



Energy Consumption Forecasting for the Nonferrous Metallurgy Industry Using Hybrid Support Vector Regression with an Adaptive State Transition Algorithm

Zhaoke Huang¹ · Chunhua Yang¹ · Xiaojun Zhou¹ · Shengxiang Yang²

Received: 18 August 2018 / Accepted: 12 April 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The nonferrous metallurgy industry is a major energy consumer in China, and accurate energy consumption forecasting for the nonferrous metallurgy industry can help government policymakers with energy planning. For this purpose, a hybrid support vector regression (HSVR) with an adaptive state transition algorithm (ASTA) named ASTA-HSVR is proposed to forecast energy consumption in the nonferrous metallurgy industry. The proposed support vector regression (SVR) model consists of a linear weighting of ϵ -SVR and ν -SVR. The ASTA was developed to optimize the parameters of the HSVR. Two cases of energy consumption from the nonferrous metallurgy industry in China are used to demonstrate the performance of the proposed method. The results indicate that the ASTA-HSVR method is superior to other methods. In this study, a hybrid support vector regression with an adaptive state transition algorithm (ASTA-HSVR) was developed and successfully applied to energy consumption forecasting for the nonferrous metallurgy industry. However, it should be noted that the outliers were not considered in this study. In the future, we expect to extend the ASTA-HSVR method to include energy consumption forecasting problems with outliers.

Keywords Energy consumption forecasting · Support vector regression · Adaptive · State transition algorithm

Introduction

Nonferrous metals are indispensable materials for the development of modern industry, modern national defense, and modern science and technology. The main task of nonferrous metallurgy is to refine them from their original sources through complex physical and chemical processes. Hence, the nonferrous metallurgy industry is of importance in the national economy. However, with the rapid development of the Chinese economy, the nonferrous metallurgy industry has become one of the seven major industrial energy consumers in China. In 2012, the power consumption of the nonferrous metallurgy industry accounted for 7.67% of the total power consumption

in China, and 10.54% of the total industrial power consumption. In the electrolytic aluminum industry, 500 billion kilowatt-hours of electricity and 180 million tons of standard coal are consumed every year. The copper-smelting industry consumes 300 kg of standard coal per ton of copper produced. Notably, the energy consumption of the nonferrous metallurgy industry is still growing from a high base, and the energy consumption structure of it is also changing dynamically. Thus, studying an energy consumption model of the nonferrous metallurgy industry has practical significance. This is expected to help government policymakers with energy planning and drafting energy policy, and thereby, to improve the relationship between energy supply and demand.

In recent years, numerous methods including such as neural networks and kernel-based methods have been proposed for use in forecasting energy consumption. Limanond et al. [10] forecast the transportation energy consumption in Thailand using log-linear regression models and feed-forward neural network models. Meng et al. [13] proposed using the partial least squares method to forecast the annual electricity consumption of China. Yuan et al.

✉ Xiaojun Zhou
michael.x.zhou@csu.edu.cn

¹ School of Automation, Central South University, Changsha 410083, China

² School of Computer Science and Informatics, De Montfort University, Leicester, UK

[26] forecast China's primary energy consumption using the ARIMA (autoregressive integrated moving average) model and GM(1,1) (Grey model). Szoplik [19] presented artificial neural networks (ANN) to forecast the natural gas consumption in Szczecin (Poland). Ruiz et al. [16] forecast the energy consumption in public buildings using Elman neural networks (ENN) with a genetic algorithm (GA). Xiao et al. [24] developed a hybrid model based on a selective ensemble to forecast energy consumption in China. Chen et al. [3] proposed a hybrid support vector regression model to predict the short-term electric demand in the building sector. To summarize, many models are available for forecasting energy consumption. Among them, support vector regression (SVR) and neural networks (NNs) are both currently research hotspots [4, 12]. However, the main disadvantage of NNs is also apparent, that is, their lack of interpretability, slow convergence, and plunging into local optima easily. Hence, SVR is a more promising method to be applied to forecast energy consumption. Moreover, the superior performance of SVR has been demonstrated in many applications [2, 15, 21, 22, 25], because SVR is theoretically guaranteed to achieve the global optimum by implementing the structural risk minimization principle. However, in the nonferrous metallurgy industry, one characteristic of the energy consumption data is that they have different time scales, which leads to a problem: a single SVR model cannot be applied to forecasting on different time scales. In addition, the performance of the SVR relies too much on the selection of its parameters.

State transition algorithm (STA) is a novel intelligent optimization method, and the powerful global search ability and flexibility of STA have been demonstrated in many real-world applications [5–9, 27–30, 32–34]. For instance, Zhang et al. [27] presented a new STA to select the optimal fractional-order PID controller parameters. Han et al. [6] developed an STA-based method to solve the multi-threshold image segmentation problem. Huang et al. [7] investigated a novel dynamic optimization method based on the STA to control zinc powder in the copper removal process. Zhou et al. [30] introduced a stochastic intelligent optimization method based on the STA to solve the sensor network localization problem. Hence, the STA is a promising method for optimizing the parameters of the SVR. In addition, in [8], Huang et al. used a novel cognitively inspired computing method based on the STA to obtain an approximate optimal solution for the linear bi-level programming problem. Thus, it is important to solve practical problems using a method like STA that associated human intelligence and nature cognition.

In this study, in order to solve the above problem, a hybrid support vector regression (HSVR) with an adaptive state transition algorithm (ASTA) named ASTA-HSVR, was

proposed to forecast energy consumption in the nonferrous metallurgy industry. On the one hand, for improving the generalization ability on different time scales, a hybrid support vector regression forecasting model was presented, which consists of a linear weighting of ϵ -SVR and ν -SVR. On the other hand, in order to optimize the parameters of the HSVR, an adaptive state transition algorithm was developed based on the STA. Finally, two cases of energy consumption from the nonferrous metallurgy industry in China are used to demonstrate the performance of the proposed method. The novelty and main contributions of the proposed method are highlighted as follows:

- A hybrid support vector regression model with adaptive state transition algorithm (ASTA-HSVR) is presented to forecast energy consumption in the nonferrous metallurgy industry.
- An adaptive state transition algorithm is proposed for parameter tuning of the hybrid SVR model including the weights.
- The proposed ASTA-HSVR method is compared with the HSVR in combination with other meta-heuristics algorithms and the non-iterative algorithms.
- Two cases of energy consumption from the nonferrous metallurgy industry in China are forecast using the proposed ASTA-HSVR method.

The rest of this paper is organized as follows. Section “[Energy Consumption in the Nonferrous Metallurgy Industry](#)” introduces the energy consumption in the nonferrous metallurgy industry. Section “[The Proposed Hybrid Forecasting Method: ASTA-HSVR](#)” presents the proposed hybrid forecasting method. Section “[Case Study from the Nonferrous Metallurgy Industry](#)” describes the efficiency of the proposed method via two cases of energy consumption in the nonferrous metallurgy industry. Finally, Section “[Conclusions](#)” presents the conclusions drawn from the work.

Energy Consumption in the Nonferrous Metallurgy Industry

The term nonferrous metallurgy refers to the production activities of extracting common nonferrous metals from raw nonferrous metal sources such as nonferrous minerals and discarded metals by methods such as smelting, refining, or electrolysis. The nonferrous metallurgy involves the smelting of aluminum, copper, nickel, lead, zinc, rare earth, gold, silver, and other metals. Because of their unique properties, nonferrous metals have been widely used in modern industrial production and are an important part of it. However, the nonferrous metallurgy industry has

become one of the seven major industrial energy consumers in China. The energy consumption of China's nonferrous metal industry is mainly concentrated in three fields: mining, smelting, and processing. China's nonferrous metal mineral resources are characterized by the following four qualities: (1) more small mines and fewer large ones; (2) more lean ore and less rich ore; (3) more cogenetic ore and less single metal ore; (4) more refractory ore and less easy-to-process ore. These properties lead directly to high energy consumption. The nonferrous metallurgy industry's power consumption accounted for 8% of the total power consumption in China annually and was 11% of the total industrial power consumption. With the continuous expansion of the overall scale of production in the nonferrous metallurgy industry, the total energy consumption is still increasing. Thus, studying an energy consumption model of the nonferrous metallurgy industry has practical significance and could help government policymakers to draw up energy plans, improve energy policy, and refine the relationship between energy supply and demand.

The energy consumption data in nonferrous metallurgy industry is characterized by the following three qualities: (1) multi-modal, (2) strong correlation, and (3) high throughput. This study focuses on the first of these. Unlike typical business datasets, in the nonferrous metallurgy industry, the datasets contain energy consumption data for a single piece of equipment that is accurate to a minute or hour, energy consumption data for a single factory that is accurate to a day or week, and energy consumption data for a single industry that is accurate to a month or year. This is determined by the nature of the industry. For instance, in the electrolytic aluminum industry, a factory worker pays attention to the production by equipment in real time. When collecting the data, the time interval of data is accurate to a minute. However, a senior manager is more focused on the business strategy and needs to see datasets that are accurate to a month or year because it is the trends in the data that are most important. Hence, the characteristic of the energy consumption data is that it has different time scales. Figure 1

shows an example of two datasets at different time scales in the nonferrous metallurgy industry. The time interval of the left figure is a year and that of the right one is a month. In the next section, we present the method proposed to handle the problem.

The Proposed Hybrid Forecasting Method: ASTA-HSVR

Hybrid Support Vector Regression (HSVR)

Support vector regression (SVR) is an important kind of machine learning method for regression problems, which was derived from the concept of a support vector machine (SVM). The difference between SVR and a SVM is that SVR has an extra ϵ -insensitive loss function. Detailed information regarding SVM or SVR can be found in the book [23]. Given a learning sample set $D = \{(x_i, y_i), i = 1, \dots, m\}$, where $x_i \in R^n$ are the input vectors, $y_i \in R^1$ are the corresponding output values, and m is the number of the samples. The goal of SVR is to find an appropriate mapping $f(x)$ between the input sample x_i and the output sample y_i . The $f(x)$ can be described as follows:

$$f(x) = w^T \varphi(x) + b, \quad (1)$$

where w and b are the parameters of the function, and φ is a nonlinear function that maps the input features to a higher feature space. The ϵ -insensitive loss function L_ϵ is defined as:

$$L_\epsilon(x_i) = \begin{cases} 0 & \text{if } |y_i - f(x_i)| \leq \epsilon \\ |y_i - f(x_i)| - \epsilon & \text{otherwise} \end{cases}, \quad (2)$$

where $\epsilon \geq 0$. The objective of this function is to measure empirical error. The parameter ϵ defines the "tube." When the two slack variables, ξ_i and ξ_i^* , are introduced to represent the distance of the y_i and y_i^* from the upper and

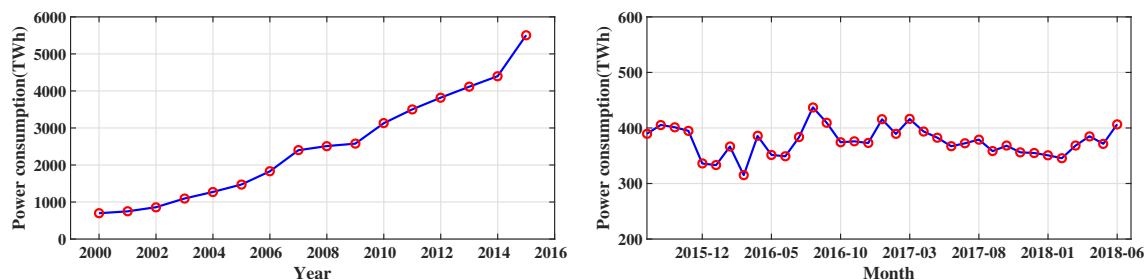


Fig. 1 Example of two datasets at different time scales

lower bound of the “tube” respectively, the SVR model is transformed to an optimization problem, which is shown as follows:

$$\begin{aligned} \min R(w, \xi, \xi^*) &= C \sum_{i=1}^m (\xi_i + \xi_i^*) + \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad &\begin{cases} y_i - w^T \varphi(x_i) - b \leq \epsilon + \xi_i \\ w^T \varphi(x_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \\ C > 0 \end{cases} \end{aligned} \quad (3)$$

The above SVR was put forward by Vapnik [23] and is called the ϵ -SVR. Later, an improved SVR model, ν -SVR, was proposed by Scholkopf et al. [17]. It used a new parameter $\nu \in (0, 1)$ to control the number of support vectors in the optimization process. Equation 3 is transformed to the following optimization problem:

$$\begin{aligned} \min R(w, \xi, \xi^*, \epsilon) &= C(\nu\epsilon + \sum_{i=1}^m (\xi_i + \xi_i^*)) + \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad &\begin{cases} y_i - w^T \varphi(x_i) - b \leq \epsilon + \xi_i \\ w^T \varphi(x_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \\ C > 0 \end{cases} \end{aligned} \quad (4)$$

Both optimization problems can be solved by introducing two Lagrangian multipliers a_i, a_i^* and a kernel function $K(x_i, x_j)$. The final regression function is as follows:

$$f(x) = \sum_{i=1}^m (a_i - a_i^*) K(x_i, x_j) + b. \quad (5)$$

The two SVR models mentioned above have been successfully applied to many applications. However, in this study, the characteristic of the energy consumption data in the nonferrous metallurgy industry is that it has different time scales. In other words, the time intervals of these data are different. For instance, the power consumption dataset for the nonferrous metallurgy industry contains yearly data from 2000 to 2015, and the power consumption dataset in the electrolytic aluminum industry consists of monthly data from January to December in different years. Hence, a hybrid support vector regression (HSVR) is presented to solve this problem. The HSVR model is shown in Fig. 2,

which contains a linear weighting of ϵ -SVR and ν -SVR. ω is a parameter to control the weights.

The performance of SVR is very sensitive to the selection of its parameters and the hybrid model is no exception. The main parameters that need to be selected are listed below: (1) the kernel function, (2) the regularization parameter C , (3) parameters of the kernel function, and (4) the tube size of the ϵ -sensitive loss function. In addition, ω in the hybrid model also needs to be optimized. Therefore, another goal of this study was to automatically determine these parameters in the hybrid model. For this purpose, the adaptive state transition algorithm is proposed to solve the above problem. This is a novel, cognitively inspired optimization method. In the next section, we describe the adaptive state transition algorithm.

Adaptive State Transition Algorithm (ASTA)

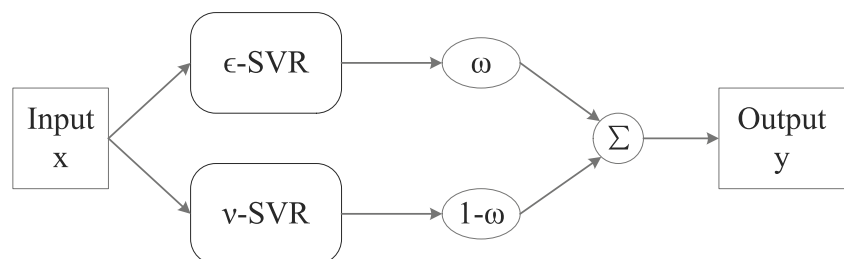
The state transition algorithm (STA) is derived from the state-space representation in control theory and establishes a unified form of solution generation based on two concepts of the theory: state and state transition [31]. In STA, each state is a solution in the optimization process and the process of updating the solutions is called state transition. Moreover, the unified form of the generation of solutions can be shown as follows:

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ y_{k+1} = f(x_{k+1}) \end{cases} \quad (6)$$

where x_k represents a state, A_k and B_k are the state transition matrices, and f is the evaluation function.

The main idea of the STA is summed up as follows: (1) produce an initial solution, which is an original state. (2) Generate a number of candidate solutions using state transformation operators according to the current state. (3) Select the current best solution as the new state according to the fitness function, which is a state transition. (4) Loop steps 2 and 3 until the termination condition stops. In addition, there are four state transformation operators in the STA, that is, rotation transformation, translation transformation, expansion transformation, and

Fig. 2 Hybrid support vector regression model



axiation transformation, of which the expressions are shown in Eq. 7, respectively.

$$\begin{cases} x_{k+1} = x_k + \alpha \frac{1}{n \|x_k\|_2} R_r x_k \\ x_{k+1} = x_k + \beta R_t \frac{x_k - x_{k-1}}{\|x_k - x_{k-1}\|_2} \\ x_{k+1} = x_k + \gamma R_e x_k \\ x_{k+1} = x_k + \delta R_a x_k \end{cases} \quad (7)$$

where R_r , R_t , R_e , and R_a are random matrices. Each operator has one parameter, and they are α , β , γ , and δ . The detailed information regarding STA can be found in [31].

The standard STA is widely used in practice, but the main disadvantage is that it may fall into a local optimum. The main reason is that the search abilities of the four search operators in the STA are limited. However, the search abilities of the operators are dependent on their parameters in the algorithm. In other words, when the parameters of the operators are different, their search abilities are quite distinct. In the STA, the two most important operators are rotation transformation and expansion transformation. The function of rotation transformation is to search the solutions in a hypersphere and α is an important parameter for controlling the radius of the hypersphere. The function of the expansion transformation is to search the solutions in a fixed range and γ is an important parameter for controlling the range. Unfortunately, in the standard STA, the value of α is divided by a fixed constant in each iteration and γ does not change. Hence, in order to search effectively and efficiently, the values of α and γ should be automatically adjusted.

In this study, an adaptive α can be automatically adjusted according to the current iteration and the number of iterations. The evolution of α is divided into three stages: (1) slow decrease, (2) rapid decrease, and (3) tendency to be stable. The expression is described as follows:

$$\alpha = \alpha \times \left(1 - \frac{0.5}{1 + |l - \frac{2}{L}|^{1.6}} \right), \quad (8)$$

where l is the current iteration and L is the number of iterations.

Meanwhile, an adaptive γ can be automatically adjusted according to the current iteration, the number of iterations, and the fitness value. When the minimum fitness in the current iteration is less than the current best fitness, γ will be decreased according to the current iteration and the number of iterations. In contrast, γ will be increased exponentially with a base e . The expression is described as follows:

$$\gamma = \begin{cases} \gamma e^{(1-\frac{l}{L})} & \text{if } f_{min} \geq f_s \\ \gamma \times (1 - \frac{l}{L+1}) & \text{if } f_{min} < f_s \end{cases} \quad (9)$$

where f_{min} represents the minimum fitness in the current iteration and f_s represents the current best fitness.

HSVR Parameter Selection Based on ASTA

In this subsection, we describe the optimization process regarding the parameter selection of the HSVR model based on the ASTA. First, because the Gaussian radial basis function (RBF) kernel function is widely used, it was selected as the kernel function in the HSVR model. Hence, there are five parameters that need to be optimized: C , σ , ϵ , ν , and ω . C is the regularization parameter and σ is a free parameter of the RBF kernel function. For ϵ -SVR, ϵ controls the width of the ϵ -insensitive zone. For ν -SVR, ν can affect the number of support vectors used to construct the regression function. ω is the weight of the HSVR model. The ASTA is used for automatic parameter tuning in the HSVR model. Thus, the whole optimization process is performed automatically, and the overall process of the ASTA-HSVR can be summed up in the following eight steps shown in Fig. 3.

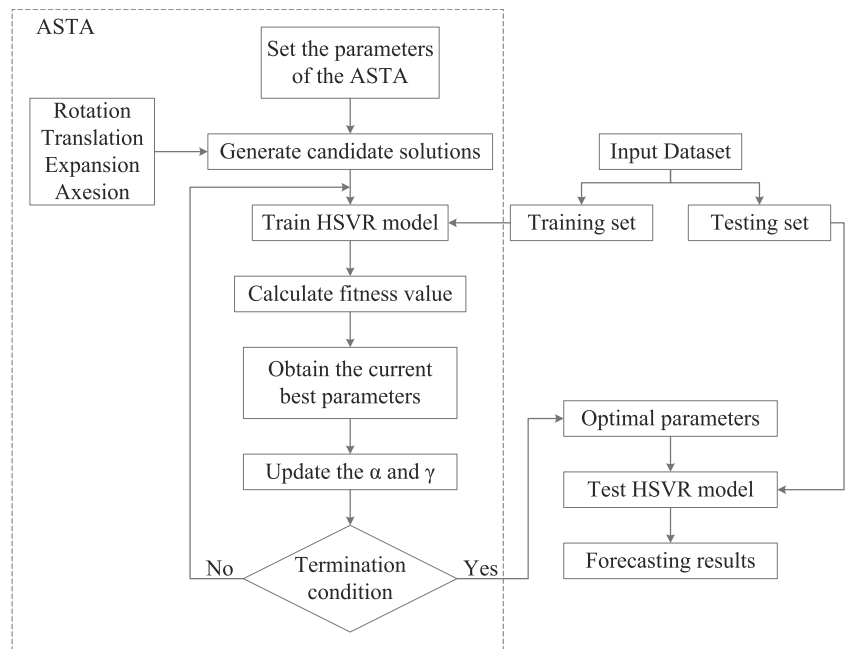
- Step 1. The input dataset is split into training and testing sets.
- Step 2. The parameters of the HSVR model, C , σ , ϵ , ν , and ω , are coded in the ASTA.
- Step 3. The parameters of the ASTA, the number of iterations, search enforcement(SE), α , β , γ , and δ are set.
- Step 4. An initial solution is generated. Each solution in the ASTA represents a choice of (C , σ , ϵ , ν , and ω).
- Step 5. At each iteration, the ASTA uses four operators in turn to generate candidate solutions. The solution that has a better fitness value with the training set survives to the next iteration. Then, α and γ are updated according to Eqs. 8 and 9.
- Step 6. The k-fold cross validation technique is employed to evaluate the fitness. The mean absolute percentage error function (MAPE) is used as the fitness function.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A(i) - F(i)}{A(i)} \right|, \quad (10)$$

where $A(i)$ is the actual value and $F(i)$ is the forecast value.

- Step 7. If the termination condition is met, the optimization process is stopped and the HSVR model has been trained completely. Otherwise, the algorithm returns to step 5.
- Step 8. The HSVR model is used to forecast energy consumption in the testing set and the forecasting results are output.

Fig. 3 Flow chart of the ASTA-HSVR



Case Study from the Nonferrous Metallurgy Industry

In this section, two cases of energy consumption from the nonferrous metallurgy industry in China are used to demonstrate the performance of the proposed method (ASTA-HSVR). Two different kind of experiments are considered: comparison with other meta-heuristic algorithms and comparison with the non-iterative algorithms. The first one is similar to a vertical comparison, which is based on the HSVR model, but with different meta-heuristics algorithms. These meta-heuristic algorithms include genetic algorithms (GA) [11], particle swarm optimization (PSO) [18], success-history-based adaptive differential evolution with linear population size reduction (LSHADE) [20], and LSHADE with a semi-parameter adaptation hybrid with covariance matrix adaptation evolutionary strategies (LSHADE-SPACMA) [14]. The second one is similar to a horizontal comparison, which is compared to the non-iterative algorithms. These non-iterative algorithms include random vector functional link (RVFL) networks, RVFL with empirical mode decomposition (EMD-RVFL), and RVFL with discrete wavelet transform and empirical mode decomposition (DWT-EMD-RVFL) [1]. To measure the forecasting accuracy experimentally, some popular statistical indicators were used, including mean error (ME), mean absolute error (MAE), mean percentage error (MPE), and mean absolute percentage error function (MAPE). The expressions used are as follows:

- $ME = \frac{1}{n} \sum_{i=1}^n (A(i) - F(i))$
- $MAE = \frac{1}{n} \sum_{i=1}^n |A(i) - F(i)|$

- $MPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{A(i) - F(i)}{A(i)} \right)$

Here, $A(i)$ is the actual value and $F(i)$ is the forecast value. Each model in the experiment was run 20 times and models

Table 1 Initial parameters of the comparative algorithms

Algorithm	Parameter
GA [11]	Size of population = 20
	The mutation probability, $p_1 = 0.01$
	The crossover probability, $p_2 = 0.4$
PSO [18]	Maximum number of iterations = 100
	Size of population = 20
	The acceleration coefficients, $c_1 = c_2 = 2$
LSHADE [20]	The inertia weight, $w = 1$
	Maximum number of iterations = 100
	The initial population size, $r^{N^{init}} = 20$
LSHADE-SPACMA [14]	Archive rate, $r^{arc} = 2.0$
	Pbest individuals rate, $p = 0.1$
	Historical memory size, $H = 5$
	The initial population size, $NP = 18 * D$
	Pbest individuals rate, $P_{best} = 0.11$
	Memory size, $H = 1.4$
ASTA	Archive rate, $arc_rate = 5$
	Probability variable, $FCP = 0.5$
	Learning rate, $c = 0.8$
	The threshold = $\max_nfes/2$
	SE = 20
	$\alpha = 1, \beta = 1, \gamma = 1, \delta = 1$
	Maximum number of iterations = 100

Table 2 Forecasting errors for the proposed model using different optimization methods for the yearly dataset

Metrics	HSVR				
	GA	PSO	LSHADE	ELSHADE-SPACMA	ASTA
ME	13.4172	16.3383	- 2.6969	- 8.3383	- 8.9161
MAE	101.7617	101.6423	69.6057	64.3412	64.0494
MPE	- 0.0043	- 0.0025	0.0025	0.0017	0.0019
MAPE	0.0443	0.0440	0.0289	0.0266	0.0265

Fig. 4 Iterative curves of the MAPE values obtained by the ASTA-HSVR for the yearly dataset

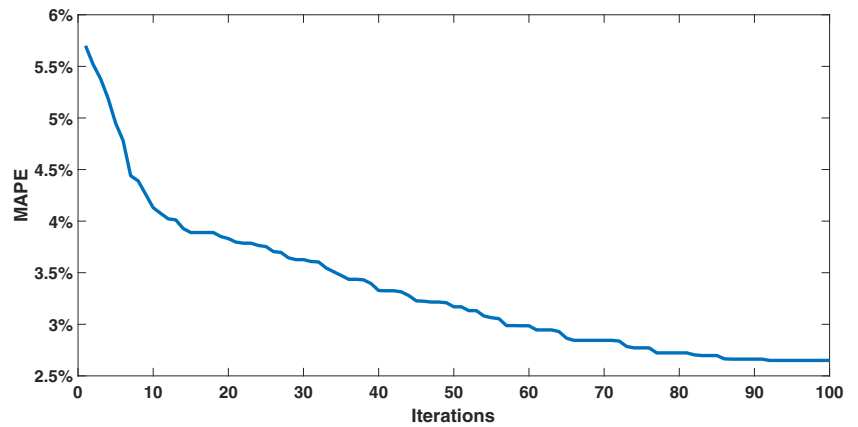


Fig. 5 Comparative curve of the actual and forecast values of the proposed hybrid model for the testing dataset

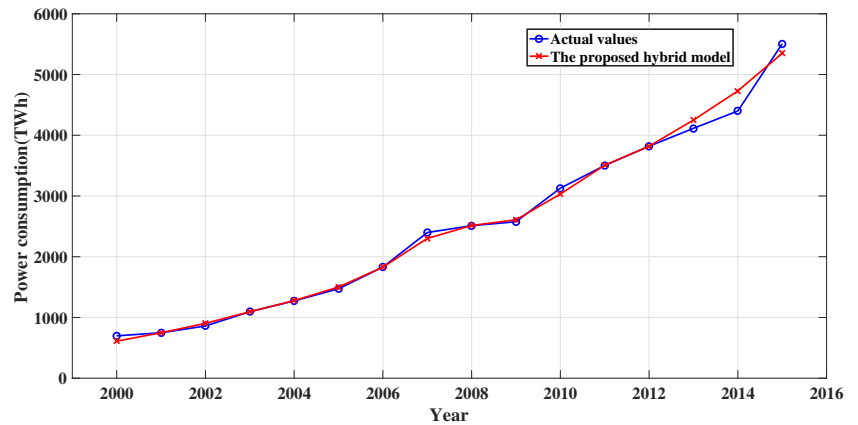
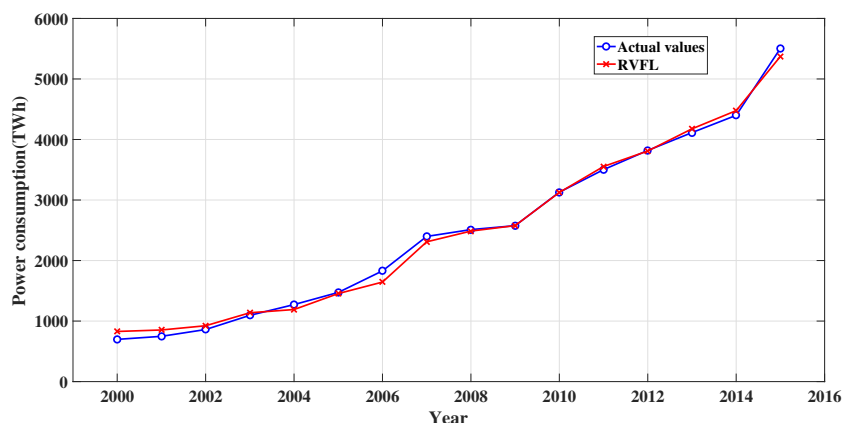


Table 3 The initial parameters of the RVFL

Algorithm	Parameter	Value
RVFL	Number of hidden neurons	10000
	Link	+
	Bias	+
	Activation function	“sig”
	Random seeds	0
	Mode	1
	Random type	Uniform
	Scale	1
	Scale mode	2

Fig. 6 Comparative curve of actual and forecast values of the RVFL for the test dataset



with the lowest MAPE among the 20 trials were selected and presented in this paper. All these experiments were implemented on a personal computer with Intel Core i7 Duo CPU at 2.8 GHz, with 16 GB RAM, using MATLAB.

Case 1: Power Consumption in Nonferrous Metallurgy Industry

In the first case, a yearly dataset about power consumption in the whole nonferrous metallurgy industry was used to evaluate the performance of the hybrid model. The power consumption dataset contains yearly data from 1978 to 2015.

Comparison with Other Meta-heuristic Algorithms

First, the proposed method is compared with the HSVR with other meta-heuristics algorithms. To guarantee that all the comparative algorithms achieved satisfactory performance, we took the suggestions of their corresponding literature sources to set the values of the related parameters, as listed in Table 1. Table 2 shows the comprehensive results of the experiment. Figure 4 presents the iterative curves of the MAPE values obtained by the ASTA-HSVR for the

experiment. Figure 5 shows the comparative curve of actual and forecast values of the proposed hybrid model for the testing dataset. We can see that the MAPE obtained by the proposed method for the testing dataset is 0.0265, which is much lower than that obtained by the hybrid model with GA or PSO, and similar to that obtained by the hybrid model with LSHADE or LSHADE-SPACMA. On the other three performance metrics, the proposed method is also better than the hybrid model with GA or PSO, and similar to the hybrid model with LSHADE or LSHADE-SPACMA. Hence, the proposed method is very effective for this case. From the results, the forecasting errors by the ASTA-HSVR method are very small. The explanation is that the HSVR model is more flexible, and the adaptive state transition algorithm improves the global search ability of the traditional state transition algorithm.

Comparison with the Non-iterative Algorithms

Next, the proposed method was compared with the non-iterative algorithms. Table 3 shows the initial parameters of the RVFL. Figures 6 and 7 show the comparative curve of actual and forecast values of the RVFL and DWT-EMD-RVFL for the test dataset. Table 4 shows the comprehensive

Fig. 7 Comparative curve of actual and forecast values of the DWT-EMD-RVFL for the test dataset

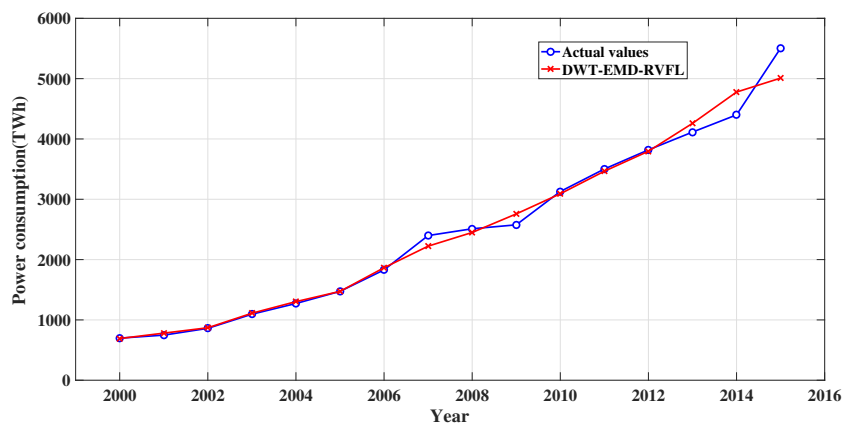


Table 4 Forecasting errors for the non-iterative algorithms and for the proposed model for the yearly dataset

Metrics	Learning models			
	RVFL	EMD-RVFL	DWT-EMD-RVFL	ASTA-HSVR
ME	0.7277	7.3620	0.4870	- 8.9161
MAE	67.4542	108.5006	104.1256	64.0494
MPE	- 0.0146	0.0048	- 0.0050	0.0019
MAPE	0.0463	0.0395	0.0331	0.0265

Table 5 Running time (in seconds) for case 1

RVFL	EMD-RVFL	DWT-EMD-RVFL	ASTA-HSVR
1.00	1.08	1.24	6.16

Table 6 Forecasting errors for the proposed model using different optimization methods for the monthly dataset

Metrics	HSVR				
	GA	PSO	LSHADE	ELSHADE-SPACMA	ASTA
ME	- 1.0622	- 1.1691	0.8895	0.9091	0.9777
MAE	16.2890	15.2524	7.6832	7.6268	7.6230
MPE	- 0.0057	- 0.0058	0.0014	0.0015	0.0016
MAPE	0.0432	0.0405	0.0204	0.0202	0.0202

Fig. 8 Iterative curves of the MAPE values obtained by the ASTA-HSVR for the monthly dataset

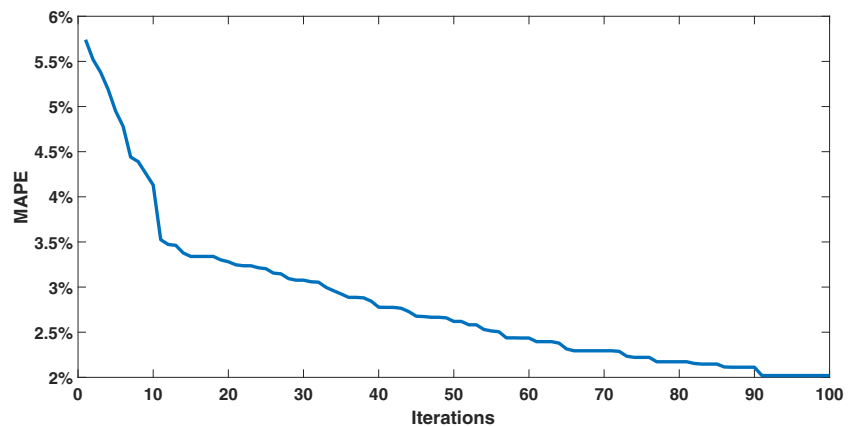


Fig. 9 Comparative curve of actual and forecast values of the proposed hybrid model for the test dataset

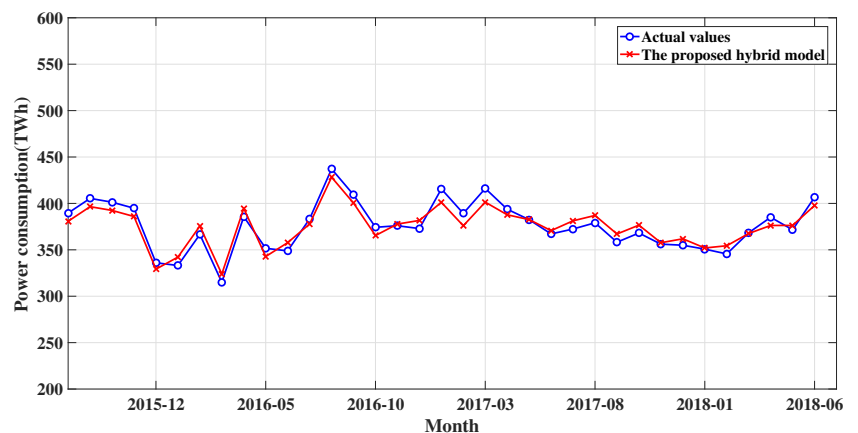


Fig. 10 Comparative curve of actual and forecast values of the RVFL for the test dataset

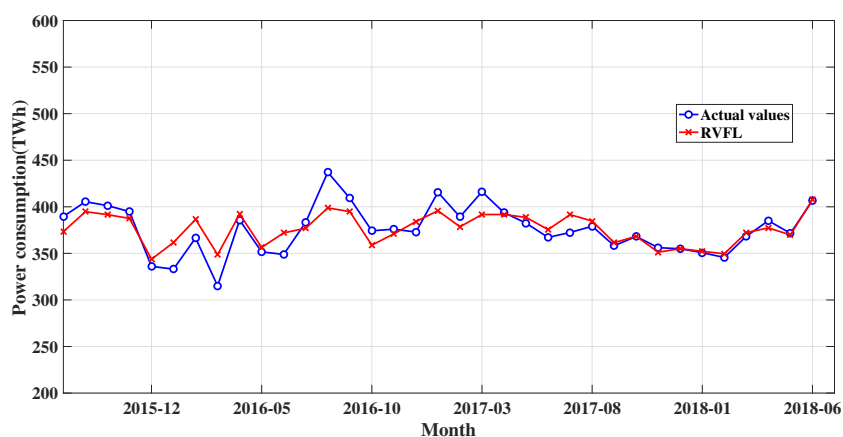


Fig. 11 Comparative curve of actual and forecast values of the DWT-EMD-RVFL for the testing dataset

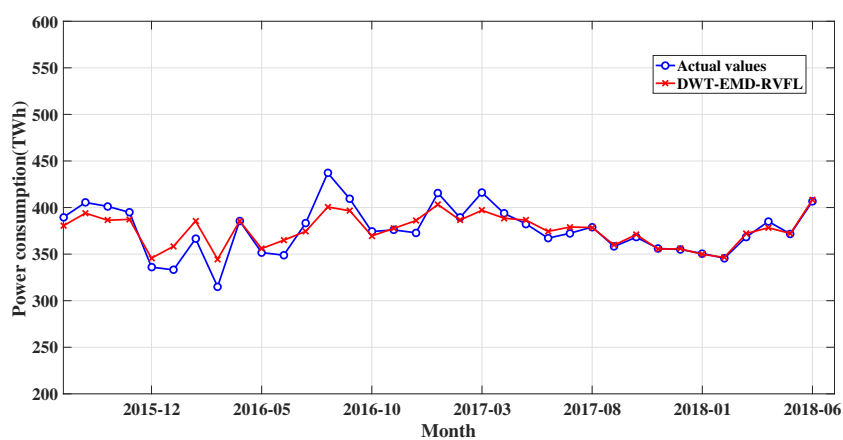


Table 7 Forecasting errors for the non-iterative algorithms and for the proposed model for the monthly dataset

Metrics	Learning models			
	RVFL	EMD-RVFL	DWT-EMD-RVFL	ASTA-HSVR
ME	0.1699	- 0.4016	0.1709	0.9777
MAE	10.9603	10.3889	8.6596	7.6230
MPE	- 0.0017	- 0.0031	- 0.0014	0.0016
MAPE	0.0292	0.0278	0.0231	0.0202

Table 8 Running time (in seconds) for Case 2

RVFL	EMD-RVFL	DWT-EMD-RVFL	ASTA-HSVR
0.94	1.10	1.32	6.98

results of the experiment. Table 5 shows the total running time of different methods. On the one hand, we can see that the ASTA-HSVR method takes more time than the non-iterative algorithms. On the other hand, although the ASTA-HSVR is worse than the non-iterative algorithms about ME, the forecasting by the ASTA-HSVR is a bit better than the non-iterative algorithms on the other three performance metrics. Hence, although the ASTA-HSVR method takes more time, the results indicate that it is still a promising method for the yearly dataset.

Case 2: Power Consumption in the Electrolytic Aluminum Industry

In the second case, a monthly dataset about power consumption in the electrolytic aluminum industry was used to evaluate the performance of the hybrid model. The power consumption dataset contains monthly data from September 2011 to June 2018.

Comparison with Other Meta-heuristic Algorithms

The proposed method was first compared with the HSVR with other meta-heuristics algorithms. The values of the parameters related to the meta-heuristic algorithms are shown in Table 1. Table 6 shows the comprehensive results of the experiment. Figure 8 presents the iterative curves of the MAPE values obtained by the ASTA-HSVR for the experiment. Figure 9 shows the comparative curve of actual and forecast values of the proposed hybrid model for the test dataset. We can see that the MAPE obtained by the proposed model for the test dataset is 0.0202, which is also much lower than that obtained by the hybrid model with GA or PSO, and similar to that obtained by the hybrid model with LSHADE or LSHADE-SPACMA. From the other three performance metrics, we can see that the conclusion is the same. Therefore, the proposed method is very effective for the monthly dataset. We can also see from the comprehensive results that the proposed method has good performance in this case.

Comparison with the Non-iterative Algorithms

Second, the proposed method was compared with the non-iterative algorithms. Table 3 shows the initial parameters of the RVFL. Figures 10 and 11 show the comparative curve of the actual and forecast values of the RVFL and DWT-EMD-RVFL for the test dataset. Table 7 shows the comprehensive results of the experiment. Table 8 shows the total running time of the different methods. As we can see from Table 7, the ranking for MAPE and MAE is ASTA-HSVR > DWT-EMD-RVFL > EMD-RVFL > RVFL, and the ranking for MPE is DWT-EMD-RVFL > ASTA-HSVR > RVFL >

EMD-RVFL. The results indicate that the ASTA-HSVR model is more promising than other models in terms of MAPE and MAE. Although the ASTA-HSVR method takes more time, it shows good performance for the monthly dataset.

Conclusions

In this study, a hybrid support vector regression with adaptive state transition algorithm (ASTA-HSVR) was developed and successfully applied for forecasting the energy consumption in the nonferrous metallurgy industry. For this method, a hybrid support vector regression (HSVR) forecasting model was presented, which consists of a linear weighting of ϵ -SVR and ν -SVR. In addition, in order to optimize the parameters of the HSVR, an adaptive state transition algorithm (ASTA) was developed based on a state transition algorithm. Two cases of energy consumption from the nonferrous metallurgy industry in China were used to demonstrate the performance of the proposed model. The results indicate that the ASTA-HSVR has good performance. However, it should be noted that outliers are not considered in this study. In the future, we expect to extend the ASTA-HSVR algorithm to include energy consumption forecasting problems with outliers.

Funding This study was funded by the National Natural Science Foundation of China (Grant Nos. 61873285, 61533020, 61751312), the 111 Project (Grant No. B17048), the Innovation-Driven Plan in Central South University (Grant No. 2018CX012), and the Hunan Provincial Natural Science Foundation of China (Grant No. 2018JJ3683).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Ethical Approval This article does not contain any experiments with human or animal participants performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Qiu X, Suganthan PN, Amaratunga GA. Ensemble incremental learning random vector functional link network for short-term electric load forecasting. *Knowl-Based Syst.* 2018;145:182–96.
2. Aljarah I, Al-Zoubi AM, Faris H, Hassonah MA, Mirjalili S, Saadeh H. Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm. *Cogn Comput.* 2018;10(3):478–95.

3. Chen Y, Tan H. Short-term prediction of electric demand in building sector via hybrid support vector regression. *Appl Energy*. 2017;204:1363–74.
4. Gao F, Huang T, Sun J, Wang J, Hussain A, Yang E. A new algorithm of SAR image target recognition based on improved deep convolutional neural network. *Cogn Comput*. 2018. <https://doi.org/10.1007/s12559-018-9563-z>.
5. Han J, Yang C, Zhou X, Gui W. Dynamic multi-objective optimization arising in iron precipitation of zinc hydrometallurgy. *Hydrometallurgy*. 2017;173:134–48.
6. Han J, Yang C, Zhou X, Gui W. A new multi-threshold image segmentation approach using state transition algorithm. *Appl Math Model*. 2017;44:588–601.
7. Huang M, Zhou X, Huang T, Yang C, Gui W. Dynamic optimization based on state transition algorithm for copper removal process. *Neural Comput & Applic*. 2017. <https://doi.org/10.1007/s00521-017-3232-0>.
8. Huang Z, Yang C, Zhou X, Gui W. A novel cognitively inspired state transition algorithm for solving the linear bi-level programming problem. *Cogn Comput*. 2018;10(5):816–26.
9. Huang Z, Yang C, Zhou X, Huang T. A hybrid feature selection method based on binary state transition algorithm and relief. *IEEE J Biomed Health Inform*. 2018. <https://doi.org/10.1109/JBHI.2018.2872811>.
10. Limanond T, Jomnonkwo S, Srikaew A. Projection of future transport energy demand of thailand. *Energy Policy*. 2011;39(5):2754–63.
11. Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew Energy*. 2014;62:592–7.
12. Ma Y, Peng H, Khan T, Cambria E, Hussain A. Sentic lstm: a hybrid network for targeted aspect-based sentiment analysis. *Cogn Comput*. 2018;10(4):639–50.
13. Meng M, Niu D. Annual electricity consumption analysis and forecasting of China based on few observations methods. *Energy Convers Manag*. 2011;52(2):953–7.
14. Mohamed AW, Hadi AA, Fattouh AM, Jambi KM. Lshade with semi-parameter adaptation hybrid with cma-es for solving cec 2017 benchmark problems. In: 2017 IEEE Congress on evolutionary computation (CEC); 2017. p. 145–152.
15. Mohandes M. Support vector machines for short-term electrical load forecasting. *Int J Energy Res*. 2002;26(4):335–45.
16. Ruiz L, Rueda R, Cuellar M, Pegalajar M. Energy consumption forecasting based on elman neural networks with evolutive optimization. *Expert Systems with Applications*. 2018;92:380–9.
17. Schölkopf B, Smola AJ, Williamson RC, Bartlett PL. New support vector algorithms. *Neural Comput*. 2000;12(5):1207–45.
18. Shrivastava NA, Khosravi A, Panigrahi BK. Prediction interval estimation of electricity prices using pso-tuned support vector machines. *IEEE Trans Ind Inf*. 2015;11(2):322–31.
19. Szoplik J. Forecasting of natural gas consumption with artificial neural networks. *Energy*. 2015;85:208–220.
20. Tanabe R, Fukunaga AS. Improving the search performance of shade using linear population size reduction. In: 2014 IEEE Congress on evolutionary computation (CEC); 2014. p. 1658–65.
21. Tanveer M. Robust and sparse linear programming twin support vector machines. *Cogn Comput*. 2015;7(1):137–49.
22. Thissen U, van Brakel R, de Weijer A, Melssen W, Buydens L. Using support vector machines for time series prediction. *Chemom Intell Lab Syst*. 2003;69(1):35–49.
23. Vapnik V. The nature of statistical learning theory. New York: Springer; 1995.
24. Xiao J, Li Y, Xie L, Liu D, Huang J. A hybrid model based on selective ensemble for energy consumption forecasting in china. *Energy*. 2018;159:534–46.
25. Xu Y, Guo R, Wang L. A twin multi-class classification support vector machine. *Cogn Comput*. 2013;5(4):580–8.
26. Yuan C, Liu S, Fang Z. Comparison of china's primary energy consumption forecasting by using arima (the autoregressive integrated moving average) model and gm(1,1) model. *Energy*. 2016;100:384–90.
27. Zhang F, Yang C, Zhou X, Gui W. Fractional-order pid controller tuning using continuous state transition algorithm. *Neural Comput & Applic*. 2018;29(10):795–804.
28. Zhou X, Gao DY, Simpson AR. Optimal design of water distribution networks by a discrete state transition algorithm. *Eng Optim*. 2016;48(4):603–28.
29. Zhou X, Gao DY, Yang C, Gui W. Discrete state transition algorithm for unconstrained integer optimization problems. *Neurocomputing*. 2016;173:864–74.
30. Zhou X, Shi P, Lim CC, Yang C, Gui W. A dynamic state transition algorithm with application to sensor network localization. *Neurocomputing*. 2018;273:237–50.
31. Zhou X, Yang C, Gui W. State transition algorithm. *Journal of Industrial and Management Optimization*. 2012;8(4):1039–56.
32. Zhou X, Yang C, Gui W. Nonlinear system identification and control using state transition algorithm. *Appl Math Comput*. 2014;226:169–79.
33. Zhou X, Yang C, Gui W. A statistical study on parameter selection of operators in continuous state transition algorithm. *IEEE Transactions on Cybernetics*. 2018. <https://doi.org/10.1109/TCYB.2018.2850350>.
34. Zhou X, Zhou J, Yang C, Gui W. Set-point tracking and multi-objective optimization-based PID control for the goethite process. *IEEE Access*. 2018;6:36683–98.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.