energies

MDPI

Article

Forecasting the Heat Load of Residential Buildings with Heat Metering Based on CEEMDAN-SVR

Xiaoyu Gao [1], Chengying Qi [1], Guixiang Xue [2,*], Jiancai Song [3,*], Yahui Zhang [2,4] and Shi-ang Yu [2,4]

1 School of Energy and Environment Engineering, Hebei University of Technology, Tianjin 300401, China; gaoxiaoyu_hebut@163.com (X.G.); qicy@163.com (C.Q.)
2 School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China; zhangyahuihebuter@163.com (Y.Z.); yushianghebuter@163.com (S.-a.Y.)
3 School of Information and Engineering, Tianjin University of Commerce, Tianjin 300134, China
4 School of Science, Hebei University of Technology, Tianjin 300401, China
* Correspondence: xueguixiang@hebut.edu.cn (G.X.); songjiancai@tjcu.edu.cn (J.S.)

Received: 15 August 2020; Accepted: 17 November 2020; Published: 20 November 2020



Abstract: The energy demand of the district heating system (DHS) occupies an important part in urban energy consumption, which has a great impact on the energy security and environmental protection of a city. With the gradual improvement of people's economic conditions, different groups of people now have different demands for thermal energy for their comfort. Hence, heat metering has emerged as an imperative for billing purposes and sustainable management of energy consumption. Therefore, forecasting the heat load of buildings with heat metering on the demand side is an important management strategy for DHSs to meet end-users' needs and maintain energy-saving regulations and safe operation. However, the non-linear and non-stationary characteristics of buildings' heat load make it difficult to predict consumption patterns accurately, thereby limiting the capacity of the DHS to deliver on its statutory functions satisfactorily. A novel ensemble prediction model is proposed to resolve this problem, which integrates the advantages of Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and support vector regression (SVR), called CEEMDAN-SVR in this paper. The proposed CEEMDAN-SVR algorithm is designed to automatically decompose the intrinsic mode according to the characteristics of heat load data to ensure an accurate representation of heat load patterns on multiple time scales. It will also be useful for developing an accurate prediction model for the buildings' heat load. In formulating the CEEMDAN-SVR model, the heat load data of three different buildings in Xingtai City were acquired during the heating season of 2019–2020 and employed to conduct detailed comparative analysis with modern algorithms, such as extreme tree regression (ETR), forest tree regression (FTR), gradient boosting regression (GBR), support vector regression (SVR, with linear, poly, radial basis function (RBF) kernel), multi-layer perception (MLP) and EMD-SVR. Experimental results reveal that the performance of the proposed CEEMDAN-SVR model is better than the existing modern algorithms and it is, therefore, more suitable for modeling heat load forecasting.

Keywords: heat load prediction; residential buildings; CEEMDAN; SVR

1. Introduction

With the improvement of urban infrastructure construction in China in recent years, the heating energy consumption of residential buildings in winter accounts for a significant proportion of urban energy consumption [1]. This is important for achieving sustainable energy supply, control, and development as well as assuring environmental protection of the city. Meeting the heating

energy demand and regulating it for environmental protection requires two important considerations. On the one hand, extreme weather events, such as haze, occur frequently during the heating period in winter with attendant risks to people's health and the environment [2]. On the other hand, people's inherent pursuit of enhanced living standards and quality of life makes the demand for thermal energy for comfort and pleasure paramount. To balance the contradiction between people's demand for quality heating energy for comfortable living and enforcing energy-saving regulations for sustainable development, the heat metering of residential buildings was carried out to capture energy heating consumption trends using a technology-based forecasting approach. The method relies on economic and technological means to urge people to adopt energy-saving measures as a lifestyle [3,4]. For buildings with heat metering, heat load forecasting is one of the important components of an intelligent district heating system (DHS) and an effective means for the DHS to realize energy-saving regulations and safe operation [5].

In recent years, the challenges of accurate forecasting of buildings' heat load have appealed to multitudinous scholars at home and abroad. Several models and algorithms of building heat load prediction were adopted to address this problem. Scholars have used traditional and non-traditional forecasting technologies in their approaches. For example, Zhang et al. [6] proposed a method to predict the heat load of buildings dynamically based on steady-state white-box models. Buttitta et al. [7] developed a heat load model with a meticulous time scale based on time use survey (TUS) data. Shen et al. [8] analyzed heat load performance based on simulation of neighborhood-scale building. Lim et al. [9] proposed forecasting models for determining heat load with the consideration of the thermal inertia delay and buildings of varying size. Dahl et al. [10] proposed an approach for improving short-term heat load forecasts with calendar and holiday data. Sholahudin et al. [11] investigated the effects of individual input parameters on heat load based on the Taguchi method. Ahmad et al. [12] demonstrated a novel predictive model for heating and cooling load demand based on a data mining approach. The models were designed to enhance the prediction accuracy and the management of heat load. Traditional heat load forecasting methods have insufficient ability to express the inherent complex characteristics, and the prediction accuracy faces difficulties in meeting the actual demand.

However, with the gradual improvement of machine learning algorithms, the field of artificial intelligence has gradually extended to the field of heat load forecasting. In Ref. [13,14], multi-layer perception (MLP) and support vector regression (SVR) models were investigated for forecasting the energy consumption of residential buildings, as well as the artificial neural network (ANN) model [15,16]. Bui et al. [17] examined the performance of a hybrid model based on machine learning and evolutionary algorithm. Le et al. [18] proposed a forecasting algorithm for buildings' heat load based on the integration of artificial neural network and heuristics algorithms. Le et al. [19] estimated the heat load of buildings based on PSO-XGBoost. Kiesel et al. [20] explored the influence of different inputs on the heat load of buildings. Guelpa et al. [21] presented a fast approach to estimate the daily thermal demand of buildings with smart meters. Moayedi et al. [22] employed six machine learning techniques for the problem of heat load energy-efficient buildings. Koschwitz et al. [23] proposed a data-driven thermal load forecasting algorithm based on epsilon-SVM Regression (epsilon-SVM-R). Chou et al. [24] estimated the heat load of buildings based on various data-mining approaches, such as SVR, ANN, and regression tree, etc. Zhao et al. [25] proposed a heat load prediction algorithm for working buildings based on machine learning approaches, such as a support vector machine (SVM), and partial least squares (PLS). Roy et al. [26] explored the predictive algorithm for the heat and cooling loads of residential buildings based on hybrid model of multivariate adaptive regression splines (MARS) and an extreme learning machine (ELM). Liu et al. [27] proposed a multi-layer model to fully utilize the intrinsic characteristic of meteorological and historical data based on auto-regressive models with exogenous inputs (ARX) and PSO neural network (PSO-NN). Bui et al. [28] evaluated the application of two robust evolutionary algorithms, namely the genetic algorithm (GA) and the imperialist competition algorithm (ICA), for heating load prediction. Roy et al. [29] proposed a heat load and cooling load forecasting model based on a deep neural network (DNN). The ensemble of

a support vector machine and a discrete wavelet transform algorithm were proposed to forecast consumers' heat load in district heating systems [30,31].

Although the traditional machine-learning based forecasting methodologies listed above have their merits, the non-linear and non-stationary characteristics of building heat load limit their effectiveness for accurate predictions. The traditional machine learning algorithm is not accurate enough to describe the non-linear and non-stationary characteristics of heat load. In this paper, the accuracy of heat load forecasting was improved based on an integrated learning strategy. The most significant feature of complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is to solve the problem of adaptive basis function, which could decompose the signal into different time scales automatically [32–35]. Compared with the wavelet decomposition algorithm, the CEEMDAN algorithm can effectively solve the problems of wavelet basis function selection and uncertainty principle limitation. It provides a data-driven signal-adaptive decomposition method. To achieve higher accuracy of heat load forecasting, the heat load forecasting model based on the integrated learning algorithm of CEEMDAN and SVR (CEEMDAN-SVR) was investigated in the present work systematically.

The organization of this paper is depicted below. Section 2 makes an introduction for the working principle of heat metering based on the on-off time-area method (OOTAM); Section 3 represents the methodology based on proposed hybrid CEEMDAN-SVR; Section 4 discusses the comprehensive experiments and prediction results; and the overall conclusion is summarized in Section 5.

2. The Principle of Heat Metering Based on an On-Off Control Strategy in DHS

In order to promote people's awareness of the need for energy conservation in daily heat consumption behavior, we designed a heat allocation measurement approach called on-off time-area method (OOTAM) for multi-story apartment buildings. The main idea of the OOTAM was to install heat meters in each building as the cost settlement point; the valves and indoor temperature controller were installed within each household, and the primary factors, such as accumulated on-off time, room area, and indoor temperature of all users in this building, are calculated by the building processor, which allocates the heat appropriately. The system schematic of the OOTAM system is shown in Figure 1.

For the household-level connected heating system (as shown in Figure 1), an on-off controller was installed on the branch of each household to adjust the circulating water of the radiator so as to achieve effective room temperature control. At the same time, a wireless indoor temperature controller was placed in each household for automatic measurement of indoor temperature and target setting. The two parameters and their different values were wirelessly sent to the controller to optimize the on/off ratio of the electric valve in a control cycle. This method can effectively alleviate actual problems, such as uneven indoor temperature distribution, vertical imbalance, and heating water quality in the heating environment. The advantage of the OOTAM is that it can reasonably allocate heat according to the thermal energy need for the comfort of each household, which helps to improve the satisfaction of thermal users and achieve the purpose of saving or conserving heating energy.

Inappropriate design of heating radiators or hydraulic imbalances in existing buildings can lead to inefficient allocation of the heat load within the buildings. Therefore, other effective, technical means must be used to correct the algorithm. Through a large number of empirical studies, our team found that appropriate secondary return water temperature correction was an effective method to improve the on-off time-area method for heat allocation measurement. Compared to other normal heat users, the difference between supply and return temperature of heat users with a larger radiator area is large. The opening time ratio of such users should be multiplied by a correction factor greater than 1, while or heat users with small flow due to hydraulic imbalance or system blockage, the opening time ratio of such users should be multiplied by a correction factor less than 1.

The heat load on the demand side of residential buildings with heat metering transformation is also a non-linear and large time-delay physical variable, which is affected by many factors, such as supply temperature, return temperature, indoor temperature, outdoor temperature, building insulation

materials, and residents' heating comfort, etc. Therefore, it is a complex and difficult task to accurately predict the demand side heat load.

The calculation method of heat allocation based on the OOTAM is shown in the Equations (1) and (2):

$$Q_i = R_i \cdot Q \cdot \frac{\tau_i \cdot S_i}{\sum_{j=1}^{n} \tau_j \cdot S_j} \tag{1}$$

$$\tau_i = \frac{\Delta\tau_i'}{\Delta\tau} \tag{2}$$

where Q_i represents the allocated heat in this heat allocation period for the ith household (kWh); Q represents the total heat consumption in this allocation period (kWh); R_i represents the correction coefficient for secondary return temperature; S_i represents the building area (m^2); τ_i represents the proportion of cumulative opening time of electronic valves in the ith household; $\Delta\tau_i'$ represents the cumulative opening time of the electronic valve of the ith household; $\Delta\tau$ represents the total allocation period; n represents the total number of households in this building.

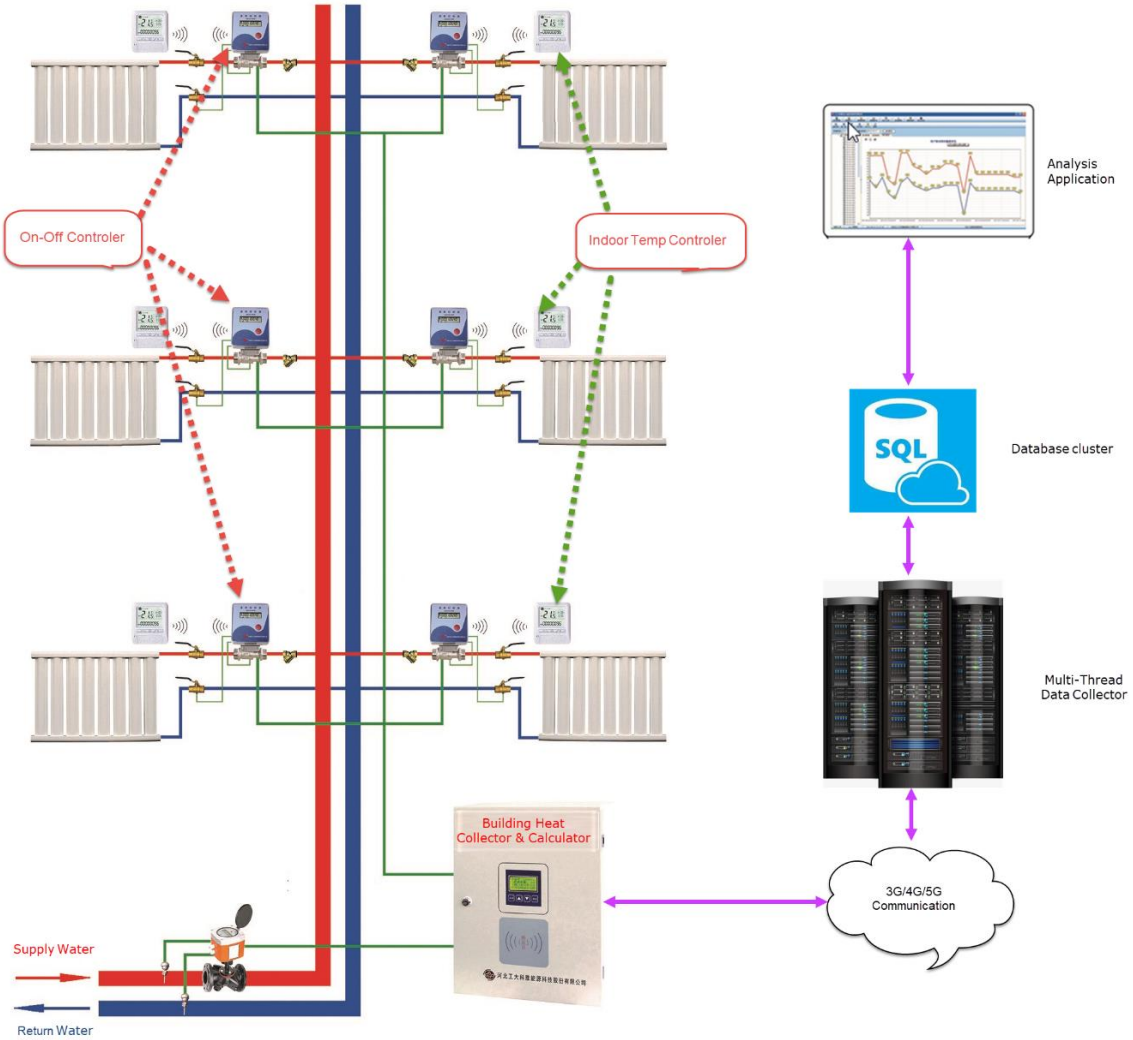


Figure 1. The system schematic diagram of the on-off time-area method (OOTAM).

3. Methodology and Analysis

3.1. The Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

The empirical mode decomposition (EMD) proposed by N. E. Huang is suitable to process the signal in the time-frequency domain adaptively [36]. In theory, the method can automatically select the basis function and decompose various types of signals according to its time scale, which is especially suitable for processing non-stationary and non-linear data. However, the EMD exhibits some defects, such as mode aliasing and nonorthogonality. Huang et al. [37] proposed ensemble empirical mode decomposition (EEMD), which introduced Gaussian white noise to the original signal for the process of decomposition. However, the Eigen mode functions obtained by decomposition varies with the change of artificial white noise, which makes the EEMD method unstable, and it is difficult to eliminate the reconstruction error caused by Gaussian white noise. To solve the above problems, Torres et al. [38] proposed an improved CEEMDAN further, which can achieve better separation of Eigen mode functions, accurately reconstructing the original signal.

Heat load with complex non-linear and non-stationary characteristics is not an intrinsic mode function, but the data may contain multiple fluctuation modes at any time point, and the frequency characteristics of heat load cannot be fully characterized simply by the Hilbert transform. We need to decompose the heat load by CEEMDAN to obtain its intrinsic mode function and then carry out the Hilbert transform on each intrinsic mode function to finally obtain the Hilbert spectrum of the heat load.

3.2. Support Vector Regression(SVR)

The support vector machine (SVM) is essentially a data-driven linear classifier. Its goal is to achieve a reasonable segmentation of samples using the maximum margin hyperplane. The SVR can be obtained by extending SVM from classification problems to a regression problem.

Given the training sample $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}(y_i \in R)$, we hope to obtain a regression model *f(x)* shown in Equation (3), which is approximately close to the target y, where W and B are the model parameters to be determined.

$$f(x) = W^T x + B \tag{3}$$

When the fitting function *f(x)* approximates y with the maximum error ϵ, the SVR can be rewritten as shown in Formula (4).

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} l_\varepsilon(f(x_i), y_i) \tag{4}$$

where C represents the regularization constant term and l_e represents the insensitive loss function:

$$l_\varepsilon(Z) = \begin{cases} 0 & if|Z| \le \varepsilon \\ |z| - \varepsilon & otherwise \end{cases} \tag{5}$$

After introducing the relaxation variables and the Lagrangian multipliers and then zeroing the partial derivative of the loss function, we can obtain the following result:

$$w = \sum_{i=1}^{m} (\hat{a}_i - a_i)x_i \tag{6}$$

The solution for SVR can be obtained as follows:

$$f(x) = \sum_{i=1}^{m} (\hat{a}_i - a_i)x_i^T x + b \tag{7}$$

The sample that can make $(\hat{a}_i - a_i) \neq 0$ is the support vector of SVR, which must fall outside the ε-septum. If we consider the following feature-mapping function $\phi(x)$,

$$w = \sum_{i=1}^{m} (\hat{a}_i - a_i)\phi(x) \tag{8}$$

The solution for SVR can be expressed as follows:

$$f(x) = \sum_{i=1}^{m} (\hat{a}_i - a_i)k\left(x_i^T x_i\right) + b \tag{9}$$

where $k\left(x_i^T x_i\right) = \phi(x_i)^T \phi(x_i)$ is the kernel function.

3.3. The Heat Load Forecasting Model Based on CEEMDAN-SVR

Heat load can be affected by many factors in a DHS, which is a random variable with non-linear, non-stationary, and time-delay characteristics. Hence, we proposed the heat load forecasting algorithm based on CEEMDAN-SVR to solve this problem by obtaining a high-precision prediction algorithm. It is based on the principle of "divide and rule" to predict the heat load for residential buildings. The CEEMDAN method is used to decompose a building's heat load time series, which is difficult to fit into several subsequences with obvious periodicity and smoothness, which avoids the difficulty of using the SVR model to predict non-linear and unstable data directly. Each subsequence is trained and fitted by different SVR models, which makes full use of the advantages of the SVR machine learning algorithm, such as strong learning ability and good adaptability. Compared with the traditional method and the SVR method, the proposed ensemble forecasting method has better performance in the short-term prediction of building heat load time series.

Firstly, we decomposed the heat load into multiple intrinsic mode functions (IMFs) and residual vectors based on the CEEMDAN algorithm so that the data characteristics of each IMF could be simplified. Subsequently, different prediction components were summed to obtain the overall, more accurate heat load forecasting results. Through the analytical characterization analysis of the heat load data, we found that it could be decomposed into the sum of five IMFs and residuals as shown in Equation (10).

$$imf_1, imf_2, imf_3, imf_4, imf_5, resi = CEEMDAN(Q) \tag{10}$$

where imf_i (i = 1, … 5) represents the ith intrinsic mode function; *resi* represents the residual error of the empirical mode decomposition; Q represents the heat load of the residual buildings; *CEEMDAN* represents the function of the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise.

The CEEMDAN decomposition results were input into the respective SVR models, and the results are as follows:

$$\begin{cases} \widehat{Q}_1 = SVR_1(imf_1) \\ \widehat{Q}_2 = SVR_2(imf_2) \\ \widehat{Q}_3 = SVR_3(imf_3) \\ \widehat{Q}_4 = SVR_4(imf_4) \\ \widehat{Q}_5 = SVR_5(imf_5) \\ \widehat{Q}_{resi} = SVR_6(resi) \end{cases} \tag{11}$$

where $\widehat{Q}_i (i = 1, \ldots 5)$ represents the prediction result of each IMF based on each SVR respectively; $\widehat{Q}_{resi}$ represents the prediction result of residual error based on SVR; SVR_i (i = 1, … 6) represents the prediction model based on each SVR respectively.

Thus, the total prediction result $\widehat{Q}$ of the heat load is as follows:

$$\widehat{Q} = \widehat{Q}_1 + \widehat{Q}_2 + \widehat{Q}_3 + \widehat{Q}_4 + \widehat{Q}_5 + \widehat{Q}_{resi} \quad (12)$$

Different kernel functions are suitable for processing data with different characteristics. The SVM with a linear kernel function for the IMF has a low-frequency component dominant and the SVM with a radial basis function (RBF) kernel function for IMFs has a high-frequency component.

The grid search method and random search method were used to perform the parameter optimization of the SVM model, such as penalty coefficients C and gamma. A higher value of coefficient C is easy to fall into overfitting; on the contrary, a smaller value of coefficient C corresponds to underfitting. An appropriate value for the coefficient C makes the algorithm have good generalization ability. However, the gamma coefficient determines the number of support vectors, which affects the convergence speed. The system schematic diagram is shown in Figure 2.

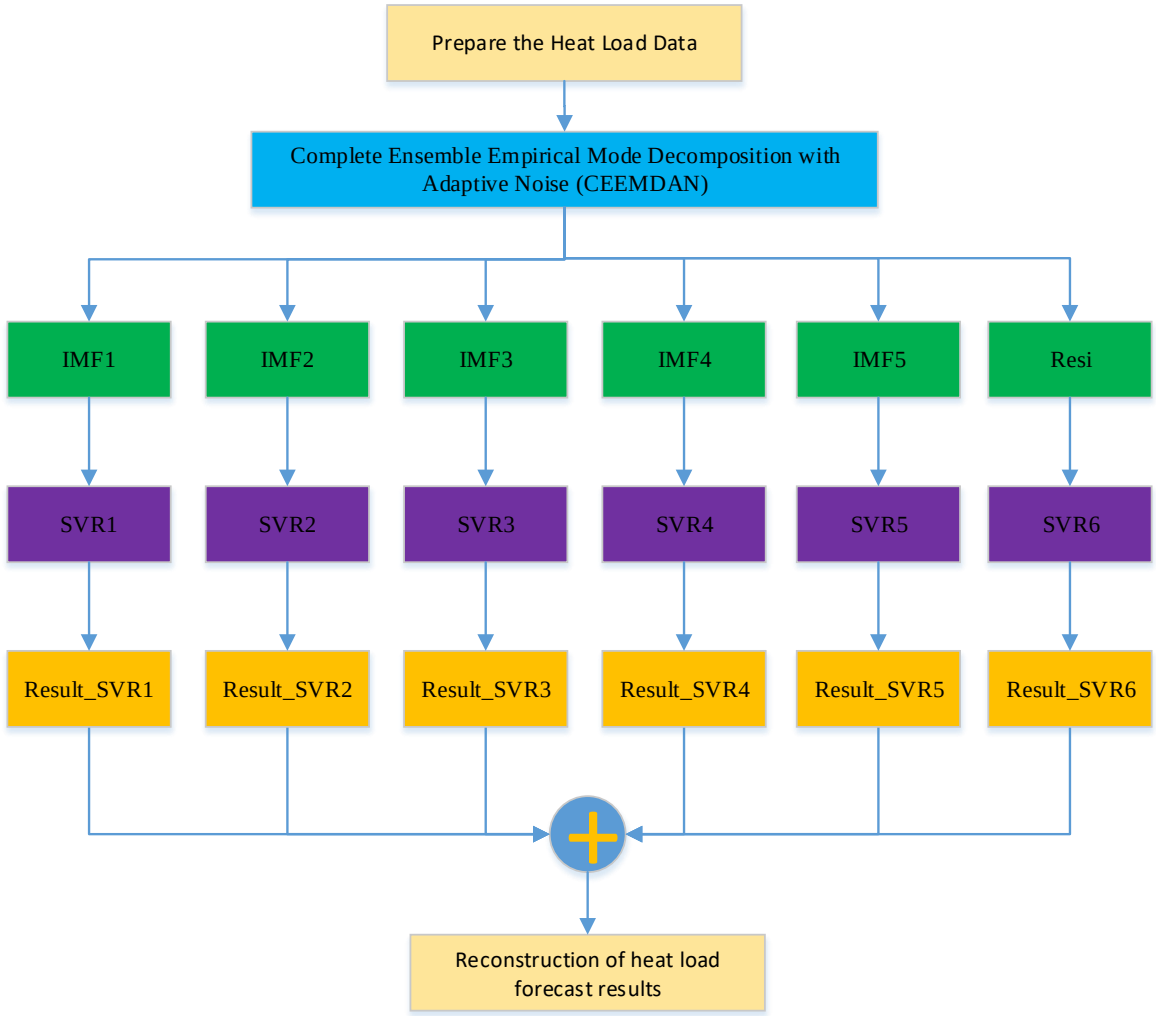


Figure 2. The system schematic of the heat load forecasting approach based on complete ensemble empirical mode decomposition with adaptive noise—support vector regression (CEEMDAN-SVR).

3.4. Evaluation Criteria

To compare the advantages and disadvantages of the algorithm more accurately, classical evaluation indices are used in this paper, such as root mean squared error(RMSE),

mean absolute error(MAE), mean squared error(MSE), and mean absolute percentage error(MAPE), in Formula (13).

$$\begin{cases} RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Q}_i - Q_i\right)^2} \\ MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{Q}_i - Q_i\right| \\ MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{Q}_i - Q_i\right)^2 \\ MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{\widehat{Q}_i - Q_i}{\widehat{Q}_i}\right| \end{cases} \quad (13)$$

4. Experiments and Discussion

4.1. Data Description

In order to test whether our algorithm has a better prediction ability for real engineering scenarios, we selected the historical heat load monitoring data of three buildings in Xingtai City in northern China in the heating season of 2019–2020. In 2011–2012, these buildings were retrofitted with the OOTAM. According to the construction specifications, ultrasonic heat meters and temperature sensors were installed at each building entrance, and a building calculator designed by our team for heat sharing and data acquisition was also installed at the building entrance. The calculator was connected with the data center through a wireless model, and all data were collected by polling regularly (once/4 h). This smart district heating system monitors and controls the heating energy consumption in thousands of buildings. Considering the building thermal inertia, control frequency, and communication cost, we set the allocation frequency of building heat consumption based on the OOTAM algorithm as once a day. It can improve the thermal comfort of building residents after regulation and reduce the communication complexity of the system, as well as improve the reliability and economy of the system. There are 120 days for formal central heating in the heating season in Xingtai City, from November 15 to March 15 of the next year. To analyze the generalization ability of the model, we selected three typical heating buildings for comparison. Each of the three sample data sets contains 120 sample data.

In the heat load forecasting of buildings with the OOTAM heat metering, the related influencing factors included supply temperature, return temperature, and outdoor temperature. To obtain the characteristic distribution of the heat load of residential buildings and its related factors, we selected the original data of one of the buildings for statistical analysis, and the results are shown in Figure 3. The dynamic curves of heat load and its three related factors are shown in Figure 3. It is evident from Figure 3 that these three parameters, including heat load, supply temperature, and return temperature, have strong non-linear and non-stationary characteristics, and the data may contain noise due to abnormal sensors and other reasons. The heat load forecasting algorithm needs to be combined with weather forecast information. Outdoor temperature is one of the important factors affecting heat load. In order to ensure the thermal user comfort and energy saving, an appropriate control strategy should be adopted to make the heat load change dynamically with changes in outdoor temperature. The average outdoor temperature data of weather forecast in this paper are from the meteorological service website of China Meteorological Administration (www.weather.com.cn), which also present complex non-stationary characteristics. We can see the daily weather forecast data for the next 7 days. The statistics of the heat load and related factors of the residential building are shown in Table 1.

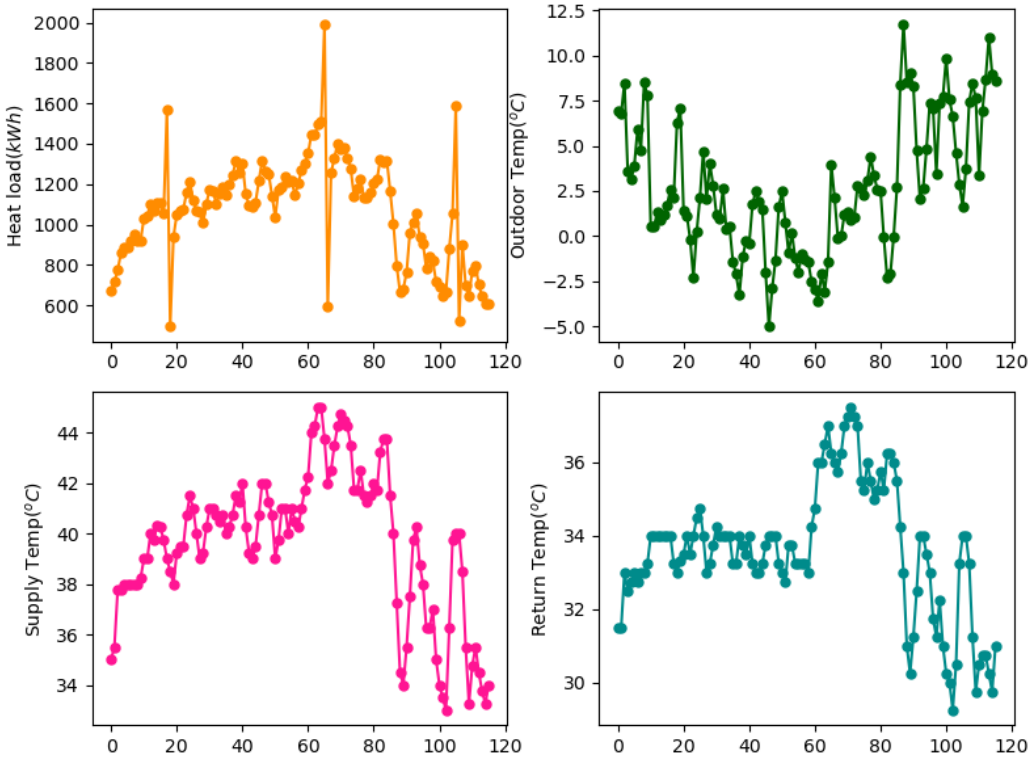


Figure 3. The historical curves of building heat load and related factors of the first residential building.

Table 1. The statistics of the heat load and related factors of the first residential building.

Parameters	Max	Min	Mean	Median	Std
Heat load (kWh)	1994	495	1063.59	1104	259.52
Outdoor temp (°C)	11.75	−5.00	2.68	2.15	3.66
Supply temp (°C)	45.00	33.00	39.58	40.00	2.9
Return temp (°C)	37.50	29.25	33.58	33.75	1.82

Scatter diagrams of the comparative analysis between building heat load and its three related factors are shown in Figure 4. It is obvious that the scatter diagram distributions of heat load and these three factors are relatively loose, which means that there is a strong non-linear characteristic between them. Therefore, the CEEMDAN-SVR-based algorithm proposed in this paper has significant advantages to deal with the heat load forecasting problems with non-linear and non-stationary features.

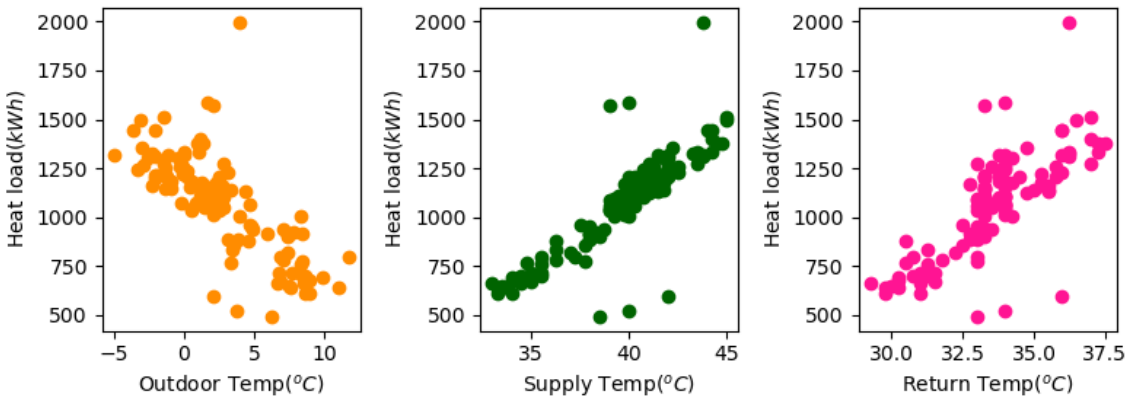


Figure 4. Scatter diagrams of building heat load vs. its related factors of the first residential building.

We decomposed the historical heat load data of building #1 by CEEMDAN and obtained five different intrinsic modal functions and residuals, as shown in Figure 5. It is evident from Figure 5 that the five IMFs have different time-domain characteristics, which reflect the internal characteristics of heat load in different time scales. The final residual term reflects the error after the CEEMDAN decomposition process. The frequency-domain characteristics of the corresponding five IMFs are shown in Figure 6. It is obvious from Figure 6 that IMF1, IMF2, and IMF3 belong to high-frequency components, while IMF4 and IMF5 are low-frequency components. Different frequency characteristics in the frequency domain correspond to the non-linear degree in the time domain. The non-linearity of the high-frequency component is higher than that of the low-frequency component. Therefore, we chose SVR models with different kernel functions for different IMFs, and the hyperparameters of each model are shown in Table 2.

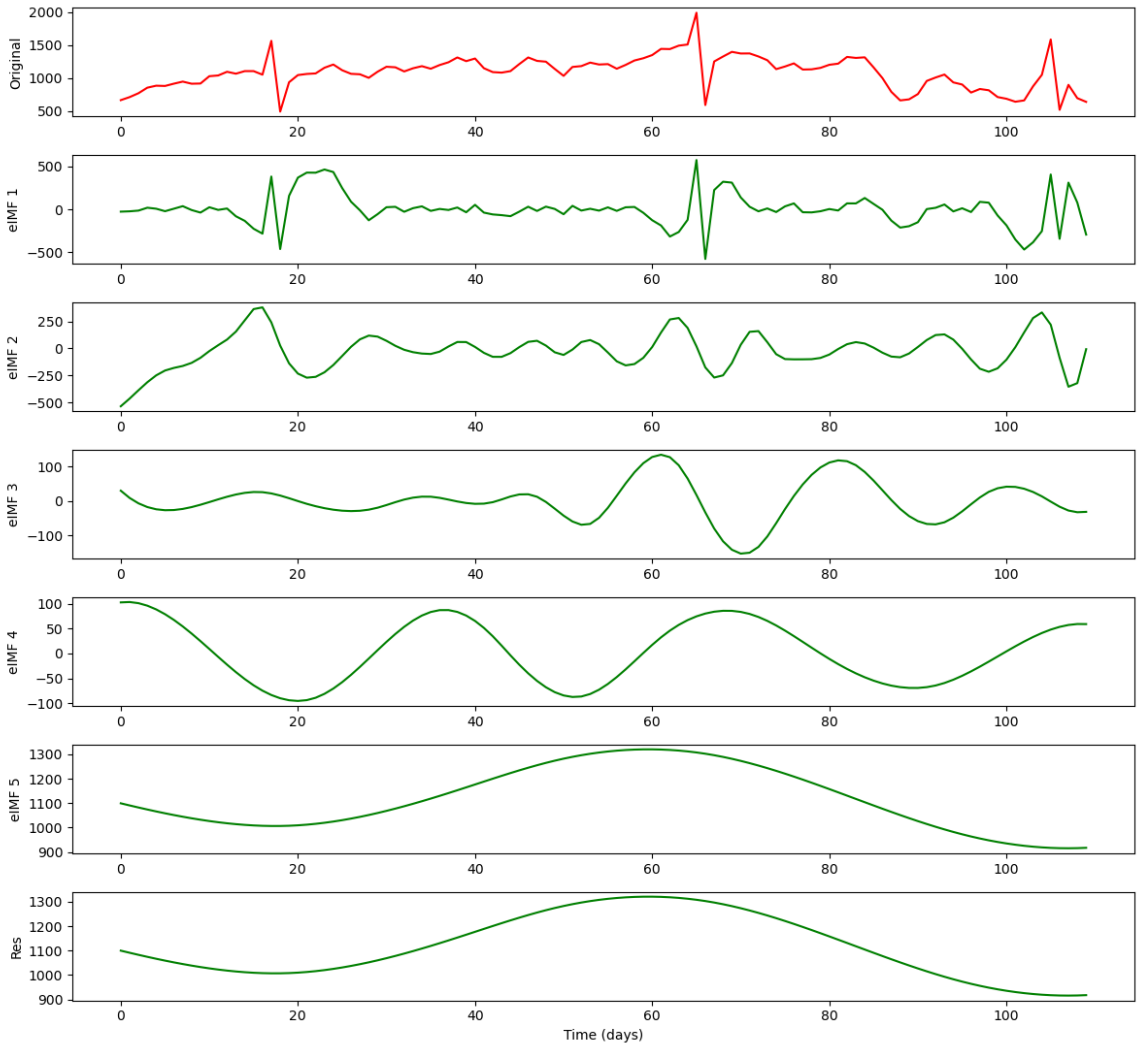


Figure 5. The curves of each component (IMF1–IMF5) obtained from CEEMDAN decomposition of the heat load of residential building #1.

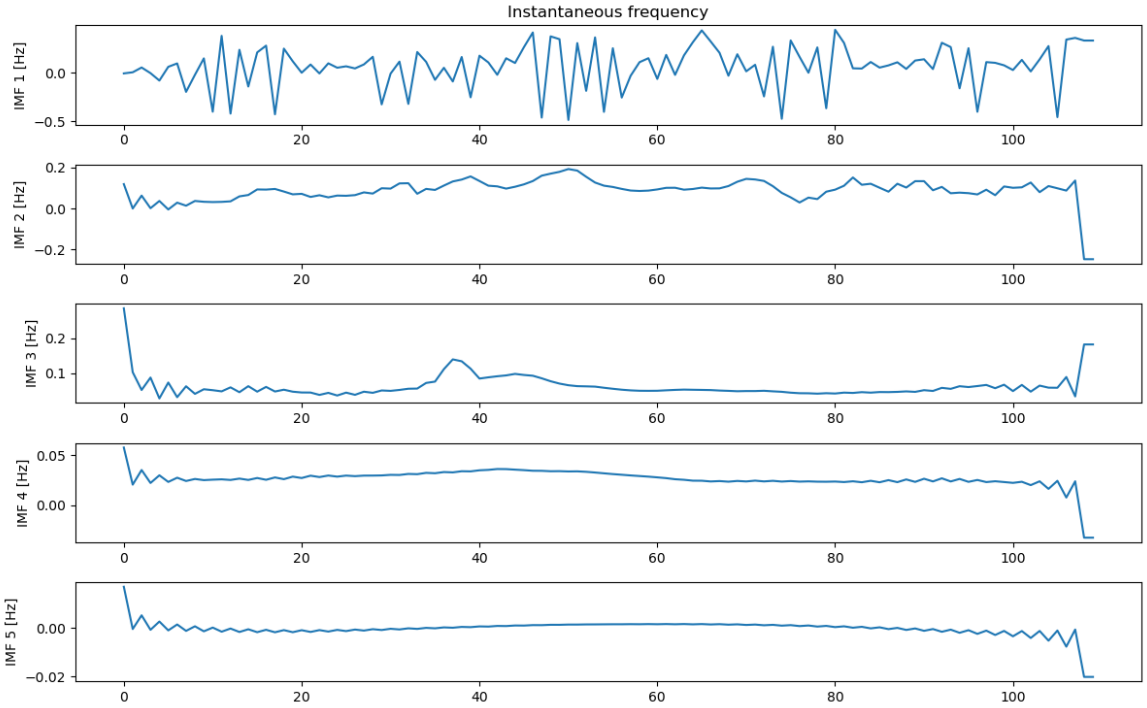


Figure 6. The frequency-domain analysis of each component (IMF1–IMF5) obtained from CEEMDAN decomposition of the heat load of residential building #1.

Table 2. The hyperparameters selection of each decomposition component.

Component	SVR Kernel Function	Frequency Component	Parameters
IMF1	RBF	High	C = 1; gamma = 0.060
IMF2	RBF	High	C = 10; gamma = 0.060
IMF3	RBF	High	C = 10; gamma = 0.060
IMF4	Poly	Low	C = 10; gamma = 0.060
IMF5	Poly	Low	C = 10; gamma = 0.060
Residual	Poly	Low	C = 10; gamma = 0.060

In order to test the prediction accuracy of the proposed ensemble algorithm based on CEEMDAN-SVR and to avoid overfitting problems, the n-fold verification method was used to predict the heat load in the next 7 days in this paper. Due to the strict time sequence characteristics of heat load history datasets, we use the three-fold experimental method to predict the heat load. As shown in Figure 3, a complete heating season roughly includes three stages: the early cold period, the high cold period, and the last cold period. Because the outdoor temperature characteristics of the early and late cold periods are similar, the training samples should generally include part of the data of the early cold period and of the high cold period. The dataset division used in the training model and prediction is shown in Figure 7. The final experimental evaluation criteria are obtained by averaging the three prediction results.

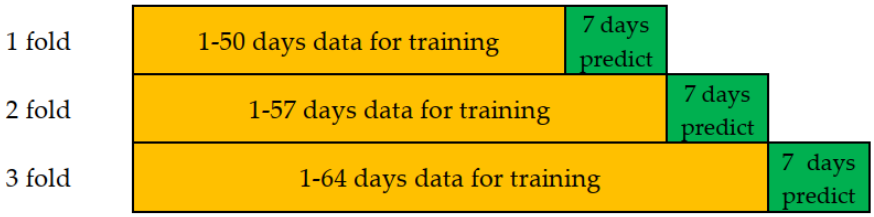


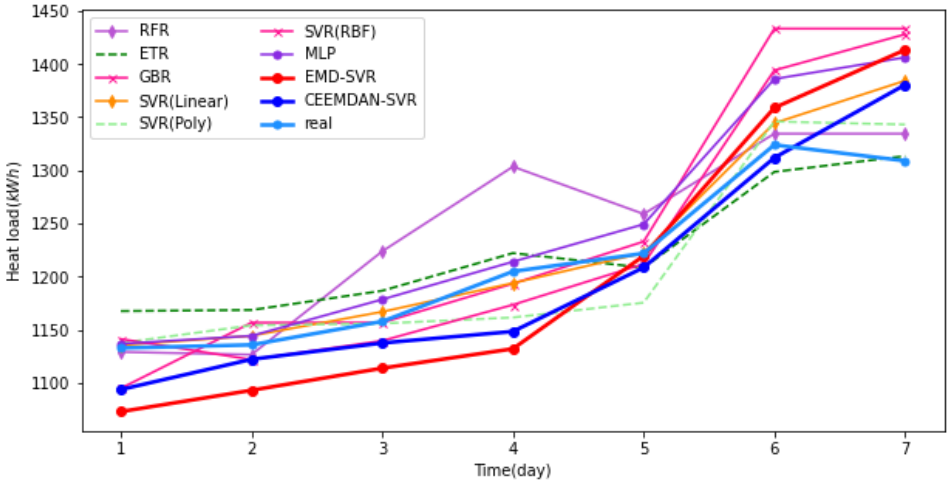
Figure 7. The split of data for training and prediction based on ensemble CEEMDAN-SVR.

4.2. Performance Evaluation

To obtain the performance evaluation of the heat load prediction algorithm based on CEEMDAN-SVR, the heat load data of three groups of different buildings in Xingtai City, collected during the heating season of 2019–2020, were used for detailed comparative experiments. The prediction performance comparisons were conducted with several other algorithms, such as the original SVR, MLP, extreme tree regression ETR, RFR, GBR, and EMD-SVR algorithms. The main parameters of the comparison algorithms are shown in Table 3. According to the data division method shown in Figure 7, this paper constructs three different training models for the historical heat load data of each building, and each model predicts the heat load change of the building in the next 7 days. The predicted 7-day heat load curves of the three buildings are shown in Figures 8–10. These results reveal that the proposed CEEMDAN-SVR algorithm outperformed other algorithms.

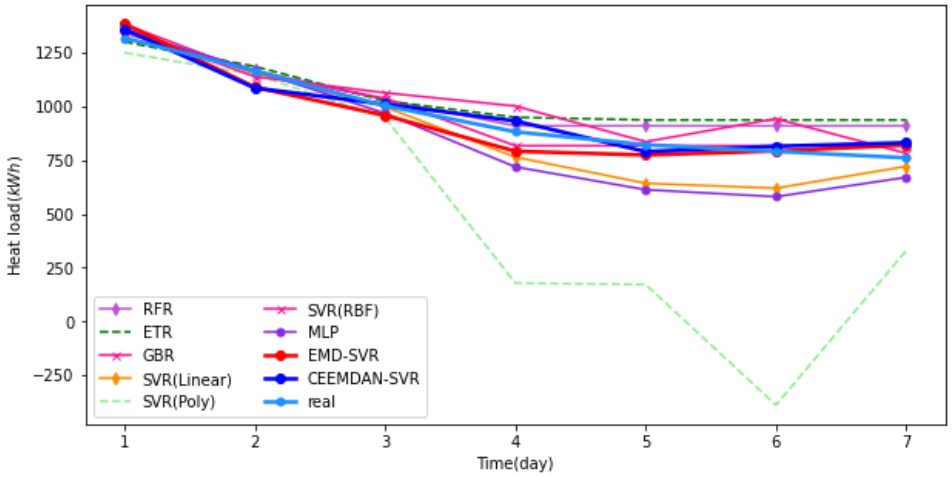
Table 3. The main parameters of comparative experiments.

Model.	Parameters	Value	Description
RFR/ETR/GBR	n_estimators	60	Number of subtrees
	max_depth	13	The maximum depth of the tree
	min_samples_split	10	Minimum number of samples needed to split internal nodes
	min_samples_leaf	10	Minimum sample leaf size
SVR(Linear)	kernel	linear	Kernel type
	gamma	scale	Kernel coefficient
SVR(Poly)	kernel	poly	Kernel type
	gamma	scale	Kernel coefficient
SVR(RBF)	kernel	rbf	Kernel type
	gamma	scale	Kernel coefficient
MLP	hidden_layer_sizes	100	Hidden layer sizes
	activation	relu	Activation function
EMD-SVR	MAX_ITERATION	1000	Maximum number of iterations
	C	10	Penalty coefficient
	gamma	0.060	Gamma

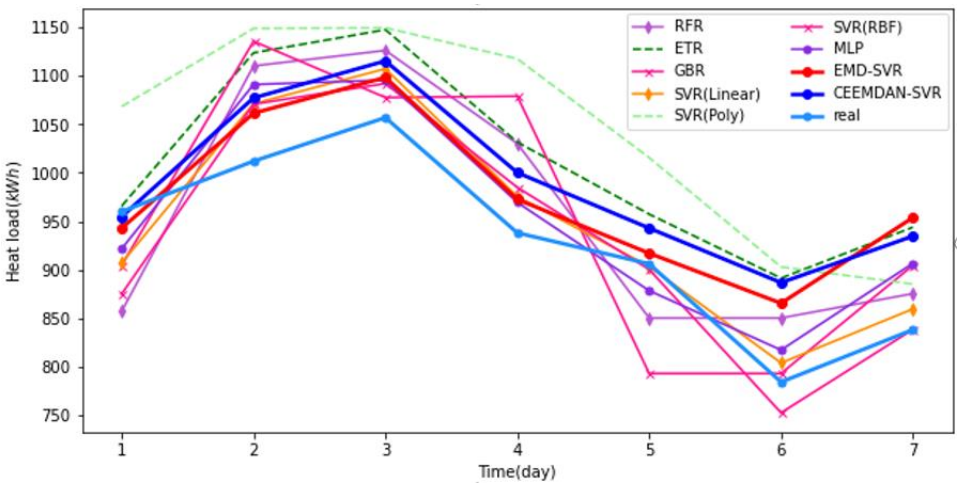


(**a**) 1-fold

Figure 8. *Cont.*

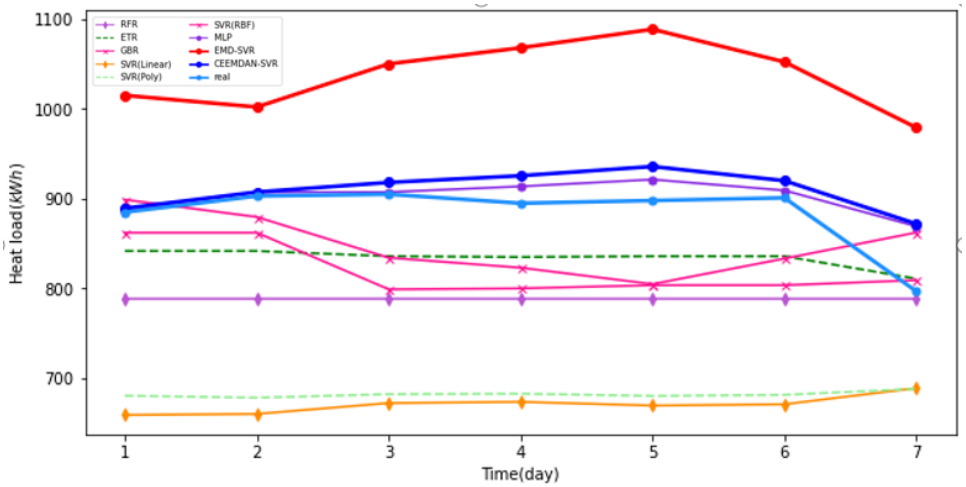


(**b**) 2-fold



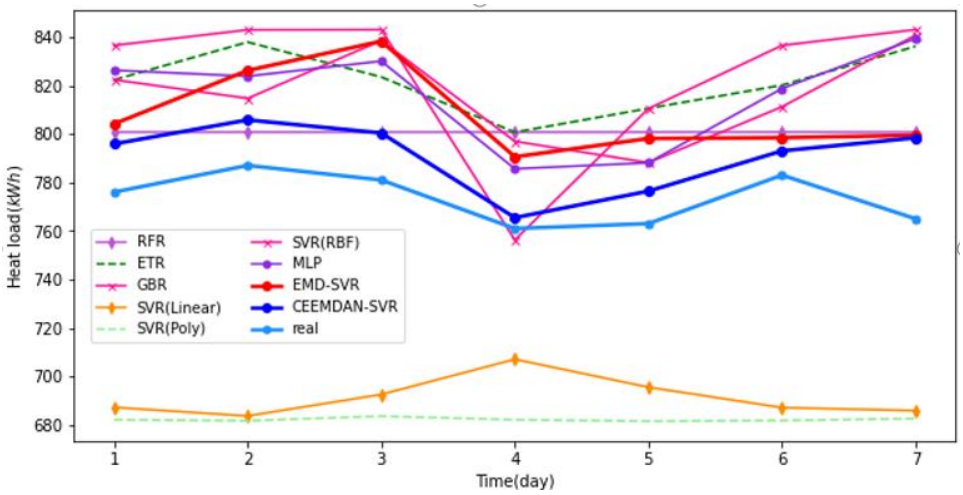
(**c**) 3-fold

Figure 8. Comparison of the heat load forecasting results of various forecasting algorithms for residential building #1 under the different data split modes.

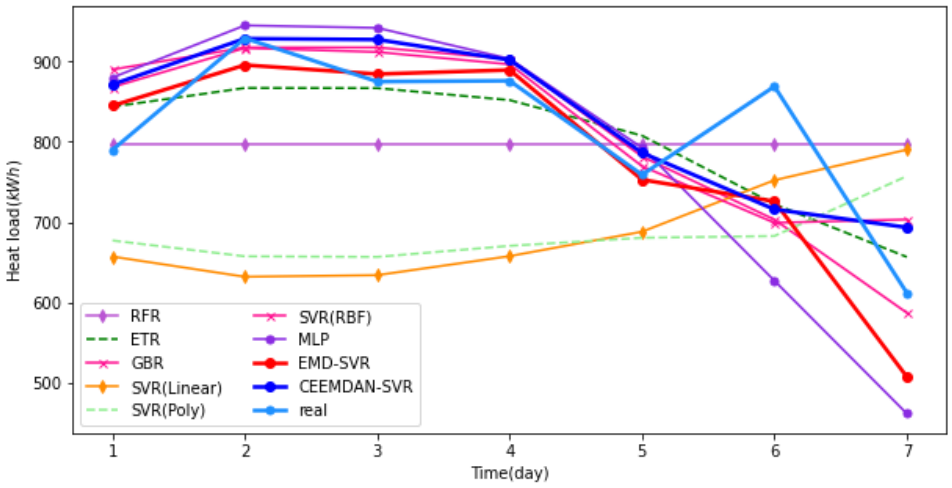


(**a**) 1-fold

Figure 9. *Cont.*

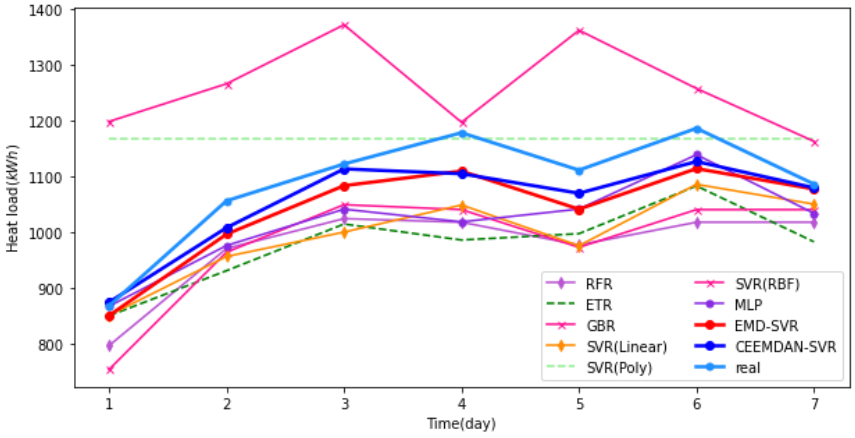


(**b**) 2-fold



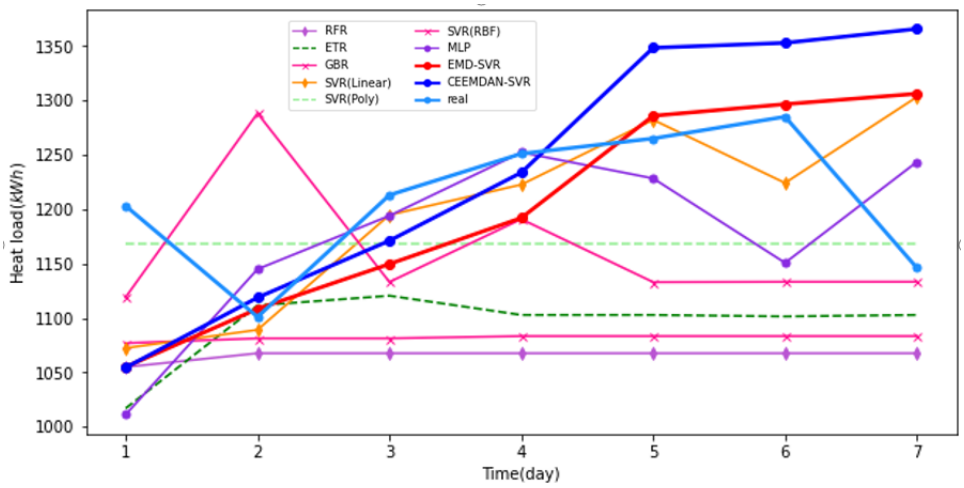
(**c**) 3-fold

Figure 9. Comparison of the heat load forecasting results of various forecasting algorithms for residential building #2 under the different data split modes.

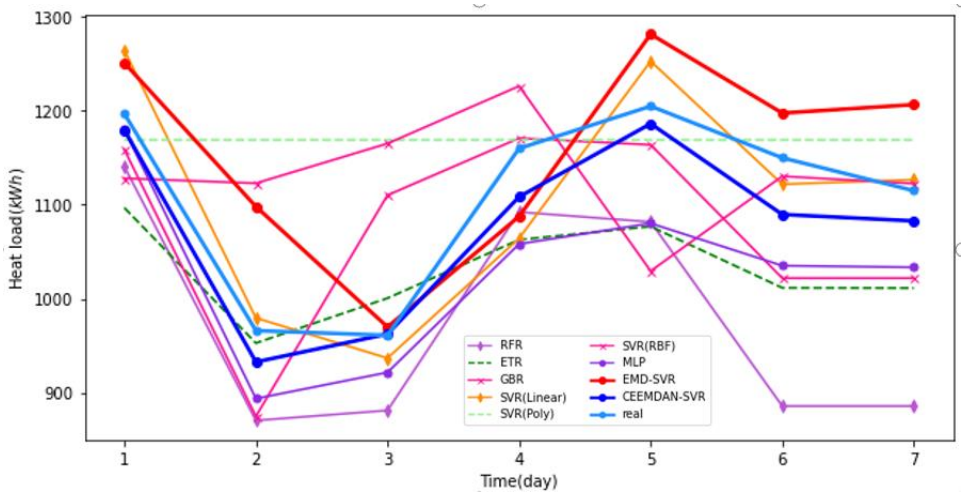


(**a**) 1-fold

Figure 10. *Cont*.



(**b**) 2-fold



(**c**) 3-fold

Figure 10. Comparison of the heat load forecasting results of various forecasting algorithms for residential building #3 under the different data split modes.

The performance results of these three buildings under different prediction algorithms are shown in Tables 4–6. It is evident from the heat load data, when compared with the original SVR algorithm with different kernel functions, that the algorithm can provide higher prediction accuracy than the original SVR algorithm. Therefore, the proposed CEEMDAN-SVR forecasting algorithm has good adaptability and prediction accuracy. It can accurately describe the non-linear and non-stationary characteristics of heating load, thereby making it more suitable for solving the heat load problem of buildings.

The prediction accuracy of MAPE based on the proposed CEEMDAN-SVR algorithm can reach 1.8%, which is significantly higher than other algorithms, where the RMSE, MAE, and MSE errors were relatively of lower magnitudes. It is evident from Table 4 that the prediction accuracy of MAPE based on the proposed CEEMDAN-SVR algorithm reaches 2.3%, which is still higher than other algorithms, comparatively. This is because the dynamic change of the heat load of building #2 is larger and the prediction is more difficult.

Table 4. The performance comparison of the various algorithms for residential building #1.

Algorithm	RMSE	MAE	MSE	MAPE
RFR	114.44	35.77	2326	7.8
ETR	92.07	22.42	608	6.2
GBR	148.94	45.14	4215	9.7
SVR(Linear)	117.73	18.19	910	7.6
SVR(Poly)	112.11	24.39	861	7.1
SVR(RBF)	139.61	38.70	2962	9.0
MLP	129.16	27.70	1741	8.4
EMD-SVR	59.42	51.52	3531	4.2
CEEMDAN-SVR	**29.10**	**22.63**	**847**	**1.8**

Table 5. The performance comparison of the various algorithms for residential building #2.

Algorithm	RMSE	MAE	MSE	MAPE
RFR	75.95	66.57	5240	8.1
ETR	54.74	52.16	2848	6.8
GBR	63.20	56.01	3708	7.6
SVR(Linear)	83.35	82.43	7043	10.6
SVR(Poly)	92.03	91.50	8468	11.8
SVR(RBF)	51.32	46.64	2466	6.1
MLP	43.68	37.18	1710	5.0
EMD-SVR	39.56	38.44	1565	5.0
CEEMDAN-SVR	**20.40**	**17.38**	**416**	**2.3**

Table 6. The performance comparison of the various algorithms for residential building #3.

Algorithm	RMSE	MAE	MSE	MAPE
RFR	170.98	110.07	18161	12.9
ETR	131.34	88.63	9648	10.3
GBR	142.63	78.84	8421	10.3
SVR(Linear)	149.99	41.28	2535	11.4
SVR(Poly)	113.06	79.35	12783	7.9
SVR(RBF)	112.45	99.86	15247	8.5
MLP	160.71	87.44	9855	12.3
EMD-SVR	77.69	69.07	6036	6.3
CEEMDAN-SVR	**36.96**	**31.82**	**1366**	**2.8**

Similarly, it is obvious from Table 5 that the prediction accuracy index of MAPE based on the proposed CEEMDAN-SVR algorithm reaches 2.8%, and the other error indices, RMSE, MAE, and MSE, are also better than other algorithms. Furthermore, the combined CEEMDAN-SVR algorithm can better model internal characteristics in different time scales through intrinsic mode decomposition of the heat load data when compared with the original SVR algorithm with different kernel functions. The algorithm can provide higher prediction accuracy than the original SVR. Therefore, the proposed CEEMDAN-SVR forecasting algorithm has good adaptability and prediction accuracy. It can accurately describe the non-linear and non-stationary characteristics of heating load, thereby making it more suitable for solving the heat load problem of buildings.

The heat load demand in the next 7 days can be obtained by the prediction model, which provides the basis for the formulation of a short-term control strategy for the smart district heating system. The first step of the intelligent control algorithm of the actual district heating system is to train the prediction model according to the historical data of the previous period and then to generate the optimal control strategy according to the predicted heat load. The machine learning prediction model based on CEEMDAN-SVR can obtain a more accurate prediction model according to the small sample dataset, and the prediction length is 7 days each time. Therefore, the training samples in this round

select the historical data of the beginning 50 days of the heating period. The length of training samples is accumulated with the increase in samples.

In our experiments, the super parameters of the prediction model from three different buildings were the same and the data set segmentation method was also completely consistent. We obtained different prediction models driven through the input data of different buildings. For different buildings under the same DHS, although the maintenance structure, insulation performance, and construction year are different, we can use a unified model structure and super parameters to obtain the accurate heat load prediction model of various buildings.

5. Conclusions

The goals of DHS intellectualization are to realize on-demand heating and balanced heating and minimize energy consumption, operation cost, and pollution emission under the premise of meeting users' thermal comfort demands. The accurate heat load prediction algorithm of buildings on the demand side is essential for the regulation and control of heat source and heat exchange stations on the supply side. However, due to different building envelopes, thermal insulation performances, and different thermal comfort requirements of residents, the buildings' heat load itself has significant non-linear and non-stationary characteristics, which are a big challenge for the accurate heat load prediction algorithm. In order to deal with this problem, this paper proposed a forecasting algorithm based on the CEEMDAN-SVR combination strategy. The algorithm can demonstrate the internal characteristics of heat load in different time scales by employing a modal decomposition process. Different SVR models are then used to calculate and predict the obtained intrinsic modal functions. Finally, a more accurate prediction model is obtained. In this paper, we compared the performance between the proposed CEEMDAN-SVR algorithm and state-of-the-art algorithms, such as forest tree regression (FTR), ETR, SVR (linear, poly, RBF kernel), MLP, and EMD-SVR, etc. The experimental results reveal that the proposed CEEMDAN-SVR model improved the prediction accuracy and it can more accurately model the complex characteristics of heat load, such as non-linear and non-stationary ones. In the practical application of the model, we will use the historical data of the previous heating season to do the pre-training of the model, and use the data of the new heating season to update and optimize the model to carry out short-term prediction of the future heat load data of the current heating season, so as to improve the efficiency of the model.

In the future, it is necessary to quantitatively analyze the thermal comfort demand of different households from the demand side and build a more accurate building heat load prediction model based on the spatial and temporal distribution characteristics of indoor temperature, and complete the optimization of the control model with the heat exchange station.

Author Contributions: This paper is the fruit of all authors' collaboration. Problem formulation and conceptualization, G.X. and J.S.; methodology, J.S., and G.X.; software, J.S., Y.Z. and S.-a.Y.; data curation, X.G., and J.S.; writing—original draft preparation, G.X., X.G. and J.S.; supervision, C.Q.; project administration, C.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 61702157, the Tianjin Technical Expert Project of China under Grant 19JCTPJC51400, the Tianjin Key Research and Development Plan of China under Grant 20YFZCSN01010, the Natural science foundation of Tianjin under Grant.20JCQNJC00430 and the Hebei Innovation Capability Improvement Project of China under Grant 20554501D.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jie, P.F.; Zhu, N.; Li, D.Y. Operation optimization of existing district heating systems. *Appl. Therm. Eng.* **2015**, *78*, 278–288.
2. Ebenstein, A.; Fan, M.; Greenstone, M.; He, G.; Zhou, M. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10384–10389. [PubMed]

3. Yuan, S.S.; Xu, W.; Liu, Y. Research of heat metering development situation and application effect evaluation model in China. *Energy Policy* **2016**, *88*, 329–342.
4. Yuan, S.S.; Xu, W. Establishment and validation of a sustainable evaluation model for heat metering technology in China. *Energy Build.* **2015**, *99*, 153–161.
5. Junke, W.; Deying, L. Economic and Technical Analysis of Several Heat Metering Modes: A Case Study from Beijing, China. *Procedia Eng.* **2015**, *121*, 1037–1043.
6. Zhang, Q.; Tian, Z.; Ma, Z.J.; Li, G.Y.; Lu, Y.K.; Niu, J.D. Development of the heating load prediction model for the residential building of district heating based on model calibration. *Energy* **2020**, *205*, 117949. [CrossRef]
7. Buttitta, G.; Finn, D.P. A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes. *Energy Build.* **2020**, *206*, 109577. [CrossRef]
8. Shen, P.Y.; Dai, M.K.; Xu, P.; Dong, W. Building heating and cooling load under different neighbourhood forms: Assessing the effect of external convective heat transfer. *Energy* **2019**, *173*, 75–91.
9. Lim, H.S.; Kim, G. Development of Regression Models considering Time-Lag and Aerosols for Predicting Heating Loads in Buildings. *Adv. Civil Eng.* **2018**, *2018*, 1–19. [CrossRef]
10. Dahl, M.; Brun, A.; Kirsebom, O.S.; Andresen, G.B. Improving Short-Term Heat Load Forecasts with Calendar and Holiday Data. *Energies* **2018**, *11*, 1678. [CrossRef]
11. Sholahudin, S.; Han, H. Simplified dynamic neural network model to predict heating load of a building using Taguchi method. *Energy* **2016**, *115*, 1672–1678. [CrossRef]
12. Ahmad, T.; Chen, H.X. Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy Build.* **2018**, *166*, 460–476. [CrossRef]
13. Ding, Y.; Zhang, Q.; Yuan, T.H.; Yang, K. Model input selection for building heating load prediction: A case study for an office building in Tianjin. *Energy Build.* **2018**, *159*, 254–270. [CrossRef]
14. Moradzadeh, A.; Mansour-Saatloo, A.; Mohammadi-Ivatloo, B.; Anvari-Moghaddam, A. Performance Evaluation of Two Machine Learning Techniques in Heating and Cooling Loads Forecasting of Residential Buildings. *Appl. Sci.* **2020**, *10*, 3829. [CrossRef]
15. Jihad, A.S.; Tahiri, M. Forecasting the heating and cooling load of residential buildings by using a learning algorithm "gradient descent", Morocco. *Case Stud. Therm. Eng.* **2018**, *12*, 85–93. [CrossRef]
16. Turhan, C.; Kazanasmaz, T.; Uygun, I.E.; Ekmen, K.E.; Akkurt, G.G. Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation. *Energy Build.* **2014**, *85*, 115–125. [CrossRef]
17. Bui, X.N.; Moayedi, H.; Rashid, A.S.A. Developing a predictive method based on optimized M5Rules-GA predicting heating load of an energy-efficient building system. *Eng. Comput.* **2020**, *36*, 931–940. [CrossRef]
18. Le, L.T.; Nguyen, H.; Dou, J.; Zhou, J. A Comparative Study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in Estimating the Heating Load of Buildings' Energy Efficiency for Smart City Planning. *Appl. Sci.* **2019**, *9*, 2630. [CrossRef]
19. Le, L.T.; Nguyen, H.; Zhou, J.; Dou, J.; Moayedi, H. Estimating the Heating Load of Buildings for Smart City Planning Using a Novel Artificial Intelligence Technique PSO-XGBoost. *Appl. Sci.* **2019**, *9*, 2714. [CrossRef]
20. Kiesel, K.; Orehounig, K.; Mandavi, A. Model assumptions and their influence on the predicted heating load of an office building. *Bauphysik* **2010**, *32*, 359–364. [CrossRef]
21. Guelpa, E.; Marincioni, L.; Verda, V. Towards 4th generation district heating: Prediction of building thermal load for optimal management. *Energy* **2019**, *171*, 510–522. [CrossRef]
22. Moayedi, H.; Bui, D.T.; Dounis, A.; Lyu, Z.; Foong, L.K. Predicting Heating Load in Energy-Efficient Buildings Through Machine Learning Techniques. *Appl. Sci.* **2019**, *9*, 4338. [CrossRef]
23. Koschwitz, D.; Frisch, J.; van Treeck, C. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale. *Energy* **2018**, *165*, 134–142. [CrossRef]
24. Chou, J.S.; Bui, D.K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy Build.* **2014**, *82*, 437–446. [CrossRef]
25. Zhao, J.; Liu, X.J. A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. *Energy Build.* **2018**, *174*, 293–308. [CrossRef]

26. Roy, S.S.; Roy, R.; Balas, V.E. Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. *Renew. Sustain. Energy Rev.* **2018**, *82*, 4256–4268.
27. Liu, K.; Liu, T.Z.; Jian, P.; Lin, Y. The re-optimization strategy of multi-layer hybrid building's cooling and heating load soft sensing technology research based on temperature interval and hierarchical modeling techniques. *Sustain. Cities Soc.* **2018**, *38*, 42–54. [CrossRef]
28. Bui, D.T.; Moayedi, H.; Anastasios, D.; Foong, L.K. Predicting Heating and Cooling Loads in Energy-Efficient Buildings Using Two Hybrid Intelligent Models. *Appl. Sci.* **2019**, *9*, 3543. [CrossRef]
29. Roy, S.S.; Samui, P.; Nagtode, I.; Jain, H.; Shivaramakrishnan, V.; Mohammadi-ivatloo, B. Forecasting heating and cooling loads of buildings: A comparative performance analysis. *J. Ambient Intell. Humaniz Comput.* **2020**, *11*, 1253–1264. [CrossRef]
30. Protic, M.; Shamshirband, S.; Petkovic, D.; Abbasi, A.; Kiah, M.L.M.; Unar, J.A.; Zivkovic, L.; Raos, M. Forecasting of consumers heat load in district heating systems using the support vector machine with a discrete wavelet transform algorithm. *Energy* **2015**, *87*, 343–351. [CrossRef]
31. Izadyar, N.; Ghadamian, H.; Ong, H.C.; Moghadam, Z.; Tong, C.W.; Shamshirband, S. Appraisal of the support vector machine to forecast residential heating demand for the District Heating system based on the monthly overall natural gas consumption. *Energy* **2015**, *93*, 1558–1567. [CrossRef]
32. Gao, X.; Li, X.; Zhao, B.; Ji, W.; Jing, X.; He, Y. Short-Term Electricity Load Forecasting Model Based on EMD-GRU with Feature Selection. *Energies* **2019**, *12*, 1140. [CrossRef]
33. Yang, J.; Chang, Y.; Gao, T.; Wang, J. Failure Prediction of the Rotating Machinery Based on CEEMDAN-ApEn Feature and AR-UKF Model. *Appl. Sci.* **2020**, *10*, 2056. [CrossRef]
34. Jiang, W.; Zhou, J.; Xu, Y.; Liu, J.; Shan, Y. Multistep Degradation Tendency Prediction for Aircraft Engines Based on CEEMDAN Permutation Entropy and Improved Grey-Markov Model. *Complexity* **2019**. [CrossRef]
35. Wang, J.; Luo, Y.; Tang, L.; Ge, P. A new weighted CEEMDAN-based prediction model: An experimental investigation of decomposition and non-decomposition approaches. *Knowl. Based Syst.* **2018**, *160*, 188–199.
36. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Ences* **1998**, *454*, 903–995. [CrossRef]
37. Zhaohua, W.U.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**. [CrossRef]
38. Colominas, M.A.; Schlotthauer, G.; Torres, M.E.; Flandrin, P. Noise-assisted emd methods in action. *Adv. Adapt. Data Anal.* **2012**, *4*. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).