

## Research Article

# An Estimation Model on Electricity Consumption of New Metro Stations

Zhao Yu,<sup>1,2</sup> Yun Bai ,<sup>1</sup> Qian Fu ,<sup>3</sup> Yao Chen ,<sup>1,4</sup> and Baohua Mao <sup>1</sup>

<sup>1</sup>Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, No. 3, Shangyuancun, Haidian District, Beijing 100044, China

<sup>2</sup>Guangzhou Transport Planning Research Institute, No. 1, Guangren Road, Yuexiu District, Guangzhou 510030, China

<sup>3</sup>Birmingham Centre for Railway Research and Education, School of Engineering, University of Birmingham, Birmingham B15 2TT, UK

<sup>4</sup>Department of Industrial Systems Engineering and Management, National University of Singapore, 1 Engineering Drive 2, Singapore 117576, Singapore

Correspondence should be addressed to Yun Bai; [yunbai@bjtu.edu.cn](mailto:yunbai@bjtu.edu.cn)

Received 6 May 2019; Accepted 29 October 2019; Published 30 January 2020

Guest Editor: Morris Brenna

Copyright © 2020 Zhao Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Electricity consumption of metro stations increases sharply with expansion of a metro network and this has been a growing cause for concern. Based on relevant historical data from existing metro stations, this paper proposes a support vector regression (SVR) model to estimate daily electricity consumption of a newly constructed metro station. The model considers some major factors influencing the electricity consumption of metro station in terms of both the interior design scheme of a station (e.g., layout of the station and allocation of facilities) and external factors (e.g., passenger volume, air temperature and relative humidity). A genetic algorithm with five-fold cross-validation is used to optimize the hyper-parameters of the SVR model in order to improve its accuracy in estimating the electricity consumption of a metro station (ECMS). With the optimized hyper-parameters, results from case studies on the Beijing Subway showed that the estimating accuracy of the proposed SVR model could reach up to 95% and the correlation coefficient was 0.89. It was demonstrated that the proposed model could outperform the traditional methods which use back-propagation neural network or multivariate linear regression. The method presented in this paper can be an adequate tool for estimating the ECMS and should further assist in the delivery of new, energy-efficient metro stations.

## 1. Introduction

A metro system plays an important role among urban mass transit systems and has a number of advantages over other public transportation modes in metropolitan areas, such as having more reliable services, being able to transport much larger volume of passengers, and being more environmentally friendly. In China, metro networks have exploded in recent decades as the population has rapidly urbanized nationwide. For example, the total length of the metro network in Beijing has reached about 591.7 km, connecting a total of 361 stations since 2017; by the end of 2020, it is projected that the network will be expanded to over 900 km, which expects to accommodate 552 stations in total.

Although the metro is one of the most energy-efficient transportation modes, the electricity consumption of a metro system rises significantly with the continuous increase in its

operation mileage. Data from the Beijing Subway shows that the whole electricity consumption of all lines added to 1.71 billion kWh in 2016, nearly three times the amount in 2010. Clearly, it has given a serious cause for concern that the overall level of the electricity consumption of the metro system will continue to go up as its network keeps expanding. Furthermore, according to statistical data derived from the Beijing Subway, the electricity consumption of a metro station (ECMS) has taken up approximately half of the total electricity consumption of an entire metro system. To reduce the ECMS is therefore of great significance to cut down the whole electricity consumption of a metro system [1].

The ECMS describes the full amount of electricity consumed within a metro station, involving all of its subsystems such as HVAC (heating, ventilation and air-conditioning), lighting, and other facilities (e.g., platform screen doors and escalators) [2], which ensures the station's normal operation.

The bulk of the ECMS is due to the use of HVAC equipment. For instance, about two thirds of the ECMS during summers is from the HVAC for most Beijing metro stations [3]. In an attempt to lower the electricity consumption by HVAC, an autonomous control system has been developed by Wang et al. [4], which could adapt cooling supply to variation of heating load. Installations of platform screen doors may help prevent additional heat brought by trains' movements and so, to some extent, lower the level of heating load within the station [5]. Apart from the HVAC, the lighting subsystem also accounts for a good portion of the ECMS [6]. Casals et al. [7] developed an adaptive, energy-saving lighting subsystem, which could adjust the illumination intensity according to passenger volume. Further, Casals et al. [8] suggested an intelligent energy management system, which integrates the lighting, ventilation and vertical transportation subsystems, so as to deliver an integrated energy saving for a metro station.

Despite the above-mentioned facilities and equipment, how the layout of a station would impact on the ECMS has not been explicitly considered in the existing studies. It turns out that the ECMS may differ markedly across stations, given similar passenger volume and the same facilities. In other words, the ECMS depends largely on the design scheme, especially the spatial structure of the station. Therefore, it is important to establish an estimation model on ECMS considering the design of metro stations, which is an essential tool for the proper design of metro stations for energy saving.

Energy audit and identifying the critical influencing factors on ECMS are the foundation of developing estimation models on ECMS. Fu and Deng [9] analysed the practical data of the energy consumption in Guangzhou railway stations and proposed a series of methods to save the energy consumed by air-conditioning system, power equipment system and lighting system. Hong and Kim [2] investigated the energy consumption of subway stations in Korean, and explored the relationship between energy consumption of station and its influencing factors. These studies pointed out the major factors influencing ECMS and revealed that the relationships between the ECMS and most factors are nonlinear. However, these studies do not serve the purpose of predicting ECMS during the station design stage to assist the energy evaluation of design schemes of a metro station.

In recent years, a series of linear models have been proposed for estimating energy consumed by stations, assuming a linear relationship between the energy consumption and its influencing factors. A few examples are as follows. Yang et al. [10] tried to identify the most important factors influencing the ECMS through correlation analysis and proposed a regression model for estimating the electricity consumption. Wang et al. [11] proposed a linear calculating model to forecast the trend of energy consumption of a given metro network, based on the indicator of monthly energy consumption. Guan et al. [12] applied a multivariate linear regression (MLR) model to analyse the contribution of different factors such as floor area of station and passenger volume to the ECMS. Ahn et al. [13] built a linear regression model to assess existing subway stations performance and predict energy consumption levels for future expansion. In fact, however, these models would be less

likely to achieve a high accuracy in predicting the ECMS, as they may fail to capture the underlying nonlinearity in the relationship between the ECMS and its influencing factors.

To improve the prediction accuracy, the nonlinear back-propagation neural network (BPNN) model has been applied to predict the ECMS based on historical data in the metro system of Hong Kong [14]. A single hidden layer BPNN can generally approximate any nonlinear function with arbitrary precision [15]. Because of the strong nonlinear mapping ability, BPNN is very popular to predict the energy consumption of buildings and other systems. Ekici and Aksoy [16] used BPNN to predict the heating energy demand of three different buildings. Yokoyama et al. [17] developed a BPNN model to predict the cooling demand of a building. However, the performance of BPNN models in prediction problems is governed by the quality and quantity of training samples. Over-fitting the data and getting stuck at some local minima, which might largely reduce the prediction accuracy, are also common issues in practical applications of the BPNN models [18].

Another issue stems from the size of data for model training. With respect to the prediction of ECMS, a metro network, which at the early stages of development may not provide sufficiently large data samples. A data sample of fairly small size would also render the BPNN model ineffective [19]. Support Vector Machine (SVM), minimizing the structure risk including training error and model complexity to achieve good generalization, is a novel machine learning method for classification based on statistical learning theory [20, 21]. The principle of SVM has also been applied to regression models called Support Vector Regression (SVR) to mimic the nonlinear relationship between the variables and results. SVR has been widely applied in energy predictions [22–25], because of its ability of nonlinear approximation and dealing with multiple inputs and small samples. Thus, SVR is considered as a promising tool to predict the ECMS during station design stage based on the limited historical data.

In this study, a SVR-based model is developed to predict the daily ECMS, considering both the internal (e.g., layout of a station) and external factors with respect to operating a station. Given the fact that hyper-parameters of the model may have significant impact on its prediction accuracy [26], a genetic algorithm (GA) with  $k$ -fold cross-validation is applied in order to optimize the hyper-parameters of the SVR model. Data collected from the Beijing metro is used to demonstrate the model performance. A comparison of the proposed method with other alternatives, including both BPNN and MLR models, is also made. This study aims to provide an adequate tool for practitioners to evaluate the performance of different station design schemes in terms of electricity consumption.

The remainder of the paper is organized as follows. Section 2 analyses the factors influencing the ECMS and determines the input variables for the SVR model. Section 3 elaborates on the fundamentals of the SVR model and the method of optimizing its associated hyper-parameters with GA. Using the proposed model, Section 4 describes a case study example of estimating daily ECMS on the Beijing metro and evaluates the model performance. Finally, Section 5 summarizes the study and concludes the paper.

TABLE 1: Summary of input variables.

Classifications	Influencing factors	Input variables	Unit	
The interior design scheme of metro station	Concourse scale	The area of concourse ( $X^1$ )	$m^2$	
	Platform scale	The area of platform ( $X^2$ )	$m^2$	
	Plant room scale	The area of plant room ( $X^3$ )	$m^2$	
	Staff accommodation room scale	The area of staff accommodation room ( $X^4$ )	$m^2$	
	VT facilities		The quantity of VT facilities ( $X^5$ )	—
			The height of VT facilities ( $X^6$ )	m
External factors	Temperature	Average temperature ( $X^7$ )	$^{\circ}C$	
	Humidity	Relative humidity ( $X^8$ )	%	
	Passenger demand	The total number of passengers ( $X^9$ )	Person/day	

## 2. Factors Influencing the ECMS

In this section, some major factors influencing the ECMS are described in detail. These factors, which serve as input variables in the proposed model, can be generally categorized into two groups: 1) factors relating to the interior design scheme and 2) other external factors to consider.

### 2.1. Interior Design Scheme of Metro Station

*2.1.1. Floor Area of Different Zones Station.* Previous studies (e.g., [11, 12]) have shown that the scale of a metro station could have significant impact on ECMS. In general, a metro station can broadly be divided into four zones: concourse, platform, plant room and staff accommodation room. In these different zones, HVAC, lighting and other facilities are installed to provide a safe, accessible and comfortable environment for passengers and staff. What models and how many of the HVAC and lighting facilities must be installed are largely dependent on the layout of the station and the area of these zones. Namely, cooling and lighting load in the metro station are both related to its structural system as a whole. In this regard, the actual area of the four different zones would influence the ECMS and should therefore be considered as input variables in model specification.

*2.1.2. Auxiliary Facilities.* In metro stations, escalators and elevators are equipped to improve the quality or security of the service during operating periods. Electrical power consumption by these vertical transportation facilities is related to their quantity and height, which will also be included in the model specification as input variables.

As mentioned above, platform screen doors installed on platforms may also be related to the ECMS, as they could effectively minimize additional heat effect from the underground tunnels when being fully enclosed at the platform edge. This factor will not be considered in this since most of the subway lines in Beijing are equipped with screen doors.

*2.2. External Factors.* Weather is a key factor influencing the cooling load of centralized air-conditioning and ventilation subsystems of the metro stations. Air from outside the stations could bring in a certain amount of heat and moisture, which increases the cooling load of the air-conditioning subsystem.

In this regard, relative humidity and temperature of the outdoor air should be taken into account as input variables of prediction model.

In addition to weather, passenger-flow volume is another major external factor that contribute to the ECMS. Internal heat of a metro station builds up as more and more passengers enter the station. Therefore, the total number of the passengers entering or leaving a metro station should also be taken as an input variable for specifying the model.

As discussed above, the input variables of the ECMS prediction model are listed in Table 1.

## 3. Development of the SVR Model

This section describes the development of a SVR-based model for estimating the ECMS, with its three hyper-parameters (denoted by  $c$ ,  $\epsilon$  and  $\sigma^2$ ) being optimized by GA.

*3.1. Introduction of SVR.* Let  $x_j$  denote a vector consisting of all normalized input variables, and  $y_j$  denote normalized value of the ECMS in the  $j$ -th sample dataset. Suppose the sample size of the dataset is  $J$ . The dataset can be defined as  $\{(x_j, y_j) | j = 1, 2, \dots, J\}$ . A SVR, which may be used to describe the nonlinear relationship between the input and output variables, can be expressed in the following form [27]:

$$f(x_i) = \omega \cdot \varphi(x_i) + b. \quad (1)$$

In Equaion (1),  $\varphi(x)$  denotes the high-dimensional feature space, which is nonlinearly mapped from the original input space  $x$ , while  $\omega$  and  $b$  are unknown parameters. According to [28], the unknown parameters can be estimated by minimizing the structure risk function as follows:

$$R_{SVR}(c) = R_{emp} + \frac{1}{2} \|\omega\|^2 = c \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \frac{1}{2} \|\omega\|^2, \quad (2)$$

where  $R_{SVR}$  and  $R_{emp}$  describe the structure risk and empirical risk, respectively;  $\|\omega\|^2/2$  describes a regularized term that control the confidence level and  $L_\epsilon(y_i, f(x_i))$  is the loss function. In addition, regularized constant  $c$  is a penalty parameter that determines the balance between the empirical risk and the regularized term. The first term of the right-hand side of

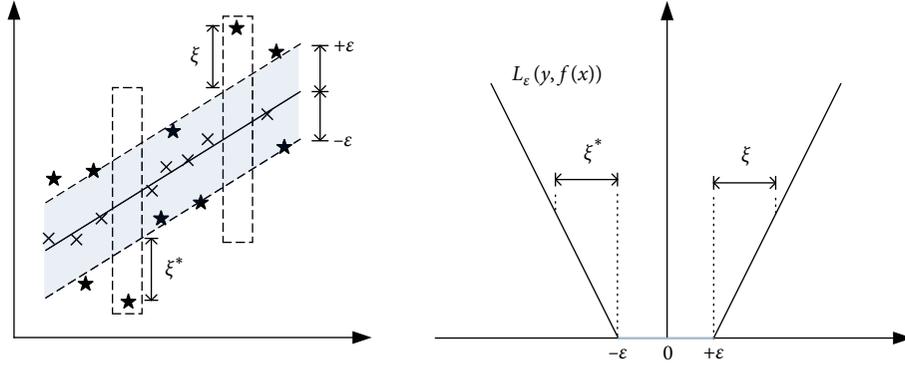
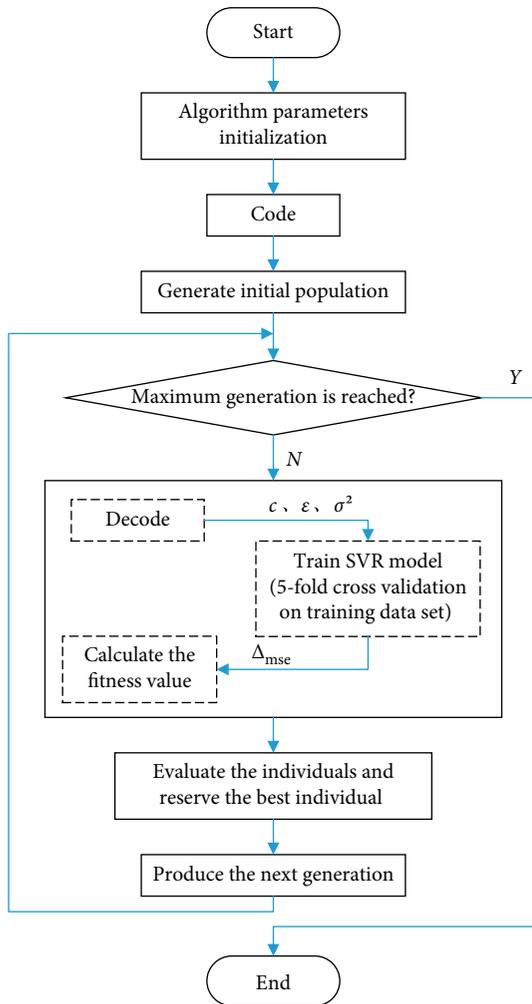
FIGURE 1: Graphical details of  $\varepsilon$ -insensitive loss function [29].

FIGURE 2: Flowchart of hyper-parameter optimization algorithm.

Equation (2) is measured by  $\varepsilon$ -insensitive loss function in  $\varepsilon$ -SVR [29], which is defined by Equation (3) as follows:

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The loss function defines a “tube” (see Figure 1), of which the tube size is denoted by  $\varepsilon$ . As shown in Figure 1, the value of

TABLE 2: GA parameter settings.

Population size	Maximum generation	Crossover probability	Mutation probability
65	60	0.9	0.2

1 1 0 0 0 0 1 1 0 0 0 0 1 0 0 1 0 1 0 1 0 0 1 1 1 0 1 0 1 0 1 0  
 ← c → \* ← ε → \* ← σ² → \*

FIGURE 3: An individual of the initial population.

the loss function would be zero if the predicted value is within the tube; otherwise, the value would be equal to the difference between prediction error and the radius of the tube. The regularized constant  $c$  and tube size  $\varepsilon$  are both user-prescribed parameters relying on empirical analysis.

By introducing the positive slack variables  $\xi_i$  and  $\xi_i^*$ , Equation (2) can be converted to the original objective function, which is formulated in Equation (4). As shown in Figure 1,  $\xi_i$  and  $\xi_i^*$  are the errors of the up and down sides, respectively, which are to be minimized so as to minimize the model training error.

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*), \\ \text{s.t.} \quad & \begin{cases} \omega\varphi(x_i) + b - y_i \leq \varepsilon + \xi_i, \\ y_i - \omega\varphi(x_i) - b \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0. \end{cases} \end{aligned} \quad (4)$$

The optimization problem formulated in Equation (4) can be solved in its dual formulation, where the constraints are handled by introducing Lagrange multipliers. The dual function is as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \varphi(x_i) \varphi(x_j), \\ & - \sum_{i=1}^n (\alpha_i^* - \alpha_i) y_i + \varepsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i), \\ \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \\ & \alpha_i, \alpha_i^* \in [0, c]. \end{aligned} \quad (5)$$

where both  $\alpha_i$  and  $\alpha_i^*$  are Lagrange multipliers. The dual problem formulated in Equation (5) is subject to the saddle point condition, which can be reduced to Equation (6).

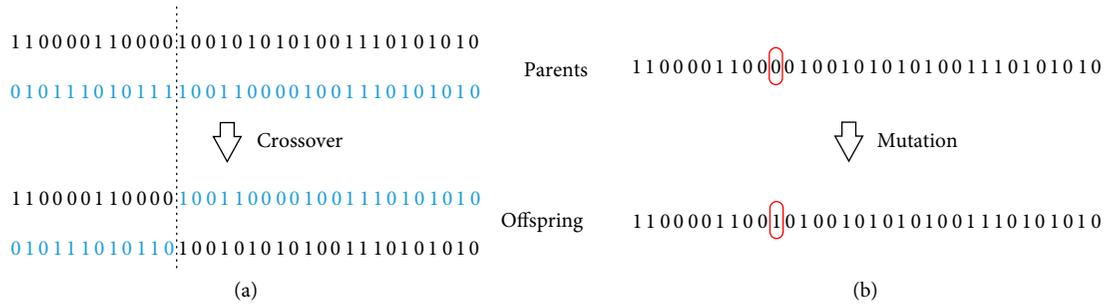


FIGURE 4: (a) Process of crossover, and (b) the process of mutation.

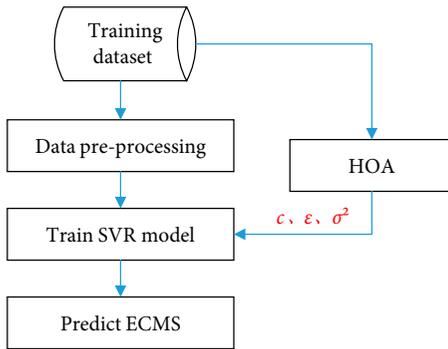


FIGURE 5: Process of building the SVR model to predict ECMS.

$$\omega = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(x_i). \quad (6)$$

Finally, by introducing a kernel function,  $K(x_i, x) = \varphi(x_i)\varphi(x)$  [27], and using Equation (6), the optimization problem (1) can be transformed into Equation (7) as follows.

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b. \quad (7)$$

Using the kernel function, a feature space of any dimension can be solved without calculating the map function  $\varphi(x)$  [28]. However, the kernel function can affect the prediction accuracy of a SVR model. In this study, a Gaussian kernel function is selected and used throughout, which is defined by Equation (8) as follows:

$$K(x_i, x) = \exp\left(\frac{-x_i - x^2}{\sigma^2}\right), \quad (8)$$

where  $\sigma^2$  describes the width of the Gaussian kernel. A Gaussian kernel has several advantages [30], it makes the model relatively simple due to having a single hyper-parameter and has less numerical difficulties, compared to others that may turn out to be invalid in some cases.

With selection of appropriate hyper-parameters,  $c$ ,  $\varepsilon$ , and  $\sigma^2$ , ECMS can be modelled as a function of its influencing factors through solving Lagrange multipliers,  $\alpha_i$  and  $\alpha_i^*$ , of the quadratic programming problem formulated above.

**3.2. Parameters Optimization Algorithm.** According to [26], the hyper-parameters  $c$ ,  $\varepsilon$  and  $\sigma^2$  have significant impact on

the performance of SVR. The hyper-parameter  $c$  determines training error and complexity of the SVR model. A fairly small value of  $c$  can give rise to a small penalty on the training error, thus resulting in the model under-fitting the training data. Whereas if  $c$  is too large, the generalization ability of the model would decrease. The hyper-parameter  $\varepsilon$  determines the width of the  $\varepsilon$ -insensitive zone. As the value of  $\varepsilon$  increases, the number of support vectors will decrease, in which case the resultant solution would be represented sparsely; however, if  $\varepsilon$  becomes too large, it will also decrease the approximation accuracy of the training data. The hyper-parameter  $\sigma^2$  specifies the structure of the high-dimensional feature space. Generally, prediction error of the model would not monotonically decrease with increasing the value of  $\sigma^2$  but increase when  $\sigma^2$  becomes too large.

Despite numerous studies looking into the optimization of the hyper-parameters, there is still a lack of metrics over which the set of these hyper-parameters would be the most suitable for the SVR model. The value of  $c$ ,  $\varepsilon$  and  $\sigma^2$  may be manually determined on an empirical basis (e.g., [20, 31]) or, often through grid search optimization (e.g., [22]). In this regard, this study employs GA, which has also been widely used in optimization problems [32], in an attempt to find an optimum set of the hyper-parameters given model performance. A flowchart of hyper-parameters optimization algorithm (HOA) is illustrated in Figure 2. The main steps of the proposed HOA are as follows.

*Step 1.* Initialize the GA parameters, including population size, crossover probability, mutation probability and maximum generation (see Table 2).

*Step 2.* Generate a random population. That is, all individuals of the initial population are randomly generated. Then, encode the values of the three hyper-parameters,  $c$ ,  $\varepsilon$  and  $\sigma^2$ , in one chromosome, given  $c \in [1, 1000]$ ,  $\varepsilon \in [0.0001, 0.1]$ ,  $\sigma^2 \in [0.001, 10]$ . A chromosome is equivalent to a combination of the hyper-parameters each coded as binary values (as illustrated in Figure 3).

*Step 3.* Repeat the evolutionary process from step (4) to step (6) until the maximum generation is reached. When the maximum generation is reached, the algorithm is terminated and the currently optimal solution is output.

*Step 4.* Calculate the fitness value of each individual. Firstly, the value of each hyper-parameter is obtained by decoding the chromosome. Then the SVR model is established with three hyper-parameters, and the sample data are applied to

TABLE 3: Partial data in training dataset.

Station ID	$X^1$ (m <sup>2</sup> )	$X^2$ (m <sup>2</sup> )	$X^3$ (m <sup>2</sup> )	$X^4$ (m <sup>2</sup> )	$X^5$	$X^6$ (m)	$X^7$ (°C)	$X^8$ (%)	$X^9$ (person/day)	$E$ (×10 <sup>6</sup> kWh)
I	975	3225	17234	8612	8	7	32	26	37803	9.990
II	872	4364	3488	1744	12	12	32	26	36510	9.769
III	656	2695	2879	1313	10	12	32	26	37695	13.149
IV	896	3498	15743	7570	5	9.2	32	26	35468	11.373
V	887	5317	3784	1623	12	12.3	32	26	38948	10.162
VI	635	3769	3562	1974	11	12.4	32	26	39683	10.965
VII	885	3191	17220	8587	8	7	32	26	36669	11.191
VIII	834	4336	3467	1682	12	12	32	26	37202	11.457
IX	602	2635	2867	1301	10	12	32	26	39211	11.353
X	841	3416	15672	7545	5	9.2	32	26	37713	6.124
XI	788	5300	3748	1610	12	12.3	32	26	37400	6.176

TABLE 4: Evaluation indicators and their calculations.

Indicators	Calculation
APE	$APE_i = \frac{ \hat{E}_i - E_i }{E_i} \times 100\%$ , $\hat{E}_i = \sum_{j=1}^J \hat{E}_{i,j} / J$
SD	$SD_i = \sqrt{\sum_{j=1}^J (\hat{E}_{i,j} - \hat{E}_i)^2 / J}$
CC	$CC = \frac{\sum_{i=1}^I (\hat{E}_i - \bar{E})(E_i - \bar{E})}{\sqrt{\sum_{i=1}^I (\hat{E}_i - \bar{E})^2 \sum_{i=1}^I (E_i - \bar{E})^2}}$ , $\bar{E} = \sum_{i=1}^I \hat{E}_i / I$ , $\bar{E} = \sum_{i=1}^I E_i / I$
RRMSE	$RRMSE = \sqrt{\sum_{i=1}^I (APE_i)^2 / I}$

model training to get the normalized root mean square error (RMSE)  $\Delta_{mse}$ . Finally, the fitness value of each individual can be calculated by Equation (9), where the value of  $\Delta_{mse}$  is given by the LibSVM toolbox [33].

$$F = \exp(-100 \cdot \Delta_{mse}). \quad (9)$$

These hyper-parameters are commonly tuned by minimizing the validation error [34]. In this paper, the  $k$ -fold cross-validation method [35] is adopted to evaluate the generalization performance of the model with different hyper-parameters in Step (4). And the normalized RMSE  $\Delta_{mse}$  is used as measurement of validation error.

To implement a  $k$ -fold cross-validation, the training datasets are partitioned into  $k$  subsets of approximately equal size in the step of model training. That is,  $k - 1$  subsets of the overall training dataset are used for model training which is called training subset, while the remaining subset serves for validation. The normalized RMSE on the validation subset indicates the performance of hyper-parameters. Repeat this process for  $k$  times, each subset will be used for validation exactly once. Then the normalized RMSE of all data in training dataset is obtained to estimate the validation error. Five-fold cross-validation adopted in this study, was suggested by literature [29, 36].

*Step 5.* All individuals need to be evaluated to identify their performance. The complete solution with the best fitness value represents the best individuals, and this solution will be saved as an offspring individual.

*Step 6.* In this step, GA operators including selection operator, crossover operator and mutation operator are implemented to generate the offspring individuals for the next generation. Firstly, the parent individuals are selected by employing the standard roulette wheel. Then the new individuals are bred through crossover process and mutation process. Figure 4(a) shows the process of crossover and Figure 4(b) shows the process of mutation.

The final SVR model was obtained by the LibSVM toolbox [33] with the optimized value of three hyper-parameters. The relationship between SVR model and HOA can be described in Figure 5.

#### 4. Case Study

This section demonstrates a case study example of estimating daily ECMS on the Beijing metro by using the above proposed method. Section 4.1 describes data collection and pre-processing for the case study, followed by the modelling results presented in Section 4.2. In Section 4.3, a comparison of the proposed SVR model with two different alternative models, including BPNN and MLR, is made in terms of their model performance; Section 4.4 presents a comparison between using holdout validation and five-fold cross-validation in terms of the prediction accuracy. Further, Section 4.5 analyses how different input variables considered in the proposed model would affect the performance of the ECMS estimation. Finally, a real case is conducted to illustrate the reliability of

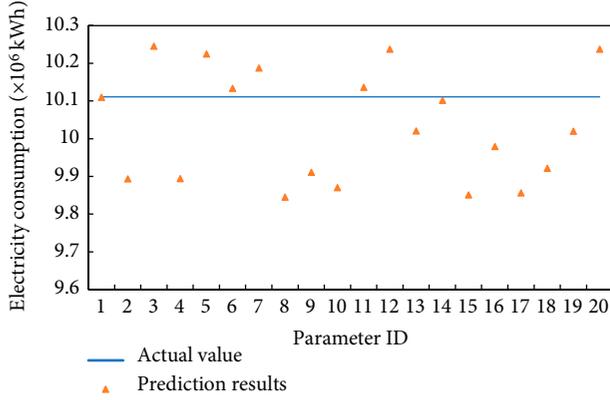


FIGURE 6: The predicted results of the 10th test sample.

TABLE 5: Prediction accuracy of the SVR model.

Sample ID	Actual value	Prediction value ( $\times 10^6$ kWh)	SD	APE (%)	RRMSE (%) /CC
1	10.430	10.243	0.201	1.79	1.39/0.89
2	10.043	10.026	0.112	0.17	
3	10.334	10.159	0.193	1.70	
4	9.820	10.032	0.119	2.16	
5	10.635	10.393	0.140	2.28	
6	10.661	10.562	0.160	0.93	
7	10.289	10.242	0.201	0.46	
8	10.193	10.073	0.142	1.17	
9	10.199	10.249	0.185	0.50	
10	10.111	10.034	0.204	0.77	

the estimation method proposed in this paper. All these studies are conducted by the MATLAB 2016a.

**4.1. Data Description.** Considering all the above specified input variables, a historical dataset of 12 metro stations (station I~ station XII) of a same Beijing metro line was available. The training dataset is composed of historical data of stations I ~ station XI, covered the period from August 1, 2016 to August 30, 2016. Table 3 lists the historical data of the 11 stations (station I~station XI) in one day. In addition, the whole test dataset contains 10 samples, which is gathered from the station XII. To assess the performance of the SVR model, daily ECMS of the test dataset will be predicted.

The data was pre-processed through the min-max normalization as follows:

$$x_i^k = \frac{X_i^k - X_{\min}^k}{X_{\max}^k - X_{\min}^k}. \quad (10)$$

$$y_i = \frac{E_i - E_{\min}}{E_{\max} - E_{\min}}. \quad (11)$$

where  $x_i^k$  represents a real value for variable  $X^k$  ( $k = 1, 2, 3, \dots, 9$ ) in the  $i$ -th sample, with  $X_{\min}^k$  and  $X_{\max}^k$  being the minimum and maximum values over all the data samples.  $x_i^k$  and  $y_i$

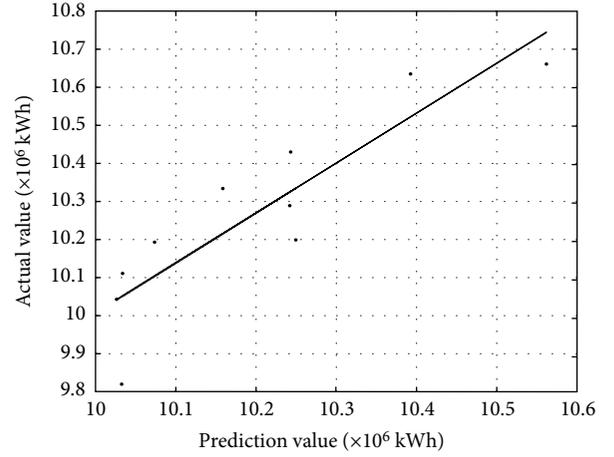


FIGURE 7: The regression curve of predicted values.

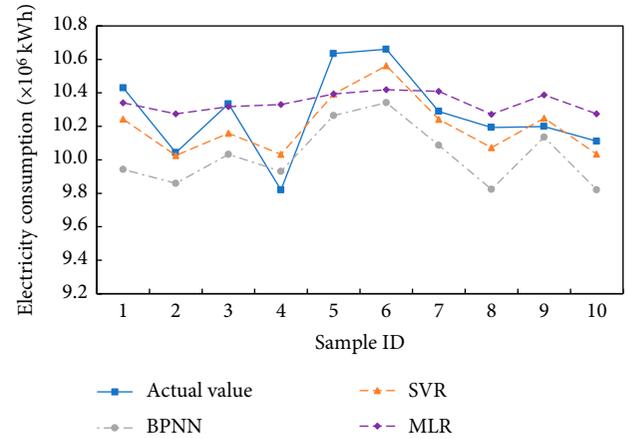


FIGURE 8: Predicted results by the SVR, BPNN, and MLR model.

denote the normalized input and output variable values,  $E_i$  means the value of the output variable under the sample  $i$ ,  $E_{\min}$  and  $E_{\max}$  are the corresponding minimum and maximum values in all samples of output variable.

**4.2. Performance of the Proposed SVR Model.** In this section, numerical cases are implemented to verify the effectiveness of the SVR model. Absolute percentage error (APE), standard deviation (SD), correlation coefficient (CC), relative root mean square error (RRMSE) were used as evaluation indicators. Table 4 lists the formulas of the above four evaluation indicators.

As mentioned above, hyper-parameters  $c$ ,  $\varepsilon$ , and  $\sigma^2$  have impact on the final model performance. In this regard, a method is proposed there to weaken the negative impact caused by the randomness of GA, which will be described in detail as follow. With the same dataset, the optimization process of hyper-parameters was repeated 20 times to obtain 20 sets of different values with respect to three hyper-parameters. In the process of ECMS prediction, the hyper-parameters are selected in turn according to parameter ID. With the selected hyper-parameters, the SVR prediction model was trained based on all training dataset. Then the estimation result of ECMS of the metro station XII can be obtained by giving the

TABLE 6: Predicted results of SVR model, BP neural model and MLR model.

Model	Prediction value			Prediction result	
	CC	RRMSE (%)	Maximum APE (%)	Average SD ( $\times 10^6$ kWh)	Maximum SD ( $\times 10^6$ kWh)
SVR	0.89	1.39	2.28	0.166	0.204
BPNN	0.75	2.86	4.67	0.345	0.465
MLR	0.61	2.27	5.19	—	—

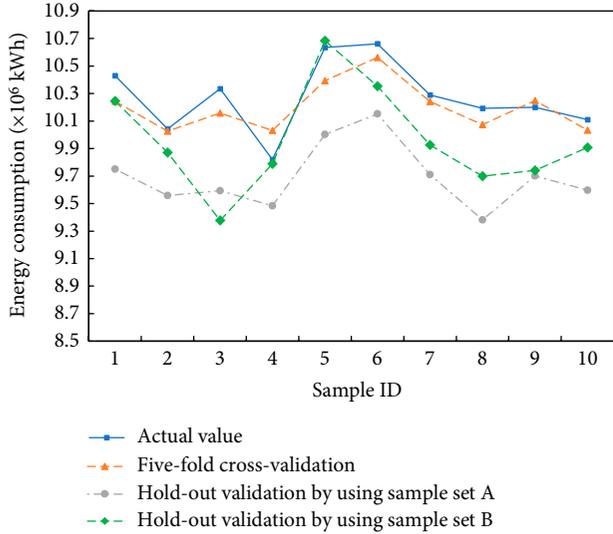


FIGURE 9: Comparison of different validation methods.

input data of test dataset. Based on the 20 sets of hyper-parameters, the ECMS prediction process also be repeated 20 times. It should be noted that the model training was conducted on the same training dataset using different sets of hyper-parameters. In the end, there are 20 prediction results for each prediction sample. And the predicted value for each test samples is equivalent to the mean value of the 20 sets of prediction results.

Figure 6 shows the 20 sets of prediction results of the 10th test sample of the metro station XII. The indicators SD, APE, RRMSE and CC are calculated to evaluate the performance of the SVR model. Table 5 presents the predicted electricity consumption for 10 samples, together with their evaluation indicators. And Figure 7 shows the regression curve of predicted values.

From Figure 6, it can be noticed that the prediction results corresponding to different hyper-parameters are different. Thus, it is necessary to establish a hyper-parameters optimization scheme of SVR model. In this study, GA is employed to optimize these hyper-parameters, since its randomness would not cause a large deviation on the prediction results.

As shown in Table 5, the predicted values by the model were close to the actual ones. The minimum and maximum APE of the predicted values were about 0.17% and 2.28%, respectively. The minimum and maximum SD of the predicted results were  $0.112 \times 10^6$  kWh and  $0.204 \times 10^6$  kWh, respectively. Moreover, the RRMSE and CC is 1.39% and 0.89. These

evaluation indicators reveal that the SVR model performs high prediction accuracy on ECMS.

4.3. *Comparison with Alternative Models.* Apart from the proposed SVR model above, both the BPNN [14] and MLR model [37] were also used to predicting the ECMS. This subsection compares their performance in terms of accuracy and applicability.

The BPNN model consists of three layers: an input layer, a hidden layer and an output layer. Let  $N_h$  denote the number of neurons in the hidden layer, which can have significant impact on the performance of the BPNN model. As a rule of thumb (cf. [14]), it can be derived by Equation (12) as follows:

$$N_h = \frac{N_{in} + N_{out}}{2} + \sqrt{N_s}. \quad (12)$$

where,  $N_{in}$  and  $N_{out}$  denote the numbers of input and output variables, respectively; and  $N_s$  the sample size of training set.

In addition, comparable result of the MLR model was obtained directly by using the IBM SPSS Statistics 20 software. Note that the sample data for training both the BPNN and MLR models were the same as the one used for SVR model. The predicted results by different models are illustrated in Figure 8 and their respective model accuracy is compared in Table 6.

In Figure 8, the curves of SVR model and BPNN model are more similar to the curve actual value than the curve of MLR model. It is proved that the BPNN model and the SVR model could fit better the sample data when dealing with nonlinear problems. The results presented in Table 6 show that the prediction accuracy of SVR model is higher than that of BPNN and MLR models. Compared with MLR model, it is found that the prediction values of SVR model is closer to the actual values due to SVR model has nonlinear approximation ability. As compared to the BPNN model, SVR model achieves better generalization since SVR adopted the principle of minimizing structure risk while BPNN model only take the empirical risk into account. The principle of minimizing empirical risk is unreasonable when the number of training dataset is limited (only 330), since it consists of the training error only. On the contrary, the SVR model has powerful ability to handle the nonlinear problem with small samples. Thus, the SVR model is more suitable for the ECMS prediction problem.

4.4. *Comparison of Validation Schemes.* In this subsection, the hold-out validation method is also implemented in the process of hyper-parameters optimizing to verify the superiority of the validation scheme applied in this paper,

TABLE 7: Comparison between five-fold cross-validation method and hold-out validation method.

Method	CC	Prediction value		Prediction result		
		RRMSE (%)	Maximum APE (%)	Average SD ( $\times 10^6$ kWh)	Maximum SD ( $\times 10^6$ kWh)	
Five-fold cross-validation	0.89	1.39	2.28	0.89	0.204	
Hold-out validation	Sample set A	0.83	5.76	7.95	0.83	0.538
	Sample set B	0.65	4.01	9.26	0.65	0.613

TABLE 8: Impact of influence factors on model prediction result.

Missing variable	Maximum APE	CC
Average temperature	9.02%	0.37
Relative humidity	8.52%	0.48
Area of concourse	6.66%	0.76
Area of platform	7.60%	0.71
Area of staff accommodation room	6.58%	0.84
Area of plant room	5.98%	0.88
Quantity of escalators or elevators	6.76%	0.81
Height of escalators or elevators	6.89%	0.81
Number of passengers	6.00%	0.74

i.e., five-fold cross-validation. Figure 9 and Table 8 show the prediction result under the implementation of the above two methods. Different from the general hold-out validation method, the training dataset is split into two parts randomly according to the ratio of 4:1 in this paper. And the large one is used as training sample set while the little one should be tested with the trained model. The verification of the hold-out validation method is performed twice because the division of the training sample set and the verification sample set was random. Table 7 shows the results of two methods. And the sample set A and sample set B in Table 7 represent the ID of different training samples and verification samples in the two experiments.

As illustrated in Figure 9 and Table 7, five-fold cross-validation outperformed holdout one. A possible reason is that the five-fold cross-validation enables each sample to be tested, and more valid information can be obtained; however, the prediction accuracy of hold-out validation differs significantly between different validation samples. That means the selection of validation samples is important to the evaluation result of SVR model. However, the selection of training samples also has great impact on the evaluation result. In summary, the validation scheme plays significant role in evaluating the model performance, which may affect the optimization results of hyper-parameters  $c$ ,  $\epsilon$  and  $\sigma^2$ .

**4.5. The Effect of Influencing Factors.** Nine influencing factors of the ECMS forecasting model are analysed in this section. Firstly, each input variables are removed in turn. Then the hyper-parameters optimization algorithm is implemented based on the surplus eight input variables of training data, and the optimization process would be repeated 20 times. With the 20 sets of hyper-parameters, the SVR model is trained based

on the training data which consists of eight input variables. Finally, the predicted results can be obtained by giving the input data of prediction samples. Table 8 lists the performance of the ECMS forecasting model after one input variable is removed.

As shown in Table 8, the removal of different input variables will lead to different degrees of decline in the accuracy of the ECMS prediction model. The average temperature and relative humidity are the most significant influencing factors, which have distinct impact on the accuracy of the model compared to other factors. The area of concourse, the area of platform and the area of staff accommodation room have a large impact on the prediction accuracy, too. It means that the scale of metro station is important to the energy-saving design of the metro system that it can be adjusted in the planning and design stage. Thus, the station size should be reduced as much as possible to reduce station electricity consumption if the constraint of accommodation capacity be satisfied.

**4.6. Prediction Results on a New Metro Station.** In this subsection, a new metro station apart from the above existing stations is employed to validate the performance of the proposed SVR model in estimating the ECMS during design stage. The real-world parameters and energy consumption of this station over ten days are collected and given in Table 9.

The prediction results by the proposed SVR model are shown in Figure 10. It is found that the prediction values are very close to the actual ones and the correlation coefficient reaches 0.88. The maximum APE, corresponding to the 7th sample, is no more than 2.75%. In other words, the SVR model is able to predict the ECMS of a new metro station with a satisfactory accuracy.

## 5. Conclusion

This paper proposes a new approach to estimating the ECMS given data of a small sample size. The major factors influencing the ECMS are discussed, including average air temperature, relative humidity, area of some key components (i.e., station concourse, platform, staff accommodation room and plant room) of a station, number of passengers, and both number and heights of escalators/elevators. All the above nine variables are proposed as the input variables of a SVR model, and the hyper-parameters of the SVR model is optimized by GA. The case studies based on actual data validated the effectiveness of

TABLE 9: Testing dataset.

Sample ID	$X^1$ (m <sup>2</sup> )	$X^2$ (m <sup>2</sup> )	$X^3$ (m <sup>2</sup> )	$X^4$ (m <sup>2</sup> )	$X^5$	$X^6$ (m)	$X^7$ (°C)	$X^8$ (%)	$X^9$ (person/day)	$E$ ( $\times 10^6$ kWh)
1	635	3769	1974	3562	11	12.4	32	26	39683	10.965
2	635	3769	1974	3562	11	12.4	19	87	38505	9.976
3	635	3769	1974	3562	11	12.4	30	63	37022	9.742
4	635	3769	1974	3562	11	12.4	27	85	36761	10.181
5	635	3769	1974	3562	11	12.4	27	85	36141	10.236
6	635	3769	1974	3562	11	12.4	25	78	35300	10.380
7	635	3769	1974	3562	11	12.4	27	85	36236	9.910
8	635	3769	1974	3562	11	12.4	26	50	34974	9.690
9	635	3769	1974	3562	11	12.4	24	57	35047	10.289
10	635	3769	1974	3562	11	12.4	26	32	36126	9.790

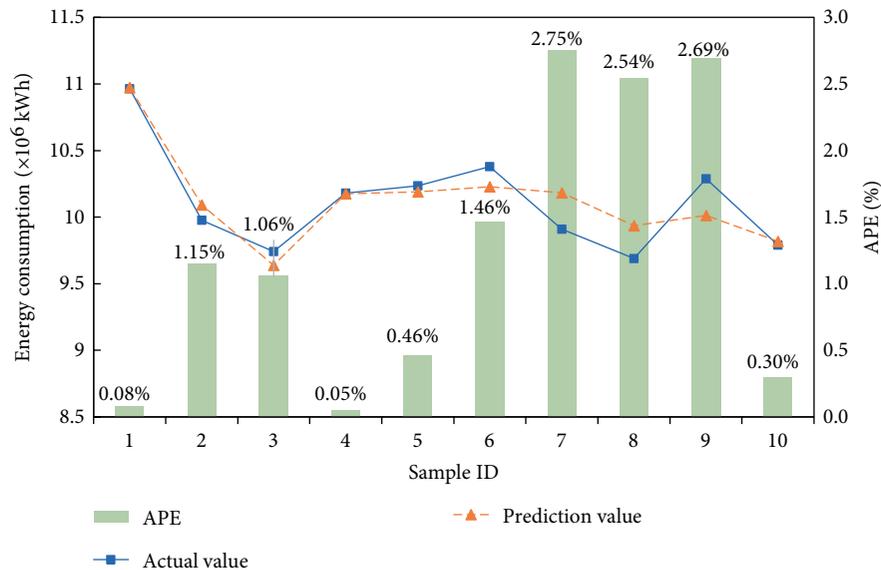


FIGURE 10: The prediction results of a real case.

the proposed SVR and demonstrated the SVR model could achieve higher prediction accuracy than a BPNN model and a MLR model. The proposed SVR model provides a promising alternative approach to predicting the ECMS of new metro stations.

The prediction of traction energy consumption for a new metro line is the next step of our research, as the amount of energy consumed by train traction also accounts for a large proportion of the overall energy consumption in a metro system.

### Data Availability

The data used to support the findings of this study are included within the article, through tables. Further details may be available to the reader, from the authors, upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This research is supported by National Natural Science Foundation of China (71571016 and 71621001).

### References

- [1] A. González-Gil, R. Palacin, P. Batty, and J. P. Powell, "A systems approach to reduce urban rail energy consumption," *Energy Conversion and Management*, vol. 80, pp. 509–524, 2014.
- [2] W. Hong and S. Kim, "A study on the energy consumption unit of subway stations in Korea," *Building and Environment*, vol. 39, no. 12, pp. 1497–1503, 2004.
- [3] W. Zhang and J. Wei, "The experimental research on Beijing subway air-conditioning system energy-saving," *Applied Mechanics and Materials*, vol. 170–173, pp. 2494–2498, 2012.
- [4] Y. Wang, H. Feng, and X. Xi, "Monitoring and autonomous control of Beijing subway HVAC system for energy sustainability," *Energy for Sustainable Development*, vol. 39, pp. 1–12, 2017.

- [5] X. Zhang, G. Gong, G. Kang, and T. Han, "The environmental influence of subway station on platform screen door system," in *International Conference on Sustainable Design and Construction*, American Society of Civil Engineers, Reston, VA, USA, 2012.
- [6] M. Casals, M. Gangoellis, N. Forcada, M. Macarulla, and A. Giretti, "A breakdown of energy consumption in an underground station," *Energy and Buildings*, vol. 78, no. 4, pp. 89–97, 2014.
- [7] M. Casals, M. Gangoellis, N. Forcada, and M. Macarulla, "Reducing lighting electricity use in underground metro stations," *Energy Conversion and Management*, vol. 119, pp. 130–141, 2016.
- [8] M. Casals, M. Gangoellis, N. Forcada, M. Macarulla, A. Giretti, and M. Vaccarini, "SEAM4US: an intelligent energy management system for underground stations," *Applied Energy*, vol. 166, pp. 150–164, 2016.
- [9] K. Fu and Z. Deng, "Current situation of energy consumption in Guangzhou railway station and analysis on the potential of energy conservation," *Journal of Sustainable Development*, vol. 2, no. 1, p. 117, 2009.
- [10] Z. Yang, J. Yue, and X. Wang, "Prediction of urban rail transit power consumption based on regression mode," *Urban Mass Transit*, vol. 13, no. 12, pp. 22–25, 2010.
- [11] Z. Wang, F. Chen, and Z. Shi, "Prediction on medium- and long-term energy consumption of urban rail transit network in Beijing," *China Railway Science*, vol. 34, no. 3, pp. 133–136, 2013.
- [12] B. Guan, X. Liu, T. Zhang, and J. Xia, "Energy consumption of subway stations in China: data and influencing factors," *Sustainable Cities and Society*, vol. 43, pp. 451–461, 2018.
- [13] J. Ahn, S. Cho, and D. Chung, "Development of a statistical analysis model to benchmark the energy use intensity of subway stations," *Applied Energy*, vol. 179, pp. 488–496, 2016.
- [14] P. C. M. Leung and E. W. M. Lee, "Estimation of electrical power consumption in subway station design by intelligent approach," *Applied Energy*, vol. 101, no. 1, pp. 634–643, 2013.
- [15] A. Aslanargun, M. Mammadov, B. Yazici, and S. Yolacan, "Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting," *Journal of Statistical Computation and Simulation*, vol. 77, no. 1, pp. 29–53, 2007.
- [16] B. Ekici and U. Aksoy, "Prediction of building energy consumption by using artificial neural networks," *Advances in Engineering Software*, vol. 40, no. 5, pp. 356–362, 2009.
- [17] R. Yokoyama, T. Wakui, and R. Satake, "Prediction of energy demands using neural network with model identification by global optimization," *Energy Conversion and Management*, vol. 50, no. 2, pp. 319–327, 2009.
- [18] H. Su, L. Zhang, and S. Yu, "Short-term traffic flow prediction based on incremental support vector regression," in *Third International Conference on Natural Computation (ICNC 2007)*, IEEE, Haikou, Hainan, China, 2007.
- [19] C. Ren, N. An, J. Wang, L. Li, B. Hu, and D. Shang, "Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting," *Knowledge-Based Systems*, vol. 56, pp. 226–239, 2014.
- [20] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, USA, 2nd edition, 1995.
- [21] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [22] B. Dong, C. Cao, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, 2005.
- [23] Q. Li, Q. L. Meng, J. J. Cai, H. Yoshino, and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Applied Energy*, vol. 86, no. 10, pp. 2249–2256, 2009.
- [24] W. C. Hong, "Application of chaotic ant swarm optimization in electric load forecasting," *Energy Policy*, vol. 38, no. 10, pp. 5830–5839, 2010.
- [25] K. Kavaklioglu, "Modeling and prediction of Turkey's electricity consumption using support vector regression," *Applied Energy*, vol. 88, no. 1, pp. 368–375, 2011.
- [26] S. Keerthi, "Efficient tuning of SVM hyper-parameters using radius/margin bound and iterative algorithms," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp. 1225–1229, 2002.
- [27] L. J. Cao and F. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003.
- [28] M. W. Ahmad, A. Mouraud, Y. Rezgui, and M. Mourshed, "Deep highway networks and tree-based ensemble for predicting short-term building energy consumption," *Energies*, vol. 11, no. 12, p. 3408, 2018.
- [29] K. Chen and C. Wang, "Support vector regression with genetic algorithms in forecasting tourism demand," *Tourism Management*, vol. 28, no. 1, pp. 215–226, 2007.
- [30] C. Hsu, C. Chang, and C. Lin, "A practical guide to support vector classification," Department of Computer Science and Information Engineering: National Taiwan University, Taiwan, 2000, <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [31] V. Cherkassky and F. Mulier, "Learning from data: concepts, theory, and methods," *Technometrics*, vol. 43, no. 1, pp. 105–106, 2001.
- [32] J. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, MA, USA, 1992.
- [33] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [34] K. Duan, S. Keerthi, and A. Poo, "Evaluation of simple performance measures for tuning SVM hyperparameters," *Neurocomputing*, vol. 51, no. 2, pp. 41–59, 2003.
- [35] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 1137–1143, Lawrence Erlbaum Associates Ltd., Montreal, QC, Canada, 1995.
- [36] R. Jain, K. Smith, P. Culligan, and J. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol. 123, no. 3, pp. 168–178, 2014.
- [37] Y. Chen, B. Mao, Y. Bai, Y. Feng, and Z. Li, "Forecasting traction energy consumption of metro based on support vector regression," *Systems Engineering: Theory and Practice*, vol. 36, pp. 2101–2107, 2016.