

Forecasting Heat Load for Smart District Heating Systems: A Machine Learning Approach

Samuel Idowu, Saguna Saguna, Christer Åhlund and Olov Schelén
Luleå University of Technology
Luleå, Sweden
Email: {firstname}.{lastname}@ltu.se

Abstract—The rapid increase in energy demand requires effective measures to plan and optimize resources for efficient energy production within a smart grid environment. This paper presents a data driven approach to forecasting heat load for multi-family apartment buildings in a District Heating System (DHS). The forecasting model is built using six and eleven weeks of data from five building substations. The external factors and internal factors influencing the heat load in substations are parameters used as our model’s input. Short-term forecast models are generated using four supervised Machine Learning (ML) techniques: Support Vector Regression (SVR), Regression Tree, Feed Forwards Neural Network (FFNN) and Multiple Linear Regression (MLR). Performance comparison among these ML methods was carried out. The effects of combining the internal and external factors influencing heat load at substations was studied. The models are evaluated with varying horizon up to 24-hours ahead. The results show that SVR has the best accuracy of 5.6% MAPE for the best-case scenario.

Keywords—Data driven modeling, district heating, energy efficiency, machine learning, smart cities, smart grid.

I. INTRODUCTION

The current increase in global demand for energy will challenge energy production and its supply. Energy sustainability and efficiency can be achieved by applying an optimal approach at the production and consumption ends. Equally important is the need for flexibility in economic and technical planning in a smart-grid. We see smart-grids encompassing both power and district heating. To achieve this goal, a load forecasting model is essential for both power and district heating grids at the consumption side.

Related work in the area of energy (e.g. thermal, electric, cooling) demand estimation in buildings is categorized into *forward (classical)* and *data-driven (inverse)* approaches [1]. The former uses equations that describes the physical behavior of a system to predict the outputwhile the latter approach relates to ML where measurements of input and output variables of a system are collected. The measured data is then used to define mathematical description of this system [1]. Recently, several researchers have considered the use of ML to develop data driven systems due to large scale availability of sensor generated data. This is due to the effectiveness of the ML approach. For estimating energy demand at the consumption environment commonly used data-driven methods are SVR [2], Multiple Regression [3], Neural Networks based methods [4][5]. In specific to heat load forecast, some of the advantages of data-driven approach over a classical approach include the

ability to discover models from large volumes of data and the ability to adapt and update models based on new data [6]. The latter point is essential due to the non-stationary nature of heat load profile.

In District Heating Systems (DHS), numerous developments have been made to obtain efficient operational management from the financial and environmental perspectives. However, most prior work are limited to the production environment [7] [8] [5]. These relate to the *top-down* approach [1]. [7] highlights the two types of heat load influencing factors as *external* (e.g., outdoor temperature) and *internal* (e.g., supply temperature) factors. The former refers to outdoor meteorological variables and the latter means the factors relating to operational characteristics of a DHS. For increased flexibility in planning, it is essential to consider the consumption environment, where local planning and process control could help reduce a) energy consumption, b) emissions such as CO_2 , c) fossil fuel consumption in Combined Heating and Power (CHP) plant and d) peak demand. These can be achieved through optimal control of local energy saving strategies, and via the shortening of control-loop between the production and consumption side. In this paper, we focus on applying a bottom-up approach using ML to estimate heat energy at the consumption side of a DHS.

Our key contributions in this paper are:

- We apply and compare different ML algorithms (SVR, FFNN, MLR and Regression Tree) in the context of forecasting heat load for multi-family apartment buildings in a DHS at various forecast horizons of 1, 3, 6, 12, 18 and 24-hours.
- We study the effects of the internal and external factors for heat load in a DHS at the consumer side.

This paper is organized as follows: Section 2 presents the DHS background and related work. Section 3 describes the system architecture and further, Section 4 describes the modeling process for heat load forecasting. Section 5 presents the evaluation and results while Section 6 presents the discussion, conclusion and the future work.

II. BACKGROUND AND RELATED WORK

A DHS produces heat energy in a centralized location (e.g. CHP plants) and supplies it to residential and commercial buildings primarily for space and domestic water heating. It has three main parts: 1) *Production-side*: usually consists of

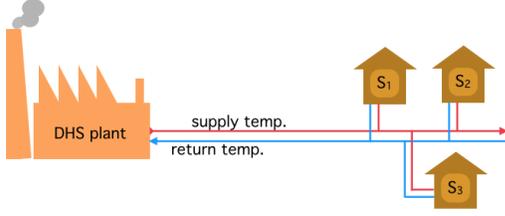


Fig. 1: A simple diagram of a DHS.

a co-generation plant (e.g., heat and electrical power) or a heat-only boiler station (run by burning fuel such as biofuel, oil or gas for heat). 2) *Distribution network*: 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. 3) *Consumer-side*: consists of a substation where heat energy is transferred from the primary network to the secondary via a heat-exchanger which may include control valve actuators, speed pump controls, flow meters and gauges. A diagram (Fig. 1) shows the basic schematic drawing of a DHS. The red and blue lines denote supply and return pipes, respectively.

The net heat energy, Q_{net} delivered to the entire grid mainly depends on the supply temperature and the flow rate. Q_{net} can be formulated as a function of heat energy at each substation as shown in Eqn. 1.

$$Q_{net} = Q_{loss} + \sum_{i=1}^n Q_{S_i} \quad (1)$$

Where Q_{S_i} is the heat demand at substation S_i . Q_{loss} , which varies based on soil temperature around the pipes, is heat lost during energy transport. The top-down approach focuses on Q_{net} while the bottom-up approach considers Q_{S_i} . Our work takes the bottom-up approach. Q_{S_i} is affected by various parameters such as meteorological parameters, building characteristics and social behavioral patterns of occupants. Apart from these attributes, internal physical dynamics within DHS can be used for forecasting heat load. Table I shows some of the identified influencing factors based on the classification described in [7].

A. Related Work

1) *Production side*: Related work that focuses on production side deals with analysis or forecast of Q_{net} in DHS (i.e., top-down approach). [8] presented a novel assessment method which describes daily variations of heat load in DHS. This assessment could be used for design and planning of storage for the DHS network. [7] [5] [4] and [9] presented a load forecasting methods in DHS with limitation to the production environment. [5] proposed a new heat load prediction method which uses a recurrent neural network to deal with the dynamic variation of heat load and its characteristics. The approach shows decent prediction accuracy for non-stationary heat load. [4] proposed a heat 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 network. [9] used a grey-box approach to identify the model that links the heat consumption in a large geographical area to

TABLE I: Influencing factors of heat load in DHS

Internal Factors	External Factors	
	Behavioral/Seasonal Factors	Meteorological Factors
- Supply temperature(ST) - Return temperature (RT) - Supply pressure - flow rate (FR)	- Hour of day (hD) - day of week (dW) - month of year (mY)	- Outdoor Temperature (OT) - Humidity - solar radiation

its climate and the calendar information. The process involved a theoretical based identification of an overall building model structure followed by data based modeling. [7] applied wavelet analysis in combination with neural network, and its evaluation shows the approach is suitable for short-term thermal load forecast.

2) *Consumer side*: Related work that focuses on consumer side deals with forecast or analysis of Q_{S_i} in DHS (i.e., bottom up approach). [10] is a recent related work which employs the *bottom-up* approach [1] with focus on the consumption environment and considers single family buildings. The work used computationally effective recursive least squares scheme with meteorological variables as input. The model presented in [10] provides forecast up to 42-hrs forecast horizon. [11] 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 authors concluded time-series analysis is powerful and appropriate for predicting thermal load in DHS.

Our work differs from related work as we focus on using ML algorithms for forecasting heat load in multi-family apartment buildings. Further, we also investigate the impact of a combination of internal and external parameters at the consumer/substation side of a DHS.

III. SYSTEM ARCHITECTURE

In this section, we present the architecture of our target system, which presents the big-picture of this paper. A diagram (Fig. 2) shows the architecture and its key components. In this work, the tasks under consideration are 1) data collection, 2) data aggregation and preprocessing, and 3) the application of ML for prediction. These tasks are described in the following sub-sections. A more detailed description of the target system is presented in [12].

A. Data Collection and Description

This work is based on data that is intrusively collected from DHS substations served by Skelleftea Kraft AB (Sweden's fourth largest energy producer). Skelleftea Kraft AB's CHP plant supplies heat to approximately 5000 substations. For this work, five different substations, which serve multi-family apartment buildings, were equipped with data acquisition sensors. At each substation, relevant parameters are monitored and collected. The substation id, the number of apartments, per annual energy consumption, the mean, minimum and maximum heat load for each building substation is shown (Table II).

The sampling interval for the data acquisition is one minute. Parameters that relate to the internal state of the substations are measured on the primary side of the DHS, these

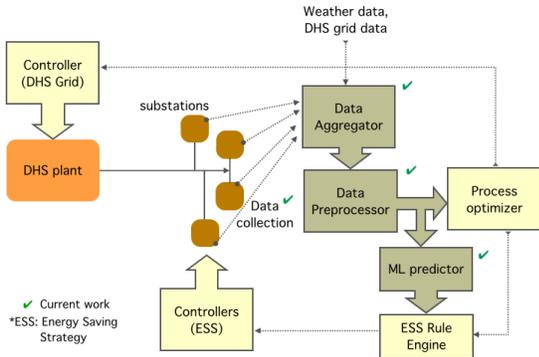


Fig. 2: A high level system architecture showing the focus of this work. The marked components are the focus of this paper.

are: supply temperature (ST), return temperature (RT), flow rate (FR) and the consumed heat load (HL). The apartment buildings are differently geolocated, and as such will have different micro-weather conditions. Hence, the outdoor temperature (OT) is collected locally at same interval with previous parameters. A plot (Fig. 3) of HL data from the substations between the period of 13/01/14 and 06/04/14 for Substation A, B, C and D is shown. Substation E has a shorter period between 17/02/14 and 06/04/14. The selected time range is a subset of data collected up to date, and it represents the period with the most consistency without missing data. We shall compare the forecast performance obtained from substations with longer data duration with Substation E which has a smaller data duration.

B. Data Aggregation and Preprocessing

The data aggregation involves merging sources of data from each substation with its weather station data and the corresponding time factor information. The output parameters from the data aggregation unit are OT , HL , ST , FR , supply and return temperature difference (DT), and the hour of day (hD). 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 hourly interval which, is a much lower frequency than one minute interval. Hence, the one minute interval data set is averaged over one hour to give a new data set of hourly interval. An essential part of the data preprocessing function is data transformation which outputs the predictor's parameters and their corresponding target variables. The transformation is mainly based on the target forecast horizon (see Section IV for details).

C. Supervised Machine Learning Methods

A supervised ML method learns a mapping from input \mathbf{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 [13]. Each example input \bar{x}_i is

TABLE II: Details of multi-family apartment building and their substations. † - units in MWh. * - units in KW.

Substation ID	Per Consumption†	Annual Consumption†	Mean*	Min*	Max*	No. of Apart.
Subst_A	365.35		59.2	21.6	124.5	60
Subst_B	1560.0		250	121.6	503.5	120
Subst_C	2400.0		367.1	182.6	675.1	185
Subst_D	77.905		12.2	2.6	41.1	12
Subst_E	1845.24		221.4	131.6	370.1	174

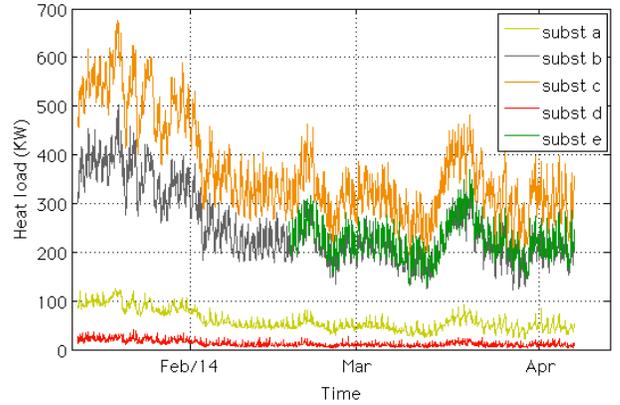


Fig. 3: Plot of hourly load pattern at five substations

a D -dimensional vector of values representing the parameters of an instance (OT , hD , ST , FR , DT). The output y_i part of the training set represents the *class* or *label* of its respective parameters (i.e. HL) [13]. When the output variable y_i is categorical or nominal, these types of ML tasks are referred to as *classification* or *pattern recognition* [13]. However, when the output variable y_i is a continuous data such as HL , the task is referred to as a *regression* task [13]. 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 [2][3][4][5]. The following subsections briefly outline the technical details for each method.

1) *Multiple linear Regression*: MLR is a learning technique based on fitting a linear function with multiple independent variables. Eqn. 2 shows the general form of MLR [14]:

$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 \cdots + \beta_n \cdot X_n. \quad (2)$$

Where Y is the target value. x_1, x_2, \dots, x_n are the input parameters, and β represents the functional weights. α is a constant offset factor used to partially reduce the effect of modeling errors.

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 inputs parameters to their respective output without the assumption about the relationships between the pairs [14]. 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 a 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. Eqn. 3 shows the mathematical representation of a FFNN with a single hidden layer for its function approximation [14].

$$f(x) = \sum_{j=1}^N w_j \Psi_j \left[\sum_{i=1}^M w_{ij} x_i + w_{io} \right] + w_{jo} \quad (3)$$

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

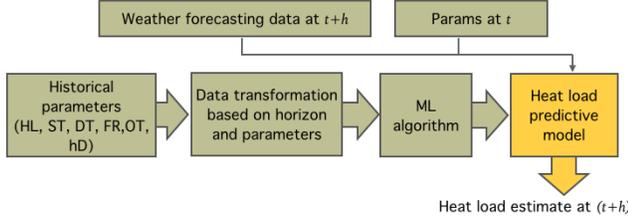


Fig. 4: Modeling process

a gradient-based approach to updates its weight by minimizing a specific error function. In this study, we used the Mean Square Error (MSE) function.

3) *Support Vector Regression*: SVR is a method of Support Vector Machines (SVM) specifically for regressions. SVMs are based on the principle of structural risk minimization [15]. 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 [14]:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^* \quad (4)$$

with the following constraints:

$$y_i - w^T \phi(\bar{x}_i) - b \leq \epsilon + \xi_i \quad (5)$$

$$w^T \phi(\bar{x}_i) + b - y_i \leq \epsilon + \xi_i^* \quad (6)$$

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.

4) *Classification and Regression Tree*: CART [16] 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.

IV. MODELING

This section describes the modeling approach taken in this work. Given the current time t , we create a model capable of estimating the HL forecast at time $t+h$ (where h is the forecast horizon). A diagram (Fig. 4) shows the processes involved in our modeling. In this paper, external input parameters OT , hD and internal input parameters ST , FR and RT are considered for predicting the thermal load. For practical reasons, it is important to reduce the number of input variables as this removes any redundant dimension and reduces complexity of the model. The choice of factors chosen in this work is based on the significance level of the influencing factors presented

TABLE III: Correlation coefficients between HL and different h at the five substations. The least correlation are shown in bold font.

Time, Horizon	Substn_A	Substn_B	Substn_C	Substn_D	Substn_E
HL_t, HL_{t+1}	0.9685	0.9600	0.9742	0.7508	0.8893
HL_t, HL_{t+3}	0.9238	0.9170	0.9313	0.6676	0.7029
HL_t, HL_{t+6}	0.8727	0.8528	0.8740	0.6111	0.4950
HL_t, HL_{t+12}	0.8389	0.8517	0.8538	0.6375	0.5565
HL_t, HL_{t+18}	0.8315	0.8109	0.8177	0.5622	0.4140
HL_T, HL_{t+24}	0.9020	0.8826	0.8782	0.6917	0.7580

in [7]. These include OT , ST , FR , and DT , where DT is the temperature difference between ST and RT . In addition to these, the time factor parameter, hD , which models the behavior and cycle pattern is included. Table I shows the categories of these influencing factors.

The historical data set of these parameters is transformed into instances of predictor variables and their corresponding target variable for h -horizon forecasting. The transformation is done such that the influencing factors for time t are paired with HL at time $t+h$. The actual HL value at time t is also considered as a predictor for HL at $t+h$. Also, the OT forecast at time $t+h$ is considered as a predictor. The resulting data set from the transformation is then fed into different supervised machine learning algorithms to output a suitable forecasting model.

A. Varying Forecast Horizon

One of the aims of the target system presented in Section III is to shorten the control-loop within a DHS. Hence, it is our interest to consider the performance of the proposed model with varying forecast horizon. We analyze the pair relationship between HL at time t and each HL at horizons 1, 3, 6, 12, 18 and 24. The results (Table III) for substation A - E show that, against the general expectation, the coefficient correlation value would decrease with an increase in horizon, the correlation of HL at 24- h is significantly higher than that of 12- h and 18- h .

B. Essential Supervised Machine Learning Steps

The essential steps taken in ML include data identification, collection and preprocessing [13]. These steps, with specifics to this paper, are discussed in Section III. Defining the training/test data is an important step in ML. The training set is the heart of supervised ML since it generates the output model. ML practice involves using a sub-set of the processed data as test data, which is used for evaluation. 11 weeks of data is available for Substation A, B, C and D, while 6 weeks of data is available for substation E. We used two phases PHASE-A and PHASE-B for the training data set selection. In PHASE-A, we chose the first 10 weeks (5 weeks for Substation E) as training data and used the 11th week (the 6th week for Substation E) as test data. The generated model using PHASE-A is used in providing visual plot for examining the performance of models. For PHASE-B, we apply a 10-fold cross-validation for providing the valid evaluation metrics. A performance metric is given as values averaged over the 10 folds with its standard deviation. The cross validation used in this work retains the sequential order of data in each fold, as it is essential to retain the temporal properties of the time-series data.

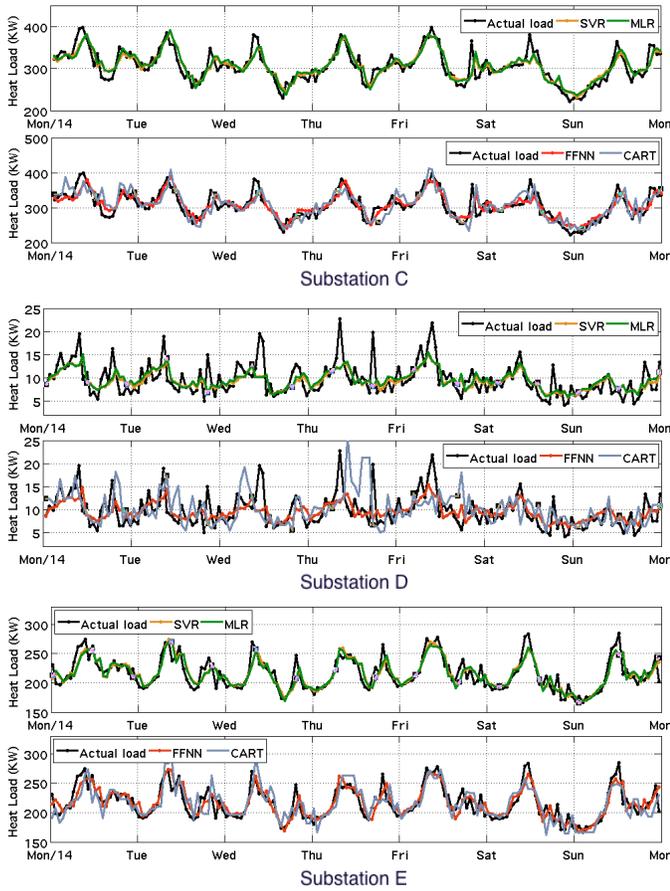


Fig. 5: 24-h forecast plot at Substations C, D and E, using PHASE-A test. Substation A and B have similar pattern as substitution C.

V. EVALUATION AND RESULTS

A. Evaluation metrics

This work considered commonly used metrics for evaluating the performance of the proposed model. These are Root Mean-Square Error (RMSE), the Mean Absolute Percentage Error (MAPE) and Correlation coefficient (Corr Coef). RMSE is commonly used to measure the difference between a model's predicted values and actual values observed. MAPE estimates how close forecast values are to actual values in percentage while Corr Coef measures the strength of correlation between the actual observed heat load and the predicted value.

B. Implementation: Modeling tool and parameters

This work was performed in Matlab R2013a [17] using its ANN Toolbox for the FFNN, while its inbuilt functions for *LinearModel* class and the Regression Tree were used for MLR and CART respectively. LIBSVM [18] was used for the SVM algorithm. MLR is a simple approach, which requires little or no parameter tweaking. For the FFNN algorithm, this study considered a FFNN with a single hidden layer, which is a similar structure used in previous studies [14]. For its activation/transfer function this work used $tansig(x)$ since several related work has shown decent performance using the $tansig(x)$ as transfer function [14]. The performance of SVR algorithm depends on the choice of the regularization parameter C and the kernel function ϕ parameters. We select

TABLE IV: 24-h forecast results for all substations. The best result at each substation is shown in bold font.

	RMSE (KW)	MAPE (%)	Corr. Coef.
Substation A			
SVR	4.7117±1.52	5.9480±1.47	0.8766±0.06
MLR	4.5120±0.72	5.8816±1.62	0.8726±0.06
FFNN	4.6642±1.08	6.0385±1.61	0.8723±0.06
CART	7.1563±2.45	9.3047±1.51	0.7662±0.11
Substation B			
SVR	6.3896±1.26	6.3896±1.26	0.8197±0.06
MLR	6.5775±1.40	6.5775±1.40	0.7992±0.06
FFNN	6.5791±1.38	6.5791±1.38	0.8007±0.56
CART	8.6496±1.66	8.6496±1.66	0.6928±0.13
Substation C			
SVR	24.7205±2.98	5.5659±1.21	0.8506±0.06
MLR	25.7431±2.50	5.7458±1.21	0.8297±0.07
FFNN	25.8459±3.02	5.8319±1.21	0.8290±0.07
CART	35.0218±7.08	7.7138±1.48	0.7434±0.08
Substation D			
SVR	3.2233±0.51	21.5455±5.71	0.5329±0.11
MLR	3.3271±0.49	23.3312±6.09	0.4818±0.13
FFNN	3.3301±0.45	24.2491±7.38	0.4772±0.15
CART	4.1839±0.59	30.0262±9.89	0.3621±0.14
Substation E			
SVR	16.5095±1.96	5.8075±0.65	0.8390±0.04
MLR	17.1391±2.26	5.9638±0.78	0.8170±0.06
FFNN	17.1365±2.23	6.0256±0.86	0.8185±0.06
CART	23.5749±3.15	8.1354±0.88	0.6724±0.08

TABLE V: Comparison of 24-h model performance for scenario 1-4. OT* refers to OT at time t and OT at time $t+h$

Substaion	Scenario 1 {OT*, hD, HL}	Scenario 2 {OT*, hD, HL ST, FR, DT}	Scenario 3 {OT*, hD, ST, FR, DT}	Scenario 4 {OT*, hD}
substn A	5.78%	5.85%	5.89%	8.79%
substn B	6.16%	6.22%	6.58%	8.37%
substn C	5.55%	5.577%	6.00%	7.10%
substn D	21.66%	21.91%	22.40%	21.73%
substn E	5.64%	5.68%	5.80%	7.29%

a Gaussian radial basis function (RBF) as the nonlinear kernel function ϕ , which has γ as its kernel parameter. In order to identify the suitable values for these parameters, we used a grid search and cross-validation method. LIBSVM [18] library provides utilities for this purpose.

C. Performance comparison of algorithms, parameters & horizon

1) *24-h forecast at each substation*: A set of plots (Fig. 5) shows the 24-h forecast results of the ML methods at substation C, D and E using the PHASE-A evaluation. Substation A and B have a similar pattern as substitution C. As can be seen in the figure, the model achieves good forecasting accuracy with the forecast trend lines close to the actual load trend lines. The performance of the model at Substation D is significantly lower than the others. Table IV shows the performance obtained from 24-h ahead forecast at substation A - E using the four methods. The results are obtained using PHASE-B evaluation. The PHASE-B evaluation resulted in 1770(1771) training data and 197(196) test data points for Substations A - D. For Substation E, there are 1016(1017) training data and 113(112) test data points. The best results are shown in bold font, where SVR performed best at four substations.

2) *Effect of internal factors as additional parameters*: To show the effect of internal influencing factors identified in [7], we compare the performance of our 24-h forecast model using four different scenarios as shown in Table V. Since the SVR method has the best result in Table IV, we used SVR method

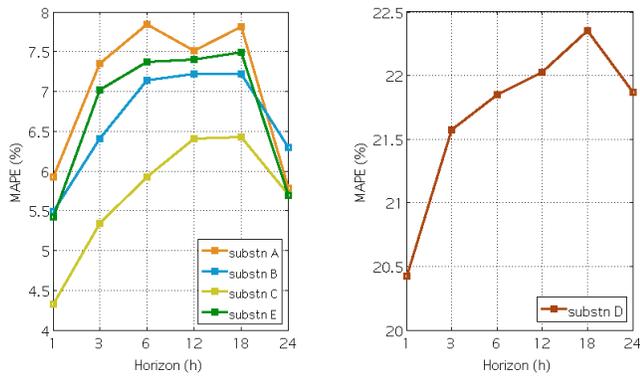


Fig. 6: SVR model performance for varying horizon at substation all substations

for this comparison. In the first scenario, we consider only the external factors as model input, while the second scenario takes both external and internal factors into consideration. The third scenario replaces the HL with internal parameters, ST , FR and DT . The fourth scenario considers OT and hD . Table V shows the obtained result using MAPE metric. Scenario 1 has the best performance and this shows internal factors has little impact on the 24-h forecast horizon.

3) *Varying horizon versus forecast performance*: Following the analysis of different horizon at different substations, which was discussed in Section IV-A, Fig. 6 shows the performance of the forecast model for horizon 1, 3, 6, 12, 18 and 24-h. The results show that MAPE values increase from horizon 1-h to horizon 18-h. The improvement in performance of 24-h forecast reflects the daily cyclic pattern of heat consumption at the substations.

VI. DISCUSSION AND CONCLUSION

The area targeted in this paper aims at providing tools for optimized energy consumption within smart grid environments. In our results, the MAPE for Substation D has a significant margin compared to the other substations. The building served by the substation has notably smaller number of apartments. The residents are people with special needs who are assisted by municipality workers. We hypothesize that this substation requires additional features which describe the load pattern at the substation. We assume that this may affect the result and we shall further investigate this issue in the future. The short period of data available for Substation E did not affect the MAPE. Meteorological parameters such as the global irradiance, humidity and wind speed shall be considered as additional parameters in future work since they are identified in [6] as influencing factors of HL .

In conclusion, we presented a data driven heat load forecasting at DHS substations for multi-family buildings. Four ML methods were used and their performances were compared. The forecasting model was evaluated for horizon values of 1, 3, 6, 12, 18 and 24-h. This paper studied the effects of the internal and external factors for heat load in DHS at the consumer side. SVR shows the best performance in terms of accuracy followed by MLR. Considering MLR's simplicity and linearity method, it confirms a stronger linear relationship between the input parameters and the heat load at the substations. Additional internal factors does not significantly affect the forecast performance of the model. There is an increase in the MAPE with increasing horizon, but this reduces significantly

when horizon is 24-h, showing the effect of daily cyclic pattern in heat demand. In future work, we aim at exploiting the results of this work for building localized energy saving strategies. These shall be used for peak shaving and maintaining a steady return temperature back to the CHP, while granting satisfactory thermal energy at homes.

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] N. Fumo, "A review on the basics of building energy estimation," *Renew. Sustain. Energy Rev.*, vol. 31, pp. 53–60, Mar. 2014.
- [2] 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," in *Syst. Appl. Technol. Conf. (LISAT), 2012 IEEE*, 2012, pp. 1–6.
- [3] T. Catalina, V. Iordache, and B. Caracaleanu, "Multiple regression model for fast prediction of the heating energy demand," *Energy Build.*, vol. 57, pp. 302–312, Feb. 2013.
- [4] M. Sakawa and S. Ushiro, "Cooling load prediction in a district heating and cooling system through simplified robust filter and multi-layered neural network," *Syst., Man and Cybern.*, pp. 995–1000, 1999.
- [5] K. Kato, M. Sakawa, and S. Ushiro, "Heat load prediction through recurrent neural network in district heating and cooling systems," *2008 IEEE Int. Conf. Syst. Man Cybern.*, pp. 1401–1406, Oct. 2008.
- [6] A. Kusiak, M. Li, and Z. Zhang, "A data-driven approach for steam load prediction in buildings," *Appl. Energy*, vol. 87, no. 3, pp. 925–933, Mar. 2010.
- [7] M. Wang and Q. Tian, "Application of wavelet neural network on thermal load forecasting," *Int. J. Wirel. Mob. Comput.*, vol. 6, no. 6, p. 608, 2013.
- [8] H. Gadd and S. Werner, "Daily heat load variations in Swedish district heating systems," *Appl. Energy*, vol. 106, pp. 47–55, Jun. 2013.
- [9] H. A. Nielsen and H. Madsen, "Modelling the heat consumption in district heating systems using a grey-box approach," *Energy Build.*, vol. 38, no. 1, pp. 63–71, Jan. 2006.
- [10] P. Bacher, H. Madsen, H. A. Nielsen, and B. Perers, "Short-term heat load forecasting for single family houses," *Energy Build.*, vol. 65, pp. 101–112, Oct. 2013.
- [11] E. Serban and D. Popescu, "Prediction of domestic warm-water consumption," *WSEAS Trans. Comput.*, vol. 7, no. 12, pp. 2032–2041, Dec. 2008.
- [12] S. Idowu, C. Åhlund, and O. Schelen, "Machine Learning in District Heating System Energy Optimization," *2014 IEEE Int. Conf. Pervasive Comput. Commun. Work Prog.*, p. 4, 2014.
- [13] S. Idowu, C. Åhlund, O. Schelen, and R. Brannstrom, "Machine Learning in Pervasive Computing," Tech. Rep. September, 2013.
- [14] R. Edwards, J. New, and L. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy Build.*, vol. 49, pp. 591–603, Jun. 2012.
- [15] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–99, Jan. 1999.
- [16] L. Breiman, H. F. J. A. O. R., and J. S. C., "*Classification and Regression Trees*". Chapman & Hall, New York, 1984.
- [17] MATLAB, version 8.1.0 (R2013a). Natick, Massachusetts: The MathWorks Inc., 2013.
- [18] C.-C. Chang and C.-J. Lin, "[LIBSVM]: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.