

Article

# Electricity Consumption Forecasting using Support Vector Regression with the Mixture Maximum Correntropy Criterion

Jiandong Duan <sup>1,2</sup>, Xuan Tian <sup>1</sup>, Wentao Ma <sup>1</sup> , Xinyu Qiu <sup>1</sup>, Peng Wang <sup>1</sup> and Lin An <sup>3,\*</sup>

<sup>1</sup> School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China

<sup>2</sup> State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710048, China

<sup>3</sup> School of Statistics, Xi'an University of Finance and Economics, Xi'an 710100, China

\* Correspondence: anlin@xaufe.edu.cn; Tel.: +861-357-209-0698

Received: 28 May 2019; Accepted: 15 July 2019; Published: 19 July 2019



**Abstract:** The electricity consumption forecasting (ECF) technology plays a crucial role in the electricity market. The support vector regression (SVR) is a nonlinear prediction model that can be used for ECF. The electricity consumption (EC) data are usually nonlinear and non-Gaussian and present outliers. The traditional SVR with the mean-square error (MSE), however, is insensitive to outliers and cannot correctly represent the statistical information of errors in non-Gaussian situations. To address this problem, a novel robust forecasting method is developed in this work by using the mixture maximum correntropy criterion (MMCC). The MMCC, as a novel cost function of information theoretic, can be used to solve non-Gaussian signal processing; therefore, in the original SVR, the MSE is replaced by the MMCC to develop a novel robust SVR method (called MMCCSVR) for ECF. Besides, the factors influencing users' EC are investigated by a data statistical analysis method. We find that the historical temperature and historical EC are the main factors affecting future EC, and thus these two factors are used as the input in the proposed model. Finally, real EC data from a shopping mall in Guangzhou, China, are utilized to test the proposed ECF method. The forecasting results show that the proposed ECF method can effectively improve the accuracy of ECF compared with the traditional SVR and other forecasting algorithms.

**Keywords:** electricity consumption forecasting; support vector regression; mixture maximum correntropy criterion; parameter optimization

---

## 1. Introduction

### 1.1. Motivation

Electricity consumption forecasting (ECF) refers to the estimation or forecast of the future monthly electricity sales by collating and analyzing historical data. Whether for the bid evaluation system of the State Grid Corporation or for the open electricity market today, electricity sales forecasting is very important, especially for power sales companies, since it will directly affect the deviation assessment. The bigger the deviation, the higher the penalty. Accurate prediction of monthly electricity sales is of vital practical significance and practical value to the power sales company.

ECF is a new problem challenging the current power sales company. In essence it consists of load forecasting. However, compared with load forecasting, it has some new characteristics, such as fewer historical load data and stronger randomness.

## 1.2. Literature Review

Extensive research has been conducted on load forecasting. Ref. [1] established a long-term power load forecasting model by using a support vector machine (SVM) model based on the comprehensive consideration of economic factors, social factors, and energy market structure and optimizing a multi-factor medium and the parameters of SVM by using particle swarm optimization (PSO) and an improved cross-validation method. In Ref. [2], the improved fly optimization algorithm (FOA) algorithm was used to optimize the parameters of the wavelet least-squares support vector machine (using the wavelet kernel function instead of the Gauss kernel function in least-squares support vector machines (LSSVM)), and the combined algorithm was applied to medium- and long-term power load forecasting. In Ref. [3], the author combined the Fuzzy Time Series with Seasonal Autoregressive Fractionally Integrated Moving model to predict load changes, and the PSO algorithm s used to optimize the weights of the model leading to a good prediction result. In Ref. [4], the author proposed a hybrid optimization grey model called Rolling-ALO-GM (1,1) for ECM based on known information from annual EC by using the grey control theory. In Ref. [5], the author used the Drosophila optimization algorithm with decreasing step size (SFOA) to optimize the extended parameters of generalized regression neural network (GRNN). By combining it with the periodic characteristics of weather factors and power load, a SFOA-GRNN power load forecasting model was proposed. The experimental results showed that the improved model was much better than the original model.

The SVM is a supervised learning algorithm which can solve the problems of small sample, non-linearity, and high latitude and is more efficient in solving pattern recognition and regression problems. Therefore, it has been widely used in short-term load forecasting. In Ref. [6], in order to improve the accuracy and efficiency of calculation, the secondary sampling strategy of support vector regression (SVR) learning process was designed, and a method of modeling and forecasting short-term load point estimation and its confidence interval length using subsampled SVR ensemble was obtained. In Ref. [7], a new grid algorithm was proposed, which provides a new method for the determination of parametric regression surface. A sequential grid method based on SVR was established for short-term load forecasting. Compared with the standard SVR model, the accuracy of short-term load forecasting was greatly improved. In order to deal with the cyclic nature of electric loads, Ref. [8] proposed the SVR with chaotic cuckoo search (SVRCCS) model, based on using a tent chaotic mapping function to enrich the cuckoo search space and diversify the population to avoid trapping in local optima, and combined the SVRCCS model with a seasonal mechanism, thus obtaining a seasonal SVRCCS model. The numerical results showed that the proposed novel model outperforms other alternative models. Ref. [9] used the proposed chaotic efficient bat algorithm, niche search, and evolution mechanisms to optimize the parameters of the hybrid kernel-based SVR model and used motion data for a real floating platform to evaluate the reliability and effectiveness of the proposed model. Ref. [10] proposed a novel vector field-based SVR method. Through multi-distortions in the sample data space or high-dimensional feature space mapped by a vector field, the optimal feature space was found, in which the high non-linearity between inputs and outputs was approximated by linearity. The results indicated that the proposed method achieves better performance than commonly used methods with regard to the accuracy, robustness, and generalization ability.

However, the loss function of the traditional SVR is the mean-square error (MSE) criterion. The MSE criterion only considers the second-order distance of the prediction error distribution. It is very effective for data with Gaussian characteristics, which results in high accuracy only when predicting a stationary sequence. Compared with the power consumption of large users and stable users, the power consumption of small users is constrained by many external factors, such as orders, operation plans, equipment, environmental factors, meteorological changes, emergencies, and so on. It produces non-linear characteristics, such as randomness and uncertainty, and prediction errors are often non-gaussian. Therefore, the application of MSE-based SVR in power consumption forecasting will be less accurate compared with that for large users and stable users. It produces non-linear characteristics, such as randomness and uncertainty, and prediction errors are often non-gaussian.

Therefore, the application of SVR based on MSE in ECF has some shortcomings. The Neural Network Laboratory of the University of Florida, USA, proposed a new definition of the correlation function of stochastic processes, namely, correntropy [11]. When it is taken as the cost function in machine learning and signal processing, a Maximum Correntropy Criterion (MCC) can be defined for non-gauss, non-linear applications. The MCC contains higher moments of probability density function and is suitable for any noise environment. So far, many robust learning algorithms have been developed using MCC [12–18]. It has been proved that they can perform a robust analysis [19–22] and effectively deal with non-Gaussian situations and outliers [23–25]. In order to further improve the learning performance, Ref. [26] proposed the concept of the maximum mixture correntropy criterion (MMCC), which uses the mixture of two Gauss functions as the core function. The learning algorithm using MMCC has a better performance than the traditional MCC-based learning algorithm and several other advanced learning algorithms. On the basis of the above analysis, MMCC was introduced into SVR for the first time, and an MMCCSVR forecasting model for EC is proposed in this paper.

### 1.3. Our Contribution

The main contributions of this study can be summarized as follows:

(1) In order to solve the problem due to the fact that traditional SVR based on MSE loss function only has high efficiency in data processing with Gaussian characteristics, which leads to its insensitivity to burst and low accuracy in predicting a non-stationary sequence, MMCC is introduced into SVR as loss function, and an MMCCSVR model is proposed.

(2) The MMCCSVR model is applied to ECF for the first time. The simulation results show that the forecasting accuracy of this method is higher than that of traditional SVR and other algorithms.

### 1.4. Organization of the Paper

This paper is structured as follows: Section 2 gives an overview of SVR and MMCC. Section 3 proposes the MMCCSVR model for ECF in detail. Section 4 introduces the main steps of forecasting. In Section 5, the model validation is introduced and next, experimental results and comparisons with traditional SVR and other algorithms are shown and discussed. Finally, a brief summary and a description of future work are given in Section 6.

## 2. Methodology

### 2.1. SVR

The SVR model has its own unique advantages in dealing with such problems as small amount of data, obvious non-linear characteristics, and high-dimensional pattern recognition. It only needs a small number of support vectors to establish decision functions, which can effectively avoid dimension disasters. Moreover, the reduction or increase of non-support vector data has little impact on the algorithm and has better robustness. SVR uses the kernel function  $\varphi(x)$  in substitution of nonlinear mapping, the controlled parameters are few, and the algorithm is simple and easy to implement [27].

In an SVR model, suppose the training set sample is  $(x_i, y_i) | i = 1, 2, \dots, n$ , and  $n$  is the number of samples. The regression problem can be transformed into an optimization problem, as follows.

$$\begin{aligned} \min C \sum_{i=1}^n \xi_i^2 + \|\omega\|^2 \\ \text{s.t. } \omega^T \varphi(x_i) + b - y_i = \xi_i, i = 1, 2, \dots, n \end{aligned} \quad (1)$$

where  $\xi$  is the error variable,  $\omega$  is the weight vector,  $\omega \in H$ ,  $b$  is the bias, and  $C$  is the penalty parameter;  $C$  is used to control the minimization of the estimation error and the function smoothness.

In order to solve the optimization problem, the Lagrange function is developed as

$$L(\omega, b, \xi, \alpha) = (\omega^T \omega) + C \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i^2 [\omega^T \varphi(x_i) + b - y_i - \xi_i] \quad (2)$$

where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$  is the Lagrange multiplier. Differentiating  $L$  with respect to the variables  $\omega, b, \xi$ , and  $\alpha$ , we obtain

$$\frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^n \alpha_i \varphi(x_i), \quad (3)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \omega = \sum_{i=1}^n \alpha_i = 0, \quad (4)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = C \xi_i, \quad (5)$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + \xi_i - y_i = 0, \quad (6)$$

and, after solving the above functions, we can obtain the solution of the problem in the following form

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

where  $K(\cdot)$  is the kernel function. In this paper, we used Gaussian kernel  $K(x, x_i) = \exp\left[-\frac{\|x-x_i\|^2}{\gamma^2}\right]$  to construct nonlinear SVR since it is widely used in the literature and has shown good learning properties in a variety of applications.

In the following subsection, we introduce the MMCC method.

## 2.2. Mixture Maximum Correntropy Criterion

First, correntropy is defined as follows

$$V(X, Y) = E[\kappa(X, Y)] \quad (8)$$

where  $\kappa(X, Y)$  is the kernel function, and  $E[\bullet]$  is the mathematical expectation. Correntropy is a generalized correlation measure. Suppose  $\{(x_i, y_i)\}_{i=1}^n$  is of the random variables  $X, Y$ , the sample mean estimator of correntropy is

$$\hat{V}(X, Y) = \frac{1}{N} \sum_{i=1}^N g(x_i - y_i, \sigma) \quad (9)$$

where  $g(x_i - y_i, \sigma) = G_\sigma(\bullet) = \exp\left(-\frac{\|x_i - y_i\|^2}{2\sigma^2}\right)$  is the Gaussian kernel with bandwidth  $\sigma > 0$ .

As a nonlinear similarity measure, correntropy has been successfully used as an efficient optimization cost in signal processing and machine learning [25], and the corresponding cost function is MCC. In order to further improve the learning performance, Ref. [25] proposed a mixture correntropy that uses the mixture of two Gaussian functions as the kernel function; the mixture correntropy is defined by

$$M(X, Y) = E[\lambda G_{\sigma_1}(X - Y) + (1 - \lambda) G_{\sigma_2}(X - Y)] \quad (10)$$

where  $\sigma_1$  and  $\sigma_2$  are the kernel bandwidths of the Gaussian functions  $G_{\sigma_1}(\bullet)$  and  $G_{\sigma_2}(\bullet)$ , and  $0 \leq \lambda \leq 1$  is the mixture coefficient. Hence, Equation (10) can be written as

$$\hat{M}(X, Y) = \frac{1}{n} \sum_{i=1}^n (\lambda g(x_i - y_i, \sigma_1) + (1 - \lambda)g(x_i - y_i, \sigma_2)) \tag{11}$$

Especially, when  $\sigma_1 = \sigma_2$ , MMCC degenerates into MCC.

### 3. SVR with MMCC

The central idea of SVR with MMCC (MMCCSVR) is to integrate the MMCC regularization technique and the kernel trick in a unified framework. Therefore, the MMCCSVR seeks the optimal regression function  $f(x) = \sum_{i=1}^n \omega_i K(x, x_i) + b$  by solving the following optimization problem

$$\max_{\omega, b} J(\omega, b) = C \sum_{i=1}^n (\lambda g(\omega^T \varphi(x_i) + b - y_i, \sigma_1) + (1 - \lambda)g(\omega^T \varphi(x_i) + b - y_i, \sigma_2)) - \|\omega\|^2 \tag{12}$$

Concretely, the first term of (12) is based on the aforementioned MMCC (11) in the  $\varphi$  – included kernel space and is used as an empirical loss function to measure the fitting error. The second term of (12) makes the required regression estimation  $f(x)$  as smooth as possible from the perspective of Tikhonov regularization. It can also be interpreted as a regularization to overcome possible overfitting.

However, there are two difficulties to solve (12) directly. The first one is that the correntropy cost (11) is hard to be optimized. Although Liu et al. proposed a steepest descend-based method in Ref. [27] to tackle this problem, it is not very efficient. The second problem concerning (12) is due to the fact that  $\omega$  may lie in a high or even infinite dimensional feature space induced by nonlinear map  $\varphi(\cdot)$ , thus limiting the application of conventional optimization methods. Fortunately, (12) can be converted to a more straightforward form by considering the following proposition to tackle the first problem [28–30].

**Proposition 1.** *There exists a convex function  $\phi: R \rightarrow R$ , such that*

$$g(x, \sigma) = \max_{p < 0} \left( p \frac{\|x\|^2}{\sigma^2} - \phi(p) \right) \tag{13}$$

and, for a fixed  $x$ , the maximum is reached at  $p = -g(x, \sigma)$ .

The proof of Proposition 1 is based on the convex conjugated function theory and can be found in Ref. [31]. Now, we construct the following optimization problem with augmented objective function in an enlarged parameter space:

$$\max_{\omega, b, p} \bar{J}(\omega^T, b, p) = C \sum_{i=1}^n \left( \lambda p_i \frac{(\omega^T \varphi(x_i) + b - y_i)^2}{\sigma_1^2} + (1 - \lambda) p_i \frac{(\omega^T \varphi(x_i) + b - y_i)^2}{\sigma_2^2} - \phi(p_i) \right) - \|\omega\|^2 \tag{14}$$

where  $p = [p_1, p_2, \dots, p_n]^T$  stores the auxiliary variables introduced in the half-quadratic optimization. Then, according to proposition 1, we notice that when  $[\omega, b]$  is fixed, the following equation holds

$$J(\omega, b) = \max_p \bar{J}(\omega, b, p) \tag{15}$$

Hence, we can further get the following equation

$$\max_{\omega, b} J(\omega, b) = \max_{\omega, b, p} \bar{J}(\omega, b, p) \tag{16}$$

That is to say, maximizing  $J(\omega, b)$  with respect to  $[\omega, b]$  is equivalent to maximizing the augmented function  $\bar{J}(\omega, b, p)$  in the enlarged parameter space  $[\omega, b, p]$ . Now, to overcome the second problem, we introduce a slack vector to convert the unconstrained optimization problem (16) to the following constrained problem  $\xi = [\xi_1, \xi_2, \dots, \xi_n]^T$ , as in standard SVR

$$\begin{aligned} \max_{\omega, b, \xi, p} \quad & \frac{C}{2} \sum_{i=1}^n \left( \lambda p_i \frac{\xi_i^2}{\sigma_1^2} + (1-\lambda) p_i \frac{\xi_i^2}{\sigma_2^2} - \phi(p_i) \right) - \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & \omega^T \varphi(x_i) + b - y_i = \xi_i, i = 1, 2, \dots, n \end{aligned} \quad (17)$$

We can choose to iteratively optimize (17) by alternating the optimization with respect to either  $[\omega, b, \xi_i]$  or  $p$ , while holding the other term fixed.

First, we hold  $p$  fixed and maximize (17) with respect to  $[\omega, b, \xi_i]$ . By dropping the other unrelated variables and introducing the variable  $q_i = -p_i$ , we obtain the following equivalent problem

$$\begin{aligned} \min_{\omega, b, \xi} \quad & \frac{C}{2} \sum_{i=1}^n \left( \lambda q_i \frac{\xi_i^2}{\sigma_1^2} + (1-\lambda) q_i \frac{\xi_i^2}{\sigma_2^2} \right) + \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & \omega^T \varphi(x_i) + b - y_i = \xi_i, i = 1, 2, \dots, n \end{aligned} \quad (18)$$

In order to solve the above problem, we introduce the following Lagrangian function

$$L = C \sum_{i=1}^n \left( \lambda q_i \frac{\xi_i^2}{2\sigma_1^2} + (1-\lambda) q_i \frac{\xi_i^2}{2\sigma_2^2} \right) + \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^n \alpha_i (\omega^T \varphi(x_i) + b - y_i - \xi_i) \quad (19)$$

where  $\alpha_i$  is a Lagrangian multiplier. Then, by following the Karush-Kuhn-Tucker (KKT) condition [32], we have

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^n \alpha_i \varphi(x_i) = 0 \rightarrow \omega = \sum_{i=1}^n \alpha_i \varphi(x_i) \quad (20)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i = 0 \quad (21)$$

$$\frac{\partial L}{\partial \xi_i} = \frac{C q_i \lambda \xi_i}{\sigma_1^2} + \frac{C q_i (1-\lambda) \xi_i}{\sigma_2^2} + \alpha_i = 0 \rightarrow \xi_i = -\frac{\alpha_i \sigma_1^2 \sigma_2^2}{\lambda C q_i \sigma_2^2 + (1-\lambda) C q_i \sigma_1^2} \quad (22)$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b - y_i = \xi_i \quad (23)$$

Combing Equations (20)–(23) leads to the following linear system of equations

$$\begin{bmatrix} K + Q & e \\ e^T & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix} \quad (24)$$

where  $K$  is an  $n \times n$  kernel matrix with  $K_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$ ,  $Y = [y_1, y_2, \dots, y_n]^T$ ,  $e = [1, 1, \dots, 1]^T$ , and  $Q$  is a diagonal matrix whose diagonal element  $Q_{ii} = \left( \frac{\sigma_1^2 \sigma_2^2}{\lambda C q_i \sigma_2^2 + (1-\lambda) C q_i \sigma_1^2} \right) = -\left( \frac{\sigma_1^2 \sigma_2^2}{\lambda C p_i \sigma_2^2 + (1-\lambda) C p_i \sigma_1^2} \right) > 0$ , since  $p_i < 0$ . Noticing that  $K + Q$  is symmetric positive-definite and thus invertible, we can efficiently solve (23) by the following equations [33].

$$b = \frac{e^T (K + Q)^{-1} Y}{e^T (K + Q)^{-1} e} \text{ and } \alpha = (K + Q)^{-1} (Y - eb) \quad (25)$$

Second, we hold  $[\omega, b, \xi]$  fixed and optimize (17) with respect to  $p$ . Actually, according to Proposition 1, the optimal  $p$  is directly given by

$$p_i = -g(\xi_i, \sigma) = -\exp\left(-\frac{\lambda \xi_i^2}{\sigma_1^2} + (1 - \lambda) \frac{\xi_i^2}{\sigma_2^2}\right) \tag{26}$$

Up to now, we have solved both sub-problems constituting the original optimization problem (17). For the sake of clarity, we prescribe the proposed MMCCSVR as follows.

First, set the parameter values  $p_i = -q_i = -1$  for all samples, tolerance  $\varepsilon = 1e-3$ , MMCC parameter  $\sigma_1$  and  $\sigma_2$ , kernel parameter  $\gamma$ , regularization parameter  $C$ , build kernel matrix  $K$ .

**Step 1.** Calculate the diagonal matrix  $Q$  and then solve (25) by Cholesky factorization to obtain  $[\alpha, b]$ .

**Step 2.** If  $[\alpha, b]$  changes less than  $\varepsilon$ , go to Step 4, otherwise go to Step 3.

**Step 3.** Calculate the error variable  $\xi_i$  by (22) and update  $p_i = -q_i$  by (26) and then go to Step 1.

**Step 4.** Determine the final regression estimation by (7).

Referring to the convergence analysis method in Ref. [28–30], the convergence of the proposed algorithm is proved as follows.

According to Step 1 and Step 2, we have  $\bar{J}(\omega^t, b^t, p^t) \leq \bar{J}(\omega^{t+1}, b^{t+1}, p^t)$  and  $\bar{J}(\omega^{t+1}, b^{t+1}, p^t) \leq \bar{J}(\omega^{t+1}, b^{t+1}, p^{t+1})$ . Hence, we can conclude that the sequence  $\bar{J}(\omega^t, b^t, p^t)$ ,  $t = 1, 2, \dots$  is non-decreasing. Based on the property of correntropy, we can verify that the objective function  $\bar{J}(\omega, b, p)$  is bounded above, since  $g(x, \sigma)$  and  $-\|\omega\|^2$  are both bounded above. Due to the above facts and the well-known monotone convergence theorem, the algorithm is guaranteed to converge to its local optimal solution.

#### 4. Electricity Consumption Forecasting Based on MMCCSVR

##### 4.1. Characteristic Analysis of Electricity Consumption Data

By analyzing the characteristics of users' EC, power sales company can accurately understand the demand response mechanism and formulate scientific marketing strategies. This is of great significance for peak shaving and valley filling, optimizing the EC curve, and improving power quality.

The monthly EC of a shopping mall in Guangdong Province in 2017 was analyzed as an example. Figure 1 shows the daily EC of the commercial property common area of the mall and its corresponding daily maximum temperature curve for 2017, which are normalized values. As shown in Figure 1, the greatest influence factor of Heating, Ventilation and Air Conditioning (HVAC) EC is the temperature, and it can be seen that the total EC of the property public area of the shopping mall fluctuated, obviously, according to the season and had strong non-linear and fluctuating characteristics. Besides the similar fluctuating trends of HVAC and the total EC of the property public area of the shopping mall, the other sub-items showed a steady trend and did not change with time.

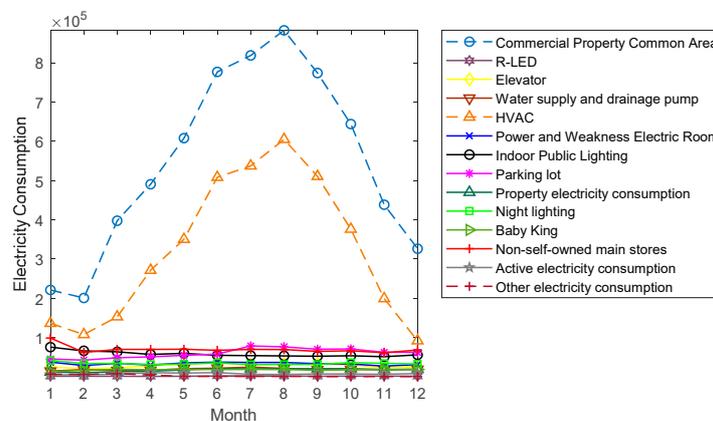
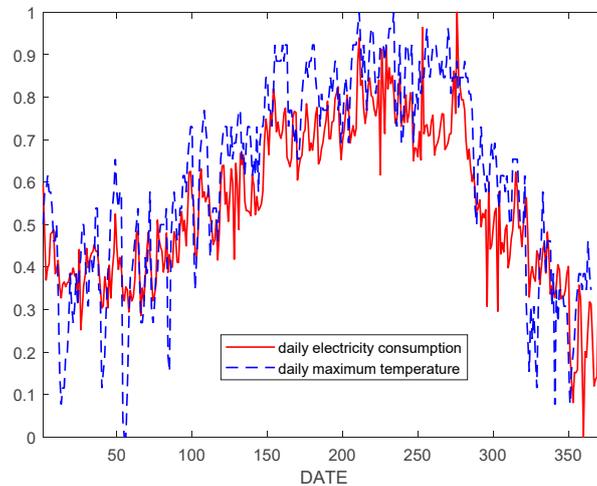


Figure 1. Monthly electricity consumption (EC) of a shopping mall in 2017.

Therefore, considering the influence of HVAC EC, it was possible to forecast the EC of the whole commercial property public area and further analyze the influencing factors of HVAC EC.

Comparing the trends of daily air conditioning power consumption and daily maximum temperature, we obtained Figure 2.



**Figure 2.** Daily EC of the commercial property common area of the mall and its corresponding daily maximum temperature.

In order to further study their correlation, we calculated Pearson correlation coefficients between historical HVAC daily EC values and their corresponding daily maximum temperature values.

$$\begin{aligned}\rho_{X,Y} &= \frac{\text{cov}(X,Y)}{\delta_X \delta_Y} \\ &= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\delta_X \delta_Y} \\ &= 0.9214\end{aligned}\quad (27)$$

The results showed that there was a great positive correlation between HVAC EC and air temperature. Therefore, air temperature and historical EC were input in the prediction model in order to improve the prediction accuracy of user EC.

#### 4.2. Data Preprocessing

In the process of forecasting, the difference of magnitude of the data may affect the accuracy of forecasting and the speed of forecasting, so the initial data should be normalized to ensure the accuracy of the forecasting results. In the equation below,  $y_i'$  indicates the normalized data. The formula for data normalization is as follows:

$$y_i' = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (28)$$

Generally, the initial data are normalized to [0,1] to ensure the speed and accuracy of the forecasting process.

After the prediction results are obtained, the obtained results should be anti-normalized to get the actual value of the predicted electricity. The inverse normalization formula is as follows:

$$y_i = (y_{\max} - y_{\min}) \times y_i' + y_{\min} \quad (29)$$

#### 4.3. Parameter Optimization

In order to ensure the accuracy of the results and the stability of the model, it is necessary to optimize the parameters of the ECF model. In this paper, the grid optimization method and K-fold cross-validation were combined to optimize the parameters.

It is convenient and simple to use the grid optimization method to optimize the parameters, because it is faster than other intelligent algorithms and can generally find a global optimal solution. The algorithm represents each set of parameters in the form of a range grid, traverses each set of parameters by using the independence of each grid, and then verifies the rationality of the set of parameters by K-fold cross-validation theory. Based on experience, most of the parameters  $\sigma_1$  and  $\sigma_2$  of MMCC were selected in  $[0, 50]$ . In the SVR model, both  $C$  and the kernel parameter  $\gamma$  have a wide range of values, but a too large range of values will lead to a lower search speed. After many experiments, the probability that  $\gamma$  falls into the  $[0, 1]$  interval is higher. Therefore, the grid range of the model parameters were specified in this paper as  $C = [0, 20], \gamma = [0, 1], \sigma_1 = [0, 50], \sigma_2 = [0, 50]$ .

#### 4.4. Model Implementation

This paper proposes an ECF model based on MMCCSVR considering the users' characteristics. Based on the robustness of MMCC in non-Gaussian sequence processing, the MMCCSVR ECF model was established to predict future EC. The parameters were optimized by combining grid optimization and K-fold cross-validation. The prediction model is

$$f(x) = \sum_{i=1}^n \omega_i K(x, x_i) + b \tag{30}$$

where  $x$  is input data, including historical EC and corresponding temperature data,  $f(x)$  is the forecasting results of EC.

The specific implementation process is shown in Figure 3.

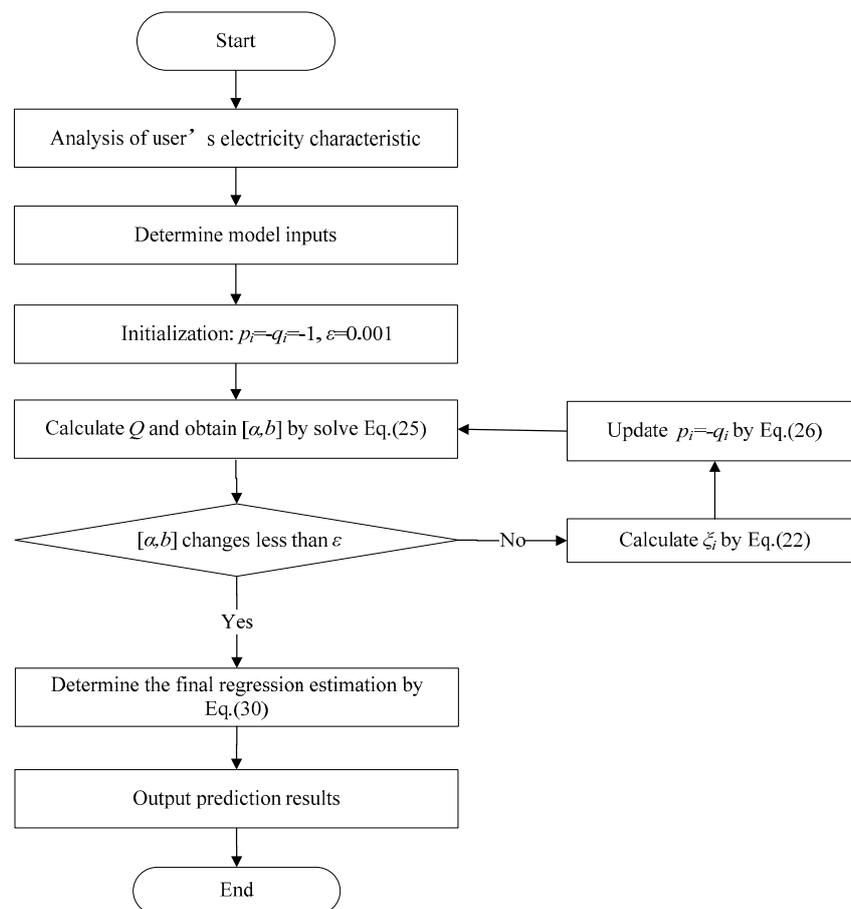


Figure 3. Flow chart of the implementation process.

#### 4.5. Evaluation Criterion

In order to comprehensively and extensively evaluate the effectiveness of the proposed forecasting model, five widely used performance metrics in statistics, i.e., mean absolute error (MAE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination ( $R^2$ ), were considered. The MAE, RMSE, and MAPE were used to measure the deviation between the actual EC and the prediction load. In addition,  $R^2$  indicates the extent to which the EC is predictable. These five metrics are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |f(x) - y_i| \times 100\% \tag{31}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|f(x) - y_i|}{y_i} \times 100\% \tag{32}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f(x) - y_i)^2} \times 100\% \tag{33}$$

$$R^2 = \frac{\left( N \sum_{i=1}^N y_i \times f(x) - \sum_{i=1}^N y_i \sum_{i=1}^N f(x) \right)^2}{\left[ N \sum_{i=1}^N f(x)^2 - \left( \sum_{i=1}^N f(x) \right)^2 \right] \cdot \left[ N \sum_{i=1}^N y_i^2 - \left( \sum_{i=1}^N y_i \right)^2 \right]} \tag{34}$$

### 5. Results

In order to verify the practicability and forecasting efficiency of the proposed ECF method, we considered the EC data of a shopping mall in Guangdong Province from 1 January 2018, to 3 June 2018, and the corresponding daily maximum temperature as the research object. We selected the daily EC from 1 January to 3 May as the training sample, forecasted the EC of 31 days from 4 May to 3 June 2018, one day ahead, and evaluated comprehensively the ECF model based on MMCCSVR.

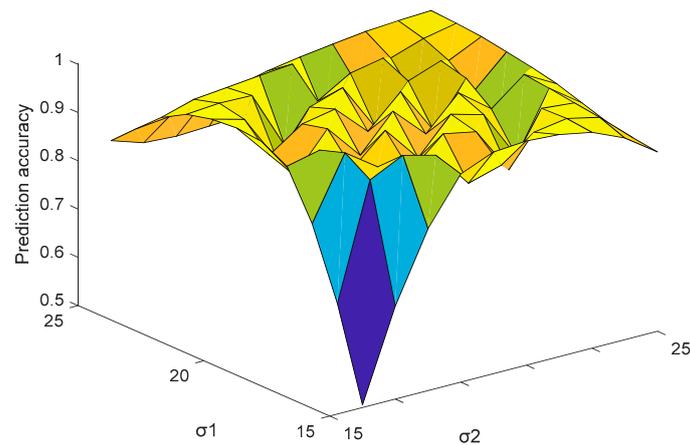
#### 5.1. Parameters Selection

In this subsection, the joint efficiency of the free parameters for the prediction accuracy was investigated by using the grid optimization and conducting many experiments. Because the penalty coefficient  $C$  showed various empirical values in different applications, and different  $\lambda$  values correspond to a set of different optimal  $\sigma_1$  and  $\sigma_2$ , which can make the prediction accuracy meet the requirements, we fixed  $C = 7$  and  $\lambda = 0.3$ , and then determined the optimal  $\sigma_1$  and  $\sigma_2$ . As can be seen in Figure 4, when  $\sigma_1$  and  $\sigma_2$  were within the interval [15,25], the prediction accuracy was higher. Taking  $\sigma_1$  as an example, we further studied the variation of prediction accuracy with kernel parameters.

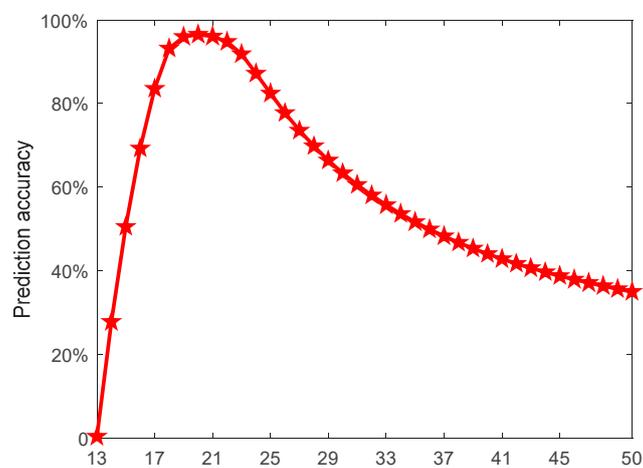
Figure 5 and Table 1 show the prediction accuracy corresponding to different  $\sigma_1$  values. The highest prediction accuracy was achieved at  $\sigma_1 = 20$ ; therefore,  $\sigma_1 = 20$  was chosen in this paper. Using the same method, we obtained the optimal  $\sigma_2 = 15$  in this case.

**Table 1.** Prediction accuracy varying with the parameter  $\sigma_1$ .

$\sigma_1$	15	16	17	18	19	20	21	22	23	24	25
<b>Prediction accuracy (%)</b>	50.5	69.2	83.6	93.2	96.0	98.2	96.1	94.8	91.8	87.2	82.4



**Figure 4.** Prediction accuracy varying with the parameters  $\sigma_1$  and  $\sigma_2$ .



**Figure 5.** Prediction accuracy varying with the parameter  $\sigma_1$ .

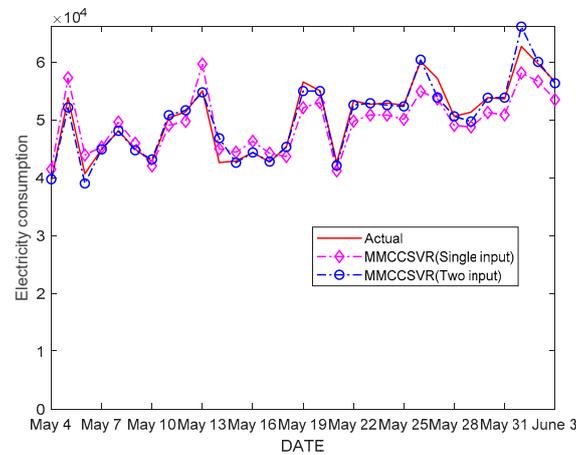
### 5.2. Comparison of the Forecast Results Obtained with Different Inputs

According to the parameter selection method described in Sections 4.3 and 5.1, the parameters were set as:  $C = 7$ ,  $\lambda = 0.3$ ,  $\sigma_1 = 20$ ,  $\sigma_2 = 15$ .

In order to further illustrate the necessity of using the temperature as an input variable, the prediction results of two kinds of inputs were compared with those of using only historical electric energy as input. The final prediction results are shown in Figure 6. It can be seen from Figure 6 that when only historical electric energy was used as input, a better prediction value could be obtained at the smooth point of the curve, but the prediction at the inflection point was not as good as that of the MMCCSVR prediction model with temperature as input. By comparing the error analysis results of the proposed method with those obtained by using only historical EC as input, it was found that the average MAPE was about 3.29% lower than that of the proposed method. Therefore, the prediction accuracy can be improved to a certain extent by using two kinds of input: EC and temperature.

The relative errors of daily forecasts are shown in Table 2. The characteristics of the shopping mall EC vary in the week. From the data, it appears that the EC is low on Monday and peaks on Saturday and Sunday. Between 4 May and 3 June 2018, the weekends were 5–6 May, 12–13 May, 19–20 May, 26–27 May, and 2–3 June. As shown in Table 2, the peak and valley values of EC were accurately predicted in the first half month, whereas the predicted values were mostly lower than the true values in the second half month. According to the analysis of historical electricity and temperature data, the temperature in May 2018 was slightly higher than that in May of the previous year, but the

thunderstorm weather in the first half of the month may have affected customer flow and activities in the shopping mall. The weather in the second half of the month continued to be cloudy, and the EC increased, since the temperature was slightly higher than in previous years.



**Figure 6.** Forecast results of the maximum mixture correntropy criterion support vector regression (MMCCSVR) method for a mall from 4 May to 3 June 2018.

**Table 2.** Forecast results of MMCCSVR for a mall from 4 May to 3 June 2018.

Day	Actual (kW·h)	Single Input (kW·h)	MAPE (%)	Two Input (kW·h)	MAPE (%)
4 May	39,442	41,463	5.12	397,19	0.70
5 May	53,823	41,463	6.43	52,141	-3.13
6 May	40,666	57,283	8.01	38,997	-4.10
7 May	40,666	43,925	11.41	44,889	10.38
8 May	48,099	45,305	3.16	48,100	0.00
9 May	44,929	49,621	2.32	44,735	-0.43
10 May	42,820	45,973	-1.82	43,160	0.79
11 May	50,363	42,039	-2.60	50,820	0.91
12 May	51,381	49,056	-3.18	51,659	0.54
13 May	55,039	59,663	8.40	54,774	-0.48
14 May	42,610	44,984	5.57	46,824	9.89
15 May	42,886	44,408	3.55	42,586	-0.70
16 May	44,358	46,295	4.37	44,364	0.01
17 May	42,699	44,224	3.57	42,777	0.18
18 May	45,329	43,647	-3.71	45,329	0.00
19 May	56,549	52,112	-7.85	54,990	-2.76
20 May	55,039	53,006	-3.69	54,984	-0.10
21 May	42,610	41,170	-3.38	42,090	-1.22
22 May	53,268	49,759	-6.59	52,574	-1.30
23 May	52,707	50,851	-3.52	52,874	0.32
24 May	52,913	50,834	-3.93	52,546	-0.69
25 May	52,547	50,085	-4.69	52,329	-0.41
26 May	60,092	54,925	-8.60	60,408	0.53
27 May	57,188	53,659	-6.17	53,875	-5.79
28 May	50,614	49,047	-3.10	50,614	0.00
29 May	51,341	48,776	-5.00	49,713	-3.17
30 May	53,857	51,274	-4.80	53,825	-0.06
31 May	53,620	50,864	-5.14	53,849	0.43
1 June	62,691	58,118	-7.29	66,143	5.51
2 June	59,699	56,645	-5.12	60,045	0.58
3 June	56,619	53,493	-5.52	56,348	-0.48
	MAPE		5.08%		1.79%

Table 2 shows the predicted values using MMCCSVR and the traditional SVR from 4 May to 3 June 2018, and the MAPE between each predicted value and the corresponding actual value. As can be seen from Table 2, compared with the SVR model, the MMCCSVR model allowed to reduce the MAPE of almost every predicted value, especially when approaching the turning point of fluctuation. The results show that using MMCC as the loss function of SVR can significantly improve the prediction accuracy.

### 5.3. Comparison of Different Forecasting Methods

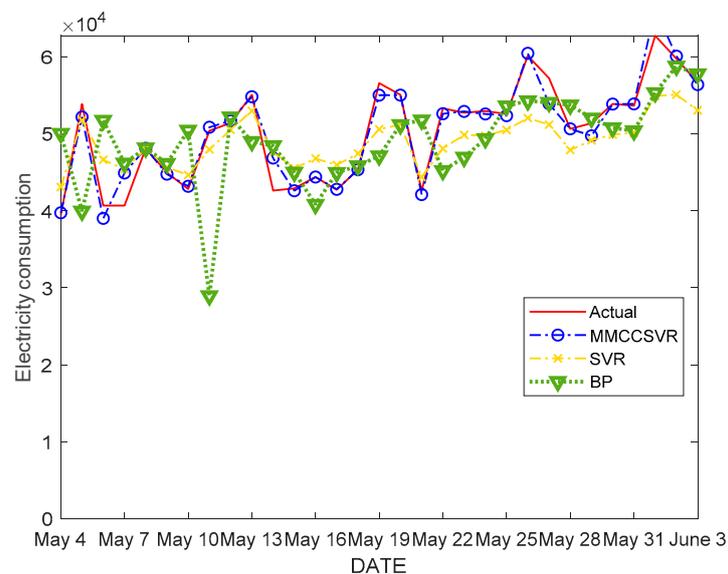
In order to test whether the proposed model can effectively improve the prediction accuracy, its prediction results were compared with those of other methods. We set the parameters as follows:

SVR:  $C = 7, \gamma = 0.1$ ;

Back-propagation algorithm (BP): 10 Hidden layer.

MMCCSVR:  $C = 7, \lambda = 0.3, \sigma_1 = 20, \sigma_2 = 15$ ;

The comparison results are shown in Figure 7. As shown in Figure 7, at the inflection point of the curve, the prediction efficiency of MMCCSVR is obviously better than those of other algorithms, showing the robustness and stability of the proposed method. Moreover, the predicted values changed appropriately during the analysis period, even coinciding with the true values at some individual points.



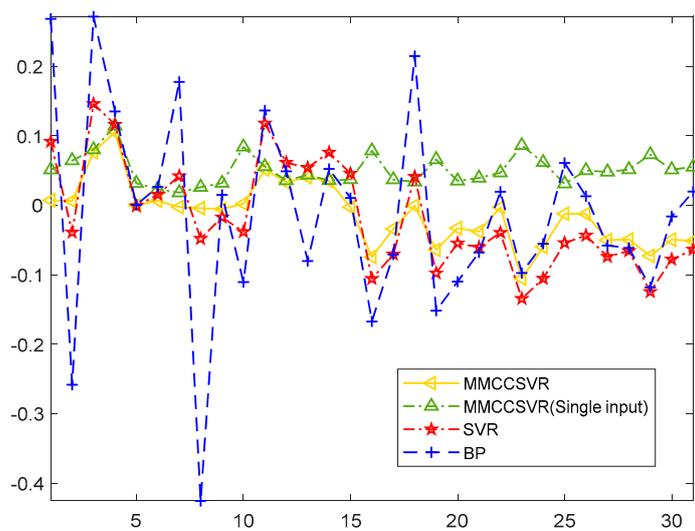
**Figure 7.** Forecast results of MMCCSVR compared with those of other methods from 4 May to 3 June 2018.

The percentages of relative error of the different methods are shown in Table 3. Table 3 shows that MMCCSVR has better performance compared with other algorithms, displaying improved prediction accuracy of both global and outlier values.

The relative errors of MMCCSVR for different inputs compared with those of other methods from 4 May to 3 June 2018, are presented in Figure 8. As shown in Figure 8, the relative error of MMCCSVR was evenly distributed on both sides of 0 and had little fluctuation. The relative error of MMCCSVR using historical electricity as input was very large, due to the fact that the temperature changes were not taken into account. BP and traditional SVR algorithms were unable to predict electricity accurately under the influence of random factors.

**Table 3.** Percentage of relative error of different methods. BP: Back-propagation.

Day	MMCCSVR	SVR	BP
4 May	0.70	9.17	26.81
5 May	-3.13	-3.88	-25.81
6 May	-4.10	14.59	27.14
7 May	10.38	11.60	13.49
8 May	0.00	-0.12	0.06
9 May	-0.43	1.54	2.64
10 May	0.79	4.21	17.74
11 May	0.91	-4.80	-42.53
12 May	0.54	-1.73	1.49
13 May	-0.48	-3.81	-11.05
14 May	9.89	11.78	13.64
15 May	-0.70	6.12	4.89
16 May	0.01	5.42	-8.01
17 May	0.18	7.59	5.22
18 May	0.00	4.58	1.04
19 May	-2.76	-10.57	-16.71
20 May	-0.10	-7.06	-7.08
21 May	-1.22	4.11	21.46
22 May	-1.30	-9.77	-15.16
23 May	0.32	-5.46	-10.99
24 May	-0.69	-6.04	-6.78
25 May	-0.41	-3.94	1.95
26 May	0.53	-13.44	-9.73
27 May	-5.79	-10.54	-5.53
28 May	0.00	-5.42	6.09
29 May	-3.17	-4.34	1.30
30 May	-0.06	-7.40	-5.79
31 May	0.43	-6.56	-6.13
1 June	5.51	-12.45	-11.81
2 June	0.58	-7.79	-1.61
3 June	-0.48	-6.56	1.96
MAPE	1.79%	6.84%	10.70%



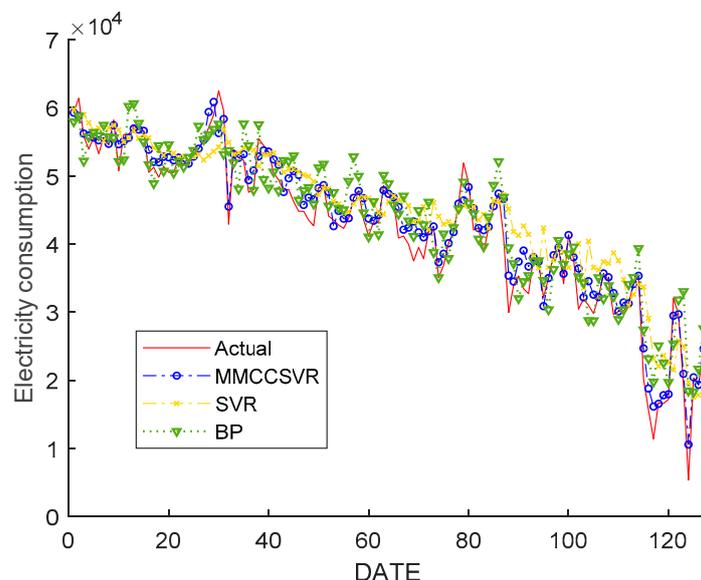
**Figure 8.** Relative error of MMCCSVR with different inputs compared with other methods from 4 May to 3 June 2018.

Table 4 proves the superiority of MMCCSVR over single-input MMCCSVR and the traditional SVR and BP algorithm for four error evaluation indexes; the four error evaluation indexes were obviously reduced, and in particular, the MAPE of this method was about 1.79%, and the  $R^2$  was about 0.9781. Therefore, compared with the SVR method and the BP neural network algorithm, the proposed method effectively improved the prediction accuracy, and when temperature was added as input, the prediction efficiency was further improved.

**Table 4.** Comparison of electricity consumption forecasting (ECF) accuracy. MAPE: mean absolute percentage error, MAE: mean absolute error, RMSE: root-mean-square error,  $R^2$ : coefficient of determination.

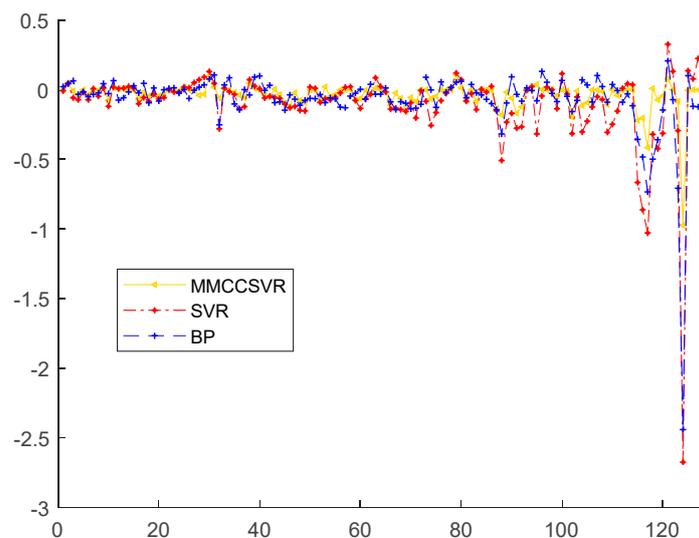
Method	MAPE	MAE	RMSE	$R^2$
MMCCSVR	1.79%	875.8387	1515.228	0.9781
MMCCSVR (Single Input)	5.08%	2582.8387	2836.0348	0.9150
SVR	6.84%	3460.8710	3951.0136	0.9304
BP	10.70%	5220.8065	6957.5602	0.3541

Considering the change of EC for a shopping mall in different months, we implemented the method by using EC data for a whole year, using 65% of the data as the training set and 35% as the testing set. We selected the EC data from 1 January to 25 August 2018 as the training sample and forecasted the EC data from 26 August to 31 December 2018, one-day ahead. The forecasted results are shown in Figure 9. As shown in Figure 9, when more data were used to validate the model proposed in this paper, the prediction curve was still close to the actual curve. Only a few curve inflection points presented inevitable errors due to random factors and the impact of emergencies, resulting in larger errors. In addition, at the inflection point of the curve, the prediction ability of MMCCSVR was obviously better than those of other algorithms, showing its robustness and stability. The MMCCSVR achieved good overall prediction, with the predicted values even coinciding with the true values at individual points.



**Figure 9.** Forecast results of MMCCSVR for a mall from 26 August to 31 December 2018.

The relative errors of MMCCSVR for different inputs compared with those of other methods for 26 August to 31 December 2018, are given in Figure 10. As shown in Figure 10, the relative error of MMCCSVR is evenly distributed on both sides of 0 and has little fluctuation. BP and traditional SVR algorithms were unable to predict electricity accurately under the influence of random factors.



**Figure 10.** Relative error of MMCCSVR for different inputs compared with those of other methods from 26 August to 31 December 2018.

Table 5 proves the superiority of MMCCSVR over the traditional SVR and BP algorithm, as shown by four error evaluation indexes; when the experimental data increased, the four error evaluation indexes were obviously reduced. We found that the MAPE of this method was about 3.86% and  $R^2$  was about 0.9846. Therefore, compared with the SVR method and BP neural network algorithm, the proposed method has good predictive performance which can effectively improve the prediction accuracy.

**Table 5.** Forecast results of MMCCSVR for a mall from 26 August to 31 December 2018.

Method	MAPE	MAE	RMSE	$R^2$
MMCCSVR	3.86%	1528.2	2289.7	0.9846
SVR	13.78%	3966.4375	6180.0521	0.9173
BP	10.43%	3123.5748	3978.9582	0.9481

## 6. Conclusions

Around the urgent demand of ECF technology for power sales companies under the wave of power system reform, a novel forecast method for ECF, called MMCCSVR, is proposed, combining the SVR with MMCC. First, the user's characteristics were analyzed to determine the factors influencing EC, indicating that the historical temperature and historical EC are the main influencing factors. Second, in view of the strong fluctuation of EC for small consumers and the large influence of random factors, the MMCCSVR algorithm with excellent performance under non-Gaussian noise was selected to construct the prediction model, which was applied to a shopping mall in Guangzhou. The following conclusions were reached:

- (1) Compared with the single-input MMCCSVR prediction model, the single-point prediction accuracy was effectively improved, and the average relative error was reduced.
- (2) Compared with the traditional SVR and other algorithms, the prediction errors of peak and valley values of EC were improved effectively.
- (3) The prediction error MAPE of this model was 1.79% and met the assessment criteria of power deviation in the location of the shopping mall and the prediction accuracy requirement of the power sales company.

**Author Contributions:** Conceptualization, J.D.; Data curation, X.T.; Investigation, L.A.; Methodology, X.T.; Validation, X.Q. and P.W.; Writing—original draft, X.T.; Writing—review & editing, W.M.

**Funding:** This work was supported by State Key Laboratory of Electrical Insulation and Power Equipment (EIPE18201), the National Natural Science Foundation of China (51877174), Key Project of Natural Science Basic Research Plan in Shaanxi Province of China (2019JZ-05), National Key R&D Program of China (2016YFB0901900), State Key Laboratory of Advanced Electromagnetic Engineering and Technology (Huazhong University of Science and Technology, AÉET2018KF001).

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

## References

1. Cao, Y.K.; Shuai, H. Medium and long-term load forecasting based on multi-factors modified PSO-SVM algorithm. Electronics, electrical engineering and information science. In Proceedings of the 2015 International Conference on Electronics, Electrical Engineering and Information Science (EEEIS2015), Guangzhou, China, 7–8 August 2015; pp. 682–693.
2. Niu, D.X.; Ma, T.N.; Liu, B.Y. Power load forecasting by wavelet least squares support vector machine with improved fruit fly optimization algorithm. *J. Comb. Optim.* **2017**, *33*, 1122–1143.
3. Sadaei, H.J.; Guimarães, F.G.; da Silva, C.J. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. *Int. J. Approx. Reason.* **2017**, *83*, 196–217. [[CrossRef](#)]
4. Zhao, H.; Guo, S. An optimized grey model for annual power load forecasting. *Energy* **2016**, *107*, 272–286. [[CrossRef](#)]
5. Hu, R.; Wen, S.; Zeng, Z. A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. *Neurocomputing* **2017**, *221*, 24–31. [[CrossRef](#)]
6. Li, Y.; Che, J.; Yang, Y. Subsampled support vector regression ensemble for short term electric load forecasting. *Energy* **2018**, *164*, 160–170. [[CrossRef](#)]
7. Yang, Y.; Che, J.; Deng, C. Sequential grid approach based support vector regression for short-term electric load forecasting. *Appl. Energy* **2019**, *238*, 1010–1021. [[CrossRef](#)]
8. Dong, Y.; Zhang, Z.; Hong, W. A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting. *Energy* **2018**, *11*, 1009. [[CrossRef](#)]
9. Hong, W.; Li, M.; Geng, J. Novel chaotic bat algorithm for forecasting complex motion of floating platforms. *Appl. Math. Model.* **2019**, *72*, 425–443. [[CrossRef](#)]
10. Zhong, H.; Wang, J.; Jia, H. Vector field-based support vector regression for building energy consumption prediction. *Appl. Energy* **2019**, *242*, 403–414. [[CrossRef](#)]
11. Liu, W.F.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [[CrossRef](#)]
12. Chen, B.D.; Xing, L.; Zhao, H.; Zheng, N.; Principe, J.C. generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]
13. Ma, W.; Chen, B.; Duan, J. Diffusion maximum correntropy criterion algorithms for robust distributed estimation. *Digit. Signal Process.* **2016**, *58*, 10–19. [[CrossRef](#)]
14. Wu, Z.; Shi, J.; Zhang, X. Kernel recursive maximum correntropy. *Signal Process.* **2015**, *117*, 11–16. [[CrossRef](#)]
15. Ma, W.; Qu, H.; Gui, G. Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments. *J. Franklin Inst.* **2015**, *352*, 2708–2727. [[CrossRef](#)]
16. Qian, G.; Wang, S. Generalized complex correntropy: Application to adaptive filtering of complex data. *IEEE Access.* **2018**, *6*, 19113–19120. [[CrossRef](#)]
17. Loza, C.A.; Principe, J.C. Generalized correntropy matching pursuit: A novel, robust algorithm for sparse decomposition. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada, 25–29 July 2016; pp. 1723–1727.
18. Luo, X.; Sun, J.; Wang, L. Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4963–4971. [[CrossRef](#)]
19. Chen, B.; Ma, R.; Yu, S. Granger causality analysis based on quantized minimum error entropy criterion. *IEEE Signal Process. Lett.* **2019**, *26*, 347–351. [[CrossRef](#)]
20. Wang, S.; Dang, L.; Chen, B. Random Fourier filters under maximum correntropy criterion. *IEEE Trans. Circuits I* **2018**, *65*, 3390–3403. [[CrossRef](#)]

21. Hou, B.; He, Z.; Zhou, X. Maximum correntropy criterion kalman filter for  $\alpha$ -jerk tracking model with non-gaussian noise. *Entropy* **2017**, *19*, 648. [[CrossRef](#)]
22. Ma, W.; Zheng, D.; Zhang, Z. Sparse-aware bias-compensated adaptive filtering algorithms using the maximum correntropy criterion for sparse system identification with noisy input. *Entropy* **2018**, *20*, 407. [[CrossRef](#)]
23. Zhou, N.; Xu, Y.; Cheng, H. Maximum correntropy criterion based sparse subspace learning for unsupervised feature selection. *IEEE Trans. Circuits Syst. Video* **2017**, *99*, 404–417. [[CrossRef](#)]
24. Wang, Y.; Li, Y.; Albu, F. Group-constrained maximum correntropy criterion algorithms for estimating sparse mix-noised channels. *Entropy* **2017**, *19*, 432. [[CrossRef](#)]
25. Wu, Z.; Peng, S.; Chen, B. Robust Hammerstein adaptive filtering under maximum correntropy criterion. *Entropy Switz.* **2015**, *17*, 7149–7166. [[CrossRef](#)]
26. Chen, B.; Wang, X.; Lu, N. Mixture correntropy for robust learning. *Pattern Recognit.* **2018**, *79*, 318–327. [[CrossRef](#)]
27. Brereton, R.G.; Lloyd, G.R. Support vector machines for classification and regression. *Analyst* **2010**, *135*, 230–267. [[CrossRef](#)]
28. He, R.; Hu, B.G.; Zhang, W.S. Robust principal component analysis based on maximum correntropy criterion. *IEEE Trans. Image Process.* **2011**, *20*, 1485–1494.
29. He, R.; Zheng, W.S.; Tan, T. Half-quadratic-based iterative minimization for robust sparse representation. *IEEE Trans. Pattern Anal.* **2013**, *36*, 261–275.
30. Wang, Y.; Tang, Y.Y.; Li, L. Correntropy matching pursuit with application to robust digit and face recognition. *IEEE Trans. Cybern.* **2016**, *47*, 1354–1366. [[CrossRef](#)]
31. Yan, H.; Yuan, X.; Yan, S. Correntropy based feature selection using binary projection. *Pattern Recognit.* **2011**, *44*, 2834–2842. [[CrossRef](#)]
32. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
33. Ojeda, F.; Suykens, J.; Moor, B.D. Low rank updated LS-SVM classifiers for fast variable selection. *Neural Netw.* **2008**, *21*, 437–449. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).