

Multiple Regression Model Based on Weather Factors for Predicting The Heat Load of A District Heating System in Dalian, China—A Case Study

Qi Cai*, Wenbiao Wang and Siyuan Wang

School of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning, 116026, China

Abstract: The managers need to have a reasonable guide on the operation and management in district heating system (DHS), so it's very necessary to predicting the heat load for DHS. In this paper, the relationships between the heat load and weather conditions have been researched in order to determine the inputs variables and output variable of the future heat load prediction model. Using the given data from the obtained database, the multiple regression modelling and analysis method was carried out so as to establish the corresponding heat load prediction models for the DHS. The results shown that the square correlation coefficient between the heat load's measured value and estimated value are all greater than 0.9000, and the mean absolute percentage error (MAPE) between the heat load's measured value and estimated value are all less than 4.00%. Moreover, the corresponding maximum absolute relative errors between the heat load's measured value and estimated value are all less than 8%. The results also indicated that the heat load prediction model's accuracy is relatively high. Furthermore, these 5 heat load prediction models can be applied in the real DHS and this multiple regression method can be promoted into the other engineering field.

Keywords: Corrected weather factors, district heating system, heat load prediction, multiple regression analysis, weather conditions.

1. INTRODUCTION

District heating system (DHS) is an inseparable infrastructure part of the agglomerations and cities in many countries, such as in Russia, Finland, Sweden, Norway, Denmark, China, and so on [1]. In China, heat energy demand for homes, businesses and industrial processes accounts is around 49% of total energy demand, and 47% of the carbon emissions demand [2]. A DHS can be divided into two types based on its structure. One type of the DHS is called the direct district heating system (DDHS), which usually has a smaller heating area ranging from 50,000 square meters to 250,000 square meters. If the heating area is larger than this range, the DHS can be called the indirect district heating system (IDHS) [3]. In recent years, like IDHS, the heating area of some systems has already been built with more than 10,000,000 square meters in China. For a DHS, heat from the heat source is transported to each terminal heat user device in the form of hot water or steam [4].

The works of predicting the heat load are the basic works for a DHS. In order to do the research on the DHS so as to design and operate the DHS well, it's very necessary to predicting the heat load for DHS [5]. Because the heating data support can be provided to the managers by this predicting method, so the managers can have a reasonable guide on the operation and management in DHS [6]. Generally speaking, a heating process is a complex nonlinear dynamic system, its transfer relationship between the energy and substance

presents the obvious nonlinear characteristic [7]. Therefore, it's very difficult to establish the physics model for a DHS by the mechanism analysis. Recently, for the heat load forecasting method in DHS, large amounts of methods are based on the statistical analysis and historical data, these historical data usually includes the heat load data and weather data [8]. During the last 30 years, for the problems of predicting the heat load in DHS, many scholars have done large amounts of researches on the relationships between the heat load and weather conditions.

In 1984, Sven Werner [9] tested 6 DHSs on the different regions in Sweden, he analyzed the main factors impacting on the heat load, such as the outdoor temperature, nature wind and solar radiation, the experimental results shown that for the outdoor temperature, its influence effect on the heat load is about 60%; for the nature wind, its influence effect on the heat load is range from 1% to 4%; for the solar radiation, its influence effect on the heat load is range from 1% to 5%.

In 2004, Fernando Simon Westphal and Roberto Lamberts [10] used different weather cycles and weather data to estimate the non-residential buildings' heat load in a DHS in Brazil, they found that the heat load's estimated results were mainly influenced by the selections of different weather cycles and data, and the mean relative error was about 18%.

In 2008, Krzysztof Wojdyga [11] used two weather factors including the nature wind speed and solar radiation to correct the outdoor temperature factor for a DHS region in Poland, and he researched how the heat load is influenced by these two weather factors. He found that for the heat load, the influence effect of the nature wind speed was weaker than the solar radiation, when the nature wind speed was not

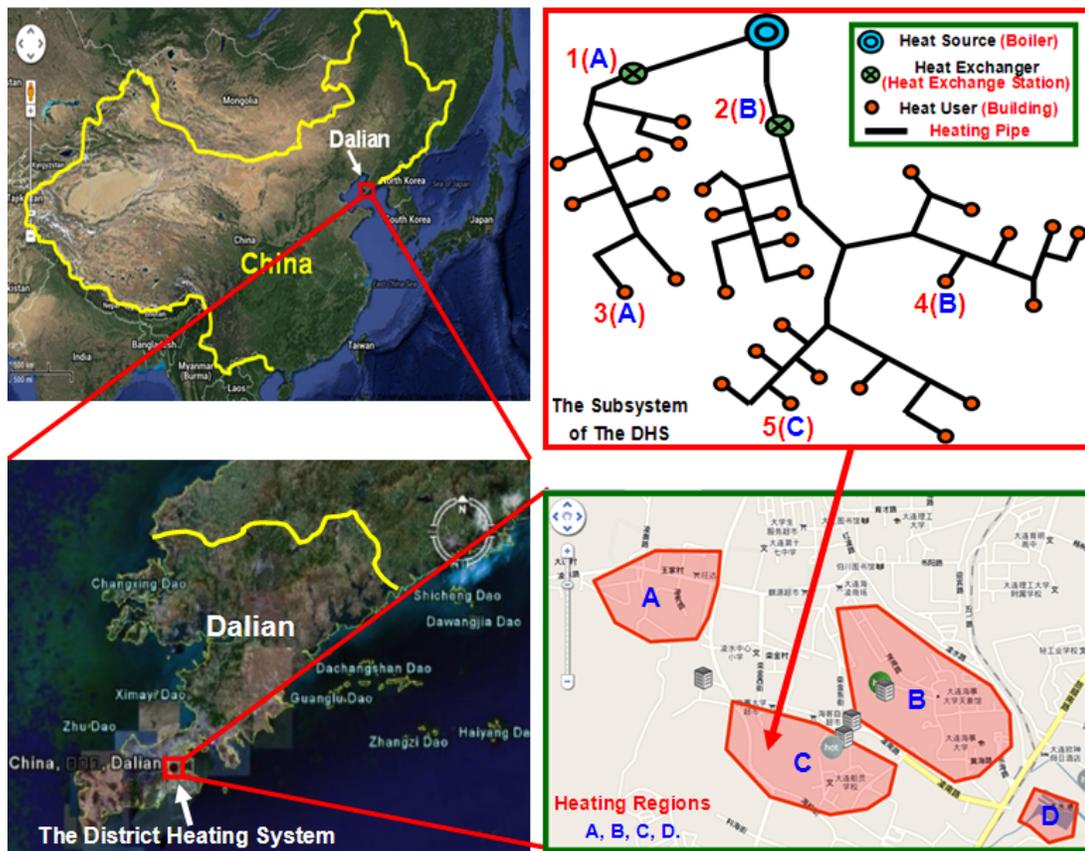


Fig. (1). The geographical position of the selected DHS in dalian, China.

more than 2 m/s, it wouldn't affect the basic heat load too much.

In 2009, Omer Yetemen and Tolga Yalcin [12] aimed at the weather conditions of a DHS region in Turkey, they selected the outdoor temperature, wind speed and solar radiation as the 3 weather factors, using the auto regression (AR) modelling method, they established a regression equation relationship between the heat load and these 3 weather factors, the output parameter is the heat load, the input parameters include the outdoor temperature, wind speed, solar radiation and the heat load form yesterday, and they found that for the long-term heat load prediction, the monsoon circulation factor should be considered into the AR model for predicting the heat load.

In this paper, based on the existing research results of many scholars during the last 30 years, the relationships between the heat load and weather conditions have been also researched in this paper. The purpose of the researches described in this paper is to find the multiple regression model based on weather factors for predicting the heat load of a DHS in Dalian, China.

2. MATERIALS

Dalian, situated on the northeast region of China (latitude $39^{\circ}01'N\sim 39^{\circ}04'N$, longitude $121^{\circ}44'E\sim 121^{\circ}49'E$), has a typical monsoon climate of medium latitudes which is characterized as high temperature and rainy summer, cold and dry winter [13]. In Dalian, the daily average temperatures during summer (June-September) vary between $20^{\circ}C$ and

$30^{\circ}C$, the daily average temperatures during winter (December-March) vary between $-5^{\circ}C$ and $5^{\circ}C$. In general, the heating season of Dalian includes 5 months from November to March [14].

2.1. Description of The DHS

Fig. (1) shows the geographical position of the selected DHS in this study. The selected DHS is located on a university of Dalian city in China, the type of the selected DHS is an IDHS, the total heating area of the selected DHS is about 800,000 square meters. In general, this DHS can serve about 40MW heat load for the heat consumers in its heating regions (Fig. 1).

Fig. (1) also shows the DHS consists of 4 heating regions including region A, region B, region C and region D. In this paper, the subsystem of the DHS was selected as the experimental object in this study, as is shown in Fig. (1), the subsystem of the DHS is located on the region C. Apparently, the subsystem of the DHS in region C is similar to the subsystems of the DHS in region A, region B and region D.

There are many optional practical equipments in this subsystem of the DHS, due to the limited paper space, therefore, only 5 practical equipments of the subsystem were selected as the experimental objects. For this subsystem of the DHS, its 5 experimental objects include 2 heat exchangers (heat exchange station) and 3 heat users (building), they are numbered by case 1, case 2, case 3, case 4 and case 5, respectively (Fig. 1). The corresponding 5 experimental objects' descriptions are shown in Table 1.

Table 1. Descriptions of the five experimental objects in this study.

Case Number	Experimental Objects
Case 1	Heat exchanger (heat exchange station) A
Case 2	Heat exchanger (heat exchange station) B
Case 3	Heat user (building) A
Case 4	Heat user (building) B
Case 5	Heat user (building) C

As can be seen from Table 1, among these 5 experimental objects, the case 1 presents the heat exchanger (heat exchange station) A; the case 2 presents heat exchanger (heat exchange station) B; the case 3 presents heat user (building) A; the case 4 presents Heat user (building) B; the case 5 presents heat user (building) C.

2.2. Overall Structure of The DHS

In general, a DHS can be grouped into two types based on its structure. One structure of a DHS is called a DDHS, another structure of a DHS is called an IDHS. An IDHS is constructed of the heat source (boiler), heat exchanger (heat exchange station) and heat user (building). In contrast, a DDHS is only constructed of the heat source (boiler) and heat user (building), not including the heat exchanger (heat exchange station). In this paper, the type of the selected DHS is an IDHS [15].

Fig. (2) shows the overall structure of the IDHS. There are 3 basic parts for this IDHS, the first part is the heat source, its corresponding practical equipment is the boiler; the second part is the heat exchanger, its corresponding prac-

tical equipment is the heat exchange station; the third part is the heat user, its corresponding practical equipment is the building.

As can be seen from Fig. (2), the IDHS is like a district heating network, this district heating network is constructed of the heating pipe, including the supply water pipe and return water pipe. Seeing from the overall structure of the district heating network, this district heating network consists of the primary pipe network and secondary pipe network. The primary pipe network includes the pressurized water pump, circulating water pump and make-up water pump, but the secondary pipe network only includes the circulating water pump and make-up water pump.

3. DATA

In this study, for the selected DHS, its whole heating information data can be collected by the corresponding heating information network platform. Based on the technology of Internet of things (IOT), this heating information network platform was built by the corresponding university of Dalian city in China, which is the location of the selected DHS in this study.

Moreover, this IOT heating information network platform can be connected to the local meteorological station network platform in Dalian city, so the local weather information can be also browsed on the IOT heating information network platform.

3.1. Data Description

The whole heating information data can be collected by the various sensors, and then, they are transferred to the corresponding measuring equipments. Using the wired data transferring method based on Modbus transmission control protocol (TCP), and using the wireless data transferring

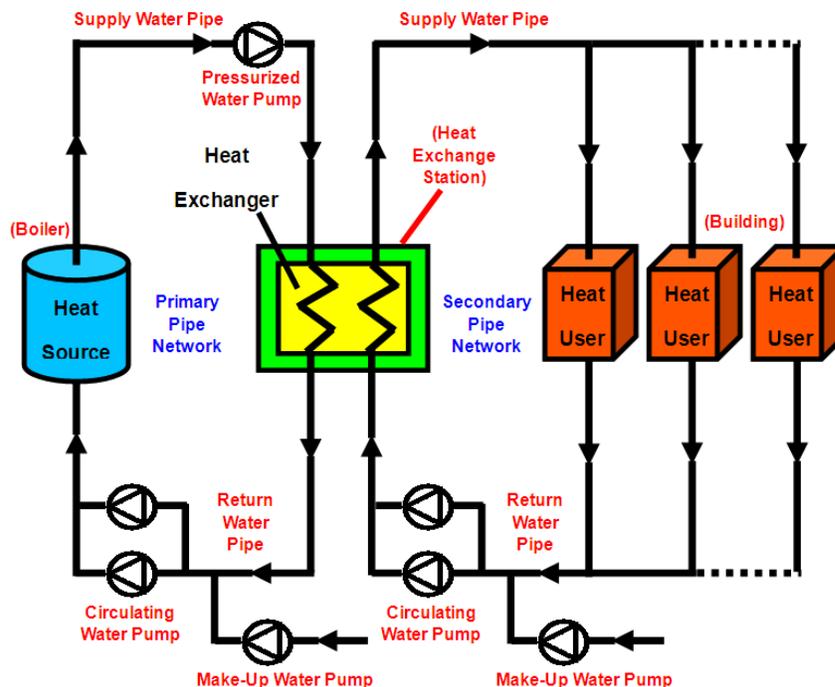


Fig. (2). The overall structure schematic diagram of the selected IDHS.

Table 2. Two types of the data parameters in the database

Types of Parameters	Data Parameters	Symbols of Parameters
Heating parameters	Supply water temperature	T_{supply} (°C)
	Return water temperature	T_{return} (°C)
	Supply water flow	F_{flow} (t/h)
Weather parameters	Outdoor temperature	T_{out} (°C)
	Wind speed	W_{wind} (m/s)
	Solar radiation	S_{solar} (w/m ²)

method based on general packet radio service (GPRS), all the collected heating information data can be transferred to the corresponding server computer, and then, all the heating information data will be stored into its database.

The whole weather information data can be collected by the local meteorological station in Dalian city. Using the C# language programming technology, the local meteorological station network platform can be connected to the IOT heating information network platform, so all the weather information data can be also transferred to the IOT heating information network platform, and then, all the weather information data will be stored into its database.

Therefore, for the database of the IOT heating information network platform, there are 2 types of the data parameters in this database, one is the heating parameters, the other is the weather parameters. The corresponding 2 types of the data parameters in the database are shown in Table 2.

As can be seen from Table 2, the heating parameters include the supply water temperature T_{supply} (°C), return water temperature T_{return} (°C), supply water flow F_{flow} (t/h), and so on; the weather parameters include the outdoor temperature T_{out} (°C), wind speed W_{wind} (m/s), solar radiation S_{solar} (w/m²), and so on. In this database, the sampling period of the whole data is per 10 minutes, because of one day includes 24 hours and one hour includes 60 minutes, so for the database, one day includes 144 sample data points.

In this study, the whole data was calculated to the daily average value for the corresponding research, including the daily average supply water temperature T_{supply} , daily average return water temperature T_{return} , daily average supply water flow F_{flow} , daily average outdoor temperature T_{out} , daily average wind speed W_{wind} and daily average solar radiation S_{solar} (Table 2). Therefore, the calculated heat load in this study is also the daily average heat load for the corresponding research.

3.2. Data Analysis

In order to find the relationship between the heating parameters and weather parameters, therefore, it's necessary to analyze the correlation between the heating parameters and weather parameters. The corresponding head load calculation equation in this paper can be written as follows:

$$Q_{load} = \frac{C_k F_{flow} (T_{supply} - T_{return})}{3600} \quad (1)$$

where Q_{load} is the head load (MW), T_{supply} is the supply water temperature (°C), T_{return} is the return water temperature (°C), F_{flow} is the supply water flow (t/h), C_k is the specific heat capacity of water (kJ/(kg·°C)), it is a constant, its value is 4.186kJ/(kg·°C) [16].

In this study, the heat exchanger (heat exchange station) B (Fig. 1) was selected as the data analysis object. For the data analysis, the sample data of the database was selected from January (2015) to March (2015) during the 2014-2015 heating season (Table 3), these 3 months include 90 days, so the number of the sample data points is 90.

For the heat exchanger (heat exchange station) B, its heat load Q_{load} during the 90 days can be calculated by the Eq. (1), the corresponding outdoor temperature T_{out} on the location of the DHS can be also acquired from the database. Then, the corresponding daily average heat load Q_{load} and daily average outdoor temperature T_{out} can be also calculated.

And then, aiming at the heat exchanger (heat exchange station) B, its daily average heat load Q_{load} with varying daily average outdoor temperature T_{out} during the 90 days are shown in Fig. (3).

As can be seen from Fig. (3), the interrelation between the daily average heat load Q_{load} and daily average outdoor

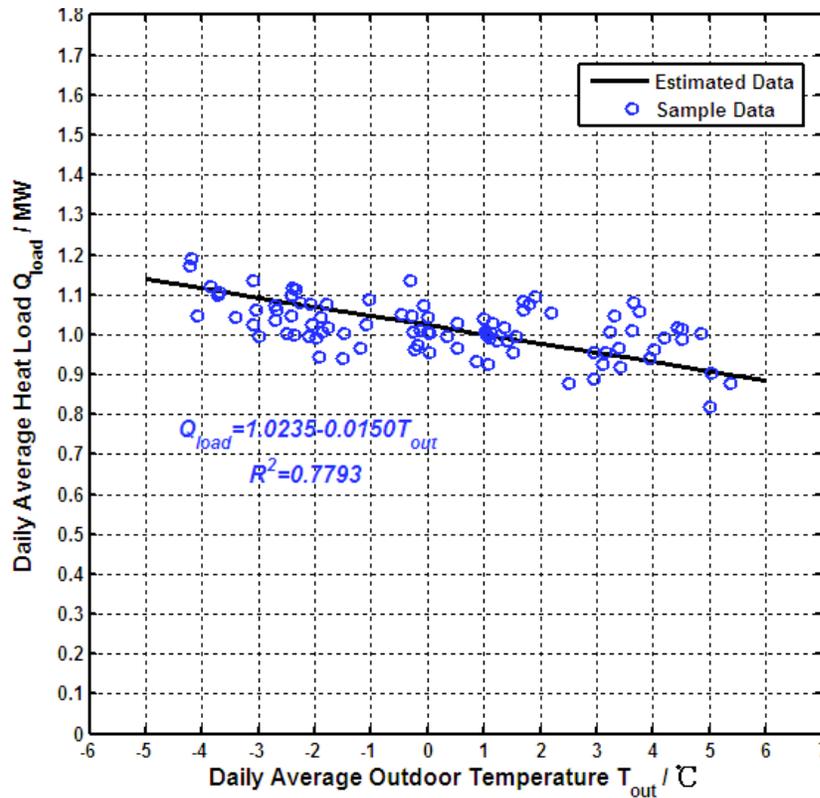


Fig. (3). The daily average heat load Q_{load} with varying daily average outdoor temperature T_{out} during the 90 days for the heat exchanger (Heat exchange station) B.

temperature T_{out} can be described as a similar linear relationship, therefore, this interrelation can be described by a linear regression equation. After calculation, the corresponding linear regression equation of the interrelation can be written as follows:

$$Q_{load} = 1.0235 - 0.0150T_{out} \tag{2}$$

where the square correlation coefficient R^2 between the sample data and estimated data is about 0.7793 (Fig. 6). This experimental result also indicated that the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature T_{out} is close to a linear relationship.

Although the heat load of a DHS can be influenced by the outdoor temperature, the local weather conditions also include the other weather factors, such as wind speed, solar radiation, and so on.

In order to confirm the wind speed's influence on the heat load of the DHS in this study, the corresponding wind speed variable can be corrected to an equivalent cooling outdoor temperature variable, the corresponding wind speed variable correcting equation can be written as follow [17]:

$$\Delta T_{wind} = 0.0246(\lg(7.23W_{wind}))^3 - 0.4525(\lg(7.23W_{wind}))^2 + 3.2398\lg(7.23W_{wind}) \tag{3}$$

where W_{wind} is the wind speed, ΔT_{wind} is the wind speed's equivalent cooling outdoor temperature. Then, this wind speed's equivalent cooling outdoor temperature variable ΔT_{wind} can be added into the outdoor temperature variable T_{out} so as to form a new variable, the corresponding calculation equation of the new variable can be written as follow:

$$T_{out-wind} = T_{out} - \Delta T_{wind} \tag{4}$$

where $T_{out-wind}$ is the outdoor temperature with wind speed added, it's the new variable. Then, the corresponding daily average outdoor temperature (with wind speed added) $T_{out-wind}$ can be also calculated.

And then, aiming at the heat exchanger (heat exchange station) B, its daily average heat load Q_{load} with varying daily average outdoor temperature (with wind speed added) $T_{out-wind}$ during the 90 days are shown in Fig. (4).

As can be seen from Fig. (4), the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature (with wind speed added) $T_{out-wind}$ can be also described as a similar linear relationship, therefore, this interrelation can be also described by a linear regression equation. After calculation, the corresponding linear regression equation of the interrelation can be written as follows:

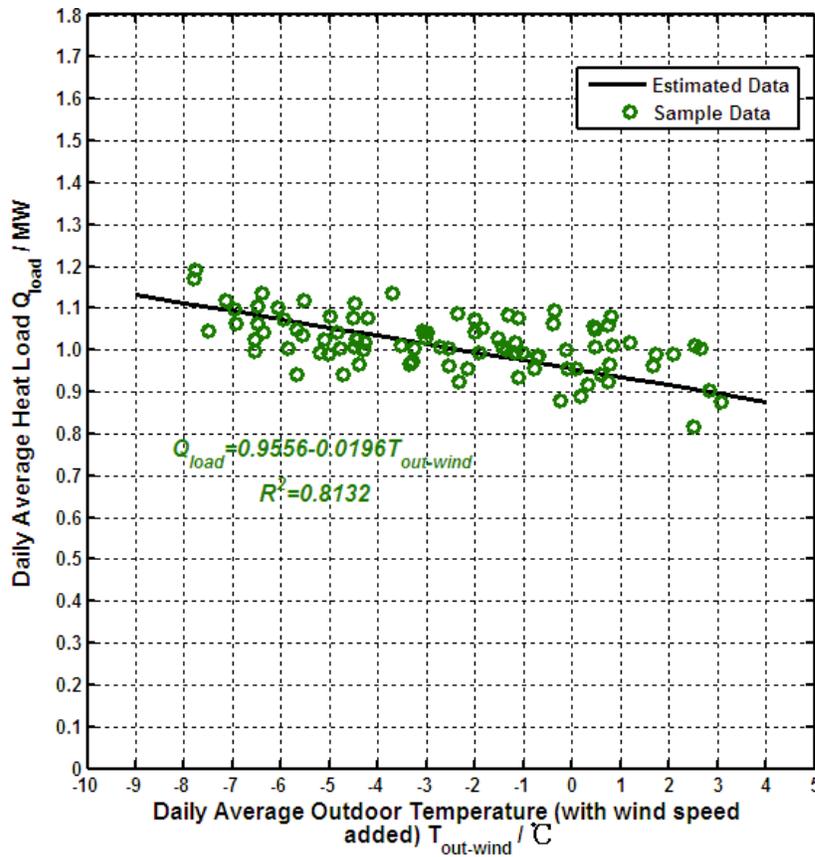


Fig. (4). The daily average heat load Q_{load} with varying daily average outdoor temperature (with wind speed added) $T_{out-wind}$ during the 90 days for the heat exchanger (Heat exchange station) B.

$$Q_{load} = 0.9556 - 0.0196T_{out-wind} \tag{5}$$

where the square correlation coefficient R^2 between the sample data and estimated data is about 0.8132 (Fig. 7). This experimental result also indicated that the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature (with wind speed added) $T_{out-wind}$ is close to a linear relationship. This experimental result also shown that the local wind speed has the influence on the local outdoor temperature so as to impact the head load of the DHS.

In order to confirm the solar radiation’s influence on the heat load of the DHS in this study, the corresponding solar radiation variable can be corrected to an equivalent warming outdoor temperature variable, the corresponding solar radiation variable correcting equation can be written as follow [18]:

$$\Delta T_{solar} = \frac{S_{solar}}{S_k} \tag{6}$$

where S_{solar} is the solar radiation, ΔT_{solar} is the solar radiation’s equivalent warming outdoor temperature, S_k is the solar radiation’s outdoor temperature converted coefficient ($w/(m^2 \cdot ^\circ C)$), it is a constant, its value is $100w/(m^2 \cdot ^\circ C)$ [19].

Then, this solar radiation’s equivalent warming outdoor temperature variable ΔT_{solar} can be added into the outdoor temperature variable T_{out} so as to form a new variable, the corresponding calculation equation of the new variable can be written as follow:

$$T_{out-solar} = T_{out} + \Delta T_{solar} \tag{7}$$

where $T_{out-solar}$ is the outdoor temperature with solar radiation added, it’s the new variable. Then, the corresponding daily average outdoor temperature (with solar radiation added) $T_{out-solar}$ can be also calculated.

And then, aiming at the heat exchanger (heat exchange station) B, its daily average heat load Q_{load} with varying daily average outdoor temperature (with solar radiation added) $T_{out-solar}$ during the 90 days are shown in Fig. (5).

As can be seen from Fig. (5), the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature (with solar radiation added) $T_{out-solar}$ can be also described as a similar linear relationship, therefore, this interrelation can be also described by a linear regression equation. After calculation, the corresponding linear regression equation of the interrelation can be written as follows:

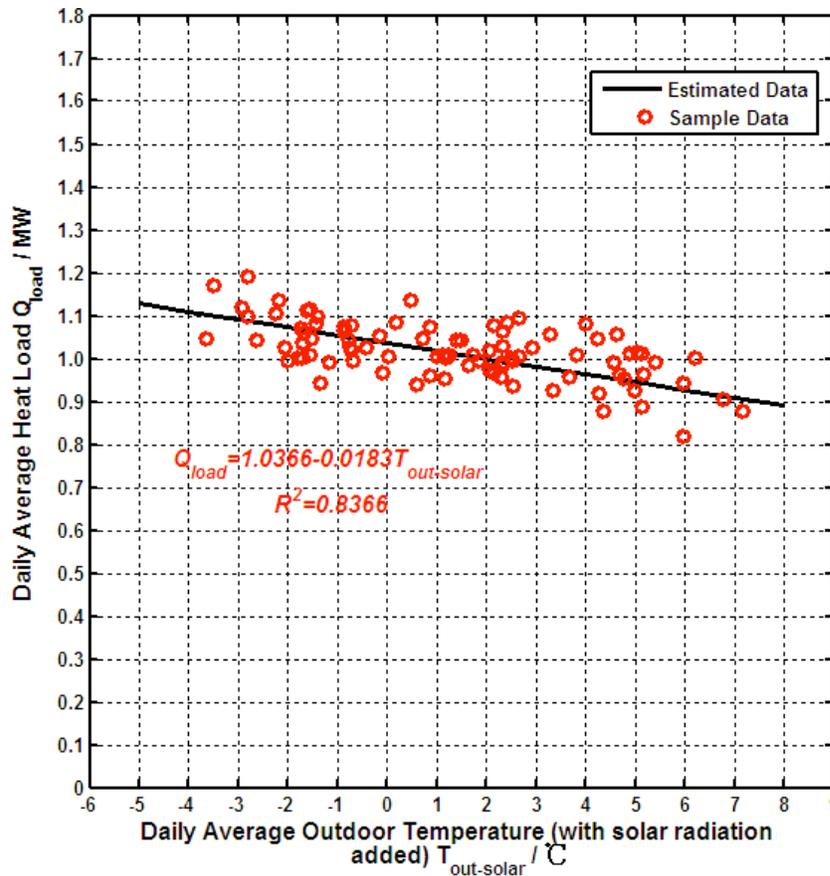


Fig. (5). The daily average heat load Q_{load} with varying daily average outdoor temperature (with solar radiation added) $T_{out-solar}$ during the 90 days for the heat exchanger (Heat exchange station) B.

$$Q_{load} = 1.0366 - 0.0183T_{out-solar} \tag{8}$$

where the square correlation coefficient R^2 between the sample data and estimated data is about 0.8366 (Fig. 8). This experimental result also indicated that the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature (with solar radiation added) $T_{out-solar}$ is close to a linear relationship. This experimental result also shown that the local solar radiation has the influence on the local outdoor temperature so as to impact the head load of the DHS.

In order to confirm the wind speed and solar radiation’s combined influence on the heat load of the DHS in this study, the corresponding wind speed’s equivalent cooling outdoor temperature variable ΔT_{wind} and solar radiation’s equivalent warming outdoor temperature variable ΔT_{solar} can be added into the outdoor temperature variable T_{out} so as to form a new variable, the corresponding calculation equation of the new variable can be written as follow:

$$T_{out-wind-solar} = T_{out} - \Delta T_{wind} + \Delta T_{solar} \tag{9}$$

where $T_{out-wind-solar}$ is the outdoor temperature with wind speed and solar radiation added, it’s the new variable. Then,

the corresponding daily average outdoor temperature (with wind speed and solar radiation added) $T_{out-wind-solar}$ can be also calculated.

And then, aiming at the heat exchanger (heat exchange station) B, its daily average heat load Q_{load} with varying daily average outdoor temperature (with wind speed and solar radiation added) $T_{out-wind-solar}$ during the 90 days are shown in Fig. (6).

As can be seen from Fig. (6), the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature (with wind speed and solar radiation added) $T_{out-wind-solar}$ can be also described as a similar linear relationship, therefore, this interrelation can be also described by a linear regression equation. After calculation, the corresponding linear regression equation of the interrelation can be written as follows:

$$Q_{load} = 1.0249 - 0.0185T_{out-wind-solar} \tag{10}$$

where the square correlation coefficient R^2 between the sample data and estimated data is about 0.8616 (Fig. 9). This experimental result also indicated that the interrelation between the daily average heat load Q_{load} and daily average outdoor temperature (with wind speed and solar radiation

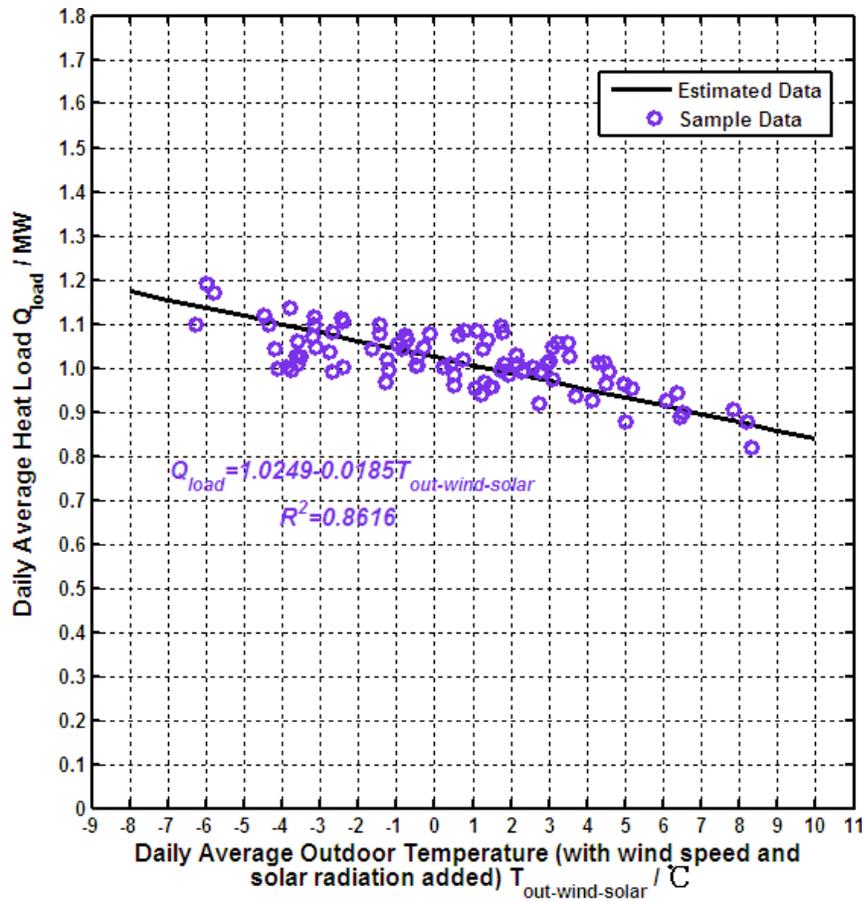


Fig. (6). The daily average heat load Q_{load} with varying daily average outdoor temperature (with wind speed and solar radiation added) $T_{out-wind-solar}$ during the 90 days for the heat exchanger (Heat exchange station) B.

added) $T_{out-wind-solar}$ is close to a linear relationship. This experimental result also shown that the local wind speed and solar radiation have the combined influences on the local outdoor temperature so as to impact the head load of the DHS.

If the weather conditions only include the outdoor temperature, the square correlation coefficient R^2 between the daily average heat load Q_{load} and daily average outdoor temperature T_{out} is only 0.7793; If the weather conditions include the outdoor temperature and wind speed, the square correlation coefficient R^2 between the daily average heat load Q_{load} and daily average outdoor temperature (with wind speed added) $T_{out-wind}$ is 0.8132; If the weather conditions include the outdoor temperature and solar radiation, the square correlation coefficient R^2 between the daily average heat load Q_{load} and daily average outdoor temperature (with solar radiation added) $T_{out-solar}$ is 0.8366; If the weather conditions include the outdoor temperature, wind speed and solar radiation, the square correlation coefficient R^2 between the daily average heat load Q_{load} and daily average

outdoor temperature (with wind speed and solar radiation added) $T_{out-wind-solar}$ is 0.8616.

From the 4 data analysis results, it can be seen that the local wind speed really has the influence on the local outdoor temperature so as to impact the heat load of the DHS, the local solar radiation really has the influence on the local outdoor temperature so as to impact the head load of the DHS. These experimental results also indicated that the local wind speed and solar radiation have the combined influences on the local outdoor temperature so as to impact the head load of the DHS.

3.3. Data Selection

For the data analysis above in this paper, it can be seen that the heat load of the DHS really can be influenced by the local weather conditions. However, for the general DHSs, its heat load from yesterday has the most important influence on its heat load from today [20]. Therefore, for the selected DHS in this study, its heat load from yesterday can be also considered as an impact factor on its heat load from today. The weather factors variables include the outdoor temperature variable, wind speed variable and solar radiation variable; the heat load variables include the yesterday and today. The corresponding selected 5 data parameters for the research in this paper are shown in Table 3.

Table 3. The selected 5 data parameters for the research in this paper.

Variable Number	Data Parameters	Symbols of Parameters
1	Daily average outdoor temperature from today	T_{out}^d (°C)
2	Daily average wind speed from today	W_{wind}^d (m/s)
3	Daily average solar radiation from today	S_{solar}^d (w/m ²)
4	Daily average heat load from yesterday	Q_{load}^{d-1} (MW)
5	Daily average heat load from today	Q_{load}^d (MW)

Table 4. Four input parameters and one output parameter of the ar model for predicting the dhs’s heat load.

Input Parameters	Output Parameters
Daily average outdoor temperature from today T_{out}^d	Daily average heat load from today Q_{load}^d
Daily average wind speed from today W_{wind}^d	
Daily average solar radiation from today S_{solar}^d	
Daily average heat load from yesterday Q_{load}^{d-1}	

As can be seen from Table 3, for the research in this paper, the whole variables are calculated to the daily average values in this paper. Where T_{out}^d is the daily average outdoor temperature from today (°C), W_{wind}^d is the daily average wind speed from today (m/s), S_{solar}^d is the daily average solar radiation from today (w/m²), Q_{load}^{d-1} is the daily average heat load from yesterday (MW), Q_{load}^d is the daily average heat load from today (MW). Where d is the day index, $d-1$ presents the previous day.

4. METHODOLOGY

In this paper, the purpose of the research is to predict the heat load of the DHS in Dalian, China. In this study, the multiple regression analysis method was used to establish the heat load prediction model for the selected DHS. The corresponding multiple regression model in this study is an auto regression (AR) model.

Aiming at the 5 experimental objects of the selected DHS, using the multiple regression analysis method and based on the weather factors, the 5 heat load prediction regression models were established in this study.

4.1. Model Development

Regression analysis method is an easier and more practical solution to the complex modelling problems. Based on the given data, the purpose of the regression analysis is to

find an appropriate mathematical model and to choose the best model’s fitting coefficients [21].

Regression analysis method is one of the most used statistical methods to describe the relationship between the dependent variable and explanatory variables, the corresponding relationship can be described by a function equation [22].

In general, using the regression technique is a clear viable method to develop a prediction model. When having the large amounts of data like the database in this study, this regression technique can be easily applied with success and good data fitting results in the modelling process [23].

In this research, based on the weather factors’ influences on the head load of the DHS, the AR model was selected as the multiple regression model to predicting the heat load for the selected DHS. As can be seen from this paper above, the 4 input parameters were considered in the AR model analysis: daily average outdoor temperature T_{out}^d , daily average wind speed W_{wind}^d , daily average solar radiation S_{solar}^d and daily average heat load from yesterday Q_{load}^{d-1} . Apparently, the purpose of the research in this paper is to predict the heat load of the selected DHS, so the daily average heat load from today Q_{load}^d was considered as the only output parameter in the AR model analysis. The corresponding 4 input parameters and 1 output parameter of the AR model for predicting the DHS’s heat load are shown in Table 4.

As can be seen from Table 4, the whole variables are calculated to the daily average values in this paper, where d is

the day index, $d-1$ presents the previous day. The corresponding AR model for predicting the heat load of the DHS can be described by a regression function, bases on the 4 input parameters and 1 output parameter, this regression function can be written as follows:

$$Q_{load}^d = f(T_{out}^d, W_{wind}^d, S_{solar}^d, Q_{load}^{d-1}) \quad (11)$$

where $f(,,,)$ is the regression function for this AR model. In this research, this regression function is a multiple variable regression function including 4 input variables, in order to reduce the complex degree of the prediction model, so this prediction model was defined in a first-order model in this research. This details equation of the multiple regression function for the AR prediction model can be written as follows:

$$Q_{load}^d = A + B \cdot T_{out}^d + C \cdot W_{wind}^d + D \cdot S_{solar}^d + E \cdot Q_{load}^{d-1} \quad (12)$$

where A, B, C, D and E are all the regression coefficients of this multiple regression function equation. Using this multiple regression function equation, based on the given data from the IOT heating information network platform's database in this study, the corresponding AR model for predicting the DHS's heat load can be easily established. The training and testing data of the heat load prediction model was selected from the IOT heating information network platform's database in 2014-2015 heating season.

In Dalian city of China, the 2014-2015 heating season includes 5 months ranging from November (2014) to March (2015). Because of the IOT heating information network platform's database was maintained during November (2014) to December (2014) in 2014-2015 heating season, so the data of these 2 months including November (2014) and December (2014) is not appropriate for modelling. However, the IOT heating information network platform was running in the normal state during January (2015) to March (2015). Therefore, the data of these 3 months including January (2015) to March (2015) can be chosen as an appropriate experimental data in this research (Table 3).

For the training data sets of the heat load prediction model, the data of January (2015) was selected as the training data. Because of the January (2015) includes 31 days, so the data points of the training data sets is 31. The corresponding training data sets can be written as follows:

$$\{T_{out}^{d|train}\} = \{T_{out}^{1|train}, T_{out}^{2|train}, \dots, T_{out}^{k|train}, \dots, T_{out}^{31|train}\} \quad (13)$$

$$\{W_{wind}^{d|train}\} = \{W_{wind}^{1|train}, W_{wind}^{2|train}, \dots, W_{wind}^{k|train}, \dots, W_{wind}^{31|train}\} \quad (14)$$

$$\{S_{solar}^{d|train}\} = \{S_{solar}^{1|train}, S_{solar}^{2|train}, \dots, S_{solar}^{k|train}, \dots, S_{solar}^{31|train}\} \quad (15)$$

$$\{Q_{load}^{d-1|train}\} = \{Q_{load}^{0|train}, Q_{load}^{1|train}, \dots, Q_{load}^{k-1|train}, \dots, Q_{load}^{30|train}\} \quad (16)$$

$$\{Q_{load}^{d|train}\} = \{Q_{load}^{1|train}, Q_{load}^{2|train}, \dots, Q_{load}^{k|train}, \dots, Q_{load}^{31|train}\} \quad (17)$$

where $\{T_{out}^{d|train}\}, \{W_{wind}^{d|train}\}, \{S_{solar}^{d|train}\}, \{Q_{load}^{d-1|train}\}$ and $\{Q_{load}^{d|train}\}$ are the training data sets. $\{T_{out}^{d|train}\}$ is the training data set of daily average outdoor temperature, $\{W_{wind}^{d|train}\}$ is the training data set of daily average wind speed, $\{S_{solar}^{d|train}\}$ is the training data set of daily average solar radiation, $\{Q_{load}^{d-1|train}\}$ is the training data set of daily average heat load from yesterday, $\{Q_{load}^{d|train}\}$ is the training data set of daily average heat load from today. Where $d = 1, 2, \dots, 31$.

The given data from the database can be converted to the training data sets like Eq. (13), Eq. (14), Eq. (15), Eq. (16) and Eq. (17). Then, aiming at the multiple regression function equation, using the levenberg-marquardt (LM) algorithm, the regression coefficients A, B, C, D and E of Eq. (12) can be identified. When these 5 regression coefficients are identified, the corresponding heat load prediction model is also established, the corresponding model development is also accomplished.

4.2. Model Validation

The 5 heat load prediction models' regression coefficients can be all identified by the LM algorithm in this research. The implementation of the LM algorithm was by a new data fitting software named 1st Opt.

1st Opt is a new introduced integrated numerical optimization software which was developed by the famous 7D-soft high technology limited company in China [24]. It has a strong global optimization capability with easy-to-use graphical user interface (GUI), and extendable programming languages' supports. Therefore, 1stOpt was selected as the data fitting software in this research.

When the whole regression coefficients A, B, C, D and E (Eq. (12)) are all identified by 1stOpt with LM algorithm, the corresponding 5 heat load prediction models are also established. And then, in order to evaluate these 5 models' accuracy, it's necessary to validate these 5 models' feasibility.

As can be seen from this paper above, for the data from the IOT heating information network platform's database in 2014-2015 heating season (including 5 months), only 3 months' data is complete and appropriate for this research. Because of the data of January (2015) was selected as the training data, and January (2015) includes 31 days as the same as March (2015), in order to keep the continuity and consistency between the training data and testing data, so the data of March (2015) was selected as the testing data. Because of the March (2015) also includes 31 days, so the data points of the testing data sets is also 31. The corresponding testing data sets can be written as follows:

$$\{T_{out}^{d|test}\} = \{T_{out}^{1|test}, T_{out}^{2|test}, \dots, T_{out}^{k|test}, \dots, T_{out}^{31|test}\} \quad (18)$$

$$\{W_{wind}^{d|test}\} = \{W_{wind}^{1|test}, W_{wind}^{2|test}, \dots, W_{wind}^{k|test}, \dots, W_{wind}^{31|test}\} \quad (19)$$

$$\{S_{solar}^{d|test}\} = \{S_{solar}^{1|test}, S_{solar}^{2|test}, \dots, S_{solar}^{k|test}, \dots, S_{solar}^{31|test}\} \quad (20)$$

$$\{Q_{load}^{d-1|test}\} = \{Q_{load}^{0|test}, Q_{load}^{1|test}, \dots, Q_{load}^{k-1|test}, \dots, Q_{load}^{30|test}\} \quad (21)$$

$$\{Q_{load}^{d|test}\} = \{Q_{load}^{1|test}, Q_{load}^{2|test}, \dots, Q_{load}^{k|test}, \dots, Q_{load}^{31|test}\} \quad (22)$$

where $\{T_{out}^{d|test}\}$, $\{W_{wind}^{d|test}\}$, $\{S_{solar}^{d|test}\}$, $\{Q_{load}^{d-1|test}\}$ and $\{Q_{load}^{d|test}\}$ are the testing data sets. $\{T_{out}^{d|test}\}$ is the testing data set of daily average outdoor temperature, $\{W_{wind}^{d|test}\}$ is the testing data set of daily average wind speed, $\{S_{solar}^{d|test}\}$ is the testing data set of daily average solar radiation, $\{Q_{load}^{d-1|test}\}$ is the testing data set of daily average heat load from yesterday, $\{Q_{load}^{d|test}\}$ is the testing data set of daily average heat load from today. Where $d = 1, 2, \dots, 31$.

The given data from the database can be converted to the testing data sets like Eq. (18), Eq. (19), Eq. (20), Eq. (21) and Eq. (22). Because of the whole regression coefficients A , B , C , D and E (Eq. (12)) have been identified by 1stOpt with the LM algorithm, so the 5 heat load prediction models have been established and the corresponding 5 multiple regression function equations have been also obtained. And then, the whole testing data sets can be put into these 5 multiple regression function equations so as to calculate the output of these 5 heat load prediction models.

4.3. Model Evaluation

In order to measure these 5 heat load prediction models' accuracy, it's necessary to have some performance indexes to evaluate these 5 models' accuracy. In this research, the square correlation coefficient R^2 and mean absolute percentage error (MAPE) are introduced into the corresponding model evaluation. The corresponding calculation equations of the R^2 and MAPE can be determined as follows:

$$R^2 = \frac{\sum_{d=1}^{31} (Q_{load}^{d|measured} - \bar{Q}_{load}^{d|measured})(Q_{load}^{d|estimated} - \bar{Q}_{load}^{d|estimated})}{\sum_{d=1}^{31} (Q_{load}^{d|measured} - \bar{Q}_{load}^{d|measured})^2 \sum_{d=1}^{31} (Q_{load}^{d|estimated} - \bar{Q}_{load}^{d|estimated})^2} \quad (23)$$

$$MAPE = \frac{\sum_{d=1}^{31} \left| \frac{Q_{load}^{d|measured} - Q_{load}^{d|estimated}}{Q_{load}^{d|measured}} \right|}{31} \times 100\% \quad (24)$$

where $d = 1, 2, \dots, 31$, d presents the day index, it is the day's number. $Q_{load}^{d|measured}$ is the measured value of the heat load, $\bar{Q}_{load}^{d|measured}$ is the average measured value of the heat

load during the 31 days; $Q_{load}^{d|estimated}$ is the estimated value of the heat load, $\bar{Q}_{load}^{d|estimated}$ is the average estimated value of the heat load during the 31 days.

When the square correlation coefficient R^2 is close to 1, it's proved that the heat load's measured value is fitted better by the heat load's estimated value, and the model's accuracy is better; at the same time, when the MAPE is close to 0%, it's proved that the heat load's measured value is also fitted better by the heat load's estimated value, and the model's accuracy is also better.

5. RESULTS AND DISCUSSION

In this study, for the selected DHS in Dalian city of China, only 5 practical equipments of the DHS were selected as the experimental objects. These 5 experimental objects include the heat exchanger (heat exchange station) A, heat exchanger (heat exchange station) B, heat user (building) A, heat user (building) B and heat user (building) C (Fig. 1). In this paper, the corresponding heat load prediction models of these 5 experimental objects can be defined as the model 1, model 2, model 3, model 4 and model 5, respectively (Fig. 1).

In this paper, for the database of the IOT heating information network platform, the data of the 2014-2015 heating season was selected in this study. The 2014-2015 heating season of Dalian in China includes 5 months: November (2014), December (2014), January (2015), February (2015) and March (2015). The data sets of January (2015) includes 31 days were selected as the training data sets for the 5 heat load prediction models; the data sets of March (2015) includes 31 days were selected as the testing data sets for the 5 heat load prediction models.

Based on the given training data sets and the multiple regression function Eq. (12), and using the LM algorithm with the 1st Opt data fitting software, the corresponding regression coefficients A , B , C , D and E of the 5 heat load prediction models can be also identified. The corresponding training results of the 5 heat load prediction models are shown in Table 5.

As can be seen from Table 5, the regression coefficients A , B , C , D and E of these 5 models are identified, the whole training results of these 5 models are very good. For the corresponding training results of these 5 models, the square correlation coefficient R^2 between the heat load's measured value and estimated value are all greater than 0.9000, and the MAPE between the heat load's measured value and estimated value are all less than 3.50%. The training experimental results shown that the established 5 heat load prediction models are reasonable.

And then, based on the identified regression coefficients A , B , C , D and E , so the 5 heat load prediction models have been established. These 5 heat load prediction models can be described by the corresponding multiple regression

Table 5. The training results of the 5 heat load prediction models in the modelling process

$$Q_{load}^d = A + B \cdot T_{out}^d + C \cdot W_{wind}^d + D \cdot S_{solar}^d + E \cdot Q_{load}^{d-1}$$

Model Number	Regression Coefficients					Error	
	A	B	C	D	E	R ²	MAPE
Model 1	0.5273	-0.0035	0.0105	-6.5987e-5	0.3579	0.9064	0.94%
Model 2	0.4418	-0.0087	0.0231	-2.2029e-4	0.5582	0.9132	3.07%
Model 3	0.0068	3.1248e-5	0.0014	-3.1434e-5	0.9560	0.9539	2.89%
Model 4	-0.0020	-6.7021e-5	-0.0036	7.7381e-5	0.9841	0.9623	3.21%
Model 5	0.0403	-0.0008	0.0054	-5.9781e-5	0.8453	0.9589	2.64%

Table 6. The testing results of the 5 heat load prediction models in the measuring process.

Model Number	Regression Function Equation	Error	
		R ²	MAPE
Model 1	$Q_{load}^d = 0.5273 - 0.0035T_{out}^d + 0.0105W_{wind}^d - 6.5987e - 5S_{solar}^d + 0.3579Q_{load}^{d-1}$	0.9449	1.75%
Model 2	$Q_{load}^d = 0.4418 - 0.0087T_{out}^d + 0.0231W_{wind}^d - 2.2029e - 4S_{solar}^d + 0.5582Q_{load}^{d-1}$	0.9015	2.70%
Model 3	$Q_{load}^d = 0.0068 + 3.1248e - 5T_{out}^d + 0.0014W_{wind}^d - 3.1434e - 5S_{solar}^d + 0.9560Q_{load}^{d-1}$	0.9072	3.09%
Model 4	$Q_{load}^d = -0.0020 - 6.7021e - 5T_{out}^d - 0.0036W_{wind}^d + 7.7381e - 5S_{solar}^d + 0.9841Q_{load}^{d-1}$	0.9649	3.16%
Model 5	$Q_{load}^d = 0.0403 - 0.0008T_{out}^d + 0.0054W_{wind}^d - 5.9781e - 5S_{solar}^d + 0.8453Q_{load}^{d-1}$	0.9483	3.76%

function equations such as the Eq. (12) proposed in this paper.

When these 5 heat load prediction models have been established, and the testing data sets are known, so it's necessary to test these 5 models' accuracy.

In this study, the corresponding testing data sets range from 1st day to 31st day during March (2015). As can be seen from Eq. (18), Eq. (19), Eq. (20) and Eq. (21), the testing data sets including $\{T_{out}^{d|test}\}$, $\{W_{wind}^{d|test}\}$, $\{S_{solar}^{d|test}\}$ and $\{Q_{load}^{d-1|test}\}$ can be put into the multiple regression function equation such as the Eq. (12) proposed in this paper, so the daily average heat load from 1st day to 31st day during March (2015) can be calculated., therefore, the estimated value of the 5 heat load prediction models for testing can be acquired. Because of the testing data set $\{Q_{load}^{d|test}\}$ is obtained, so the measured value of the 5 heat load prediction models for testing is also known. The corresponding testing results of the 5 heat load prediction models are shown in Table 6.

As can be seen from Table 6, the whole testing results of these 5 models are very good. For the corresponding testing results of these 5 models, the square correlation coefficient R² between the heat load's measured value and estimated value are all greater than 0.9000, and the MAPE between the

heat load's measured value and estimated value are all less than 4.00%. The testing experimental results shown that the established 5 heat load prediction models are reasonable. Moreover, for the corresponding testing results of these 5 models, the corresponding data fitting curves between the heat load's measured value and estimated value for the 5 models' testing are shown in Figs. (7, 9, 11, 13 and 15); the corresponding relative errors variations between the heat load's measured value and estimated value for the 5 models' testing are shown in Figs. (8, 10, 12, 14 and 16).

As can be seen from Figs. (7, 9, 11, 13 and 15), the corresponding heat load's measured value and estimated value are all the daily average heat load, for the testing of the 5 models, the corresponding daily average heat load is ranging from 1st day to 31st day during March (2015). The data fitting results between the heat load's measured value and estimated value for the 5 models' testing are very well. As can be seen from Figs. (8, 10, 12, 14 and 16), the relative errors variations between the heat load's measured value and estimated value for the 5 models' testing are also reasonable, the maximum absolute relative errors between the heat load's measured value and estimated value for the 5 models' testing are less than 8%.

In order to have the better analysis and discussion on the experimental results, it's necessary to summarize the training and testing results of the 5 heat load prediction models. The corresponding maximum and minimum relative errors results of these 5 models' training and testing are shown in Table 7.

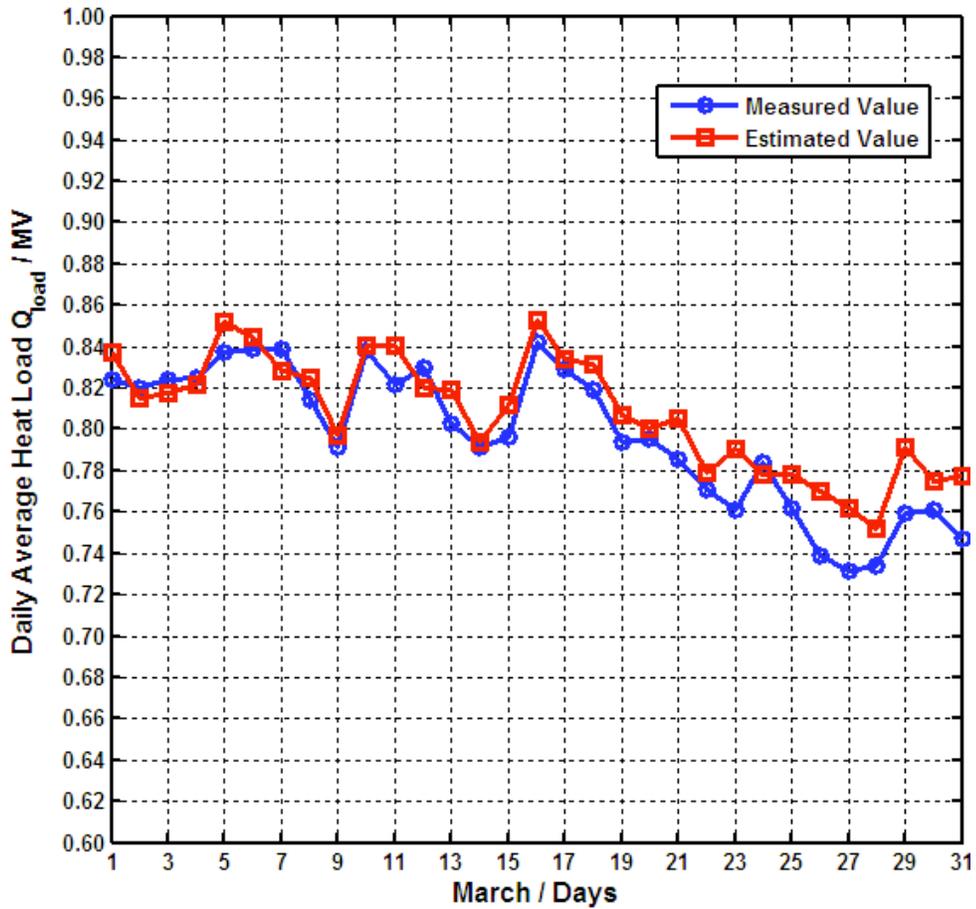


Fig. (7). The data fitting curves between the heat load's measured value and estimated value for the model 1's testing.

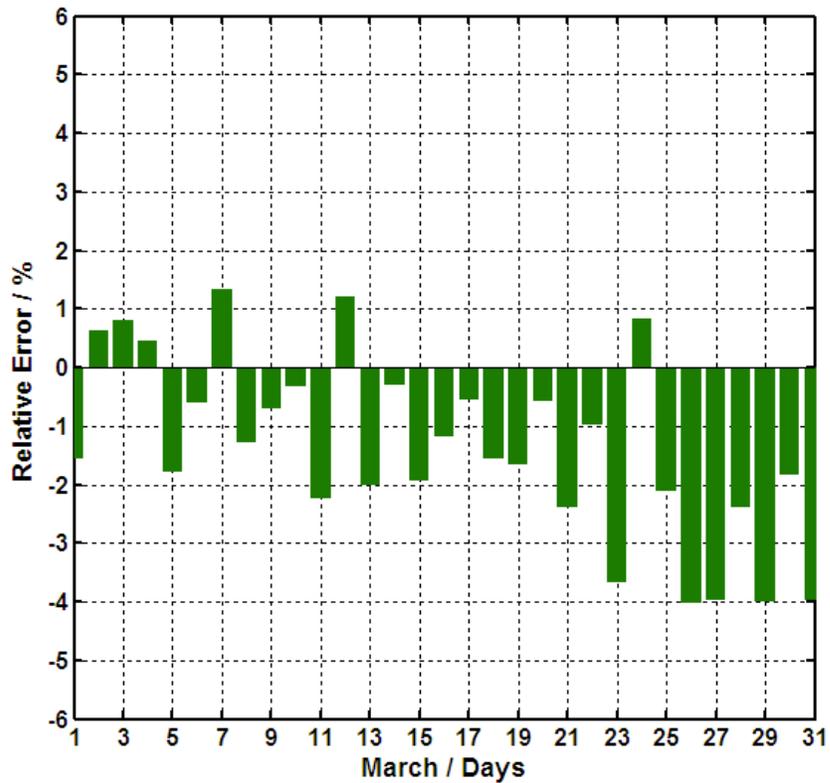


Fig. (8). The relative errors variations between the heat load's measured value and estimated value for the model 1's testing.

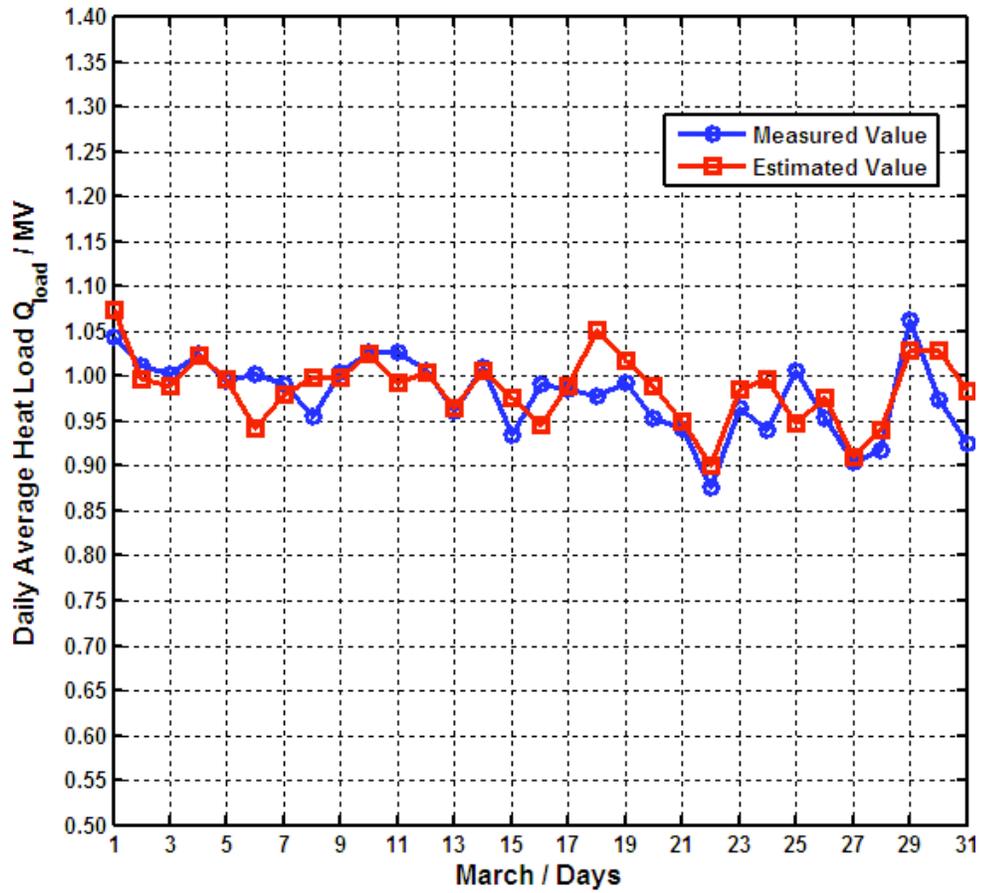


Fig. (9). The data fitting curves between the heat load's measured value and estimated value for the model 2's testing.

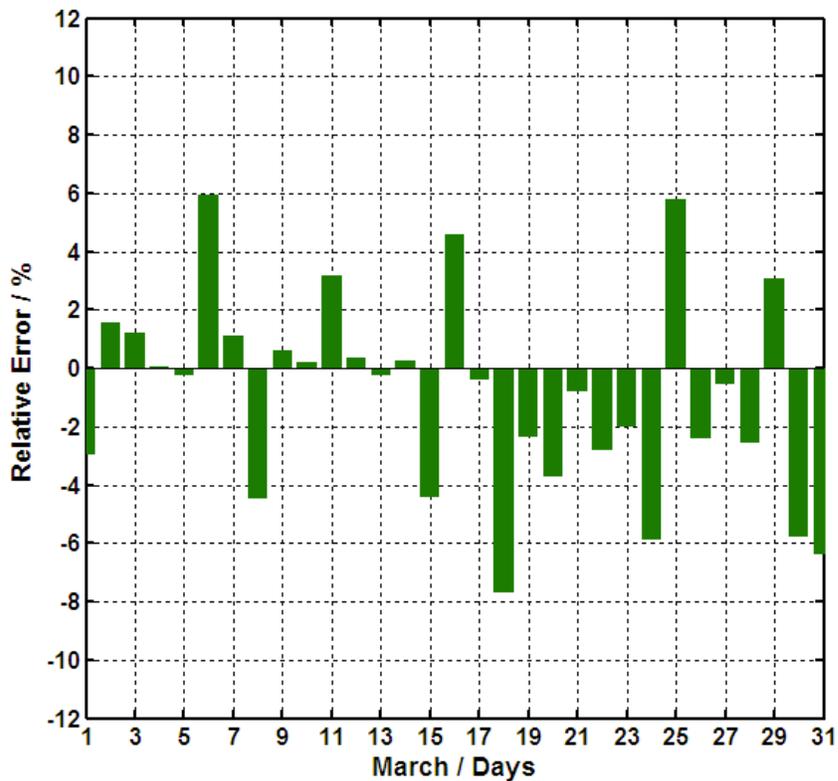


Fig. (10). The relative errors variations between the heat load's measured value and estimated value for the model 2's testing.

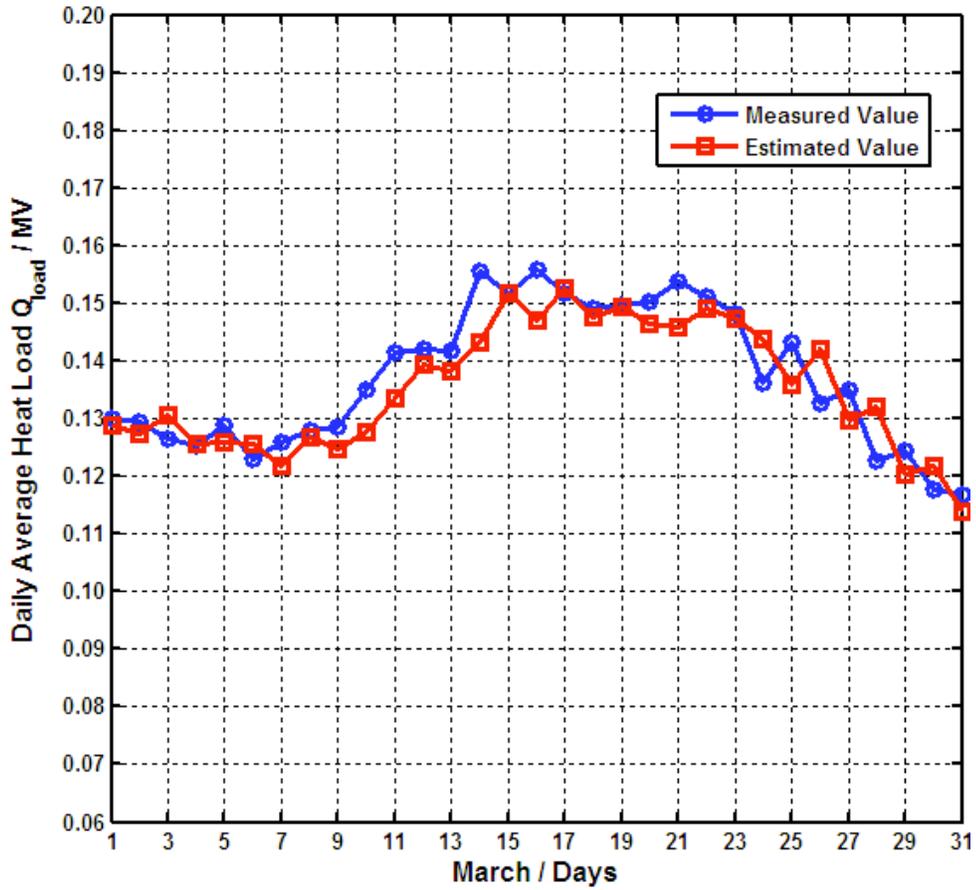


Fig. (11). The data fitting curves between the heat load’s measured value and estimated value for the model 3’s testing.

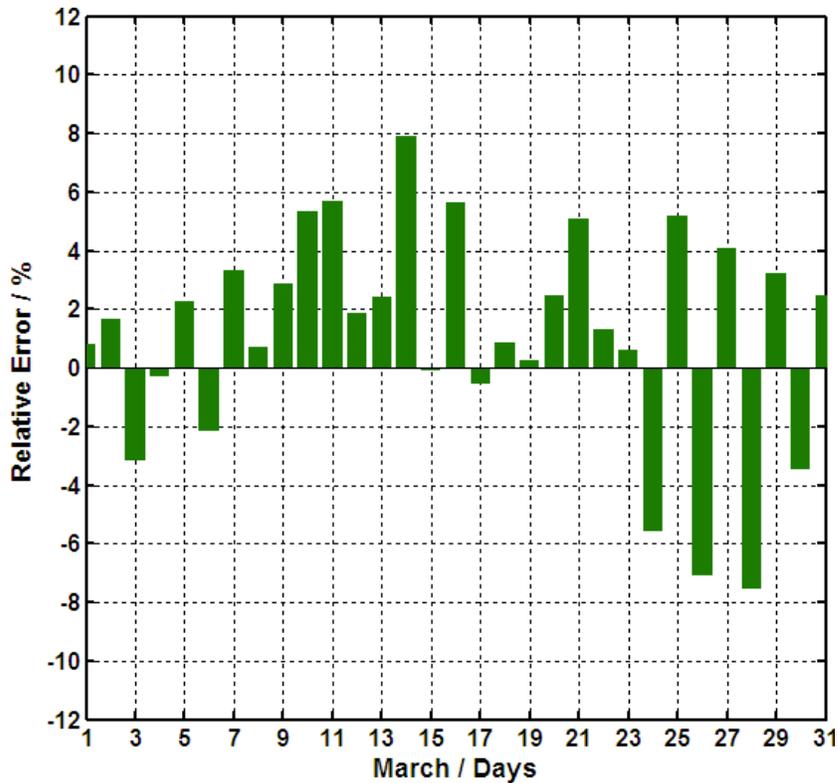


Fig. (12). The relative errors variations between the heat load’s measured value and estimated value for the model 3’s testing.

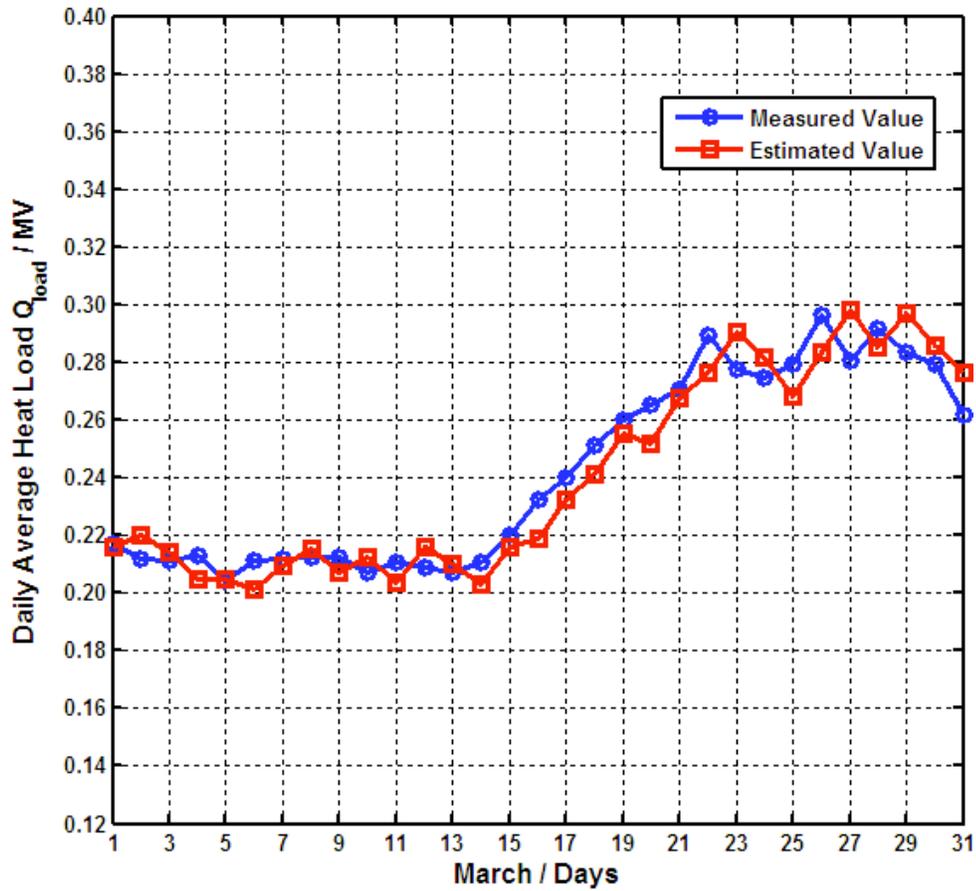


Fig. (13). The data fitting curves between the heat load's measured value and estimated value for the model 4's testing.

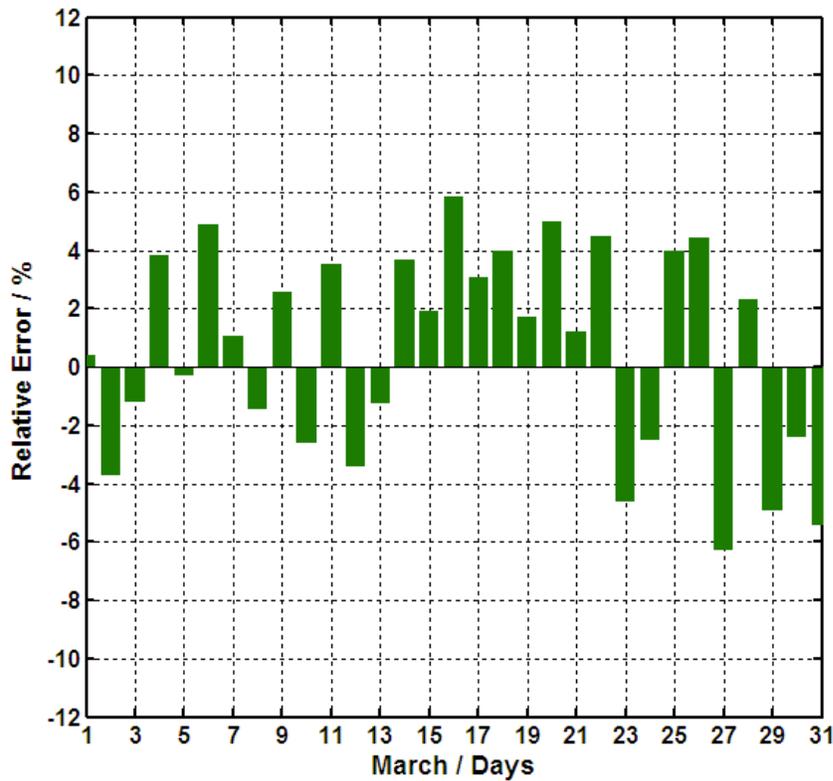


Fig. (14). The relative errors variations between the heat load's measured value and estimated value for the model 4's testing.

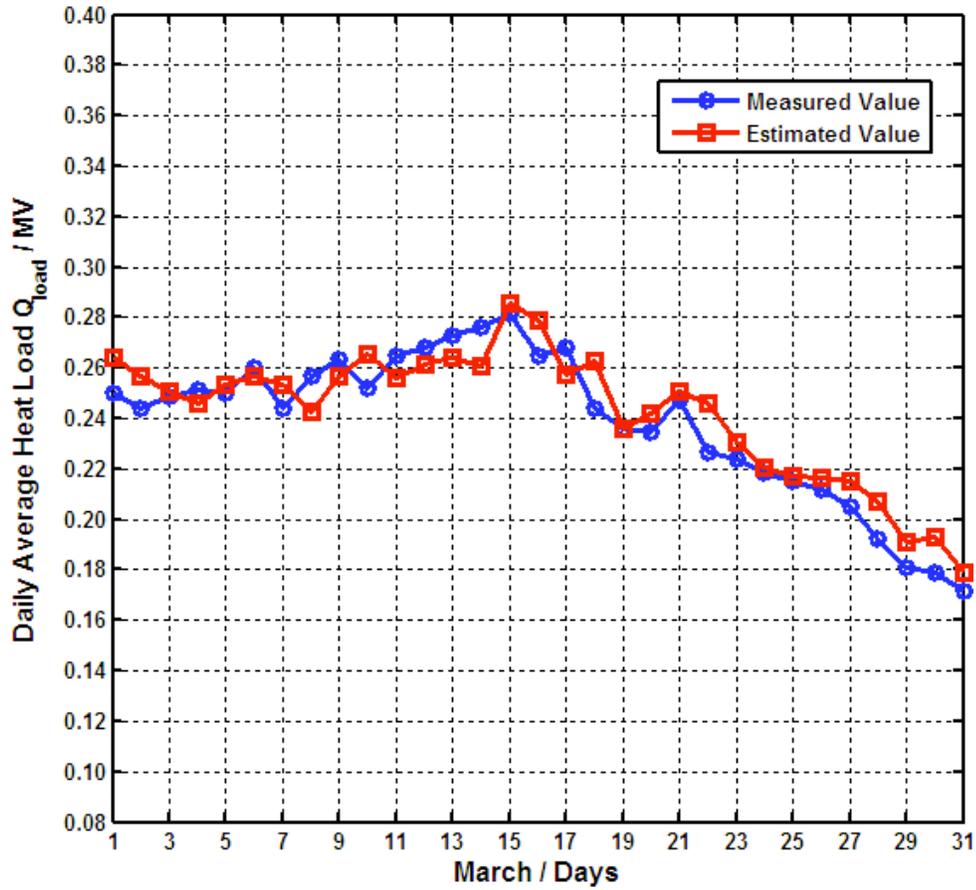


Fig. (15). The data fitting curves between the heat load’s measured value and estimated value for the model 5’s testing.

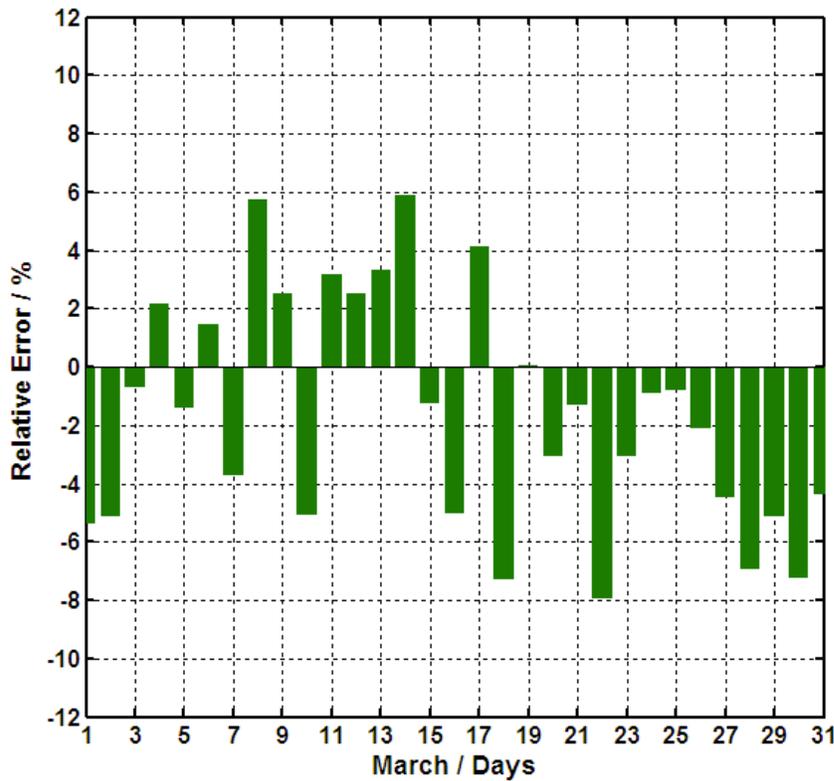


Fig. (16). The relative errors variations between the heat load’s measured value and estimated value for the model 5’s testing.

Table 7. The maximum and minimum relative errors results of the 5 models' training and testing.

Model Number	Model Training		Model Testing	
	Maximum Relative Error	Minimum Relative Error	Maximum Relative Error	Minimum Relative Error
Model 1	2.74%	-2.27%	1.32%	-4.21%
Model 2	7.98%	-7.82%	5.94%	-7.69%
Model 3	5.72%	-7.78%	7.90%	-7.54%
Model 4	6.44%	-7.73%	5.22%	-6.28%
Model 5	5.40%	-6.22%	5.91%	-7.93%

As can be seen from Table 7, for these 5 heat load prediction models' training and testing, the maximum relative errors between the heat load's measured value and estimated value are all less than 8%, the minimum relative errors between the heat load's measured value and estimated value are all greater than 8%. These experimental results shown that for the 5 models' training and testing, the corresponding maximum absolute relative errors between the heat load's measured value and estimated value are all less than 8%.

Moreover, considering the experimental results of Tables 5-7, for the 5 models' training and testing, the square correlation coefficient R^2 between the heat load's measured value and estimated value are all greater than 0.9000, and the MAPE between the heat load's measured value and estimated value are all less than 4.00%. These experimental results also indicated that the 5 heat load prediction models' accuracy is relatively high, and these 5 heat load prediction models can be applied in the real heating engineering, the most importantly, this multiple regression modelling and analysis method can be promoted into the practical engineering applications.

CONCLUSION

In summary, it can be concluded that the proposed heat load prediction models presented promising features to be easy and efficient heat load forecast tools for the DHS.

The results shown that the heat load of the DHS can be really influenced by the outdoor temperature. Moreover, the wind speed variable and solar radiation variable were corrected to the equivalent outdoor temperature variable so as to be added into the real outdoor temperature variable, and the results shown that the local wind speed and solar radiation have the combined influences on the local outdoor temperature so as to impact the heat load of the DHS.

The training and testing results shown that the square correlation coefficient R^2 between the heat load's measured value and estimated value are all greater than 0.9000, and the MAPE between the heat load's measured value and estimated value are all less than 4.00%. Moreover, the training and testing results also shown that the corresponding maximum absolute relative errors between the heat load's measured value and estimated value are all less than 8%. These experimental results also indicated that the 5 heat load prediction models' accuracy is relatively high, and these 5 heat

load prediction models can be applied in the real DHS engineering. In a word, this multiple regression modelling and analysis method can be promoted into the practical engineering applications.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The authors would like to thank "The Fundamental Research Funds for The Central Universities" for Dalian Maritime University's financial supports. The authors also would like to thank "The Science and Technology Plan of Dalian City in 2014" of China with Grant No. 2014E11SF059.

REFERENCES

- [1] U. Persson and S. Werner, "Heat distribution and the future competitiveness of district heating," *Applied Energy*, vol. 88, pp. 568-576, 2011.
- [2] X. Xu, S. J. You, X. J. Zheng, and H. Li, "A survey of district heating systems in the heating regions of northern China," *Energy*, vol. 77, pp. 909-925, 2014.
- [3] P. F. Jie, Z. Tian, S. S. Yuan, and N. Zhu, "Modeling the dynamic characteristics of a district heating network," *Energy*, vol. 39, pp. 126-134, 2012.
- [4] F. T. Sun, L. Fu, S. G. Zhang, and J. Sun, "New waste heat district heating system with combined heat and power based on absorption heat exchange cycle in China," *Applied Thermal Engineering*, vol. 37, pp. 136-144, 2012.
- [5] S. Pusat and H. H. Erdem, "Techno-economic model for district heating systems," *Energy and Buildings*, vol. 72, pp. 177-185, 2014.
- [6] N. Yıldırım, M. Toksoy, and G. Gökçen, "District heating system design for a university campus," *Energy and Buildings*, vol. 38, pp. 1111-1119, 2006.
- [7] I. Gabrielaitiene, B. Bøhm, and B. Sunden, "Modelling temperature dynamics of a district heating system in Naestved, Denmark—A case study," *Energy Conversion and Management*, vol. 48, pp. 78-86, 2007.
- [8] A. Hepbasli and A. Keçebaş, "A comparative study on conventional and advanced exergetic analyses of geothermal district heating systems based on actual operational data," *Energy and Buildings*, vol. 61, pp. 193-201, 2013.
- [9] S. Werner, "The heat load in district heating systems," Chalmers University of Technology, Gothenburg, 1984 (Ph.D. Thesis).
- [10] F. S. Westphal and R. Lamberts, "The use of simplified weather data to estimate thermal loads of non-residential buildings," *Energy and Buildings*, vol. 36, pp. 847-854, 2004.

- [11] K. Wojdyga, "An influence of weather conditions on heat demand in district heating systems," *Energy and Buildings*, vol. 40, pp. 2009-2014, 2008.
- [12] O. Yetemen and T. Yalcin, "Climatic parameters and evaluation of energy consumption of the Afyon geothermal district heating system, Afyon, Turkey," *Renewable Energy*, vol. 34, pp. 706-710, 2009.
- [13] R. Mu and M. D. Jong, "Establishing the conditions for effective transit-oriented development in China: the case of Dalian," *Journal of Transport Geography*, vol. 24, pp. 234-249, 2012.
- [14] Y. Zhao, B. Chen, Y. L. Guo, F. F. Peng, and J. L. Zhao, "Indoor air environment of residential buildings in Dalian, China," *Energy and Buildings*, vol. 36, pp. 1235-1239, 2004.
- [15] I. Vallios, T. Tsoutsos, and G. Papadakis, "Design of biomass district heating systems," *Biomass and Bioenergy*, vol. 33, pp. 659-678, 2009.
- [16] S. J. Zhou, "Operational parameters prediction and optimization research of district heating system based on pipe network dynamic model," Shandong University, Jinan, 2012 (Ph.D. Thesis in Chinese).
- [17] D. H. Xu and R. Zhu, "The sensation of human to temperature, humidity, wind speed and the clothing," *Journal of Applied Meteorology*, vol. 11, pp. 430-439, 2000 (in Chinese).
- [18] 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.
- [19] Z. Y. Shi, *Operation regulation and control of heating system*, Beijing: Tsinghua University Press, 1994 (in Chinese).
- [20] K. Mahapatra and L. Gustavsson, "Influencing Swedish homeowners to adopt district heating system," *Applied Energy*, vol. 86, pp. 144-154, 2009.
- [21] Y. Cho and H. B. Awbi, "A study of the effect of heat source location in a ventilated room using multiple regression analysis," *Building and Environment*, vol. 42, pp. 2072-2082, 2007.
- [22] T. Catalina, J. Virgone, and E. Blanco, "Development and validation of regression models to predict monthly heating demand for residential buildings," *Energy and Buildings*, vol. 40, pp. 1825-1832, 2008.
- [23] I. Korolija, Y. Zhang, L. M. Halburd, and V. I. Hanby, "Regression models for predicting UK office building energy consumption from heating and cooling demands," *Energy and Buildings*, vol. 59, pp. 214-227, 2013.
- [24] X. Y. Cheng, F. X. Chai, J. Gao, and K. L. Zhang, "1stOpt and global optimization platform—comparison and case study," in: Proceedings of the 4th IEEE International Conference on Computer Science and Information Technology, Chengdu, China, 2011.

Received: June 10, 2015

Revised: July 29, 2015

Accepted: August 15, 2015

© Cai *et al.*; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.