

Research Article

Swarm Intelligence-Based Hybrid Models for Short-Term Power Load Prediction

Jianzhou Wang,¹ Shiqiang Jin,² Shanshan Qin,³ and Haiyan Jiang²

¹ School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

² School of Mathematics & Statistics, Lanzhou University, Lanzhou 73000, China

³ MOE Key Laboratory of Western China's Environmental Systems, Research School of Arid Environment & Climate Change, Lanzhou University, Lanzhou 73000, China

Correspondence should be addressed to Shiqiang Jin; jinsql@gmail.com

Received 6 June 2014; Revised 18 July 2014; Accepted 1 August 2014; Published 30 September 2014

Academic Editor: Fang Zong

Copyright © 2014 Jianzhou Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Swarm intelligence (SI) is widely and successfully applied in the engineering field to solve practical optimization problems because various hybrid models, which are based on the SI algorithm and statistical models, are developed to further improve the predictive abilities. In this paper, hybrid intelligent forecasting models based on the cuckoo search (CS) as well as the singular spectrum analysis (SSA), time series, and machine learning methods are proposed to conduct short-term power load prediction. The forecasting performance of the proposed models is augmented by a rolling multistep strategy over the prediction horizon. The test results are representative of the out-performance of the SSA and CS in tuning the seasonal autoregressive integrated moving average (SARIMA) and support vector regression (SVR) in improving load forecasting, which indicates that both the SSA-based data denoising and SI-based intelligent optimization strategy can effectively improve the model's predictive performance. Additionally, the proposed CS-SSA-SARIMA and CS-SSA-SVR models provide very impressive forecasting results, demonstrating their strong robustness and universal forecasting capacities in terms of short-term power load prediction 24 hours in advance.

1. Introduction

With the additional types of energy integration into the power grid and the development of generation technologies, power utilities are going through a crucial challenge stemming from maintaining the desired security and reliability of the electricity supply [1]. Various technologies in power generation, transmission, distribution, and utility are researched by organizations, and the power grid is facing a major change. The smart grid, an internationally popular topic recently, is pushing the power grid into an open system with the characteristics of robustness and dynamics, which provides a great chance to improve and enhance the load demand response programs [2]. However, the high accuracy of load prediction is the key factor in a power intelligence system, which determines the quality of the smart grid. If the load demands are overestimated, it will induce a conservative operation with excessive energy purchased, thereby resulting in energy waste and unnecessary cost. It has been estimated

that a 1% increase in the forecasting error will result in a 10 million dollar increase in operation costs [3]. If the load demands are underestimated, it will induce a risky operation with a deficient preparation of the spinning reserve, causing the power system to operate in a vulnerable region to the disturbance and power cut [4].

Therefore, according to the aforementioned analysis, the power load prediction should be urgently conducted with high accuracy to guarantee the operational performance of the power system. Over the past few decades, large efforts have been devoted to improving the power load forecasting accuracy. The various methods utilized for load prediction range from the traditional statistical models to the complicated artificial intelligence-based models [5]. For intelligence-based methods, the evolutionary programming, expert systems, artificial neural networks (ANN), and fuzzy inferences are included [6]. Among the intelligent approaches, the ANN method has received more attention because of its good flexibility, easy implementation, and nonlinear

mapping ability between the powers loads and weather variables such as humidity, temperature, wind speed, and historical load patterns [7]. Moreover, to perform a better training process and improve the load forecasting accuracy, the ANN is usually exploited in a hybrid model by combining it with other methods or techniques, such as the support vector regression (SVR). The SVR [8] model is used successfully to address forecasting problems in many fields, such as the financial time series (exchange rate and stocks index) prediction [9, 10], hydroinformatic forecasting [11, 12], and tourist arrival forecasting [13]. Additionally, the SVR model is also successfully applied in forecasting the power load [14–17]. The empirical results demonstrate that the selection of the two parameters C (to trade off the training errors and large weights) and σ (the parameter for the Gaussian kernel function) in an SVR model influences the forecasting accuracy considerably. To conquer the difficult problems in the selection of the parameters for the SVR, the author conducted a series of relevant experiments by employing a hybrid sequence with swarm intelligent optimization algorithms for the parameter determination to overcome the problem of immature convergence (trapped in local optimum) [11, 15, 17]. To continue testing the superiority of the hybrid sequence with swarm intelligent optimization algorithms, this paper tried to employ the cuckoo search (CS) and particle swarm optimization (PSO) algorithm to determine the values of three parameters in an SVR model.

The cuckoo search (CS) algorithm is a new optimization metaheuristic algorithm [21] based on a stochastic global search and the obligate brood-parasitic behavior of cuckoos in combination with the Lévy flight behavior of several birds and fruit flies. It is widely and successfully used in a number of practical problems, such as knapsack problems [22], software test effort estimations [23], scheduling problems [24], test sequence optimization problems [25], convex and nonconvex ED (economic dispatch) problems, and microgrid power dispatch problems [26]. In [26], the author implemented the CS algorithm to solve both the convex and nonconvex ED problems and the micro grid power dispatch problem. Moreover, the author compared the CS algorithm with many other artificial optimization algorithms, such as simulated annealing (SA), evolutionary programming (EP), genetic algorithm (GA), PSO, differential evolution (DE), and bacterial foraging algorithm (BFA). It is seen that the proposed CS algorithm has the ability to converge to a better quality solution than all of the artificial optimization algorithms mentioned above. One of the main advantages of the CS algorithm is that there are fewer parameters to be fine-tuned in the CS algorithm than in the GA and PSO [21]. In [21], Yang and Deb formulated the CS algorithm to search the minimum values of the multimodal objective functions: the bivariate Michalewicz function, De Jong's first function, Shubert's bivariate function, Griewangk's test function, Ackley's function, generalized Rosenbrock's function, Schwefel's test function, Rastrigin's test function, and Michalewicz's test function. Then, by comparing the results searched by the CS algorithm with the PSO and GA, Yang and Deb concluded that the CS algorithm is superior to these existing algorithms for multimodal objective functions. The preliminary studies indicate that it is very

promising and could outperform the existing algorithms, such as the artificial bee colony (ABC), GA, PSO, bacterial foraging (BF), ant colony optimization (ACO), and honey-bee mating optimization (HBMO) [27–38]. However, few papers employ the CS algorithm to optimize the parameters of the SARIMA and compare the convergence speed with the PSO, which has a fast convergence speed. In this paper, we conduct the CS algorithm to optimize the parameters of the SVR and SARIMA and meanwhile compare the accuracy of the forecast and convergence speed of the CS algorithm with that of the PSO in predicting the power load.

The most popular and classic statistical models include the linear or nonlinear regression models, time series models, state estimation, and Kalman filtering technology [7]. The multiregression models consider certain factors as the explanatory variables, such as the weather factors, climatic conditions, social activities, and seasonal factors, with the regression coefficients estimated by the least squares estimation or modern regression method. The time series models assume that the load is related to its past values, which can be regarded as an autoregressive process that is therefore forecasted by time series models. Based on the aforementioned literature, however, most studies have been regardless of the seasonal patterns, the fluctuation of which can lead to a deviation in the load forecasts. To solve this, Wang et al. [39] proposed a hybrid model by incorporating the SARIMA and ANN to address the periodic relationships and nonlinear patterns, respectively. SARIMA [40, 41] mainly addresses the linear relationships and considers the periodicity of the time series in a real-life scenario. Considering, insufficient work that is related to the periodicity of load data has been performed. Therefore, a new hybrid model based on SARIMA is proposed in this paper.

The SSA [42, 43] is a powerful and novel technique of time-series analysis and forecasting incorporating the elements of classical time-series analysis, multivariate geometry, multivariate statistics, dynamical systems, and signal processing [44]. The main purpose of the SSA is to develop a decomposition of the raw time series into the sum of several interpretable and independent components, such as a gradually varying trend, oscillatory components (periodic or quasi-periodic components) and a structure-less noise. Then, several of these components are used for time series forecasting. The SSA technique has been successfully used in several fields such as geophysics, hydrology, climatology [45, 46], and economics [47, 48]. Golyandina and Korobeynikov [49] described very detailed steps to show how the methodology of the SSA analysis, forecasting, and parameter estimation can be implemented with the help of the package *Rssa*. Chen et al. [50] assessed the value of the SSA for extracting the time-variable seasonal signals from the GPS time series and compared the SSA-based results to the least-squares analysis and Kalman filtering. The results demonstrate that the SSA is a viable and complementary tool for extracting modulated oscillations from the GPS time series. In [51], Wu and Chau attempts to eliminate the lag effect, often appearing in the ANN modeling process, by using the singular spectrum analysis (SSA). Two watersheds from China are explored with daily collected data. The results demonstrate that the SSA

can significantly enhance the performance of the prediction model and eliminate the lag effect.

If the original load data are directly applied to a train model without eliminating noise, the high-frequency components may disturb the forecasted load patterns [6]. To improve the load forecasting accuracy, a wavelet denoising technique is usually adopted to extract the low frequency component of the load pattern. Therefore, these data are utilized in load forecasting methods. However, many parameters should be determined during the procedure of wavelet analysis, such as the determination of the decomposition layers, wavelet basis function, and threshold function; furthermore, each of the parameters is a large amount and is hard to determine objectively. Because of the subjective selection of the parameters for the wavelet, different researchers obtain various denoising effects; thus, researchers often need to conduct a large amount of data experiments to gain a satisfactory result before continuing the study. However, one of the major advantages of the SSA compared to other approaches is that only two parameters are required to model the time series under analysis [44]. It is a relatively new data-driven or nonparametric technique developed to model a nonlinear and/or nonstationary as well as noisy short time series [52]. Moreover, it is considered from Claudio [44] that the SSA does not depend on a priori defined functions, such as the Fourier approach (based on sine and cosine functions), but it generates a set of components directly from the time series under study. Additionally, the SSA technique can compute periodic or quasiperiodic components and a slowly changing trend. To the best of our knowledge, the SSA is not applied to denoise the power load time series. Most of the references predict the components that are decomposed by the SSA and then reconstruct them. In this paper, we denoise the power load by the SSA and then conduct the linear and nonlinear methods to validate whether this denoising technique can help the linear and nonlinear models further improve the accuracy of forecasting.

This paper starts with a brief description of the related methodology in Section 2, specifies the procedure of the SSA in Section 3.1, and presents the results of the case study for simulating the power load in Section 3.2.

2. Related Methodology

In this section, SSA, SARIMA, the SVR model, the SI algorithm (CS and PSO), and the design of the proposed hybrid SI-based predictive models are summarized as the foundation to construct the proposed hybrid model.

2.1. Singular Spectral Analysis (SSA) Technique. The SSA technique, a well-developed method of time series analysis, can extract major information components from a time series, such as the trend and periodicities components without prior knowledge regarding the trend as well as period values [50]. In this section, the information regarding the SSA, which is critical for understanding the implementations of the SSA, is described, and the SSA tool is used for the analysis of the power load series.

The basic SSA consists of two complementary parts: decomposition and reconstruction. For the decomposition part, it comprises two steps: embedding and singular values decomposition. For the reconstruction part, two steps are also involved, which are the Eigentriple grouping and diagonal averaging [51]. Assume a time series $X = \{x_1, x_2, \dots, x_N\}$. The processes of the SSA are given as follows [49].

Part 1 (decomposition). Consider the following.

Step 1 (embedding). The original time series X is mapped into a sequence of multidimensional lagged vectors of size L by performing the embedding procedure

$$X_i = (x_1, x_2, \dots, x_{i+L-1})^T, \quad i = 1, 2, \dots, K, \quad (1)$$

where L ($1 < L < N$) is a positive integer called the window length and the delay time is $K = N - L + 1$. The trajectory matrix of the series X is

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{bmatrix}. \quad (2)$$

For the trajectory matrix \mathbf{X} , there are two vital properties, which are the following.

- (i) Both the columns and rows of \mathbf{X} are a subseries of the original series X .
- (ii) The trajectory matrix \mathbf{X} is Hankel with equal elements on the antidiagonals.

Step 2 (singular values decomposition (SVD)). Decompose the trajectory matrix \mathbf{X} by using the following:

$$\mathbf{X} = \sum_{i=1}^L P_i Q_i^T = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_L, \quad (3)$$

where $\{P_i\}_{i=1}^L$ is an orthonormal basis in R^L and $Q_i = \mathbf{X}^T P_i$. For the orthonormal basis $\{P_i\}_{i=1}^L$, here we have two versions for consideration:

- (A) basic: $\{P_i\}_{i=1}^L$ are the eigenvectors of $\mathbf{X}\mathbf{X}^T$,
- (B) Toeplitz: $\{P_i\}_{i=1}^L$ are the eigenvectors of the matrix \mathbf{C} , which is given by

$$c_{ij} = \frac{1}{N - |i - j|} \sum_{m=1}^{N-|i-j|} x_m x_{m+|i-j|}, \quad 1 \leq i, j \leq L. \quad (4)$$

In both cases, the eigenvectors are ordered, which can thereby guarantee that the corresponding eigenvalues are placed in decreasing order.

Note that the Case (B) version is only suitable for the analysis of the stationary time series with mean value zero.

Note also that the Case (A) version corresponds to the SVD of \mathbf{X} ; that is, $\mathbf{X} = \sum_i \sqrt{\lambda_i} U_i V_i^T$, $P_i = U_i$ are the left

singular vectors of \mathbf{X} . Consider $Q_i = \sqrt{\lambda_i} V_i$, where λ_i are the eigenvalues of $\mathbf{X}\mathbf{X}^T$. $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$ and V_i are the orthonormal systems of the eigenvectors corresponding to these eigenvalues as well as the right singular vectors of \mathbf{X} . The collection (λ_i, U_i, V_i) is called the i th eigentriple of the SVD.

Part 2 (reconstruction). Consider the following.

Step 3 (eigentriple grouping). Let $d = \max\{j : \lambda_j \neq 0\}$. After decomposition, the grouping procedure splits the set of indices $\{1, \dots, d\}$ into m disjoint subsets I_1, \dots, I_m . Let $I = \{I_1, I_2, \dots, I_m\}$, and define the resultant matrix \mathbf{X}_I corresponding to the group I by the following equation:

$$\mathbf{X}_I = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_m}. \quad (5)$$

The above procedure of selecting the sets I_1, \dots, I_m is called the eigentriple grouping. If $m = d$ and $I_j = \{j\}$, $j = 1, \dots, d$, then the corresponding grouping is called elementary. The selection of several leading eigentriples for the Case (A) version determines the approximation of the original time series and corresponds to the well-known optimality property of the SVD.

Step 4 (diagonal averaging). In this step, each matrix \mathbf{X}_{I_j} is transformed into a new series of length N . Let \mathbf{Y} be the $L \times K$ matrix with elements y_{ij} , $1 \leq i \leq L$, $1 \leq j \leq K$, and $L \leq K$. By performing diagonal averaging, we transfer the matrix \mathbf{Y} into the series $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)$ using the following:

$$\hat{y}_s = \sum_{(l,k) \in A_s} \frac{y_{lk}}{|A_s|}, \quad (6)$$

where $A_s = \{(l, k) : l + k = s + 1, 1 \leq l \leq L, 1 \leq k \leq K\}$ and $|A_s|$ denote the number of elements in the set A_s . The diagonal averaging utilized for a resultant matrix \mathbf{X}_{I_k} generates a reconstructed series $\hat{X}^{(k)} = (\hat{X}_1^{(k)}, \hat{X}_2^{(k)}, \dots, \hat{X}_N^{(k)})$. Then, the initial series $X = \{x_1, x_2, \dots, x_N\}$ is decomposed into a sum of m reconstructed series:

$$x_n = \sum_{k=1}^m \hat{X}_n^{(k)}, \quad n = 1, 2, \dots, N. \quad (7)$$

The reconstructed series produced by the elementary grouping is called the elementary reconstructed series.

The SSA is a data-driven technique that can extract information from a short and noisy time series without prior knowledge of the dynamics affecting the time series. A significant characteristic of the SSA is that trend patterns obtained in this way are not necessarily linear [50]. Most importantly, the intricate periodicities lying in the time series can be modulated and extracted out.

2.2. Seasonal Autoregressive Moving Average (SARIMA). The time series that mainly contains the periodic and stochastic components can be forecasted by the SARIMA model, which is the most popular linear model for a seasonal time series and has achieved great success in both academic research and industrial applications over the past few decades [53].

A time series $\{X_t \mid t = 1, 2, \dots, N\}$ is generated by a SARIMA $(p, d, q)(P, D, Q)_s$ process if

$$\phi_p(B) \Phi_P(B^s) (1 - B)^d (1 - B^s)^D X_t = \theta_q(B) \Theta_Q(B^s) \varepsilon_t, \quad (8)$$

where p, d, q, P, D, Q are integers and s is the periodicity length. Consider the following:

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \\ \Phi_P(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}, \\ \theta_q(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \\ \Theta_Q(B^s) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}, \end{aligned} \quad (9)$$

are polynomials in B of degree p, q, P , and Q . Herein, B, d , and D denote the backward shift operator, the regular differential order, and the seasonal differential order, respectively. ε_t and X_t are the estimated residual and the observed value at time t ; $t = 1, 2, \dots, k$, separately. The residual series should be identical and independent as a white noise with an average value equal to zero and a constant variance value.

When fitting a SARIMA model, the following four steps are involved [54].

Step 1 (model identification). Identify the SARIMA $(p, d, q)(P, D, Q)_s$, which determines the most relevant combination of the autoregression and moving average process.

Step 2 (model estimation). These parameters are estimated and determined by the maximum likelihood estimation (MLE).

Step 3 (model validation). The precision of the chosen model is tested, and possible enhancements are also established during this step.

Step 4 (model forecasting). The future values that are reforecasted based on the constructed SARIMA $(p, d, q)(P, D, Q)_s$.

2.3. Basic Description of the Support Vector Regression (SVR). The SVR is an adaptation of a recently developed machine learning theory (MLT) known as the support vector machine (SVM) proposed by Vapnik et al. [55]. In the SVR model, a regression function $y = f(x)$ is fit, and then it is applied to predict the outputs based on a new input set. A brief review of the SVR is introduced as follows [56].

Step 1. A nonlinear mapping $\varphi(\cdot) : \mathcal{R}^n \rightarrow \mathcal{R}^{n_h}$ is defined to solve a nonlinear regression problem by mapping the training sets $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ into a high dimensional feature space \mathcal{R}^{n_h} .

Step 2. In the high dimensional feature space, the nonlinear regression problem in the lower dimension space is transformed into a linear one by a linear function, namely, the SVR function

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \varphi(\mathbf{x}) + b, \quad (10)$$

where $f(\mathbf{x})$ denotes the forecasting values; the coefficients \mathbf{w} ($\mathbf{w} \in \mathcal{R}^{n_h}$) and b ($b \in \mathcal{R}$) are adjustable.

Step 3. Define the empirical risk $R_{\text{emp}}(f)$:

$$R_{\text{emp}}(f) = \frac{1}{N} \sum_{i=1}^N \Theta_{\varepsilon}(y_i, \mathbf{w}^T \varphi(\mathbf{x}_i) + b), \quad (11)$$

where $\Theta_{\varepsilon}(y, f(\mathbf{x}))$ is the ε -intensive loss function and is given by

$$\Theta_{\varepsilon}(y, f(\mathbf{x})) = \begin{cases} |f(\mathbf{x}) - y| - \varepsilon, & \text{if } |f(\mathbf{x}) - y| \geq \varepsilon \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

The ε -intensive loss function is utilized to control the sparsity of the solutions and generalization of the models.

Step 4. Determine the training overall errors between the training data and ε -insensitive loss function, which can be defined as a quadratic optimization problem with inequality constraints. Consider the following:

$$\text{Min}_{\mathbf{w}, b, \xi^*, \xi} R_{\varepsilon}(\mathbf{w}, \xi^*, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N (\xi_i + \xi_i^*). \quad (13)$$

Subject to

$$\begin{aligned} y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b &\leq \varepsilon + \xi_i^*, \quad i = 1, 2, \dots, N, \\ \mathbf{w}^T \varphi(\mathbf{x}_i) - y_i + b &\leq \varepsilon + \xi_i, \quad i = 1, 2, \dots, N, \\ \xi_i^*, \xi_i &\geq 0, \quad i = 1, 2, \dots, N. \end{aligned} \quad (14)$$

The first term in (13) is employed to regularize weight sizes, to penalize large weights, and to maintain regression function flatness. The second term in (13) penalizes the training errors of $f(\mathbf{x})$ and y by exploiting the ε -intensive loss function. Herein, C is a parameter to balance these two terms. Training errors below $-\varepsilon$ are denoted as ξ_i . Otherwise, they are denoted as ξ_i^* .

Step 5. Obtain the parameter vector \mathbf{w} by solving the quadratic optimization problem defined in Step 4:

$$\mathbf{w} = \sum_{i=1}^N (\beta_i^* - \beta_i) \varphi(\mathbf{x}_i), \quad (15)$$

where β_i^*, β_i are the Lagrangian multipliers.

Step 6. Establish the SVR regression function by using the following equations:

$$f(\mathbf{x}) = \sum_{i=1}^N (\beta_i^* - \beta_i) K(\mathbf{x}_i, \mathbf{x}_j) + b, \quad (16)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \quad \gamma > 0,$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function and $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$. Several types of kernel functions can be selected to build the model. However, the most commonly used kernel functions are the Gaussian radial basis functions (RBF) and the polynomial kernel functions. Until now, it has been difficult to determine which type of kernel functions for specific

data patterns is suitable [57]. In this paper, the RBF is selected as the kernel function because of its easy implementation and its strong capability of nonlinearly mapping the training sets into an infinite dimensional space, which is suitable to handle nonlinear relationship problems. Therefore, the Gaussian RBF kernel function is specified in this study.

2.4. *Swarm Intelligence Optimization Algorithms.* In recent years, the metaheuristic optimization algorithms and evolutionary computation have been a noticeable part for solving real mathematics and engineering problems [19], especially in the field of parameter determination. In this section, several of the optimization algorithms utilized in this study are described briefly.

2.4.1. *Cuckoo Search (CS).* The CS algorithm, inspired by the breeding behavior of cuckoos, is a recently developed metaheuristic algorithm by Yang and Deb [21]. For the CS algorithm, two behaviors are adapted and combined from nature that fulfills the criteria of a metaheuristic algorithm, which are described as follows [22–24].

Breeding Behavior. Many species of cuckoos lay their eggs in communal nests, but to increase the hatching probability of their own eggs, they always remove other cuckoos' eggs. Once a host cuckoo discovers an alien egg (does not belong to itself), then it will either throw the egg away or discard the current nest and build another nest elsewhere. For the CS algorithm in each step with the new solutions generated, the poorer solutions are abandoned.

Lévy Flight. Generally, the flight path of many birds is effectively a random walk, which is representative of Lévy flights, with a step length drawn from the Lévy distribution. In the CS-based algorithm for producing a new solution $\mathbf{x}^{(t+1)}$ for a cuckoo, a Lévy flight is defined as the following expression:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \alpha \oplus \text{Lévy}(\lambda), \quad (17)$$

$$\alpha = \alpha_0 \left(\mathbf{x}_j^{(t)} - \mathbf{x}_i^{(t)} \right), \quad (18)$$

$$\text{Lévy}(\lambda) \sim l^{-\lambda}, \quad (1 < \lambda < 3), \quad (19)$$

where \mathbf{x}_i^t is the eggs (samples), i is the sample size, t is the iterations, and α is the step size that is mostly utilized as in (18). The symbol \oplus denotes an entry-wise multiplication, while the $\text{Lévy}(\lambda)$ values are found in the Lévy distribution defined in (19).

Figure 1 illustrates the basic steps of the CS algorithm and Lévy flight together with a detailed immigration of a cuckoo toward a goal habitat. More information regarding the CS method can be found in literature [21].

2.4.2. *Particle Swarm Optimization (PSO).* The PSO algorithm, inspired by the social behaviors of animal movements, investigates the search space by applying a flock of potential solutions named particle swarms, characterized by their corresponding position and velocity [58]. The individual

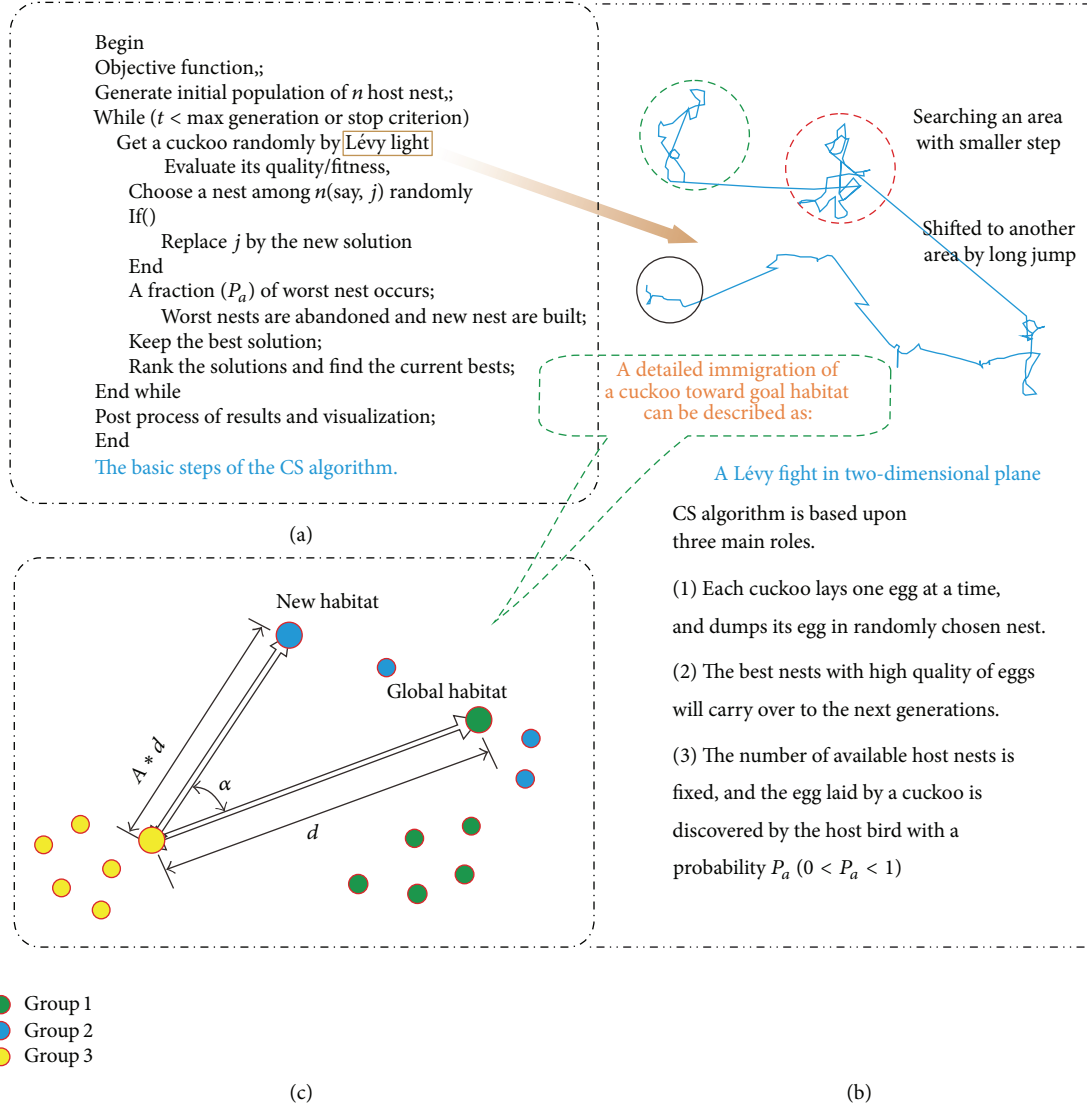


FIGURE 1: The basic steps of the CS algorithm and Lévy flight together with a detailed immigration of a cuckoo toward a goal habitat [18–20]. In this paper, we set the initial population size as 25 and the iteration to 100; an alien egg laid by a cuckoo bird is discovered by the host cuckoo with a probability $P_a = 0.25$.

particles' position and the velocity are given by (20) and (21), respectively [20],

$$x_i^{k+1} = x_i^k + v_i^k, \quad (20)$$

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 (P_{\text{best},i} - x_i^k) + c_2 r_2 (G_{\text{best}} - x_i^k), \quad (21)$$

where ω is the inertia weight; c_1 and c_2 denote the acceleration constants; $P_{\text{best},i}$ and G_{best} are the individual and global best position, separately. For each step, the velocity of each particle is changed toward the best local and global locations. The process is repeated until a predefined termination condition is reached. More details on the PSO are available in literature [32]. In this paper, we set the initial population size as 25 and the iteration as 100, which are the same as the CS algorithm, and the acceleration c_1 and c_2 equals 1.49455.

2.5. Design of the Proposed Hybrid SI-Based Predictive Models.

The original power load data have some noisy information. If not eliminating noise and directly training and building models, the high frequency components may disturb the forecasted load patterns [6]. To improve the load forecasting accuracy, the SSA technique is adopted to extract the trend and periodic components and remove the useless information from the load pattern. After this procedure of data preprocessing, a smoother power loads time series is obtained for forecasting. Furthermore, it can also calculate the period. It is well known that the predictive accuracy of an SVR model largely relies on a reasonable setting of the kernel parameter σ and hyperparameters C . Therefore, the determination of two parameters is a significant issue [59]. However, there is no structural approach or any shortage of opinions on the efficient selection of SVR parameters. The traditional

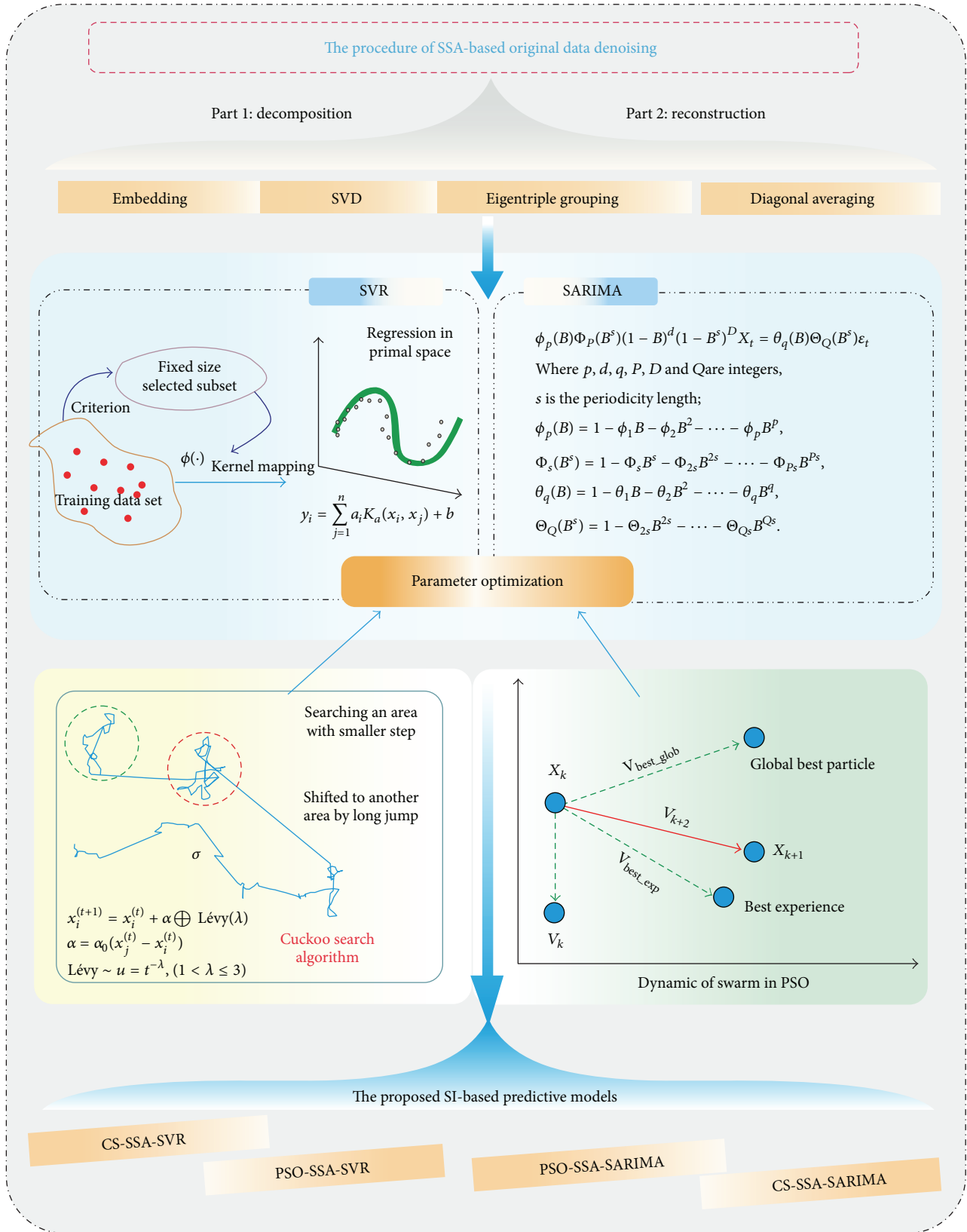


FIGURE 2: The structure of the proposed SI-based hybrid models.

method to determine the parameters of the SVR is a grid search, which is time-consuming and not effective enough to obtain a satisfactory result. In this study, the artificial intelligence optimization (AIO) algorithm, that is, the CS and PSO, is adopted to optimize the parameter selection of the proposed SVR model. Additionally, the traditional method of estimating the parameters of SARIMA is the maximum likelihood estimation, which has an assumption of a normal distribution or another known distribution. However, in the real-world scenario, the power load data are not strictly normally distributed. Thus, employing CS and PSO to optimize the parameters of SARIMA is necessary to build a more proper time series model. The structure of the proposed hybrid models is given in Figure 2.

2.6. Evaluation Criteria. The determination of which prediction model outperforms the other models is of prime concern. In most study cases, model performance is evaluated by numerous error evaluation criteria that can be classified into two main types: absolute error and relative error. For the absolute error, there are the mean absolute error (MAE) and root mean square error (RMSE). For the relative error, there are the mean absolute percentage error (MAPE) and symmetrical mean absolute percentage error (SMPAE). All of them are commonly used to evaluate the accuracy. In this paper, the mean absolute error (MAE) and mean absolute percentage error (MAPE) are used to measure the prediction accuracy of these models. The smaller these values are, the better the predictive performance is.

The MAE can be defined as

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|. \quad (22)$$

The MAPE can be defined as

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\%, \quad (23)$$

where y_t and \hat{y}_t denote the real and predictive values at time t , respectively.

3. Case Study

As described previously, the SSA technique is widely used in extracting principal information; that is, its trend and oscillation components, which are then effectively used for time series forecasting. Understanding that extracting leading information by the SSA is also a procedure of denoising such as wavelet denoising, in this work, we decompose the power load time series and then reconstruct the principal components into a smoother time series. To demonstrate the performance of the SSA denoising used in the power load, in Section 3.1, we specify the procedure of simulating the power load time series from NSW. In Section 3.2, to eliminate noisy information that may disturb the forecasted accuracy, the results of the SARIMA and SVR prediction after the SSA-based denoising are displayed to test the capability of SSA noise elimination. Moreover, forecasting models optimized

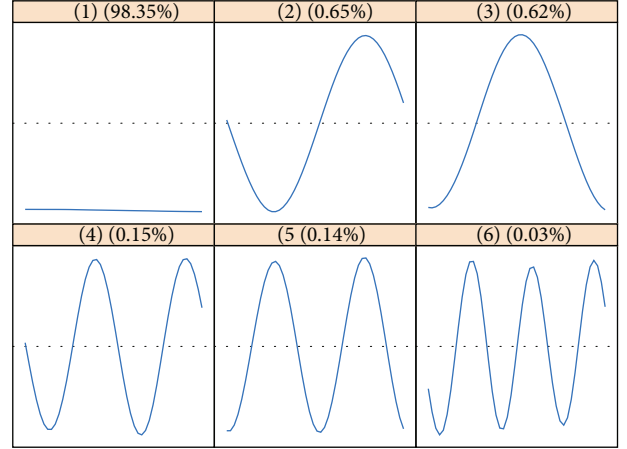


FIGURE 3: 1st stage: eigenvectors ($L = 48$).

by the PSO and CS algorithms are proposed in this section to further improve the predictive ability.

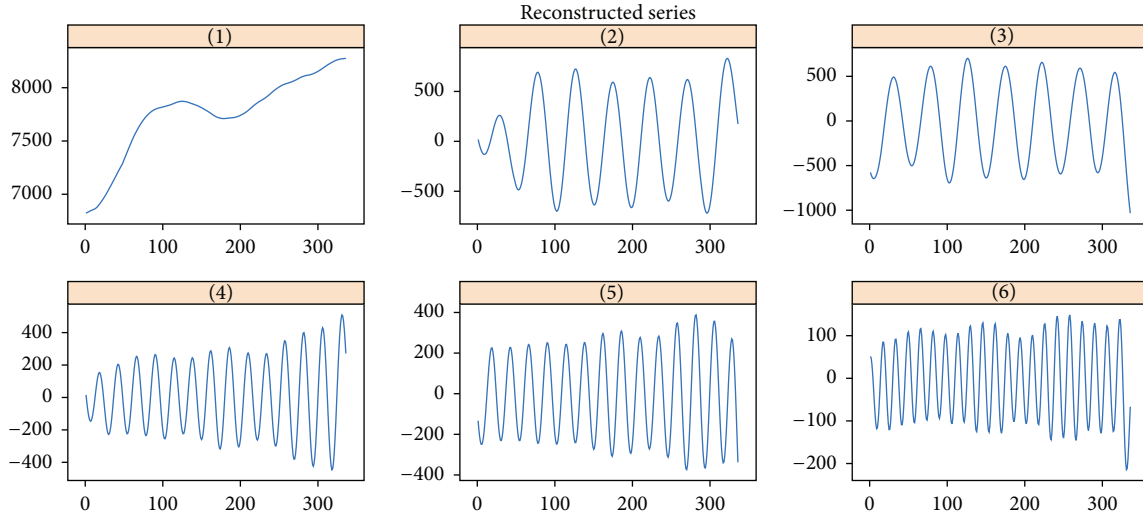
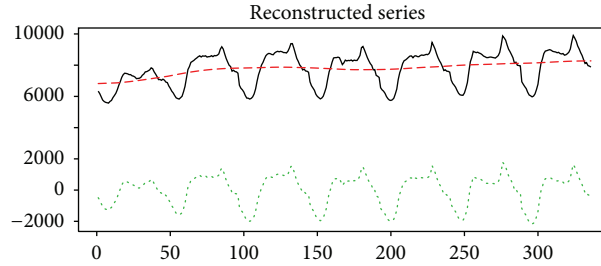
We choose two samples of a half-hour power load, each of which contains 336 training data and 48 test data from April to May in NSW of Australia.

3.1. SSA-Based Denoising. The power load time series in Case 1 is chosen as an example to show the detailed process of SSA-based denoising. The data in Case 2 are denoised by the same procedure as described below. SSA-based denoising has two main stages: the 1st stage is to extract the trend; the 2nd stage is to extract the seasonal components from the residuals from the first step and then reconstruct it.

1st Stage (decomposition). In this stage, the main task is to extract the trend component from the original time series with a small window length. For a varying form of the trend, its extraction is similar to a smoothing effect, and we begin with choosing a possibly minimal window length which in this case is $L = 48$. The SSA with small L performs a smoothing effect of the series by using a filter of order $2L - 1$. The reason we choose this window length is similar to that in moving the averaging procedure: because the smoothing time series includes a periodic component, the window length should be divisible by the period.

Six leading eigenvectors are displayed in Figure 3, which reflects a large contribution of the leading eigentriple. The leading eigenvector contains nearly constant coordinates, and thus, it corresponds to a pure smoothing by the Bartlett filter.

1st Stage (reconstruction). In Figure 4, the result of reconstruction by each of the six eigentriple is illustrated. Combining both figures confirms that the first eigentriple corresponds to the trend, while the rest of the eigentriples contains high frequency components and thus are not related to the first component, the trend. Additionally, Figure 4(1) is roughly considered as the weekly trend, having a share of 98.35% of the power load time series according to Figure 3(1). The trend in Figure 5 is precisely the trend depicted by one

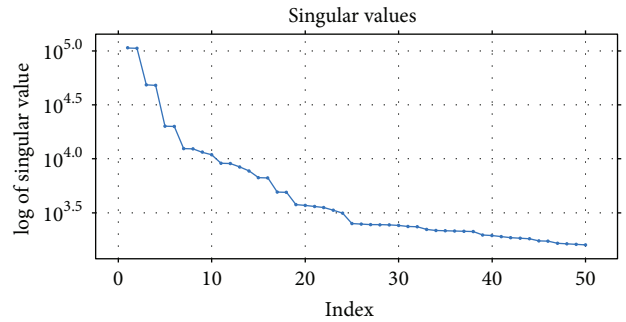
FIGURE 4: 1st stage: elementary reconstructed series ($L = 48$).FIGURE 5: 1st stage: initial series, estimated trend, and residuals ($L = 48$, ET1).

leading eigentriple and coincides with the first reconstructed component from Figure 4(1).

2nd Stage (decomposition and visual information). Because we have extracted out the trend in the first stage, this stage is the extraction of seasonality from the residual component depicted as a green dotted line in Figure 5.

To properly identify the sine waves, we use the graph of eigenvalues (Figure 6), scatterplots of eigenvectors (Figure 7), periodogram (Figure 8), and w -correlation matrix of the elementary components (Figure 9). In Figure 6, we see that several steps are obtained by approximately equal eigenvalues. Eight pairs of eigenvectors are illustrated in Figure 7, showing four nearly regular polygons. The number of edges of polygons represents their periods. ET1-2, ET3-4, ET5-6, and ET7-8 correspond to the periods of 48, 24, 16, and 12 in Figure 8, respectively, and correspond to F1-F2, F3-F4, F5-F6, and F7-F8 in Figure 9. These periods are obtained by the seasonality, are clearly explained by the periodogram (Figure 8), and are estimated as 48.14255, 23.94845, 16.04126, and 12.00288. Figure 9 displays the considered pairs of components for a high correlation within and for a low correlation between.

2nd Stage (reconstruction and plotting of the results). The extracted seasonality (ET1-8) is illustrated in Figure 10(c). A slow change in the sine wave phases and amplitudes is

FIGURE 6: 2nd stage: eigenvalues of residuals ($L = 168$).

observed in Figure 7 and produces a periodic performance with a complicated form. Figure 10 demonstrates the resultant decomposition of both stages of SSA. Figure 11 depicts the reconstructed time series in comparison with the actual power load time series.

3.2. Forecasting Results Analysis of Proposed SI-Based Forecasting Models. In this section, we employ SARIMA and SVR to build the forecasting models after denoising the useless information in the power load by SSA. To further improve

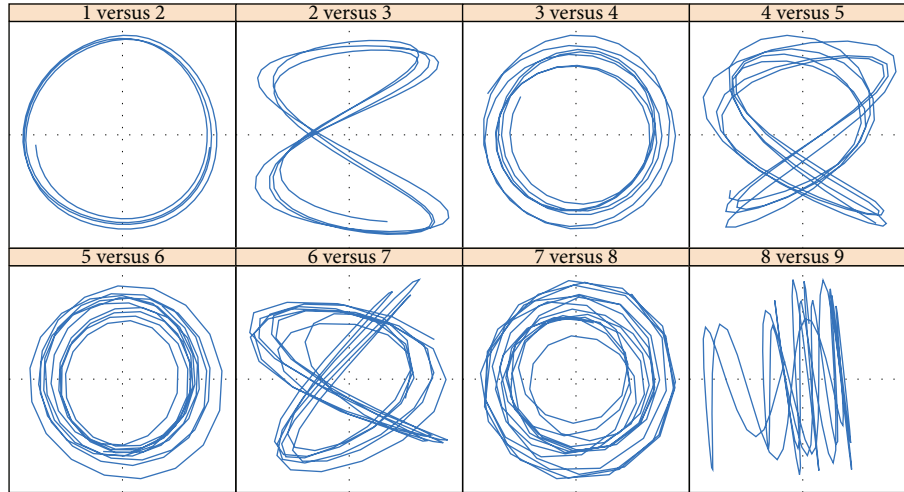
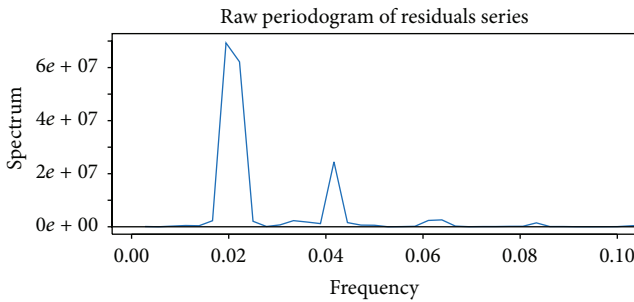
FIGURE 7: 2nd stage: scatterplot for eigenvector pairs ($L = 168$).

FIGURE 8: 2nd stage: periodogram of the series (i.e., of the residual at the 1st stage).

the accuracy of forecasting, we optimize the parameters of SAMRIA and SVR by the CS and PSO.

The proposed hybrid models are applied for short-term (half an hour) load forecasting with a 48-step ahead of NSW in Australia over a prediction of one day in two cases. The performance of the proposed methods is marked as CS-SSA-SARIMA and CS-SSA-SVR. The comparisons of the power load forecasting results are intuitively shown in Figures 12 and 13. The forecasts for Case 1 based on the PSO-SSA-SARIMA, CS-SSA-SARIMA, PSO-SSA-SVR, and CS-SSA-SVR models are illustrated in Figure 12, while Figure 13 shows the results of the proposed methods for Case 2. Figures 12(b) and 12(c) are boxplots of the percent error (PE) and error (E), respectively. In Figure 12(b), the first to fourth boxplots correspond to the models of SARIMA, SSA-SARIMA, PSO-SSA-SARIMA, and CS-SSA-SARIMA. The positive values of PE and E mean the forecasting values are underestimated, while the negative values indicate the forecasting values are overestimated. The rest of the graphs and subgraphs can be identified in the same manner.

It can be seen from Figures 12 to 13 over the predictive horizon that the forecasting series obtained by the proposed hybrid models CS-SSA-SARIMA and CS-SSA-SVR perform substantially better with synchronicity and a range of

the vibration of the series approximating to the original series. In Figures 12(a) and 12(d), it is obvious that both of the curves for SSA-SARIMA and SSA-SVR are substantially closer to the original data. A consistent conclusion can be seen in Figures 12(b), 12(c), 12(e), and 12(f). In Figures 12(b) and 12(c), the boxplots from left to right represent the accuracy of SARIMA, SSA-SARIMA, PSO-SSA-SARIMA, and CS-SSA-SARIMA. Through observing the trend of these four boxplots from Figures 12(b) and 12(c), we discover a trend evidently decreasing progressively. Similarly, a decreasing tendency can be observed from Figures 12(e) and 12(f). Furthermore, when observing each of boxplot from Figure 12(b), we notice that many forecasting values of SARMIA model are underestimated, while that of SSA-SARIMA are less underestimated because the median of SSA-SARIMA's boxplot is much smaller than that of SARIMA's boxplot. These results combined with the similar results obtained from Figure 13 demonstrate the excellent denoising performance of the SSA and that processed data by the SSA can enhance the forecasting estimate of the models built by SARIMA and SVR.

In Figures 12(a)–12(c), it can obviously be observed that the forecasting performance of SARIMA, SSA-SARIMA, and PSO-SSA-SARIMA are substantially lower than that obtained by the proposed model, especially during the periods of peak load, which is marked with black dotted rectangles. Moreover, the same phenomenon can be obtained from Figures 12(d)–12(f) with the poorest forecasting results obtained by a single SVR model. Additionally, from Figures 12(e) and 12(f) the median of CS-SSA-SVR's boxplot is close to zero, while median of other models are not. This indicates that the forecasting values of proposed model are only slightly underestimated. In sum, the forecasting values of model SSA-SVR are substantially better than that of the SVR but are inferior to that generated by the model CS-SSA-SVR. As for Case 2, the similar results can be concluded from Figure 13. These figures reveal that the models of SSA-SARIMA and SSA-SVR based on the optimization of the CS have a higher predictive precision.

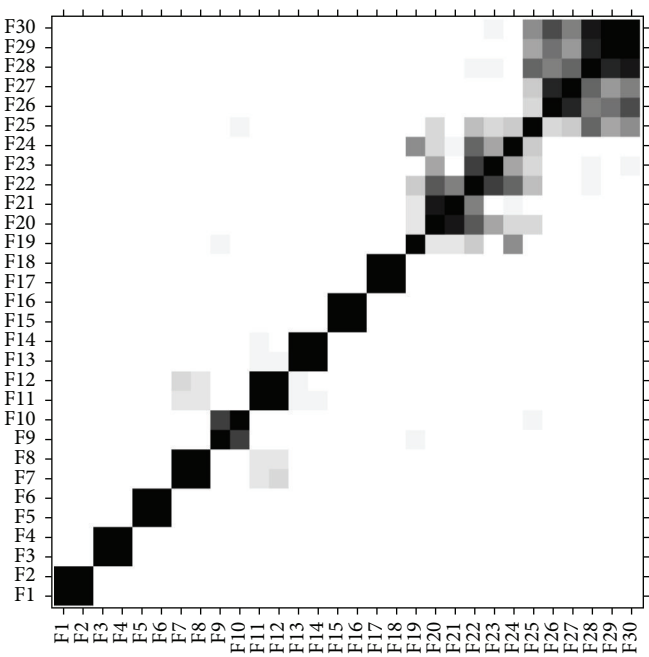


FIGURE 9: 2nd stage: w-correlation matrix ($L = 168$).

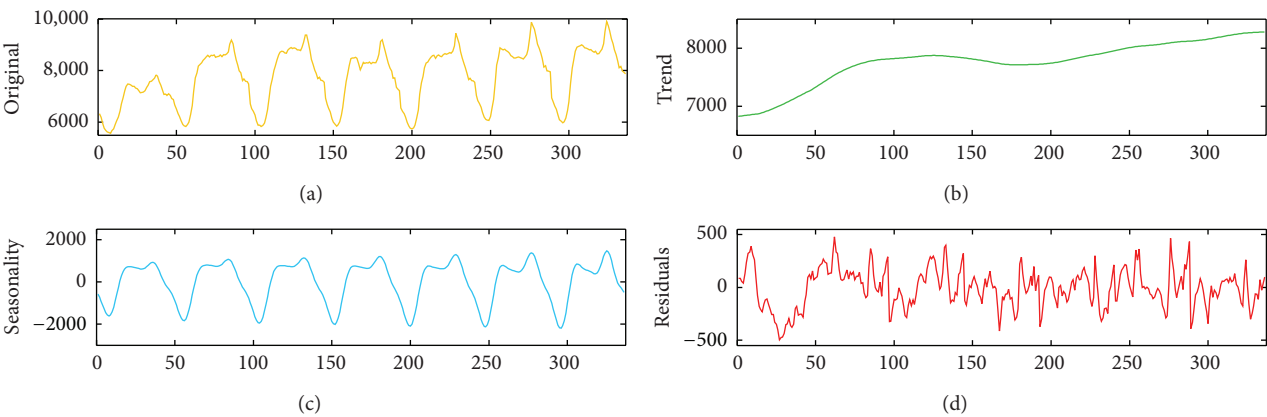


FIGURE 10: 2nd stages: original series and its trend-periodic-residuals decomposition.

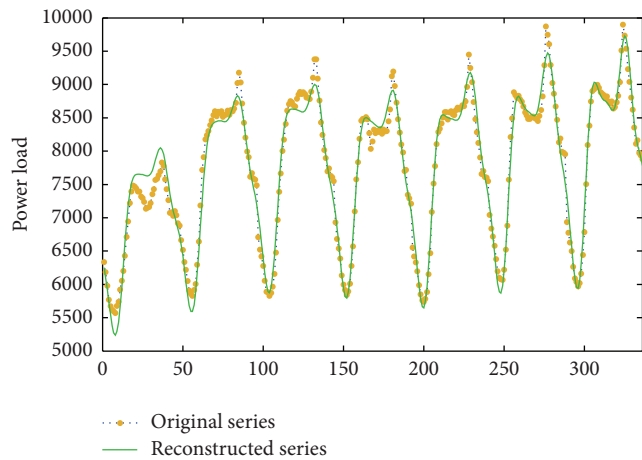


FIGURE 11: Reconstruction series and original series.

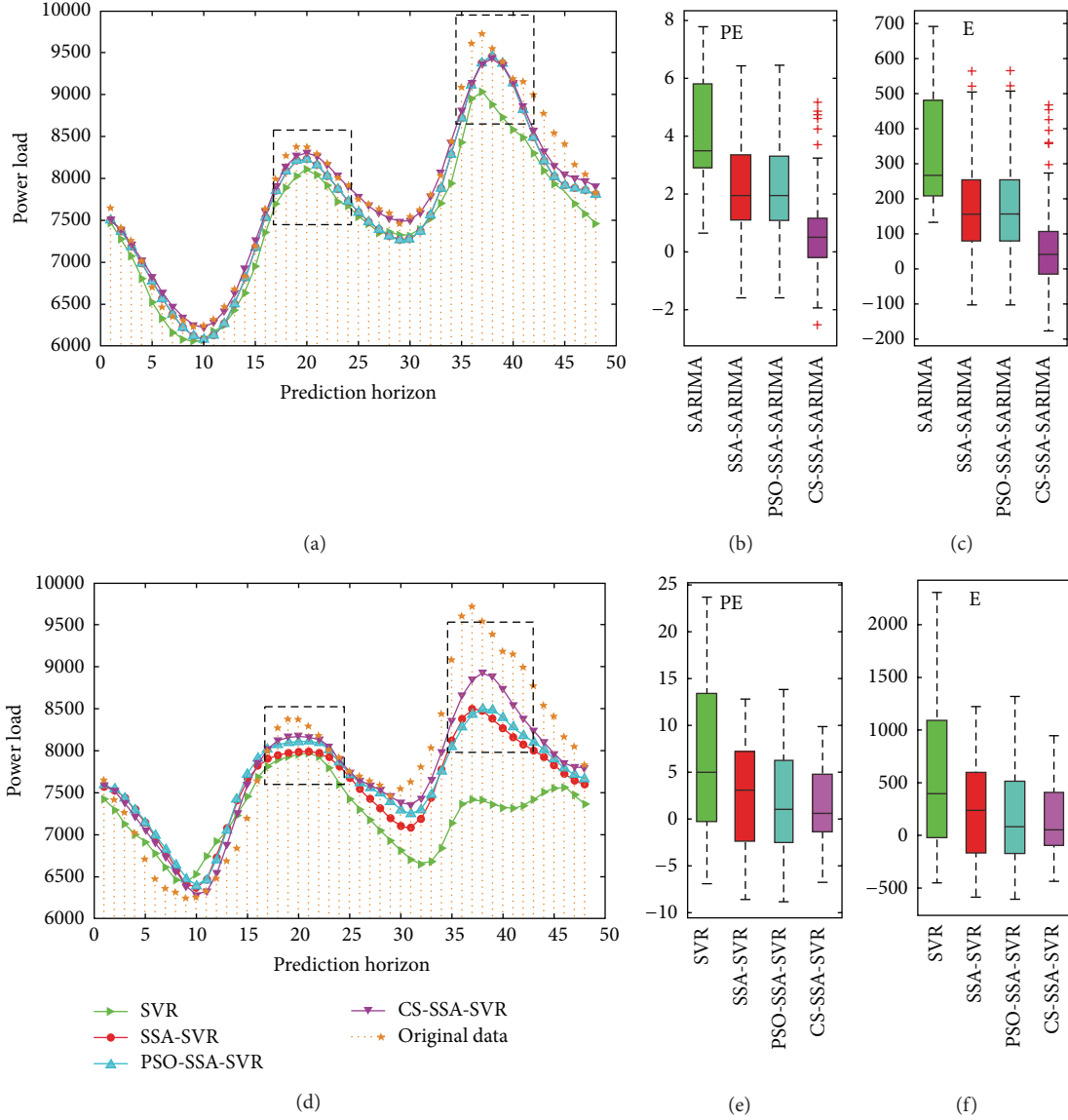


FIGURE 12: The forecasting results of the proposed PSO-SSA-SARIMA, CS-SSA-SARIMA, PSO-SSA-SVR, and CS-SSA-SVR models for Case 1 ((b, e): $PE(t) = (\hat{y}(t) - y(t))/y(t)$, (c, f): $E(t) = \hat{y}(t) - y(t)$).

Figure 14 illustrates the convergence speed by PSO and CS optimizing the SSA-SARIMA model in Case 1 and Case 2, respectively. It is evident that in both cases, both the convergence point and MSE of the PSO is smaller than that of the CS, while the precision of the prediction for the CS-SSA-SARIMA model is higher than that of the PSO illustrated from Figures 12 and 13 and listed in Table 1. This demonstrates that PSO has a faster convergence speed but a lower accuracy than the CS in these two cases, implying that the PSO appears to be overfitting.

To evaluate the forecasting model quantitatively, the statistical errors are computed in testing datasets over the forecasting horizon. Table 1 and Figure 15 give the statistical error measures' comparisons between different models. As shown in Table 1 and Figure 15, MAPE of SARIMA decreases from 4.22% to 2.37% in Case 1 and from 6.87% to 4.60% in Case

TABLE 1: Statistical error measures' comparison between different models.

Model	MAPE (%)		MAE	
	Case 1	Case 2	Case 1	Case 2
SARIMA	4.22	6.87	344.32	613.62
SSA-SARIMA	2.37	4.60	190.68	404.10
PSO-SSA-SARIMA	2.37	4.47	190.79	392.06
CS-SSA-SARIMA	1.38	3.82	113.72	333.12
SVR	8.75	6.64	734.83	580.10
SSA-SVR	5.53	5.76	451.26	516.21
PSO-SSA-SVR	4.96	4.68	404.26	392.42
CS-SSA-SVR	3.42	3.73	278.86	305.88

2. The result of the MAE is similar to the MAPE, decreasing

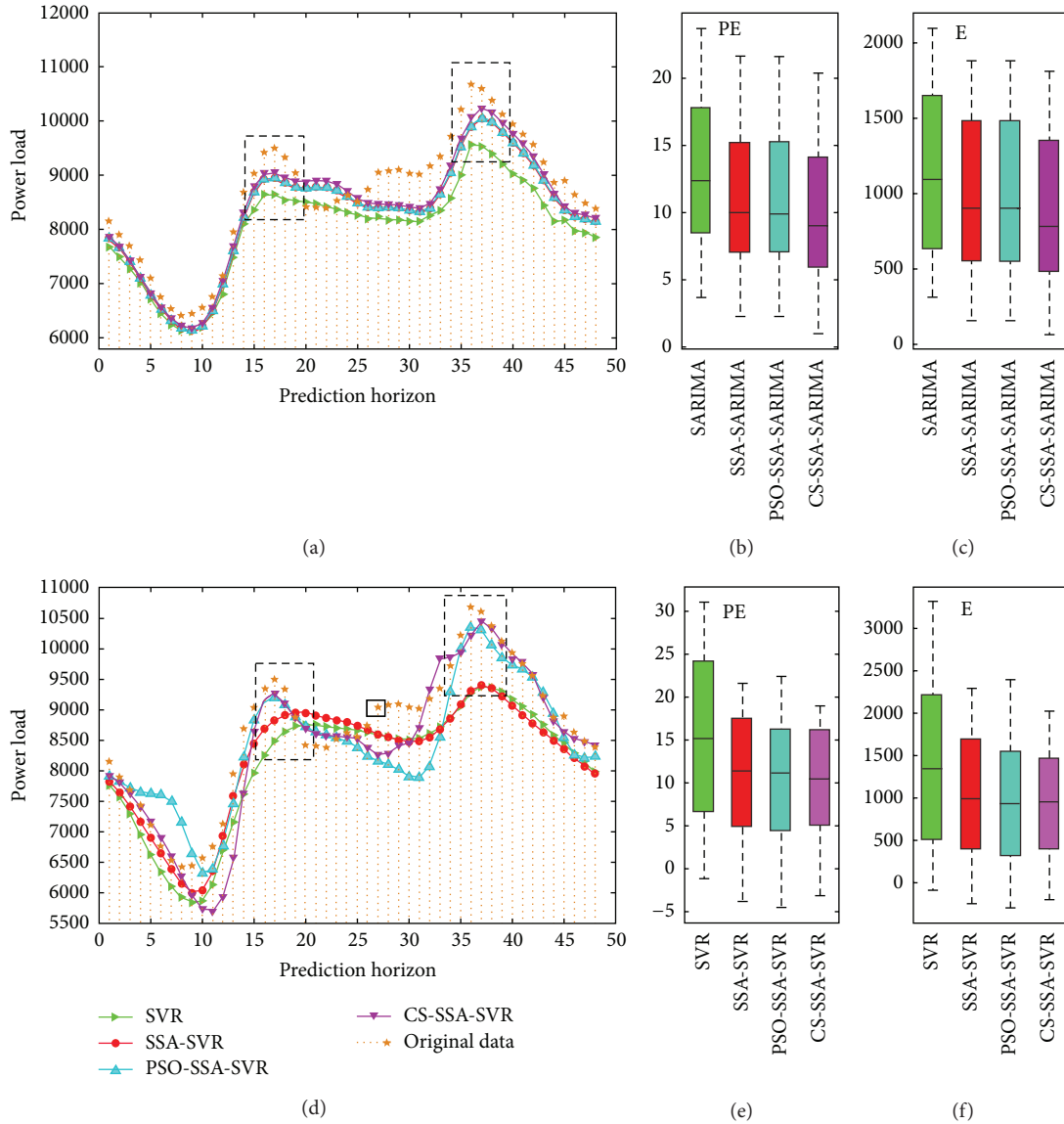


FIGURE 13: The forecasting results of the proposed PSO-SSA-SARIMA, CS-SSA-SARIMA, PSO-SSA-SVR, and CS-SSA-SVR models for Case 2 ((b, e): $PE(t) = (y(t) - \hat{y}(t))/y(t)$, (c, f): $E(t) = y(t) - \hat{y}(t)$).

from 344.32 to 190.68 in Case 1 and from 613.62 to 404.10 in Case 2. As for the SVR, the SSA-SVR outperforms the single SVR method because both the MAPE and MAE decrease in Case 1 and Case 2. This result demonstrates that the SSA has a strong capacity for noise elimination, which can improve the forecasting performance of SARIMA and SVR. Considering that the cuckoo search can be a very effective tool in parameter searching for further improving the accuracy of the SARIMA and SVR models, in this study, the CS-SSA-SARIMA and CS-SSA-SVR models outperform the SSA-SARIMA, PSO-SSA-SARIMA, SSA-SVR, and PSO-SSA-SVR models, respectively. These results prove that the cuckoo search is capable of improving the accuracy of model forecasting. Then, a comparison between the CS-SSA-SARIMA and PSO-SSA-SARIMA models and a comparison between the CS-SSA-SVR and PSO-SSA-SVR models indicate that the CS

algorithm outperforms the PSO algorithm in the application of improving the forecasting capacity of the proposed hybrid models.

4. Conclusions and Future Work

This paper presents hybrid swarm intelligent forecasting model strategies to accurately predict the short-term power load. The results obtained in this study illustrate that the SSA technique can be successfully used as a noise eliminating technique for time series similar to the short-term power load time series used here. The SSA-based denoising technique is capable of extracting important trend and seasonal components and then reconstructing it into smooth data to enhance the forecasting accuracy for SARIMA and SVR. In addition, the good noise eliminating ability via SSA could

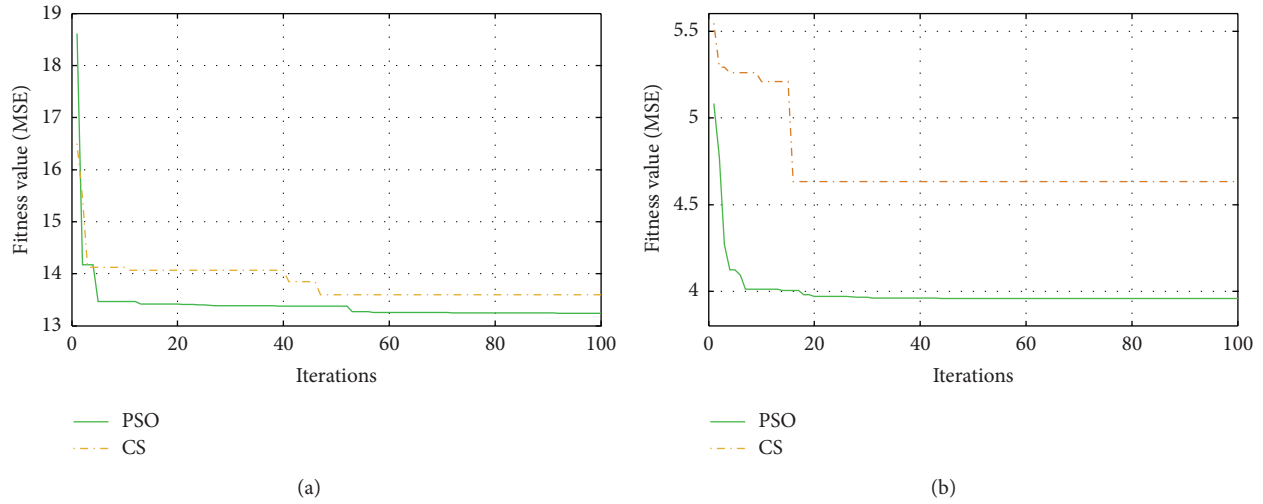


FIGURE 14: CS and PSO convergence procedure during the training for the NSW load.

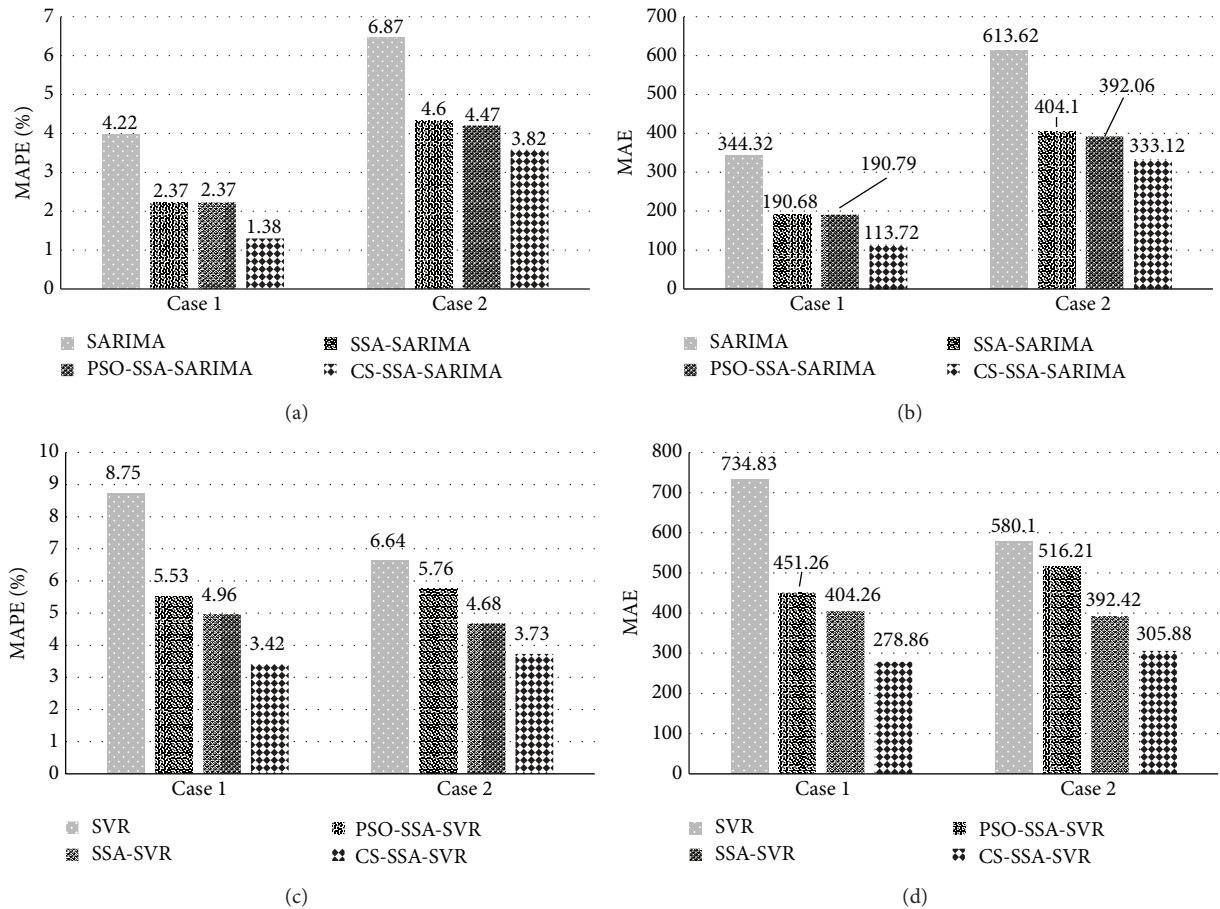


FIGURE 15: Statistical error measures' comparison between different models.

make these characteristics more obvious when modeling and could provide a more accurate forecast by SARIMA and SVR.

The CS algorithm is a recently developed metaheuristic artificial intelligence algorithm of parameter optimization. It has the ability to search parameters outstrips, that of

the maximum likelihood estimation method and that of the traditional optimization algorithm (PSO) when estimating the parameters of SARIMA. Similarly, its capability of expanding the scope of the search intelligently provides an optimization that is more effective and efficient than a grid

search when searching the optimal hyperparameters in SVR. Although PSO has a faster convergence velocity, it appears to have an overfitting problem. In our cases, it is revealed that by optimizing the parameters of the SSA-SARIMA and SSA-SVR models, the CS algorithm can further enhance the accuracy of prediction in short-term power loads and obtains a higher precision than PSO. The proposed hybrid swarm intelligent forecasting model could predict the short-term power load in a real-world scenario, which helps to enhance the predictive accuracy of the power system.

In this paper, our contribution is that an SI-based forecasting model is proposed to highly increase the accuracy. However, we did not sufficiently compare other feasible forecasting models, data preprocessing methods, and AI algorithms, such as BP, autoregressive integrated moving average (ARIMA), wavelet analysis, and GA. A more detailed comparison between the proposed method and other feasible forecasting models, data preprocessing methods, and SI optimizations is required. This is a very heavy workload but is very meaningful research; thus, it is necessary to perform additional research in future work.

For future work, we outline four directions. The first direction is to study the use of the other feasible forecasting models mentioned above within our framework. The second direction is to study in detail the other feasible SI optimization algorithms mentioned in this paper to search parameters of various forecasting models. The third direction is to study the use of other feasible data preprocessing methods, including the methods of not only denoising useless information but also removing outliers. The fourth direction is to explore the ability of different SI optimization algorithms to search certain parameters of certain forecasting models.

Highlights

- (i) A novel swarm intelligence-based hybrid approach is proposed for short-term load forecasting.
- (ii) The proposed approach consists of three steps to increase the forecasting accuracy.
- (iii) SSA is used for removing noised information in the first step.
- (iv) SARIMA and SVR are used for forecasting in the second step.
- (v) CS is employed to optimize the parameters of SARIMA and SVR.
- (vi) The proposed approach can improve the forecasting accuracy.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant no. 7117110).

References

- [1] A. U. Haque, J. Meng, P. Mandal, and R. L. Pineda, "Performance evaluation of different optimization algorithms for power demand forecasting applications in a smart grid environment," *Procedia Computer Science*, vol. 12, pp. 320–325, 2012.
- [2] F. Javed, N. Arshad, F. Wallin, I. Vassileva, and E. Dahlquist, "Forecasting for demand response in smart grids: an analysis on use of anthropologic and structural data and short term multiple loads forecasting," *Applied Energy*, vol. 96, pp. 150–160, 2012.
- [3] D. Niu, Y. Wang, and D. D. Wu, "Power load forecasting using support vector machine and ant colony optimization," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2531–2539, 2010.
- [4] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 134–141, 2012.
- [5] M. López, S. Valero, C. Senabre, J. Aparicio, and A. Gabaldon, "Application of SOM neural networks to short-term load forecasting: the Spanish electricity market case study," *Electric Power Systems Research*, vol. 91, pp. 18–27, 2012.
- [6] R.-A. Hooshmand, H. Amooshahi, and M. Parastegari, "A hybrid intelligent algorithm based short-term load forecasting approach," *International Journal of Electrical Power and Energy Systems*, vol. 45, no. 1, pp. 313–324, 2013.
- [7] M. Moazzami, A. Khodabakhshian, and R. Hooshmand, "A new hybrid day-ahead peak load forecasting method for Iran's National Grid," *Applied Energy*, vol. 101, pp. 489–501, 2013.
- [8] V. Vapnik, S. Golowich, and A. Smola, "Support vector machine for function approximation, regression estimation, and signal processing," *Advances in Neural Information Processing Systems*, vol. 9, pp. 281–287, 1996.
- [9] W.-M. Hung and W.-C. Hong, "Application of SVR with improved ant colony optimization algorithms in exchange rate forecasting," *Control and Cybernetics*, vol. 38, no. 3, pp. 863–891, 2009.
- [10] P.-F. Pai, C.-S. Lin, W.-C. Hong, and C.-T. Chen, "A hybrid support vector machine regression for exchange rate prediction," *International Journal of Information and Management Sciences*, vol. 17, no. 2, pp. 19–32, 2006.
- [11] S. Ch, N. Anand, B. K. Panigrahi, and S. Mathur, "Streamflow forecasting by SVM with quantum behaved particle swarm optimization," *Neurocomputing*, vol. 101, pp. 18–23, 2013.
- [12] G.-F. Lin, Y.-C. Chou, and M.-C. Wu, "Typhoon flood forecasting using integrated two-stage support vector machine approach," *Journal of Hydrology*, vol. 486, pp. 334–342, 2013.
- [13] W.-C. Hong, Y. Dong, L.-Y. Chen, and S.-Y. Wei, "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting," *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 1881–1890, 2011.
- [14] W.-C. Hong, "Hybrid evolutionary algorithms in a SVR-based electric load forecasting model," *International Journal of Electrical Power and Energy Systems*, vol. 31, no. 7-8, pp. 409–417, 2009.
- [15] W.-C. Hong, "Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model," *Energy Conversion and Management*, vol. 50, no. 1, pp. 105–117, 2009.
- [16] W.-C. Hong, "Electric load forecasting by support vector model," *Applied Mathematical Modelling*, vol. 33, no. 5, pp. 2444–2454, 2009.

- [17] W.-C. Hong, "Application of chaotic ant swarm optimization in electric load forecasting," *Energy Policy*, vol. 38, no. 10, pp. 5830–5839, 2010.
- [18] S. Pino-Povedano and F.-J. González-Serrano, "Comparison of optimization algorithms in the sensor selection for predictive target tracking," *Ad Hoc Networks*, vol. 20, pp. 182–192, 2014.
- [19] H. Kaydani and A. Mohebbi, "A comparison study of using optimization algorithms and artificial neural networks for predicting permeability," *Journal of Petroleum Science and Engineering*, vol. 112, pp. 17–23, 2013.
- [20] J. Ahmed and Z. Salam, "A Maximum Power Point Tracking (MPPT) for PV system using Cuckoo Search with partial shading capability," *Applied Energy*, vol. 119, pp. 118–130, 2014.
- [21] X. S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *World Congress on Nature & Biologically Inspired Computing*, pp. 210–214, IEEE, 2009.
- [22] A. Gherboudj, A. Layeb, and S. Chikhi, "Solving 0-1 knapsack problems by a discrete binary version of cuckoo search algorithm," *International Journal of Bio-Inspired Computation*, vol. 4, no. 4, pp. 229–236, 2012.
- [23] P. R. Srivastava, A. Varshney, P. Nama, and X.-S. Yang, "Software test effort estimation: a model based on cuckoo search," *International Journal of Bio-Inspired Computation*, vol. 4, no. 5, pp. 278–285, 2012.
- [24] M. K. Marichelvam, "An improved hybrid Cuckoo Search (IHCS) metaheuristics algorithm for permutation flow shop scheduling problems," *International Journal of Bio-Inspired Computation*, vol. 4, no. 4, pp. 200–205, 2012.
- [25] P. R. Srivastava, C. Sravya, A. S. Kamiseti, and M. Lakshmi, "Test sequence optimisation: an intelligent approach via cuckoo search," *International Journal of Bio-Inspired Computation*, vol. 4, no. 3, pp. 139–148, 2012.
- [26] M. Basu and A. Chowdhury, "Cuckoo search algorithm for economic dispatch," *Energy*, vol. 60, pp. 99–108, 2013.
- [27] E. Cuevas, F. Sención, D. Zaldivar, M. Pérez-Cisneros, and H. Sossa, "A multi-threshold segmentation approach based on artificial bee colony optimization," *Applied Intelligence*, vol. 37, no. 3, pp. 321–336, 2012.
- [28] R. C. Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceeding IEEE International Conference on Evolutionary Computation*, vol. 1, pp. 81–86, 2001.
- [29] K. Hammouche, M. Diaf, and P. Siarry, "A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation," *Computer Vision and Image Understanding*, vol. 109, no. 2, pp. 163–175, 2008.
- [30] M.-H. Horng, "Multilevel minimum cross entropy threshold selection based on the honey bee mating optimization," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4580–4592, 2010.
- [31] M.-H. Horng and T.-W. Jiang, "Multilevel image thresholding selection using the artificial bee colony algorithm," in *Artificial Intelligence and Computational Intelligence*, vol. 6320 of *Lecture Notes in Computer Science*, pp. 318–325, Springer, Berlin, Germany, 2010.
- [32] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, December 1995.
- [33] E. García-Gonzalo and J. L. Fernández-Martínez, "A Brief historical review of particle swarm optimization (PSO)," *Journal of Bioinformatics and Intelligent Control*, vol. 1, no. 1, pp. 3–16, 2012.
- [34] L. Ali and S. L. Sabat, "Particle swarm optimization based universal solver for global optimization," *Journal of Bioinformatics and Intelligent Control*, vol. 1, no. 1, pp. 95–105, 2012.
- [35] Z. H. Cui and X. J. Cai, "Integral particle swarm optimization with dispersed accelerator information," *Fundamenta Informaticae*, vol. 95, no. 4, pp. 427–447, 2009.
- [36] Z. H. Cui, X. J. Cai, J. C. Zeng, and Y. F. Yin, "PID-controlled particle swarm optimization," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 16, no. 6, pp. 585–609, 2010.
- [37] P. D. Sathya and R. Kayalvizhi, "Optimum multi level image thresholding based on Tsallis entropy method with bacterial foraging algorithm," *International Journal of Computer Science*, vol. 7, no. 5, pp. 336–343, 2010.
- [38] W. Tao, H. Jin, and L. Liu, "Object segmentation using ant colony optimization algorithm and fuzzy entropy," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 788–796, 2007.
- [39] J. Wang, J. Wang, Y. Li, S. Zhu, and J. Zhao, "Techniques of applying wavelet de-noising into a combined model for short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 816–824, 2014.
- [40] G. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 1970.
- [41] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 1976.
- [42] J. B. Elsner and A. A. Tsonis, *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, Plenum Press, New York, NY, USA, 1996.
- [43] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure—SSA and Related Techniques*, vol. 90, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2001.
- [44] M. R. S. Claudio, "Singular spectrum analysis and forecasting of failure time series," *Reliability Engineering & System Safety*, vol. 114, no. 1, pp. 126–136, 2013.
- [45] R. Vautard and M. Ghil, "Singular spectrum analysis in non-linear dynamics, with applications to paleoclimatic time series," *Physica D: Nonlinear Phenomena*, vol. 35, no. 3, pp. 395–424, 1989.
- [46] R. Vautard, P. Yiou, and M. Ghil, "Singular-spectrum analysis: a toolkit for short, noisy chaotic signals," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1–4, pp. 95–126, 1992.
- [47] M. de Carvalho, P. C. Rodrigues, and A. Rua, "Tracking the US business cycle with a singular spectrum analysis," *Economics Letters*, vol. 114, no. 1, pp. 32–35, 2012.
- [48] H. Hassani, S. Heravi, and A. Zhigljavsky, "Forecasting European industrial production with singular spectrum analysis," *International Journal of Forecasting*, vol. 25, no. 1, pp. 103–118, 2009.
- [49] N. Golyandina and A. Korobeynikov, "Basic singular spectrum analysis and forecasting with R," *Computational Statistics & Data Analysis*, vol. 71, pp. 934–954, 2014.
- [50] Q. Chen, T. van Dam, N. Sneeuw, X. Collilieux, M. Weigelt, and P. Rebischung, "Singular spectrum analysis for modeling seasonal signals from GPS time series," *Journal of Geodynamics*, vol. 72, pp. 25–35, 2013.
- [51] C. L. Wu and K. W. Chau, "Rainfall-runoff modeling using artificial neural network coupled with singular spectrum analysis," *Journal of Hydrology*, vol. 399, no. 3–4, pp. 394–409, 2011.
- [52] H. Hassani, "Singular spectrum analysis: methodology and comparison," *Journal of Data Science*, vol. 5, pp. 239–257, 2007.
- [53] K.-Y. Chen and C.-H. Wang, "A hybrid SARIMA and support vector machines in forecasting the production values of the

- machinery industry in Taiwan,” *Expert Systems with Applications*, vol. 32, no. 1, pp. 254–264, 2007.
- [54] M. Bouzerdoum, A. Mellit, and A. M. Pavan, “A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant,” *Solar Energy*, vol. 98, pp. 226–235, 2013.
- [55] V. Vapnik, S. Golowich, and A. J. Smola, “Support vector method for function approximation, regression estimation and signal processing,” *Advanced Neural Information Processing System*, vol. 9, pp. 281–287, 1996.
- [56] 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,” *International Journal of Electrical Power and Energy Systems*, vol. 44, no. 1, pp. 604–614, 2013.
- [57] S. Amari and S. Wu, “Improving support vector machine classifiers by modifying kernel functions,” *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.
- [58] S. Pino-Povedano and F. J. González-Serrano, “Comparison of optimization algorithms in the sensor selection for predictive target tracking,” *Ad Hoc Networks*, vol. 20, pp. 182–192, 2014.
- [59] 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,” *Electrical Power and Energy Systems*, vol. 44, no. 1, pp. 604–614, 2013.

