

Received December 27, 2019, accepted January 9, 2020, date of publication January 15, 2020, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966712

# Electric Load Forecasting by Hybrid Self-Recurrent Support Vector Regression Model With Variational Mode Decomposition and Improved Cuckoo Search Algorithm

ZICHEN ZHANG<sup>1</sup>, WEI-CHIANG HONG<sup>1</sup> , (Senior Member, IEEE), AND JUNCHI LI<sup>2</sup>

<sup>1</sup>School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, China

<sup>2</sup>Department of Medical Ultrasonics, Xuzhou No.1 Peoples Hospital, Xuzhou 221006, China

Corresponding author: Wei-Chiang Hong (samuelsonhong@gmail.com)

This work was supported by the Jiangsu Normal University, China, under Grant 9213618401.


**ABSTRACT** Accurate electric load forecasting is critical not only in preventing wasting electricity production but also in facilitating the reasonable integration of clean energy resources. Hybridizing the variational mode decomposition (VMD) method, the chaotic mapping mechanism, and improved meta-heuristic algorithm with the support vector regression (SVR) model is crucial to preventing the premature problem and providing satisfactory forecasting accuracy. To solve the boundary handling problem of the cuckoo search (CS) algorithm in the cuckoo birds' searching processes, this investigation proposes a simple method, called the out-bound-back mechanism, to help those out-bounded cuckoo birds return to their previous (the most recent iteration) optimal location. The proposed self-recurrent (SR) mechanism, inspired from the combination of Jordan's and Elman's recurrent neural networks, is used to collect comprehensive and useful information from the training and testing data. Therefore, the self-recurrent mechanism is hybridized with the SVR-based model. Ultimately, this investigation presents the VMD-SR-SVRCBCS model, by hybridizing the VMD method, the SVR model with the self-recurrent mechanism, the Tent chaotic mapping function, the out-bound-back mechanism, and the cuckoo search algorithm. Two real-world datasets are used to demonstrate that the proposed model has greater forecasting accuracy than other models.

**INDEX TERMS** Support vector regression, variational mode decomposition, self-recurrent mechanism, tent chaotic mapping function, out-bound-back mechanism, cuckoo search algorithm.

## I. INTRODUCTION

Along with the huge economic growth in China, the electricity consumption in each sector, such as industrial production, mining exploration, economic business administrations, educational activities, and the residential usages, has simultaneously increased. Therefore, providing sufficient electricity reserves for all users will be the important functions of the government. On the other hand, excessive electricity generation will waste the country's non-renewable resources and cause environmental pollution. To successfully prevent the

production of waste, it is an essential challenge to provide suitable electricity production planning to meet the equilibrium from both supply and demand sides. In addition, energy reserves reducing caused from inaccurate electric load forecasting would lower other industries' energy budgets, while it is an unacceptable problem for those resource-saving developing countries like China [1]. Therefore, more accurate electric load forecasting is not only the critical success factor among the highly competitive electricity markets, but also the essential support for national energy policy [2]. However, due to lots of exogenous variables, such as weather statuses, economical productions, promotion activities, and special dates (e.g., weekends, holidays and festivals) [3], the electric load

The associate editor coordinating the review of this manuscript and approving it for publication was Wuhui Chen .

data demonstrate fluctuation tendency, nonlinear characteristics, and chaotic nature, which let the electric load forecasting become a complicate research topic to conduct the electric load forecasting [4], [5].

### A. TRADITIONAL FORECASTING MODELS

Many improving approaches to enhance higher load forecasting accuracy have been widely explored in past decades. Those load forecasting approaches have two categories: the first one is the traditional forecasting models which are mostly based on statistics, and the other is based on artificial intelligent technologies. The traditional models are devoted to look for the recurrent relationships of the electric loads themselves among the previous time periods. The well-known statistical models include ARIMA models [6]–[8], exponential smoothing models [9], [10], regression models [11]–[13], Kalman filter models [14], [15], and Bayesian models [16], [17]. For instance, Yuan *et al.* [6] employ the ARIMA model to conduct the energy demand forecasting in China. The forecasting results indicate that the fitted values of ARIMA model respond less to the fluctuations in the long-term trend. Dong *et al.* [10] propose a partially linear additive quantile regression model to forecast the short-term electricity demand during the peak hours using South African data from 2009 to 2012. The results show that the proposed model could provide a minimal cost of unit scheduling and electricity dispatching to system operators during the peak period. Dudek [11] proposes a new forecasting framework by using the ensemble Kalman filter technique. The employed ensemble Kalman filter technique can extract detail analytical information (e.g., temperature response rate) by modeling the nonlinear electric load. The forecasting results show that the forecasting performance of the proposed models significantly outperforms other alternative models. Those statistical models could be established easily, however, based on their theoretical definitions, these models are difficult to get the substantial improvements, which largely limit their forecasting ability and reduce their forecasting accuracy. These models are based on linear definitions; thus, they could hardly deal with nonlinear modeling, particularly suffering from very complex tendency changes in electric load [18].

### B. ARTIFICIAL INTELLIGENT FORECASTING MODELS

With the significant technology development in recent years, the artificial intelligent models have received lots of attentions to improve forecasting accuracy, such as, knowledge-based system (KBS) models [19], [20], artificial neural network (ANN) [21]–[24], and fuzzy inference models [25]–[28]. For instance, Karimi *et al.* [19] provide a new priority index to determine the similar days by considering the similarity of the temperature and the special date. The issued model illustrates its superior advantages than other alternative models in terms of modeling time and forecasting accuracy. Singh and Dwivedi [21] provide a novel algorithm (i.e., follow the leader), they hybridize the algorithm with the ANN to properly tune its parameters to conduct the load fore-

casting problem. Three real-world electric load data sets are used to compare the forecasting performance, and the results show that the superiority of the proposed model. Hippert and Taylor [16] provide a novel framework by hybridizing fuzzy model with optimized parameter, which is a bio-inspired optimizer combination between two heuristic approaches. The experimental results illustrate that the suitability and the superiority of the proposed approach in load forecasting. However, these proposes artificial intelligent models also have several drawbacks, such as time consuming, slow convergent speed, subjective to determine suitable network architecture [29], and easily trapping into local optimum [30]; more relevant discussions of artificial intelligent technologies in electricity demand forecasting could be found in [31], [32]. Thus, to continue exploring innovative forecasting frameworks, approaches would still be an essential issue.

### C. SUPPORT VECTOR REGRESSION MODEL WITH META-HEURISTIC ALGORITHMS IN FORECASTING

Due to simultaneously considering the minimization principle of the structural risk and the novel loss function of Vapnik's  $\epsilon$ -insensitivity, support vector machines (SVMs) have been successfully extended to solve the nonlinear problems, particularly for time series forecasting, i.e., the support vector regression (SVR) model [33]. Recently, the SVR model has received widely explorations in many professional fields, such as financial time series forecasting [34]–[37], tourism forecasting [38], [39], atmospheric rainfall forecasting [40]–[42]. For electric load forecasting fields, authors also provide a series research results, by hybridizing novel searching techniques, such as hybridizing with cloud generator to simulate the continuous process of the temperature annealing [33], hybridizing with the classical chaotic mapping functions to avoid higher homogeneity of the population for each meta-heuristic algorithms [43], [44], and quantum searching mechanism [45], [46]. Based on the conclusions of these researches, selecting an appropriate parameter combination of the SVR model would receive better forecasting performances, i.e., to simultaneously select suitable values of  $C$  (the tradeoff among the flatness of function and training errors),  $\epsilon$  (the width of the  $\epsilon$ -insensitive function), and  $\sigma$  (the critical parameter of the Gaussian kernel function) is still a deserved issue in the SVR research fields. To continue exploring the feasibility in terms of hybridizing meta-heuristic algorithm with an SVR model and providing associate improving arrangements to conquer the critical drawbacks of the employed approach, this investigation tries to apply the cuckoo search (CS) algorithm [47] to select appropriate parameters of the SVR model. The primary concepts of the CS algorithm are inspired from the phenomenon of the interesting parasitic mechanism of the cuckoo birds, which includes the followings actions: imitating the pattern and the color from the hosts, throwing the eggs out of the hosts' nests, and constructing a new nest. As a meta-heuristic algorithm, the CS algorithm also has some problems as the same as other algorithms, such as slow convergent

performances, homogeneity of the population, and the premature convergent phenomenon [48], [49]. Besides, another outstanding problem of the CS algorithm is boundary handling problem. Therefore, this paper firstly hybridizes the Tent mapping function to avoid homogeneity of the cuckoo population in the searching space and to guarantee to escape from the local optima. Secondly, this investigation proposes an out-bound-back mechanism to overcome the boundary handling problem, by keeping those cuckoo birds which were out-bounded come back to previous (last iteration) optimal location.

#### D. SELF-RECURRENT MECHANISMS

In addition, to share the past and current information of electricity data, this article proposed a combined two famous recurrent mechanisms (Jordan's [50] and Elman's [51] recurrent neural networks), namely self-recurrent (SR) mechanism, to learn about: (1) more output differences (i.e., forecasting errors) from the testing results, which is based on Jordan's mechanism; and (2) the inherent characteristics (i.e., information structure of the training data), which is based on Elman's mechanism. By applying those learned knowledge (experiences) from these two mechanisms, the SR mechanism can help current training work to achieve more precise information explanation of the employed electricity data. As known that the basic design of the recurrent framework often uses input layer and the memory of previous layer (called 'context layer') as the input in the training stage [52]. This kind of recurrent framework could also be hybridized with the SVR model. Furthermore, the recurrent framework not only possess the link weight between units in different layers, but also has the special link weight from the hidden layer to the context layer, which is the critical factor that they are extensively applied to sequence-relational data forecasting [53]. Jordan's and Elman's recurrent mechanisms both contain the multilayer perceptron (MLP) with one hidden layer. The former one sets the 'context layer' to be represented the memory from output layer; the latter one sets 'context layer' to be represented the memory from hidden layer. They both implement the historical information interaction from layers and obtain more details from past information or trained knowledge. However, as mentioned in [54] that both Jordan's and Elman's mechanisms are with simple structure and outstanding dynamic characteristic. Therefore, in this investigation, the proposed SR mechanism combines both recurrent mechanisms (i.e., the 'context layer' is represented the memory from both output layer and from hidden layer) is employed to hybridize with an SVR model.

#### E. APPLICATIONS OF VARIATIONAL MODE DECOMPOSITION (VMD)

On the other hand, it is helpful to implement the data pre-processing procedure to reduce the non-linear and non-stationary characteristics, to obtain a more regular sub-series, such as the empirical mode decomposition (EMD) method [55]–[57]. After this kind of pre-processing

operation, the electric load forecasting model can closer fit each decomposed component (namely intrinsic mode function, briefed as IMF) to improve the forecasting performance. However, while data set suffers from the characteristics of mode aliasing (mode mixing), false modes, and many components with similar frequencies [58], the EMD method is limited to provide significant decomposed components to be further analyzed. Thus, the improvement of forecasting accuracy is also limited. Variational mode decomposition (VMD) method, proposed by Dragomiretskiy and Zosso [58], can adaptively decompose the data into the non-recursive frequency domain and transform them into variational modes with strong continuity and correlation [58], [59]. The VMD method has been employed to solve the data decomposition problems in many fields, such as financial analysis [60] and energy forecasting [61]. This paper would also employ the VMD method to successfully reduce the non-linearity and non-stationarity of the electric load data, and separate the data into feature component precisely.

#### F. CONTRIBUTIONS OF THIS PAPER

Therefore, the proposed hybrid the VMD method and the SR-SVR-based model with the CBCS algorithm, the so-called VMD-SR-SVRCBCS model, is used to receive higher forecasting accuracy while exploring nonlinear electric load data. This paper also employs ARIMA, SARIMA, BPNN, and GRNN models as alternative models to compare the forecasting accuracy with the proposed VMD-SR-SVRCBCS model. In the meanwhile, two datasets from worldwide famous electric load datasets are used to validate the performances of all compared models. The principal contributions are stated as follows,

- 1) An innovative hybrid electricity demand forecasting model is proposed, by applying the VMD method, the self-recurrent mechanism, the Tent mapping function, the out-bound-back mechanism, the CS algorithm, and the SVR model, namely VMD-SR-SVRCBCS model.
- 2) The VMD method is hybridized with the SR-SVR-based model to conduct data preprocessing to decompose more accurate IMFs (not aliasing in a certain range, less false modes, and less components with similar frequencies), and then apply SR-SVRCBCS model to model each decomposed IMFs to receive more accurate forecasting results. Please refer Section II.A.
- 3) The Tent mapping function is introduced to avoid homogeneity of the population during the searching periods of the CS algorithm to help cuckoo birds escape from the local optima. The out-bound-back mechanism is proposed to provide a guideline of these cuckoo birds while they fly out bound the defined domain, and capture them back to previous better location to ensure the searching quality. Please refer Section II.B. and D.
- 4) The self-recurrent mechanism is applied to the SVRCBCS model to help to capture more embedded data information during the training processes among

layers. Eventually, select more suitable parameters of the SVR model and to receive higher forecasting accuracy than other compared models; in addition, the performances also receive the significance under 95% confident levels. Please refer Section II.D.

## G. ORGANIZATION OF THIS PAPER

The rest of this paper is organized as follows. Section II presents the modeling details of the VMD-SR-SVRCBCS model, including the theoretical introduction of the SVR model, the CBCS algorithm, the out-bound-back mechanism, the VMD method, and the self-recurrent mechanism. Two numerical examples in real world are presented in Section III. Finally, the conclusion is proposed in Section IV.

## II. THE PROPOSED VMD-SR-SVRCBCS MODEL

### A. THE VARIATIONAL MODE DECOMPOSITION (VMD) METHOD

The variational mode decomposition (VMD) method, proposed by Dragomiretskiy and Zosso [58], is a newly non-recursive data processing approach to adaptively decompose an input data series into  $k$  discrete number of sub-series (modes) by obtaining the optimal solution of the constrained variational model, where each mode ( $u_k$ ) has a limited bandwidth with a unique center frequency ( $\omega_k$ ) in the frequency domain. During the VMD processes, the center frequency and bandwidth of each mode are constantly determined and the sum value of the estimated bandwidth is minimized [58]. In addition, the sum of the IMFs must be equal to the input data series,  $F(t)$ , which is the constraint condition. The process of estimating the bandwidth of each mode is as follows:

- 1) Hilbert transform is used to decompose the electric load data series,  $F(t)$ ; the analytic signal of each mode ( $u_k$ ) is calculated to obtain the associated unilateral frequency spectrum;
- 2) Apply an exponential tuned operator,  $e^{-j\omega_k t}$ , to mix with the estimated center frequency to modulate the mode's frequency spectrum to baseband;
- 3) Apply the  $H^1$  Gaussian smoothness (the  $L^2$  norm) of the demodulated signal gradient to estimate the bandwidth for each mode.

Subsequently, the constrained variational problem is described as (1),

$$\begin{aligned} \text{Min}_{\{u_k\}, \{\omega_k\}} & \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ \text{s.t.} & \sum_{k=1}^K u_k = F(t) \end{aligned} \quad (1)$$

where  $\{u_k\} = \{u_1, u_2, \dots, u_k\}$  is the set of all modes, i.e.,  $u_k$  represents the  $k$ th mode;  $\{\omega_k\} = \{\omega_1, \omega_2, \dots, \omega_k\}$  are the center frequencies of each corresponding mode, i.e.,  $\omega_k$  is the  $k$ th center frequency of  $u_k$ ;  $F(t)$  is the original electric load time series;  $\delta(t)$  represents the Dirac distribution;  $j$  is an

imaginary number, i.e.,  $j^2 = -1$ ; represents the convolution operator.

By introducing a quadratic penalty term,  $\mu$ , and Lagrange multipliers,  $\lambda(t)$ , the constrained variational problem can be converted to an unconstrained variable problem, as shown in (2),

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \ell) &= \mu \sum_{k=1}^K \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \left\| \mathcal{F}(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \langle \ell(t), \mathcal{F}(t) - \sum_{k=1}^K u_k(t) \rangle \end{aligned} \quad (2)$$

The alternate direction method of multipliers (ADMM) is used to find out the saddle point of Eq. (2) by updating  $u_k$  and  $\omega_k$ . As mentioned in [58], the convergent criterion is  $\sum_k \left( \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \left\| \hat{u}_k^n \right\|_2^2 \right) < \gamma$ , where  $\gamma$  is the convergence tolerance;  $\wedge$  represents the Fourier transforms. Consequently, the solutions for  $u_k$ ,  $\omega_k$ , and  $l(t)$  can be demonstrated as (3) to (5) [58],

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{\mathcal{F}}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega} \quad (4)$$

$$\hat{\ell}^{n+1}(\omega) = \hat{\ell}^n(\omega) + \tau \left[ \hat{\mathcal{F}}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right] \quad (5)$$

where  $\hat{\mathcal{F}}(\omega)$ ,  $\hat{u}_i(\omega)$ ,  $\hat{\ell}(\omega)$ ,  $\hat{u}_k^{n+1}(\omega)$  represent the Fourier transforms of  $\mathcal{F}(t)$ ,  $u_i(t)$ ,  $\ell(t)$ , and  $u_k^{n+1}(t)$ , respectively;  $n$  represents the number of iteration;  $\tau$  is the time step of the dual ascent.

### B. THE SUPPORT VECTOR REGRESSION (SVR) MODEL

The modeling details of the SVR model are briefed as follows. A nonlinear mapping function,  $\varphi(\cdot) : \mathfrak{R}^n \rightarrow \mathfrak{R}^{n_h}$ , is defined to map the original data(training data) into a so-called high dimensional feature space (which may have infinite dimensions),  $\mathfrak{R}^{n_h}$ . Given a set of data,  $G = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ , where  $\mathbf{x}_i$  is the input data;  $\mathbf{y}_i$  is the actual value, and  $N$  is the total number of the data set. Then, in the high dimensional feature space, there theoretically exists a linear function,  $f$ , i.e., the SVR function, to formulate the nonlinear relationship between input data and output data, as shown in (6),

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \quad (6)$$

where  $f(\mathbf{x})$  denotes the forecasting values; both  $\mathbf{w}$  ( $\mathbf{w} \in \mathfrak{R}^{n_h}$ ) and  $b$  ( $b \in \mathfrak{R}$ ) are adjustable coefficients, they are estimated by minimizing the following empirical risk function as shown

in (7),

$$R(f) = \frac{1}{N} \sum_{i=1}^N L_\varepsilon \left( \mathbf{y}_i, \mathbf{w}^T \varphi(\mathbf{x}_i) + b \right) \quad (7)$$

where  $L_\varepsilon(\mathbf{y}, f(\mathbf{x}))$  is the  $\varepsilon$ -insensitive loss function to penalize the training errors between  $f(\mathbf{x})$  and  $\mathbf{y}$ , and is defined as (8),

$$L_\varepsilon(\mathbf{y}, f(\mathbf{x})) = \begin{cases} 0 & \text{if } |\mathbf{y} - f(\mathbf{x})| \leq \varepsilon \\ |\mathbf{y} - f(\mathbf{x})| - \varepsilon & \text{otherwise} \end{cases} \quad (8)$$

The loss equals zero if the forecasted value is within the  $\varepsilon$ -insensitive tube (8). In addition,  $L_\varepsilon(\mathbf{y}, f(\mathbf{x}))$  is also employed to excellently divide the training data into two subsets in the so-called feature space by the optimized hyper plane. Thus, the SVR model is devoted to look for the optimal hyper plane and minimize the training errors between the training data ( $\mathbf{y}$ ) and the  $\varepsilon$ -insensitive tube ( $f(\mathbf{x})$ ).

After solving the quadratic optimization problem with inequality constraints, the SVR regression function is computed as (9),

$$f(\mathbf{x}) = \sum_{i=1}^N (\beta_i - \beta_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (9)$$

where  $\beta_i, \beta_i^*$  are the so-called Lagrangian multipliers, and they are obtained in the quadratic programming processes;  $K(\mathbf{x}_i, \mathbf{x}_j)$  is the so-called kernel function, it is defined as the inner product calculating for the feature values,  $\varphi(\mathbf{x}_i)$  and  $\varphi(\mathbf{x}_j)$ , from these two vectors,  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , respectively, i.e.,  $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ .

The Gaussian function with a width of  $\sigma$ :  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ , is well-known kernel function. Due to easily implementations and powerful nonlinear mapping capability [43], the Gaussian exponential kernel function (another classical Gaussian kernel function),  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\| / 2\sigma^2)$ , is employed in this investigation.

As mentioned above that the better forecasting performance of an SVR model always comes from an appropriate selection of its three parameters. Recently, authors have employed the Tent chaotic mapping function with the famous bionic-based meta-heuristic algorithm, i.e., the CS algorithm, to provide appropriate parameter combination of the SVR model, and receive some degree improvement in terms of forecasting accuracy [44]. However, some outstanding problem of the CS algorithm, i.e., the boundary handling problem, still deserves to be explored. Thus, authors propose the out-bound-back mechanism hybridizing with the Tent chaotic mapping function, namely CBCS algorithm, to receive more approximate parameter combination of the SVR model to enhance the forecasting accurate level.

### C. CBCS ALGORITHM

#### 1) TENT MAPPING FUNCTION

As defined as highly unstable/unpredictable motion in finite phase space, chaos often occurs in deterministic nonlinear dynamic systems [4]. It is sensitive caused from the

minute changes in initial condition (the so-called butterfly effect), and generally demonstrates multiple elements with non-linear interactions. Chaos could also be observed via a very simple equation, which is the so-called chaotic mapping function.

The chaotic mapping function could transform the data to have the chaotic ergodicity mentioned above (i.e., sensitive dependence on initial conditions), therefore, it is easily to avoid the homogeneity of population during the optimization procedures of any algorithms, to abound the searching tracks in the definition domain, and eventually to avoid premature problem. Based on the chaotic characteristics analysis, the Tent mapping function demonstrates faster iteration speed than the Logistic mapping function, more autocorrelation, and more suitable for a large number of sequences, i.e., with good ergodic uniformity. The Tent mapping function not only has better uniformity traversal and optimization efficiency, but also its initial value sensitivity of probability density distribution function is not strong, which is suitable for computing. In addition, randomly set the initial values in the interval (0,1) for the Logistic mapping function and the Tent mapping function, and set the iteration times as 10,000. Then, record the obtained chaotic values for each mapping function, and the statistical results are plotted as **Figs. 1** and **2**, respectively. It is obviously that the Tent mapping function has more chaotic features than the Logistic mapping function. Therefore, this paper uses the Tent mapping function during the modeling processes of the CS algorithm to assist those cuckoo birds to look for more suitable parameters of an SVR model.

The Tent mapping function is a one dimension piecewise-linear mapping defined by (10),

$$x_{n+1} = \begin{cases} x_n/a & x \in [0, a] \\ (1 - x_n)/(1 - a) & x \in (a, 1] \end{cases} \quad (10)$$

where  $x_n$  represents the  $n^{\text{th}}$  iterative value of the variable  $x$ , and  $n$  represents the number of the iteration.

When  $a = 0.5$ , it is the standard Tent mapping and shown in (11),

$$x_{n+1} = \begin{cases} 2x_n & x \in [0, 0.5] \\ 2(1 - x_n) & x \in (0.5, 1] \end{cases} \quad (11)$$

#### 2) CUCKOO SEARCH ALGORITHM

The primary concepts of the CS algorithm are based on the brood parasitic mechanism: the cuckoos' eggs are laid in the neighborhood nests. In addition, due to application of Lévy flight behaviors, cuckoo birds could accelerate the searching speed than the normal random walk. Therefore, the number of iterations could be successfully reduced and improve the local search performances [47]. The implementation of the CS algorithm, each egg in a nest represents a potential solution. Based on the Lévy flight behaviors, the cuckoo birds can soon choose to leave their eggs in the recently-spawned nests (host nests) to make sure that the left eggs could receive the priority to be hatched. Due to the natural hatching habits of the cuckoo

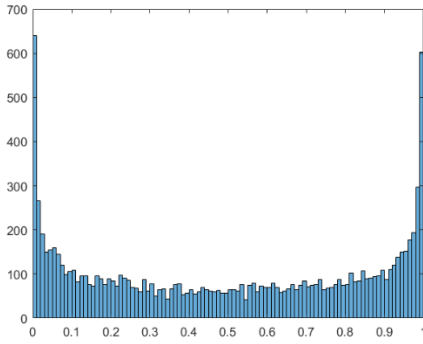


FIGURE 1. Histogram of 10,000 iterative distributions for Logistic mapping function.

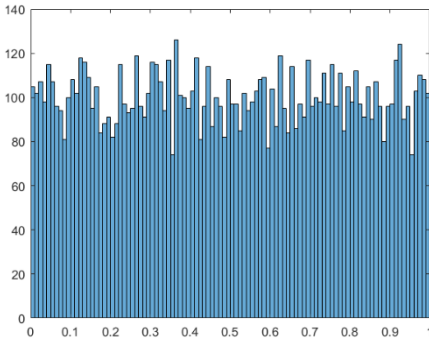


FIGURE 2. Histogram of 10,000 iterative distributions for Tent mapping function.

birds, they are usually without doubt to hatch the eggs in the host nests. Of course, the host cuckoo birds would have the probability,  $p_a$ , to detect out those eggs in their nests but not belong to them [49]. Once those “stranger” eggs have been detected out, the host cuckoo birds have two strategies, one is directly thrown out all eggs in their nest; the second one is to leave the old nest, then, establish a new nest in a completely new location. Then, based on Lévy flight behaviors, new eggs (solutions) would be laid continuously by the cuckoo birds around the current best solutions [62].

Based on the cuckoo birds’ behaviors analysis, three useful rules are required [62]: 1) each cuckoo is not permitted to lay more than one egg at a time in its selected host nest; 2) only those eggs with high-quality and the associate host nests could be survived for the next generations; 3) the number of the host nests should not be changed, and the detection probability of the host cuckoo birds is set as  $p_a \in [0, 1]$ . The last rule could be represented by using a probability ( $p_a$ ) of the  $n$  host nests will be replaced by new random nests. Here,  $p_a$  is often set as 0.25 [49]. Therefore, the core searching ability of the CS algorithm is the usage of the parameter,  $p_a$ , to trade off the balance between two kinds of searches, the local search and the global search, i.e., let the cuckoo birds discover new or more prospective regions. These two searches are defined as (12) and (13), respectively,

$$x_i^{t+1} = x_i^t + \alpha s \otimes H(p_a - \delta) \otimes (x_j^t - x_k^t) \quad (12)$$

$$x_i^{t+1} = x_i^t + \alpha \mathcal{L}(s, \lambda) \quad (13)$$

where  $x_j^t$  and  $x_k^t$  are the current positions which are randomly determined;  $\alpha$  is the positive scale variable and is defined as the size of Lévy flight step;  $s$  is defined as the step size for each local search;  $H(\cdot)$  is the Heavy-side function;  $\delta$  is a uniformly random number;  $\otimes$  is the product of  $H(\cdot)$  and the current positions;  $\mathcal{L}(s, \lambda)$  is the Lévy distribution as the step size of search, it is often defined as (14),

$$\mathcal{L}(s, \lambda) = \frac{\lambda \Gamma(\lambda) \sin(\pi \lambda / 2)}{\pi} \frac{1}{s^{1+\lambda}} \quad (14)$$

where  $\lambda$  is the standard deviation of these step sizes in the global search; the Gamma function,  $\Gamma(\lambda)$ , is defined as  $\Gamma(\lambda) = \int_0^\infty t^{\lambda-1} e^{-t} dt$ . Only when  $\lambda$  is a positive integer, it can also be represented as  $\Gamma(\lambda) = (\lambda - 1)!$ . The Lévy flight distribution guarantees the system to avoid premature problem [63].

### 3) THE OUT-BOUND-BACK MECHANISM

Based on the details of the CS algorithm illustrated above, the search space of any cuckoo birds is not limited to a specified range. However, after hybridizing the CS algorithm with an SVR model, the searching range for each parameter would be limited according to the employed electric load data sets. Thus, it is possible to suffer from the out-bound problem, i.e., the cuckoo birds may search for the host nest out of the limited bound set by the SVR modeling requirements, and the worst situation is that those cuckoo birds continue looking for the solution to the wrong direction which is far from the defined domain. The best improving method is to propose a warning mechanism, namely out-bound back mechanism in this paper, to let those out-bound cuckoo birds come back to the best solution (the location of the nest) in previous iteration.

The so-called out-bound behavior means that any cuckoo birds fly out of the defined domain in any dimension of the coordinate would be noticed and all behaviors in this iteration would also be cleaned, i.e., the best solution would not be updated and the locations of these birds would be turned back to the previous locations (not back to the bound) associate with the previous best solution, as shown in (15) and (16).

$$x_i^{t+1} = \begin{cases} x_i^t + \alpha s \otimes H(p_a - \delta) \otimes (x_j^t - x_k^t), & \text{if } x_j^t, x_k^t \in \text{defined domain} \\ x_i^t, & \text{otherwise} \end{cases} \quad (15)$$

$$x_i^{t+1} = \begin{cases} x_i^t + \alpha \mathcal{L}(s, \lambda), & \text{if } \alpha \mathcal{L}(s, \lambda) \in \text{defined domain} \\ x_i^t, & \text{otherwise} \end{cases} \quad (16)$$

More details of the proposed out-bound-back mechanism are demonstrated in Fig. 3. In which, it is clearly to see that the  $i^{\text{th}}$  cuckoo bird has conducted  $t$  iterations searching jobs, for the  $(t + 1)^{\text{th}}$  iteration searching, it flies out of the defined domain. At that moment, it would receive some warning from the mechanism and its searching status would be cleaned by the mechanism. In addition, it would be turned back to the previous locations associate with the previous best solution,



best location of the desired nest,  $x_{k,best}^{(i+1)}$ . In this paper, the forecasting error is computed based on the mean absolute percentage error (MAPE), as illustrated in (19),

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - f_i}{y_i} \right| \times 100\% \quad (19)$$

where  $N$  is the total number of data;  $y_i$  is the actual electricity value at the  $i^{\text{th}}$  point;  $f_i$  is the forecasted electricity value at the  $i^{\text{th}}$  point.

**Step 4 Cuckoo Global Search With the Out-Bound-Back Mechanism:** Use the best nest position,  $x_{k,best}^{(i+1)}$ , to implement cuckoo global search with the out-bound-back mechanism (16), and use (14) to update other nest positions, and compute the associate forecasting errors.

**Step 5 Determine New Nest Location:** Update and select the new nest location only with more appropriate fitness value among the generated nest positions in **Step 4** and from the previous iteration, and denote it as,  $x_{k,j}^{(t)} = [x_{k,1}^{(t)}, x_{k,2}^{(t)}, \dots, x_{k,n}^{(t)}]^T$ .

**Step 6 Cuckoo Local Search With the Out-Bound-Back Mechanism:** If  $p_a$  is smaller than a critical number  $r$ , then, use the new nest position,  $x_{k,j}^{(t)}$ , and turn to local search with out-bound-back mechanism (Eq. (15)). And, continue updating the new nests,  $x_{k,j}^{(t)}$ , only with smaller MAPE value.

**Step 7 Determine the Best Nest Location:** Determine the best nest location,  $x_{k,best}^{(t)}$ , only with the smallest fitness value among the new nest locations in **Step 6**,  $x_{k,j}^{(t)}$ , and the best nest position,  $x_{k,best}^{(i+1)}$ .

**Step 8 Stop Criteria:** If the number of searching iterative loops exceeds the given criterion, then, the best nest location,  $x_{k,best}^{(t)}$ , in **Step 7** could be verified as the most appropriate parameters (i.e.,  $C$ ,  $\sigma$ , and  $\varepsilon$ ) of the SVR model; otherwise, go back to **Step 2** and start the next iteration.

#### D. SELF-RECURRENT (SR) MECHANISM

As mentioned above, the proposed self-recurrent (SR) mechanism combines the superiorities of both Jordan's and Elman's recurrent mechanisms as the learning framework. It is clearly to see that, except the context layer, the neurons in the same layer are connected with all neurons in their forward layer. The context layer is a specified hidden layer. It interacts with itself, the hidden layer, and the output layer.

For the proposed SR mechanism with an SVR model, suppose there are  $p$  input neurons ( $\mathbf{x} = [x_1, x_2, \dots, x_p]$ ),  $q$  hidden neurons,  $r$  output neurons and  $s$  context neurons, then, the output of the  $n^{\text{th}}$  neuron in the  $t^{\text{th}}$  iteration,  $\hat{f}_n(t)$ , is shown as (20),

$$\hat{f}_n(t) = \sum_{i=1}^q \omega_i^{ho} \Phi_i(t) + b \quad (20)$$

where  $\omega_i^{ho}$  is the weight linking with the  $i^{\text{th}}$  hidden neuron and output neuron,  $b$  is the same as in (6), and  $\Phi_i$  is the hidden layer output of the  $i^{\text{th}}$  hidden neuron in the  $t^{\text{th}}$  iteration, which

is defined as (21),

$$\Phi_i(t) = g \left( \sum_{j=1}^p \omega_{ji}^{ih} x_j + \sum_{m=1}^s \omega_{mi}^{ch} \Phi_i(t-1) + \sum_{m=1}^s \sum_{v=1}^r \omega_{mvi}^{ch} f_v(t-m) \right) \quad (21)$$

where  $\omega_{ji}^{ih}$  is the weight linking with the  $j^{\text{th}}$  input neuron and the  $i^{\text{th}}$  hidden neuron;  $\omega_{mi}^{ch}$  is the weight linking with the  $m^{\text{th}}$  context neuron and the  $i^{\text{th}}$  hidden neuron;  $\omega_{mvi}^{ch}$  is the weight linking with the  $m^{\text{th}}$  context neuron and the  $v^{\text{th}}$  output neuron with delay periods;  $\Phi_i(t-1)$  is the context layer output of the  $i^{\text{th}}$  neuron in the  $t^{\text{th}}$  iteration, which is the same with the hidden layer output of the  $i^{\text{th}}$  neuron in the  $(t-1)^{\text{th}}$  iteration;  $f_v(t-m)$  is the output layer output of the  $v^{\text{th}}$  neuron in the  $(t-m)^{\text{th}}$  iteration, which is the same with the context layer output of the  $m^{\text{th}}$  neuron in the  $t^{\text{th}}$  iteration.

Therefore, the output of the SVR model in the  $t^{\text{th}}$  iteration,  $\hat{f}_n(t)$  could be computed as (22),

$$\hat{f}_n(t) = \sum_{i=1}^q \omega_i^{ho} g \left( \sum_{j=1}^p \omega_{ji}^{ih} x_j + \sum_{m=1}^s \omega_{mi}^{ch} \Phi_i(t-1) + \sum_{m=1}^s \sum_{v=1}^r \omega_{mvi}^{ch} f_v(t-m) \right) + b \quad (22)$$

Then, (21) replaces (6) in the SVR modeling processes, to hybridize with the proposed CBCS algorithm and implement the parameter combination optimization procedure, and eventually, receive the forecasting values.

### III. NUMERICAL EXAMPLE

#### A. THE EMPLOYED DATA SETS

To comprehensively illustrate the advantages of the proposed VMD-SR-SVRCBCS model, two famous electric load data sets are employed in this paper. The first one is collected from the National Electricity Market (NEM), Queensland region, Australia, namely Queensland Example, which is based on half-hour load data type. The second example is acquired from the New York Independent System Operator (NYISO), New York, USA, namely New York Example, which is based on hourly load data type.

In Queensland Example, the data set has 1,200 half-hour electric load values in total, for detail time period is from the half of 0 o'clock on 1 January 2017 to the end of 24 o'clock on 25 January 2017. As Schalkoff [64] recommends that the ratio of the number of the validation data set to the number of the training data set should approximate to one to four. Thus, in this example, the electricity demand data set is partitioned into three parts. The first part is the training set, which contains 768 half-hour demand values, from the half of 0 o'clock on 1 January 2017 to the end of 24 o'clock on 16 January 2017. The second part is the validation set, which includes 192 half-hour demand values, the half of 0 o'clock on 17 January 2017 to the end of 24 o'clock on



**TABLE 1. Training, validation and testing data sets of the proposed model.**

Data sets	Regions	
	Queensland*	New York#
Training data set	Jan 1 <sup>st</sup> - 16 <sup>th</sup> , 2017	Jan 1 <sup>st</sup> - Feb 1 <sup>st</sup> , 2018
Validation data set	Jan 17 <sup>th</sup> -20 <sup>th</sup> , 2017	Feb 2 <sup>nd</sup> -9 <sup>th</sup> , 2018
Testing data set	Jan 21 <sup>th</sup> -25 <sup>th</sup> , 2017	Feb 10 <sup>th</sup> -19 <sup>th</sup> , 2018

\*National Electricity Market (NEM), Queensland region, Australia.  
 #New York Independent System Operator (NYISO), New York, USA.

20 January 2017. The third part is the testing set, which contains 240 half-hour demand values, from the half of 0 o'clock on 21 January 2017 to the end of 24 o'clock on 25 January 2017.

In New York Example, the employed data set has 1,200 hourly electricity demand values in total, from 1 o'clock on 1 January 2018 to the end of 24 o'clock on 19 February 2018. In this example, the electricity demand data set is also partitioned into three parts, based on Schalkoff's suggestion, the training set is with 768 hourly demand values (i.e., from 1 o'clock on 1 January 2018 to the end of 24 o'clock on 1 February 2018), the validation set is with 192 hourly load values (i.e., from 1 o'clock on 2 February 2018 to the end of 24 o'clock on 9 February 2018), and the testing set is with 240 hourly load values (i.e., from 1 o'clock on 10 February 2018 to the end of 24 o'clock on 19 February 2018). The data set divisions for these two examples are demonstrated in **Table 1**. To compare with the same modeling conditions, other alternative models also have the same data partitions.

**B. FORECASTING RESULTS OF THE VMD-SR-SVRCBCS MODEL**

**1) THE SETTINGS OF THE VMD METHOD AND CBCS ALGORITHM'S INTERNAL PARAMETERS**

In the VMD processing, the quadratic penalty term ( $\mu$ ) for these two examples are both set to be 2,000, and the time step of the dual ascent ( $\tau$ ) are both set to be 1.

The settings of the internal parameters in the CBCS algorithm are briefed as followings. The maximal number of nests is put as fifty; the maximal number of computing iterative loops is 10; the detection probability,  $p_a$ , is recommended as 0.25 [49]. The three parameters of an SVR model are set as,  $C \in [0, 18000]$ ,  $\sigma \in [0, 5]$  and  $\varepsilon \in [0, 1]$ . In the meanwhile, to take the iterative time into account that it would seriously affect the forecasting performances of each compared model, thus, the modeling time for each compared model is equal as far as possible.

**2) FORECASTING ACCURACY INDEXES**

Four forecasting performance indexes are employed to comprehensively evaluate the forecasting accuracy for those compared models, they are: 1) the MAPE, as mentioned in Eq. (19); 2) the mean absolute error (MAE); 3) the mean

square error (MSE); and 4) the root mean square error (RMSE). The latter three indexes are shown in (23) to (25), respectively,

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - f_i| \tag{23}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2 \tag{24}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2} \tag{25}$$

where  $N$  is the total number of data;  $y_i$  is the actual electricity value at the  $i^{\text{th}}$  point;  $f_i$  is the forecasted electricity value at the  $i^{\text{th}}$  point.

**3) SIGNIFICANT TESTS FOR OUTSTANDING FORECASTING PERFORMANCES**

To verify the forecasting significance of the VMD-SR-SVRCBCS model, some useful statistical tests are recommended to be conducted. Derrac *et al.* [65] indicate that for large sample size test (240 half-hour/hour loads forecasting in both examples), Friedman test is suitable. In addition, the Wilcoxon signed-rank test can be used to make simple pairwise comparisons. Therefore, these two statistical tests are used herein.

Wilcoxon signed-rank test is used to determine the significance of the forecasting errors made by two forecasting models with the same sample number. Let  $e_i$  be the absolute forecasting errors in the  $i$ th forecasts from these two forecasting models. If  $e_i > 0$ , then let  $r^+$  be the sum of ranks; if  $e_i < 0$ , let  $r^-$  be the sum of ranks, and if  $e_i = 0$ , eliminate this comparison, and delete the sample size. The statistic  $W$  is defined as in (26):

$$W = \min \{r^+, r^-\} \tag{26}$$

Friedman test is one of the non-parameter statistical tests in ANOVA analysis procedure, it devotes to conduct the comparison of the significance between two models or among more models. The statistic ( $F$ ) of Friedman test is illustrated as (27),

$$F = \frac{12N}{u(u+1)} \left[ \sum_{j=1}^u R_j^2 - \frac{u(u+1)^2}{4} \right] \tag{27}$$

where  $N$  is the total number of forecasted values;  $u$  is the total number of alternative models;  $R_j$  is the average rank sum received in each forecasted value for each alternative model as demonstrated in (28),

$$R_j = \frac{1}{N} \sum_{i=1}^N r_i^j \tag{28}$$

where  $r_i^j$  is the  $i$ th forecasted result among the  $j$ th compared model, which is based on the rank sum from the best forecast-

ing result (ranked as the  $I^{st}$ ) to the worst forecasting result (ranked as the  $k^{th}$ ).

Eventually, if the  $p$ -value does not meet the acceptable criterion, then, the null hypothesis could not be held, i.e., the significance of the forecasting performance from the proposed VMD-SR-SVRCBCS model could be verified.

#### 4) FORECASTING RESULTS AND ANALYSIS FOR QUEENSLAND EXAMPLE

With respect to Queensland Example, the CBCS algorithm is firstly applied to search for the appropriate parameter combination of an SVR model that is with the smallest forecasting error in terms of the MAPE index value. During the training stage, the famous rolling-based procedure [64] is implemented to assist the CBCS algorithm to carefully consider the fluctuation change between training points, and to well look for appropriate parameter of the SVR model. In this example, in the first run, the first 576 load data in the training set (totally 768 load data) is used by the CBCS algorithm to minimize the empirical risk, then, the first forecasting load (i.e., the 577<sup>th</sup> forecasting load) by the SVRCBCS model is received. Secondly, the next 576 load data (from 2<sup>nd</sup> to 577<sup>th</sup> load values) are employed by the SVRCBCS model to receive the second forecasting load (i.e., the 578<sup>th</sup> forecasting load). Repeat this procedure till the remained 192 forecasting loads are eventually received, then, the training error could be further calculated. Similarly, based on the validation dataset, the validation error could also be calculated after iteration stopped in the validation stage. The potential parameters with the smallest training and validation errors would be selected as the appropriate parameters for the SVRCBCS model. Then, the 240 half-hour forecasting demand would be eventually forecasted by the SVRCBCS model.

In addition, the finalized SVRCBCS model with the smallest testing MAPE value would be used to implement the SR-SVRCBCS model. Eventually, the forecasting results and the suitable parameters for the SVRCS model (hybridizing the CS algorithm with the SVR model), for the SVRCBCS model (hybridizing the CBCS algorithm with the SVR model), and for the SR-SVRCBCS model (hybridizing SR mechanism with the SVRCBCS model) in the Queensland Example would be illustrated in **Table 2**. In which, it is clearly to illustrate that the SVRCBCS model receives more accurate forecasting performances than the SVRCS model. In addition, the SR-SVRCBCS model obtains the highest forecasting accuracy among these hybrid CS with SVR-based models.

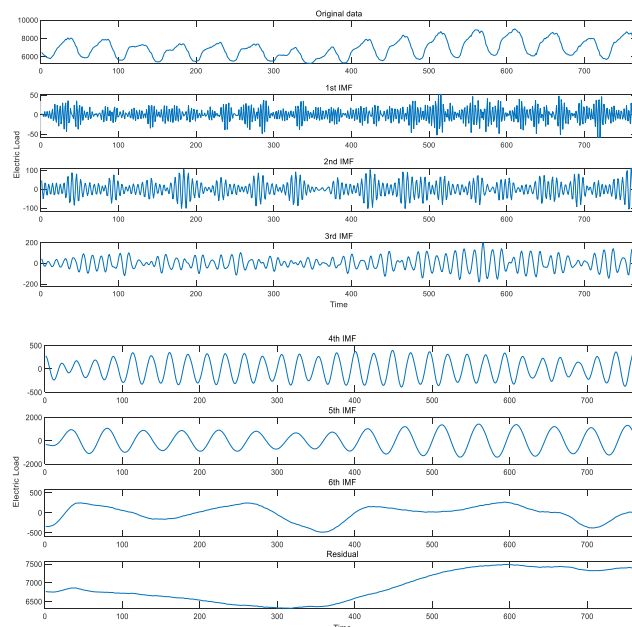
Secondly, the VMD method is employed to the original electric load data from the National Electricity Market (NEM), Queensland region, Australia, and six IMFs and one residual term are obtained. The decomposition result is shown in **Fig. 5**. Then, the SR-SVRCBCS model is used to forecast these six IMFs and the residual term, separately. **Table 3** illustrates the best suitable parameters for each IMF and the residual term.

**TABLE 2. Parameters and forecasting accuracy indexes of SVRCS, SVRCBCS, and SR-SVRCBCS models (Queensland example).**

Models	Parameters			Forecasting accuracy indexes			
	C	$\sigma$	$\epsilon$	MAE	MSE	MAPE(%)	RMSE
SVRCS	13,650	0.16	0.33	175.5	38,240	2.5	195.5
SVRCBCS	14,670	0.50	0.17	92.2	17,230	1.3	131.3
SR-SVRCBCS	17,180	0.33	0.01	77.7	11,630	1.1	107.9

**TABLE 3. The optimized parameters of the VMD-SR-SVRCBCS model for each IMF and residual (Queensland example).**

IMFs/Residual	Parameters		
	C	$\sigma$	$\epsilon$
IMF1	4,324.9	1.76	0.45
IMF2	4,217.0	4.22	0.02
IMF3	3,734.3	0.26	0.24
IMF4	7,430.2	0.41	0.06
IMF5	17,302.0	0.36	0.45
IMF6	9,263.0	0.91	0.02
Residual	10,983.0	1.81	0.48



**FIGURE 5. The decomposed IMFs and the residual term in Queensland Example.**

The forecasting results of these IMFs and the residual term are reconstructed, then, the electric load forecasting results can be obtained by adding the forecasting values of these IMFs and the residual term. For comparing the forecasting superiority of the proposed VMD-SR-SVRCBCS model, authors provide a four-level comparison structure to comprehensively compare the superiority of the proposed VMD-SR-SVRCBCS model. Thus, for the first comparison level, it is devoted to compare the superiority of the original SVR model against other artificial intelligent approaches, such as  $ARIMA(7,0,8)$ ,  $SARIMA(7,0,8) \times (5,0,4)$ , GRNN ( $\sigma = 1.0$ ), and BPNN models. For the second comparison level, it is devoted to compare the superiority of the proposed SVRCS model

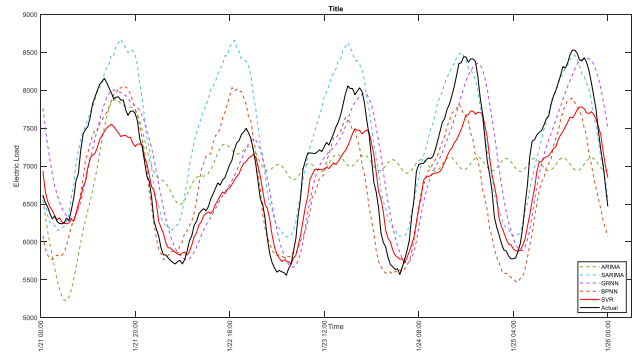
**TABLE 4.** Forecasting accuracy indexes of the VMD-SR-SVRCBCS model and other models (Queensland example).

Compared Models	Forecasting accuracy indexes			
	MAE	MSE	MAPE(%)	RMSE
<i>ARIMA</i> <sub>(7,0,8)</sub>	684.6	$6.67 \times 10^5$	10.1	816.9
<i>SARIMA</i> <sub>(7,0,8)×(5,0,4)</sub>	439.3	$3.38 \times 10^5$	6.5	581.3
GRNN ( $\sigma=1.0$ )	409.6	$2.39 \times 10^5$	6.0	489.3
BPNN	452.2	$2.92 \times 10^5$	6.3	540.5
SVR	361.3	$1.86 \times 10^5$	4.9	430.9
SVRPSO	272.3	$1.31 \times 10^5$	3.9	361.5
SVRBA	298.4	$1.63 \times 10^5$	4.4	403.6
SVRFFA	267.3	$1.37 \times 10^5$	3.9	370.4
SVRCS	175.5	$3.82 \times 10^4$	2.5	195.5
SVRCBPSO	156.8	$4.68 \times 10^4$	2.2	216.3
SVRCBBA	159.0	$5.12 \times 10^4$	2.3	226.3
SVRCBFFA	151.4	$4.70 \times 10^4$	2.2	216.7
SVRCBCS	92.2	$1.72 \times 10^4$	1.3	131.3
SR-SVRCBCS	77.7	$1.16 \times 10^4$	1.1	107.9
VMD-SR-SVRCBCS	<b>62.3</b>	<b><math>7.32 \times 10^3</math></b>	<b>0.9</b>	<b>85.5</b>

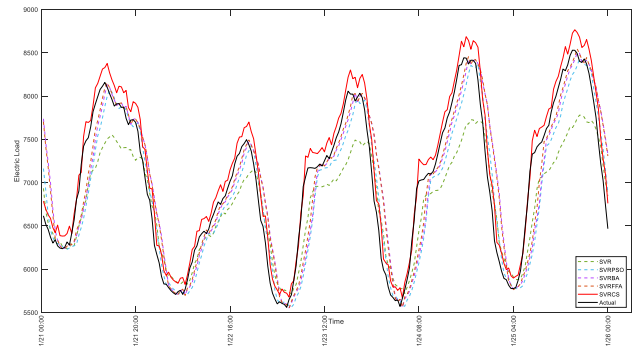
against other SVR-based model with different swarm-based algorithms (including particle swarm optimization (PSO) algorithm, bat algorithm (BA), cuckoo search (CS) algorithm, and firefly algorithm (FFA)). For the third comparison level, it is devoted to compare the superiority of the proposed chaotic out-bound-back mechanism among different SVR-swarm-based models. For the fourth comparison level, it is devoted to compare the superiority of the proposed SR mechanism and the VMD method among other SVR-based models. **Table 4** presents the forecasting performance evaluation indexes for these four comparison levels.

For the first comparison level, the original SVR model receives superior forecasting accuracy among *ARIMA*<sub>(7,0,8)</sub>, *SARIMA*<sub>(7,0,8)×(5,0,4)</sub>, GRNN ( $\sigma = 1.0$ ), and BPNN models. For the second comparison level, the proposed SVRCS model also receives superior forecasting accuracy among other SVR models with different swarm-based algorithms. For the third comparison level, the forecasting error of the proposed SVR-CBCS model is smaller than other SVR-CB-based models with different swarm-based algorithms. Finally, for the fourth comparison level, the proposed VMD-SR-SVRCBCS and SR-SVRCBCS models receive the best and the second-best forecasting performance among other alternative models.

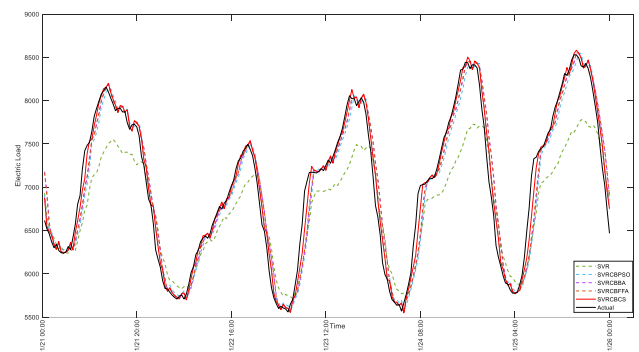
These comparison results of the four comparison levels obviously reveal the advantages of the VMD method, the SR mechanism, the chaotic-out-bound-back mechanism, and CS algorithm, respectively. In addition, the comparison figures of these four levels are demonstrated in **Fig. 6** to **9**, respectively. In **Fig. 6**, it is obviously to see that the SVR model (red line) can effectively catch the trend of the actual load and is superior to other models. In **Fig. 7**, the SVRCS model (red line) is able to accurately simulate the changing patterns of the actual load than other SVR-based models with different swarm-based algorithms. In **Fig. 8**, the chaotic-out-bound-back mechanism can significantly improve the forecasting performance of each employed swarm-based algorithms. In which, the proposed SVRCBCS model is superior to other SVR-CB-based models. In **Fig. 9**, the



**FIGURE 6.** Forecasting comparison level 1: the original SVR model v.s. other models for Queensland Example.



**FIGURE 7.** Forecasting comparison level 2: the SVRCS model v.s. other models for Queensland Example.



**FIGURE 8.** Forecasting comparison level 3: the SVRCBCS model v.s. other models for Queensland Example.

VMD-SR-SVRCBCS model is the closest one to the actual electricity demand than other alternative models. The VMD method really plays well the role in decomposing the electric load data into six different IMFs and the residual term.

Finally, to verify the forecasting accuracy improvement is with significance in each comparison level. As mentioned above, the Friedman test is conducted based on one-tail style, under one significant level,  $\alpha = 0.05$ . The significant test results in each comparison level are shown in **Tables 5** to **8**, respectively. These tables clearly indicate that the SVR, the SVRCS, the SVRCBCS, and the VMD-SR-SVRCBCS

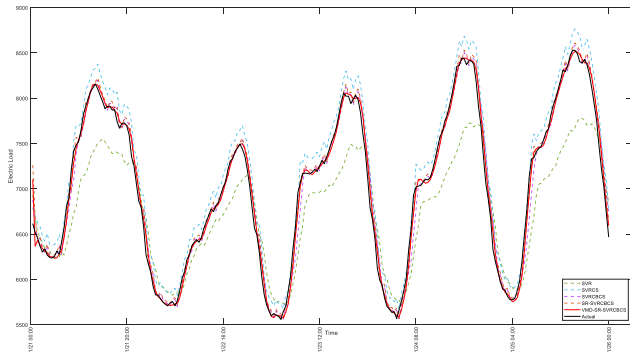


FIGURE 9. Forecasting comparison level 4: the VMD-SR-SVRCBCS model v.s. other models for Queensland Example.

TABLE 5. Results of Wilcoxon signed-rank test and Friedman test for comparison level 1 (Queensland example).

Compared models	Wilcoxon Signed-rank Test $\alpha = 0.05; p\text{-value}$	Friedman Test ( $\alpha = 0.05$ , one-tail)
SVR vs. $ARIMA_{(7,0,8)}$	0.0000**	$H_0: e_1 = e_2 = e_3 = e_4 = e_5$
SVR vs. $SARIMA_{(7,0,8) \times (5,0,4)}$	0.0000**	$F = 276.8300$
SVR vs. GRNN ( $\sigma=1.0$ )	0.0000**	$p = 0.00001^{**}$
SVR vs. BPNN	0.0000**	(Reject $H_0$ )

\*\* represents the test result does not accept the null hypothesis under  $\alpha = 0.025$ .

TABLE 6. Results of Wilcoxon signed-rank test and Friedman test for comparison level 2 (Queensland example).

Compared models	Wilcoxon Signed-rank Test $\alpha = 0.05; p\text{-value}$	Friedman Test ( $\alpha = 0.05$ , one-tail)
SVRCBCS vs. SVR	0.0000**	$H_0: e_1 = e_2 = e_3 = e_4 = e_5$
SVRCBCS vs. SVRPSO	0.0000**	$F = 363.8033$
SVRCBCS vs. SVRBA	0.0000**	$p = 0.00001^{**}$
SVRCBCS vs. SVRFFA	0.0000**	(Reject $H_0$ )

\*\* represents the test result does not accept the null hypothesis under  $\alpha = 0.025$ .

TABLE 7. Results of Wilcoxon signed-rank test and Friedman test for comparison level 3 (queensland example).

Compared models	Wilcoxon Signed-rank Test $\alpha = 0.05; p\text{-value}$	Friedman Test ( $\alpha = 0.05$ , one-tail)
SVRCBCS vs. SVR	0.0000**	$H_0: e_1 = e_2 = e_3 = e_4 = e_5$
SVRCBCS vs. SVRBPPO	0.0000**	$F = 137.0767$
SVRCBCS vs. SVRCBBA	0.0000**	$p = 0.00001^{**}$
SVRCBCS vs. SVRCBFFA	0.0000**	(Reject $H_0$ )

\*\* represents the test result does not accept the null hypothesis under  $\alpha = 0.025$ .

models receive the significance than other alternative models in each comparison level, respectively.

### 5) FORECASTING RESULTS AND ANALYSIS FOR NEW YORK EXAMPLE

In New York Example, the CBCS algorithm is also used to determine the appropriate parameter combination of an SVR model with the smallest MAPE index value. The rolling-based procedure is also employed in the training stage of the VMD-SR-SVRCBCS modeling process. The forecasting results and the suitable parameters for the SVRCBCS

TABLE 8. Results of Wilcoxon signed-rank test and Friedman test for comparison level 4 (queensland example).

Compared models	Wilcoxon Signed-rank Test $\alpha = 0.05; p\text{-value}$	Friedman Test ( $\alpha = 0.05$ , one-tail)
VMD-SR-SVRCBCS vs. SVR	0.0000**	$H_0: e_1 = e_2 = e_3 = e_4 = e_5$
VMD-SR-SVRCBCS vs. SVRCBCS	0.0000**	$F = 499.0833$
VMD-SR-SVRCBCS vs. SVRCBCS	0.0000**	$p = 0.00001^{**}$
VMD-SR-SVRCBCS vs. SR-SVRCBCS	0.0000**	(Reject $H_0$ )

\*\* represents the test result does not accept the null hypothesis under  $\alpha = 0.025$ .

TABLE 9. Parameters and forecasting accuracy indexes of SVRCBCS, SVRCBCS, and SR-SVRCBCS models (New York example).

Models	Parameters			Forecasting accuracy indexes			
	C	$\sigma$	$\epsilon$	MAE	MSE	MAPE(%)	RMSE
SVRCBCS	70,810	0.09	0.40	872.8	1,142,000	5.2	1,068.6
SVRCBCS	14,620	0.22	0.03	515.5	529,800	2.9	727.9
SR-SVRCBCS	18,770	0.22	0.07	458.0	41,080	2.6	640.9

TABLE 10. The optimized parameters of the VMD-SR-SVRCBCS model for each IMF and residual (New York example).

IMFs/Residual	Parameters		
	C	$\sigma$	$\epsilon$
IMF1	2576.4	1.71	0.35
IMF2	4313.3	1.70	0.36
IMF3	2354.2	0.93	0.03
IMF4	4544.9	0.49	0.50
IMF5	10772.0	0.23	0.96
IMF6	7786.7	0.40	0.22
Residual	16057.0	1.89	0.82

model, the SVRCBCS model, and the SR-SVRCBCS model in this example are shown in Table 9. In which, it also clearly illustrates the same results as in previous example, i.e., the SVRCBCS model once again receives higher forecasting accuracy than the SVRCBCS model. In addition, the SR-SVRCBCS model obtains the highest forecasting accuracy among these hybrid CS with SVR-based models.

Secondly, the VMD method is also applied to the original electric load data from the New York Independent System Operator (NYISO), New York, USA, and six IMFs and one residual term are also obtained. The decomposition result is shown in Fig. 10. The VMD-SR-SVRCBCS model is then employed to forecast these six IMFs and the residual term, separately. Table 10 presents the best suitable parameters for each IMF and the residual term.

The forecasting results of these IMFs and the residual term are further reconstructed, the electric load forecasting results can be obtained by adding the forecasting values of these IMFs and the residual term. For demonstrating the forecasting superiority of the VMD-SR-SVRCBCS model, four comparison levels as shown in Section III-B-4) are also conducted. The comparison figures of the forecasting results for these four levels are demonstrated in Figs. 11 to 14, respectively. Similar to Queensland Example, for each comparison level, the proposed mechanisms (including the chaotic-outbound-back mechanism, the recurrent mechanism, and the VMD method) can significantly improve the forecasting

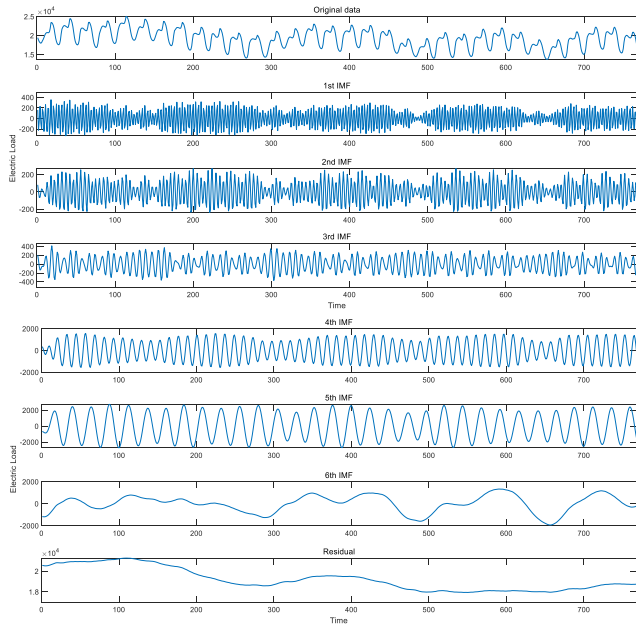


FIGURE 10. The decomposed IMFs and the residual term in New York example.

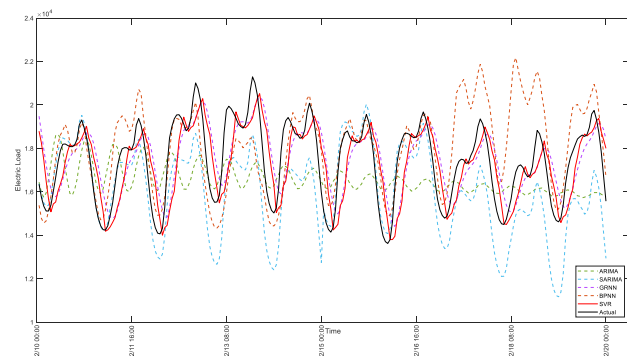


FIGURE 11. Forecasting comparison level 1: the original SVR model v.s. other models for New York example.

performance. In which, the proposed VMD-SR-SVRCBCS model is also the closest one to the actual electricity demand than other alternative models. Table 11 illustrates the forecasting performance evaluation indexes for these four comparison levels. Similar to Queensland Example, the comparison results of these four comparison levels once again reveal the advantages of the VMD method, the SR mechanism, the chaotic-out-bound-back mechanism, and the CS algorithm, respectively.

Finally, to verify the forecasting accuracy improvement is with significance, once again, the Friedman test is implemented based on one-tail style, under one significant level,  $\alpha = 0.05$ . The test results are list in Tables 12 to 15, respectively, which indicate clearly that the SVR, the SVRCS, the SVRCBCS, and the VMD-SR-SVRCBCS models receive the significance than other alternative models in each comparison level, respectively.

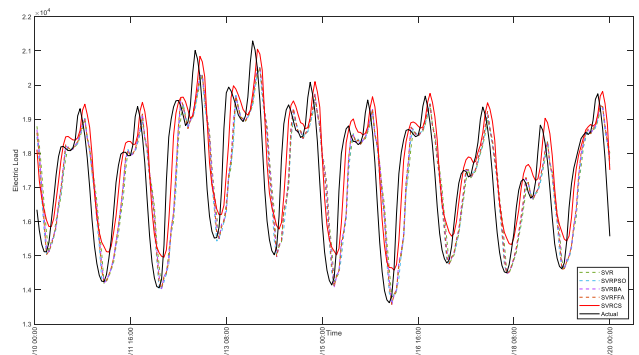


FIGURE 12. Forecasting comparison level 2: the SVRCS model v.s. other models for New York example.

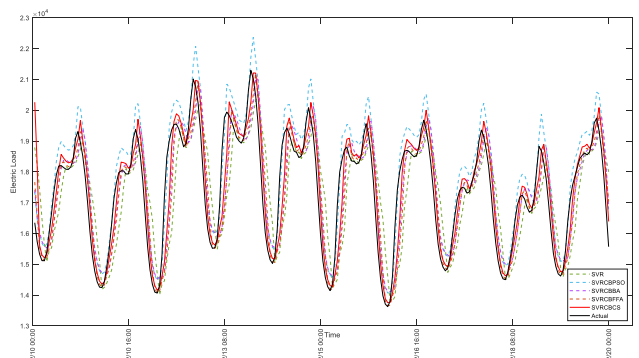


FIGURE 13. Forecasting comparison level 3: the SVRCSBCS model v.s. other models for New York example.

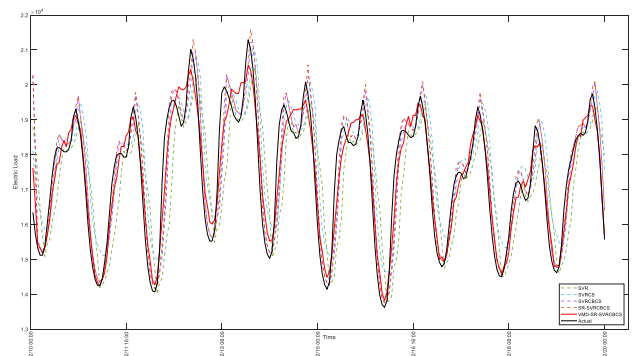


FIGURE 14. Forecasting comparison level 4: the VMD-SR-SVRCBCS model v.s. other models for New York example.

### C. DISCUSSIONS

The proposed VMD-SR-SVRCBCS model has obtained significant highest forecasting accuracy indexes than other alternative models (ARIMA, SARIMA, GRNN ( $\sigma = 1.0$ ), BPNN, SVRCS, SVRPSO, SVRBA, SVRFFA, SVRCSBCS, SVRCBPSO, SVRCBBA, SVRCBFFA, and SR-SVRCSBCS models). It is contributed from the following points. 1) The embedded superior nonlinear capabilities and structural risk minimization of the SVR model. 2) The excellent decomposition capability of the VMD to reduce the

**TABLE 11. Forecasting accuracy indexes of the VMD-SR-SVRCBCS model and other models (New York example).**

Compared Models	Forecasting accuracy indexes			
	MAE	MSE	MAPE(%)	RMSE
ARIMA <sub>(9,2,12)</sub>	1,688.1	3.96 × 10 <sup>6</sup>	9.5	1,921.3
SARIMA <sub>(9,2,12)×(8,1,5)</sub>	1,567.3	3.56 × 10 <sup>6</sup>	9.0	1,885.6
GRNN (σ=1.0)	1,304.0	2.48 × 10 <sup>6</sup>	7.8	1,574.5
BPNN	1,301.9	2.85 × 10 <sup>6</sup>	7.6	1,688.1
SVR	1,080.0	1.87 × 10 <sup>6</sup>	6.3	1,367.4
SVRPSO	975.7	1.64 × 10 <sup>6</sup>	5.6	1,280.9
SVRBA	990.0	1.65 × 10 <sup>6</sup>	5.8	1,285.2
SVRFFA	963.9	1.59 × 10 <sup>6</sup>	5.6	1,261.7
SVRCS	872.8	1.14 × 10 <sup>6</sup>	5.2	1,068.6
SVRCBPSO	673.8	5.76 × 10 <sup>5</sup>	3.9	758.7
SVRCBBA	767.9	9.07 × 10 <sup>5</sup>	4.4	952.4
SVRCBFFA	653.8	8.41 × 10 <sup>5</sup>	3.7	917.2
SVRCBCS	515.5	5.30 × 10 <sup>5</sup>	2.9	727.9
SR-SVRCBCS	458.0	4.11 × 10 <sup>5</sup>	2.6	640.9
VMD-SR-SVRCBCS	<b>427.8</b>	<b>2.76 × 10<sup>5</sup></b>	<b>2.5</b>	<b>525.6</b>

**TABLE 12. Results of Wilcoxon signed-rank test and Friedman test for comparison level 1 (New York example).**

Compared models	Wilcoxon Signed-rank Test α = 0.05; p-value	Friedman Test (α = 0.05, one-tail)
SVR vs. ARIMA <sub>(9,2,12)</sub>	0.0000**	H <sub>0</sub> : e <sub>1</sub> = e <sub>2</sub> = e <sub>3</sub> = e <sub>4</sub> = e <sub>5</sub>
SVR vs. SARIMA <sub>(9,2,12)×(8,1,5)</sub>	0.0000**	F = 193.8600
SVR vs. GRNN (σ=1.0)	0.0000**	p = 0.00001**
SVR vs. BPNN	0.0000**	(Reject H <sub>0</sub> )

\*\* represents the test result does not accept the null hypothesis under α = 0.025.

**TABLE 13. Results of Wilcoxon signed-rank test and Friedman test for comparison level 2 (New York example).**

Compared models	Wilcoxon Signed-rank Test α = 0.05; p-value	Friedman Test (α = 0.05, one-tail)
SVRCS vs. SVR	0.0000**	H <sub>0</sub> : e <sub>1</sub> = e <sub>2</sub> = e <sub>3</sub> = e <sub>4</sub> = e <sub>5</sub>
SVRCS vs. SVRPSO	0.0000**	F = 323.2733
SVRCS vs. SVRBA	0.0000**	p = 0.00001**
SVRCS vs. SVRFFA	0.0000**	(Reject H <sub>0</sub> )

\*\* represents the test result does not accept the null hypothesis under α = 0.025.

**TABLE 14. Results of Wilcoxon signed-rank test and Friedman test for comparison level 3 (New York example).**

Compared models	Wilcoxon Signed-rank Test α = 0.05; p-value	Friedman Test (α = 0.05, one-tail)
SVRCBCS vs. SVR	0.0000**	H <sub>0</sub> : e <sub>1</sub> = e <sub>2</sub> = e <sub>3</sub> = e <sub>4</sub> = e <sub>5</sub>
SVRCBCS vs. SVRBPPO	0.0000**	F = 322.9800
SVRCBCS vs. SVRCBBA	0.0000**	p = 0.00001**
SVRCBCS vs. SVRCBFFA	0.0000**	(Reject H <sub>0</sub> )

\*\* represents the test result does not accept the null hypothesis under α = 0.025.

characteristics of mode aliasing, each IMF can comprehensively represent its decomposed characteristics, then, the SR-SVRCBCS model can separately simulate the decomposed data pattern of each IMF, eventually, receive outstanding forecasting performance. 3) The superior searching capabilities of the CBCS algorithm that employs the Tent mapping function to enrich the diversity of the population, to avoid premature problem, which are the general drawbacks

**TABLE 15. Results of Wilcoxon signed-rank test and Friedman test for comparison level 4 (New York example).**

Compared models	Wilcoxon Signed-rank Test α = 0.05; p-value	Friedman Test (α = 0.05, one-tail)
VMD-SR-SVRCBCS vs. SVR	0.0000**	H <sub>0</sub> : e <sub>1</sub> = e <sub>2</sub> = e <sub>3</sub> = e <sub>4</sub> = e <sub>5</sub>
VMD-SR-SVRCBCS vs. SVRCS	0.0000**	F = 192.3067
VMD-SR-SVRCBCS vs. SVRCBCS	0.0000**	p = 0.00001**
VMD-SR-SVRCBCS vs. SR-SVRCBCS	0.0000**	(Reject H <sub>0</sub> )

\*\* represents the test result does not accept the null hypothesis under α = 0.025.

of any meta-heuristic algorithms. In addition, the CBCS algorithm applies both global search and local search in each iteration to look for better solution (parameters of the SVR model) and proposes a simple out-bound-back mechanism to provide successful actions to avoid these cuckoo birds flying continuously to the undefined domain, eventually deteriorate the searching performance. For example, in Fig. 3, the out-bound birds could be easily captured back to the previous better location, and continue qualified searching behaviors. (4) The SR mechanism has more superior capability to capture more data structures between the hidden layer and the context layer embedded in the past electric load data. For example, in Table 2, for the Queensland Example, the CBCS algorithm is excellently to shift the local solution of the SVRCS model, (C, σ, ε) = (1.265 × 10<sup>4</sup>, 0.1638, 0.3265) with local optimal forecasting errors, in terms of MAPE (2.502%), to another better solution, (C, σ, ε) = (1.467 × 10<sup>4</sup>, 0.4967, 0.1660) of the SVRCBCS model to be another more accurate forecasting results in terms of MAPE (1.338%). Thus, it proves the superiority of the Tent mapping function to avoid premature problem by hybridized into the SVR model. In addition, the SR mechanism also provides a significant contribution to continue improving the better solution of the SVRCBCS model to another better solution, (C, σ, ε) = (1.718 × 10<sup>4</sup>, 0.3342, 0.0059) of the SR-SVRCBCS model to receive the most appropriate solution in terms of MAPE (1.123%).

In the meanwhile, based on the results of the proposed four comparison levels, it is clearly to see that 1) the superiority of the original SVR model than other artificial intelligent approaches; 2) the superiority of the SVRCS model than other SVR- based models with different swarm-based algorithms; 3) the superiority of the SVRCBCS model than other SVR-CB-based models, i.e., the very contributions from the proposed chaotic-out-bound-back mechanism; and 4) the superiority of the SR-SVRCBCS model and the VMD-SR-SVRCBCS model than other SVRCS-based models, i.e., the very contributions of the proposed SR mechanism and the VMD method.

For forecasting accuracy improvement significant test, based on Tables 6 to 8 and 12 to 15, the VMD-SR-SVRCBCS model significantly outperforms other alternative models. By comparing the SVRCS model with the SVR-CBCS model (Tables 7 and 14), they recognize that the Tent mapping function and the out-bound-back mechanism could

significantly improve the premature problem. By comparing the SVRCBCS model with the SR-SVRCBCS model (Tables 8 and 15), they also indicate the SR mechanism could also significantly improve the forecasting performances, even it really costs more computing time, it is valuable to conduct the SR mechanism while modeling. By comparing the SR-SVRCBCS model with the VMD-SR-SVRCBCS model (Tables 8 and 15), it indicates the VMD method could significantly improve the forecasting performances, even it really costs some pre-processing time, it is deserved to conduct the VMD method.

Finally, compare the differences between the SR-SVRCBCS model and authors' previous proposed SSVR-CCS model [44], which also hybridizes the CS algorithm with the SVR model. For the same theoretical mechanism of these two models, they both benefit from the Tent mapping function to enrich the diversity of cuckoo's population (i.e., the CCS algorithm) while trapping into local optimum. For the differences, the SR-SVRCBCS model furtherly hybridizes the out-bound-back mechanism to improve the searching quality by turning the out-bound cuckoo birds back to the previous location, to guarantee the continuous searching on the right way. In addition, the hybridized SR mechanism also improves the searching quality by enhancing the past searching experiences. Eventually, the SR-SVRCBCS model could receive more satisfied forecasting accuracy. On the other hand, the SSVRCCS model only furtherly combines the seasonal mechanism to resolve the seasonal/cyclic trends embedded in the electric load data, then, calculate the seasonal indexes to receive higher forecasting accuracy. The improvement is related minor than the SR-SVRCBCS model.

For the theoretical modeling, it is no doubt that the SR-SVRCBCS model is superior to the SSVRCCS model. Because the SR-SVRCBCS model additionally demonstrates the theoretical improvements (out-bound-back mechanism and SR mechanism) by hybridizing with the original CS algorithm and the original SVR model.

#### IV. CONCLUSION

For a resource-saving developing country, like China, accurate electric load forecasting plays an important role to effectively implementations of national energy policy planning. In this paper, authors hybridize several novel intelligent techniques, including the SVR model, the cuckoo search algorithm, the Tent chaotic mapping function, the out-bound-back mechanism, the VMD method, and the SR mechanism, namely VMD-SR-SVRCBCS model, to receive more satisfied forecasting performances. Another six alternative models, ARIMA, SARIMA, BPNN, GRNN, SVRCS, SVRPSO, SVRBA, SVRFFA, SVRCBCS, SVRCBPSO, SVRCBBA, SVRCBFFA, and SR-SVRCBCS models are employed to compare the forecasting results. Experiment results indicate that the proposed VMD-SR-SVRCBCS model has significantly outperformed other six alternative models. The conclusions of this paper could be as the followings:

- 1) For the Ten chaotic mapping function, in Tables 2 and 9, they demonstrate clearly that the Tent function could improve the forecasting performances of the SVRCS models, i.e., to enrich the diversity of the population to avoid premature problems. The forecasting accuracy improves 1.164% (Queensland Example) and 2.223% (New York Example) in terms of MAPE, respectively.
- 2) For the searching quality, also from Tables 2 and 9, the eventually determined appropriate parameters of the SVR model are all within the defined domains. And the status is never occurred that the searching quality deteriorates to the point where it can't be cleaned up.
- 3) For the SR mechanism, it also can be revealed in Tables 2 and 9, it could learn more recurrent information from previous hidden layer of the SVRCBCS models during the recurrent mechanism processes. The forecasting accuracy improves 0.215% (Queensland Example) and 0.319% (New York Example) in terms of MAPE, respectively.
- 4) For the VMD effects, it can be revealed in Fig. 10 and 15 that each decomposed IMF and the residual demonstrate simpler tendency in both Examples, which could be simulated by the SR-SVRCBCS model more accurately. The forecasting performance demonstrates in Tables 4 and 11, respectively, that the forecasting accuracy improves 0.220% (Queensland Example) and 0.171% (New York Example) in terms of MAPE.
- 5) Eventually, the forecasting results for these two examples are shown in Fig. 6 to 9 and 11 to 14; the comparison results of the VMD-SR-SVRCBCS model with other alternative models are shown in Tables 4 and 11. The significant test results are shown in Tables 5 to 8 and 12 to 15. These findings all demonstrate the VMD-SR-SVRCBCS models all significantly receive the highest forecasting accuracy in terms of MAE, MSE, MAPE and RMSE.

Based on the forecasting results illustrate in above sections, it is significantly to indicate that the performance of the VMD-SR-SVRCBCS model is superior to other alternative models. It can also be applied in other forecasting fields, such as stock price forecasting which heavily depends on the embedded past information, i.e., the SR mechanism could be extended to this field successfully. The VMD method also could be employed to decompose the intrinsic mode of the financial data set, to reduce the characteristics of mode aliasing (mode mixing), false modes, and many IMFs with similar frequencies. The out-bound-back mechanism could also be applied to any meta-heuristic algorithm which suffers from boundary handling problem. It is easily implemented and could effectively to deal with these out-bound behaviors. In the future, authors would like to apply these novel proposed intelligent techniques to be hybridized with other alternative algorithms to continue this interesting exploration.

## REFERENCES

- [1] G.-F. Fan, A. Wang, and W.-C. Hong, "Combining grey model and self-adapting intelligent grey model with genetic algorithm and annual share changes in natural gas demand forecasting," *Energies*, vol. 11, no. 7, p. 1625, Jun. 2018.
- [2] G.-F. Fan, L.-L. Peng, and W.-C. Hong, "Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model," *Appl. Energy*, vol. 224, pp. 13–33, Aug. 2018.
- [3] G.-F. Fan, L.-L. Peng, W.-C. Hong, and F. Sun, "Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression," *Neurocomputing*, vol. 173, pp. 958–970, Jan. 2016.
- [4] W.-C. Hong, Y. Dong, W. Y. Zhang, L.-Y. Chen, and B. K. Panigrahi, "Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 44, no. 1, pp. 604–614, Jan. 2013.
- [5] F.-Y. Ju and W.-C. Hong, "Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting," *Appl. Math. Model.*, vol. 37, no. 23, pp. 9643–9651, Dec. 2013.
- [6] C. Yuan, S. Liu, and Z. Fang, "Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model," *Energy*, vol. 100, pp. 384–390, Apr. 2016.
- [7] P. Sen, M. Roy, and P. Pal, "Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization," *Energy*, vol. 116, pp. 1031–1038, Dec. 2016.
- [8] S. Barak and S. S. Sadegh, "Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 82, pp. 92–104, Nov. 2016.
- [9] D. Yang, V. Sharma, Z. Ye, L. I. Lim, L. Zhao, and A. W. Aryaputera, "Forecasting of global horizontal irradiance by exponential smoothing, using decompositions," *Energy*, vol. 81, pp. 111–119, Mar. 2015.
- [10] Z. Dong, D. Yang, T. Reindl, and W. M. Walsh, "Short-term solar irradiance forecasting using exponential smoothing state space model," *Energy*, vol. 55, pp. 1104–1113, Jun. 2013.
- [11] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electric Power Syst. Res.*, vol. 130, pp. 139–147, Jan. 2016.
- [12] M. E. Lebotsa, C. Sigauke, A. Bere, R. Fildes, and J. E. Boylan, "Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem," *Appl. Energy*, vol. 222, pp. 104–118, Jul. 2018.
- [13] D. Vu, K. Muttaqi, and A. Agalgaonkar, "A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables," *Appl. Energy*, vol. 140, pp. 385–394, Feb. 2015.
- [14] H. Takeda, Y. Tamura, and S. Sato, "Using the ensemble Kalman filter for electricity load forecasting and analysis," *Energy*, vol. 104, pp. 184–198, Jun. 2016.
- [15] C.-N. Ko and C.-M. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," *Energy*, vol. 49, pp. 413–422, Jan. 2013.
- [16] H. S. Hippert and J. W. Taylor, "An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting," *Neural Netw.*, vol. 23, no. 3, pp. 386–395, Apr. 2010.
- [17] Y. Ohtsuka, T. Oga, and K. Kakamu, "Forecasting electricity demand in Japan: A Bayesian spatial autoregressive ARMA approach," *Comput. Statist. Data Anal.*, vol. 54, no. 11, pp. 2721–2735, Nov. 2010.
- [18] S. Kelo and S. Dudul, "A wavelet Elman neural network for short-term electrical load prediction under the influence of temperature," *Int. J. Electr. Power Energy Syst.*, vol. 43, no. 1, pp. 1063–1071, Dec. 2012.
- [19] M. Karimi, H. Karami, M. Gholami, H. Khatibzadehazad, and N. Moslemi, "Priority index considering temperature and date proximity for selection of similar days in knowledge-based short term load forecasting method," *Energy*, vol. 144, pp. 928–940, Feb. 2018.
- [20] P. Lulis, K. R. Khalilpour, L. Andrew, and A. Liebman, "Short-term residential load forecasting: Impact of calendar effects and forecast granularity," *Appl. Energy*, vol. 205, pp. 654–669, Nov. 2017.
- [21] P. Singh and P. Dwivedi, "Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem," *Appl. Energy*, vol. 217, pp. 537–549, May 2018.
- [22] A. Khwaja, X. Zhang, A. Anpalagan, and B. Venkatesh, "Boosted neural networks for improved short-term electric load forecasting," *Electr. Power Syst. Res.*, vol. 143, pp. 431–437, Feb. 2017.
- [23] A. Khwaja, M. Naeem, A. Anpalagan, A. Venetsanopoulos, and B. Venkatesh, "Improved short-term load forecasting using bagged neural networks," *Electr. Power Syst. Res.*, vol. 125, pp. 109–115, Aug. 2015.
- [24] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, A. Sánchez-Esguevillas, and J. Lloret, "Artificial neural networks for short-term load forecasting in microgrids environment," *Energy*, vol. 75, pp. 252–264, Oct. 2014.
- [25] D. Chaturvedi, A. Sinha, and O. Malik, "Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network," *Int. J. Electr. Power Energy Syst.*, vol. 67, pp. 230–237, May 2015.
- [26] V. N. Coelho, I. M. Coelho, B. N. Coelho, A. J. Reis, R. Enayatifar, M. J. Souza, and F. G. Guimarães, "A self-adaptive evolutionary fuzzy model for load forecasting problems on smart grid environment," *Appl. Energy*, vol. 169, pp. 567–584, May 2016.
- [27] C. W. Lou and M. C. Dong, "A novel random fuzzy neural networks for tackling uncertainties of electric load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 73, pp. 34–44, Dec. 2015.
- [28] M. Amina, V. Kodogiannis, I. Petrounias, and D. Tomtsis, "A hybrid intelligent approach for the prediction of electricity consumption," *Int. J. Electr. Power Energy Syst.*, vol. 43, no. 1, pp. 99–108, Dec. 2012.
- [29] S. Aras and İ. D. Kocakoç, "A new model selection strategy in time series forecasting with artificial neural networks: IHFS," *Neurocomputing*, vol. 174, pp. 974–987, Jan. 2016.
- [30] H. K. Ghritlahre and R. K. Prasad, "Application of ANN technique to predict the performance of solar collector systems—A review," *Renew. Sustain. Energy Rev.*, vol. 84, pp. 75–88, Mar. 2018.
- [31] Ö. F. Ertugrul, "Forecasting electricity load by a novel recurrent extreme learning machines approach," *Int. J. Electr. Power Energy Syst.*, vol. 78, pp. 429–435, Jun. 2016.
- [32] W.-C. Hong, *Hybrid Intelligent Technologies in Energy Demand Forecasting*. Cham, Switzerland: Switzerland: Springer, 2020.
- [33] J. Geng, M.-L. Huang, M.-W. Li, and W.-C. Hong, "Hybridization of seasonal chaotic cloud simulated annealing algorithm in a SVR-based load forecasting model," *Neurocomputing*, vol. 151, pp. 1362–1373, Mar. 2015.
- [34] S. Lahmiri, "Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression," *Appl. Math. Comput.*, vol. 320, pp. 444–451, Mar. 2018.
- [35] G. Sermpinis, C. Stasinakis, K. Theofilatos, and A. Karathanasopoulos, "Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms—Support vector regression forecast combinations," *Eur. J. Oper. Res.*, vol. 247, no. 3, pp. 831–846, Dec. 2015.
- [36] H. Jiang and Z. Wang, "GMRVm-SVR model for financial time series forecasting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7813–7818, Dec. 2010.
- [37] J. Wang, R. Hou, C. Wang, and L. Shen, "Improved v-Support vector regression model based on variable selection and brain storm optimization for stock price forecasting," *Appl. Soft Comput.*, vol. 49, pp. 164–178, Dec. 2016.
- [38] W.-C. Hong, Y. Dong, L.-Y. Chen, and S.-Y. Wei, "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1881–1890, Mar. 2011.
- [39] R. Chen, C.-Y. Liang, W.-C. Hong, and D.-X. Gu, "Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm," *Appl. Soft Comput.*, vol. 26, pp. 435–443, Jan. 2015.
- [40] X. Yu, X. Zhang, and H. Qin, "A data-driven model based on Fourier transform and support vector regression for monthly reservoir inflow forecasting," *J. Hydro-Environ. Res.*, vol. 18, pp. 12–24, Feb. 2018.
- [41] P.-S. Yu, T.-C. Yang, S.-Y. Chen, C.-M. Kuo, and H.-W. Tseng, "Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting," *J. Hydrol.*, vol. 552, pp. 92–104, Sep. 2017.
- [42] S. M. Hosseini and N. Mahjouri, "Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling," *Appl. Soft Comput.*, vol. 38, pp. 329–345, Jan. 2016.
- [43] W.-C. Hong, "Application of chaotic ant swarm optimization in electric load forecasting," *Energy Policy*, vol. 38, no. 10, pp. 5830–5839, Oct. 2010.
- [44] Y. Dong, Z. Zhang, and W.-C. Hong, "A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting," *Energies*, vol. 11, no. 4, p. 1009, Apr. 2018.
- [45] M.-W. Li, J. Geng, S. Wang, and W.-C. Hong, "Hybrid chaotic quantum bat algorithm with SVR in electric load forecasting," *Energies*, vol. 10, no. 12, p. 2180, Dec. 2017.



- [46] Z. Zhang and W.-C. Hong, "Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm," *Nonlinear Dyn.*, vol. 98, no. 2, pp. 1107–1136, Oct. 2019.
- [47] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, "Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems," *Eng. Comput.*, vol. 29, no. 1, pp. 17–35, Jan. 2013.
- [48] S. Ishak Boushaki, N. Kamel, and O. Bendjeghaba, "A new quantum chaotic cuckoo search algorithm for data clustering," *Expert Syst. Appl.*, vol. 96, pp. 358–372, Apr. 2018.
- [49] L. Huang, S. Ding, S. Yu, J. Wang, and K. Lu, "Chaos-enhanced Cuckoo search optimization algorithms for global optimization," *Appl. Math. Model.*, vol. 40, nos. 5–6, pp. 3860–3875, Mar. 2016.
- [50] A. M. Rather, A. Agarwal, and V. Sastry, "Recurrent neural network and a hybrid model for prediction of stock returns," *Expert Syst. with Appl.*, vol. 42, no. 6, pp. 3234–3241, Apr. 2015.
- [51] L. Ruiz, R. Rueda, M. Cuéllar, and M. Pegalajar, "Energy consumption forecasting based on Elman neural networks with evolutive optimization," *Expert Syst. Appl.*, vol. 92, pp. 380–389, Feb. 2018.
- [52] H. Zhang, Z. Wang, and D. Liu, "A Comprehensive Review of Stability Analysis of Continuous-Time Recurrent Neural Networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 7, pp. 1229–1262, Jul. 2014.
- [53] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," 2015, *arXiv:1506.00019*. [Online]. Available: <https://arxiv.org/abs/1506.00019>
- [54] X. Chen, X. Chen, J. She, and M. Wu, "A hybrid time series prediction model based on recurrent neural network and double joint linear–nonlinear extreme learning network for prediction of carbon efficiency in iron ore sintering process," *Neurocomputing*, vol. 249, pp. 128–139, Aug. 2017.
- [55] G.-F. Fan, S. Qing, H. Wang, W.-C. Hong, and H.-J. Li, "Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting," *Energies*, vol. 6, no. 4, pp. 1887–1901, Apr. 2013.
- [56] G.-F. Fan, L.-L. Peng, X. Zhao, and W.-C. Hong, "Applications of hybrid EMD with PSO and GA for an SVR-based load forecasting model," *Energies*, vol. 10, no. 11, p. 1713, Oct. 2017.
- [57] W.-C. Hong and G.-F. Fan, "Hybrid empirical mode decomposition with support vector regression model for short term load forecasting," *Energies*, vol. 12, no. 6, pp. 1093–1108, Mar. 2019.
- [58] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [59] Y. Zhang, K. Liu, L. Qin, and X. An, "Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods," *Energy Convers. Manage.*, vol. 112, pp. 208–219, Mar. 2016.
- [60] S. Lahmiri, "A variational mode decomposition approach for analysis and forecasting of economic and financial time series," *Expert Syst. Appl.*, vol. 55, pp. 268–273, Aug. 2016.
- [61] D. Wang, H. Luo, O. Grunder, and Y. Lin, "Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction," *Renew. Energy*, vol. 113, pp. 1345–1358, Dec. 2017.
- [62] X.-S. Yang and S. Deb, "Multiobjective cuckoo search for design optimization," *Comput. Oper. Res.*, vol. 40, no. 6, pp. 1616–1624, Jun. 2013.
- [63] X.-S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 169–174, Jan. 2014.
- [64] R. J. Schalkoff, *Artificial Neural Networks*. New York, NY, USA: McGraw-Hill, 2011.
- [65] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 3–18, Mar. 2011.



**ZICHEN ZHANG** received the M.S. degree in management science and engineering from Jiangsu Normal University, Xuzhou, China, in 2018. He is currently pursuing the Ph.D. degree in computer science and technology with the School of Computer Science and Technology, China University of Mining and Technology, China. His research interests include machine learning, meta-heuristic algorithms, and AI-based forecasting models.



**WEI-CHIANG HONG** (Senior Member, IEEE) received the Ph.D. degree in management from Dayeh University, Changhua, Taiwan, in 2008. He is currently a Professor with the School of Computer Science and Technology, Jiangsu Normal University, Xuzhou, China. His research interests mainly include computational intelligence (neural networks and evolutionary computation), and application of forecasting technology (ARIMA, support vector regression, and chaos theory) and machine learning algorithms. His articles have been published in *Energy Policy*, *Applied Energy*, *Energy*, *Energy Conversion and Management*, *Applied Mathematical Modelling*, the *International Journal of Electrical Power & Energy Systems*, *Electric Power Systems Research*, and the *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, among others. His research interest mainly includes computational intelligence and application of forecasting technology.



**JUNCHI LI** received the Bachelor of Medicine degree from Xuzhou Medical University, Xuzhou, China, in 2016. He is currently pursuing the Master of Medicine degree with the Department of Medical Ultrasonics, Xuzhou No. 1 Peoples Hospital. His research interest includes deep learning and its application in medical imaging.