Selecting the model and influencing variables for DHW heat use prediction in hotels in Norway

Dmytro Ivanko [a,*], Åse Lekang Sørensen [b], Natasa Nord [a]

[a] *Department of Energy and Process Technology, Norwegian University of Science and Technology (NTNU), Kolbjørn Hejes vei 1 B, Trondheim 7491, Norway*
[b] *SINTEF, SINTEF Community, P.O. Box 124 Blindern, 0314 Oslo, Norway*

ARTICLE INFO

ABSTRACT

Domestic hot water heat use prediction modelling is an important instrument for increasing energy efficiency in many buildings. This article addressed hourly domestic hot water heat use prediction, using a Norwegian hotel as a case study. Since the information available for buildings may vary, two widespread situations with different input variables were studied. For the first situation, the prediction is based only on data obtained from historical measured domestic hot water heat use. For the second situation, additional variables that affect domestic hot water heat use were applied. These variables were determined using the Wrapper approach. The Wrapper approach showed that factors related to the guests presence have the most significant influence on the domestic hot water heat use in the hotel. Nevertheless, daily data about the number of guests booked at the hotel did not appear to be informative enough for precise hourly modelling. Therefore, to improve the accuracy of the prediction, it was proposed to use an artificial variable. This artificial variable explained the hourly intensity of the guests domestic hot water use. In order to select the best model for the domestic hot water heat use prediction, ten advanced time series and machine learning techniques were tested based on the criteria of models adequacy. For both considered situations, the Prophet model showed the best results with R2 equal to 0.76 for the first situation, and 0.83 the second situation.

1. Introduction

Buildings are one of the largest categories of energy consumers in the European Union (EU) [1]. Buildings are currently responsible for approximately 36% of global energy use [2]. Therefore, increasing energy efficiency in buildings is an essential step for reducing fossil fuel use and improving the environmental situation.

Nowadays, most building constructions have complex technical systems, to realize a comfortable living condition for people. Among these systems, the domestic hot water (DHW) system is an integrated component of every building. DHW systems are significant consumers of energy. According to [3], 15% of the total heat demand in the EU is associated with DHW use. In regular buildings, DHW systems typically consume 25–35% of the total energy use [4]. However, in highly insulated constructions, the share of DHW heat use is increasing and may exceed the space heating [5]. Therefore, substantial opportunities for energy savings in buildings can be achieved by improving the performance of DHW systems [6]. The investigation [7] shows that DHW account for almost 26% of total energy use in the hotel, and therefore it should be prioritized in energy-saving measures.

Data-driven analysis and predictive modelling are powerful instruments for increasing the efficiency of heat use in DHW systems. Improving the design and operation of DHW systems requires both validated forecasting models, heat use profiles, effective utilization of monitoring and control systems. In order to solve all these issues, accurate predictive models of DHW heat use should be developed.

The introduction of modern technical energy solutions in DHW systems is essential for energy efficiency in buildings [8]. The proper implementation of these solutions requires the application of data analysis for DHW heat use. For example, the conceptual designs for DHW heating systems in a hotel with the application of wastewater technologies are considered in [9]. The research shows that the DHW system control is prioritized to operate with the wastewater technologies and heat pumps. This control can be performed based on DHW predictive models. Using a solar-assisted DHW water heating systems in hotels becomes popular all over the world [10]. The prediction of DHW heat use is necessary for the optimal operation of these systems [11]. Different types of DHW heating systems are investigated in [12]. This study

* Corresponding author.
E-mail address: dmytro.ivanko@ntnu.no (D. Ivanko).

summarises that DHW energy use can be reduced through using combined systems based on traditional and renewable energy solutions. However, due to unstable behaviour of renewable energy sources, development of accurate profile and prediction of DHW heat use becoming crucial for successful operation of combined DHW heating systems.

In recent years, increasing attention is paid to the investigation for the modelling of space heating heat use and the development of Energy Signature Diagrams [13]. On the contrary, the DHW heat use predictive modelling has not been studied sufficiently [6]. It is important to stress that the majority of existing publications are focused mainly on the modelling of DHW volumetric use rather than heat use. These two parameters have a strong positive correlation. Besides, the factors that affect the DHW volumetric use have a similar effect on the DHW heat use. Since not so many publications are dedicated to DHW heat use prediction, both previous experience of the predictive modelling for DHW volumetric and heat use are considered in this introduction.

Traditionally, predictive modelling includes the following main steps: identifying influencing variables, selecting the method for prediction, and determining the parameters of the model.

1.1. Identifying influencing variables

Identifying influencing variables with significant impact on the DHW heat use in the building is an initial step for prediction. There is a number of scientific papers analyzing the influence of different factors on DHW volumetric and heat use, as shown in Table 1.

Most of the articles represented in Table 1 assume that the number of occupants, seasons, day of the week and time of the day have a significant influence on the DHW heat use. The information about activities, such as occupant's presence, sleeping, hygiene and cooking, as well as a time when appliances are in use (sinks, showers, baths, clothes washer, and dishwasher) gives a better understanding of the DHW heat use [19]. It should be noticed that the factors influencing DHW heat use can vary from one building type to another, and also depending on the location of the building. For example, in the investigation [15], it is concluded that the influence of seasons, outdoor temperature, and rainy days on DHW in the dwellings is negligible. However, in the articles [23], the seasons and outdoor temperature are considered as essential variables and taken into account. Therefore, it is necessary to evaluate the influence of variables on the DHW heat use for each building type in Norway based on reliable statistical methods.

1.2. Selecting the method for prediction and determining parameters

In accordance with selected influencing factors, the model of DHW energy use should be built. Machine learning and deep learning techniques show high accuracy for solving prediction and data analysis problems in DHW systems [27]. The review of prediction techniques that different researchers use for solving this issue is represented below.

The application of artificial neural networks (ANNs) for DHW modelling in Canadian households is considered in [31]. The DHW heat use as ANNs model of draw-off temperatures is presented in [26]. The model is tested in three residential DHW systems. The archived ANNs model accuracy is more than 89% for the trained data. However, the use of the ANN model for new data obtained from other systems shows significant inaccuracy.

Creation of easy to use forecasting model of DHW use is considered in [32]. Autoregressive moving average (ARMA) model as a solution to this problem is proposed. The ARMA model takes into consideration the periodicity of the week, the water use of the days before and random fluctuations of DHW use. The model based on data from eight apartments in France is examined [32].

Table 1
Investigations of variables that have a significant impact on DHW volumetric use and heat use.

Influencing variables	Authors
Number of occupants, day of the week	Ferrantelli, Ahmed, Pylsy and Kurnitski [14]
Day of the week	de Santiago, Rodriguez-Villalón and Sicre [15]
Number of rooms, area	Chmielewska, Szulgowska-Zgrzywa and Danielewicz [16]
The magnitude of the drains, the start times of DHW use, the time between two successive drains	Beeker, Malisani and Petit [17]]
Occupancy in the hotel and regulation of the system	Todorovic, Tomic, Bojanic, Bajatovic and Andelkovic [9]
Hotel star rating, DHW system type, occupancy	Priyadarsini, Xuchao and Eang [18]
Activities, number of DHW tap starts, time of tapping, the duration of tapping	Fischer, Wolf, Scherer and Wille-Haussmann [19]
Flow rates, cold and supply temperatures	Verhaert, Bleys, Binnemans and Janssen [20]
Type of the tap (conventional mixer tap or low flow electronic tap)	Fidar, Memon and Butler [21]
Activities, appliances	Good, Zhang, Navarro-Espinosa and Mancarella [22]
Outdoor temperature, season, number of tenants, type of building (apartment or detached), the location, the household area, month, density of water, specific heat of water, reference temperatures, cold inlet temperature	Gutierrez-Escolar, Castillo-Martinez, Gomez-Pulido, Gutierrez-Martinez and Stapic [23]
Socioeconomic characteristics, activities, appliances, and type of apparatuses that use water	Fan, Liu, Wang, Geissen, Ritsema and Tong [24]
Occupant behaviour, appliances, demographic conditions, and occupancy rate	Swan, Ugursal and Beausoleil-Morrison [25]
Draw-off temperatures	Barteczko-Hibbert, Gillott and Kendall [26]
Activities	Widen, Lundh, Vassileva, Dahlquist, Ellegard and Wackelgard [27]
Appliances, flow rates and times of DHW use	Hendron and Burch [28]
The day of the week, time of the day, season, appliances, age of occupants (seniors or not), pay or does not pay for hot water	Lutz, Liu, McMahon, Dunham, Shown and McCure [29]
Family size, season, day of the week, time of the day	Papakostas, Papageorgiou and Sotiropoulos [30]

The linear regression models were used for DHW energy use identification in apartment blocks in Norway [33].

A bottom-up model that estimates the day ahead DHW use for end-users is investigated in [34]. The type of facilities and timing of DHW use is applied as an input in the model. The prediction for the next day of the total DHW use in the system is calculated as a sum of end-users DHW use.

The survey of DHW use in 626 apartments in Poland is carried out in [16]. The authors create a database of DHW use for residential buildings with different parameters. The configuration of apartments in these buildings is randomly selected by using the bootstrap method. Based on the database, the regression model is constructed. This model considers DHW use as a function of the number of rooms and the floor area.

The stochastic analysis of DHW use for 65 apartments is performed in Hungary [35]. As an input for the stochastic model, the authors use the number of apartments in the building, the duration curve, daily average, minimum and maximum values of DHW use.

The issue of DHW use forecasting for demand-side management in residential buildings in the UK is reviewed in [36]. Various time

series forecasting techniques, such as exponential smoothing, seasonal autoregressive integrated moving average, seasonal decomposition by Loess model and a combination of them, were tested on data from 120 dwellings.

A model for DHW use prediction that consists of 16 equations is proposed in [29]. These equations take in account season, day of the week, and hours with similar DHW use. To improve the model, the authors propose to consider additional factors to adjust the predicted hot water use. These factors include the availability of dishwashers, cloth washers, age of occupants, and if the residents should pay for hot water or not.

The Long-Short Term Memory (LSTM) neural networks were used for DHW heat use prediction in [11]. The performance of simple LSTM neural network, Attention-based LSTM neural network (ALSTM) and Attention-based LSTM using decomposed data (ALSTM-D) are compared. The authors claims that the Long-Short Term Memory (LSTM) neural network shows the best results for DHW heat use prediction in the case of solar-assisted DHW systems.

As we can see, the largest part of the above-mentioned studies performed investigations for residential buildings. Practice shows that for residential buildings, information about the DHW heat use is more opened and accessible [6]. Despite this fact, the share of DHW heat use in non-residential buildings is also significant and cannot be neglected [37]. Among non-residential buildings, hotels [38] are those with the most energy-consuming categories [39]. In hotels, the specific DHW heat use, the regimes of work, and available information about factors affecting DHW heat use are substantially different from the residential buildings [6]. Accordingly, the approaches proposed in the above-mentioned studies cannot be directly applied for the DHW heat use modelling in hotels. Therefore, more reliable prediction models of DHW heat use for non-residential buildings, including hotels, should be created.

1.3. Contribution and organization of the paper

The purpose of this article is to develop an accurate and reliable hourly DHW heat use prediction model for hotels, using a hotel in Norway as a case. In order to make the results of the investigations applicable to other buildings, two alternative situations with available inputs for prediction were considered.

Situation 1 assumed that information about influencing variables for the DHW heat use was not available. Only historical data about DHW heat use weres known. For these conditions, the article investigated the various methods to handle the prediction based on the time series of the DHW heat use only. In general, Situation 1 is less common for hotels. Usually, measurement systems in hotels collect data about building energy performance. In addition, useful information about guest presence can be obtained from the hotel booking system. However, for certain non-residential buildings, these variables are unknown. The results of the investigation and developed models for Situation 1 may be useful and applicable to such buildings.

In Situation 2, the research focused on identifying factors affecting DHW heat use and developing a prediction model based on these variables. The influencing variables on DHW heat use is identified based on the wrapped approach. In order to improve the accuracy of the prediction, the article proposes procedure for preprocessing data of daily guests presence and extracting information of their influence on DHW heat use on an hourly basis. Finally, advanced time series and machine learning techniques were tested, to find the best prediction model among them.

The paper is organized as the following. Section 2 describes the main characteristics of the hotel for which the prediction of DHW heat use was made. Section 3 introduces the methodology for DHW heat use prediction in the following situations: for Situation 1, only retrospective Time Series of DHW heat use is known. For Situation 2, also other parameters that could influence DHW heat use were available. In Section 3, the methodology was applied for the DHW heat use prediction in a hotel located in Oslo, Norway. Among considered modelling techniques, the model that gives the most accurate and robust prediction for Situation 1 and Situation 2 was identified.

2. Description of the hotel

The investigations in this article were performed based on data obtained from an urban hotel, located on the west side of Oslo, Norway. The characteristics of the hotel are typical for Scandinavian conditions. The building was built in 1938. There has been several renovation projects, where the most recent was in 2007. The total area of the building is 4 939 m². The building has eight floors with 164 guest rooms. All the guest rooms are equipped with bathrooms that have toilet facilities, washbasin, and a shower. The check-in time for the guests is between 15:00o'clock and midnight, and check out before 12:00o'clock.

The considered hotel well represents the general tendency of the DHW heat use in similar building types. According to hotel management, employees use hot water for cleaning and guests use hot water for personal hygiene. In the DHW system, the hot water is circulated to ensure fast delivery at each tap. The hotel uses electric water heaters for DHW production. Data on heat use for DHW production was collected within several years from a stationary energy meter in the hotel. The meter measures electricity delivered to the DHW tanks, which means that both DHW needs and heat losses in the DHW system are included in the presented DHW heat use. The data about electrical use for other needs in the hotel are also measured. The hotel booking system allows us to obtain daily information about the number of arriving guests and booked rooms in the hotel. The influence of weather conditions on hourly DHW heat use was investigated, too. For this purpose, data obtained from the nearest meteorological station located in Oslo were used [40].

3. Methods

This chapter consists of two subsections that are dedicated to modelling in Situation 1 and Situation 2. Subsection 3.1 investigates the hourly prediction based on the historical time series of DHW heat use. Subsection 3.2 considers the issue of identifying variables that affect DHW heat use, followed by making prediction when using these variables. For this purpose, time series and machine learning techniques were used. In addition, in Subsection 3.2, a method which introduced the artificial variable reflecting the hourly intensity of the guests DHW use and improved the accuracy of the hourly DHW models was proposed.

3.1. Prediction based on the historical time series of DHW heat use

For certain types of buildings, information about users presence and other explanatory variables are unknown. In these conditions, only DHW heat use data from previous periods of time can be used for prediction. Practice shows that the DHW heat use may vary at different hours of the day, day of the week, and months. For this reason, the preference was given to methods that allowed us to make a prediction based on the historical time series of DHW heat use and additionally take in account the day, week, and month when the DHW heat use occurred. Among different methods such as classical methods for time series analyses, Exponential Smoothing (ES) and Autoregressive Integrated Moving Average (ARIMA),

and modern methods of machine learning, Neural Network (ANN), Prophet and XGBoost, were considered.

The ES method uses recurrence relations between the current and the previous values of the parameter. According to ES, predictions are calculated by applying weighted averages where the weights are exponentially decreasing as observations come further from the past [41]. In detail, the ES method is presented in [41]. According to [41], exponential smoothing uses the following equation for prediction:

$$\hat{E}_{T+1|t} = \alpha \cdot E_T + (1 - \alpha) \cdot \hat{E}_{T|t-1} \tag{1}$$

where $\hat{E}_{T+1|t}$ is the predicted value and $\hat{E}_{T|t-1}$ is the prediction for the previous moment of the time. E_T is the most recent observation. α is the smoothing parameter, accepted from 0 to 1.

The ARIMA method predicts the next step in the sequence as a linear function of the differenced observations and residual errors at previous time steps [42]. This method combines autoregressive (AR), Moving Average (MA) and the integrated (I) parts in one model. An integrated part of the model performs a differentiation pre-processing step of modelling that removes the non-stationarity of the time series. AR and MA are the core of prediction. The algorithm and theoretical bases of ARIMA modelling technique are well explained in [42].

The Prophet is a package for time series prediction developed by Facebook [43]. Prophet uses additive regression model $E(t)$ that includes the following components:

$$E(t) = g(t) + s(t) + h(t) \tag{2}$$

where $g(t)$ is a trend for non-periodic changes that may be obtained by a simple Piecewise Linear Model. $s(t)$ is a seasonal (periodical) component of the model obtained based on Fourier series. $h(t)$ is a component of the model that takes into account the effects of holidays and other untypical days with irregular schedules of DHW heat use.

XGBoost is a machine learning prediction technique based on gradient boosting decision tree method [44]. XGBoost sequentially sums the prediction of multiple weak learners, such as regression trees models, in order to ensemble a robust prediction model [45]. By adding additional regression trees models in such a way, the errors made by the initial model are reduced. The regression trees models are added until further improvements of the initial model can no longer be obtained. The gradient boosting is related to a gradient descent algorithm that is used in XGBoost to minimize the loss when adding new models [46]. Mathematically, gradient boosting can be represented by the following equation [46]:

$$\hat{E}_i = \sum_{k=1}^{K} f_k(X_i), f_k \in F \tag{3}$$

where $\hat{E}_i$ is predicted DHW heat use. X_i are influencing variables. K is the number of functions (regression trees) in the function space F.

In XGBoost the parameters of the functions can be found automatically by solving the following optimization function [46]:

$$obj(\theta) = \sum_{i}^{n} l(\hat{E}_i, E_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4}$$

where l is a differentiable loss function. Ω is the regularizing function that introduces penalties for the complexity of the model. A more extensive introduction to XGBoost modelling technique and its mathematical apparatus are given in [47].

Artificial Neural Network (ANN) is a powerful modelling technique that mimics the behaviour of the brain with its homogeneous elements - neurons. For prediction, classification and solving of other tasks, ANN uses the number of simple nonlinear functional blocks that are called neurons. Multiple neurons are organized into layers [48], where the actual processing of data is performed via a system of weighted connections [47]. The ANN represents the group of mathematical models of high complexity. This method demonstrates good results for nonlinear relationships among between variables. In this article, the ANN model with the two-layer feed-forward network [49] was used for DHW heat use prediction.

In order to estimate the accuracy of DHW heat use models, cross-validation was used. Hourly data of DHW heat use in 2015 were used in a training set, and data from 2016 were applied to test the models. The prediction for all the above-mentioned methods, except ANN, was performed in Python, using Statsmodels, XGBoost, and Prophet packages [50]. For Neural Networks modelling, the Neural Network Toolbox in Matlab software was utilized [49]. The comparison of the models was performed based on the Coefficient of Determination (R2), Mean Absolute Error (MAE), and Mean Squared Error (MSE) criteria of the model adequacy [50].

3.2. Prediction based on the variables that have a significant influence on the DHW heat use

Compared to Subsection 3.1, Subsection 3.2 considers more favourable conditions for DHW heat use prediction. In these conditions, in addition to DHW heat use data from previous periods of time, information about the guest's presence and other explanatory variables are known. The procedure for DHW heat use prediction in this subsection includes three main steps: data preprocessing, identifying variables that affect DHW heat use, and selection of the best model for hourly prediction of DHW use. The preprocessing step included removing outliers and unrealistic data. Finally, as a part of preprocessing, a method for introducing an artificial variable, which reflects the influence of hourly guest presence on DHW heat use, was proposed. This method, in detail, is explained in Section 3.2.1. The set of variables that affect the DHW heat use was selected according to the Wrapper approach. This approach is explained in Section 3.2.2. After, the selected set of influencing variables was used as an input for modelling. The accuracy of various machine learning methods for the DHW heat use prediction was carried out. The general information about the considered methods is presented in Section 3.2.3.

3.2.1. Preprocessing the daily data of the guest presence

It is well known that occupancy has a significant effect on the DHW heat use in buildings [6]. Among all influencing factors, the number of guests being present in a hotel is typically the key factor that affects DHW heat use the most.

Traditionally, a hotel booking system stores information about the number of guests who were booked into the hotel for each particular day. For a given date, both the number of guests booked in one day before (Gst_{Lag1}) as well as on the date itself (Gst) are influencing the DHW heat use. In general, Gst shows the number of guests who are staying in the hotel after 15.00o'clock, and Gst_{Lag1} reflects information about people who are leaving before 12:00o'clock. Nevertheless, despite the official check-in/out time, in practice, the actual time when guests are arriving and leaving can vary. Sometimes guests arrive before the set time of check-in, and it happens that some guests can stay longer in the building after the check-out time.

The daily profiles in the hotel showed that the highest DHW heat use occurs before 12:00o'clock. Consequently, the influence of Gst_{Lag1} on daily DHW heat use can be more significant than Gst. For this reason, it is crucial to take both factors Gst and Gst_{Lag1} into account in the model.

The investigation showed that using Gst and Gst_{Lag1} allows us to perform a quite accurate daily prediction of DHW heat use. How-

ever, if we consider hourly analysis of the DHW heat use, Gst and Gst_{Lag1} do not give sufficient information about hourly occupancy in the hotel. These parameters do not show whether the guests are present in the hotel at certain hours or not. For this reason, the considered factors cannot substantially enhance the accuracy of the hourly model of the DHW heat use. To increase the accuracy of the hourly model, we propose to introduce an additional artificial variable (Gst_{art}) that reflects the hourly influence of the guests presence on DHW heat use. The following equation proposed to use to determine the numerical value of the Gst_{art} for each separate hour:

$$Gst_{art} = Gst \cdot Cgp_i + Gst_{Lag1} \cdot Cgp_{Lag1.i} \tag{5}$$

where Cgp_i and $Cgp_{Lag1.i}$ are the coefficients for the guest DHW use intensity for ith-hour, which were identified based on the number of people booked into the hotel on the given day Gst and one day before Gst_{Lag1}.

In order to identify the coefficients of the guest DHW use intensity for ith-hour the following optimization problem was solved:

$$\max(corr(\left\{ \begin{array}{c} Cgp_{i=1} \cdot \left(\overrightarrow{Gst}\right) + Cgp_{Lag1.i=1} \cdot \left(\overrightarrow{Gst_{Lag1}}\right), \cdots, Cgp_{i=24} \cdot \left(\overrightarrow{Gst}\right) \\ + Cgp_{Lag1.i=24} \cdot \left(\overrightarrow{Gst_{Lag1}}\right) \end{array} \right\},$$

$$\left\{ \overrightarrow{E}_{i=1}, \cdots, \overrightarrow{E}_{i=24} \right\})) \tag{6}$$

where Cgp_i and $Cgp_{Lag1.i}$ are the target variables. $\overrightarrow{E}_i$ is the vector of the DHW energy use data in the hotel in ith-hour, $\overrightarrow{Gst}$, $\overrightarrow{Gst_{Lag1}}$ are vectors of the daily number of guests booked into the hotel on the given day and one day before.

By solving the optimization problem in Equation (6), the values of the coefficients of the guest DHW use intensity for each hour of the day can be obtained. These Cgp_i and $Cgp_{Lag1.i}$ coefficients are maximizing the correlation between Gst_{art} and the DHW heat use. Application of the coefficients makes Gst_{art} based predictions more accurate. For the considered hotel, the values of Cgp_i and $Cgp_{Lag1.i}$ were calculated for several years. The obtained coefficients for 2015 and 2016 years are shown in Fig. 6. and Fig. 7. in Section 4.2. The investigation indicated that the changes in the values of the Cgp_i and $Cgp_{Lag1.i}$, see Fig. 6. and Fig. 7., in different years, were not substantial. Thus, their values from previous years may be used for the identification of the variable Gst_{art} and prediction for the next year. In this article, the numerical values of the coefficients were calculated based on the year of 2015, and they were used for predicting the DHW heat use in 2016. Besides, to conduct a thorough investigation, both cases for modelling with application of the artificial variable Gst_{art}, and without it, were considered.

3.2.2. Wrapper approach for selecting the influencing variables on the DHW heat use

Choosing the proper set of influencing variables is a crucial step for the DHW heat use prediction. The use of irrelevant and redundant input variables in the model leads to an increase in computational demand, an inadequate interpretation of the model, and generally makes prediction more complicated and less accurate. Traditionally, three different approaches may be used for feature selection: Filtering, Wrapper, and Embedded method [51].

In this article, the Wrapper method was used for optimal variables selection. This method is one of the most precise methods, because it detects possible interactions between variables and takes into account the specific characteristics of the prediction algorithm [51]. According to the Wrapper method, first, all the variables were sorted by the absolute value of the correlation criteria between a variable and the DHW energy use. Afterwards, an

iteration algorithm was applied. In each iteration step, one additional variable from the sorted list of variables was added to the model. For each step, parameters and accuracy criteria of the model were recalculated. The obtained criteria of model accuracy on a current step were compared with criteria on a previous step. Thus, parameters that do not improve the accuracy of the model were determined and eliminated from the model, and a set of variables that makes predictions more precise was selected. Despite the higher computational time compared to commonly used analysis based on the correlation matrix (Filtering method), the application of the Wrapper method is a more potent instrument for assessing the impact of different combinations of variables on the DHW heat use and selecting their proper set for accurate prediction [51].

3.2.3. Prediction techniques for modelling DHW heat use based on influencing factors

The prediction techniques for the considered case are presented in Table 3, see Section 4.2. The advanced time series techniques have the ability to take into account explanatory variables. For this reason, some models in Subsection 3.1 were also used for prediction in current conditions. In addition to the models in Subsection 3.1, the availability of data on influencing factors allowed us to apply more diverse prediction techniques.

Group Method of Data Handling (GMDH) is a computer-based method for calculating complex multivariable models. GMDH stands on self-organization theory of mathematical models. The method recursively combines selective submodels (base function) to obtain a more accurate predictive model. On each step of the modelling, the number of submodels included in the main model is gradually growing. In this way, the accuracy and complexity of the model are increasing. The GMDH allows us to find a model structure with optimal complexity based on the minimum value of an external criterion [52]. As base functions in GMDH can be used various models: linear, polynomials, exponential, etc.

Partial Least Squares Regression (PLSR) is a powerful instrument for prediction in conditions when a large number of independent variables is used in the model. PLSR works well with highly collinear variables, too. This method performs the decomposition of the initial data into a subspace of latent variables (scores and loadings). Latent variables are representing the main features of covariance among the dependent and the independent variables [53]. PLSR calculates the linear regression model via the projection of the predicted variables and the observable variables to a subspace of the latent variables [53].

Support Vector Regression (SVR) is based on the computation of a linear regression function in high dimensional feature space [54], where the input data are mapped via a nonlinear function. SVR is minimizing the generalized error bound [55]. The generalization error bound includes the training error and a regularization term that controls the complexity of the hypothesis space [55]. The comprehensive overview of this method is given in [56].

Ridge and LASSO methods are used to deal with overfitting and variables that may be affected by multicollinearity [57]. Both these methods are based on principals of regularization, i.e. introduction penalties to the coefficients of features. Ridge Regression is penalizing the square of the magnitude of coefficients [58]. LASSO introduces penalties to the absolute value of the magnitude of the coefficients [58].

In Subsection 3.2, the general principles for the DHW heat use modelling were applied in the same way as in Subsection 3.1. The data about DHW heat use and influencing variables from 2015 were used in a training set and data from 2016 were used for testing. The best model was selected based on R2, MAE, and MSE criteria of the model adequacy. The prediction for the meth-

ods mentioned above, was performed in Python, using Statsmodels and GmdhPy packages.

4. Results

This section is divided into two subsections, which examines two situations for modelling with different input data. The hourly prediction based on information from the historical DHW heat use is investigated in Section 4.1. A more favourable situation with using additional influencing variables is shown in Section 4.2.

4.1. Results on hourly DHW heat use based on the historical time series

DHW heat use measurements are widely used for paying utility bills in non-residential buildings in Norway. As a consequence, historical data about hourly DHW heat use are available for building owners for many types of non-residential buildings in Norway, including hotels. Historical data about hourly DHW use provide us with a valuable basis for DHW heat use modelling.

For more precise prediction, the variation of DHW heat use in different periods of time should be taken into account. Certain factors, which explain appropriate variation, can be identified based on the descriptive statistical analysis of the retrospective time series. Box plot is a statistical method, that graphically depicts the median, first quartile and third quartiles, minimum and maximum, and outliers for the data. A visual study of the box plots showed that hourly DHW heat use in the hotel varies depending on the hour of the day, day of the week, and the month, as shown in Fig. 1.-Fig. 3. Fig. 1. and Fig. 3 shows hourly heat use in kW, while Fig. 2. shows average hourly DHW heat use for each day in kW.

It is generally known that changes in the DHW heat use during the day normally is associated with personal hygiene activities. The box plot of the hourly DHW heat use in Fig. 1. indicates that the significant peak of the DHW use could be observed in the morning from 7:00o'clock to 10:00o'clock. The heat use for DHW in the evening looks pretty even, with the small spikes from 22:00o'clock to 23:00o'clock. The minimum of the DHW heat use occurred at night time from 1:00o'clock to 5:00o'clock in the morning.

Weekly variation of the DHW heat use, See Fig. 2., is usually related to the preferences of visitors to make trips on different days of the week. The days of the week in Fig. 2. are displayed from Monday to Sunday. Fig. 2. shows that heat use for the DHW may vary depending on the day of the week. For this specific hotel, the highest average daily DHW heat use in 2015 was observed on Saturdays and the smallest on Mondays.

The box plot of DHW heat use from January till December 2015 is shown Fig. 3. From Fig. 3. the seasonal changes in DHW heat use
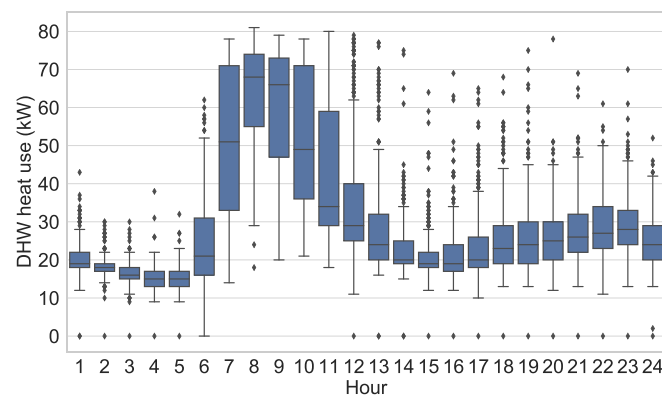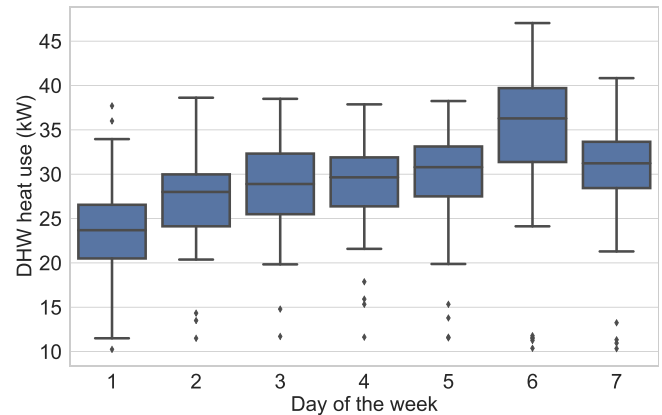

Fig. 2. Box plot of the average hourly DHW heat use for different days of the week in 2015.
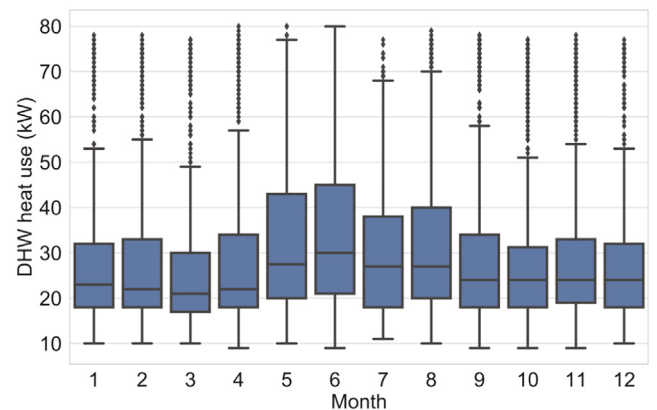

Fig. 3. Box plot of hourly DHW heat use for the different month of the year in 2015.

can be noted. The highest monthly heat use took place from May to September. Such a pattern may arise due to an increase in the number of tourists in the warm season. Another parameter that affecting the monthly heat use is the variation in cold freshwater inlet temperature in the DHW system.

The box plots gave us only rough information about the variation of heat use in different periods of time. However, this method clearly shows that parameters such as hour, day of the week and month should be included in the model. Accordingly, in Situation 1, the retrospective time series of DHW heat use and the hour, day and month were used as inputs for different prediction techniques.

The classical time series modelling techniques, ES and ARIMA showed high values of MAE and MSE, and R2 less than 0.6. Due to the low accuracy of ES and ARIMA models, they were not considered for DHW heat use modelling in further analysis. The NN, Prophet, and XGBoost techniques showed better outcomes. The MAE, MSE, and R2 criteria for these models are presented in Table. 2.

Among the models considered for Situation 1, see Table 2, the Prophet had the best accuracy for hourly DHW heat use modelling. In addition, this model stays robust. The R2 remained equals to 0.76 for both the training and the testing set. The results of hourly prediction based on the Prophet model are shown in Fig. 4. The analysis indicates that most of the actual values of DHW heat use lie within the confidence intervals [59] of the model, as shown in Fig. 4. The predicted versus actual values are distributed around the ideal line, as shown in Fig. 5. This means that the Prophet model developed for Situation 1, can be used for forecasting

Fig. 1. Box plot of hourly DHW heat use in 2015.

Table 2
Prediction modelling based on historical time series of DHW heat use.

Prediction technique	Training set			Testing set		
	R2	MAE	MSE	R2	MAE	MSE
Prophet	0.76	3.8	67.7	0.76	4.46	73.27
NN	0.73	4.4	78.6	0.73	4.7	79.13
XGBoost	0.73	3.56	59.67	0.68	4.11	71.14
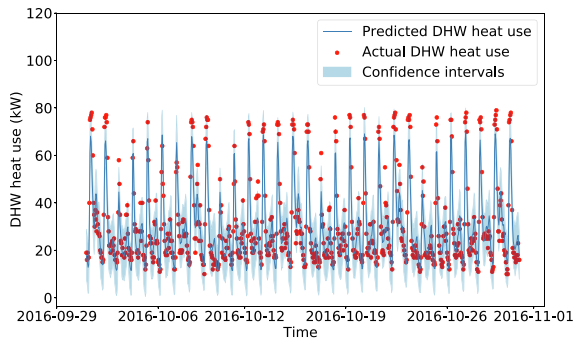
Fig. 4. Hourly modelling of DHW heat use based on the Prophet method in Situation 1.
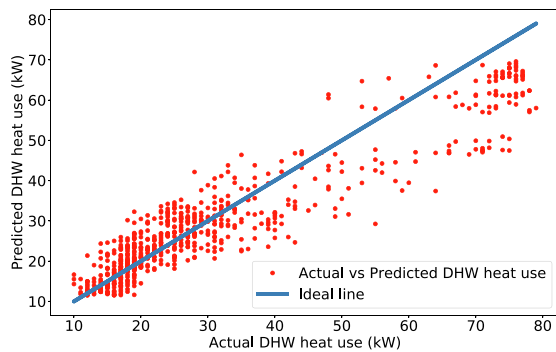
Fig. 5. Predicted by the Prophet method vs. actual values of DHW heat use in Situation 1.

DHW heat use in the hotel. However, despite this fact, the model can be improved. For this purpose, additional variables that affect the DHW heat use should be taken into account. The results of the prediction for corresponding conditions (Situation 2) are presented in Section 4.2.

### 4.2. Results of hourly DHW heat use based on influencing variables

As a part of the investigation for Situation 2, the feasibility of using different variables for DHW heat use modelling was tested. In order to identify the variables that may affect the DHW heat use, data from the hotel's measurement and booking systems were collected, as well as climate data from a weather station located nearby the building. The following variables were considered as potential inputs for the DHW heat use prediction modelling: Gst and Gst_{Lag1} – the number of guests on a given day and the day before, Rm – number of booked rooms in the hotel. Eon – energy use for other needs in the building, T – outdoor air temperature, Rh – relative humidity, Ff – mean wind speed, Pa – atmospheric pressure, H – hour of the week, DoW – day of the week, Mth – month of the year.

The Gst and Gst_{Lag1} are representing only the daily values of the guests presence. To take into account the daily variation of the

guests presence and improve the prediction, the artificial variable Gst_{art} was used. Gst_{art} was identified based on Equation (1). The coefficients of the guests DHW use intensity in Equation (5) were calculated by solving the optimization problem in Equation (6). These coefficients for a given day and the day before are shown in Fig. 6. and Fig. 7. The patterns in Fig. 6. and Fig. 7. coincided with a shape of the box plot of hourly DHW heat use in Fig. 1, which represents the hourly habits of DHW use in the hotel. The coefficients calculated on the basis of the data for 2015 were used to determine Gst_{art} in 2016. Models with and without application of artificial variable Gst_{art} were tested to determine the most accurate.

The Wrapper algorithm was applied to categorise the best set of influencing variables. It was found that the most influencing parameters for all models are related to the guest presence in the building. Gst and Gst_{Lag1} showed the best result for the models created only based on measured data, and Gst_{art} for models where this artificial variable was applied. These three parameters allowed us to receive quite reliable models of DHW heat use in the hotel.

Rm, number of booked rooms, is highly correlated with a number of guests. It does not give additional information and quality to the models. For this reason, Rm was taken out of consideration. Application of mean wind speed, Ff, and atmospheric pressure, Pa in the models, did not increase their accuracy. In this regard, these parameters also should be excluded from modelling. When relative humidity, Rh, was used, only a few models showed insignificant improvement. Thus, application of Rh is usually not reasonable.T, outdoor air temperature *and Eon,* energy use for other needs improved the models, but not much. For example, when adding these parameters to certain models, $R2$ coefficient was increased by several percents. In some instances, the application of T and Eon may be useful for modelling. However, it should be mentioned that when choosing these parameters, we also must take into consideration that some data, such as weather data, will not be readily available when we are running the prediction. For analysis of the historical data, knowledge about all the data is available, but for forecasting, meteorological and energy data must be forecasted as well, which brings additional uncertainty into the prediction.
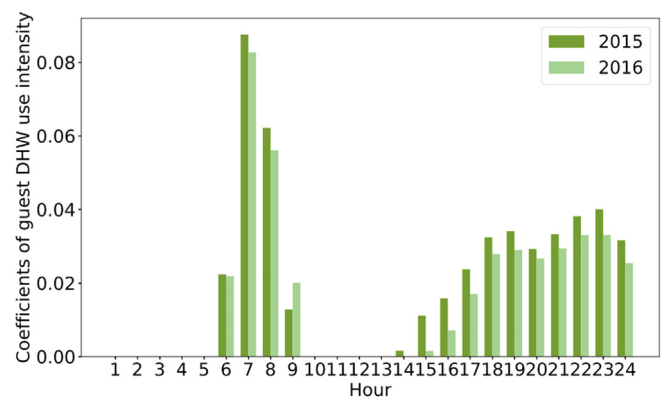
Fig. 6. Coefficients of the guest DHW use intensity based on the booking in the given day in the hotel in 2015–2016.
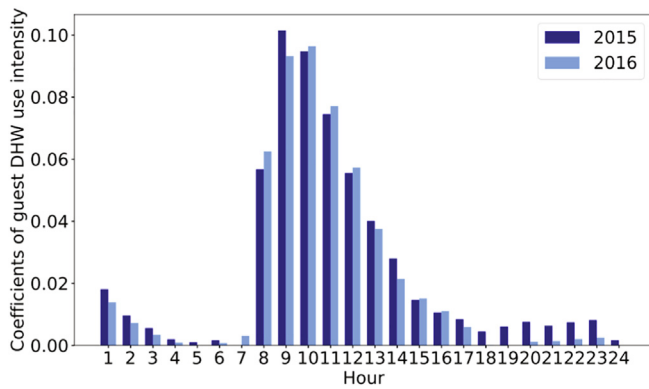
Fig. 7. Coefficients of the guest DHW use intensity based on the booking one day before in the hotel in 2015–2016.

The parameters hour (*H*), day of the week (*DoW*), and month (*Mth)* represented changes in the DHW heat use in different periods of time. In complex and accurate models such as Prophet, NN and XGBoost, applying these parameters gave us good effects. However, some models were unable to extract useful information from *H*, *DoW*, and *Mth* for DHW heat use prediction.

Since the main target of modelling was to build a more accurate model, all parameters that may improve the accuracy of modelling were taken into account. Generally, two sets of influencing variables showed the best outcomes:

a) the set of variables without using the artificial variable Gst_{art}: Gst, Gst_{Lag1}, T, Eon, H, DoW, and Mth;

b) the set of variables with using the artificial variable Gst_{art}: Gst_{art}, T, Eon, H, DoW, and Mth.

In order to select the most accurate DHW heat use prediction model, nine different prediction techniques, see Table 4, were tested. For the set of variables that do not include Gst_{art}, the MAE, MSE and R2 criteria of models adequacy were specified in Table 3. On the other hand, Table 3 contains the same criteria for prediction based on Gst_{art}.

It should be noted that unacceptably inaccurate models were removed from consideration. Therefore, such models are not included in Table 3. When the set of the variables without Gst_{art} was used, only Prophet, NN, XGBoost, and GMDH models showed satisfactory results of prediction. On the contrary, the application of artificial variable Gst_{art} allowed us to improve the accuracy of prediction. Therefore, more models met the minimum acceptable criteria with R2 greater than 0.65. In general, the models in Table 4 showed better outcomes compared to the models in Table 3. However, for advanced and complex prediction techniques, the effect of application of Gst_{art} was less visible. These consequences can be explained by the fact that Prophet, NN, XGBoost, and GMDH models can better reflect hidden relationships in explanatory variables than the other models in Table 4. Accordingly, these models may give us a quite reliable forecast based on both sets of variables, both with and without the application of Gst_{art}.

Table 3 and Table 4 indicate that Prophet and NN are the best models for hourly prediction DHW heat use in the hotel. The NN model showed better performance on the training set, while Prophet on a testing set. For the NN model, R2 calculated on a training set was 0.89. Nevertheless, for the testing set, this criterion was reduced to 0.8. Such changes of R2 may indicate a tendency of the given model to overfitting.

Compering to the NN model, the Prophet model allowed us to obtain more robust results with minor changes in R2, MAE, and MSE. For this reason, the Prophet method was selected as the best model for the DHW heat use prediction in the considered hotel. The result of the hourly modelling based on the testing data set is shown in Fig. 8. Fig. 8. and Fig. 9. confirm the adequate performance of the model. As shown in Fig. 8., the actual values of DHW heat use were within the confidence intervals of the Prophet model. The predicted versus actual values lies close to the ideal line, as shown in Fig. 9.

The study confirmed that by means of easily accessible data, it is possible to obtain a fairly accurate model for the DHW heat use prediction for a hotel. Comparing the results in Situation 2 with a model that uses only historical DHW heat use data (Situation 1), the application of additional variables (Situation 2) allowed us to improve the accuracy of prediction. For example, R2 was increased from 0.76 to 0.83 in the testing set, if using an artificial variable. For all considered cases, the Prophet model proved to be an accurate and reliable model that can reflect periodical changes in DHW heat use. The developed models are useful for the DHW heat use modelling for other hotels under similar conditions.

5. Conclusions

Predictive modelling is a powerful instrument for increasing the efficiency of the DHW heat use in buildings. The modelling involves the following tasks: selecting input variables for prediction, determining the prediction technique, and parameters for the model. This article highlightes the issue of the DHW heat use prediction for a hotel located in Norway.

For accurate prediction, it is crucial to select a proper set of input variables. These variables should include the main factors that affect the DHW heat use in the building. Yet, the data availability may vary from one building to another. Therefore, two common situations with data availability were considered. Situation 1 assumed that only information from the historical DHW heat use might be used for prediction. Situation 2 demonstrated more favourable conditions, where also additional variables that affect DHW heat use were included in the model.

The Wrapper approach showed high efficiency in determining the variables that should be included in the prediction model. This approach indicated that the main factor that affected the DHW heat use in the hotel were number of guests booked in the hotel on the given day and the day before. Nevertheless, the number of guests are collected on a daily basis, which makes them less efficient for hourly modelling. Therefore, to improve the accuracy of the hourly model, the introduction of an additional artificial variable was proposed. This artificial variable reflects the hourly inten-

Table 3
Prediction modelling without using the artificial variable $Gst_{art.}$

Prediction technique	Training set			Testing set		
	R2	MAE	MSE	R2	MAE	MSE
Prophet	0.8	3.7	56	0.79	4.6	63
NN	0.88	3.18	33.65	0.8	4	59
XGBoost	0.89	2.5	25	0.78	3.8	51
GMDH	0.81	4.35	58.9	0.64	4.7	116.3

Table 4
Prediction modelling with using the artificial variable $Gst_{art.}$

Prediction technique	Training set			Testing set		
	R2	MAE	MSE	R2	MAE	MSE
Prophet	0.82	3.26	52.4	0.83	3.67	52.46
NN	0.89	2.85	31.04	0.8	3.48	59
XGBoost	0.88	2.64	28.7	0.81	3.12	45.04
GMDH	0.82	3.57	56	0.8	3.67	62
SVR	0.77	3.88	69.52	0.72	4.90	79.03
ARMAX	0.85	3.8	43.3	0.66	5.33	83.2
PLSR	0.77	4.13	69.25	0.73	4.91	77.21
Ridge	0.79	4.52	64.39	0.68	6.52	90.50
Lasso	0.79	4.47	64.64	0.70	6.19	84.97
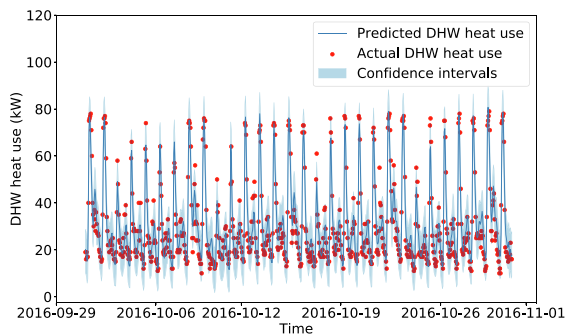
Fig. 8. Hourly DHW heat use model based on the Prophet method in Situation 2.
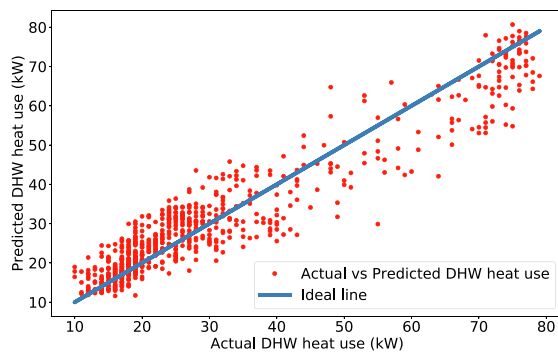
Fig. 9. Predicted by the Prophet method vs. actual values of DHW heat use in Situation 2.

sity of the guests DHW use with a major peak of the heat use in the morning and a small peak in the evening. The method for identifying this variable was based on an optimization problem, presented in the article. In addition, several other factors were identified, that may increase the accuracy of the prediction to a certain extent.

Identifying the DHW heat use model requires a comparison of various prediction methods. Selection of the best method among those considered should be based on the criteria of model adequacy. In order to obtain an accurate and reliable DHW heat use model for a hotel, ten different time series and machine learning prediction techniques were tested. Among considered methods, the Prophet model showed the best accuracy and robustness for the DHW heat use prediction in the case study. In Situation 1, the R2 criterion for testing set obtained via the Prophet model was 0.76. However, with the introduction of additional explanatory variables in the model (Situation 2), the R2 criterion was increased to 0.83. The outcomes of the hourly DHW heat use predictive modelling for the hotel could also find application in similar building types.

CRediT authorship contribution statement

Dmytro Ivanko: Conceptualization, Methodology, Formal analysis, Software, Investigation, Writing - original draft, Visualization, Writing - review & editing. **Åse Lekang Sørensen:** Formal analysis, Writing - review & editing. **Natasa Nord:** Conceptualization, Formal analysis, Writing - original draft, 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.

Acknowledgment

References

[1] European Commission, Energy Performance of Buildings Directive, Available from: https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-performance-of-buildings (Accessed October 16, 2019).

[2] T. Abergel, B. Dean, J. Dulac, I. Hamilton, Global Status Report: Towards a Zero-Emission, Efficient, and Resilient Buildings and Construction Sector, in: Global Alliance for Buildings and Construction, 2008.

[3] S. Pezzutto, S. Croce, S. Zambotti, L. Kranzl, A. Novelli, P. Zambelli, Assessment of the Space Heating and Domestic Hot Water Market in Europe - Open Data and Results, Energies 12 (9) (2019) 1760.

[4] B. Bohm, Production and distribution of domestic hot water in selected Danish apartment buildings and institutions. Analysis of consumption, energy

efficiency and the significance for energy design requirements of buildings, Energy conversion and management, 67 (2013) 152-159.

[5] S. Medved, S. Domjan, C. Arkar, Sustainable Technologies for Nearly Zero Energy Buildings, Springer International Publishing, 2019.

[6] E. Fuentes, L. Arce, J. Salom, A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis, Renew. Sustain. Energy Rev. 81 (2018) 1530–1547.

[7] J. Xing, P. Ren, J. Ling, Analysis of energy efficiency retrofit scheme for hotel buildings using eQuest software: A case study from Tianjin, China, Energy and Buildings 87 (2015) 14–24.

[8] T. Tsoutsos, S. Tournaki, M. Frangou, M. Tsitoura, Creating paradigms for nearly zero energy hotels in South Europe, AIMS Energy 6 (1) (2018) 1–18.

[9] D. Todorovic, M. Tomic, R. Bojanic, D. Bajatovic, A.S. Andelkovic, A comparative analysis of a heat pump application with grey wastewater source for domestic hot water preparation in hotels, J. Therm. Anal. Calorim. (2020) 1–14.

[10] E. Kyriaki, E. Giama, A. Papadopoulou, V. Drosou, A. Papadopoulos, Energy and environmental performance of solar thermal systems in hotel buildings, Procedia Environ. Sci. 38 (2017) 36–43.

[11] A. Heidari, D. Khovalyg, Short-term energy use prediction of solar-assisted water heating system: Application case of combined attention-based LSTM and time-series decomposition, Sol. Energy 207 (2020) 626–639.

[12] H. Ahmad, Energy Efficiency Solutions for Domestic Hot Water, Online publisher: bigee.net, (2017) p. 1-5.

[13] T. Tereshchenko, D. Ivanko, N. Nord, I. Sartori, Analysis of energy signatures and planning of heating and domestic hot water energy use in buildings in Norway, E3S Web Conf. 111 (06009) (2019) 1–8.

[14] A. Ferrantelli, K. Ahmed, P. Pylsy, J. Kurnitski, Analytical modelling and prediction formulas for domestic hot water consumption in residential Finnish apartments, Energy Build. 143 (2017) 53–60.

[15] J. de Santiago, O. Rodriguez-Villalón, B. Sicre, The generation of domestic hot water load profiles in Swiss residential buildings through statistical predictions, Energy Build. 141 (2017) 341–348.

[16] A. Chmielewska, M. Szulgowska-Zgrzywa, J. Danielewicz, Domestic hot water consumption in multi-apartment buildings, in: E3S Web of Conferences, EDP Sciences, 2017, pp. 1-8.

[17] N. Beeker, P. Malisani, N. Petit, Statistical properties of domestic hot water consumption, in: 12th REHVA World Congress, CLIMA 2016, 2016, pp. 1-10.

[18] R. Priyadarsini, W. Xuchao, L.S. Eang, A study on energy performance of hotel buildings in Singapore, Energy Build. 41 (12) (2009) 1319–1324.

[19] D. Fischer, T. Wolf, J. Scherer, B. Wille-Haussmann, A stochastic bottom-up model for space heating and domestic hot water load profiles for German households, Energy Build. 124 (2016) 120–128.

[20] I. Verhaert, B. Bleys, S. Binnemans, E. Janssen, A methodology to design domestic hot water production systems based on tap patterns, in: Proceedings of CLIMA 2016: 12th REHVA World Congress, Aalborg, Denmark, 2016, pp. 1-10.

[21] A. Fidar, F. Memon, D. Butler, Performance evaluation of conventional and water saving taps, Sci. Total Environ. 541 (2016) 815–824.

[22] N. Good, L. Zhang, A. Navarro-Espinosa, P. Mancarella, High resolution modelling of multi-energy domestic demand profiles, Appl. Energy 137 (2015) 193–210.

[23] A. Gutierrez-Escolar, A. Castillo-Martinez, J.M. Gomez-Pulido, J.-M. Gutierrez-Martinez, Z. Stapic, A new system to estimate and reduce electrical energy consumption of domestic hot water in spain, Energies 7 (11) (2014) 6837–6855.

[24] L. Fan, G. Liu, F. Wang, V. Geissen, C.J. Ritsema, Y. Tong, Water use patterns and conservation in households of Wei River Basin, China, Resources, Conservation and Recycling, 74 (2013) 45-53.

[25] L.G. Swan, V.I. Ugursal, I. Beausoleil-Morrison, Occupant related household energy consumption in Canada: Estimation using a bottom-up neural-network technique, Energy Build. 43 (2011) 326–337.

[26] C. Barteczko-Hibbert, M. Gillott, G. Kendall, An artificial neural network for predicting domestic hot water characteristics, Int. J. Low-Carbon Technol. 4 (2) (2009) 112–119.

[27] J. Widen, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegard, E. Wackelgard, Constructing load profiles for household electricity and hot water from time-use data - Modelling approach and validation, Energy Build. 41 (7) (2009) 753–768.

[28] R. Hendron, J. Burch, Development of standardized domestic hot water event schedules for residential buildings, in: ASME 2007 Energy Sustainability Conference, American Society of Mechanical Engineers, 2007, pp. 531-539.

[29] J.D. Lutz, X. Liu, J.E. McMahon, C. Dunham, L.J. Shown, Q.T. McCure, Modeling patterns of hot water use in households, in, Lawrence Berkeley National Lab, (LBNL), Berkeley, CA (United States) (1996).

[30] K. Papakostas, N. Papageorgiou, B. Sotiropoulos, Residential hot water use patterns in Greece, Sol. Energy 54 (6) (1995) 369–374.

[31] M. Aydinalp, V.I. Ugursal, A.S. Fung, Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks, Appl. Energy 79 (2) (2004) 159–178.

[32] A. Lomet, F. Suard, D. Chèze, Statistical Modeling for Real Domestic Hot Water Consumption Forecasting, Energy Procedia 70 (2015) 379–387.

[33] A.L. Sorensen, K.B. Lindberg, H.T. Walnum, I. Sartori, U.R. Aakenes, I. Andresen, Heat Analysis for Energy Management in Neighbourhoods: Case Study of a Large Housing Cooperative in Norway, IOP Conf. Ser. Mater. Sci. Eng. 609 (2019) 1–7.

[34] H. Aki, T. Wakui, R. Yokoyama, Development of a domestic hot water demand prediction model based on a bottom-up approach for residential energy management systems, Appl. Therm. Eng. 108 (2016) 697–708.

[35] L. Garbai, A. Jasper, Z. Magyar, Probability theory description of domestic hot water and heating demands, Energy Build. 75 (2014) 483–492.

[36] L. Gelazanskas, K.A. Gamage, Forecasting hot water consumption in residential houses, Energies 8 (11) (2015) 12702–12717.

[37] H. Kauko, O. Stavset, M. Bantle, N. Nord, Energy use in Norwegian non-residential buildings: building regulations, calculations and measurements, in: Proceedings of 8th International Conference on Heating, Ventilation and Air-Conditioning : Cold Climate HVAC, 2015, pp. 1-8.

[38] Z. Pang, Z. O'Neill, Uncertainty quantification and sensitivity analysis of the domestic hot water usage in hotels, Appl. Energy 232 (2018) 424–442.

[39] D. Ivanko, N. Nord, Å.L. Sørensen, T.S.W. Plesser, H.T. Walnum, I. Sartori, Identifying typical hourly DHW energy use profiles in a hotel in Norway by using statistical methods, E3S Web Conf. 111 (2019) 1–8.

[40] Climate data from Norwegian Meteorological Institute. Available from: http://eklima.met.no (Accessed January 13, 2020). in.

[41] R.J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts, 2018.

[42] J. Brownlee, Introduction to time series forecasting with python: how to prepare data and develop models to predict the future, Mach. Learn. Mast. (2017).

[43] Prophet, Available from: https://facebook.github.io/prophet/docs/contributing.html#documentation (Accessed Match 12, 2020).

[44] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.

[45] B. Pan, Application of XGBoost algorithm in hourly PM2. 5 concentration prediction, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2018, pp. 1-8.

[46] D. Chakraborty, H. Elzarka, Early detection of faults in HVAC systems using an XGBoost model with a dynamic threshold, Energy Build. 185 (2019) 326–344.

[47] Introduction to Boosted Trees, Available from: https://xgboost.readthedocs.io/en/latest/tutorials/model.html (Accessed Match 12, 2020).

[48] Z. Raida, Modeling EM structures in the neural network toolbox of MATLAB, IEEE Antennas Propag. Mag. 44 (6) (2002) 46–67.

[49] M.H. Beale, M.T. Hagan, H.B. Demuth, Neural network toolbox, User's Guide, MathWorks, 2 (2010) 77-81.

[50] S. Raschka, Python machine learning, Packt Publishing Ltd, 2015.

[51] A. Miller, Subset selection in regression, Chapman and Hall/CRC, 2002.

[52] H.R. Madala, Inductive learning algorithms for complex systems modeling, cRc press, 2019.

[53] K.H. Esbensen, D. Guyot, F. Westad, L.P. Houmoller, Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design, Multivariate Data Analysis, 2002.

[54] V. Vapnik, S.E. Golowich, A.J. Smola, Support vector method for function approximation, regression estimation and signal processing, in: Advances in neural information processing systems, 1997, pp. 281-287.

[55] Basak Debasish, Srimanta Pal, Dipak Chandra Patranabis, Support vector regression, Neural Information Processing-Letters and Reviews, 11 (1) (2007) 203-224.

[56] I. Steinwart, A. Christmann, Support vector machines, Springer Science & Business Media, 2008.

[57] L. Melkumova, S.Y. Shatskikh, Comparing Ridge and LASSO estimators for data analysis, Procedia Eng. 201 (2017) 746–755.

[58] A. Jain, A Complete Tutorial on Ridge and Lasso Regression in Python, 2016, Available from: https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/ (Accessed March 25, 2020).

[59] M. Gardner, D. Altman, Calculating confidence intervals for regression and correlation, Statistics with confidence. British Medical Journal, London, United Kingdom, (1989) 34-49.