Energy and Buildings 133 (2016) 478–488

Contents lists available at ScienceDirect

Energy and Buildings

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

Applied machine learning: Forecasting heat load in district heating system

Samuel Idowu*, Saguna Saguna, Christer Åhlund, Olov Schelén

Luleå University of Technology, Luleå, Sweden

ARTICLE INFO

Article history:
Received 23 January 2015
Received in revised form
28 September 2016
Accepted 29 September 2016
Available online 8 October 2016

Keywords:
Data driven modeling
District heating
Energy efficiency
Machine learning
Smart cities

ABSTRACT

Forecasting energy consumption in buildings is a key step towards the realization of optimized energy production, distribution and consumption. This paper presents a data driven approach for analysis and forecast of aggregate space and water thermal load in buildings. The analysis and the forecast models are built using district heating data unobtrusively collected from 10 residential and commercial buildings located in Skellefteå, Sweden. The load forecast models are generated using supervised machine learning techniques, namely, support vector machine, regression tree, feed forward neural network, and multiple linear regression. The model takes the outdoor temperature, historical values of heat load, time factor variables and physical parameters of district heating substations as its input. A performance comparison among the machine learning methods and identification of the importance of models input variables is carried out. The models are evaluated with varying forecast horizons of every hour from 1 up to 48 h. Our results show that support vector machine, feed forward neural network and multiple linear regression are more suitable machine learning methods with lower performance errors than the regression tree. Support vector machine has the least normalized root mean square error of 0.07 for a forecast horizon of 24 h.

1. Introduction

The current increase in global demand for energy is a major challenge for energy production and its distribution [1]. To improve energy sustainability, there is a need to approach the consumer end rather than focus at the production side alone. This is an important consideration since a fully optimized energy grid can only be realized with optimization at both the production and the consumer ends [2]. Forecasting total energy consumption is essential for the economical and technical planning of power generation. Also, forecasting energy consumption in buildings is crucial for optimizing its energy use. Energy consumption optimization can either be carried out on a grid scale (e.g., large central accumulators in district heating systems (DHS) and the conversion of electricity energy to thermal energy storage) or on a smaller scale such as in buildings.

A survey of recent methods employed in the estimation of energy-demand, for instance, thermal and electric energy in buildings show that there are two broad categories, the *forward (classical)* and the *data-driven (inverse)* methods [3]. The former uses equations that describe the physical behavior of a system to predict an output, while the latter approach relates to supervised machine learning (ML) methods where measurements of input and output variables of a system are collected, and then used to mathematically describe the system [3]. Recently, several ML methods are used for predicting energy-demand, such as, support vector machine (SVM) [4], multiple regression [5,6] and neural networks based methods [7,8]. In specific to thermal load forecast in DHS, some of the advantages of a data-driven approach over a classical approach include the ability to discover models from large volume of data and the ability to adapt and update models based on new data [9].

A more comprehensive classification of methods used in building load forecast can be broadly grouped into physical models, black-box models and Grey-box models [10,11]. In relation to the two categories mentioned in [3], physical models relate to the classical methods, while black-box models relate to data-driven methods. In comparison to black-box models, physical models are costly and time consuming due to their requirement of very large number of parameters and building or system information as input. Black-box models on the other hand are easy to build and usually require data over a long period for its training purpose. Grey-box models are generally used to address the drawbacks in physical models and the difficulty in determining optimal parameters associated with black-box models. While physical models suffer from poor generalization capabilities, grey-box models provide a balance between high accuracy and good generalization capabilities, by extracting the mathematical model/structure from system's

* Corresponding author.
E-mail addresses: sam.idowu@outlook.com, samup4web@gmail.com (S. Idowu).

http://dx.doi.org/10.1016/j.enbuild.2016.09.068
0378-7788/

physics, and estimation of model parameters from measured data. As a consequence, grey-box models require more effort to develop, have good generalization capabilities in reference to black-box models, and demonstrate higher accuracy compared to white-box models [11].

Several DHS related work are limited to predicting net consumption energy at the production end [12–14,8]. These studies fall under the category of the *top-down* approach [3]. Forecasting thermal load in buildings can be prone to high error rates due to the highly stochastic nature of the consumer load pattern [12]. Grosswindhager et al. [12] stated that it is required to build several individual models for the consumer ends of a DHS due to the stochastic nature of its data. Gadd and Werner [14] presented a novel assessment method which describes daily variations of heat load in DHS. It describes variation in heat load patterns and could be used for the design and planning of storage in a DHS network.

In related work, a number of studies have looked more closely at thermal load forecasting methods for DHS [12,8,7,15,13]. Grosswindhager et al. [12] presented an approach for on-line short term load forecast using *seasonal autoregressive integrated moving average* models in state space representation. Kato et al. [8] proposed a thermal load prediction method which uses a recurrent neural network to deal with the dynamic variations of heat load and its characteristics. The approach shows decent prediction accuracy for non-stationary heat load. Sakawa and Ushiro [7] proposed a load prediction method which is robust enough to handle cases of outliers and missing data. The method uses a simplified robust filter and a three-layered neural networks. Nielsen et al. [15] used a grey-box approach to identify the model that links the heat consumption in a large geographical area to its climate and the calendar information. The process involved a theoretical based identification of an overall building model structure followed by a data based modeling.

Zhou et al. [10] integrated air temperature/relative humidity and solar radiation prediction modules with a grey-box model for hourly building thermal load prediction. The author reported performance of the building load prediction model is satisfactory, while the weather prediction is occasionally affected by temporary weather changes. Afram et al. [11] developed and compared the performance of black-box models and grey-box models of a residential heating, ventilation and air conditioning (HVAC) system. All models performed well and were able to predict the outputs correctly, while black-box model, artificial neural network (ANN) performed overall best in most prediction tasks. Wang and Tian [13] applied wavelet analysis in combination with neural network, and its evaluation shows that the approach is suitable for short-term heat load forecast.

Furthermore, while the top-down approach emphasizes on the total energy consumption in a grid at the production side, the bottom-up approach on the other hand focuses on individual consumption end [3]. A comparative review of recently developed models for the prediction of building energy consumption is provided in Zhao and Magoulès [16]. The identified methods in the review are classified into engineering, statistical, neural networks, support vector machines and grey models. A recent study [17], applied probabilistic methods and heuristics for fault detection and ranking of anomalies. The methods are applied on hourly data from district heating substations. Bacher et al. [18] recently employed a *bottom-up* approach with focus on the consumer end and specifically consider single family buildings. The work used computationally effective recursive least squares scheme with meteorological variables as model input. The model presented provides forecast up to 42 h horizon [18]. Serban and Popescu [19] presented a methodology for prognosis of domestic hot water consumption in DHS using time-series analysis. The work modeled hot water heating load in a block of flats with 60 apartments. The study concluded that time-series analysis is powerful and appropriate for predicting thermal load in DHS. Nikovski et al. [20] proposed a general method for controlling building zone air temperature by setting temperature set-points. The method uses building thermal model based on thermal circuit identified from collected sensor data. The building thermal dynamics was reduced to a markov decision process (MDP) whose decision variables are the sequence of temperature set-points over a suitable horizon. This method saved cost significantly, sometimes exceeding 50% with respect to control strategies in buildings such as the night set-up and demand limiting strategy.

Generally, previous studies consider meteorological variables as major influencing factors to forecasting load in buildings. Likewise, Wang and Tian [13] classifies the thermal load influencing factors in a DHS as *external* and *internal* factors. The former refers to meteorological variables and building occupant's behavioral factors. The latter refers to factors, which relates to the operational characteristics of a district heating substation such as the supply temperature.

The research in this paper is a step further from the preliminary research project described in Idowu et al. [21]. The project aims to achieve a more efficient co-operation of a combined heat and power (CHP) plant and a DHS, where the key focus is to reduce (a) energy consumption, (b) emissions such as CO_2, (c) fossil fuel consumption in CHP plant and (d) peak demands. It will employ energy saving strategy (ESS) in buildings. The ESS will maintain the operation of substations within set values or ranges, e.g., maintaining a set value of return temperature for a specific condition. The forecast of heat demand estimate is key for the optimal control of the ESS.

This paper presents a bottom-up approach to the analysis and forecasting of heat load (building space and domestic hot water) in buildings using ML techniques. The ML methods used are SVM, feed forward neural network (FFNN), multiple linear regression (MLR) and regression tree. The forecasting models are built using district heating data collected in a non intrusive manner from five multi-family and five commercial buildings located in Skellefteå, Sweden, combined with outdoor temperature measurements and forecast. The data collection period is between mid February and early April 2014. The basic model inputs used include outdoor temperature, historical values of heat load, time factor variables and parameters of district heating substations. As a contribution, the application and comparison among the ML methods and the identification of the importance of the internal factors as model input variables is carried out. The forecast model for each building is evaluated with varying forecast horizon up to 48 h. This paper shows a comparative analysis among the energy consumption patterns in building using methods such as partial least square method. The outcome of this paper is intended to be used for the optimal control systems in a building energy management system, which can also be applied in other related work.

The content of this paper is organized as follows. Section 2 presents the background theory of DHS, the description of the target system and theory of ML methods used. Section 3 describes the analysis and modeling processes in this work. Section 4 presents the results while Section 5 presents the discussion. Section 6 presents the conclusion and the future work.

2. Theory

This section presents a technical overview of district heating systems. A high-level description of the target system corresponding to the research project under which this work is done is presented. In this paper, our focus lies mainly on the consumption side. Finally, this section also presents four ML methods used in this paper.

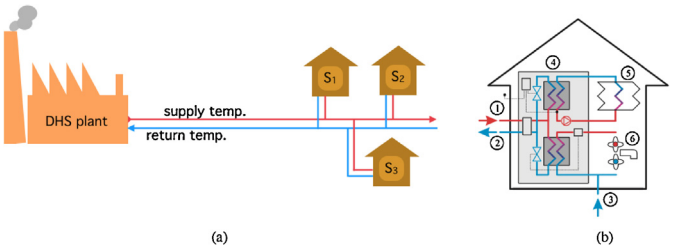

Fig. 1. (a) A schematic diagram showing a district heating system network and its plant. (b) Schematic illustration of a building with a district heating substation. Indicated in the figure are the primary water supply (1); primary return water (2); tap water supply (3); district heating substation including heat exchangers, electronic energy meter and control system with related sensors (4); heating system (5); and tap water (6) [17].

2.1. District heating systems

DHS are dominantly used for space and water heating in residential and commercial buildings, and for industrial heating purposes. For instance, about 90% of apartment buildings in Sweden are heated using this technology [17]. District heating produces heat energy at a centralized location (e.g., CHP plant) and supplies it to residential and commercial buildings primarily for space and domestic water heating. A typical DHS has three main parts. The *Production-side*, which usually consists of a co-generation plant and/or a heat-only boiler station, which is run by fuel such as biofuel, oil or gas. The *Distribution network*, which consists of insulated pipes of varying diameters carrying hot water through the entire grid. The supply pipelines transport hot water to substations while return pipelines transport used water back to the production side. The last part is the *Consumer-side*, it consists of a substation where thermal energy is transferred from the primary to the secondary network via a heat-exchanger. Fig. 1a shows the basic schematic drawing of a CHP plant and a DHS. The red and blue lines denote supply and return pipes respectively. Fig. 1b shows the schematic illustration of a building with a district heating substation highlighting the heat exchangers, which are used for the thermal energy transfer between the primary and the secondary network (Table 1).

The net heat energy, Q_{net} delivered to the entire grid is mainly a function of the supply temperature, T_{ps}, the return temperature, T_{pr}, and the flow rate, m, measured at the production side as shown in Eq. (1). Q_{net} can be formulated as a function of heat energy delivered at each substation as shown in Eq. (2).

$$Q_{net} = c \times m \times (T_{ps} - T_{pr}) \tag{1}$$

$$Q_{net} = Q_{loss} + \sum_{i=1}^{n} Q_{S_i} \tag{2}$$

where Q_{S_i} is the heat demand at substation S_i. Q_{loss}, which varies based on soil temperature around the pipes, is the heat lost during energy transport. An electronic energy meter in a substation calculates the thermal power, Q_{S_i}, received from the distribution network. This includes three main sensors, which measure m, T_{ps} and T_{pr}. The thermal energy is calculated from the flow rate and primary temperature difference

$$Q_{S_i} = c \times mt' \times (T_{ps}t' - T_{pr}t')\,dt', \tag{3}$$

where c is the specific heat of the liquid in the distribution network [17]. The top-down approach focuses on Q_{net} while the bottom-up approach considers Q_{S_i}. The heat demand, Q_{S_i} at the substation is affected by various external and internal factors. The external factors are meteorological variables, building characteristics and social behavioral factors of occupants. The internal factors are those relating to the physical dynamics within a district heating substation. Table 2 shows the two groups of identified influencing factors based on the classification described in [13].

Table 1
Table of notations.

Symbol	Description
t	Time
T_{ps}	Primary supply temperature
T_{pr}	Primary return temperature
ΔT	Primary temperature difference, $T_{ps} - T_{pr}$
m	Mass, flow rate
P	Thermal power. Derived from the energy calculated by the energy meter
D_w	Day of the week
H_d	Hour of the day

Table 2
Influencing factors of heat load in DHS.

Internal factors	External factors	
	Behavioral/ seasonal factors	Meteorological factors
Supply temperature (T_{ps})	Hour of day (H_d)	Outdoor temperature (T_{out})
Return temperature (T_{pr})	Day of week (D_w)	Humidity
Supply pressure	Month of year	Solar radiation
Flow rate (m)		Wind

2.2. System architecture

In this section, we present the high-level architecture of proposed system under which this work is done. Fig. 2 shows the architecture and its key components. In this work, the tasks under consideration, as highlighted in the figure by "√", are data collection and analysis, data aggregation and preprocessing, and the application of ML for thermal load prediction.

A primary objective of this architecture is to improve the overall fuel efficiency of a CHP plant by increasing the ΔT, as more electricity can be produced by colder cooling water [17]. To achieve this, it is necessary for each substation to return the least possible T_{pr} back to the CHP. In the proposed system, we implement an ESS at the substation end of a DHS network. The ESS controller primarily depends on the generated outcome from the ML predictor and the current state of the district heating network.

The data aggregation involves merging sources of data from each substation with its locally collected weather information. In addition, the necessary time information are merged with their corresponding data instances. The output variables from the data

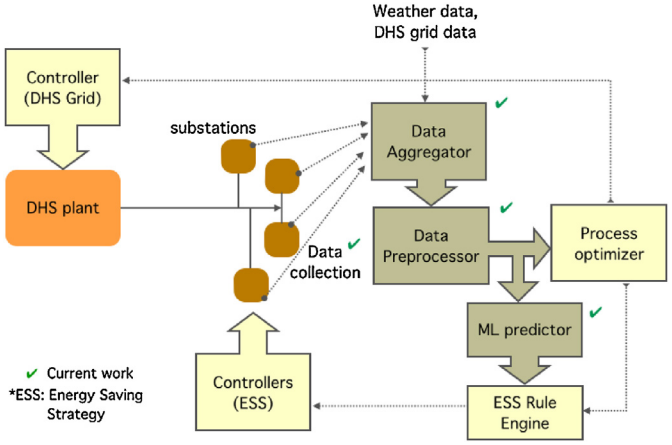

Fig. 2. A high level system architecture of the target system, highlighting the aspects of this work [21].

aggregation unit (see Fig. 2) are T_{out}, P, T_{ps}, T_pr, ΔT, H_d and D_w. The preprocessing function converts the original sampling interval to a target forecast interval. This work aims at a varying forecast interval (e.g., 15 min or 30 min), however, we validated the model with an hourly interval which, is a much lower frequency than the original 1 min sampling interval. Hence, the original dataset is averaged over 1 h to give a new dataset with hourly interval. An essential part of the data preprocessing function is data transformation which outputs the predictor's input variables and their corresponding target variables. The transformation is based on the intended forecast horizon. To be more precise, a transformation required for a 24 h horizon is different from that required for a 6 h horizon. The transformation process is further described in Section 4.1 where the forecast model is presented in detail.

2.3. Supervised machine learning methods

A supervised ML method learns a mapping from input **x** to output y, given a labeled set of input-output pairs $\mathcal{D} = \{(\bar{x}_i, y_i)\}_{i=1}^{N}$, where $\mathcal{D}$ is the training set, and N is the cardinality of the training set [22]. Each example input $\bar{x}_i$ is a D-dimensional vector of values representing the parameters of an instance (e.g., T_{out}, H_d, D_w, M_y, T_{ps}, ΔT and m). The output y_i part of the training set represents the *class* or *label* of its respective parameters (e.g. P) [22]. When the output variable y_i is categorical or nominal, these types of ML tasks are predominantly referred to as *classification* or *pattern recognition* [22]. However, when the output variable y_i, is a continuous data such as P, the task is referred to as a *regression* task [22]. Neural network based algorithms, MLR, SVR as well as regression trees (CART) are examples of commonly used regression algorithms that have been applied in prior work [4,5,7,8]. The following subsections briefly outline the technical details for each method.

2.3.1. Multiple linear regression

MLR [6] is a learning technique based on fitting a linear function with multiple independent variables. Eq. (4) shows the general form of MLR in relation to related variables:

$$P = \alpha + \beta_1 T_{out} + \beta_2 \Delta T + \cdots + \beta_n H_d, \tag{4}$$

where P is the response variable. T_{out}, ΔT, …, H_d are examples of input variables, and β represents the functional weights. α is a constant offset factor used to partially reduce the effect of modeling errors. The model creates a linear relationship in the form of a straight line, which best approximates all the individual variables. The relationship describes how the mean response variable, i.e. P changes with the explanatory variables e.g., T_{out}, ΔT, H_d.

2.3.2. Feed forward neural network

In previous studies, neural network based methods have been extensively used for forecasting energy consumption. FFNN is a general-purpose neural network for approximation of function f, which maps a set of input parameters to their respective output without the assumption about the relationships between the pair [6]. FFNN requires a definition of its model structure by stating the number of hidden layers, hidden units within the network and other related parameters. Deciding the ideal size of hidden layer is highly essential because an underestimation may lead to poor approximation and issues with generalization. On the other hand, overestimation may result in overfitting and makes the search for global minimal highly difficult. Eq. (5) shows the mathematical representation of a FFNN with a single hidden layer for its function approximation [6].

$$f(x) = \sum_{j=1}^{N} w_j \Psi_j \left[\sum_{i=1}^{M} w_{ij} x_i + w_{io} \right] + w_{jo} \tag{5}$$

where N denotes the total number of hidden units, M denotes the total number of inputs, and ψ represents the transfer function for each hidden unit. A FFNN training algorithm uses a gradient-based approach to update its weight by minimizing a specific error function. In this study, we used the mean square error (MSE) function.

2.3.3. Support vector regression

SVR is a method of support vector machines (SVMs) specifically for regressions. SVMs are based on the principle of structural risk minimization [23]. SVM constructs one or more hyperplanes in a high dimensional space. The objective of SVR minimizing the probability that the model generated from input data set will make an error on an unseen data instance. The objective is achieved by finding a solution which, best generalizes the training examples. The best solution is obtained by minimizing the following convex criterion function [6]:

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{l} \xi_i + \xi_i^* \tag{6}$$

with the following constraints:

$$y_i - w^T \phi(\bar{x}_i) - b \leq \epsilon + \xi_i \tag{7}$$

$$w^T \phi(\bar{x}_i) + b - y_i \leq \epsilon + \xi_i^* \tag{8}$$

where ϵ denotes the desired error range for all points. The variables ξ_i and ξ_i^* are the slack variables which guarantee that a solution exists for all ϵ. C is the penalty term used to balance between data fitting and smoothness. ϕ represents a kernel function for mapping the input space to a higher dimensional feature space.

2.3.4. Classification and regression tree

CART [24] is a general algorithm for generating statistical tree models. CART builds a binary tree for both classification and regression tasks, used for categorical and continuous target variables respectively. It adopts a greedy (i.e. nonbacktracking) approach that constructs trees in a top-down recursive divide-and-conquer manner. The training set is recursively partitioned into smaller subsets as the tree is being built. CART uses the minimization of prediction error as the split criterion. In this study, we used MSE as the split criterion.

3. ML application in DHS thermal forecast

This section describes in detail the process of data collection and analysis. The later part of this section provides details on the thermal load forecast model. Finally, the implementation details including modeling tools and chosen model parameters are presented.

3.1. Real-world dataset: data collection and pre-processing

This work is based on data collected unobtrusively from 10 district heating substations served by Skellefteå Kraft AB. Skellefteå Kraft AB's CHP plant supplies heat energy to approximately 5000 substations. In this paper, we considered a total of 10 substations from two building categories namely, multi-family apartment and commercial office buildings. Five of the substations belong to the multi-family apartment buildings and the other five substations belong to the commercial office building. These substations are equipped with data acquisition sensors, which monitor and facilitate the collection of relevant information associated with thermal load for each building. For reference purposes, Table 3 shows the substation reference ID (which is used henceforth in this paper), the number of apartments, per annual energy consumption and the thermal load statistics for each building substation.

Table 3
Statistics of energy consumption at the residential and commercial buildings under study.

Substation ref. ID	Per annual consumption (MWh)	Mean (kW)	Min (kW)	Max (kW)	No. of apartments
Res A	365.35	47.92	21.63	91.43	60
Res B	1560.0	210.95	119.68	344.67	120
Res C	2400.0	313.58	184.86	480.52	185
Res D	77.91	9.77	2.62	25.78	12
Res E	1845.24	221.62	131.63	370.07	174
Com A	1053.47	138.63	14.30	375.74	–
Com B	1561.99	198.04	15.96	473.14	–
Com C	694.57	95.69	36.02	228.73	–
Com D	203.86	32.04	5.11	97.23	–
Com E	493.54	64.60	19.47	182.46	–

The thermal load values and variables that relate to the internal state of a substation are measured on the primary side of the district heating network at each substation. These are: P, m, T_{ps}, T_{pr} and T_{out}. Fig. 3 shows the hourly averaged P values from the substations between the period of February 18, 2014 and April 6, 2014.

3.2. Analysis of influencing factors

One of the aim of this paper is to employ data analysis to provide insight into variations in thermal load at each substation. The understanding of relationships among variables with respect to the target variable, P, could assist in eliminating redundant ones and hence reduce the dimension of a model's input variables.

3.2.1. Variability of heat load in buildings

In this paper, the foremost analysis attempts to establish the daily load profile at each substation over 24 h. This is vital as it reveals the daily variations in thermal energy consumption at each building. This is accomplished by obtaining the mean P values of each D_w grouped by H_d for each substation. The obtained results show the differences among the substations, more notably, between the load pattern for the commercial buildings and the load pattern for the residential buildings.

Fig. 4 shows the obtained plots for all residential and commercial buildings. The figures show the mean thermal load grouped by H_d, and their 90% confidence intervals for each D_w. The figures display the differences between the consumption patterns of the two building categories. All buildings have a major peak in the early hours of the day, i.e., 6:00 and 11:00. Furthermore, the residential buildings also have another conspicuous consumption peak towards the evening hours, i.e., 18:00 to 22:00. In contrast a second peak is missing or very subtle in commercial buildings. The early peak in residential buildings is characterized by the early morning need for domestic water-heating. In consequence, the DHS plant supplies more energy for the peak period. In other words, the T_{ps} and m are usually increased at the production end to serve the heating network. Owing to the increase in thermal energy within the DHS grid, the commercial buildings as a result have an increase in delivered energy, and this serves a purpose of pre-heating for commercial buildings. In addition, there is a prominent drop in thermal load towards the business closing hours in commercial buildings. This is partly due to the average increase in T_{out} during the afternoon, which causes lower heat demand and partly due to the approaching closing office hours. Furthermore, the load profile of all buildings on the weekends show a prominent difference between the commercial buildings and the residential buildings. In all commercial buildings, Saturdays and Sundays have the least average load values for all hours of the day, while in residential buildings, the weekends generally have a shift in the first peak of the day by few hours.

3.2.2. Pearson product-moment correlation analysis

This paper employs a classical Pearson product-moment correlation coefficient analysis. A correlation analysis is used to obtain metric for linear relationship between two variables X and Y.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{9}$$

where r is the relationship between variables x_i and y_i. Here, we are interested in the correlation between thermal load, P and each of T_{ps}, T_{pr}, ΔT, m and T_{out}. We use a similar approach described in [13] to describe their relationships. Table 4 shows the resulting analysis for both residential and commercial buildings. The cell color highlighting in the table reflects the significance level of each parameter for its respective substation. The table distinctly reveals the domi-

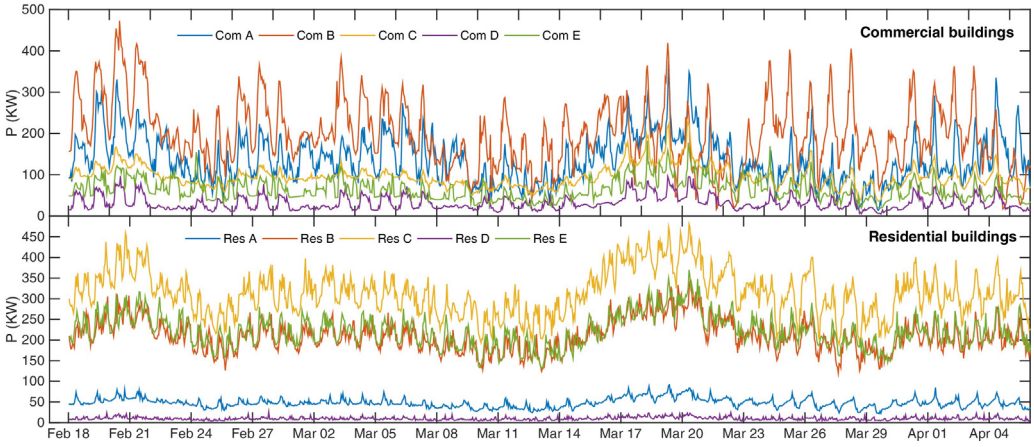

Fig. 3. Two plots showing the hourly thermal energy consumption for each building types.

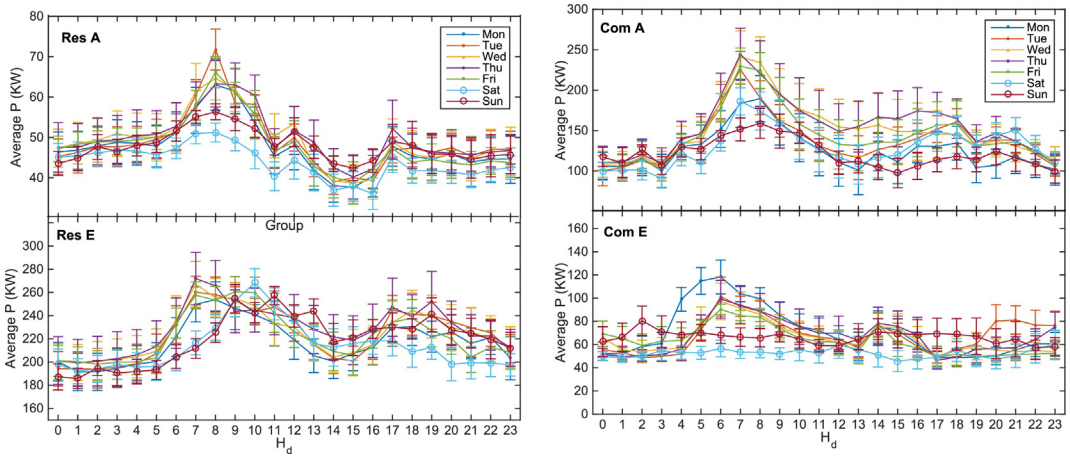

Fig. 4. Thermal energy consumption analysis for each building: the mean and confidence interval of P grouped over H_d for each D_w.

nant influencing variables at each substation and it shows that the significance level for the variables differ from one substation to the other.

3.2.3. Detecting energy consumption differences using PLS

The analysis presented in this section applies a commonly used multivariate analysis method, namely, partial least square (PLS). PLS is used to identify the degree of stochastic nature in the data obtained from each building. PLS method has a number of advantages. For this study, in order to gain insight to the level of randomness per building, the percentage variance explained in a response variable, P, as a function of the number of PLS components is used. A row of observed variables at time, t is matched with its respective response variable P at time, t. The percentage of variance explained by the observed variables per substation is obtained. We attempt to classify the PLS responses into two categories based on the % value of variance explanation and the number of PLS components. We chose 90% variance explanation in P as a boundary of interest and four components as a boundary region between two groups - few (1–4) and many (4–10) components. As a consequence, buildings that requires more than four PLS components to obtain 90% variance explanation in P are considered to be more stochastic in nature.

A simple PLS analysis carried out on the data from each substation produced the result shown in Fig. 5, which shows the number of PLS components as a function of percentage variance explained in P. It is observed that the percentage variance explained in P varies from one building to the other. A total of 7 substations have above 90% variance interpreted using only four components. These buildings are observed to have consistent energy consumption pattern which can be described using fewer variables. As indicated in the plot, *Com Bldg D*, *Com Bldg E* and *Res Bldg D* vary from the others. These buildings require more number of components for establishing the underlying relationship of the energy consumption patterns. Also, the plot implies that these buildings have higher occurrence of randomness in their energy consumption patterns.

4. Results

In this section the results obtained from thermal energy forecasting for each building are presented and discussed. First, we present the results relating to the ML methods employed and the variable inputs used, followed by the result obtained for a varying forecast horizon.

4.1. Model description: forecasting thermal load

For the forecast model presented in this paper, given the current time t, we create a model capable of estimating the value of variable P at time $t+h$ (where h is the forecast horizon). Fig. 6b shows the processes involved in our modeling. The model's input variables T_{out}, H_d, D_w (external factors) are considered as essential model inputs. The DHS operational parameters are also used as the internal influencing factors in the load forecast. This work also includes variables derived from T_{ps}, T_{pr} and m (DHS operational parameters) as input variables for the forecast model. The derived variables as used are ΔT, T_{ps}, T_{pr} and m.

To account for the temporal relationship between the variables, especially the response variable P, the model includes previously observed values of P variables as model's input. In order to accomplish a time series model capable of capturing the temporal properties of the variables, a data transformation process lags the data in relation to the target horizon. Lagging implies that new variables are created from the old ones by forward shifting data (Fig. 6a) [25]. The dataset of these variables is transformed into instances of X predictor variables and their corresponding Y target variable for a h-horizon forecast. The transformation is done such that a set of identified influencing factors at time t is paired with the response variable P at time $t+h$. This follows from the need to model how

Table 4
Commercial buildings: correlation analysis between influencing factors and heat load. The four most influential variables are shown.

	Res A	Res B	Res C	Res D	Res E	Com A	Com B	Com C	Com D	Com E
T_{ps}	0.76	0.70	0.74	0.67	0.69	0.59	0.67	0.83	0.59	0.62
T_{pr}	-0.11	0.58	-0.48	-0.51	0.36	0.70	0.76	0.84	0.28	-0.21
ΔT	0.85	0.69	0.74	0.72	0.44	-0.08	-0.05	0.29	0.41	0.78
m	0.92	0.32	-0.10	0.76	0.76	0.98	0.99	0.94	0.98	0.96
T_{out}	-0.82	-0.69	-0.76	-0.49	-0.69	-0.73	-0.83	-0.95	-0.59	-0.57

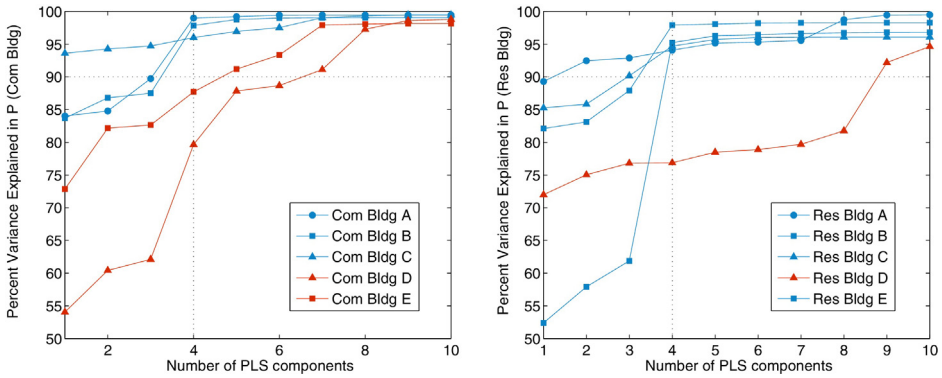

Fig. 5. Plot showing percentage variance explained in P using observed parameters as a function of the number of PLS components.

the current process situation at time t is influenced by the process variables at earlier time. A variable is lagged by the horizon h. In Fig. 6a, variable $T_{out}(t)$ is shifted by h steps down, variable $T_{out}(t-h)$ is shifted by $h \times 2$ steps down. A variable $T_{out}(t-2h)$ will be shifted down by $h \times 3$ steps. The variable $T_{out}(t+h)$ is the weather temperature forecast for the same time level with the response variable, Y ($P(t+h)$). Furthermore, the time variables input, D_w and H_d are related to the time ($t+h$). The resulting dataset X and Y from the transformation is then fed into the four supervised machine learning algorithms considered in this paper to output the forecasting models.

4.2. Implementation: modeling tool and parameters

This work was performed in Matlab R2013a [26] using its ANN Toolbox for the FFNN, while its inbuilt functions for *LinearModel* class and the Regression Tree were used for MLR and regression tree respectively. LIBSVM [27] was used for the SVM algorithm. MLR and regression tree both requires little or no parameter tweaking. For the FFNN algorithm, this study considered a FFNN with N hidden layers. The best value of N is chosen for each building dataset. Building Res A, Res C and Com C have the value of $N = 1$, which is a similar structure used in a previous study [6]. Res B, Res D, Res E and Com A, B work best with a value of $N = 2$, while Com D and Com E performed best with $N = 4$. For the FFNN activation function, this paper used *tansig*(x) owing to similar approach being used in related work and has shown decent performance as a transfer function [6]. For the SVM algorithm, the performance largely depends on the choice of the regularization parameter C and the kernel function ϕ parameters. Several related research chose Gaussian function which is included in radial basis function (RBF) as the kernel model for SVM regression [28]. RBF has a number of advantages [28], for instance, its ability to handle non-linear relationship between input and the response variables. In addition, the BRF kernel has less numerical difficulties in contrast to polynomial kernels [28]. Similarly, for the SVM method, this paper used Gaussian RBF as the nonlinear kernel function ϕ, which has γ as its kernel parameter. The parameter C is a constant for penalizing the training errors. A high value of C indicates a very narrow Gaussian, which tries to fit all the training data (i.e., memorize almost all trying data). This paper used optimal pair of C and γ, which is selected through the use of search grid. A series of search grid was carried out with varying steps (i.e. from larger steps to smaller steps) to determine the final suitable search range. To avoid overfitting, the value of C is set lower than values used in some related work. The upper bound in the parameter search for C is set to 2^2, and this is returned as the best value of the search for all datasets. The value of γ is set to the optimal value of 0.005.

4.3. Model sample selection

For a dataset $\mathcal{D}$ of a period range t_i to t_j, we define a split time t_s which represent a time within the dataset $\mathcal{D}$. A model is generated using data from t_i to t_s, the model is then used to forecast hourly P values from t_s to t_j. This approach relates to the typical use of a forecast model. In this paper, the length of t_s to t_j is taken as 1 week. The dataset for each building is separated into a training and a test sample. The data obtained from February 18 to March 29 (39 days) is used as training sample, while the last week of the data set, namely March 30 to April 6 (7 days), is used as the testing sample.

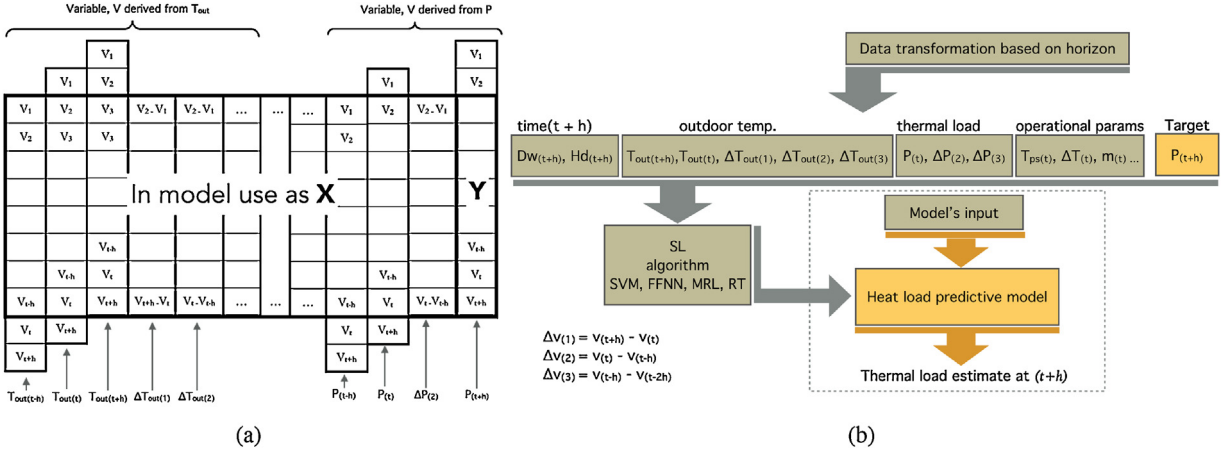

Fig. 6. (a) Data transformation with variable lagging. (b) Modeling process.

The effect of variables associated with thermal load in a DHS is investigated by creating four sets of identified input variables, the description is listed as follows (see Table 5).

- Parameter set 1: A variable set of all identified input variables. These are variables that have been included as input set based on intuition from knowledge acquired in the DHS domain.
- Parameter set 2: A set of all input variables excluding the operational parameters of a substation.
- Parameter set 3: A set of all input variables excluding the previous thermal load variables (e.g., P at time t, $t-h$).
- Parameter set 4: A set of all input variables excluding both the operational parameters and the previous thermal load values.

4.4. Forecast performance and analysis

In this section the forecasting performance is analyzed and discussed, especially the differences among the ML methods, the substation buildings and the input parameter sets. The performance evaluation indices to SVM, FFNN, MLR and regression tree adopted in this paper is the normalized root mean square error (NRMSE), which is defined as

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{10}$$

$$NRMSE = \frac{RMSE}{\bar{y}_t} \tag{11}$$

where y_t is the actual value of P at a building, $\hat{y}_t$ is the forecast value of y_t and $\bar{y}_t$ is the average thermal load for the building.

Based on the input parameter sets, a 24 h horizon forecast models are established on the basis of each building. Their NRMSE performance error obtained in the testing phase are presented as follows.

Fig. 7 shows four plots for the performance error obtained for each ML approach. Each plot shows the error rate observed at the 10 buildings. Furthermore, the specific performance for each input parameter set is also displayed. It can be seen that SVM, MLR and FFNN have lower performance error with very small margin, while regression tree approach produced models with higher performance error in all buildings. The values obtained for the four input parameter sets have similar results except in a few cases where for example the "*parameter set 4*" generated slightly higher performance errors.

Next, the trend plots for evaluating all models are presented. A trend plot evaluation provides a visual means of model evaluation for a time-series data. Fig. 8 shows the trend plot results for 1 week of 24-h forecast at the residential and commercial buildings respectively. The plots also include the last 48 h from the training dataset (data points to the right of vertical red line). The plot of observed P values and the forecast values of P using SVM, FFNN, MLR and regression tree are all shown. Also, the plot shows the trend fittings for the last two days of the training phase. It can be seen that SVM generated forecast values that closely match with the actual P values in all buildings while regression tree models produce much lesser fitted forecast values. FFNN and MLR also produce closely fitted trend lines.

In addition to the above result, we investigate the performance of the model for varying horizons of every hour from 1 h up to 48 h. Based on the varying horizon values, SVM models are generated and the performance errors are identified. The results obtained for the models at each building are displayed in Fig. 9. The figure is comparable to result shown in Bacher et al. [18], and it shows the error rate in relation to the forecast horizon for each building.

Table 5
Four input parameter sets for evaluating their influencing significance in a thermal load forecast.

	Time factor		Outdoor temperature						Previous thermal load				DHS operational variables					
	H_d	D_w	$T_{out(t+h)}$ forecast	$T_{out(t)}$	$T_{out(t-h)}$	$\Delta T_{out(1)}$	$\Delta T_{out(2)}$	$P_{(t)}$	$P_{(t-h)}$	$P_{(t-2h)}$	$\Delta P_{(2)}$	$\Delta P_{(3)}$	$T_{ps(t)}$	$\Delta T(t)$	$m_{(t)}$	$\Delta T_{ps(t-h)}$	$\Delta T(t-h)$	$m_{(t-h)}$
Parameter set 1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Parameter set 2	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	✓	✓	✓	✓	✓	✓
Parameter set 3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	×
Parameter set 4	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×	×

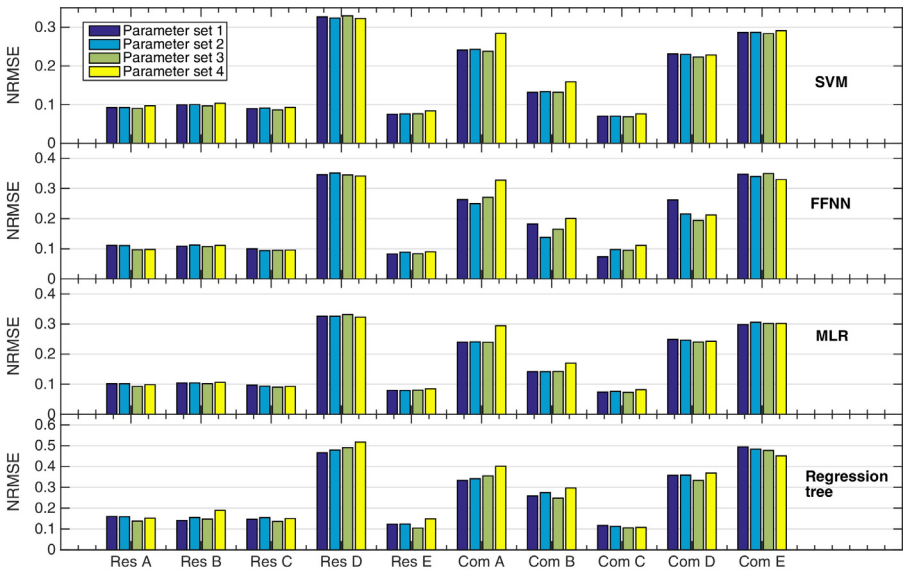

Fig. 7. Plots showing the performance error obtained for each ML approach applied at all the buildings based on a 24-hour horizon forecast model.

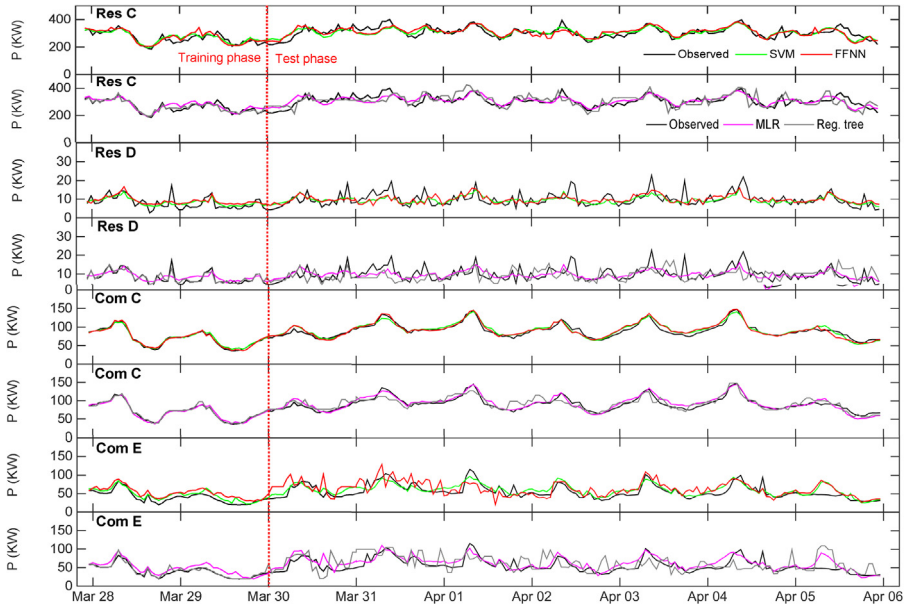

Fig. 8. Trend plots of a 24-h forecast values at four buildings using SVM, FFNN, MLR and regression tree based on parameter set 1. The plotted graphs are from buildings with the best and worst forecast performance for both residential and commercial buildings. The plots on the left side of the vertical red line corresponds to the trailing part of the training phase (i.e. the last 2 days in the training data), while the data on the right of the red vertical line is the data for all the testing period.

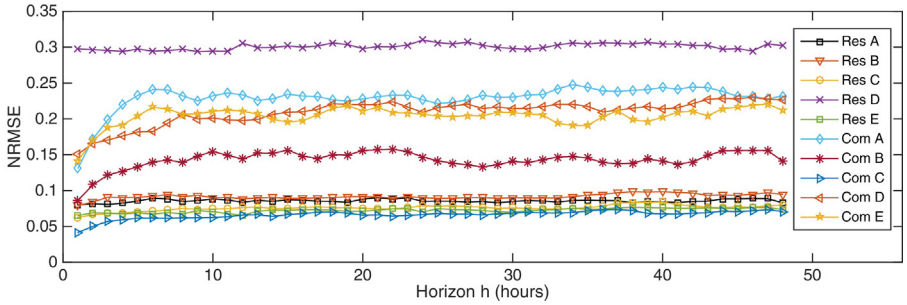

Fig. 9. The NRMSE as a function of the horizon *h* for SVM model for each building.

5. Discussion

In this section, a discussion based on our approach and the obtained results is provided. Furthermore, discussion about the results and comparison with results from a closely related work is presented.

First, the different results obtained from each building indicate that they are fundamentally different and hence exhibit varying energy response patterns. This is also true even for buildings within the same category. This supports the claims in existing literatures such as Edwards and Parker [6]. The SVM, FFNN and MLR methods considered in this work produced models with similar performance error. Based on the model performance evaluation, the result presented in this paper is comparable to result in related work on the forecast of thermal energy consumption in buildings. MLR prediction technique is suitable for input variables and response variables with strong linear relationship. In this case, since there exists a strong linear relationship between the model's input variables and the response variable, MLR performance is comparable to SVM and FFNN. A SVM model without computational overhead is used, this is achieved by keeping parameters C and γ relatively lower. The regression tree models produced relatively higher performance errors. This shows that regression tree is not suitable for this problem. A regression tree would be more suitable for discrete data types with smaller amount of dataset. In essence, the results show that SVM, FFNN and MLR are more suitable for a DHS thermal load forecast task. This reinforces the reason behind the application of SVM and neural network based methods in related papers in building energy forecast.

Second, our work shows comparable results with existing literature such as Bacher et al. [18], where the authors presented a forecasting method for space heating in 16 single-family houses. In their results, they show the performance error for each of the 16 houses based on forecast horizons ranging from 1 to 42 h. Fifteen houses produced NRMSE approximately between 0.060 and 0.32, with one house having a higher NRMSE value of about 0.55 for horizons above 5 h. In our work, lower performance errors were achieved as compared to Bacher et al. [18]. Half of the buildings produced NRMSE below 0.10. Four buildings have NRMSE approximately between 0.1 and 0.25, while the highest performance error obtained is approximately 0.30.

Third, effect of DHS internal parameters. The forecast of energy consumption mostly requires a weather forecast information as its input variable. In few cases where additional useful variables are accessible, such as the internal system information of a DHS, it is worthwhile to identify the impact of such variables in forecasting energy consumption. A related work [13] has made use of internal variables using a neural network based approach for forecasting total heat energy consumption in a DHS. Nevertheless, the results presented in this paper have shown that the importance of these additional variables varies from one building to the other. Furthermore, the impacts are very minimal in most cases. However, the use of these internal variables can be of significant impact when applying a localized energy saving strategies e.g., controlling the charge and discharge in heat accumulators or in hybrid energy systems. Energy saving strategies are essential for minimizing energy cost and there are various methods proposed in other work, such as the Markov decision process based strategy proposed in Nikovski et al. [20].

Referring back to the third and fourth points, the essence of an accurate weather forecast information and time information is shown. In this paper, we considered outdoor temperature while it is possible to take into account other weather related aspects too. Other weather information, for instance, solar irradiance, humidity and windspeed are essential factors [9]. In accordance, since outdoor temperature shows high impact on thermal consumption, it follows that other weather variables can improve the model's performance in buildings. Occupancy behaviour in buildings is another factor which can impact energy consumption, however, this work is restricted to existing parameters within the target DHS under consideration. This restriction however does not limit the application of the proposed model, i.e., the model can be easily extended for use with additional parameters. Grey-box models are ideal in cases where important parameters are missing and forecast accuracy is of great interest. This is in consequence to grey-box's ability to leverage on the physical model of the building in addition to statistical model obtained from data. In this paper however, the interest is on data-driven approach which utilizes data available within an existing DHS.

Furthermore, a model created from a set of data fits the underlying patterns in the data. In this paper, although the results have been generated from only one test site with a restricted time range, the model approach used can still be used in different scenarios, such as a different city and season. It is expected that a model generated from data collected from a different scenario will reflect the scenario specifically. A new city or different time range will require training a model using data corresponding to the city or time accordingly.

Finally, the forecast produced results with very low error rate for substations with lower variation in energy consumption. However, substations with higher stochastic nature are more difficult to forecast, producing higher error rates in the model. This is partly caused by the nature of the problem and the high variation in weather parameters, especially in the Spring season, in North Sweden which is the test site.

6. Conclusion

An ML approach for forecasting the thermal load in district heating substations is presented. We evaluated and compared different ML methods such as SVM, FFNN, MLR and regression tree. The forecast models are produced and evaluated using data observed from 10 district heating substations serving five multi-family apartment and five commercial buildings. We consider four key parameters as input to our model. These include weather (outdoor temperature), temporal aspects (day of the week, hour of the day), historical values of thermal load and the physical parameters of a substation (supply temperature, difference between supply and return temperature and flow rate). The effect of the substation operational parameters as additional inputs is also shown. Lastly, a performance comparison of the SVM models is presented based on forecast horizons of up to 48-h.

We achieve a high forecast performance for seven buildings, which have regular patterns for energy consumption while three buildings have relatively higher performance error which are attributed to unpredictable high frequency variations in the thermal load signal. The results show that using the parameters associated with a district heating substation adds subtle impact to the forecasting error. SVM gave the best prediction performance, with FFNN and MLR having very similar error rate as SVM. Regression tree models have higher performance error rate and hence are less suitable. Finally, it is established that buildings exhibit different energy consumption patterns, which differentiate their response to forecasting models. In future work, there will be a focus on exploiting the results of this paper for building localized energy saving strategies. These strategies shall be used for peak shaving and maintaining a steady return temperature back to the CHP while achieving satisfactory thermal energy in buildings.

Acknowledgment

The authors would like to thank Skelleftea Kraft for providing the data used for this work. And also Chih-Jen Lin for LIBSVM and

for his generous sharing of his expertise and code. The presented research work is funded by the European Commission within the Seventh Framework Programme FP7 Project, OPtimising Hybrid Energy grids for smart citieS (OrPHEuS), grant agreement 608930.

References

[1] F. Marechal, D. Favrat, E. Jochem, Energy in the perspective of the sustainable development: the 2000W society challenge, Resour. Conserv. Recycl. 44 (3) (2005) 245–262, http://dx.doi.org/10.1016/j.resconrec.2005.01.008.
[2] A. Molderink, V. Bakker, M.G.C. Bosman, J.L. Hurink, G.J.M. Smit, Management and control of domestic smart grid technology, IEEE Trans. Smart Grid 1 (2) (2010) 109–119, http://dx.doi.org/10.1109/TSG.2010.2055904.
[3] N. Fumo, A review on the basics of building energy estimation, Renew. Sustain. Energy Rev. 31 (2014) 53–60, http://dx.doi.org/10.1016/j.rser.2013.11.040.
[4] L. Wu, G. Kaiser, D. Solomon, R. Winter, A. Boulanger, R. Anderson, Improving efficiency and reliability of building systems using machine learning and automated online evaluation, in: 2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2012, pp. 1–6.
[5] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy Build. 57 (2013) 302–312, http://dx.doi.org/10.1016/j.enbuild.2012.11.010.
[6] R. Edwards, J. New, L. Parker, Predicting future hourly residential electrical consumption: a machine learning case study, Energy Build. 49 (2012) 591–603, http://dx.doi.org/10.1016/j.enbuild.2012.03.010.
[7] M. Sakawa, S. Ushiro, Cooling load prediction in a district heating and cooling system through simplified robust filter and multi-layered neural network, Syst. Man Cybern. (1999) 995–1000 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=823364.
[8] K. Kato, M. Sakawa, S. Ushiro, Heat load prediction through recurrent neural network in district heating and cooling systems, in: 2008 IEEE Int. Conf. Syst. Man Cybern., 2008, pp. 1401–1406, http://dx.doi.org/10.1109/ICSMC.2008.4811482.
[9] A. Kusiak, M. Li, Z. Zhang, A data-driven approach for steam load prediction in buildings, Appl. Energy 87 (3) (2010) 925–933, http://dx.doi.org/10.1016/j.apenergy.2009.09.004.
[10] Q. Zhou, S. Wang, X. Xu, F. Xiao, A grey-box model of next-day building thermal load prediction for energy-efficient control, Int. J. Energy Res. 32 (15) (2008) 1418–1431, http://dx.doi.org/10.1002/er.1458.
[11] A. Afram, F. Janabi-Sharifi, Black-box modeling of residential {HVAC} system and comparison of gray-box and black-box modeling methods, Energy Build. 94 (2015) 121–149, http://dx.doi.org/10.1016/j.enbuild.2015.02.045.
[12] S. Grosswindhager, A. Voigt, M. Kozek, Online short-term forecast of system heat load in district heating networks, in: proceedings of the 31st international symposium on forecasting, Prag, Czech Republic, 2011, pp. 1–8, http://www.forecasters.org/submissions/GROSSWINDHAGERSTEFANISF2011.pdf.
[13] M. Wang, Q. Tian, Application of wavelet neural network on thermal load forecasting, Int. J. Wirel. Mob. Comput. 6 (6) (2013) 608, http://dx.doi.org/10.1504/IJWMC.2013.057579.
[14] H. Gadd, S. Werner, Daily heat load variations in Swedish district heating systems, Appl. Energy 106 (2013) 47–55, http://dx.doi.org/10.1016/j.apenergy.2013.01.030.
[15] H.A. Nielsen, H. Madsen, Modelling the heat consumption in district heating systems using a grey-box approach, Energy Build. 38 (1) (2006) 63–71, http://dx.doi.org/10.1016/j.enbuild.2005.05.002.
[16] H.-x. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (6) (2012) 3586–3592, http://dx.doi.org/10.1016/j.rser.2012.02.049.
[17] F. Sandin, J. Gustafsson, J. Delsing, Fault Detection with Hourly District Energy Data: Probabilistic Methods and Heuristics for Automated Detection and Ranking of Anomalies, Svensk Fjärrvärme, 2013.
[18] P. Bacher, H. Madsen, H.A. Nielsen, B. Perers, Short-term heat load forecasting for single family houses, Energy Build. 65 (2013) 101–112, http://dx.doi.org/10.1016/j.enbuild.2013.04.022.
[19] E. Serban, D. Popescu, Prediction of domestic warm-water consumption, WSEAS Trans. Comput. 7 (12) (2008) 2032–2041 http://dl.acm.org/citation.cfm?id=1486811.1486827.
[20] D. Nikovski, J. Xu, M. Nonaka, A method for computing optimal set-point schedules for HVAC systems, in: REHVA World Congress CLIMA'13, 2013.
[21] S.O. Idowu, C. Åhlund, S. Olov, Machine learning in district heating system energy optimization, in: 2014 IEEE Int. Conf. Pervasive Comput. Commun. Work Prog., 2014, p. 4.
[22] S. Idowu, C. Ahlund, O. Schelen, R. Brannstrom, Machine Learning in Pervasive Computing, Tech. Rep. September, 2013 http://pure.ltu.se/portal/files/43907990/Report_ML_in_Pervasive_Computing.pdf.
[23] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10 (5) (1999) 988–999, http://dx.doi.org/10.1109/72.788640.
[24] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, Chapman & Hall, New York, 1984.
[25] L. Eriksson, Multi- and Megavariate Data Analysis, no. p. 1 in Training in multivariate technology, Umetrics AB, 2006 http://books.google.se/books?id=B-1NNMLLoo8C.
[26] MATLAB, version 8.1.0 (R2013a), The MathWorks Inc., Natick, MA, 2013.
[27] C.-C. Chang, C.-J. Lin, {LIBSVM}: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 27:1–27:27.
[28] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy Build. 37 (5) (2005) 545–553, http://dx.doi.org/10.1016/j.enbuild.2004.09.009.