Prediction of DHW energy use in a hotel in Norway

IOP Conf. Series: Materials Science and Engineering **609** (2019) 052018 doi:10.1088/1757-899X/609/5/052018

Prediction of DHW energy use in a hotel in Norway

Ivanko Dmytro[1,*], Nord Natasa[1], Sørensen Åse Lekang[2], Sartori Igor[2], Plesser Thale Sofie Wester[2], Harald Taxt Walnum[2]

[1] Norwegian University of Science and Technology, Norway
[2] SINTEF Buildings and Infrastructure, Oslo, Norway
[*] dmytro.ivanko@ntnu.no

Abstract. Domestic hot water (DHW) systems are significant consumers of energy in buildings. This article is dedicated to hourly and daily DHW energy use modeling, with the aim of achieving energy savings in buildings. The methods investigated in the article were tested using statistical data obtained from a hotel located in Oslo, Norway. For better modeling, the influence of various factors on DHW energy use in the hotel was studied. For this purpose, the wrapper approach was used. The analysis indicates that the most important variable that should be used in the model is number of guests. There are also other factors that can be taken in account, even though they do not have such strong influence. Traditionally, only daily data about number of guests are available in the hotels. These data do not allow us to develop accurate hourly model of DHW energy. The article therefore proposes a method which, based on introduction of artificial variables, improve accuracy of the hourly DHW model. Eight models are compered, based on criteria of their adequacy. The Support vector machine model shows the best results for daily modeling and the Partial least squares (PLS) regression for hourly modeling.

1. Introduction

Buildings are responsible for approximately one third of the energy use in the world [1]. Thus, efficient use of energy in buildings is a topical issue from both an environmental and economic point of view. A domestic hot water (DHW) system is an essential part of most buildings, and contribute to 25-35% of the total energy use [2]. Many studies claim that a large potential for future energy savings in buildings lies in improving operation and design of DHW systems [3]. Mathematical modeling of energy use is a powerful tool for achieving energy saving in buildings. Prediction, data recovery, monitoring of energy use and other important tasks could be solved via accurate and physically valid mathematical modeling.

Recently, much attention was paid to the modeling of energy needs required for heating [4]. Meanwhile, the issue of DHW energy use modeling and prediction has not been studied well enough [3]. The majority of publications in this area are dedicated to the modeling of DHW volumetric consumption in building rather than energy use. These two parameters are strongly positively correlated. The knowledge obtained from the studies about DHW volumetric consumption modeling is valuable for development of advanced models of DHW energy use. Therefore, articles considered in this introduction are dedicated to both prediction of DHW volumetric consumption and energy use in buildings.

For instance, according to [5], DHW energy loads are modeled as a function of draw-off temperatures. For three different systems, the models based on application of neural networks (NNs) are calculated. The results show that the models trained on their associated systems produce errors less than 11%. However, when obtained models were used with new systems, they had significant errors.

A bottom-up approach for DHW energy use prediction is proposed in [6]. The developed prediction model calculates the quantity of hot water and timing of each end-use for the next day from historical data and summarizes these as prediction data.

The necessity of development of daily DHW use models, which do not require strong computation time and information about the residents in the buildings, is stressed in [7]. The authors proposed

application of Auto Regressive Moving Average method for solving this issue. The developed model takes into account the periodicity of one week, the water use of the previous days and random fluctuations. The results are tested on one-year data of DHW use in eight residential homes in France.

DHW use in flats based on the number of rooms and area of the flat are described in [8]. This study is conducted in 626 apartments in Wroclaw, Poland. To build a model, the bootstrap technique is used. Based on the obtained statistical data, a database is created, consisting of randomly simulated buildings with randomly selected flats in different configurations. After this, the regression model is developed, explaining the relationship between DHW use, number of rooms and floor area.

In [9] proposed to consider the DHW use as stochastic variables. The statistical data for the research has been collected from 65 apartments in Budapest, Hungary. The model presents the relationship between DHW use and number of apartments, sorted duration curve of DHW consumption, as well as minimum, average and maximum daily values of DHW consumption in the buildings.

The necessity of accurate hot water use forecasting for future development of demand-side management in residential dwellings is stressed in [10]. Various forecasting models, such as exponential smoothing, seasonal autoregressive integrated moving average, seasonal decomposition by Loess model and a combination of them, were tested on data obtained from 120 houses in UK.

16 equations for prediction of average hot water use in different times of the day are proposed in [11]. The authors consider weekdays and weekends separately. Each day is divided into periods by combining hours with similar DHW use by time of day, type of day and season. It is also proposed to take additional variables into account in the model, to adjust the predicted hot water use: if the household has a dishwasher, clothes washer, only seniors as occupants, or if the residents pay for hot water or not.

Most of the above mentioned studies are focused on residential buildings, because this type of building is taking a big share in national building stocks. The characteristics, regimes of work, and available data in the hotels are significantly different from residential buildings [3]. Therefore, considered methods cannot be directly applied for DHW energy use prediction in hotels. A better methods of DHW energy use prediction in hotels should be developed.

The aim of this article was to model hourly and daily DHW energy use that may be used for achieving energy savings in hotels. The analysis in this article is based on two years hourly data of DHW energy use collected in a hotel located in Oslo, Norway. The focus was on the statement that obtained model of DHW energy use should be accurate, reliable and take into account particular characteristic of the buildings. In order to meet these requirements, the factors that have significant influence on DHW energy use in the hotel were investigated. To improve the accuracy of an hourly model, the procedure of preprocessing daily data for number of guests and extracting information of their influence on DHW energy use on hourly basis was proposed. After that, various methods for daily and hourly DHW energy use modeling were compared. The comparison was carried out according to the following criteria: the coefficient of determination (R2), the average absolute error (MAE) and the mean square error (MSE). The most accurate models of DHW energy use were identified.

2. Description of the hotel and available statistical data

The characteristics of the analyzed hotel are typical for Scandinavian conditions, and it well reflect the trends of DHW energy use in similar types of buildings. The hotel, located in Oslo, Norway, was built in 1938, and reconstructed in 2007. The total area of the building is 4 939 m^2. The building consists of eight floors with 164 guest rooms. All the guest rooms have bathrooms with toilet facilities and shower. Guests usually arrive between 3 p.m. and midnight and check out before noon. According to the 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 energy use for DHW production was collected during several years from a energy meter mouted by the hotel owner. The meters measure electricity delivered to the DHW tanks, which mean that both DHW needs and heat losses in the DHW system are included in the presented DHW energy use. The data about energy for other needs are also known. The daily data about arriving guests and booked rooms in the hotel are available from the hotel reservation system. In addition, in order to investigate the influence of weather conditions on DHW energy use in the hotel, data from the meteorological station in Oslo (Blindern) were used.

3. Methods

We started with the task of choosing the variables which should be taken into the DHW energy model. To determine the proper subset of variables, and taking into consideration characteristics of each modeling method, a wrapper approach of optimal variables selection was used [12]. According to this approach, an iteration algorithm was applied. First, all the variables were sorted by the absolute value of the correlation criteria between a variable and DHW energy use. Then, 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. Thus, parameters that do not improve the accuracy of the model significantly can be determined and eliminated. Despite the higher computational time comparing to correlation matrix analysis, the application of wrapper algorithms is a powerful instrument for assessing the impact of different combinations of variables on DHW energy use and development of accurate prediction models.

3.1 Preprocessing daily data for guest presence

It is known from previous studies that the main factor affecting DHW energy use in a hotel is the number of guests presence [3]. Most hotels have a reservation system, which register number of visitors that check in at the set time, usually after 12 a.m. Thus, the hotel reservation system tells us the number of guests booked into the hotel. However, whether the visitors are actually in the hotel at any given time or not remains unknown.

The peak of DHW energy use in the hotel occurs before 12 a.m. The actual time when visitors are arriving and leaving can vary. Some people can arrive before the set time of check in, and some of them can stay a bit longer in the building after the check-out time. Therefore, the model should take into account both number of guests registered in the reservation system on a given day (***Gst***) and one day before (***Gst*** $_{Lag1}$).

The use of daily data of the number of guests in the hotel cannot significantly improve the accuracy of the hourly model of DHW energy use. Therefore, we propose to introduce an additional artificial variable ***Gst*** $_{art}$**.** We introduce this variable to increase accuracy of hourly DHW model. Eq. (1) was used to identify the numerical value of ***Gst*** $_{art}$ for each separate hour:

$$Gst_{art} = Gst \cdot Cgp_i + Gst_{Lag1} \cdot Cgp_{Lag1.i} \tag{1}$$

where Cgp_i and $Cgp_{Lag1.i}$ were the coefficients of guests DHW use intensity for *i*-hour based on the number of people booked into the hotel on the given day and one day before.

It was suggested to calculate the coefficients of guests DHW use intensity for *i*-hour by solving the following optimization problem:

$$\max\left(corr\left(\begin{matrix} Cgp_{i=1} \cdot \left(\overrightarrow{Gst}\right) + Cgp_{Lag1.i=1} \cdot \left(\overrightarrow{Gst_{Lag1}}\right), \dots, Cgp_{i=24} \cdot \left(\overrightarrow{Gst}\right) \\ +Cgp_{Lag1.i=24} \cdot \left(\overrightarrow{Gst_{Lag1}}\right) \end{matrix}\right), \left\{\vec{E}_{i=1}, \dots, \vec{E}_{i=24}\right\}\right) \tag{2}$$

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

The optimization problem in Eq. (2) gave the values of the coefficients of guests DHW use intensity for each hour of the day. These coefficients are maximizing the correlation between ***Gst*** $_{art}$ and DHW energy use, which makes ***Gst*** $_{art}$-based predictions more accurate. The obtained coefficients for 2015 and 2016 years are shown in Figure 1. Variation of the coefficients values, Figure 1, in different years was not significant. Thus, the values of coefficients from previous years can be used for identification of variable ***Gst*** $_{art}$ in the prediction model. In this article, the values of coefficients were calculated based on the year of 2015, and they were used for predicting the year of 2016.

3.2 DHW energy use modeling

The selection of the DHW energy use model and modeling techniques should be done individually for each building, taking into consideration its characteristics. In this article, the number of models shown in Table 1 was investigated. The detailed explanation of these models can be found in [13, 14]. The best model can be selected by comparing different modeling techniques obtained on the same set of data. In order to compare models, cross validation was used. 70 % of yearly data in 2015 were used in a training

set and 30% in testing of the model. Besides, the models were tested on one-year data from the year 2016. The comparison of the models is performed based on R2, MAE and MSE criteria of the model adequacy. The modeling is performed in Python using the Scikit-learn tool [14].

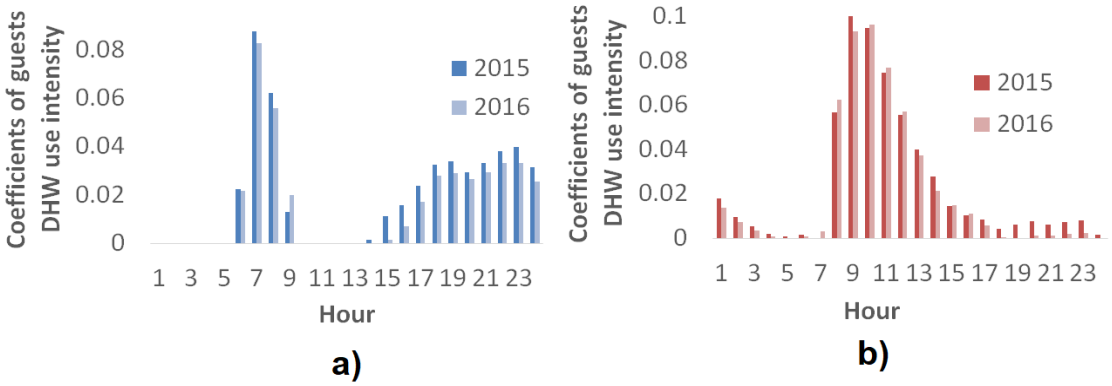

Figure 1. Coefficients of guests DHW use intensity based on the reservation in the given day (a) and one day before (b) in the hotel in 2015-2016

4. Results

The variables ***Gst*** and $\boldsymbol{Gst_{Lag1}}$, which represent number of guests on a given day, and the day before were investigated. The data of energy use for other needs (***Eon***) and number of booked rooms (***Rm***) were also examined. In addition, the influence of the following meteorological parameters were analised: outdoor air temperature (***T***), relative humidity (**Rh**), mean wind speed (***Ff***), atmospheric pressure (***Pa***). The influence of day of the week (***DoW***) and month (***Mth***) was also considered. In addition, the artificial variable $\boldsymbol{Gst_{art}}$ was introduced in the hourly model.

Application of the wrapper algorithm for all the considered models, shown in Table 1, showed approximately the same results. The main parameters for daily DHW energy use modeling in the hotel were ***Gst*** and $\boldsymbol{Gst_{Lag1}}$, and for the hourly model it was $\boldsymbol{Gst_{art}}$. Application of these parameters allowed us to get quite reliable models of the DHW energy use in the hotel. ***Rm*** is highly correlated with number of guests and was taken out of the model, because it does not give additional information and quality to the model. Generally, ***Pa***, ***Ff*** *and* ***Mth*** (in hourly model model) did not increase the accuracy of any model and were therefore eliminated.

DoW, ***T***, ***Rh***, ***Eon*** and ***Mth*** (in daily model) improved the models, but not much. For example, when adding all these parameters to the model, depending on the modeling approach, R2 coefficient increased by 5-15%. Thus, if the target of modeling is to build more accurate model, then these parameters can be taken into account, as we have done in this article. However, if the simple model is more preferable, then only data about number of guests in the hotel can be used.

When choosing our parameters we also must take into consideration that some data, such as weather data, will not be readily available when we are running prediction models. For analysis of 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. In this work we had accurate values of these data, since the models were tested on previous years.

To choose the most appropriate prediction model for DHW energy use in the hotel, eight different models were used, see Table 1. We tested the models using both the cross validation approach and one year ahead prediction. Based on Table 1 for these data sets, the best model for daily modeling was the Support vector machine method. The result of the daily modeling based on the cross validation testing of the data set is shown in Figure 2. For daily model, R2 equals 0.881 for the Support vector machine model based on the cross validation of the data set, and 0.777 for one year ahead data set. For hourly model, Ridge regression gave the best results based on the cross validation of the data. However, for one year ahead prediction Partial least squares (PLS) regression was more accurate. Since PLS regression was more stable, the preference was given to this model. The hourly DHW energy use modeling based on PLS regression is shown in Figure 3.

Table 1. Comparison of different hourly and daily models of DHW energy use

Period	Daily model						Hourly model					
Testing data	Cross validation			Testing based on the next year data			Cross validation			Testing based on the next year data		
Type of regression model	R2	MAE	MSE	R2	MAE	MSE	R2	MAE	MSE	R2	MAE	MSE
Support vector machine	0.881	30	2485	0.777	39	3919	0.781	4	66	0.725	5	79
Partial least squares	0.855	32	3030	0.776	34	3928	0.780	4	66	0.731	5	77
Ridge	0.855	34	3029	0.777	34	3922	0.794	5	62	0.685	7	90
Lasso	0.855	33	3030	0.777	34	3923	0.794	5	62	0.704	6	85
Linear Discriminant Analysis	0.776	40	4683	0.664	48	5903	0.768	4	70	0.670	4	95
Stochastic Gradient Descent	0.855	32	3030	0.777	34	3923	0.794	5	62	0.686	6	90
Bayesian Ridge	0.855	33	3030	0.777	34	3919	0.794	5	62	0.685	7	90
Passive Aggressive	0.840	36	3342	0.735	38	4662	0.720	4	84	0.712	5	83

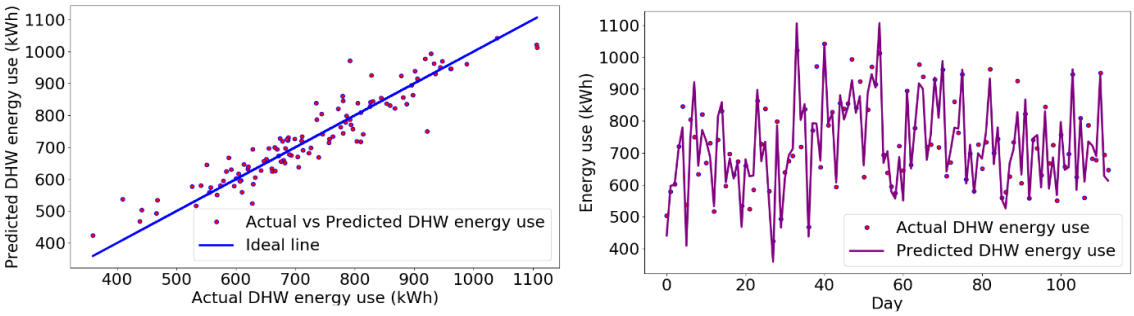

Figure 2. Daily modeling of DHW energy based on Support vector machine method

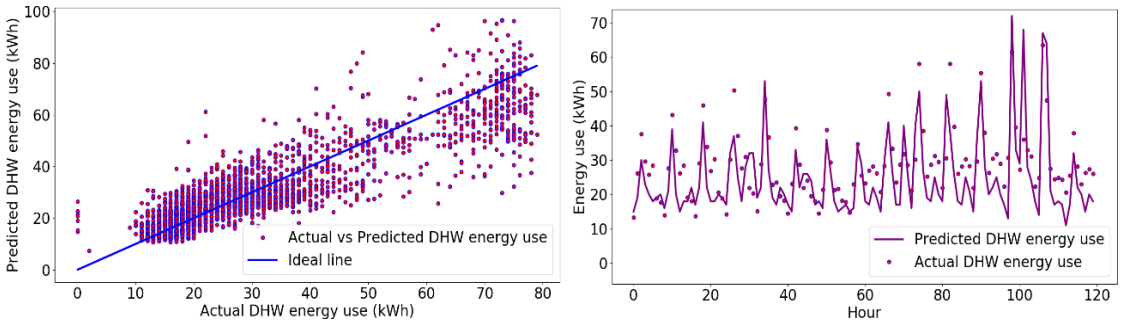

Figure 3. Hourly modeling of DHW energy based on PLS method

The investigated methods of the hourly and daily models could find application for the prediction of DHW energy in similar types of buildings. In addition, these models are useful for DHW energy use modelling in hotels in Norway under the similar conditions.

5. Conclusion

Prediction of the DHW energy use in buildings is a complex task, due to previously lower focus on the DHW energy use and high requirement for relevant, but not easily available data. This article focused on modelling DHW for a typical hotel located in Norway. The wrapper approach shows its high efficiency for determining variables affecting DHW energy use in the hotel. The analysis indicated that the main variables that influence the DHW energy use were numbers of guests registered in the reservation system during the given day and the day before. However, the daily values of the guest numbers did not allow us to develop an accurate hourly model for the DHW energy use. Therefore, introduction of the additional artificial variables, which explain the hourly intensity of the guests DHW

use was proposed. The method of identifying these variables based on solving optimization problem was shown in the article. Selection of the best DHW energy use model requires comparison of different models based on the criteria of models adequacy. Appropriate comparison of the models for the hotel showed that the best daily model was based on the support vector machine method, and the hourly model obtained by using the PLS regression.

Acknowledgement
This article has been written within the research project "Energy for domestic hot water in the Norwegian low emission society". The authors gratefully acknowledge the support from the Research Council of Norway (ENERGIX-programme), SINTEF Building and Infrastructure, Department of Energy and Process Engineering at NTNU, Drammen Eiendom, Omsorgsbygg, Boligbygg, OBOS, Olav Thon Gruppen, Armaturjonsson, Høiax, Geberit, Uponor and FM Mattsson.

References
[1] Dean B, Dulac J, Petrichenko K, Graham P. Towards zero-emission efficient and resilient buildings.: Global Status Report. 2016.
[2] Bøhm B. 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. 2013;67:152-9.
[3] Fuentes E, Arce L, Salom J. A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis. Renewable and Sustainable Energy Reviews. 2018;81:1530-47.
[4] Sretenović A, Jovanović R, Novaković V, Nord N, Živković B. Prediction of hourly heating energy use for HVAC using feedforward neural networks.
[5] Barteczko-Hibbert C, Gillott M, Kendall G. An artificial neural network for predicting domestic hot water characteristics. International Journal of Low-Carbon Technologies. 2009;4(2):112-9.
[6] Aki H, Wakui T, Yokoyama R. Development of a domestic hot water demand prediction model based on a bottom-up approach for residential energy management systems. Applied Thermal Engineering. 2016;108:697-708.
[7] Lomet A, Suard F, Chèze D. Statistical Modeling for Real Domestic Hot Water Consumption Forecasting. Energy Procedia. 2015;70:379-87.
[8] Chmielewska A, Szulgowska-Zgrzywa M, Danielewicz J. Domestic hot water consumption in multi-apartment buildings. Conference Domestic hot water consumption in multi-apartment buildings, vol. 17. EDP Sciences, p. 00014.
[9] Garbai L, Jasper A, Magyar Z. Probability theory description of domestic hot water and heating demands. Energy and Buildings. 2014;75:483-92.
[10] Gelažanskas L, Gamage KA. Forecasting hot water consumption in residential houses. Energies. 2015;8(11):12702-17.
[11] Lutz JD, Liu X, McMahon JE, Dunham C, Shown LJ, McCure QT. Modeling patterns of hot water use in households. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); 1996.
[12] Miller A. Subset selection in regression: Chapman and Hall/CRC, 2002.
[13] Everitt BS, Dunn G. Applied multivariate data analysis: Arnold London, 2001.
[14] Raschka S. Python machine learning: Packt Publishing Ltd, 2015.