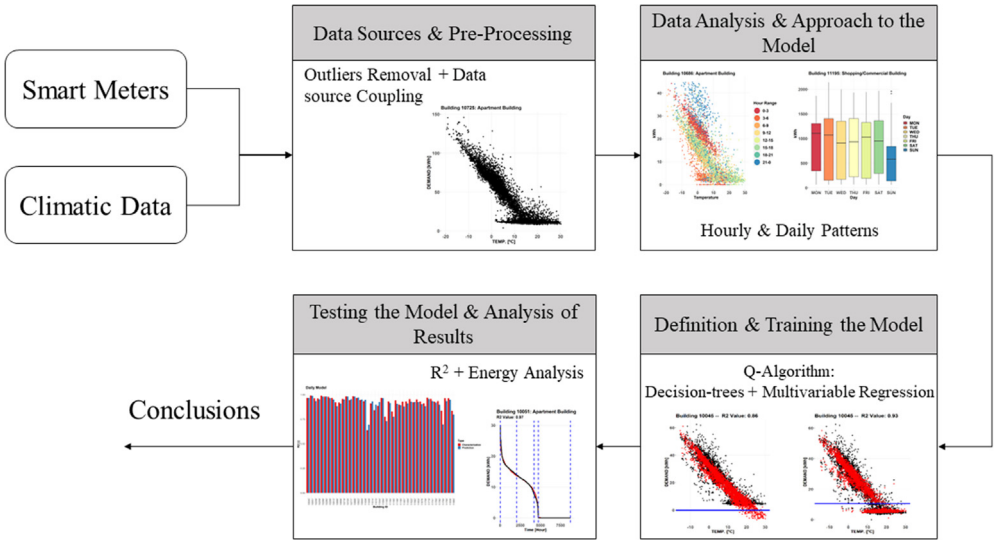

Fig. 1. General Methodology of the study.

Table 1
Summary of type of Buildings analysed.

Use of Buildings	Number of DH Substations
Residential apartments	25
Private House	7
Commercial Buildings	1
Educational Buildings	8
Offices	1

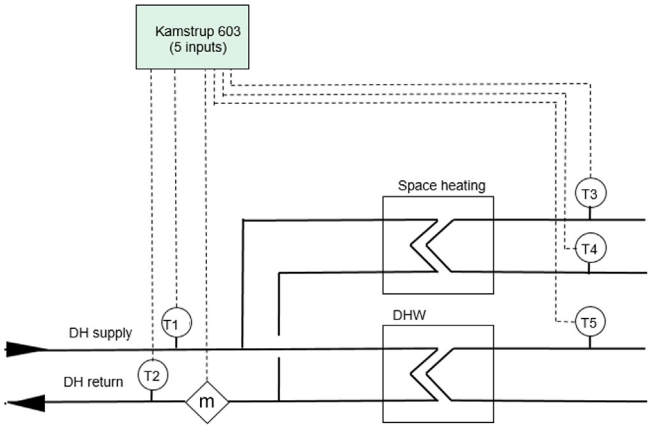

Fig. 2. Location and lay out of the smart energy meters in the DH in Tartu. Source [45].

return temperatures (T_i in °C presented in Fig. 2) in the heat distribution loop of the primary and secondary sides of the network. As presented in Fig. 2, five different temperatures (T_i in °C) and a volumetric flow ($\dot{m}$ in m^3/h) are measured. For this study, the total heat consumption is used, calculated using T1, T2, m and the specific heat of water. The corresponding unit changes are made to obtain the heat consumption in kWh. The energy meter installed in the buildings is the Multical® 603 from Kamstrup [46]. The accuracy of this devices is always higher than the one fixed in the European directive for this purpose (EN-1434–1:2015 [47]) and the measuring error remains below 5% in all the variables read. Heat energy consumption is saved as a cumulated variable and is read hourly. Consequently, the hourly energy use is calculated as the measured reading in that hour minus the measured value in the previous hour. Each substation corresponds to one building. Among the substations under study, different types of thermal zones can be found in terms of the final use. In this sense, residential apartments, offices, educational buildings and commercial buildings are included.

Regarding the climatic variables, data from a weather station located and managed by the University of Tartu [48] has been sourced with a 15-min frequency. The parameters used in this work are the outdoor temperature, T_{OUT} [°C], the global solar irradiance on a horizontal plane, G_T [W/m^2], the wind speed, W_S [m/s], and the wind direction, W_D [°]. As previously stated, the introduction of relative humidity into the model has been discarded due to the specific climate of Tartu.

The process for coupling both data sources is a calendar-based process developed in Ref. [49]. Data for 2019 is obtained for both heat load and weather data, but with different time-frequency and format. The weather data in a 15-min frequency was resampled to hourly intervals, obtaining 8760 readings representing each hour of the year. This matches the time format of the data from the smart energy meters, which was provided with an hourly frequency. Both data sources of the study were coupled by the exact date and time (month, day and hour).

In both data sources, outliers and reading errors are found. Reading errors are directly removed from the original dataset, reducing the total data points available. For the identification of the outliers, quartiles of each variable are calculated using boxplots in Ref. [49]. Interquartile range (IQR) has been used as a criterion so that all observations above the third quartile + 1.5·IQR and the values below first quartile − 1.5·IQR are considered potential outliers, where IQR is the difference between first and third quartile. This process for outlier removal has been widely applied in different studies ([50,51]).

For daily data, an additional process is followed, where variables are aggregated for each day, obtaining a smaller data array with 365 readings. In this process, and due to the large amount of data available, the incomplete days (24 measures) are directly removed. Thus, there is no need to calculate these missing values.

Finally, training and test datasets are determined. As is discussed in the next section, different consumption patterns have been recognized with respect to the season of the year. In order not to exclude these consumption patterns, training and testing data

are defined containing odd and even days, respectively. The data from odd days have been used to calibrate and train the models; whereas data from the even days have been used to test and verify the model's performance.

2.2. Initial data analysis and modelling approach

In a first observation of the heating energy demand, a range of different heat profiles are found among the different buildings under study. These can be attributed to the different final uses of the buildings and the energy consumption patterns of the users in their respective dwellings. However, the model presented in the following section aims for a general application to any building, independently of usage or heat profile.

Some of the buildings included in the study show thermal energy consumption only for space heating purposes (e.g. Building 10051, Building 10512, etc.); whereas other buildings consume energy for both space heating and hot water production (e.g. Building 10045, Building 10718, etc.). In all cases, the energy metered is the total heat consumption of the building (see location of the energy meter in Fig. 2). The energy required to satisfy space heating demand is dependent on both the climatic variables and the physical characteristics of the building (such as geometry and thermal envelope). Thus, it can be anticipated that when the outdoor temperature is low or the solar irradiance is low, the demand for space heating consumption will be higher. It is concluded that weather variables and SH demand show a large correlation. However, hot water consumption shows little to no dependence on climatic variables and primarily responds to use patterns and seasonal variations. Thus, a young worker and a retired person are expected to have quite different DHW consumption profiles. To illustrate the variation in the heating profiles of different buildings, Fig. 3 plots the hourly total heat consumption of two buildings (both apartments) against the outdoor temperature. The heat demand of Building 10051 (Fig. 3a) is not affected by DHW consumption; whereas, in Building 10725 (Fig. 3b), part of the heat demand is dedicated to that purpose. This effect can be observed in the lowest part of both figures, where a roughly horizontal profile is identified.

A night setback or a reduction in the demand has been identified in certain buildings, where heat energy consumption patterns differ along different hours, independently from the climatic variables of that moment, incorporating a time dependency into the consumption. Thus, calendar variables and heating demand variables are somehow correlated. The night setback can be used by the DH operator to reduce energy production in periods when a low heat load is expected, regardless of the climate conditions. Moreover, the high thermal inertia of the DH network could be used to satisfy the possible heat energy demand at night. In this context, Fig. 4 shows the heat profile of two buildings under study where a night setback has been identified. In both buildings, a reduction of the heat load is identified more or less between 3AM and 5AM. Even though the size of the buildings under study is unknown, large buildings have high thermal inertia that permits to maintain an indoor temperature for some hours. These two thermal inertias in the system (building and network) enable to control the demand peaks. The effect of the night setback in the demand leads the author to the definition of the first level for the decision-tree (LVL3 in Section 2.3) for application in the model.

Furthermore, the daily aggregated heating energy consumption profiles allow the energy share used for daily DHW to be identified, as shown in the next section. However, in the same way as has been done for the hourly data, additional time-dependent patterns have been identified in the daily aggregated data. In buildings that have no occupation at the weekends (e.g., offices or schools), this phenomenon is more noticeable. Fig. 5 shows an example of how heat energy consumption varies with respect to the day of the week, by means of a boxplot of the quartiles of daily heat energy demand. It can be observed that Building 11166 (Fig. 5a) presents a lower demand on Saturdays and Sundays, matching the days of no occupancy. In the same manner, Building 11195 (Fig. 5b), which corresponds to a commercial building, only presents lower demand on Sundays, as this type of building is usually closed on this day. As a general conclusion, heat consumption at the weekends is lower than on weekdays in some of the buildings. This is caused by the lower or non-occupancy of the buildings those days. This effect leads to the definition of the second variable of the decision-tree (LVL2 in Section 2.3).

Finally, specific seasonal patterns have been identified in different buildings, identifying two main periods: Summer & Rest of the year/winter. For instance, despite there being relatively low external temperatures at some moments of the summer, the monitored heat energy consumption does not correspond to expectations for similar climatic conditions outside this season. This divergence could be motivated by a reduction of the heat load by the DH operator in this period. However, this phenomenon is not

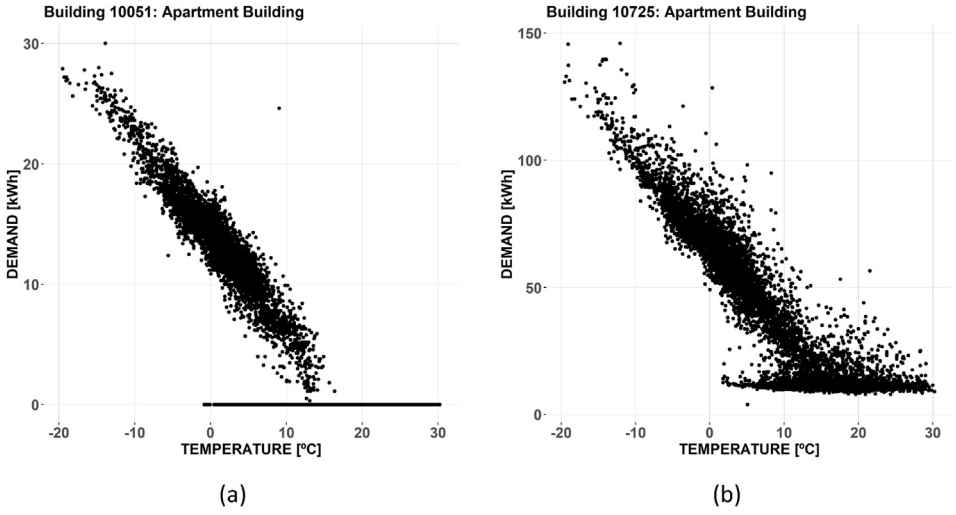

Fig. 3. Hourly heat consumption (vertical axis) vs outdoor temperature (horizontal axis) in Building 10051 (a) and Building 10725 (b).

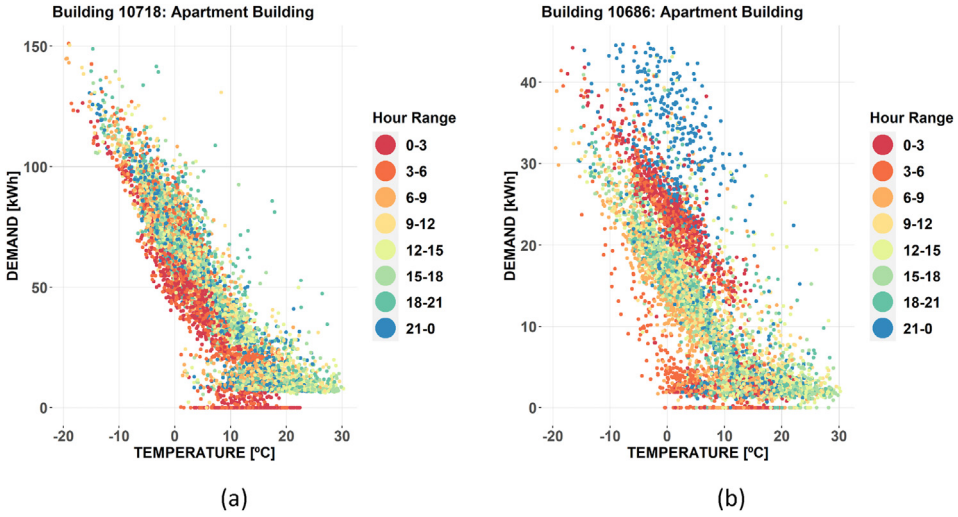

Fig. 4. Identification of night setback in Building 10718 (a) and Building 10686 (b).

Fig. 5. Daily consumption patterns in Building 11166 (a) and Building 11195 (b).

identified in all the buildings under study and, in consequence, a general methodology is necessary for the identification of the summer period performance. It has been observed that the variability of the demand in the summer period is much lower than the variability in other periods of the year. Consequently, the summer heat consumption follows a more stable (less varying) profile through time. Indeed, the comparison of the standard deviation between data periods was found to be an accurate method for identifying the relevant summer period for each building. Batches of 15 days were selected, and this methodology was applied to all the buildings under study. As a result, from the application of this method, the start and finish day of the summer are obtained for each of the buildings. These periods differ from building to building and this is way this methodology is applied independently. This list of days will serve as the input variable in the decision-tree (LVL1 in Section 2.3).

All in all, it can be concluded that energy consumption in the analysed buildings is highly dependent on the weather parameters and also on the specific user-behaviour. The latter factor is frequently omitted in modelling tools due to its random nature, which adds a significant complexity to the problem. However, it has been considered here for the sake of accuracy. The developed model is detailed in the following sections.

2.3. Definition of the model

As a first approach for the mathematic characterization of the model, it was proposed to split the data in two parts by a specific temperature threshold. The consumption data matching an external temperature above that threshold was attributed to periods with no space heating consumption (no consumption or DHW consumption only); whereas the data below that temperature threshold would also entail SH consumption. However, unsuccessful results were obtained, since this initial premise was not representative of most of the buildings and a large part of the data was not included in the characterization process by the model.

As a more suitable alternative, we decided to use the heat load as the threshold and the following type of equation is proposed, the so-called Q-algorithm:

$$Q_{alg}\ Q = \begin{cases} \alpha_1 \cdot T_{OUT} + \alpha_2 \cdot G_T + \alpha_3 \cdot W_S + \alpha_4 \cdot W_D, & Q < Q_{REF} \\ \alpha_0, & Q \geq Q_{REF} \end{cases} \quad (1)$$

In this algorithm, a calibration process must be performed using training data in order to obtain the coefficients needed for the application of the same model to testing data. For this calibration process, the data are split by a reference heat load, Q_{REF}. The data below this reference load would not be weather dependent; whereas the data above this point is assumed to follow a linear correlation with the abovementioned set of climatic variables. The process for the calculation of Q_{REF} is carried out in an iterative manner by using a range of different heat load thresholds to split the data, ranging from a minimum of $Q = 0$ to a maximum of $0.5 \cdot Q_{MAX}$. Observing data, the instant DHW consumption never exceeds 40% of the maximum load in any building. Thus, 50% is taken as the maximum limit for this iterative process. The absolute error of the regression is calculated in each step, so the heat load that minimizes the error in the second part of the equation ($Q \geq Q_{REF}$) determines the Q_{REF} value. This same algorithm logic is applied to both hourly and daily data.

The iterative process proposed for the calibration of the model is replicated for all the buildings under study. As expected, different calibration coefficients are obtained for each of the buildings in the district. Fig. 6 illustrates 4 steps (the number of iterations for each building are 50) of the iterative calibration process for one of the buildings studied: in this example, the third one (bottom left) would represent the most accurate choice. Together with the figure of the iterative process, the R^2 value obtained in each of the regressions is shown. Note that the Q_{REF} value is not necessarily equal to the base DHW consumption. In all the cases the heat consumption for DHW is equal or less than Q_{REF}. In other words, the optimal Q_{REF} is the same or higher than the constant part of the demand in Fig. 6 (third step).

As concluded from the previous section, heat consumption data is not only weather dependent but also time dependent, following different consumption patterns as a function of the hour of the day, day of the week and day of the year.

Therefore, and in order to obtain a more accurate result, decision trees (DT) are applied in the algorithm, in three different levels. Decision trees are non-parametric supervised techniques that predict values of responses by learning decision rules derived from features. For this model, the following three time-variables or features are introduced:

- **LVL1:** Variable season, divided into summer and rest of the year (SUM/REST)
- **LVL2:** Day of the week (MON, TUE, WED …)
- **LVL3:** Hour of the Day (1AM, 2AM, 3AM …)

This supervised classification process enables the characterization of a dynamic problem using stationary equations. The first

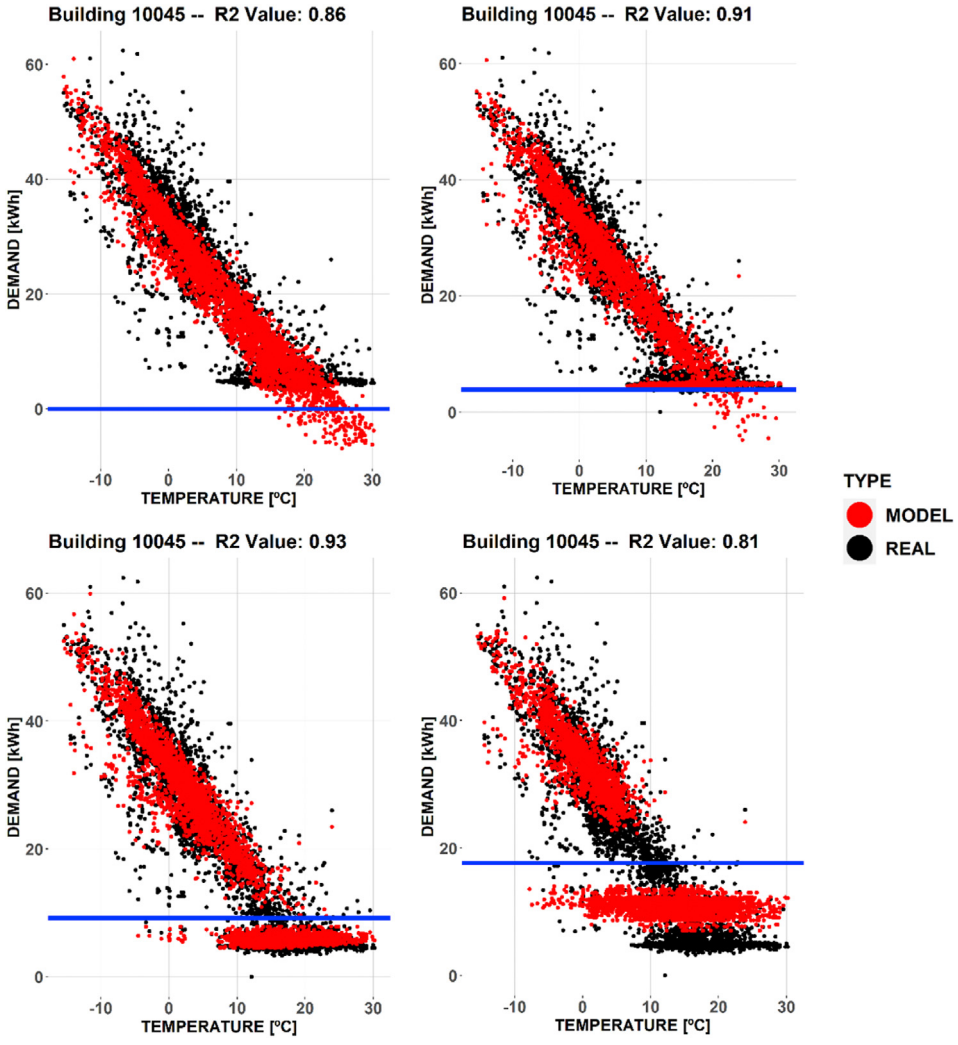

Fig. 6. Four steps of the calibration iterative process of one building (Building 10045) using hourly data.

level of the DT enables the characterization of the possible seasonal variations in the demand, as observed in some of the buildings under study. Besides, daily and hourly levels of the DT allow to introduce the influence of user-behaviour in the demand and identifying the different heat consumption patterns shown in Figs. 4 and 5 for some of the buildings.

Therefore, for the hourly model, each hour is classified by the consecutive application of the three levels of DT, whereas for the daily data model, only the two first levels of the DT are used for the corresponding classification. The classification by means of the supervised clustering method results in different equation coefficients for each data subset, increasing both the calculation cost and the accuracy of the proposed model.

The whole process, including the decision trees and the abovementioned iterative process of the Q-algorithm, is applied to the training data to obtain the parameters that make up the model for each of the buildings. Then, the fitted model is applied over testing data to verify the accuracy of the model.

2.4. Analysis of results

The accuracy and efficiency of the model is numerically evaluated by the R squared value or coefficient of determination, R^2. This value represents the proportion of the variance that is predictable using the predictors of the model. The R^2 variable is calculated as follows:

$$R^2 = 1 - \frac{SSE}{SSYY} \tag{2}$$

$$SSE = \sum_{i=1}^{N} (X_i - Y_i)^2 \tag{3}$$

$$SSYY = \sum_{i=1}^{N} (X_i - mean(X))^2 \tag{4}$$

where Y_i is the value obtained from the prediction model, X_i represents the measured data and N is the number of observations.

However, the approach for the evaluation of the accuracy of the model is not only based on the calculation of the R^2 value and its analysis. The practicality of the model resides in the prediction of the heating demand so that the heat generation process can be optimized. The DH operator is responsible for the management of the heat production process in the entire DH network and, in this context, the analysis of the model's accuracy also has to be evaluated in energy terms. Adopting the R^2 value as the only criterion can favour an overfitted or biased model. For the application assessed in this study, the high thermal inertia of the DH network could assume these fluctuations and, therefore, the analysis focuses on global energy results.

Thus, the total yearly energy consumption deviation (YEC) is calculated as follows, where 0% indicates a perfect match between measurement and prediction. This metric is comparable with the abovementioned MAPE.

$$YEC = 100 \cdot \frac{\left| \sum_{i=1}^{N} X_i - \sum_{i=1}^{N} Y_i \right|}{\sum_{i=1}^{N} X_i} \tag{5}$$

3. Results and discussion

3.1. General results

For this study, data from 42 buildings connected to the DH in Tartu have been used and this section shows the results arising from the application of the models explained in Section 2. Even if the basis of the model is the same, the results obtained for daily and hourly data are separately shown and discussed. When the model is applied to the training data (odd days) again, the results measure the accuracy of the model to characterize the heat load of the building. If the model is applied to testing data (even days), the results measure the accuracy to predict the demand.

First, and in order to evaluate the accuracy of the model, Fig. 7 presents the R^2 values obtained from the application of the model to daily and hourly data. In these plots, both results for characterization and prediction of heat loads are shown.

For a daily resolution, the model yields an excellent fit to the monitored data: the minimum value for the R^2 among the 42 buildings studied is around 0.70, with the maximum value very close to one. The daily aggregation filters out the hardly predictable intra-daily variations, thereby reducing the inherent uncertainty of demand prediction. In general, R^2 values in characterization of the heat load are higher than the ones for prediction because the data used for tuning the parameters of the model is the one applied for characterization. However, in some of the buildings (e.g. Building 10922 and Building 10949) where the model obtains R^2 values above 0.90, prediction results are even better than those for characterization.

Lower accuracy is obtained for hourly data resolution. The low linearity and changing variability of the consumption patterns of the users in the building reduces the accuracy of the model. Nevertheless, favourable accuracy results (R^2 values above 0.60) are obtained for around 90% of the buildings. The minimum R^2 value is obtained in Building 11676 (residential apartment) and the maximum R^2 is reached in Building 10051 (residential apartment). As it occurs in daily data, the prediction accuracy results to be lower than characterization.

Some of the biggest deviations between model estimations and monitored data correspond to private dwellings (e.g., Buildings 11795, 11009 & 11860). From the authors' belief and experience, the implementation of statistical models on this type of buildings can be challenging, especially if they feature manual heat switching systems with an intermittent usage. These activities are hardly predictable for a data interval as low as 1 h. For this purpose, Fig. 8 presents the correlation between R^2 and YEC (defined in Section 2.4). The correlation between a purely statistic variable (R^2) and the variable including energy management is observed (YEC), classified by the final use of the building. This figure is divided into results for daily data (Fig. 8a) and hourly data (Fig. 8b).

As illustrated in Fig. 8, a slightly linear negative correlation is observed between R^2 and YEC values. Thus, lower R^2 values mean that the yearly energy predicted to be used in the building deviates more from the real energy use. This figure confirms that buildings used as private houses (purple) present the lowest accuracy results, both for daily and hourly data. It is remarkable that some of the buildings with relatively low R^2 values show almost no error for YEC. This means that despite that the prediction deviations throughout the year are offset by each other, reaching an almost perfect result for the annual energy consumption at the end of the year.

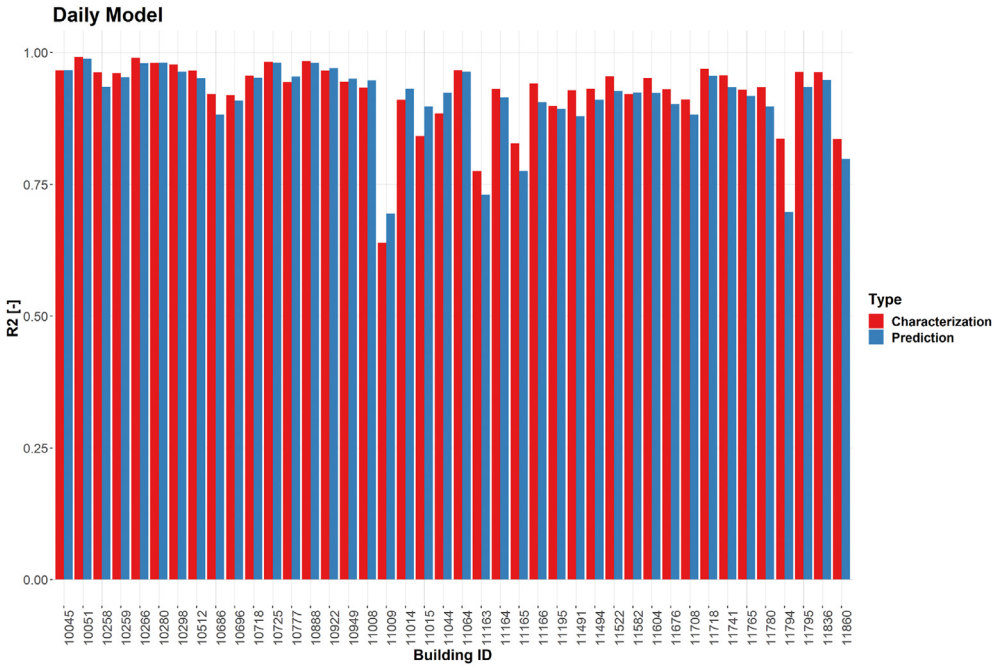

Fig. 7. R^2 Values in all the cases from (a) daily model and (b) hourly model.

As the buildings are connected to a DH network, the proposed model can be used to improve the control of the heat production system in the network. The modelling of individual buildings' demand enable to characterize the particular heat load patterns in each dwelling. This methodology provides an individual demand characterization and the demand of the whole district or specific branches could be obtained by the aggregation of the relevant buildings' demand. Thus, one of the most important advantages of this methodology is that the demand of the network can be adjusted if one building is disconnected from the network or if a new building is connected to the heating grid. Therefore, the heat production can be continuously optimized by matching the production to the predicted demand.

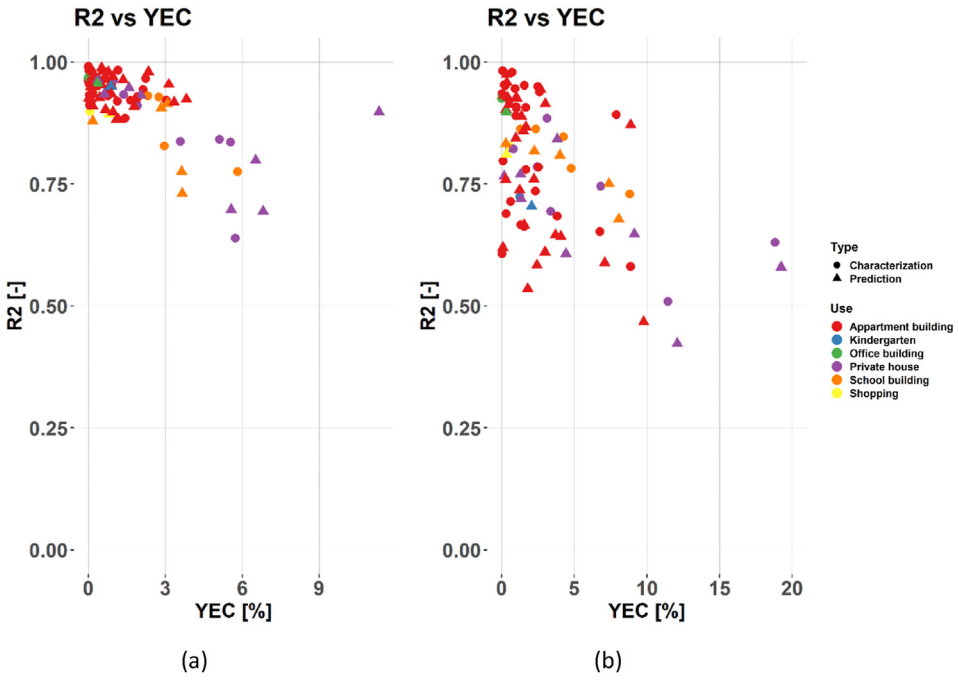

Fig. 8. R^2 vs YEC classified by type of building for (a) daily model and (b) hourly model.

3.2. Particular buildings

A deeper focus has been placed on three buildings covering a range of uses. These buildings have been selected for a deeper analysis because they cover a range of different heat load profiles, as requirements for their associated building uses are completely different. The results shown in Fig. 8 demonstrate that the model provides a suitable fit to the heat consumption profile in all three cases, regardless of the particularities of each building, as shown in Table 2. The following three buildings are chosen:

- Building 10051, a residential building with no demand neither for SH nor DHW in summer.
- Building 11164, an educational building/school where low DHW consumption is found in summer.
- Building 11718, a building with office use in which SH and DHW is consumed.

In Fig. 9, the hourly heat loads of these buildings are presented, comparing the monitored heat loads (black points) to the model estimations (red points). Fig. 9 presents a plot of the heat load against the outdoor temperature for the selected buildings, while Fig. 10 shows a monotonic plot of their heat loads. In Fig. 10, the quartiles (0%, 25%, 50%, 75% and 100% percentiles) of the demand are also included as vertical blue lines. Monotonic functions represent the ordered hourly heat profile from maximum (peak) to minimum load. These are valuable for DH operators as they portray a good overview of the heat consumption patterns of a building, such as maximum peak load, number of hours at peak load, number of hours at summer consumption pattern, etc. They convey the most important variables for managing and controlling heat production in the district by means of the different heat production plants along the network.

From Fig. 9, it can be concluded that the present model fits the general shape of the real data in the three buildings, with a minimum R^2 value of 0.85 in the school and a maximum R^2 value of 0.97 in Building 10051 (Fig. 9a). A low scattering of the demand points in Building 10051 facilitates the gathering of very accurate results when applying the model to predict the heating demand. Thus, the high scattering of the demand in Building 11164 (Fig. 9c) results in a lower R^2 value, probably caused by the greater variation of the set-point in the heating system due to the larger size of the building under study.

The monotonic function of the hourly heating demand shown in Fig. 10 presents the general trend of the prediction profile from the model. Note that hour zero corresponds with the 00:00AM of January of 2019. In Building 10051 (Fig. 10a), both lines representing the real consumption profile and the result from the model almost match. However, in the other two buildings under study, similar results are obtained. In peak demand (first blue line starting from the left, 100% quartile) moments, the demand from the model and the real data are very similar. At high demand moments up to the 3rd quartile, the model slightly underestimates the demand, as can be observed when the red line is below the black line in Fig. 10 (b) and Fig. 10 (c). The inflection point in both cases is located in the hour 2500, after which the model slightly overestimates the real

Table 2
R2 values for the three buildings selected for a deeper analysis.

	DAILY DATA		HOURLY DATA	
	Characterization	Prediction	Characterization	Prediction
Building 10051	0.99	0.99	0.98	0.97
Building 11164	0.96	0.92	0.85	0.81
Building 11718	0.91	0.96	0.93	0.90

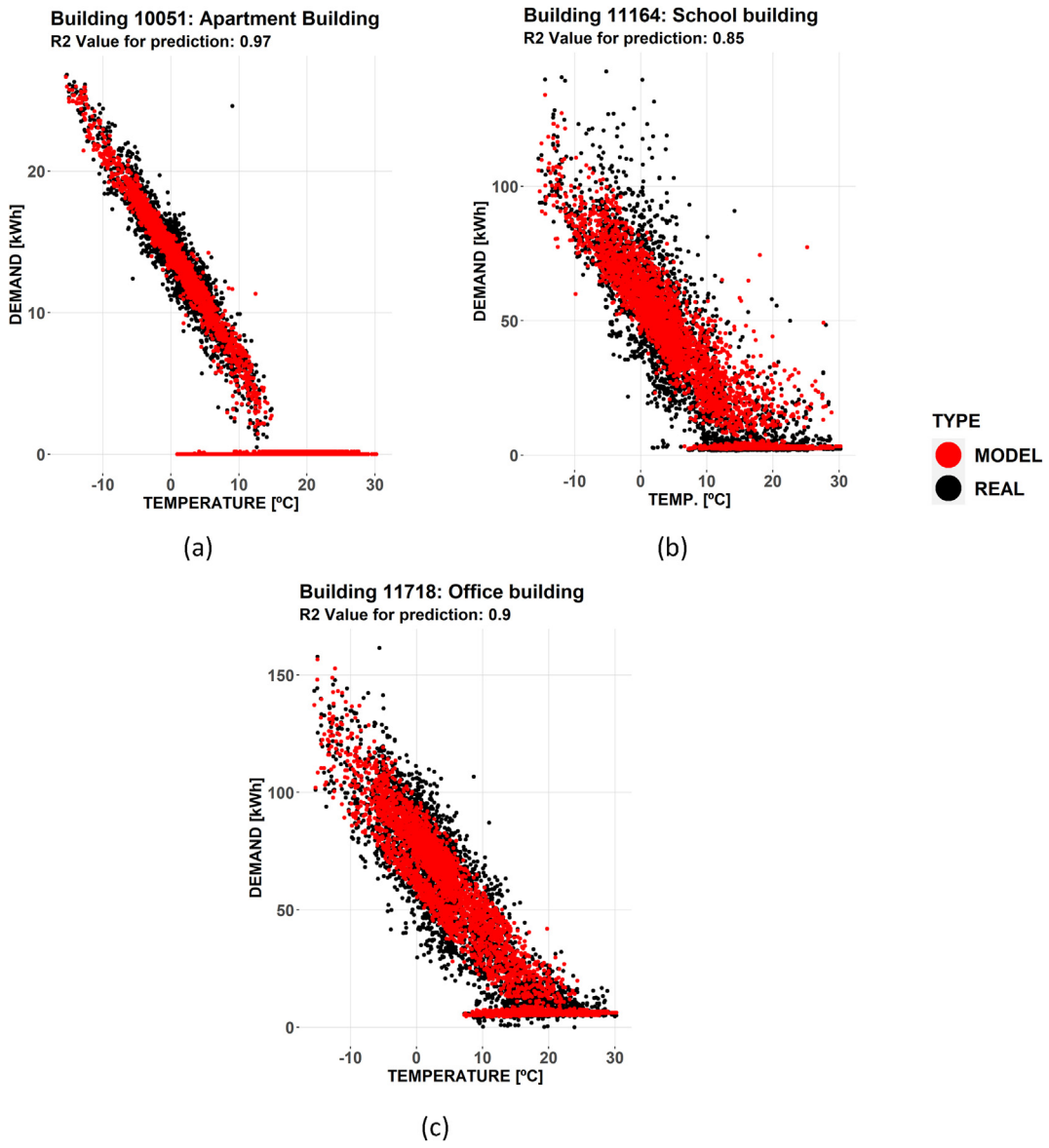

Fig. 9. Hourly heat load vs outdoor temperature for Building 10051 (a), Building 11164 (b) and Building 11718 (c). Black points represent real data and red points represent the predictions from the model. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

demand. Lastly, in the summer period, the model again fits the real demand.

Finally, an additional variable for measuring the accuracy of the model for the energy management of the DH network is the total yearly aggregated consumption estimated for each building. The sum of the estimations of each building would anticipate the total energy required to be produced and distributed by the network. Due to the large thermal inertia within the network, the variation in hourly demand could be compensated with heat storage. However, the annual heat production requirement is a key variable for avoiding the overuse of resources to produce heat for the network. Biomass is the dominant fuel in the assessed district, covering more than 50% of the primary energy share. Table 3 shows the total annual delivered heat monitored and estimated for each of the three buildings considered for the analysis.

Small variations between the real demand data and demand resulting from the model are observed. Table 3 presents the yearly energy consumption divided into training and testing data. The relative error of the real data is set at 0%. It can be seen that the total heat demand error remains below 5% of its real value and, in both Building 10051 & Building 11718, the error is very near the top zero. Moreover, with the exception of one case (Building 11718 and training data), the rest always show a positive relative error; in other words, the model estimates a slightly higher demand than the real one, which ensures the comfort conditions in the buildings.

On the whole, the proposed model appears to be viable for both daily and hourly heat consumption; considering the ease of application and the good accuracy of the estimations for most of the buildings. The application of this type of data-driven models in the operation and management of DH networks would be useful to

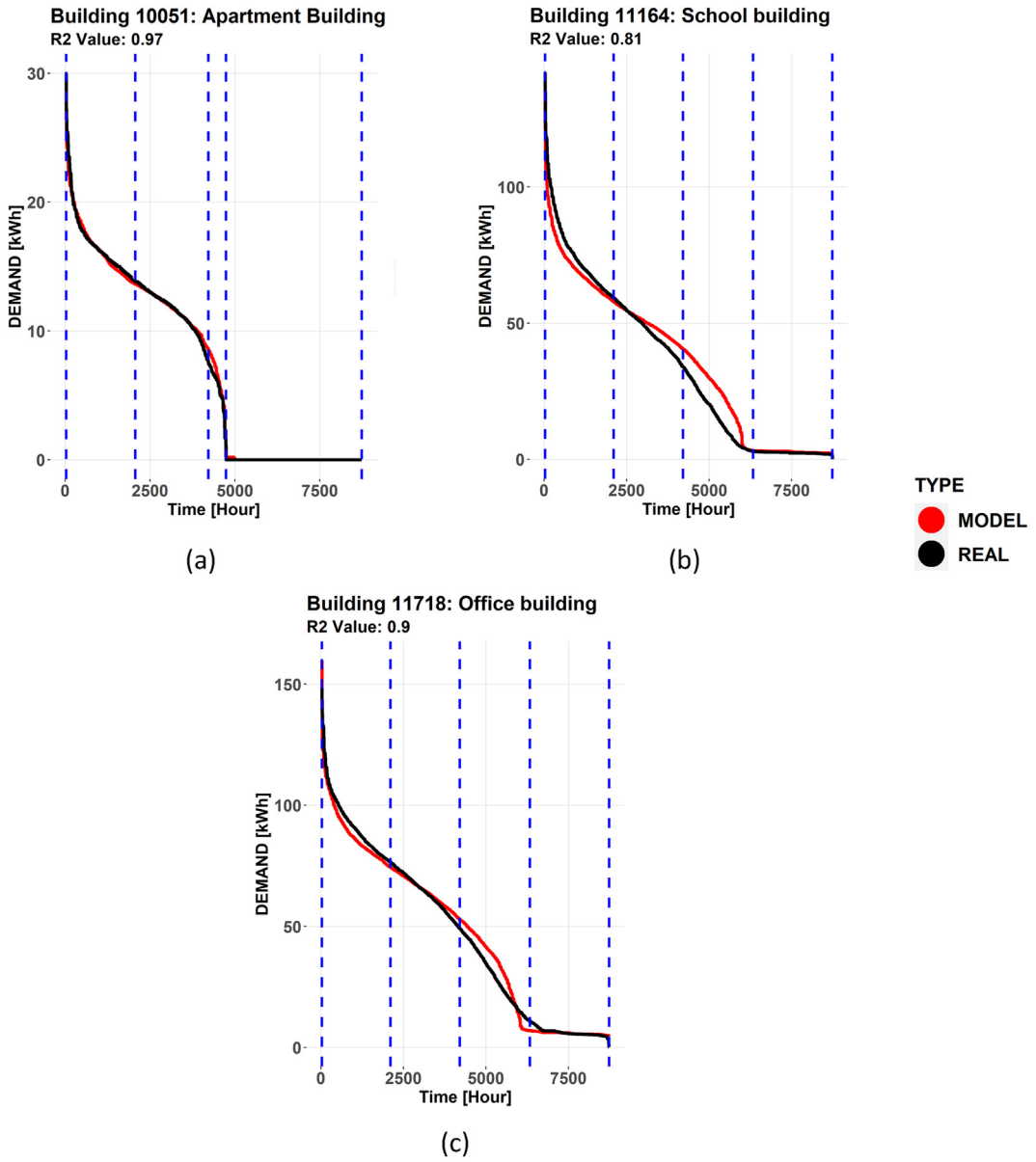

Fig. 10. Monotonic function of Building 10051 (a), Building 11718 (b) & Building 11164 (c). Real data in black & data retrieved from the model in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Yearly demand in GWh for real data and results from the model.

	TRAINING DATA				TESTING DATA			
	REAL DATA		FROM MODEL		REAL DATA		FROM MODEL	
	GWh/Year	YEC	GWh/Year	YEC	GWh/Year	YEC	GWh/Year	YEC
Building 10051	31.70	0	31.93	0.71	31.13	0	31.23	0.32
Building 11164	150.55	0	156.96	4.25	148.52	0	154.48	4.01
Building 11718	204.02	0	204.01	0.01	201.00	0	201.61	0.30

reduce primary energy consumption, as well as to achieve a more efficient operation within the flexibility allowed by the network.

4. Conclusions

In this study, a data-driven model for the characterization and prediction of heating loads in buildings connected to a DH network has been presented. In a preliminary analysis of these heat loads, an additional time dependency was found, related to time-varying consumption patterns and transient effects with a great effect on the instantaneous value of the heat demand. Time dependencies have been captured using decision trees with three levels, thus

maintaining the simplicity and stationarity of the model. This supervised clustering method allows, among others, to characterize the impact of users' behaviour on the heat energy demand.

The main contribution of the study is the development of a relatively simple model that could be deployed over large sets of buildings. This implies that the model needs to be generally applicable to any building, regardless of its usage pattern or construction characteristics. For this reason, no prior knowledge of the building has been incorporated into the model. Model inputs are limited to weather variables and calendar information, with hourly or daily heating consumption being obtained as a prediction output. Real data obtained from heat meters has been used for the validation of the model.

As a result of the abovementioned process, the following conclusions can be drawn from the study:

- The part of the heat demand corresponding to SH is weather and time dependent, while demand for DHW is solely dependent on the heat consumption patterns of the building. Supervised clustering enables the incorporation of this time-dependent consumption patterns into the model.
- When the presented model is applied to hourly data for a full year, the results have shown good agreement with metered data in predicting yearly and daily heat load profiles. Therefore, the model presented in this study could have a good use in applications that require the long-term energy performance of buildings.
- Weekly patterns are affected by occupancy schedules, mostly due to the weekday-weekend cycle. Generally, lower heat loads are found when the building remains unoccupied, with peak consumptions on the initial day of the week.
- Intra-daily patterns are also related to occupancy schedules, mostly business and leisure hours. However, additional variations have been found in heating patterns due to night setbacks.
- Statistically, the model obtains more accurate results in the prediction process for daily data resolution than for an hourly resolution. This can be attributed to the uncertainty of intra-daily consumption patterns.
- The model shows a good performance in predicting the total yearly aggregated heat demand in each of the buildings, with a maximum deviation of around 15% for the worst-fitted building.

The data-driven model presented in this study is straightforward to implement and does not require a large computational capacity. The results of the study demonstrate that an accurate hourly heat load prediction is obtained for most of the buildings under study. The availability of such estimations for a range of different buildings in a DH network could enable the optimization of the resources for heat generation, deriving in both primary energy and economic savings.

Credit author statement

Mikel Lumbreras: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing Original Draft, Writing Review & Editing, Visualization. Roberto Garay-Martinez: Conceptualization, Methodology, Writing Review & Editing, Formal analysis, Project administration. Beñat Arregi: Conceptualization, Methodology, Writing Review & Editing, Formal analysis. Gonzalo DIARCE: Methodology, Writing – review & editing. Koldobika Martin-Escudero: Methodology, Writing – review & editing. Margus Raud: Resources. Indrek Hagu: Resources

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.

Acknowledgement

This study has been carried out in the context of RELaTED project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 768567.

This publication reflects only the authors' views and neither the Agency nor the Commission are responsible for any use that may be made of the information contained therein.

References

[1] Pérez-Lombard Luis, Ortiz José, Pout Christine. A review on buildings energy consumption information. Energy Build 2008;40(Issue 3):394–8. https://doi.org/10.1016/j.enbuild.2007.03.007. ISSN 0378-7788.
[2] Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency.
[3] Directive 2012/27/EU of the European parliament and of the council of 25 october 2012 on energy efficiency, amending directives 2009/125/EC and 2010/30/EU and repealing directives 2004/8/EC and 2006/32/EC text with EEA relevance. Orkesterjournalen L 14.11.2012;315:1–56.
[4] Werner Sven. International review of district heating and cooling. Energy 2017;137:617–31. https://doi.org/10.1016/j.energy.2017.04.045. ISSN 0360-5442.
[5] Lund Henrik, Werner Sven, Wiltshire Robin, Svendsen Svend, Eric Thorsen Jan, Hvelplund Frede, Vad Mathiesen Brian. 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. ISSN 0360-5442.
[6] Li Haoran, Nord Natasa. Transition to the 4th generation district heating - possibilities, bottlenecks, and challenges. Energy Procedia 2018;149:483–98. https://doi.org/10.1016/j.egypro.2018.08.213. ISSN 1876-6102.
[7] Lumbreras Mikel, Garay Roberto. Energy & economic assessment of façade-integrated solar thermal systems combined with ultra-low temperature district-heating. Renew Energy 2020;159:1000–14. https://doi.org/10.1016/j.renene.2020.06.019. ISSN 0960-1481.
[8] Wahlroos Mikko, Pärssinen Matti, Manner Jukka, Syri Sanna. Utilizing data center waste heat in district heating Impacts on energy efficiency and prospects for low-temperature district heating networks. Energy 2017;140(Part 1):1228–38. https://doi.org/10.1016/j.energy.2017.08.078. ISSN 0360-5442.
[9] Ziemele Jelena, Roberts Kalnins, Vigants Girts, Vigants Edgars, Veidenbergs Ivars. Evaluation of the industrial waste heat potential for its recovery and integration into a fourth generation district heating system. Energy Procedia 2018;147:315–21. https://doi.org/10.1016/j.egypro.2018.07.098. ISSN 1876-6102.
[10] Open district Heating™. 2019. https://www.opendistrictheating.com.
[11] Fitó Jaume, Hodencq Sacha, Ramousse Julien, Wurtz Frédéric, Stutz Benoit, Debray François, Vincent Benjamin. Energy- and exergy-based optimal designs of a low-temperature industrial waste heat recovery system in district heating. Energy Convers Manag 2020;211. https://doi.org/10.1016/j.enconman.2020.112753. ISSN 0196-8904.
[12] Darby S. Smart metering: what potential for householder engagement? Build Res Inf 2010;38(5):442–57.
[13] Liu X, Golab W, Golab W, Ilyas IF. Benchmarking smart meter data analytics. In: Proc of the 18th international conference on extending database technology; 2015. p. 385–96.
[14] Lichtenegger Klaus, Wöss David, Halmdienst Christian, Höftberger Ernst, Schmidl Christoph, Tobias Pröll. Intelligent heat networks: first results of an energy-information-cost-model. Sustainable Energy, Grids and Networks 2017;11:1–12. https://doi.org/10.1016/j.segan.2017.05.001. ISSN 2352-4677.
[15] Vesterlund Mattias, Toffolo Andrea, Dahl Jan. Optimization of multi-source complex district heating network, a case study. Energy 2017;126:53–63. https://doi.org/10.1016/j.energy.2017.03.018. ISSN 0360-5442.
[16] U.S. Department. Of energy. EnergyPlus™; 2018. https://energyplus.net/.
[17] Klein SA, et al. TRNSYS 18: a transient system simulation program. Madison, USA: Solar Energy Laboratory, University of Wisconsin; 2017. http://sel.me.wisc.edu/trnsys.
[18] Hammarsten Stig. A critical appraisal of energy-signature models. Appl Energy 1987;26(Issue 2):97–110. https://doi.org/10.1016/0306-2619(87)90012-2. ISSN 0306-2619.
[19] Margaret F. Fels, PRISM: an introduction. Energy Build 1986;9(Issues 1–2):5–18. https://doi.org/10.1016/0378-7. ISSN 0378-7788.

[20] Ferbar Tratar Liljana, Strmčnik Ervin. The comparison of Holt–Winters method and Multiple regression method: a case study. Energy 2016;109:266–76. https://doi.org/10.1016/j.energy.2016.04.115. ISSN 0360-5442.
[21] Aalborg Nielsen Henrik, Madsen Henrik. Modelling the heat consumption in district heating systems using a grey-box approach. Energy Build 2006;38(Issue 1):63–71. https://doi.org/10.1016/j.enbuild.2005.05.002. ISSN 0378-7788.
[22] Madsen H, Holst J. Estimation of continuous-time models for the heat dynamics of a building. Energy Build 1995;22(Issue 1):67–79. https://doi.org/10.1016/0378-7788(94)00904-X.
[23] Andersen Klaus Kaae, Madsen Henrik, Hansen Lars H. Modelling the heat dynamics of a building using stochastic differential equations. Energy Build 2000;31(Issue 1):13–24. https://doi.org/10.1016/S0378-7788(98)00069-3. ISSN 0378-7788.
[24] Bacher Peder, Madsen Henrik. Identifying suitable models for the heat dynamics of buildings. Energy Build 2011;43(Issue 7):1511–22. https://doi.org/10.1016/j.enbuild.2011.02.005. ISSN 0378-7788.
[25] Tureczek Alexander. Structured literature review of electricity consumption classification using smart meter data. Energies 2017;10:584. https://doi.org/10.3390/en10050584.
[26] McLoughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterization using smart metering data. Appl Energy 2015;141:190–9. https://doi.org/10.1016/j.apenergy.2014.12.039.
[27] Andersen FM, Larsen HV, Boomsma TK. Long-term forecasting of hourly electricity load: identification of consumption profiles and segmentation of customers. Energy Convers Manag 2013;68:244–52. https://doi.org/10.1016/j.enconman.2013.01.018. ISSN 0196-8904.
[28] do Carmo Carolina Madeira R, Christensen Toke Haunstrup. Cluster analysis of residential heat load profiles and the role of technical and household characteristics. Energy Build 2016;125:171–80. https://doi.org/10.1016/j.enbuild.2016.04.079. ISSN 0378-7788.
[29] Dotzauer Erik. Simple model for prediction of loads in district-heating systems. Appl Energy 2002;73(Issues 3–4):277–84. https://doi.org/10.1016/S0306-2619(02)00078-8. ISSN 0306-2619.
[30] Heller AJ. Heat-load modelling for large systems. Appl Energy 2002;72(Issue 1):371–87. https://doi.org/10.1016/S0306-2619(02)00020-X. ISSN 0306-2619.
[31] Strachan PA, Vandaele L. Case studies of outdoor testing and analysis of building components. Build Environ 2008;43(Issue 2):129–42. https://doi.org/10.1016/j.buildenv.2006.10.043. ISSN 0360-1323.
[32] Powell Kody M, Sriprasad Akshay, Cole Wesley J, Edgar Thomas F. Heating, cooling, and electrical load forecasting for a large-scale district energy system. Energy 2014;74:877–85. https://doi.org/10.1016/j.energy.2014.07.064. ISSN 0360-5442.
[33] Potočnik Primož, Škerl Primož, Govekar Edvard. Machine-learning-based multi-step heat demand forecasting in a district heating system. Energy Build 2021;233:110673. https://doi.org/10.1016/j.enbuild.2020.110673. ISSN 0378-7788.
[34] Grosswindhager S, Voigt A, Kozek Martin. Online short-term forecast of system heat load in district heating networks. In: Proceedings of the 31st international symposium on forecasting, Prag, Czech Republic; 2011.
[35] Fang Tingting, Lahdelma Risto. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Appl Energy 2016;179:544–52. https://doi.org/10.1016/j.apenergy.2016.06.133. ISSN 0306-2619.
[36] Dagdougui Hanane, Bagheri Fatemeh, Le Hieu, Dessaint Louis. Neural network model for short-term and very-short-term load forecasting in district buildings. Energy Build 2019;203:109408. https://doi.org/10.1016/j.enbuild.2019.109408. ISSN 0378-7788.
[37] Alexander Sandberg, Wallin Fredrik, Li Hailong, Maher Azaza. An analyze of long-term hourly district heat demand forecasting of a commercial building using neural networks. Energy Procedia 2017;105:3784–90. https://doi.org/10.1016/j.egypro.2017.03.884. ISSN 1876-6102.
[38] Cholewa Tomasz, Siuta-Olcha Alicja, Smolarz Andrzej, Muryjas Piotr, Wolszczak Piotr, Guz Łukasz, Constantinos A. Balaras, on the short term forecasting of heat power for heating of building. J Clean Prod 2021;307:127232. https://doi.org/10.1016/j.jclepro.2021.127232. ISSN 0959-6526.
[39] Heine Kristensen Martin, Elbæk Hedegaard Rasmus, Petersen Steffen. Long-term forecasting of hourly district heating loads in urban areas using hierarchical archetype modeling. Energy 2020;201:117687. https://doi.org/10.1016/j.energy.2020.117687. ISSN 0360-5442.
[40] Ciulla G, D'Amico A. Building energy performance forecasting: a multiple linear regression approach. Appl Energy 2019;253:113500. https://doi.org/10.1016/j.apenergy.2019.113500. ISSN 0306-2619.
[41] Catalina Tiberiu, Joseph Virgone, Blanco Eric. Development and validation of regression models to predict monthly heating demand for residential buildings. Energy Build 2008;40(Issue 10):1825–32. https://doi.org/10.1016/j.enbuild.2008.04.001. ISSN 0378-7788.
[42] White JA, Reichmuth R. Simplified method for predicting building energy consumption using average monthly temperatures. In: Iecec 96. Proceedings of the 31st intersociety energy conversion engineering conference, Washington, DC, USA. vol. 3; 1996. p. 1834–9. https://doi.org/10.1109/IECEC.1996.553381.
[43] Xue Puning, Jiang Yi, Zhou Zhigang, Chen Xin, Fang Xiumu, Liu Jing. Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. Energy 2019;188:116085. https://doi.org/10.1016/j.energy.2019.116085. ISSN 0360-5442.
[44] Liang Xin, Hong Tianzhen, Shen Geoffrey Qiping. Improving the accuracy of energy baseline models for commercial buildings with occupancy data. Appl Energy 2016;179:247–60. https://doi.org/10.1016/j.apenergy.2016.06.141. ISSN 0306-2619.
[45] GREN Eesti, https://gren.com/ee/(Accessed in 2020).
[46] Karmstrup, https://www.kamstrup.com/en-us/heat-solutions/heat-meters/multical-603 (Accessed in January 2021).
[47] EN 1434-1:2015, Heat meters. Part 1: general requirements.
[48] University of Tartu, Institute of Physics, Laboratory of Environmental Physics, http://meteo.physic.ut.ee/?lang=en (Accessed in 2019).
[49] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. http://www.Rproject.org/.
[50] Li Aihua, Feng Mengyan, Li Yanruyu, Liu Zhidong. Application of outlier mining in insider identification based on boxplot method. Procedia Computer Science 2016;91:245–51. https://doi.org/10.1016/j.procs.2016.07.069. ISSN 1877-0509.
[51] Schwertman Neil C, Ann Owens Margaret, Robiah Adnan. A simple more general boxplot method for identifying outliers. Comput Stat Data Anal 2004;47(Issue 1):165–74. https://doi.org/10.1016/j.csda.2003.10.012. ISSN 0167-9473.