Energy & Buildings 247 (2021) 111160

Contents lists available at ScienceDirect

Energy & Buildings

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

A systematic approach for data analysis and prediction methods for annual energy profiles: An example for school buildings in Norway

Yiyu Ding *, Helge Brattebø, Natasa Nord

Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), Kolbjørn Hejes vei 1 a, Trondheim 7491, Norway

ARTICLE INFO

Article history:
Received 20 March 2021
Revised 12 May 2021
Accepted 31 May 2021
Available online 4 June 2021

Keywords:
District heating
Electricity
modified Z-Score
Hourly profile
Regression analysis
Schools

ABSTRACT

Current research on energy supply systems and building energy demand presents positive impacts from the two sides, with potentials of combining top-down and bottom-up modelling. Mostly, the energy demand input has been employed directly from energy utility companies as a package of information, without considering energy use patterns regarding building type. There lacks a bridge between demand profiles on building stock functions and urban energy supply systems. Accordingly, this article proposes a framework that enables the prediction of annual energy profiles, applied to one educational building type on an hourly basis. The work consists of five steps: (1) getting energy information of 40 district heating (DH) supplied schools in Norway, (2, 3) processing data for getting the modified average hourly demand per m^2 and holiday breakpoints through a modified Z-Score, (4) energy forecast of DH and electricity load profiles through temperature moving average, correlation, and linear regression analysis, (5) validation of the predicted yearly profiles by three criteria, and further with the cluster methods for DH profiles. The results showed that the suggested methods for annual energy forecast were satisfying. The defined load profiles might represent the current energy demand of the Nordic school and the methods could be transferred to other building types. With energy analysis of typical building types, the proposed method enables the planners to better understand the energy needs for different building functions.

1. Introduction

1.1. Background

Every year approximately 36–40% of the energy supply is used to serve buildings around the world, which is responsible for nearly one-third of global CO_2 emissions [1]. Energy efficiency strategies in urban building stocks are expected to make a high contribution to reductions in energy use and greenhouse gas emissions.

There is fruitful research suggesting how to make positive impacts from both energy system side and building side on the total building stock use. Several relevant publications are introduced below. Thellufsen et al. [2] investigated the possibilities of achieving a smart energy city within a 100% renewable energy context of Denmark and Europe. Averfalk et al. [3] found that lower distribution temperatures provide higher profitability and would facilitate transition to renewable and recycled heat supply in district heating (DH) systems. Lund et al. [4] gave perspectives on 5th generation DH that has a strong focus on combined heating and cooling as the main driver and may coexist with 4th generation technologies. From the view of enhancing building energy efficiency, Moschetti et al. [5] proposed a pathway for transiting building from zero energy to zero emission with a life cycle assessment on the most influential buildings' factors. When the techno-economic benefits analysis of building refurbishment is performed, Ascione et al. [6] emphasized the uncertainty in building occupant behavior shall not be neglected.

However, the energy demand input has usually been employed directly from utility companies as a package of information for energy system modelling. Without energy data mining or targeting energy use patterns on different building types, abnormal energy use could be misleadingly used as the input, which may deviate the anticipated benefits. According to the International Energy Agency Energy in Buildings and Communities Programme Annex 53, the offset issues from the designed demand to the real building energy were analyzed on different climates [7]. Meanwhile, Reynders et al. [8] explored the methodologies that enable residential buildings to offer flexibility to the energy system by utilizing building thermal storage. Therefore, a better understanding of energy usage and profiles in different building categories is needed and possible, with the aim to treat energy system and building in a holistic view.

* Corresponding author.
E-mail address: yiyu.ding@ntnu.no (Y. Ding).

https://doi.org/10.1016/j.enbuild.2021.111160
0378-7788/

Nomenclature

CV(RMSE)	coefficient of variation of the root mean squared error
DH	district heating
DHW	domestic hot water
ES curve	energy signature curve
GESD	generalized extreme studentized deviate
MAD	median absolute deviation
MAPE	mean absolute percentage error
NMBE	normalized mean bias error
PAA	piecewise aggregate approximation
SAX	symbolic aggregate approximation
SH	space heating
TMA	temperature moving average
TMY	typical meteorological year

el	Electricity
no.	Number
t_{ot}	outdoor temperature
R^2	coefficient of determination
N	number of observations
W	number of data in the new series in PAA

Greek letters

σ	standard deviation
μ	average value

1.2. Previous studies

As an important subset of non-residential buildings, the educational building aims to educate students with knowledge and social mind, on the basis of laws on day-care seats and rights to education in Norway and many other countries. This building category contains kindergarten, school, and college/university. Meanwhile, the building operators are responsible to maintain the required indoor environment in energy-efficient ways [9]. It has been found that energy spending is the second biggest operating cost for schools in the U.S., only beneath the employees' salaries [10]. In Italy, 60% of the educational buildings were built before 1976, most of which fail to meet the current energy performance requirements despite of extraordinary retrofit [11].

According to the building conditions in Norwegian local governments [12], schools make up nearly 50% of the total local public building mass and constitute the most important building type. Population is one of the key drivers for developing educational buildings. Under the expectation of population increase and urbanization to come, there is a growing need for educational building expansion [13].

The aim of this study was to understand and identify load profiles of one typical Norwegian educational building type: the district heating supplied school buildings. Local municipalities are responsible to monitor and manage the operation of public buildings, hence energy data for such buildings are often available. Finding appropriate energy statistical methods and prediction methods are needed to achieve the study goal.

Sun et al. [14] developed a cooling load prediction strategy for a super high-rise building in Hong Kong, by combining selection of reference day, calibration with weather data, and model accuracy enhancement. This method requires low computation load and is feasible for online application of building load prediction in order to optimize equipment operation and guide load shifting [14].

Fan et al. [15] proposed a transfer learning-based methodology by utilizing the massive well-measured building operational data to predict other buildings. A quantitative assessment of this methodology for 24-hour ahead building energy demand was studied on building types of office, schools, and universities. Comparing with individual models, this methodology could reduce 15–78% of prediction errors and give insights to realize the value of existing data in building energy management [15].

Liu et al. [16] suggested a method with two-step clustering analysis to identify the typical electricity load patterns (TELPs) at individual building level. Density-based spatial clustering application with noise (DBSCAN) algorithm clustering technique is used in the first step to detect daily outliers. The second step uses the k-means algorithm to group similar TELPs. The effectiveness of this framework was confirmed with the time-series electricity data of office buildings in Chongqing [16]. Another k-means algorithm-based clustering analysis was made by Gianniou et al. [17] for studying residential district heating data. These single-family houses in Aarhus were segmented based on heat use intensity and representative patterns, by examining with the characteristics of buildings and occupants, load profiles of households, and use behavior changes.

Raza et al. [18] reviewed a number of artificial intelligence based short term load forecasting techniques, among which artificial neural network (ANN) was praised with great performance for complex problems. The findings show that improvement of training capability of neural network is needed to achieve promising forecasting results.

From the literature study, the main efforts regarding prediction have been put on short term energy load forecasting (time-series based techniques and artificial intelligence based techniques), with the contribution to optimal scheduling of energy equipment, spinning reserve, evaluation of economic dispatch, etc. Medium to long term load forecast for typical building types has not been stressed well. Medium to long term load forecast, especially for the public buildings, can be used for efficient operation/maintenance of the local energy system and building energy planning policies. However, there are few studies on annual load profiles for specific building types. Melillo et al. [19] suggested a method enabling to automatically generate and reproduce the annual heat demand for residential buildings. However, data from a reference building is required for achieving the recoverability of relevant building and system parameters, and scaling up the proposed method to a larger building set has not been validated yet. Bao et al. [20] proposed a probabilistic approach to formulate annual cooling load profile for office building, by involving tens of thousands of generation outputs from weather data files by Monte Carlo simulations. This prediction method can save much more overcooling hours, which was validated by measurements. Lundström et al. [21] proposed a heat load weather normalization, by segmenting the weather data and heat load with heating degree days and performing multivariable linear regression. The work was validated with two multi-family residential buildings.

The novel contributions of this work may be summarized as follows. The defined specific hourly-based load profiles were built on the analysis of a group of schools. These typical load profiles give an insight to the general energy situation of schools and they can be used for building simulations and model calibration. The proposed systematic approach is simple and robust, and it can be easily transferred to a large number of other public buildings in a fast way. The findings and method are thought to benefit public administrations and energy planners regarding local energy planning.

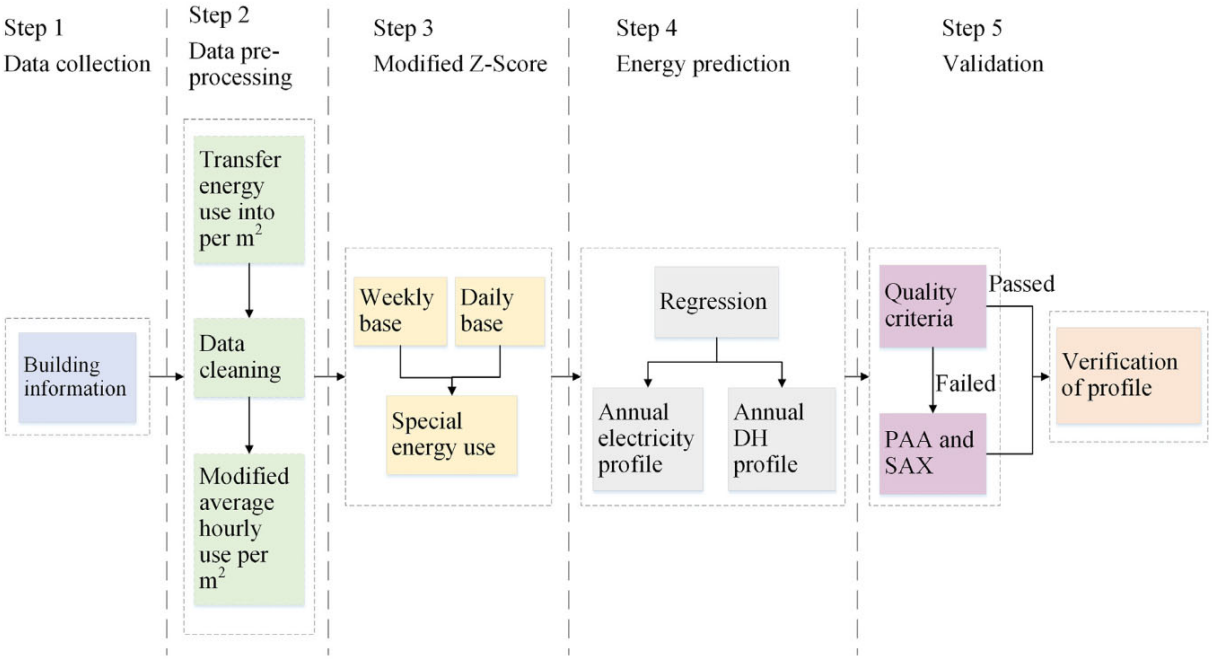

Fig. 1. The workflow of the data analysis of energy use profiles.

The rest of the article is organized as follows. Section 2 briefs the study methods about the data information of the observed buildings and the prediction process of energy profiles. Section 3 shows the periods of special energy use by adopting a modified Z-Score. Section 4 shows the prediction results of the typical annual DH profile and electricity profile. The accuracies of the predicted profiles were evaluated in Section 5. The application and limitations were discussed in Section 6. Finally, Section 7 concludes the main findings of this study.

2. Methodology

The outline of the five main steps of the energy analysis is illustrated in Fig. 1. Section 2.1 gives the building information. Section 2.2 explains how the energy data of the observed buildings were processed for getting the modified specific hourly demand for an average school building. Section 2.3 presents a modified Z-Score method to describe the energy use trend, which can identify the possible holiday breakpoints. Section 2.4, 2.5, and 2.6 describe the specific energy use forecast of DH and electricity that was proceeded through regression analysis, by using the modified specific hourly demand and considering the energy differences between normal days and holiday. Finally, Section 2.7 briefly introduces three quality criteria and cluster methods for validation. MATLAB was used for the energy data analysis.

2.1. Description of the observed buildings

The observed DH supplied schools are located in Trondheim, Norway. As the third largest city in Norway, Trondheim Municipality has been committed to improving strategies for a better living environment under the pressure of urbanization, population growth, and mitigation of anthropological carbon footprint.

The historical data of schools from 2015 to 2018 were retrieved from the energy monitoring platform of Trondheim Municipality [22]. Besides schools, other public buildings such as kindergartens, health/nursing centers, sports centers are also monitored in the platform. During the four years, the number of these schools registered in the monitoring platform increased from 36 to 40. The retrieved annual data file of each building included annual DH demand, electricity demand, and outdoor temperature in hourly resolution [22]. A total of 153 annual data files were used in the analysis. Table 1 summarizes the building information regarding average measured energy demand, energy labelling, building year, and building number in the data platform. The building year and energy labelling of the buildings were obtained from the Norwegian Energy Efficiency Agency (Enova) [23], which provides the information of the energy labelling scheme. Most of the buildings were built between 1980 and 2010 and labelled with C and D[1]. This building composition of school group can represent the current Norwegian situation that most of the schools have medium aged buildings and medium energy demand.

In the observed buildings, DH delivered the space heating (SH) and domestic hot water (DHW), while electricity was mainly used for ventilation, lighting, computers, and other electric appliances. The building area of each school varied from 1 822 m^2 to 8 996 m^2, and their average annual demand for DH and electricity of each building are shown in Fig. 2. The buildings regarding areas versus their annual demand were clustered with k-means.

It can be seen that a larger building floor area accounted for higher annual demand in general, but the relationship cannot be simply explained by linear regression models, due to a poor correlation. Hence, the energy analysis was performed on specific energy use (kWh per m^2) and the hourly resolution, in a search for representative energy use profiles for the school building in the Nordic climate.

2.2. Representative hourly energy demand

To find the representative hourly energy use, the simple way is to use the average hourly energy value of the n buildings as $\frac{\sum_{i=1}^{n} e_i}{n}$,

[1] The energy labelling was performed during 2011–2015, and possible update of equipment operation might happen afterwards, which may explain the building number difference when considering the measured annual energy demand in the platform and the labelling requirement. Contrarily, new buildings may have nevertheless higher energy use than stipulated by the building code [24]. Hereby the energy labelling result was used as reference.

Table 1
Information of the observed buildings.

Average measured annual energy demand (kWh/m^2)	≤70	≤100	≤135	≤175	≤220	≤280	
Building number	/	5	17	14	4	/	
Energy labelling	A	B	C	D	E	F, G	No infor.
Building number	1	5	8	14	7	1	4
Building year	Before 1950	1950–1979	1980–2010	After 2010			No infor.
Building number	3	8	20	5			4
Monitored year in data platform	2015	2016	2017	2018			
Building number	36	38	39	40			

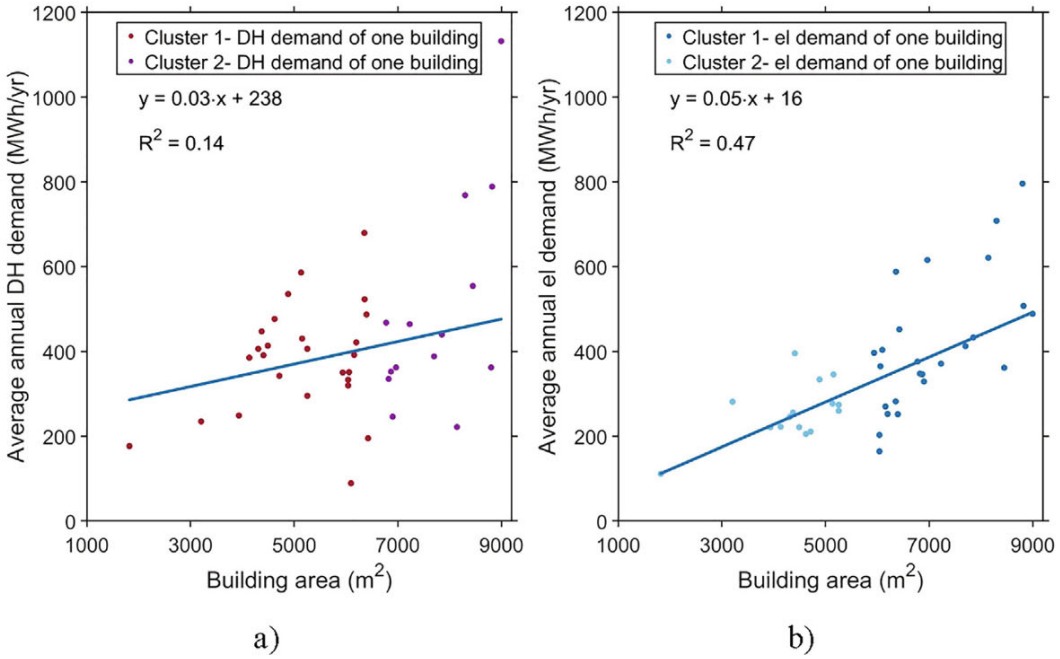

Fig. 2. a) Average annual DH demand vs building area, b) Average annual electricity demand vs building area.

where e_i is the hourly energy demand of building i, n is the total building numbers. However, this might mislead to high or low energy use. As stated in [15], the overall power use distribution of buildings (including schools and universities) is observed as right skewed, as few buildings have extremely high powers. The authors considered such rare measurements should be removed to avoid undesired model instability, with a threshold of maximum 10% extreme data of high powers being removed. In this study, there was one possible situation leading to atypical energy use. According to the report from Norwegian Water Resources and Energy Directorate (NVE) [13], student halls and canteens can be rented out for catering during evenings and weekends, which is meant to maximize the utilization of public resources. In other countries, such as the UK and the U.S., it is also seen opening part of schools to the public after normal school hours [25]. Educational buildings "can be used as communication means towards students and their families, and can thus reach many different society groups" [26]. Additionally, the working paper [12] addressed the importance of maintenance of public buildings. The annual report of Oslo [27], the Norwegian capital, specially emphasized the development and maintenance of schools must be planned and implemented with good functionality. Therefore, it was reasonably assumed some atypical energy might be caused by renting out and maintaining schools.

There is no widely accepted method to clear outliers automatically. Several methods of locating the outliers have been developed, such as the methods of median, mean, grubbs, generalized extreme studentized deviate (GESD), and observations beyond the range of quartiles, and each method has its own characteristics [28]. In this study, the method of "quartiles" was used. By following this method, data beyond 1.5 inter-quartile ranges of the upper (75 percentile) and lower quartiles (25 percentile) were detected as outliers. This method has proved useful when data is not normally distributed [29,30], which also fits the monitored dataset of the school group.

After clearing outliers of each hour, the average value of the remaining dataset was used. The typical energy demand for each hour was calculated as $\frac{\sum_{i=1}^{n'} e_i}{n'}$, where e_i is the hourly energy demand of building i among the remaining datasets, and n' is the total remaining building numbers. For simplicity, this modified, but still monitored average hourly energy demand was named as modified average energy demand in this study. This process was applied to both DH use and electricity use. By comparing the raw average hourly energy demand $\frac{\sum_{i=1}^{n} e_i}{n}$ in the original datasets with the modified average hourly energy demand $\frac{\sum_{i=1}^{n'} e_i}{n'}$ in the remaining datasets during the four years, only 9.5% of the differences ($|\frac{\sum_{i=1}^{n} e_i}{n} - \frac{\sum_{i=1}^{n'} e_i}{n'}|/\frac{\sum_{i=1}^{n} e_i}{n}$) were higher than 10% and most of the differences were minor. Hereby, it is appropriate to use the modified average hourly energy demand without a bias.

2.3. Modified Z-Score

The Z-Score is the number of standard deviations from the mean value, by subtracting the mean value from a raw data point and then dividing the difference by the standard deviation. If the absolute value of Z-Score is higher, it implies the raw data point is farther than the average. In this study, a modified Z-Score method defined by Iglewicz et al. was introduced.

As recommended by Iglewicz et al., this modified Z-Score is preferred to identify possible outliers over the common practice of Z-Score [31]. The modified Z-Score is defined as Eq. (1), x_i refers to the value of monitored sample, where $\tilde{x}$ refers to the median value of the samples, and MAD denotes the median absolute deviation in the dataset, see Eq. (2). The detailed information regarding this method can be found in [31].

$$M_i = \frac{0.6745 \cdot (x_i - \tilde{x})}{MAD} \tag{1}$$

$$MAD = median(|x_i - \tilde{x}|) \tag{2}$$

If the absolute value of the M_i to one data point is higher than 3.5, the point shall be marked as a potential outlier. This method is robust to detect outliers ranging from small to large sample size [31]. Accordingly, this method was used in this study to explore whether the building operation followed the schedule by considering the energy difference between the normal days and special periods, by showing the energy profile trend and observing "unusual" energy conditions during each year, such as school weeks, public holidays and so on. For instance, if there was an "unusually" low energy demand during school week noted with a negative M_i in a series, it might indicate the building operation followed the low attendance. Contrarily, if a high positive M_i in a series, a high energy use might be needed.

2.4. Temperature moving average and energy signature curve

Energy demand for DHW is much less sensitive to outdoor temperature than SH, and its use is relatively stable throughout a year. According to the report from NVE, the energy demand for the annual DHW in schools is less than 6% of their total heating needs (SH + DHW) [13]. The DH meter in the platform did not have submeters for monitoring DWH and SH, separately, thus energy for DHW was not excluded in the heating analysis. Unlike other building types such as nursing home, where DHW may account for over 20% of total heating needs and shall be exclusively studied, such as the work by Ivanko et al. [32].

Besides thermal inertia of the piping system, by considering the impact of building thermal inertia, the concept of a temperature moving average (TMA) was introduced to define a more accurate mathematical relation between outdoor temperature and DH demand. Depending on building physics, there is a big variety of capacity for temporary heat storage. It is simply saying that better insulation plus internal heat gains may yield longer time lag. The practices of considering TMA have been addressed in literature [33,34], and it is proven that the time lag hour shall be estimated as per the building physical characteristics [33]. The empirical time lag hour such as 24 or 48 h [33,34] was initially used as the reference range, and extended to 55 h. Within the range, the outdoor temperature was shifted backward by each hour to find the highest correlation between the outdoor temperature and the DH demand and to get accurate model. When the correlation has a higher absolute value, it implies a better fit between the moved outdoor temperature by TMA and the DH demand. Fig. 3a) compares the temperature lag with the highest correlation of each building. Fig. 3b) presents the effect of temperature lag on the DH based on the 4-year modified average hourly values. It can be found that the lag of five hours yielded the highest correlation for the building group in this study. Therefore, the outdoor temperature of five hours ago at each time clock was used to identify the relationship between outdoor temperature and DH demand. Meanwhile, there was a second peak noted at 29 h, implying the outdoor temperature of 29 h ago might also have a good correlation with DH demand. Similar finding was at 53 h. This may be explained that the outdoor temperature at the same time clock between neighboring days are usually similar. However, these correlations were not as high as at five hours. This differed from both the empirical number and the lag of 14 h found in [33], where a low-energy building of better building physics was analyzed.

Instead of constructing a physical model with the detailed building input data and parameters, data-driven approach defines building energy demand through statistical relationships. Among the data-driven approaches, there are regression analysis and advanced techniques such as artificial neural networks and decision trees by requiring long computation time and sophisticated knowledge. One of the important applications of linear regression is the energy signature curve (ES curve), which is a function of the outdoor temperature to predict heating energy demand [35–37]. It has been applied to fruitful research and is welcomed by utility companies. From the ES curve model, the DH use is usually described by the composition of two parts, temperature-dependent and temperature-independent. These two parts are divided by heating effective temperature or changing point temperature (CPT). The ES curve models may be expressed as:

$$\text{If } t_{ot} \le \text{CPT},\ P(t_{ot}) = p_1 \cdot t_{ot} + p_2 + \varepsilon \tag{3}$$

$$\text{If } t_{ot} > \text{CPT},\ P(t_{ot}) = p_1 \cdot t_{ot} + p_2 + \varepsilon; \approx p_2 \tag{4}$$

In Eqs. (3) and (4), p_1 and p_2 are the coeffiients of each ES curve model. p_1 denotes the slope, p_2 denotes the intercept, and ε is the residual error. The regression model of $P(t_{ot})$ as a function of t_{ot} may be used to estimate heat use. When the outdoor temperature is above the CPT, the heating demand is usually either at a small and constant volume or under a mild slope with the outdoor temperature. To solve the equations above, the ordinary least squares method is traditionally adopted. The coefficients are determined through the aim that the error between predicted and observed values is minimized. The higher the R^2, the better the model fits the data. Technically, R^2 shall not be less than 0.75 for a satisfying model in accordance with ASHRAE [38,39].

Boxplot method can reveal descriptive statistics about dispersion [30]. As illustrated in Fig. 4, these boxplots for the daily DH profiles were used to separate working and off-working hours, in order to segment ES curve models with each operational period. The observed DH data were assigned into four groups by considering seasonal and operational issues. Group I was for January, February, March, and December. Group II was for April, October, and November. Group III was for May, first-half June, second-half August, and September. Group IV was for the whole July. Fig. 4 shows the considerable heat variation between the working and off-working hours, the weekdays and weekends (without other short or long holidays). It also shows the heating loads during nights and weekends were close. From the boxplots, the DH system was generally under operation from 6 to 18 o'clock on weekdays. This is similar as depicted in [15], where the schools' daily power use typically rises at 6 and fall at 17 o'clock. It is a common practice that building operators start the system early before occupants arrive, and reduce the heat supply after they leave, to provide satisfying thermal comfort. Ma et. al mentioned the extended heating hours for coping with the complaints of freezing mornings [40].

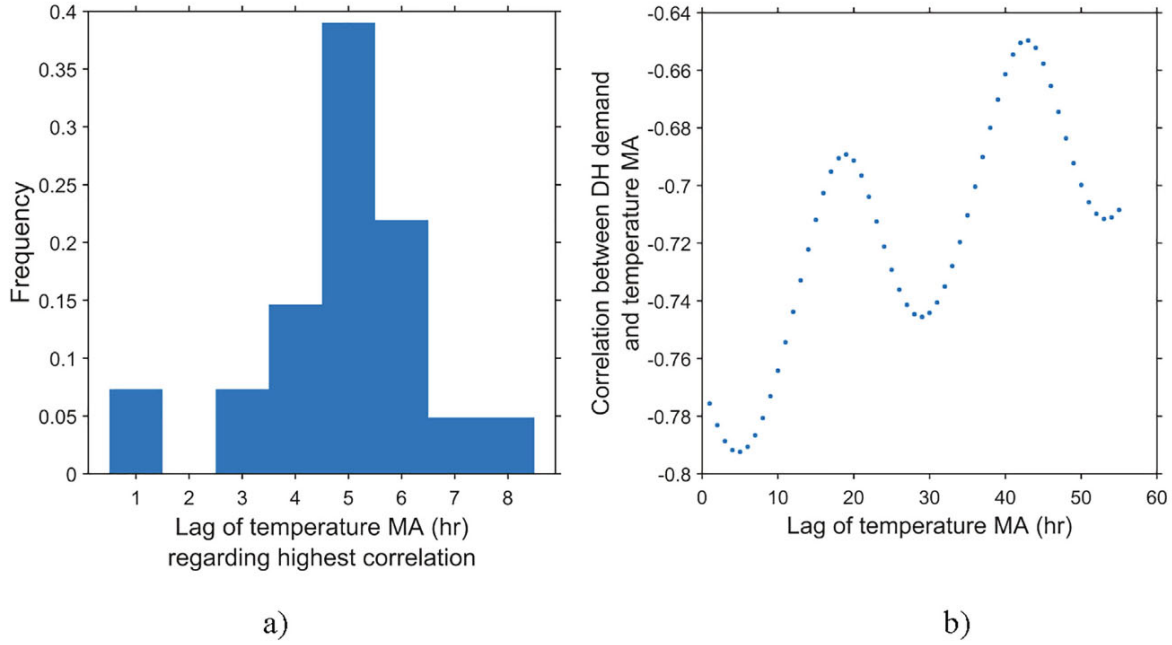

Fig. 3. a) Distribution of temperature lag regarding highest correlation of each building, where 5-hour lag had the highest frequency among the building group; b) Temperature lag moving average based on the modified average DH demand, where 5-hour lag yielded the highest correlation.

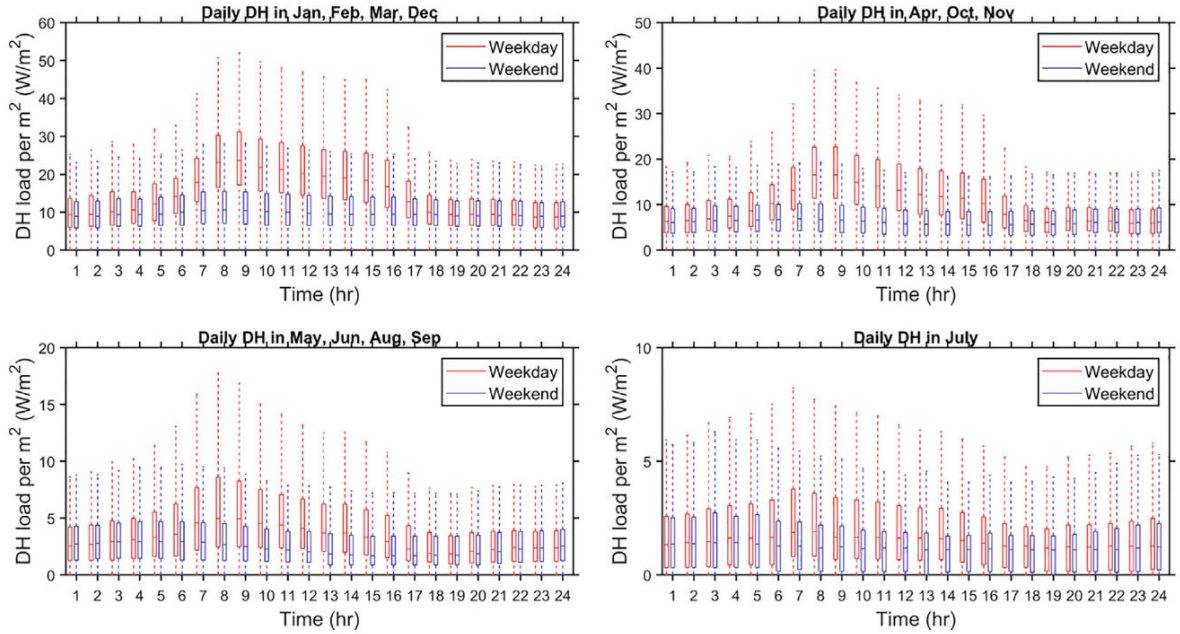

Fig. 4. Boxplots for the daily DH profiles on weekdays and weekends during four groups, from top left to bottom right, they are Group I, Group II, Group III, and Group IV.

The minor heat fluctuation of demand in July reflected the minimum water flow circulating in the piping system of the DHW system when the school was normally closed for the summer holiday.

2.5. Typical meteorological year and prediction of DH profile

The importance of typical meteorological year (TMY) that allows estimation of long term performance from a single year analysis is elaborated in literature [41]. From the European Union website, TMY 2007–2016 for Trondheim was retrieved and used for heating prediction in this study [42]. The calendar of 2025 in Norway was used as a reference year for acquiring information on public holidays, and the methods can be adopted to calendars of other years.

2.6. Electricity profile

It is acknowledged that each year starts and ends with different day numbers. For example, it started with Friday in 2016 and with Monday in 2018. Therefore, the 4-year similarity regarding the electricity use was examined with the correlation for every two years' electricity profile between the week 1 and the week 52 (only week 2 to week 53 for 2015). The remaining 1 or 2 days were discarded, the correlation plot is shown in Fig. 5. The high correlation numbers within every two years imply the similar electricity use

Fig. 5. Correlation of electricity profile during 2015–2018 (52 weeks).

during the four years. The small deviations can be explained by the situation that study trips, activities, and public holidays were arranged on different dates and weeks in each year. After the modified Z-Score mentioned in Section 2.3, typical profiles for normal days and special days can be separately identified. The 4-year value was treated as the predictor and their average value was treated as the response, and they were together trained in the linear regression learner. Additionally, cross-validation was selected for increasing the accuracy of the final model by using the full data. Then these typical profiles can be combined to extrapolate and estimate the demand for future years.

2.7. Validation procedures

To prove that the predicted energy profiles are convincing, an evaluation process is needed for measure and verification. In this study, all the profile models were verified through three mostly-used quality criteria as the following: mean absolute percentage error (MAPE), normalized mean bias error (NMBE), and coefficient of variation of the root mean squared error (CV(RMSE)). If the model meets the requirements of all the three criteria, it may be regarded as a satisfying model. If the model fails to one of the criteria, it is further checked with the cluster methods of piecewise aggregate approximation (PAA) and symbolic aggregate approximation (SAX). The explanation and validation results of these indices are presented in Section 5.1 and 5.2.

3. Results of modified Z-Score

The daily DH and electricity demand on weekdays are shown in Fig. 6a) and Fig. 6b), separately, which was made based on the equal division of every year's 52 weeks into four seasons. It can be seen that there were considerable seasonal energy demand variations and different requirements for DH and electricity. The red and pink boxes for the winter and early spring required very high energy demand, while the demand in the green boxes was lowest and the blue boxes in between. The DH and electricity use was then analyzed on seasonal and weekly base, separately, see Section 3.1 and 3.2.

3.1. Analysis of DH use

The weekly DH demand based on the modified average DH demand from 2015 to 2018 is shown in Fig. 7a), and the modified Z-Scores within each season are given in Fig. 7b). A summary of the seasonal Z-Scores is given in Appendix Table A1, and the abnormal DH use with $|Z| \geq 3.5$ are highlighted with numbers.

From the weekly DH aspect, most of the data were within the threshold of ± 3.5. The high DH demand marked with No. 1, 2, 3 and 4 was mainly due to the extremely cold condition. From Fig. 7a) and b), it was obvious to identify seasonality, but it was not clear to detect the breaking points of weekly breaks such as the school week, Easter week, when the DH demand was expected to be low during the temporary low attendance. There was only No. 6 of 2018 found with very low DH demand even though the weekly coldest temperature was −11.1 °C.[2]

As for the heating operation of short and one-day holiday, the analysis was performed within the week by comparing with the adjacent days. This concerned that heat demand was dependent on the outdoor temperature, while the outdoor temperature difference between neighboring days were not big. Thus, in Fig. 8 there were five days in each line, and their modified Z-Scores are given in Appendix Table A2, where the holidays were marked with numbers. Most of the DH demand was under the upper threshold, only No.6 in 2017 had an abnormally high demand when it was supposed to be closed for Whit Monday. There were only No. 1 and 3 in 2015, No. 1 in 2016, No. 1, 4 and 5 in 2017, and No. 1 and 6 in 2018 followed the expectation of the low DH demand on public holidays, although some of them were not as low as to achieve the lower threshold by comparing with neighboring days.

3.2. Analysis of electricity use

The 4-year weekly electricity is given in Fig. 9a). The corresponding modified Z-Scores during each season are given in Fig. 9b), and the breaking points are listed in Appendix Table A3. Although the electricity demand also had seasonal variation, the demand was much lower than DH, especially in early spring and winter.

By comparing with the DH use, there was no weekly electricity demand beyond the upper threshold, moreover, it was straightforward to distinguish the breaking points where weekly low demand occurred on holidays.

As suggested in [14], for the days having similar occupancies as their previous days, the previous days can be selected as their reference days. For example, the previous Tuesday is selected as the reference day of this Tuesday due to their similar occupancy periods for electric chiller operation management [14]. Hence, regarding the electricity demand during a short and one-day holiday, the analysis was made by comparing the same weekday number within each season. Since the electricity use was less sensitive to the outdoor temperature, it was assumed that the electricity use followed weekday schedule in each season. For example, all the Mondays in season 1 were assumed having similar electricity use if there were no holidays. Accordingly, in Fig. 10 each line presents 13 points of the same weekday. Except for the abnormally high use of No. 4 in 2015, the other electricity demand was under the upper threshold line. In contrast to the DH use, all the public holidays were detected with the low electricity demand as the "outliers" or the local minimum. Besides that, when the one-day holiday was on Thursday, it would be probably that the following Friday was also on holiday or had reduced school hours. For example, No. 7 in 2015, No. 5 in 2016, and No. 6 in 2017 were found with a local minimum, see Fig. 10. The corresponding modified Z-Scores are given in Appendix Table A4.

3.3. Similarities and dissimilarities from the modified Z-Scores on DH and electricity

Based on the weekly energy use diagrams of the DH and the electricity use above, it was found that both Spring and Autumn

[2] Week 2 to week 53 of 2015 was in the analysis, the week number of 2015 in Fig. 7a) and b) shall be plus 1 accordingly.

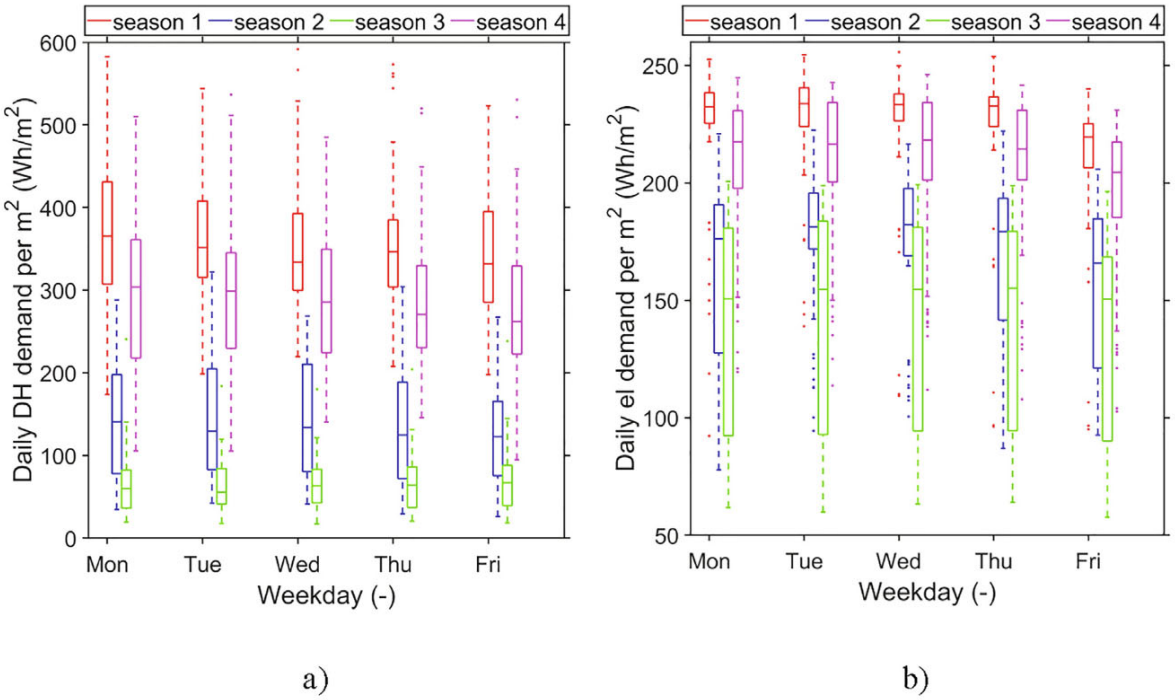

Fig. 6. Boxplots of energy use on weekdays, where a) DH use, b) electricity use. The red box refers to season 1 (week 1–13), the blue box refers to season 2 (week 14–26), the green box refers to season 3 (week 27–39), and the pink box refers to season 4 (week 40–52). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

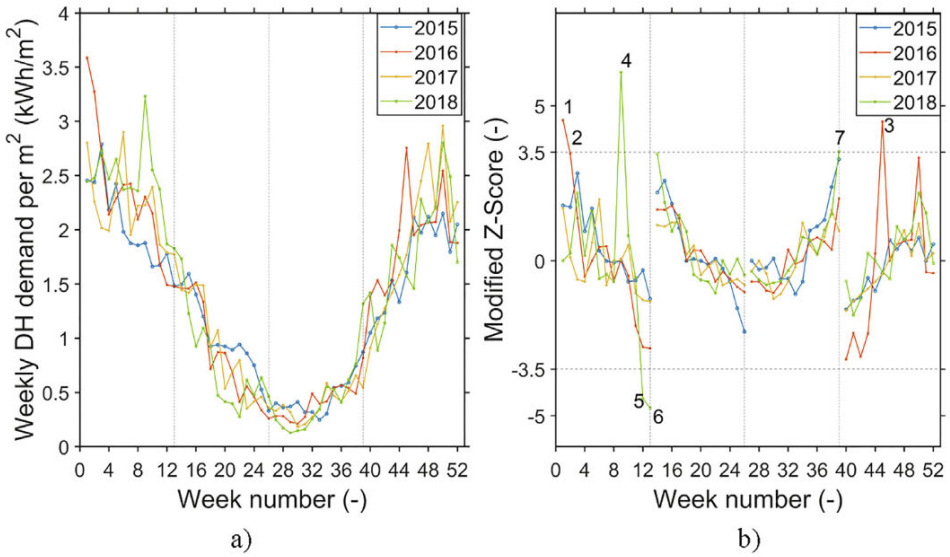

Fig. 7. a) Weekly DH use 2015–2018, b) Modified Z-Score I of weekly DH use during 2015–2018.

school week were arranged in the same week number during the four years, week 8 and week 41, respectively, while the Easter weeks and other public holidays were of different week numbers, causing various energy impacts on the neighboring weeks. The school closure for summer vacation was approximately arranged from week 26 to week 32, and the Christmas holiday during week 51 and week 52.

There was no extremely high electricity use generally, but pit low use during the breaks. Season 1 and Season 2 had more "outliers" since most of the holidays were in the first half of the year, see Fig. 9a) where there were considerable variations of weekly electricity use, and Fig. 9b) where there were six out of eight "unusually" low demand. However, the modified Z-Scores for the DH demand could not disclose the holiday breaks as in the case of the electricity use, as shown in Fig. 7a) and b) where most of the points were not below −3.5.

Based on the above introductory analyses, it can be concluded that the control of electrical appliances in the school made reasonably fast responses by closely following the attendance and schedule. Nevertheless, the response of hydronic DH systems was a relatively slow process mainly due to the long transport of heating fluid and complex control of the DH sub-stations, as mentioned by others [36,43]. From the modified Z-Scores, it was most likely that the heating operators in the school set mode between weekdays and weekends, and long holiday mode for Summer vacation and Christmas week, without disturbing for

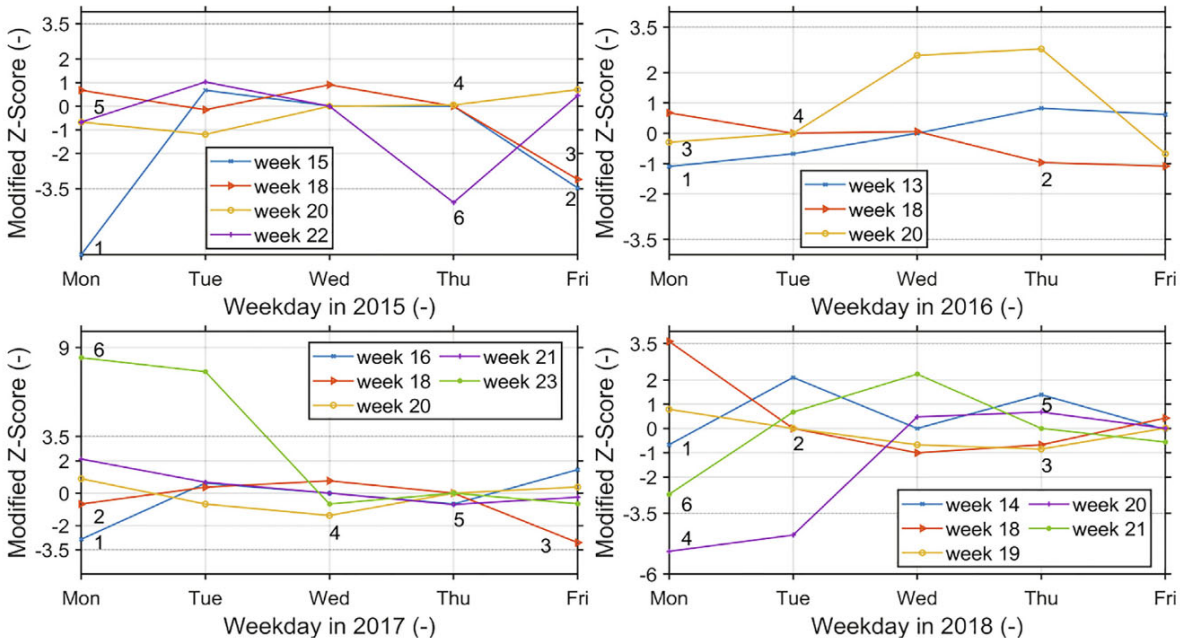

Fig. 8. Modified Z-Score I of DH use regarding short holidays during 2015–2018.

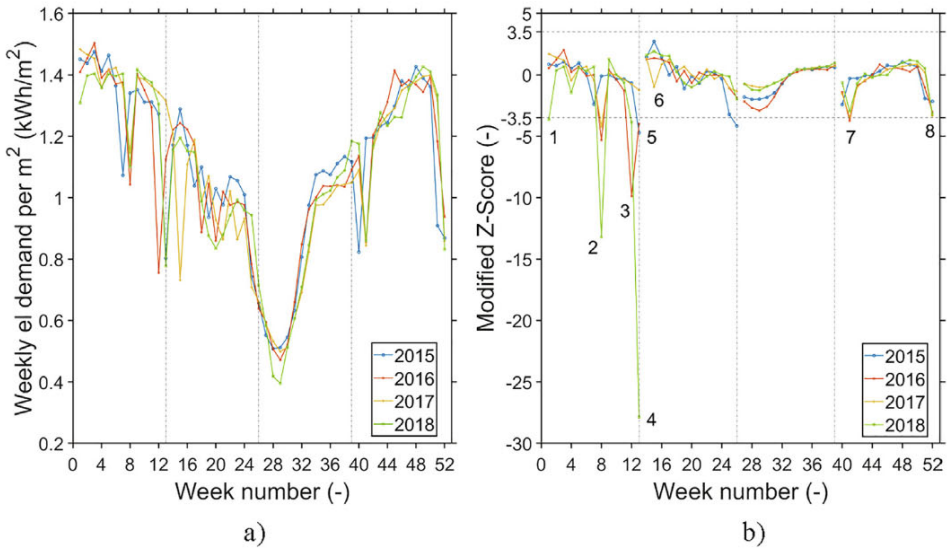

Fig. 9. a) Weekly electricity use during 2015–2018, b) Modified Z-Score I of weekly electricity use during 2015–2018.

the short public holidays. This implied that the DH demand was more likely prone to the outdoor temperature over the schedule on short breaks.

4. Results on energy prediction

4.1. Results on ES curve models and prediction of DH profiles

The ES curve for DH demand is illustrated in Fig. 11, where the CPT was found at 13 °C by giving the adequate piece-wise approximation for an average school building. This CPT is in the range of typical average national threshold temperatures for the current European building stock, which vary between 10 and 15 °C [36]. In the area below 13 °C, it was the temperature-dependent DH part. In Fig. 11, the red dots represent the stable working hour from 8 to 16 o'clock, the green dots represent the ramping hour at 6, 7 17, and 18 o'clock, and the blue dots represent the non-working hour. The non-working hours included weekends and nights on weekdays, since these two parts had close demand as mentioned in Section 2.4. In the area above the CPT, the light blue dots only covered a small share of the DH demand.

Fig. 11 shows the strong differences in energy demand among the four parts, from dominating high to nearly negligible heating use. The coefficients of Eqs. (3) and (4), and R^2 of each part are given in Table 2. In the outdoor temperature dependent area, only the model of ramp period showed a bit weaker fitting and the other two models were capable to predict the DH demand well. In the area above CPT, the heating needs were trivial and had minor impact on energy supply system. Fig. 12 is the logistic diagram by giving the relevant coefficients under different conditions. Then by inserting the weather data of TMY (see Section 2.5) and the coeffects into Eqs. (3) and (4), the typical annual hourly DH load can be predicted. The outdoor temperatures in the analysis were

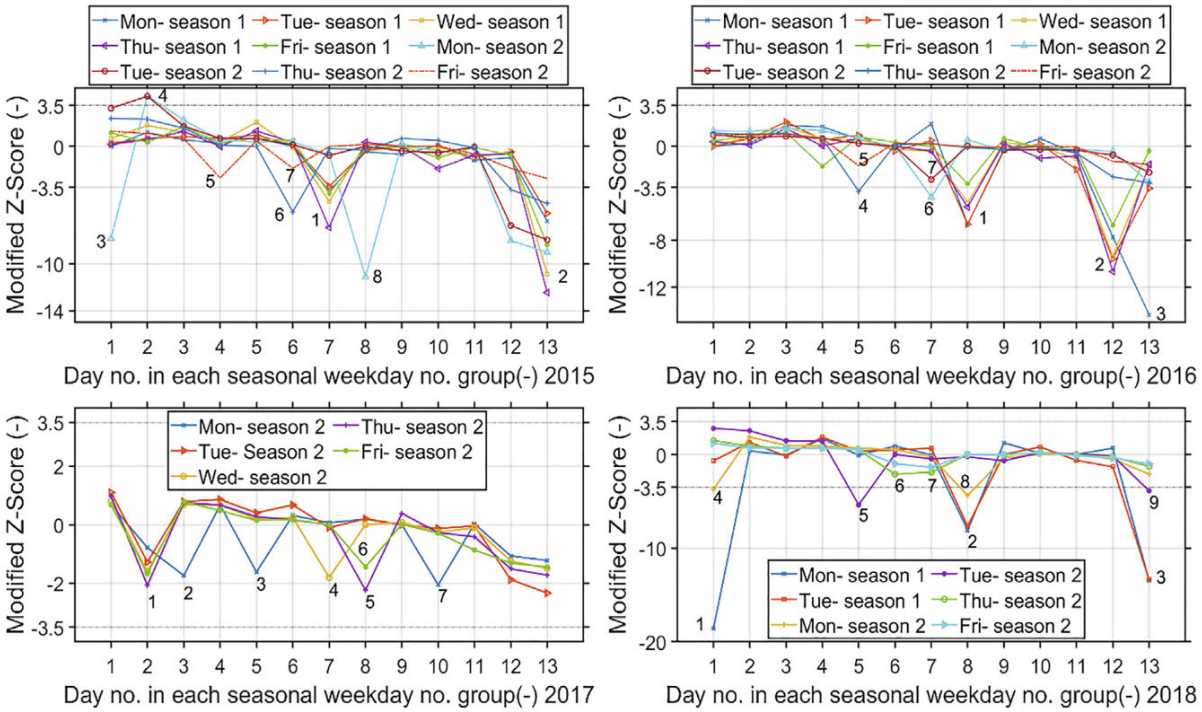

Fig. 10. Modified Z-Score I of electricity use regarding short holidays during 2015–2018.

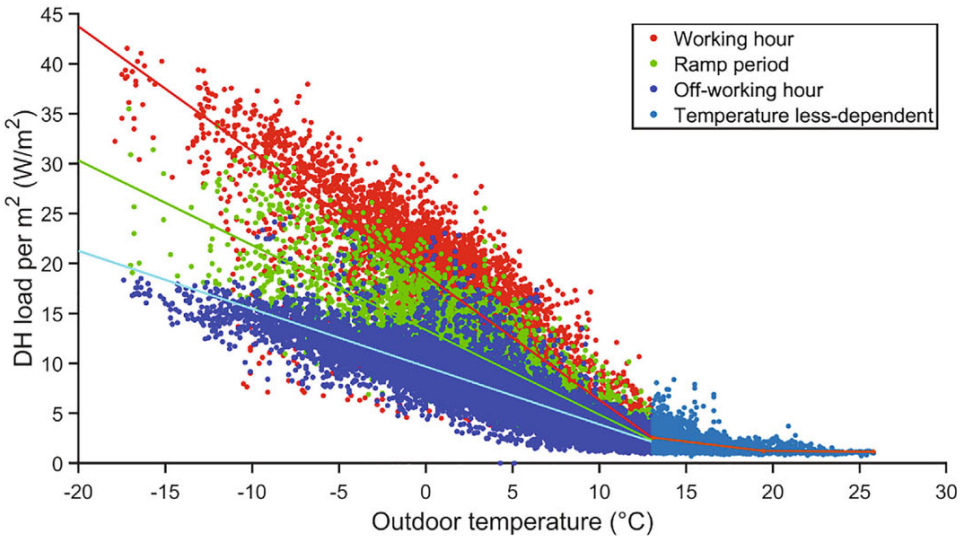

Fig. 11. Energy signature curve models of DH demand.

Table 2
Coefficients of Eq.(3) and Eq.(4), and the corresponding R^2.

	Outdoor temperature dependent ≤ 13 °C			Outdoor temperature less dependent	
	Working hour	Ramp period	Off-working hour	(13, 20 °C]	>20 °C
p_1	−1.2	−0.85	−0.58	−0.23	/
p_2	18.3	13.4	9.5	5.6	1.2
R^2	0.76	0.73 (↓)	0.78	0.35 (↓)	

between −18 and 26 °C, covering the cold design temperature of several major Nordic cities, such as Stockholm (−18 °C), Copenhagen (−11 °C), Gothenburg (−17 °C), etc. [36].

Finally, the monitored and the predicted annual profiles for the DH demand from 2015 to 2018 are compared in Fig. 13, where the purple lines present for the monitored profiles and the green lines present the predicted profiles. Meanwhile, the predicted typical annual DH load profile in TMY is presented in Fig. 14. Owing to the large dependence on the outdoor temperature, the heat load fluctuated throughout the year with oscillation of peaks. The peak load was around 48 W/m^2 and the minimum load was close to 1 W/m^2, and the total annual heat demand was 72 kW h/m^2.

4.2. Prediction of electricity profiles

Fig. 15 shows the 4-year monitored and predicted annual profiles of electricity. The comparison plots were of the same color pattern as in Fig. 13. There was one hour at around 2000-hour with

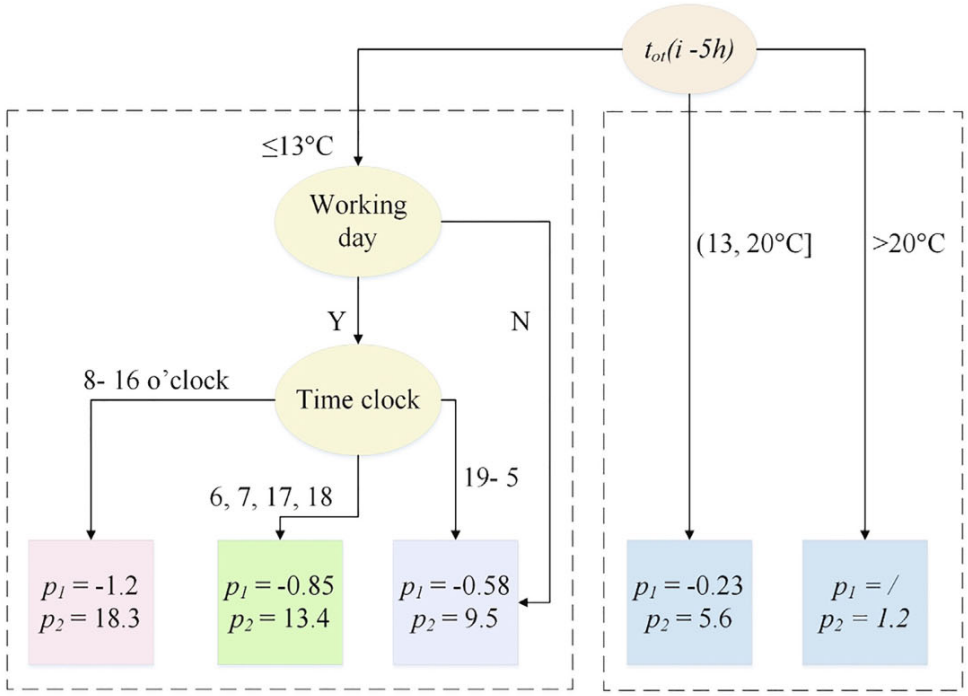

Fig. 12. Logistic diagram of predicting DH demand under different conditions.

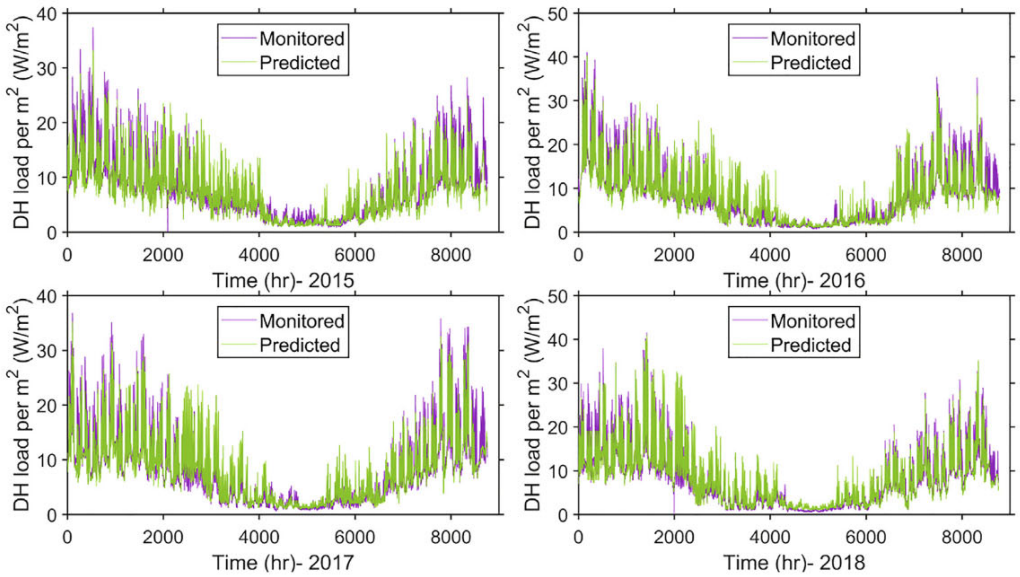

Fig. 13. Monitored vs. predicted DH profile during 2015–2018.

a monitoring failure, and this was fixed in the prediction model. There were 3.2–8.6% of predicted load differed from the monitored load by 30–40% in the four years, and most of the predicted load were close to the monitored load. The defined typical hourly profiles for school week, Easter week, short holiday, and the remaining normal days (without the above special days) are shown in Fig. 16.

The predicted typical annual electricity load profile is presented in Fig. 17, which showed the relatively stable pattern of electricity. The peak load was around 18 W/m^2 and the minimum load was 2 W/m^2, and the total annual demand was 57 kWh/m^2. The minimum 2 or 3 W/m^2 electricity load was mostly used for some plug-in equipment and low ventilation in meeting the air quality requirement during the unoccupied period. Additionally, it is seen as a conventional custom in the Nordic region that a few lights in the main entrance or hallways are kept on after a school is closed. The peak load for electricity was only one-third of the peak DH load (48 W/m^2), and the total demand was 79% of the DH demand (72 kWh/(m^2·yr)). Hence, it remarkably reduced the strains of the power grid for the buildings with DH comparing with the buildings those were solely supplied with electricity. It was of significance especially in the winter season when both heating and electricity called for high energy supply. Moreover, by analyzing heat and electricity profile separately, it presents the different requirements for thermal and power grid in relation to sizing and production.

The total specific energy demand for the observed average school was 129 kW h/(m^2·yr), with nearly 56% for heating needs. The energy share for heating purpose was almost the same as the average situation [13]. The total energy demand was slightly

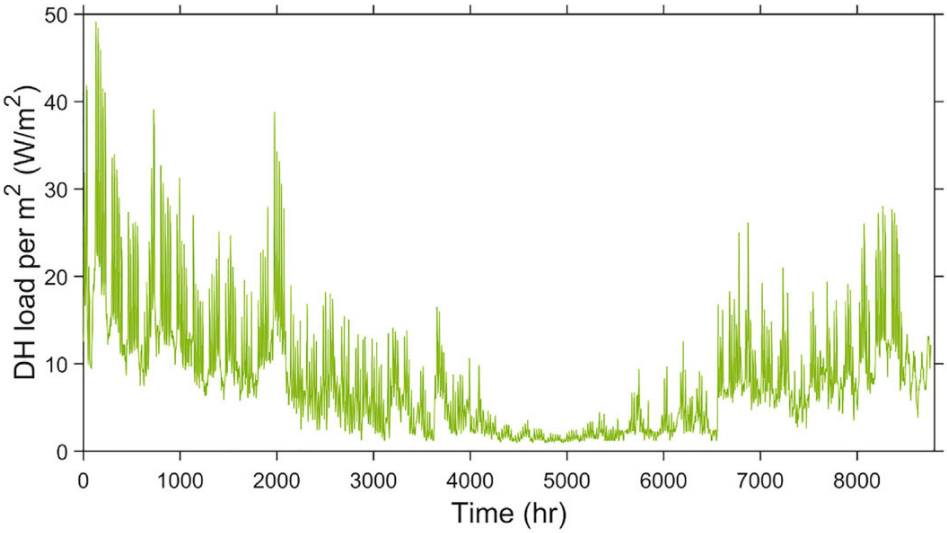

Fig. 14. Predicted typical annual DH load profile.

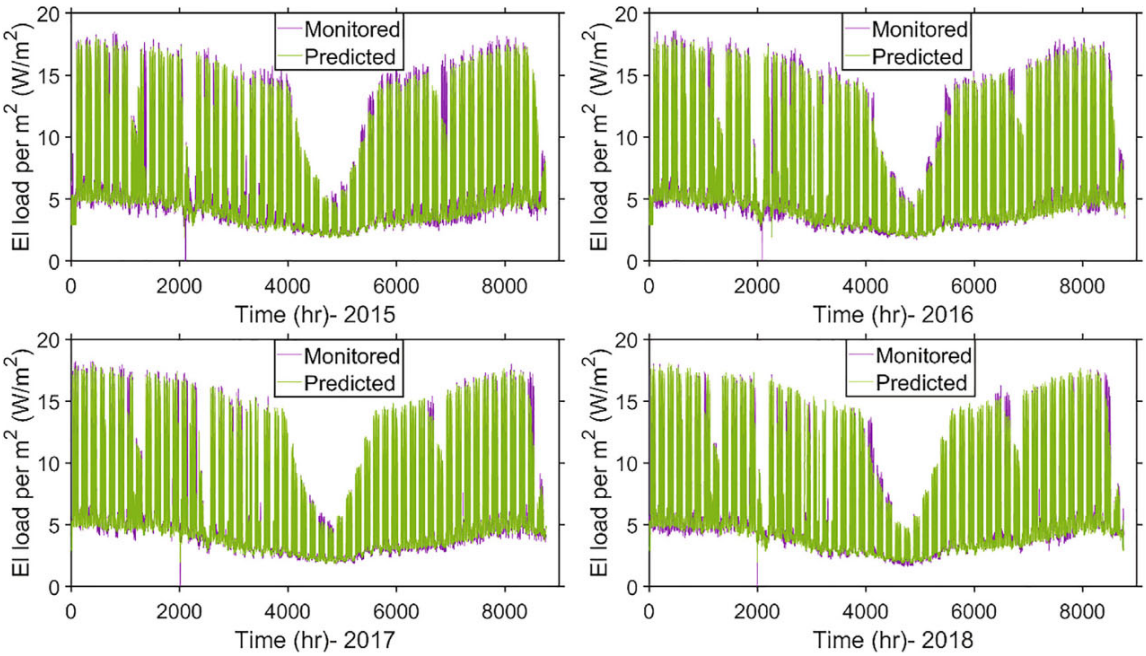

Fig. 15. Monitored vs. predicted electricity profile during 2015–2018.

lower than the annual average energy use in Norwegian schools [13,44], but same as the mean value of Swedish schools according to the Energy Statistics [24]. This predicted demand was also approaching to the proposed nZEB energy performance target level for Finnish educational buildings in FInZEB project, 104 kW h/(m^2·yr) [45].

5. Validation results

5.1. Criteria results of MAPE, NMBE, and CV(RMSE)

After the regression analysis, the deviations between the predicted and observed profiles were examined through the verification process with MAPE, NMBE, and CV(RMSE). If the verification results are within the recommended range, it means the prediction methods are reliable to be used for future modelling.

MAPE means the average error between the actual and predicted data to the actual data, and it is usually expressed in relative numbers. The expression of MAPE is given as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_i - F_i}{A_i}\right| \cdot 100\% \quad (5)$$

where A_i is the actual and monitored value, F_i is the predicted value, and n is the number of the observations. The absolute value of MAPE avoids the possible offset among positive and negative errors. MAPE is used as a common measure in the forecast of wide areas such as finance, business, energy sectors and so on [18,46,47]. Table 3 lists the prediction quality when using MAPE criterion, where the MAPE result is recommended less than 20% to verify an accurate model [47].

NMBE calculates the total percentage of the error during the evaluation period and this criterion is given as:

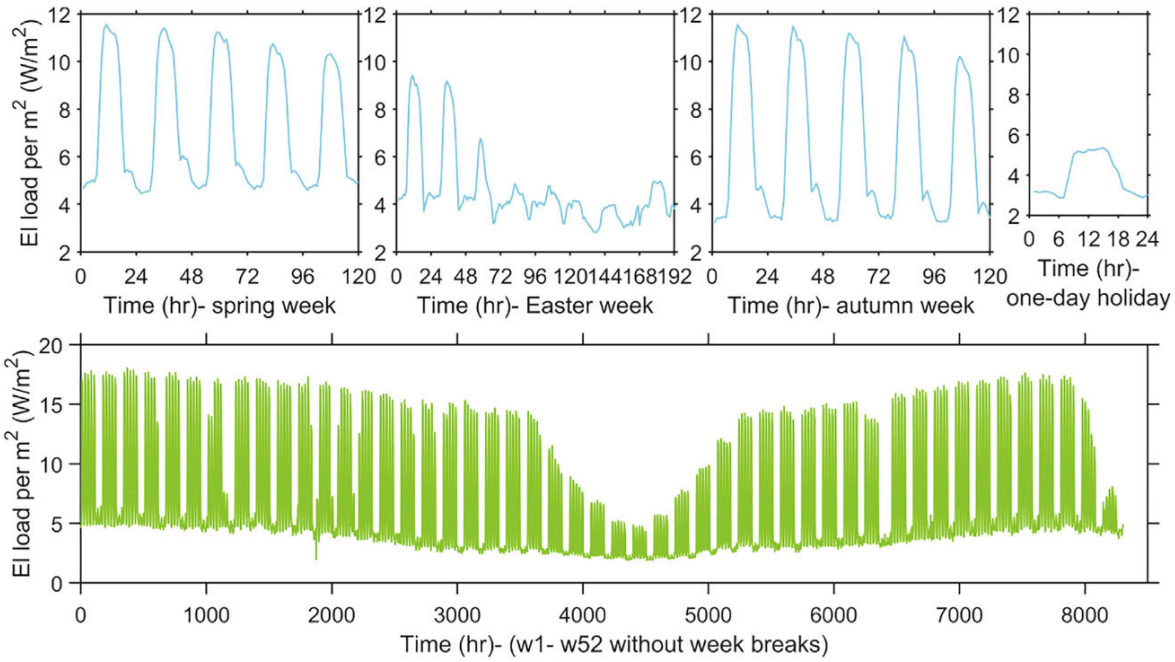

Fig. 16. Typical hourly profiles for school week, Easter week, short holiday, and the normal days without these special days; for easy reading, the profiles for Easter, autumn week, and one-day holiday have the same Y-axis title of spring week.

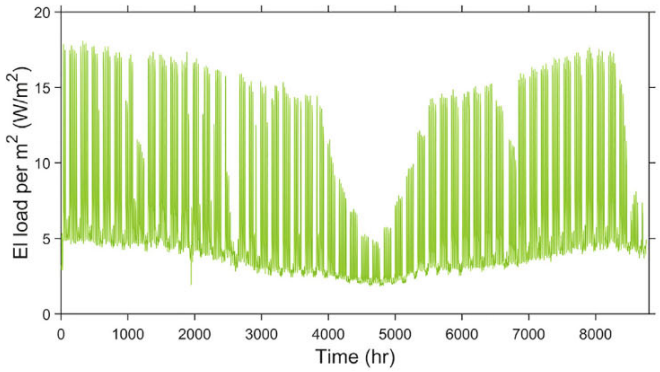

Fig. 17. Predicted typical annual electricity load profile.

$$NMBE = \frac{1}{n}\frac{\sum_{i=1}^{n}(A_i - F_i)}{\bar{A}} \cdot 100\% \quad (6)$$

where $\bar{A}$ refers to the average value of the monitored data, the other denotations are the same as for the MAPE. If a negative result is obtained for the NMBE, it implies that the energy demand is over-predicted, on the contrary, an under-prediction is made. The directionality of the NMBE shows the difference between the actual and predicted use [46,48].

The Root Mean Squared Error (RMSE) assesses the mean squared error, and CV(RMSE) normalizes the RMSE with the average energy demand during the evaluation time [46,48]. CV(RMSE) is given as:

Table 3
MAPE criterion of evaluating forecast quality.

MAPE (%)	Forecast quality
<10	Highly accurate forecasting
10–20	Good forecasting
20–50	Reasonable forecasting
>50	Inaccurate forecasting

$$CV(RMSE) = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2}}{\bar{A}} \cdot 100\% \quad (7)$$

CV(RMSE) indicates whether the forecast model can reflect the real load shape. The two metrics, NMBE and CV are commonly used together to find out prediction performance. Under ASHRAE Guideline 14, the hourly criteria of NMBE and CV are limited within ± 10% and 30% respectively for verifying a satisfying model [38,39].

As listed in Table 4, all the MAPE results regarding the electricity demand were less than 10%, NMBE were with ±1%, and CV were less than 20%, which meant the forecast of the electricity profile was of high accuracy. Accordingly, it was reliable to estimate the future profile for electricity by extrapolating the regression results with adjustment of the calendar considering such as school weeks, public holidays, etc., as mentioned in Section 3.2. However, from the validation results on the DH profile that MAPEs were higher than 20% and CV in 2018 was slightly beyond the ASHRAE criterion, it can be said that the ES curve model was to some extent convincing but not very accurate. Further validation was thus needed for the DH profile.

5.2. Discretization results for PAA and SAX

Since the prediction of the electricity profile passed via all the three quality criteria, only the predicted DH profile was further checked with piecewise aggregate approximation (PAA) and symbolic aggregate approximation (SAX). In this study, weekdays from the predicted typical annual DH profile were extracted to compare with the 4-year monitored weekdays on Winter, Spring and Autumn, and Summer, respectively. Each weekday (24-h) was treated as a time-series.

First, Fig. 18 illustrates the comparison of the daily DH profiles during weekdays between the predicted profile (shown in Fig. 14) and the 4-year monitored profiles. In Fig. 18, the red lines show the Winter season, the greens line show the Spring and Autumn season, and the blue lines show the Summer season. In addition, the thick lines depict the predicted profiles, while the thin dashed lines describe the monitored profiles. It was noted that there was a peak

Table 4
Evaluation results of the energy forecast by three criteria.

Year	prediction of DH load profile			prediction of electricity load profile		
	MAPE (%)	NMBE (%)	CV(RMSE) (%)	MAPE (%)	NMBE (%)	CV(RMSE) (%)
2015	20.2	2.9	25.7	8.8	0.9	16.0
2016	23.2	−1.4	26.7	6.8	0.2	10.6
2017	24.0	−2.4	26.9	6.6	−0.3	11.7
2018	29.6	−4.9	30.2 (↑)	7.5	−0.4	10.1

load arising at 9 o'clock. After manually adding 2 W/m^2 at 9 o'clock during the heating season (thick dotted line), the predicted profile was seen closer to the observed one for the Winter season. The peak heat load at 9 o'clock was also seen in the average weekly heat load patterns of public administration buildings in [49]. However, since the outdoor temperature in the measured years and TMY were not same, the predicted profile was not anticipated to be the same as the measured ones.

Second, during the approximation process of PAA, the predicted data in the time-series were initially Z-normalized with $\frac{C_i-\mu}{\sigma}$, where μ and σ referred to the mean value and the standard deviation of the time-series, respectively. More specifically, in this study the μ and σ were the daily average DH load and daily standard deviation, respectively; and the daily DH load were normalized to [−2, 2]. It is an essential step allowing the mining algorithm to focus on the patterns' similarities/dissimilarities instead of on the data amplitudes in the time-series. This normalization step is different from the modified Z-score in Section 2.3. Next, the Z normalized time-series $C = C_1, C_2, \cdots, C_n$, proceeded through the PAA. PAA is one of the algorithms that is designed to reduce the dimensionality of the raw time-series, and the basic idea is to split them into equally sized intervals [50,51]. Each interval is computed by averaging the values within the interval. The time-series C is represented by a dimensionally reduced new series as $\overline{C} = \overline{C_1}, \overline{C_2}, \cdots, \overline{C_w} (w \leq n)$, w is the number of data in the new series. Theoretically, if w is close to n, it has a high accuracy of data representation, however, it would lose the meaning of dimensionality reduction from the PAA. On the contrary, if w is very small with for example only one or two intervals, it would hardly make a sounding representation of the original dataset. Typically, it has $w \ll n$. The ith element of new time-series $\overline{C}$, or say the mean value of the data falling within the ith interval, is calculated as Eq.(8) [51]:

$$\overline{C_i} = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} C_j \tag{8}$$

Accordingly, in this study the 24-hour data (n) of each weekday was split into 8 (w) equally sized segments in the new time-series. The representation was aimed at approximating the raw 24-hour time-series by a linear combination of 8 box functions.

Finally, after deciding the value of w that suits the dataset, the PAA coefficients are assigned with a string representation graph through symbolization process of SAX. As recommended in literature [51], three breakpoints (−0.67, 0, 0.67) were chosen here. The PAA coefficients for those below the −0.67 were symbolized with the string "a", the coefficients between −0.67 and 0 were with "b", the coefficients between 0 and 0.67 were with "c", and those higher than 0.67 were with "d". In each interval, if the compared data from different datasets have the same SAX strings, they may be clustered as the same group and the PAA coefficients do not have to be exactly same.

The comparison of the PAA coefficients between the predicted and the monitored DH daily profiles is shown in Fig. 19, and the corresponding SAX symbols are listed in Table 5. The PAA coefficients and the SAX symbols of the predicted profile were calculated based on the thick dotted line in Fig. 18. In the SAX Table, those had different SAX strings within the same PAA interval were marked with the italic font, implying they cannot be clustered as the same group. It can be concluded that the predicted and monitored load profile for the whole Winter had very high similarity, since all the eight intervals between predicted and monitored load profile on each weekday had same clustering strings. And Winter also consumed most of the heating energy. However, the similarities for the other seasons were not as strong as Winter, but there were still more than half belonged to the same group within the same

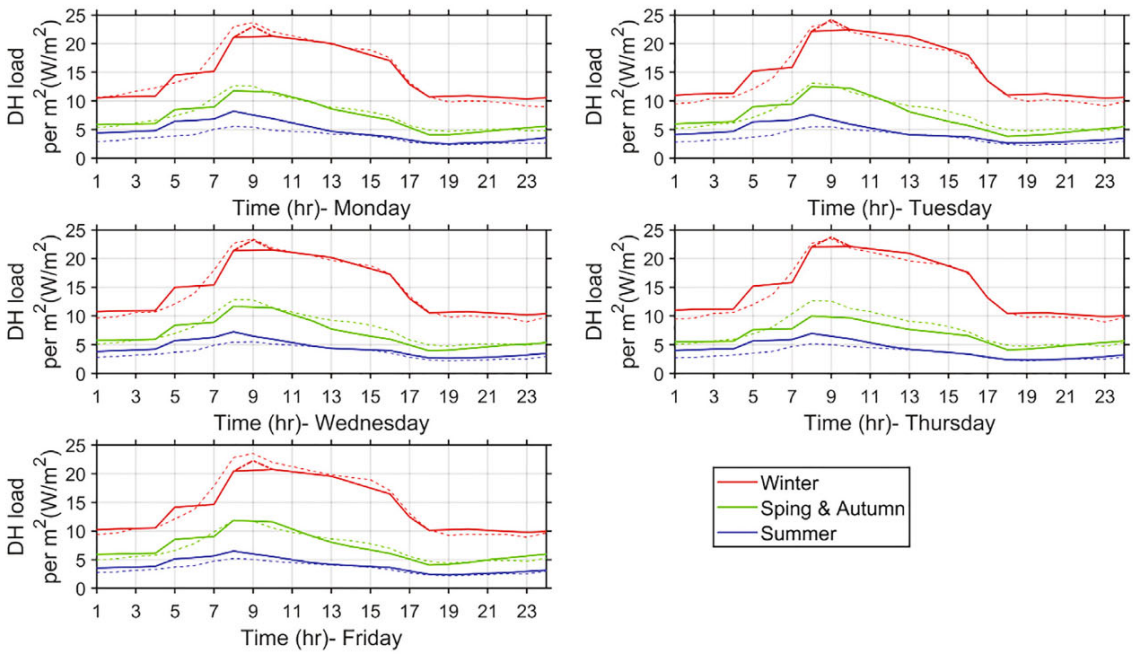

Fig. 18. Comparison between the predicted and monitored DH profiles.

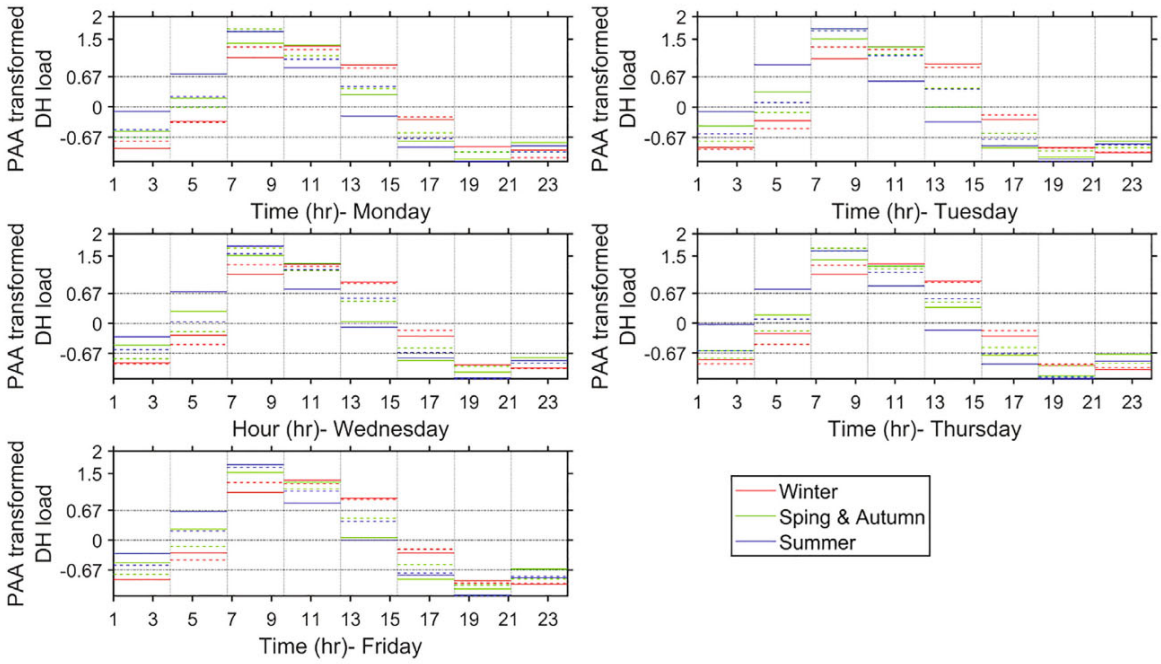

Fig. 19. PAA coefficients results of the comparison related to Fig. 18.

interval, (23 + 29) out of (40 + 40). The relatively large deviation and error during the Spring and Autumn seasons might be explained by the changing seasons with unstable outdoor temperatures, which might cause larger temperature differences. From the approximation of the PAA and the SAX symbol results, the predicted DH profile for the average school was proved convincing and representative. Besides the advantages of using PAA for time-series discretization and amplitude normalization, it also enabled us to present seasonal load patterns and shape comparisons through extracting the required day types. For example, from the transferred SAX strings, it shows the daily heat load varies in all the seasons with peak intervals "d", valley intervals "a", and transition intervals "b" and "c", even when the heat load was low in Summer. Miller et al. praised the benefits of the PAA and SAX, which may accelerate clustering building performance for building commissioning and fault tracking in response to the rapid increase of building data amount [52].

6. Application and limitations

The suggested analysis method for the ES curves and identified load profiles could be used as a reference example for stating energy use of other building types in the Nordic climate. An artificial urban area may be aggregated and synthesized by involving different building types. The method may be used as input data for modelling energy supply optimization.

Electricity has been traded freely in the European market after the regulation being removed since the 1990s, and the CO_2 emission of electricity counts on market production. To estimate the annual CO_2 emission and economic cost of electricity use in an urban area, there are two approaches to follow. The first one is to use the annual energy demand multiplied with the annual average CO_2 factor and average electricity price. For example, within the Nordic region, the CO_2 factor is approximately 110 g CO_2/kWh and the number can be higher in the wider European region when fossil fuels are involved [53]. The second approach is to adopt the hourly energy demand multiplied with the hourly CO_2 factor and price in the spot market. The real-time data of the CO_2 factor of electricity and spot price can be retrieved from electricityMap and Nord Pool separately [54,55]. The benefit of employing the hourly energy demand with spot market information is that it can better locate the critical impacts from the peak load and define the energy-saving potentials, for example how to perform load shedding and energy storage, and transition to higher DH coverage system and so on.

Due to the length of this paper, only the most representative load profiles for the average Nordic school were identified, without separating the buildings with typical high energy–density profiles and typical low energy–density profiles. In addition, the evaluation

Table 5
PAA coefficients results in Fig. 19 transferred into SAX symbols.

weekday		Winter	Spring& Autumn	Summer
Mon	predicted	a b d d d b a a	*b* c d d c *a* a a	b *d* d d *b* a a a
	monitored	a b d d d b a a	*a b* d d c *b* a a	b c d d c a a a
Tue	predicted	a b d d d b a a	*b* c d d *b a* a a	b *d* d c *b* a a a
	monitored	a b d d d b a a	*a b* d d c *b* a a	b c d *d* c a a a
Wed	predicted	a b d d d b a a	*b* c d d *c a* a a	b d d d *b a* a a
	monitored	a b d d d b a a	*a b* d d c *b* a a	b c d d c *b* a a
Thur	predicted	a b d d d b a a	*b* c d d c *a* a a	b d d d *b* a a a
	monitored	a b d d d b a a	*a b* d d c *b* a a	b c d d c a a a
Fri	predicted	a b d d d b a a	*b* c d d c *a* a *b*	b c d d *b* a a a
	monitored	a b d d d b a a	*a b* d d c *b* a *a*	b c d d c a a a

process was made by using the modified average data that were extracted for the regression analysis. It may also be necessary to use the current statistically typical profiles and individual buildings to compare and reflect the energy use trend and load scales. And the selection of the sample individual buildings shall be carefully considered with regards to sample sizes, representativeness, etc.

Although the proposed approach is robust and transferable, the model accuracies for annual heat load profiles shall be further improved. More advanced statistical based techniques and artificial intelligence based prediction techniques shall be selected. These AI methods are mostly developed for short term load prediction, it is still worthwhile to extrapolate the time-series prediction to yearly ones.

7. Conclusions

This study proposed a systematic approach to identify the typical energy load profiles of the average Nordic school connected to the DH system, which was made up of the modified average load of the 40 schools. The observed buildings involved different building ages, areas, energy labelling levels, and operation in the Nordic climate.

The work was done through five steps, including data collection and processing, detection of special energy use period, prediction of annual DH and electricity load profiles, and validation process. The main findings are the following:

- The selected modified Z-Score may point out the special energy use periods and show the energy demand trend. The electric appliances in the schools might be concluded with reasonably fast responses by following the attendance. While the DH demand mainly followed the outdoor temperature and the daily work schedule, with a slow control response to the short holidays, which caused part of heating energy being wasted.
- The ES curve models combining temperature moving average and segmented piece-wise linear regression gave satisfying descriptive results.
- The identified specific load profiles may present the current energy use of schools in the Nordic climate. The predicted peak load of DH was 48 W/m^2 and the annual demand was 72 kWh/m^2. The predicted peak load of electricity was 18 W/m^2 and the annual demand was 57 kWh/m^2. Accordingly, the buildings with DH may largely reduce the power grid strains.
- The symbolization cluster methods of PAA and SAX were efficient and robust for validating building energy prediction.
- The suggested approach does not require much time-consumed computation and can be efficiently applied to other public buildings under the similar climate. This benefits public administrations to have a better understanding of energy needs for different building functions and project future demand changes by varying penetration of various building types.

In the future work, the authors are going to synthesize an artificial urban area by aggregating representative load profiles for different building types, which is to be used as input for modelling and optimization of the energy system. Advanced statistical and

Table A1
Results of modified Z-Score of unusual weekly DH use.

Year	No. in figure	Week no.	Remarks	|Modified Z-Score| ≥3.5 (Y/N)
2016	1	1	Coldest temp was −17.3 °C	Y
	2	2	Coldest temp was −15 °C	Y
	3	45	Coldest temp was −10.9 °C	Y
2018	4	9	Coldest temp was −17.5 °C	Y
	5	12		Y
	6	13	Easter; Coldest temp was −11.1 °C	Y
	7	39		Y

Table A2
Results of modified Z-Score of DH use regarding short holidays 2015–2018.

Year	No. in figure	Date	Week no.	Holiday (Y/N)	|Modified Z-Score|≥3.5 (Y/N)	Remarks
2015	1	4.6	W15	Y	Y	Easter
	2	4.10	W15	N	Y	
	3	5.1	W18	Y	N (but local minimum)	Labor Day
	4	5.14	W20	Y	N	Ascension Day
	5	5.25	W22	Y	N	Whit Monday
	6	5.29	W22	N	Y	
2016	1	3.28	W13	Y	N (but local minimum)	Easter
	2	5.5	W18	Y	N	Ascension Day
	3	5.16	W20	Y	N	Whit Monday
	4	5.17	W20	Y	N	Constitution Day
2017	1	4.17	W16	Y	N (but local minimum)	Easter
	2	5.1	W18	Y	N	Labor Day
	3	5.5	W18	N	N (but local minimum)	
	4	5.17	W20	Y	N (but local minimum)	Constitution Day
	5	5.25	W21	Y	N (but local minimum)	Ascension Day
	6	6.5	W23	Y	Y	Whit Monday, abnormally high demand
2018	1	4.2	W14	Y	N (but local minimum)	Easter
	2	5.1	W18	Y	N	Labor Day
	3	5.10	W19	Y	N	Ascension Day
	4	5.14	W20	N	Y	Abnormally low demand
	5	5.17	W20	Y	N	Constitution Day
	6	5.21	W21	Y	N (but local minimum)	Whit Monday

Table A3
Results of modified Z-Score of unusual weekly electricity use.

Year	No. in figure	Week no.	Remarks	|Modified Z-Score|≥3.5 (Y/N)
2018	1	1	Involve holiday	Y
2015–2018	2	8	School week	N in 2015 (local minimum); Y
2016	3	12	Easter	Y
2018	4	13	Easter	Y
2015, 2016	5	14 in 2015; 13 in 2016	Easter	Y
2017	6	15	Easter	N (local minimum)
2015–2018	7	41	School week	N in 2015 (local minimum); Y
2015–2018	8	53 in 2015; 52	Christmas	N in 2015 (local minimum); Y

Table A4
Results of modified Z-Score of electricity use regarding short holidays during 2015–2018.

Year	No. in figure	Date	Week no.	Holiday (Y/N)	|Modified Z-Score|≥3.5 (Y/N)	Remarks
2015	1	2.16–2.20	W8	Y	Y	school week
	2	3.30–4.3				
6.29–7.3	W14; W27	Y	Y	Easter; summer vacation		
	3	4.6	W15	Y	Y	Easter
	4	4.13, 4.14	W16	N	Y	Abnormally higher use
	5	5.1	W18	Y	N (but local minimum)	Labor Day
	6	5.14	W20	Y	Y	Ascension Day
	7	5.15	W20	N	N (but local minimum)	After holiday 5.14, some schools might have reduced school hour
	8	5.25	W22	Y	Y	Whit Monday
2016	1	2.22–2.26	W8	Y	2.22–2.25 Y; 2.26 N (but local minimum)	school week
	2	3.21–3.25	W12	Y	Y	Easter
	3	3.28	W13	Y	Y	Easter
	4	5.5	W18	Y	Y	Ascension Day
	5	5.6	W18	N	N (but local minimum)	After holiday 5.5, some schools might have reduced school hour
	6	5.16	W20	Y	Y	Whit Monday
	7	5.17	W20	Y	N (but local minimum)	Constitution Day
2017	1	4.11–4.14	W15	Y	N (but local minimum)	Easter
	2	4.17	W16	Y	N (but local minimum)	Easter
	3	5.1	W18	Y	N (but local minimum)	Labor Day
	4	5.17	W20	Y	N (but local minimum)	Constitution Day
	5	5.25	W21	Y	N (but local minimum)	Ascension Day
	6	5.26	W21	N	N (but local minimum)	After holiday 5.25, some schools might have reduced school hour
	7	6.5	W23	Y	N (but local minimum)	Whit Monday
2018	1	1.1	W1	Y	Y	
	2	2.19, 2.20	W8	Y	Y	school week
	3	3.26, 3.27	W13	Y	Y	Easter
	4	4.2	W14	Y	Y	Easter
	5	5.1	W18	Y	Y	Labor Day
	6	5.10	W19	Y	N (but local minimum)	Ascension Day
	7	5.17	W20	Y	N (but local minimum)	Constitution Day
	8	5.21	W21	Y	Y	Whit Monday
	9	6.26	W26	Y	Y	summer vacation

artificial intelligence based techniques shall be selected and performed to enhance the accuracies of the load profiles. Moreover, in order to make more precise energy planning, on the trend of more efficient buildings into the market, which is dominated by the medium-aged buildings, it is needed to make dynamic building stock forecast and separate the representative load profiles based on building energy density.

CRediT authorship contribution statement

Yiyu Ding: Conceptualization, Methodology, Formal analysis, Software, Investigation, Validation, Visualization, Writing - original draft. **Helge Brattebø:** Formal analysis, Supervision, Writing - review & editing. **Natasa Nord:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing - review & editing.

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 article has been written within the research project "Methods for Transparent Energy Planning of Urban Building Stocks–ExPOSe". The authors gratefully acknowledge the support from the Research Council of Norway, Department of Energy and Process Engineering of NTNU, and Trondheim municipality.

Appendix

Table A1–A4 are listed in Appendix.

References

[1] T. Abergel, B. Dean, J. Dulac, "Towards a zero-emission, efficient, and resilient buildings and construction sector, Global status report (2017",) 48.

[2] J.Z. Thellufsen et al., Smart energy cities in a 100% renewable energy context, Renewable and Sustainable Energy Reviews 129 (Sep. 2020), https://doi.org/10.1016/j.rser.2020.109922 109922.

[3] H. Averfalk, S. Werner, Economic benefits of fourth generation district heating, Energy 193 (Feb. 2020), https://doi.org/10.1016/j.energy.2019.116727 116727.

[4] H. Lund et al., Perspectives on fourth and fifth generation district heating, Energy 227 (Jul. 2021), https://doi.org/10.1016/j.energy.2021.120520 120520.

[5] R. Moschetti, H. Brattebø, M. Sparrevik, Exploring the pathway from zero-energy to zero-emission building solutions: A case study of a Norwegian office building, Energy and Buildings 188–189 (Apr. 2019) 84–97, https://doi.org/10.1016/j.enbuild.2019.01.047.

[6] F. Ascione, N. Bianco, R.F. De Masi, M. Mastellone, G.M. Mauro, G.P. Vanoli, The role of the occupant behavior in affecting the feasibility of energy refurbishment of residential buildings: Typical effective retrofits compromised by typical wrong habits, Energy and Buildings 223 (Sep. 2020), https://doi.org/10.1016/j.enbuild.2020.110217 110217.

[7] H. Yoshino, T. Hong, N. Nord, IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods, Energy and Buildings 152 (Oct. 2017) 124–136, https://doi.org/10.1016/j.enbuild.2017.07.038.

[8] G. Reynders, R. Amaral Lopes, A. Marszal-Pomianowska, D. Aelenei, J. Martins, D. Saelens, Energy flexible buildings: An evaluation of definitions and quantification methodologies applied to thermal storage, Energy and Buildings 166 (May 2018) 372–390, https://doi.org/10.1016/j.enbuild.2018.02.040.

[9] L. Dias Pereira, D. Raimondo, S. P. Corgnati, and M. Gameiro da Silva, "Energy consumption in schools – A review paper," Renewable and Sustainable Energy Reviews, vol. 40, pp. 911–922, Dec. 2014, 10.1016/j.rser.2014.08.010.

[10] "EE Noon: Back to School with Energy Efficiency," Alliance to Save Energy, Aug. 14, 2013. https://www.ase.org/events/ee-noon-back-school-energy-efficiency (accessed Feb. 24, 2021).

[11] L. Pistore, G. Pernigotto, F. Cappelletti, A. Gasparella, P. Romagnoni, A stepwise approach integrating feature selection, regression techniques and cluster analysis to identify primary retrofit interventions on large stocks of buildings, Sustainable Cities and Society 47 (May 2019), https://doi.org/10.1016/j.scs.2019.101438 101438.

[12] A.O. Hopland, S. Kvamsdal, Building conditions in Norwegian local governments: trends and determinants, Facilities 37 (3/4) (Jan. 2018) 141–156, https://doi.org/10.1108/F-10-2017-0101.

[13] Norges vassdrags- og energidirektorat, "Analyse av energibruk i undervisningsbygg," p. 120.

[14] Y. Sun, S. Wang, F. Xiao, Development and validation of a simplified online cooling load prediction strategy for a super high-rise building in Hong Kong, Energy Conversion and Management 68 (Apr. 2013) 20–27, https://doi.org/10.1016/j.enconman.2013.01.002.

[15] C. Fan et al., Statistical investigations of transfer learning-based methodology for short-term building energy predictions, Applied Energy 262 (Mar. 2020), https://doi.org/10.1016/j.apenergy.2020.114499 114499.

[16] X. Liu, Y. Ding, H. Tang, F. Xiao, A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data, Energy and Buildings 231 (Jan. 2021), https://doi.org/10.1016/j.enbuild.2020.110601 110601.

[17] P. Gianniou, X. Liu, A. Heller, P.S. Nielsen, C. Rode, Clustering-based analysis for residential district heating data, Energy Conversion and Management 165 (Jun. 2018) 840–850, https://doi.org/10.1016/j.enconman.2018.03.015.

[18] M.Q. Raza, A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, Renewable and Sustainable Energy Reviews 50 (Oct. 2015) 1352–1372, https://doi.org/10.1016/j.rser.2015.04.065.

[19] A. Melillo, R. Durrer, J. Worlitschek, P. Schütz, First results of remote building characterisation based on smart meter measurement data, Energy 200 (Jun. 2020), https://doi.org/10.1016/j.energy.2020.117525 117525.

[20] Y. Bao, W.L. Lee, J. Jia, Probabilistic assessment of overcooling risk for a novel extra-low temperature dedicated outdoor air system for Hong Kong office buildings, Build. Simul. 14 (3) (Jun. 2021) 633–648, https://doi.org/10.1007/s12273-020-0684-4.

[21] L. Lundström, F. Wallin, Heat demand profiles of energy conservation measures in buildings and their impact on a district heating system, Applied Energy 161 (Jan. 2016) 290–299, https://doi.org/10.1016/j.apenergy.2015.10.024.

[22] "iEOS - Planning." https://www2.esave.no/Esave.nsf/iEOS_Hovedbilde.xsp (accessed Feb. 24, 2021).

[23] "Enova Offentlig søk etter energiattester." https://attest.energimerking.no/ (accessed May 10, 2021).

[24] C. Hjortling, F. Björk, M. Berg, and T. af Klintberg, "Energy mapping of existing building stock in Sweden – Analysis of data from Energy Performance Certificates," Energy and Buildings, vol. 153, pp. 341–355, Oct. 2017, 10.1016/j.enbuild.2017.06.073.

[25] BRE Global, "BRE Environmental & Sustainability Standard, BES 5051: ISSUE 1.0, BREEAM Education 2008 Assessor Manual." [Online]. Available: https://tools.breeam.com/filelibrary/Non%20Domestic%20Manuals/BREEAM_Education_2008.pdf

[26] M. Davis, "Implementing the Energy Performance of Buildings Directive (EPBD) - Featuring Country Reports 2012," Build Up, Jul. 02, 2013. https://www.buildup.eu/en/practices/publications/implementing-energy-performance-buildings-directive-epbd-featuring-country-0 (accessed Mar. 12, 2021).

[27] "Årsberetning 2017," Oslo kommune. https://www.oslo.kommune.no/politikk/byradet/arsberetning-2017/ (accessed May 10, 2021).

[28] J.E. Seem, Using intelligent data analysis to detect abnormal energy consumption in buildings, Energy and Buildings 39 (1) (Jan. 2007) 52–58, https://doi.org/10.1016/j.enbuild.2006.03.033.

[29] S. Lydersen, Mean and standard deviation or median and quartiles?, Tidsskrift for Den norske legeforening (Jun. 2020), https://doi.org/10.4045/tidsskr.20.0032.

[30] M. Frigge, D.C. Hoaglin, B. Iglewicz, Some Implementations of the Boxplot, The American Statistician 43 (1) (Feb. 1989) 50–54, https://doi.org/10.1080/00031305.1989.10475612.

[31] B. Iglewicz, D.C. Hoaglin, How to detect and handle outliers, ASQC Quality Press, Milwaukee, Wis, 1993.

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

[33] T. Tereshchenko, D. Ivanko, N. Nord, and I. Sartori, "Analysis of energy signatures and planning of heating and domestic hot water energy use in buildings in Norway," in E3S Web of Conferences, Les Ulis, France, 2019, vol. 111. http://dx.doi.org/10.1051/e3sconf/201911106009.

[34] K. B. Lindberg and G. Doorman, "Hourly load modelling of non-residential building stock," in 2013 IEEE Grenoble Conference, Jun. 2013, pp. 1–6. 10.1109/PTC.2013.6652495.

[35] R. Hitchin, I. Knight, Daily energy consumption signatures and control charts for air-conditioned buildings, Energy and Buildings 112 (Jan. 2016) 101–109, https://doi.org/10.1016/j.enbuild.2015.11.059.

[36] S. Frederiksen, S. Werner, *District Heating and Cooling*. Studentlitteratur AB (2013).

[37] C. Ghiaus, Experimental estimation of building energy performance by robust regression, Energy and Buildings 38 (6) (Jun. 2006) 582–587, https://doi.org/10.1016/j.enbuild.2005.08.014.

[38] American Society of Heating Refrigerating and Air Conditioning Engineers, ASHRAE handbook: fundamentals Accessed: Feb. 27, 2021. [Online]. Available: http://app.knovel.com/hotlink/toc/id:kpASHRAEC1/2013-ashrae-handbook 2013, 2013.

[39] S. Menard, Coefficients of Determination for Multiple Logistic Regression Analysis, The American Statistician 54 (1) (Feb. 2000) 17–24, https://doi.org/10.1080/00031305.2000.10474502.

[40] Z. Ma, R. Yan, N. Nord, A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings, Energy 134 (2017) 90–102, https://doi.org/10.1016/j.energy.2017.05.191.

[41] J. L. M. Hansen and R. Lamberts, "Building Performance Simulation for Design and Operation," Routledge & CRC Press. https://www.routledge.com/Building-Performance-Simulation-for-Design-and-Operation/Hensen-Lamberts/p/book/9781138392199 (accessed Feb. 27, 2021).

[42] "JRC Photovoltaic Geographical Information System (PVGIS) - European Commission." https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html#TMY (accessed Feb. 24, 2021).

[43] D. Rutz et al., "UpgradeDH: Upgrading the performance of district heating networks in Europe," Euroheat & Power, Jun. 14, 2018. https://www.euroheat.org/our-projects/upgradedh/ (accessed Mar. 12, 2021).

[44] "Rapport: Enovas byggstatistikk,, Mynewsdesk, (accessed Mar. 12 (2016",) 2021). https://presse.enova.no/documents/rapport-enovas-byggstatistikk-2016-73172.

[45] T. Niemelä, R. Kosonen, J. Jokisalo, Cost-optimal energy performance renovation measures of educational buildings in cold climate, Applied Energy 183 (Dec. 2016) 1005–1020, https://doi.org/10.1016/j.apenergy.2016.09.044.

[46] J. Granderson, S. Touzani, C. Custodio, M. Sohn, S. Fernandes, D. Jump, "Assessment of Automated Measurement and Verification (M&V), Methods" (2015).

[47] Nigel Meade, Industrial and business forecasting methods, Lewis, C.D., Borough Green, Sevenoaks, Kent: Butterworth, 1982. Price: £9.25. Pages: 144, Journal of Forecasting 2 (2) (1983) 194–196, https://doi.org/10.1002/(ISSN)1099-131X10.1002/for.v2:210.1002/for.3980020210.

[48] G. R. Ruiz and C. F. Bandera, "Validation of Calibrated Energy Models: Common Errors," Energies, vol. 10, no. 10, Art. no. 10, Oct. 2017, 10.3390/en10101587.

[49] H. Gadd and S. Werner, "Heat load patterns in district heating substations," Applied Energy, vol. 108, pp. 176–183, Aug. 2013, 10.1016/j.apenergy.2013.02.062.

[50] E. Keogh, J. Lin, and A. Fu, "HOT SAX: efficiently finding the most unusual time series subsequence," in Fifth IEEE International Conference on Data Mining (ICDM'05), Nov. 2005, pp. 8 pp.-. 10.1109/ICDM.2005.79.

[51] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing SAX: a novel symbolic representation of time series, Data Min Knowl Disc 15 (2) (Oct. 2007) 107–144, https://doi.org/10.1007/s10618-007-0064-z.

[52] C. Miller, Z. Nagy, A. Schlueter, Automated daily pattern filtering of measured building performance data, Automation in Construction 49 (Jan. 2015) 1–17, https://doi.org/10.1016/j.autcon.2014.09.004.
[53] Norsk Eenergi, "Klimaregnskap for fjernvarme." http://www.fjernvarme.no/uploads/Rapport_Klimaregnskap%20for%20fjernvarme_2.pdf (accessed Jan. 28, 2021).
[54] "Live CO_2 emissions of electricity consumption." http://electricitymap.tmrow.co (accessed Mar. 12, 2021).
[55] "See market data for all areas." https://www.nordpoolgroup.com/Market-data1/ (accessed Feb. 24, 2021).