

Research Article

Data-Driven Machine-Learning Model in District Heating System for Heat Load Prediction: A Comparison Study

Fisnik Dalipi,¹ Sule Yildirim Yayilgan,¹ and Alemayehu Gebremedhin²

¹Faculty of Computer Science and Media Technology, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

²Faculty of Technology and Management, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Correspondence should be addressed to Fisnik Dalipi; fisnik.dalipi@ntnu.no

Received 21 February 2016; Revised 11 May 2016; Accepted 16 May 2016

Academic Editor: Shyi-Ming Chen

Copyright © 2016 Fisnik Dalipi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present our data-driven supervised machine-learning (ML) model to predict heat load for buildings in a district heating system (DHS). Even though ML has been used as an approach to heat load prediction in literature, it is hard to select an approach that will qualify as a solution for our case as existing solutions are quite problem specific. For that reason, we compared and evaluated three ML algorithms within a framework on operational data from a DH system in order to generate the required prediction model. The algorithms examined are Support Vector Regression (SVR), Partial Least Square (PLS), and random forest (RF). We use the data collected from buildings at several locations for a period of 29 weeks. Concerning the accuracy of predicting the heat load, we evaluate the performance of the proposed algorithms using mean absolute error (MAE), mean absolute percentage error (MAPE), and correlation coefficient. In order to determine which algorithm had the best accuracy, we conducted performance comparison among these ML algorithms. The comparison of the algorithms indicates that, for DH heat load prediction, SVR method presented in this paper is the most efficient one out of the three also compared to other methods found in the literature.

1. Introduction

As stated in the report of European Commission strategy for energy, the continuous growing of energy demand worldwide has made energy security a major concern for EU citizens. This demand is expected to increase by 27% by 2030, with important changes to energy supply and trade [1]. Being the largest energy and CO₂ emitter in the EU, the building sector is responsible for 40–50% of energy consumption in Europe and about 30–40% worldwide [2]. The North European countries have proved themselves as forerunners in the development and application of clean and sustainable energy solutions. Their excellent performance on adopting such solutions enables them to achieve ambitious national climate objectives and requirements and to serve as key players in the entire European energy system [3].

District heating (DH) system is an optimal way of supplying heat to various sectors of the society such as industrial, public, or private buildings. DH network offers functional, economic, and ecological advantages and is also instrumental in reducing the global and local CO₂ emissions. It offers an

enormous adaptability to combine different types of energy sources efficiently [4]. Considering the recent technological trends of progressing to smart energy infrastructures, the development of the fourth generation of district heating implies meeting the objective of more energy-efficient buildings. Moreover, this also envisions DH networks to be as an integrated part of the operation of smart energy systems, that is, integrated smart electricity, gas, and thermal grids [5]. The application of new and innovative technology in district heating is therefore considered essential to improve energy efficiency [6].

The deregulation of the electricity market and the increasing share of energy-efficient buildings have put district heating in a more vulnerable position with regard to challenges in terms of cost effectiveness, supply security, and energy sustainability within the local heat market. With this background, it is therefore important for district heating sector to maintain an efficient and competitive district heating system which is able to meet the various requirements which characterize the heat market. In a flexible district heating system with multiple energy sources and production technologies,

the need for accurate load forecasting has become more and more important. This is especially important in a district heating system with simultaneous production of heat, steam, and electricity.

In this paper, with the application of three different ML algorithms to predict heat consumption, we investigate the performance of *Support Vector Regression* (SVR), *Partial Least Squares* (PLS), and *random forest* (RF) approach to develop heat load forecasting models by making a comparative study. Our focus is on low error, high accuracy, and validating our approach with real data. We also compare the error analysis of each algorithm with existing techniques (models) and also find the most efficient one out of the three.

The rest of the paper is organized as follows: Section 2 outlines the related work, where we provide an overview of many approaches to load prediction that are found in the literature. In Section 3, we provide some background information about DH concepts. This is followed by a presentation of the system framework and related prediction models, given in Section 4. Further, in Section 5, we present and discuss the evaluation and results. Finally, Section 6 concludes the paper.

2. Related Work

The state of the art in the area of energy (heating, cooling, and electric energy) demand estimation in buildings is classified as *forward (classical)* and *data-driven (inverse)* approaches [7]. While the forward modelling approach generally uses equations with physical parameters that describe the building as input, the inverse modelling approach uses machine-learning techniques. Here, the model takes the monitored building energy consumption data as inputs, which are expressed in terms of one or more driving variables and a set of empirical parameters and are widely applied for various measurements and other aspects of building performance [8]. The main advantage of data-driven models is that they can also operate online, making the process very easily updatable based on new data. Considering the fact that ML models offer powerful tools for discovery of patterns from large volumes of data and their ability to capture nonlinear behavior of the heat demand, they represent a suitable technique to predict the energy demand at the consumer side.

Numerous ML models and methods have been applied for heat load prediction during the last decade. A good overview of some recent references is given by Mestekemper [6, 9]. The former also built his own prediction models using dynamic factor models. A simple model proposed by Dotzauer [10] uses the ambient temperature and a weekly pattern for prediction of the heat demand in DH. The author makes the social component equal to a constant value for all days of the week. There is another interesting model, which address the utilization of a grey box that combines physical knowledge with mathematical modelling [11]. Some approaches to predict the heat load discussed in the literature include artificial neural networks (ANN) [12–15]. In [12], a backpropagation three-layered ANN is used for the prediction of the heat demand of different building samples. The inputs of the network for training and testing are building

transparency ratio (%), orientation angles (degrees), and insulation thickness (cm) and the output is building heating energy needs (Wh). When ANN's outputs of this study are compared with numerical results, average 94.8–98.5% accuracy is achieved. The authors have shown that ANN is a powerful tool for prediction of building energy needs. In [13], the authors discuss the way self-organizing maps (SOMs) and multilayer perceptrons (MLP) can be used to develop a two-stage algorithm for autonomous construction of prediction models. The problem of heat demand prediction in a district heating company is used as a case study where SOM is used as a means of grouping similar customer profiles in the first stage and MLP is used for predicting heat demand in the second stage. However, the authors do not provide any information related to the error rates obtained during the predictions.

In [14], recurrent neural networks (RNNs) are used for heat load prediction in district heating and cooling systems. The authors compare their prediction results from RNN with the prediction results obtained from a three-layered feed forward neural network (TLNN). The mean squared error between the TLNN and the stationary actual heat load is reported to be 21.05² whereas it is 11.82² between the RNN and the actual heat load data. In the nonstationary case, RNN still provides lower mean squared error. The use of RNNs rises the expectation to capture the trend of heat load since it uses heat load data for several days as the input.

In [15], time, historical consumption data, and ambient temperatures were used as input parameters to forecast heat consumption for one week in the future. The authors compared the performances of three black-box modelling techniques SVR, PLS, and ANN for the prediction of heat consumption in the Suseo DH network and analyzed the accuracy of each method by comparing forecasting errors. The authors report that in one-day-ahead overall average error of PLS is 3.87% while that of ANN and SVR is 6.54% and 4.95%, respectively. The maximum error of SVR is 9.82%, which is lower than that of PLS (16.47%) and ANN (13.20%). In terms of the overall error, the authors indicate that PLS exhibits better forecasting performance than ANN or SVR.

In [16], a multiple regression (MR) model is used for heat load forecasting. The reported MAE is 9.30. The model described in [17] uses an online machine-learning approach named Fast Incremental Model Trees with Drift Detection (FIMT-DD) for heat load prediction and hence allows the flexibility of updating the model when the distribution of target variable changes. The results of the study indicate that MAE and MAPE for FIMT-DD (using Bagging) have lower values in comparison to Adaptive Model Rules (AMRules) and Instance Based Learner on Streams (IBLStreams).

Authors in [18] compare the performance of four supervised ML algorithms (MLR, FFN, SVR, and Regression Tree (RT)) by studying the effect of internal and external factors. The external factors include outdoor temperature, solar radiation, wind speed, and wind direction. The internal factors are related to the district heating system and include supply and return water pressure, supply and return water temperature, the difference of supply and return temperature, and circular flow. Their study shows that SVR showed the best

TABLE 1: ML models for heat demand prediction in the literature.

Applied algorithms	[12]	[13]	[14]	[15]	[16]	[17]	[32]	[33]	[34]	[35]	[36]	Our work
ANN/FNN/SOM/RNN	+	+	+	+	-	-	+	-	-	+	-	-
MLR/MR/PLS	-	-	-	+	+	-	+	-	-	-	-	+
SVM/SVR	-	-	-	+	-	-	+	+	-	-	+	+
BN	-	-	-	-	-	-	-	-	+	-	-	-
DT/RF/RT	-	-	-	-	-	+	+	-	-	-	-	+
Ensembles	-	-	-	-	-	+	-	-	-	-	-	+

accuracy on heat load prediction for 1- to 24-hour horizons. However, the prediction accuracy decreases with the rise in horizon from 1 to 18 hours.

Wu et al. [19] discuss and implement SVR as a predictive model to the building's historical energy use. Their predictive model proved to approximate current energy use with some seasonal and customer-specific variations in the approximations. Another work [20] discusses the importance of prediction of load in a smart energy grid network. The authors propose a BN to predict the total consumer water heat consumption in households. Shamshirband et al. [21] construct an adaptive neurofuzzy inference system (ANFIS), which is a special case of the ANN family, to predict heat load for individual consumers in a DH system. Their result indicates that more improvements of the model are required for prediction horizons greater than 1 hour. Protić et al. [22] study the relevance of short-term heat load prediction for operation control in DH network. Here, authors apply SVR for heat load prediction for only one substation for time horizon of every 15 minutes. To improve the predictive model, authors also add a dummy variable to define the state of DH operation.

In literature, the research towards developing load forecasting models is also discussed from different perspectives and used in different energy related applications, such as head load in district heating, wind turbine reaction torque prediction [23], and wind power forecasting [24, 25].

In [23], SVR is employed for wind turbine torque prediction. The results show that an improvement in accuracy can be achieved and conclude that SVR can be considered as a suitable alternative for prediction. It can be also seen that the proposed SVR prediction models produce higher accuracy compared to ANN and ANFIS (adaptive neurofuzzy inference system). The work discussed in [24] considers the penetrations of renewable energies in electrical power systems by increasing the level of uncertainty. In such situations, traditional methods for forecasting of load demand cannot properly handle these uncertainties. Hence, they implement a neural network method for constructing prediction intervals by using a low upper bound estimation (LUBE) approach. The authors conduct a comparative analysis and show that this method can increase the prediction intervals quality for load and wind power generation predictions.

Bhaskar and Singh [25] perform a statistical based wind power prediction using numerical weather prediction (NWP). In order to validate the effectiveness of the proposed method, the authors compared it with benchmark models, such as persistence (PER) and new-reference (NR), and show

that the proposed model outperforms these benchmark models.

Additionally, due to innovations in the future sustainable and smart energy systems and recent technological trends with IoT (Internet of Things), many research works [5, 26] consider DH systems as being an integral part in Smart Grid, within the smart city concept. Moreover, such a DH system model will require high computation time and resources for knowledge representation, knowledge inference, and operational optimization problems. Thus, in response to this, researchers are continuously focusing on the development and use of fast and efficient algorithms for real-time processing of energy and behavior related data.

As a summary, previous research on heat load prediction points to various training algorithms: ANN including RNN, FFN (Feedforward Neural Network)/MLP, and SOM; MR including MLR and PLS; SVM including SVR; Bayesian networks (BN); decision trees (DT); ensemble methods [27]; FIMT-DD; AMRules; and IBLStreams.

In spite of the interest and the considerable efforts given by the research community so far, there is no consensus among researchers on neither selecting the most suitable training model for heat load prediction nor selecting an appropriate set of input parameters for training the model with [16] in order to achieve high level of prediction accuracy. This is due to the fact that superiority of one model over another in heat load prediction cannot be asserted in general because performance of each model rather depends on the structure of the prediction problem and the type of data available. The comparison in [15] pointed to the superior performance of SVR already; however, as our problem structure and inputs are different from theirs, we chose to do a comparison of several up-to-date models to find the most promising approach for our case. Table 1 lists models from the literature. The “plus” sign indicates that a particular algorithm has been applied, while “minus” means the opposite. Based on the table, we concluded that SVR, PLS, and RF provide us with a unique combination of models to compare with each other. Simplicity and efficiency of each model in our combination are preferred such that rapid and simple assessment of energy demand with high accuracy can be obtained.

3. District Heating Systems

District heating is a well-proven technology for heat supply from different energy sources through heat generation and distribution to heat consumers. DH systems are valuable

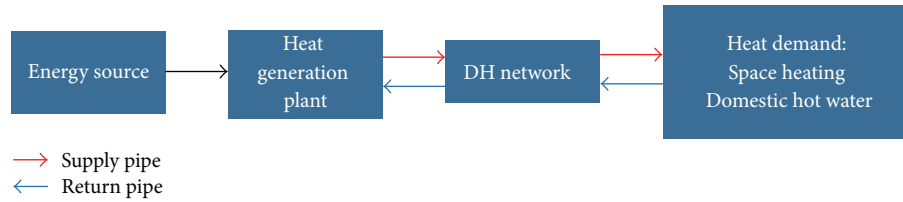


FIGURE 1: District heating block diagram.

infrastructure assets, which enable effective resource utilization by incorporating the use of various energy sources. One of the main advantages of DH system is that it facilitates the use of combined heat and power (CHP) generation and thereby makes the overall system efficient.

District heating can play a crucial role in reaching some of the energy and environmental objectives by reducing CO₂ emissions and improve overall energy efficiency. In a district heating system, heat is distributed through a network of hot-water pipes from heat-supplying plants to end users. The heat is mostly used for space heating and domestic hot water. A simplified schematic picture of DH system is shown in Figure 1.

The main components of district heating system are heat generation units, distribution network, and customer substations. The heat generation unit may use heat-only boilers or CHP plants or a combination of these two for heat generation. Various types of energy sources, like biomass, municipal solid waste, and industrial waste heat, can be used for heat production. The heat is then distributed to different customers through a network of pipeline. In the customer substations, the heat energy from the network is transferred to the end users internal heating system.

The heat-supplying units are designed to meet the heat demand. The heat output to the network depends on the mass flow of the hot water and the temperature difference between the supply and the return line. The supply temperature of the hot water is controlled directly from the plant's control room based on the outdoor temperature and it follows mostly a given operation temperature curve. The return temperature, on the other hand, depends mainly on the customer's heat usage and also other network specific constraints. The level of the supply temperature differs from country to country. For instance, in Sweden, the temperature level varies between 70 and 120°C depending on season and weather [28].

The heat load in district heating systems is the sum of all heat loads that are connected to the network and distribution and other losses in the network.

With increased concerns about the environment, climate change, and energy economy, DH is an obvious choice to be used. Nowadays, district heating systems are equipped with advanced and cutting-edge technology systems and sensors that monitor and control production units from a control room remotely. From a smart city perspective, one of the future challenges that now remains is to integrate district heating with the electricity sector as well as the transport sector. Heat load forecasting models with high accuracy are

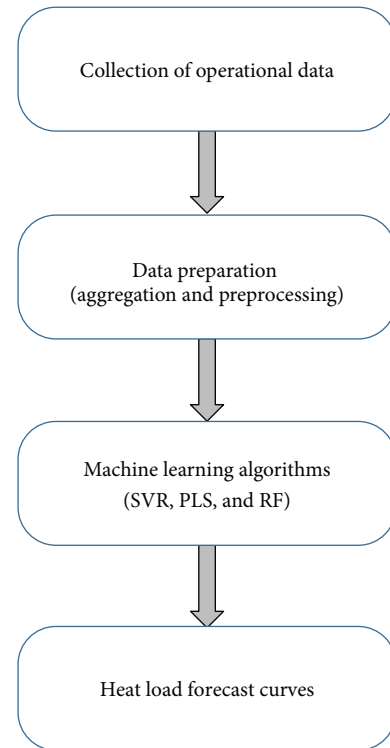


FIGURE 2: Workflow scenario of the proposed heat load approach.

important to keep up with the rapid development in this direction.

4. System Design and Application

As mentioned earlier, in this study, we perform short time prediction for heat consumption and evaluate the three ML methods. For the development of the heat load system presented in our previous work [6], in this section, we present and describe our heat load prediction approach in detail, as shown in Figure 2, which includes collection of operational data, data preparation, and the examined ML algorithms.

In this work, there are two main tasks, which are relevant in the system implementation: (a) data aggregation and preprocessing and (b) ML application, where the heat load prediction is approached with the supervised ML algorithms.

4.1. Operational Data Collection. The data we have used in this study is provided by Eidsiva Bioenergi AS which operates

TABLE 2: Typical data samples for one day from the Eidsiva dataset.

Time (h)	FT (°C)	RT (°C)	Flow (m ³ /h)	HL (MW)
1	100.9	60.4	316.5	14.6
2	99.4	57.5	279.1	13.4
3	99.6	58.3	228.3	13.2
4	100.3	59.8	261.9	15.1
5	100.5	59.6	276.9	14.9
6	100.1	58.9	270.5	12.8
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
20	98.7	59.7	353	14.8
21	101.4	58.9	330.9	16.1
22	101.1	59.1	250.8	14.2
23	100.8	59.8	260.2	15.4
24	101.9	59.3	247.9	14.1

one of Europe’s most modern waste incineration plants located in the city of Hamar, Norway. The plant produces district heating, process steam, and electricity. These data are collected by regular measurements that are part of the control system in a DH plant. The measurements consist of 24 measurements per each day, that is, every hour. The dataset contains values of the parameters: time of day (tD), forward temperature (FT), return temperature (RT), flow rate (FR), and heat load (HL). The data are collected in the period between October 1st 2014 and April 30th 2015. In Table 2, we present a portion of typical data samples for one day.

4.2. Data Preparation. In this module, activities related to preparing the data to be compatible with the ML module are performed. The module includes data aggregation and preprocessing. During the process of data aggregation, we combine the sources of this data with the weather data (outdoor temperature), which is collected at the same interval with previous parameters. Consequently, we obtain from the aggregation process these output parameters: outdoor temperature (OT), heat load (HL), forward temperature (FT), time of day (tD), and the difference between forward temperature (FT) and return temperature (RT), namely, DT.

4.3. Machine-Learning Predictive Modelling. Machine learning (ML) is a very broad subject; it goes from very abstract theory to extreme practice. It turns out that even amongst machine-learning practitioners there is no very well accepted definition of what is machine learning. As a subfield of artificial intelligence, with its objective on building models that learn from data, machine learning has made tremendous improvements and applications in the last decade.

In general, ML can be clearly defined as a set of methods that can automatically detect and extract knowledge patterns in empirical data, such as sensor data or databases, and then use the discovered knowledge patterns to predict future data or execute other types of decision-making under uncertainty. ML is divided into three principal groups: supervised learning (predictive learning approach), unsupervised learning

(descriptive learning approach), and reinforcement learning [29]. In supervised learning, the algorithm is given data in which the “correct answer” for each example is told and the main property of the supervised learning is that the main criteria of the target function $y = f(x)$ are unknown. At the very high level, the two steps of supervised learning are as follows: (i) train a machine-learning model using labeled data that consist of N data pairs $(x_1, y_1), \dots, (x_N, y_N)$, called instances, and (ii) make predictions on new data for which the label is unknown. Each instance is described by an input vector x_i which incorporates a set of attributes $A = \{A_1, A_2, \dots, A_m\}$ and a label y_i of the target attribute that represents the wanted output. To summarize these two steps, the predictive model is learning from past examples made up of inputs and outputs and then applying what is learned to future inputs, in order to predict future outputs. Since we are making predictions on unseen data, which is data that is not used to train the model, it is often said that the primary goal of supervised learning is to build models that generalizes; that is, the built machine-learning model accurately predicts the future rather than the past. Therefore, the goal is to train a model that can afterwards predict the label of new instances and to figure out the target function.

Based on the type of output variable y_i , supervised learning tasks are further divided into two types, as classification and regression problems. In problems where the output variable y_i is categorical or nominal (or belongs to a finite set), the ML tasks are known as classification problems or pattern recognition, whereas in regression problems the output variable is a real valued scalar or takes continuous values.

4.3.1. Support Vector Regression (SVR). Support vector machines (SVM), as a set of supervised learning algorithms based on a statistical learning theory, are one of the most successful and widely applied machine-learning methods, for both solving regression and pattern recognition problems. Since the formulation of SVMs is based on structural risk minimization and not on empirical risk minimization, this algorithm shows better performance than the traditional ones. Support Vector Regression (SVR) is a method of SVM, specifically for regressions. In SVR, the objective function (e.g., the error function that may need to be minimized) is convex, meaning that the global optimum is always reached and satisfied. This is sharply in contrast to artificial neural networks (ANNs), where, for instance, the classical backpropagation learning algorithm is prone to convergence to “bad” local minima [30, 31], which makes them harder to analyze theoretically. In practice, SVR have greatly outperformed ANNs in a wide range of applications [31].

In SVR, the input x is mapped first into an m -dimensional feature space by using nonlinear mapping. As a subsequent step, we construct a linear model in that feature space. Mathematically, the linear model $f(x, w)$ is given by

$$f(x, w) = \sum_{j=1}^m w_j g_j(x) + b, \quad (1)$$

where $g_j(x)$, $j = 1, \dots, m$, represents the set of nonlinear transformations, while b is the bias term, and most of the time is assumed to be zero; hence, we omit this term.

The model obtained by SVR depends exclusively on a subset of the training data; at the same time, SVR tries to reduce model complexity by minimizing $\|w\|^2$. Consequently, the objective of SVR is to minimize the following function [32]:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{such that} \quad & y_i - f(x_i, w) \leq \varepsilon + \xi_i^* \\ & f(x_i, w) - y_i \leq \varepsilon + \xi_i \\ & \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n. \end{aligned} \quad (2)$$

In these equations, ε is a new type of (insensitive) loss function or a threshold, which denotes the desired error range for all points. The nonnegative variables ξ_i and ξ_i^* are called slack variables; they measure the deviation of training samples outside ε , that is, guaranteeing that a solution exists for all ε . The parameter $C > 0$ is a penalty term used to determine the tradeoff between data fitting and flatness, and w are the regression weights. In most cases, the optimization problem can be easily solved if transformed into a dual problem. By utilizing Lagrange multipliers, the dualization method is applied as follows:

$$\begin{aligned} L = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - y_i + f(x_i, w)) \\ - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + y_i - f(x_i, w)) \\ - \sum_{i=1}^n (\lambda_i \xi_i + \lambda_i^* \xi_i^*), \end{aligned} \quad (3)$$

where L is the Lagrangian and $\alpha_i, \alpha_i^*, \lambda_i, \lambda_i^* \geq 0$, are called the Lagrange multipliers.

Considering the saddle point condition, it follows that the partial derivatives of L in relation to variables (w, b, ξ_i, ξ_i^*) will disappear for optimality. By proceeding with similar steps, we end up with the dual optimization problem. Finally, the solution of the dual problem is given by

$$f(x) = \sum_{i=1}^{n_{SV}} (\alpha_i - \alpha_i^*) K(x_i - x), \quad (4)$$

where $0 \leq \alpha_i^* \leq C$, $0 \leq \alpha_i \leq C$, n_{SV} is the number of space vectors, and K is the kernel function, which for given two vectors in input space will return, to a higher dimensional

feature space, the dot product of their images. The kernel is given by

$$K(x, x_i) = \sum_{j=1}^m g_j(x) g_j(x_i). \quad (5)$$

In order to map the input data to a higher dimensional space and to handle nonlinearities between input vectors and their respective class, we use as a kernel the Gaussian radial basis function (RBF), which has γ as its kernel parameter. Once the kernel is selected, we used grid search to identify the best pair of the regularization parameters C and γ , that is, the pair with the best cross-validation accuracy.

4.3.2. Partial Least Squares (PLS). The Partial Least Squares (PLS) technique is a learning method based on multiple linear regression model that takes into account the latent structure in both datasets. The dataset consists of explanatory variables X_i and dependent variables Y_j . The model is linear, as can be seen in (6), that, for each sample n , the value y_{nj} is

$$y_{nj} = \sum_{i=0}^k \beta_i x_{ni} + \varepsilon_{nj}. \quad (6)$$

The PLS model is similar to a model from a linear regression; however, the way of calculating β_i is different. The principle of PLS regression is that the data tables or matrices X and Y are decomposed into latent structure in an iterative process. The latent structure corresponding to the most variation of Y is extracted and explained by a latent structure of X that explains it the best.

The Partial Least Squares (PLS) technique is a learning method based on multivariate regression model, which can be used for correlating the information in one data matrix X to the information in another matrix Y . More specifically, PLS is used to find the fundamental relations between two matrices (X and Y), which are projected onto several key factors, such as T and U , and linear regression is performed for the relation between these factors. Factor U represents the most variations for Y whereas factor T denotes the variations for X , but it is not necessarily explaining the most variation in X .

The first results of PLS are the model equations showing the β coefficients that give the relationship between variables X and Y . These model equations are as follows:

$$\begin{aligned} Y &= X\beta + \varepsilon, \\ Y &= T_h C_h' + \varepsilon_h = XW_h^* C_h' + \varepsilon_h \\ &= XW_h (P_h' W_h)^{-1} C_h' + \varepsilon_h, \end{aligned} \quad (7)$$

where Y is the matrix of dependent variables, X is the matrix of explanatory variables, T_h, C_h, W_h^*, W_h , and P_h are the matrices generated by the PLS algorithm, and ε_h is the matrix of residuals. Matrix β of the regression coefficients of Y on X , with h components generated by the PLS algorithm, is given by

$$\beta = W_h (P_h' W_h)^{-1} C_h'. \quad (8)$$

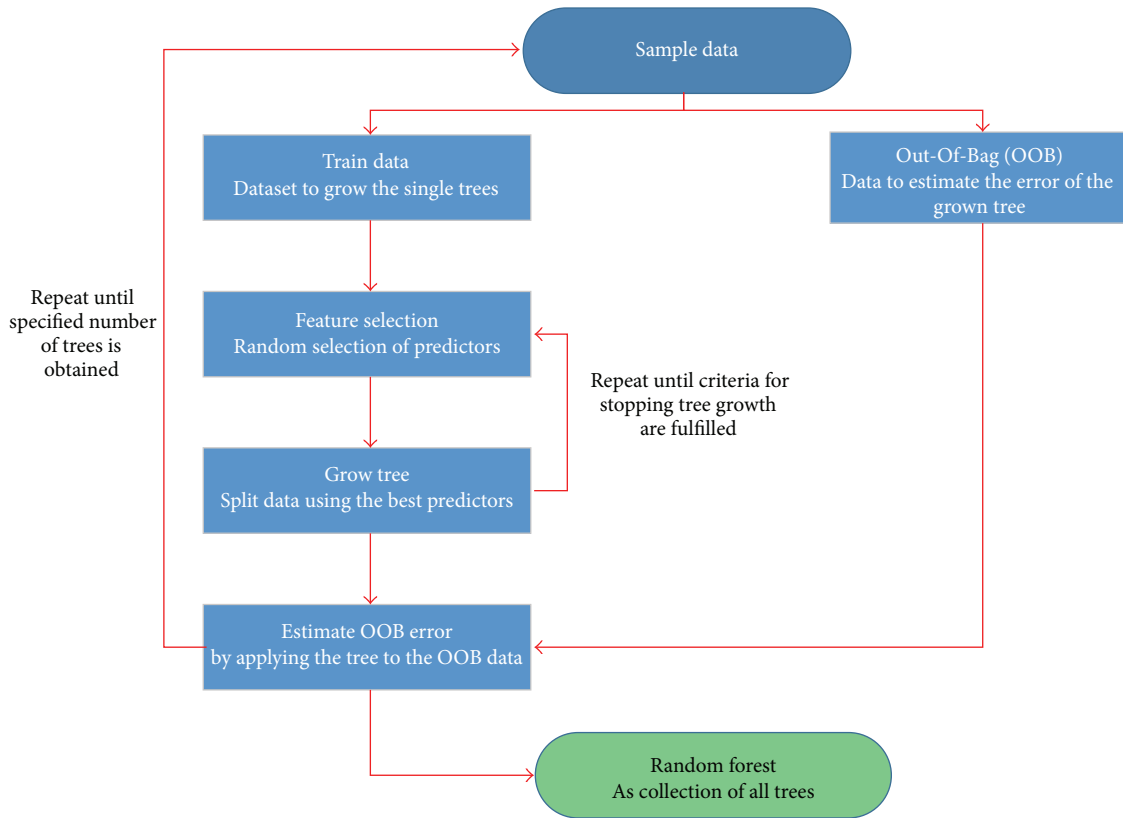


FIGURE 3: Random forest algorithm process flow.

The advantage of PLS is that this algorithm allows taking into account the data structure in both X and Y matrices. It also provides great visual results that help the interpretation of data. Finally, yet importantly, PLS can model several response variables at the same time taking into account their structure.

4.3.3. Random Forest (RF). Random forest algorithm, proposed by Breiman [33], is an ensemble-learning algorithm consisting of three predictors where the trees are formulated based on various random features. It develops lots of decision trees based on random selection of data and random selection of variables, providing the class of dependent variable based on many trees.

This method is based on the combination of a large collection of decorrelated decision trees (i.e., Classification and Regression Trees (CART) [34]). Since all the trees are based on random selection of data as well as variables, these are random trees and many such random trees lead to a random forest. The name forest means that we use many decision trees to make a better classification of the dependent variable. The CART technique divides the learning sample using an algorithm known as binary recursive partitioning. This splitting or partitioning starts from the most important variable to the less important ones and it is applied to each of the new branches of the tree [35].

In order to increase the algorithm accuracy and reduce the generalization error of the ensemble trees, another

technique called *Bagging* is incorporated. The estimation for the generalization error is performed with the Out-Of-Bag (OOB) method. Bagging is used on the training dataset to generate a lot of copies of it, where each one corresponds to a decision tree.

With the RF algorithm, each tree is grown as follows [36]:

- If the number of cases (observations) in the training set is N , sample N cases at random, but with replacement, from the original data, this sample will be the training set for the growing tree.
- If there are M input variables, a number $m \ll M$ is specified such that, at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. This value m is held constant during the forest growing.
- Each tree is grown to the largest extent possible.

The process flow in the random forest models is shown in Figure 3.

The error rate of RF primarily depends on the correlation degree between any two trees and the prediction accuracy of an individual tree. The principal advantages of this method are the ease of parallelization, robustness, and indifference to noise and outliers in most of the dataset. Due to its unbiased nature of execution, this method avoids overfitting.

TABLE 3: Performance results for one week.

Algorithm	Training			Testing		
	MAE (%)	MAPE (%)	Correlation coefficient	MAE (%)	MAPE (%)	Correlation coefficient
SVR	64.91	3.11	0.90	70.32	3.43	0.91
PLS	265.24	12.73	0.81	238.68	10.42	0.79
RF	173.82	19.77	0.83	159.32	18.61	0.84

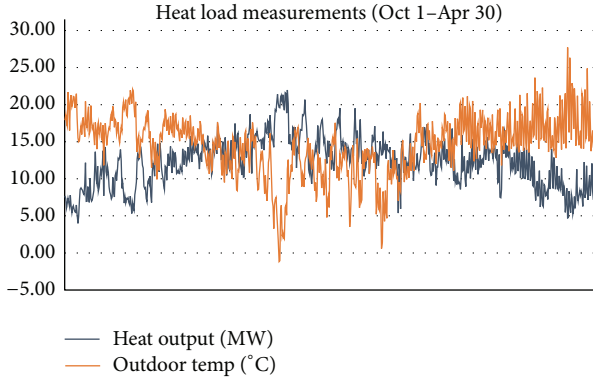


FIGURE 4: Hourly heat load pattern in relation to outdoor temperature in the DH network.

5. Performance Evaluation and Results

The proposed approach is implemented in MATLAB R2014a [37] and executed in a PC with Intel® Core i7 processor with 2.7 GHz speed and 8 GB of RAM. In this work, as training dataset, we select the data measured during the first 28 weeks, which consist of 4872 instances. As for prediction period, we choose the 29th week as test data, that is, 148 instances. In order to evaluate the performance of the proposed algorithms in terms of accuracy of the results, we use the mean absolute average (MAE), the mean average percentage error (MAPE), and the correlation coefficient, which measures the correlation between the actual and predicted heat load values. MAE and MAPE are defined as follows:

$$\begin{aligned} \text{MAE} &= \frac{\sum_{i=1}^n |p_i - y_i|}{n}, \\ \text{MAPE} &= \frac{\sum_{i=1}^n |(p_i - y'_i) / p_i|}{n}, \end{aligned} \quad (9)$$

where p_i is the actual value, y_i is the predicted value, and n is the number of samples in the training set.

We apply a 10-fold cross-validation for obtaining the valid evaluation metrics. Figure 4 presents the heat load in the network with respect to the outdoor temperature. We see that the higher values of the heat load occur during the days with lower outdoor temperature values, which in fact reflects the increased consumption of heat.

Figure 5(a) shows results of actual heat load and heat load prediction for one week, based on SVR algorithm. In Figure 5(b), results of heat prediction production for one week based on PLS are shown with data on actual heat load,

whereas results of heat load forecasting for one week based on RF are shown in Figure 5(c).

As can be seen from Figure 5, the predicted HL with SVR is closer to the actual energy consumption, with an MAPE value of about 3.43% and a correlation coefficient of 0.91. Correlation has been consistent throughout training and testing. On the other hand, graphs presented in Figures 5(b) and 5(c) for PLS and RF, respectively, are less accurate with higher errors. The performance of the PLS is significantly lower compared to SVR. Concerning RF, as the trees started to become progressively uncorrelated, the error rate also declined significantly. The best performance of the SVR over the other two methods is attributed to the efficient modelling of feature space and to the fact that SVR is less prone to overfitting, and that does not require any procedure to determine the explicit form as in ordinary regression analysis.

Table 3 outlines the results of the investigated ML algorithms for heat consumption prediction for both training phase (where we developed the models using supervised learning schemes) and testing phase (to generalize newly unseen data). From Table 3, it is evident that SVR shows the best prediction performance in terms of average errors and correlation coefficient, confirming the superiority of SVR over the other machine-learning methods. Therefore, based on the assumption that the number and type of the operating facilities should be determined, SVR can be effectively applied in the management of a DHS. The mean absolute percentage error value of 3.43% obtained in our approach with the SVR is lower than or equal to the mean absolute percentage error of state-of-the-art approaches for heat load prediction in DHS. Moreover, SVR is also better than PLS and RF in terms of mean average errors and correlation coefficient. Nevertheless, sometimes it is impossible to perform direct comparison with other works due to different system implementation and input data and different structure of the experimental setup.

5.1. Comparison with State-of-the-Art Methods. Some prior research work is carried out to predict and analyze the head demand in DH. However, due to different system design and input data and different architecture implementation or structure of the experimental setup, sometimes it is difficult to perform direct comparison. Our approach uses operational data from DHS to model the heat demand, and since the SVR exhibited the best prediction performance, we use this method to perform the comparison. In our case, we obtained smaller MAPE compared to [17], where the experimental results showed MAPE of 4.77% which is lower than or at least equal to the mean percentage error of state-of-the-art regression approaches that have been proposed for heat load forecasting in DH systems. Furthermore, reported results

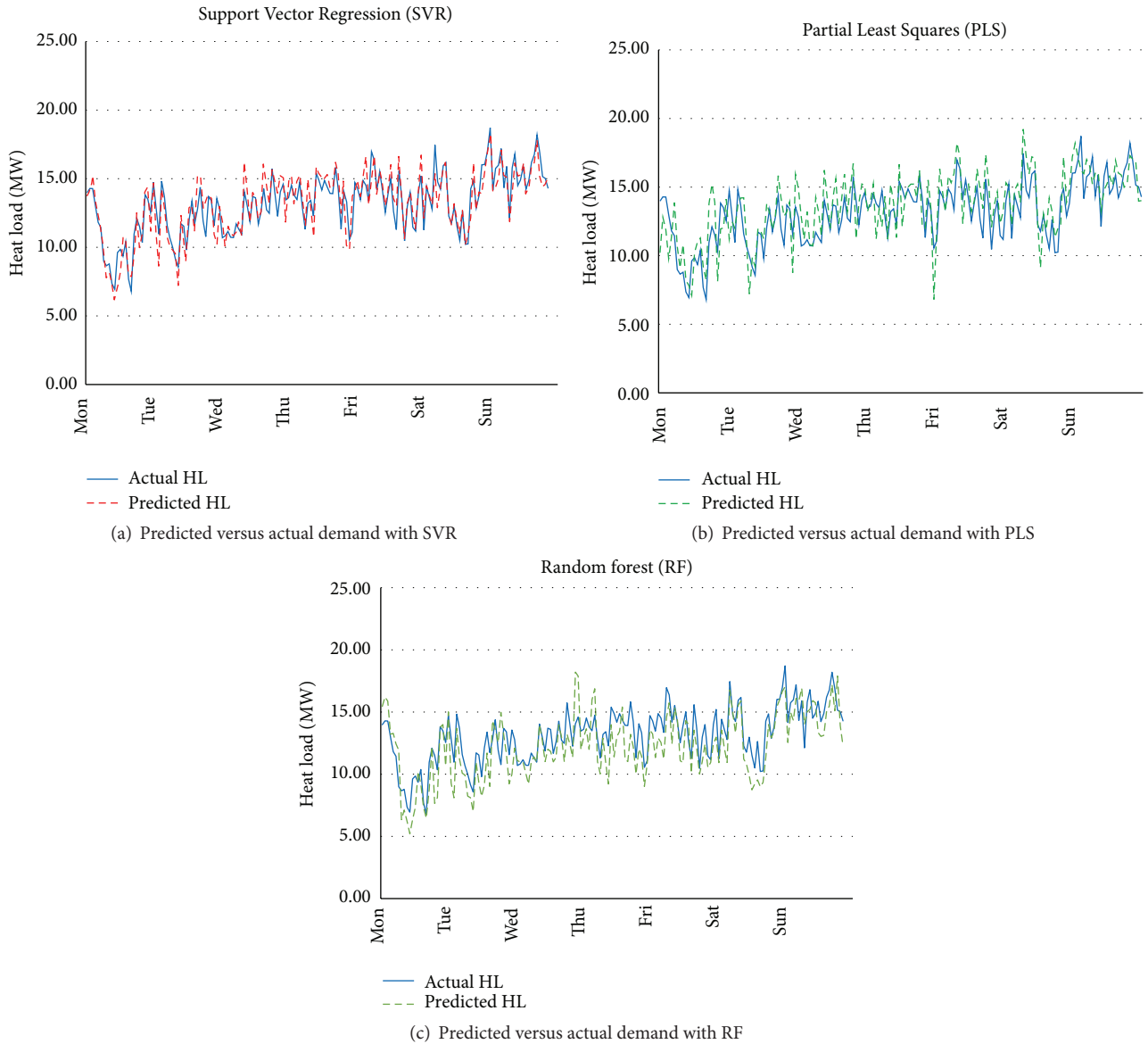


FIGURE 5: Hourly heat load prediction based on SVR, PLS, and RF for one week.

from our study are also superior compared to the results presented in [15, 32]. More specifically, the SVR method we apply shows better results in terms of MAPE and correlation coefficient compared to [32], where MAPE in this paper is 5.54% and the correlation coefficient is 0.87. As far as the comparison with work [15] is concerned, in terms of MAPE, our SVR method exhibits better hourly prediction performance, where the MAPE for one week is 5.63%. On the other hand, the PLS method performs better than in our case, having the MAPE value of about 8.99%.

6. Conclusion

District heating (DH) sector can play an indispensable role in the current and future sustainable energy system of the North European countries, where the share of DH in the total

European heat market is significantly high. The innovations and emergence of new technology and the increasing focus on smart buildings impose energy systems to face a challenge in adapting to customers' more flexible and individual solutions. Consequently, one needs to know more about what drives customers' requirements, choices, and priorities. Therefore, the creation and application of innovative IT ecosystems in DH are considered essential to improve energy efficiency.

Heat load prediction in the past decade has attracted a lot of interest to the researchers, since it can assist in energy efficiency of the DHS, which also leads to cost reduction for heat suppliers and to many environmental benefits.

In this paper, three ML algorithms for heat load prediction in a DH network are developed and presented. The algorithms are SVR, PLS, and RF. The heat load prediction models

were developed using data from 29 weeks. The predicted hourly results were compared with actual heat load data. Performances of these three different ML algorithms were studied, compared, and analyzed. SVR algorithm proved to be the most efficient one, producing the best performance in terms of average errors and correlation coefficient. Moreover, the prediction results were also compared against existing SVR and PLS methods in literature, showing that the SVR presented in this paper produces better accuracy.

In conclusion, the comparison results validate the notion that the developed SVR method is appropriate for application in heat load prediction or it can serve as a promising alternative for existing models.

As for the future work, apart from outdoor temperature, we intend to incorporate other meteorological parameters influencing the heat load, such as wind speed and humidity.

Competing Interests

The authors declare that they have no competing interests.

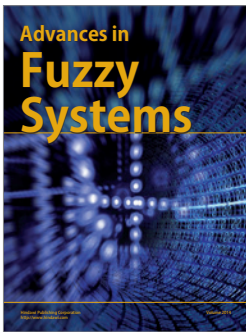
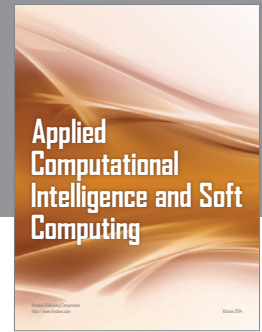
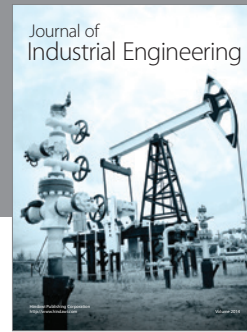
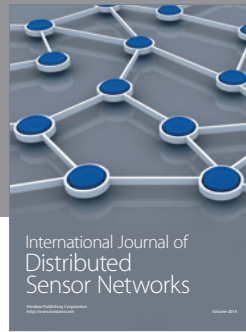
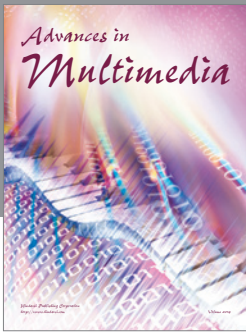
Acknowledgments

The authors would like to thank *Eidsiva Bioenergi AS*, Norway, which kindly provided the data for this research work.

References

- [1] Report on European Energy Security Strategy, May 2014, Brussels, <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014DC0330&from=EN>.
- [2] A. R. Day, P. Ogumka, P. G. Jones, and A. Dunsdon, "The use of the planning system to encourage low carbon energy technologies in buildings," *Renewable Energy*, vol. 34, no. 9, pp. 2016–2021, 2009.
- [3] IEA (International Energy Agency), *Nordic Energy Technology Perspectives*, OECD/IEA, Paris, France, 2013.
- [4] A. Gebremedhin and H. Zinko, "Seasonal heat storages in district heating systems," in *Proceedings of the 11th International Thermal Energy Storage for Energy Efficiency and Sustainability Conference (Effstock '09)*, Stockholm, Sweden, June 2009.
- [5] H. Lund, S. Werner, R. Wiltshire et al., "4th Generation District Heating (4GDH): integrating smart thermal grids into future sustainable energy systems," *Energy*, vol. 68, pp. 1–11, 2014.
- [6] F. Dalipi, S. Y. Yayilgan, and A. Gebremedhin, "A machine learning approach to increase energy efficiency in district heating systems," in *Proceedings of the International Conference on Environmental Engineering and Computer Application (ICEECA '14)*, pp. 223–226, CRC Press, Hong Kong, December 2014.
- [7] N. Fumo, "A review on the basics of building energy estimation," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 53–60, 2014.
- [8] D. E. Claridge, "Using simulation models for building commissioning," in *Proceedings of the International Conference on Enhanced Building Operation*, Energy Systems Laboratory, Texas A&M University, October 2004.
- [9] T. Mestekemper, *Energy demand forecasting and dynamic water temperature management [Ph.D. thesis]*, Bielefeld University, Bielefeld, Germany, 2011.
- [10] E. Dotzauer, "Simple model for prediction of loads in district-heating systems," *Applied Energy*, vol. 73, no. 3-4, pp. 277–284, 2002.
- [11] H. A. Nielsen and H. Madsen, "Modelling the heat consumption in district heating systems using a grey-box approach," *Energy and Buildings*, vol. 38, no. 1, pp. 63–71, 2006.
- [12] B. B. Ekici and U. T. Aksoy, "Prediction of building energy consumption by using artificial neural networks," *Advances in Engineering Software*, vol. 40, no. 5, pp. 356–362, 2009.
- [13] M. Grzenda and B. Macukow, "Demand prediction with multi-stage neural processing," in *Advances in Natural Computation and Data Mining*, pp. 131–141, 2006.
- [14] K. Kato, M. Sakawa, K. Ishimaru, S. Ushiro, and T. Shibano, "Heat load prediction through recurrent neural network in district heating and cooling systems," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '08)*, pp. 1401–1406, Singapore, October 2008.
- [15] T. C. Park, U. S. Kim, L.-H. Kim, B. W. Jo, and Y. K. Yeo, "Heat consumption forecasting using partial least squares, artificial neural network and support vector regression techniques in district heating systems," *Korean Journal of Chemical Engineering*, vol. 27, no. 4, pp. 1063–1071, 2010.
- [16] T. Catalina, V. Iordache, and B. Caracaleanu, "Multiple regression model for fast prediction of the heating energy demand," *Energy and Buildings*, vol. 57, pp. 302–312, 2013.
- [17] S. Provasas, N. Lavesson, and C. Johansson, "An online machine learning algorithm for heat load forecasting in district heating systems," in *Proceedings of the 14th International Symposium on District Heating and Cooling*, Stockholm, Sweden, September 2014.
- [18] S. Idowu, S. Saguna, C. Åhlund, and O. Schelén, "Forecasting heat load for smart district heating systems: a machine learning approach," in *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm '14)*, pp. 554–559, IEEE, Venice, Italy, November 2014.
- [19] L. Wu, G. Kaiser, D. Solomon, R. Winter, A. Boulanger, and R. Anderson, *Improving Efficiency and Reliability of Building Systems Using Machine Learning and Automated Online Evaluation*, Columbia University Academic Commons, 2012, <http://hdl.handle.net/10022/AC:P:13213>.
- [20] M. Vlachopoulou, G. Chin, J. C. Fuller, S. Lu, and K. Kalsi, "Model for aggregated water heater load using dynamic Bayesian networks," Tech. Rep., Pacific Northwest National Laboratory (PNNL), Richland, Wash, USA, 2012.
- [21] S. Shamshirband, D. Petković, R. Enayatifar et al., "Heat load prediction in district heating systems with adaptive neuro-fuzzy method," *Renewable and Sustainable Energy Reviews*, vol. 48, pp. 760–767, 2015.
- [22] M. Protić, S. Shamshirband, M. H. Anisi et al., "Appraisal of soft computing methods for short term consumers' heat load prediction in district heating systems," *Energy*, vol. 82, pp. 697–704, 2015.
- [23] S. Shamshirband, D. Petković, A. Amini et al., "Support vector regression methodology for wind turbine reaction torque prediction with power-split hydrostatic continuous variable transmission," *Energy*, vol. 67, pp. 623–630, 2014.
- [24] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 303–315, 2014.

- [25] K. Bhaskar and S. N. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 306–315, 2012.
- [26] Heat Pump Centre Newsletters, International Energy Agency (IEA) Heat Pumping Technologies, Vol. 30, Nr. 2/2012.
- [27] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization," *Machine Learning*, vol. 40, no. 2, pp. 139–157, 2000.
- [28] Swedish District Heating Association, 2016, <http://www.svenskfjarrvarme.se/>.
- [29] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, Adaptive Computation and Machine Learning Series, MIT Press, 2012.
- [30] L. Wang, *Support Vector Machines: Theory and Applications*, Springer, New York, NY, USA, 2005.
- [31] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regressions," *Neural Information Processing—Letters and Reviews*, vol. 11, no. 10, pp. 203–224, 2007.
- [32] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: a machine learning case study," *Energy and Buildings*, vol. 49, pp. 591–603, 2012.
- [33] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [34] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth and Brooks, Monterey, Calif, USA, 1984.
- [35] P. Vezza, C. Comoglio, M. Rosso, and A. Viglione, "Low flows regionalization in North-Western Italy," *Water Resources Management*, vol. 24, no. 14, pp. 4049–4074, 2010.
- [36] L. Breiman and A. Cutler, "Random forests," January 2016, <http://www.stat.berkeley.edu/users/breiman/RandomForests/>.
- [37] *MATLAB, Version 8.3.0*, MathWorks, Natick, Mass, USA, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

