Research Article

Global Solar Radiation Forecasting Using Square Root Regularization-Based Ensemble

Yao Dong ^{1,2} and He Jiang ^{1,2}

¹*School of Statistics, JiangXi University of Finance and Economics, Nanchang JiangXi 330013, China*

²*Applied Statistics Research Center, Nanchang Jiangxi 330013, China*

Correspondence should be addressed to He Jiang; jiangsky2005@aliyun.com

Received 26 February 2019; Revised 30 April 2019; Accepted 2 May 2019; Published 14 May 2019

Academic Editor: Bogdan Dumitrescu

Copyright © 2019 Yao Dong and He Jiang. 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.

In recent decades, the integration of solar energy sources has gradually become the main challenge for global energy consumption. Therefore, it is essential to predict global solar radiation in an accurate and efficient way when estimating outputs of the solar system. Inaccurate predictions either cause load overestimation that results in increased cost or failure to gather adequate supplies. However, accurate forecasting is a challenging task because solar resources are intermittent and uncontrollable. To tackle this difficulty, several machine learning models have been established; however, the forecasting outcomes of these models are not sufficiently accurate. Therefore, in this study, we investigate ensemble learning with square root regularization and intelligent optimization to forecast hourly global solar radiation. The main structure of the proposed method is constructed based on ensemble learning with a random subspace (RS) method that divides the original data into several covariate subspaces. A novel covariate-selection method called square root smoothly clipped absolute deviation (SRSCAD) is proposed and is applied to each subspace with efficient extraction of relevant covariates. To combine the forecasts obtained using RS and SRSCAD, a firefly algorithm (FA) is used to estimate the weights assigned to individual forecasts. To handle the complexity of the proposed ensemble system, a simple and efficient algorithm is derived based on a thresholding rule and accelerated gradient method. To illustrate the validity and effectiveness of the proposed method, global solar radiation datasets of eight locations of Xinjiang province in China are considered. The experimental results show that the proposed RS-SRSCAD-FA achieves the best performances with a mean absolute percentage error, root-mean-square error, Theil inequality coefficient, and correlation coefficient of 0.066, 20.21 W/m², 0.016, 3.40 s, and 0.98 in site 1, respectively. For the other seven datasets, RS-SRSCAD-FA still outperforms other approaches. Finally, a nonparametric Friedman test is applied to perform statistical comparisons of results over eight datasets.

1. Introduction

An increasing number of countries are paying substantial attention to environmental issues such as global warming, climate change, and greenhouse gas emissions. A major cause of global warming is the burning of fossil fuels including coal and oil for supplying traditional electricity generation. This encourages the exploitation and utilization of renewable energy sources including solar, wind, hydro, tidal, wave, biomass, and geothermal sources as alternatives [1]. Nevertheless, solar energy, as a typical renewable energy source, has stochastic, intermittence, time-varying, and uncertainty properties, which may result in defects in the stability and reliability of power grid systems [2–4]. This introduces new

challenges for the incorporation of solar energy sources into the power grids. In addition, solar radiation can be measured based on sensors including pyrheliometers, pyranometers, and radiometers combined with data-acquisition hardware and software. However, it is time-consuming, cumbersome, and expensive to install sensors across the world [5].

To tackle these challenges, it is necessary to develop accurate global solar radiation forecasting models. A wide variety of models have been established using machine learning techniques. Machine learning, which is classified as an artificial intelligence technology, is a subfield of computer science. It can be applied to numerous domains, and its characteristic is that the model can identify relations between inputs and outputs notwithstanding whether the

representation is infeasible. This advantage promotes the application of machine learning models in classification, pattern recognition, spam filtering, data mining, and forecasting [6]. At present, a popular method for forecasting global solar radiation is to use an artificial neural network (ANN). ANNs have been used in various areas with different input covariates (multiple meteorological, geographical, and astronomical parameters such as temperature, humidity, wind speed, precipitation, altitude, declination, and zenith angle) and multiple time scales (hourly, daily, and monthly) [7, 8]. The main advantage of ANNs is that they do not require as many adjustable parameters as other classical techniques and provide higher forecasting accuracy. Ozgoren et al. [9] proposed an ANN model based on a multinomial regression (MNLE) algorithm to forecast the monthly mean daily sum global solar radiation for Turkey. Hejase et al. used a multilayer perceptron and radial basis function (RBF) techniques with comprehensive training architectures and different combinations of inputs to forecast global horizontal irradiance for three major cities in the United Arab Emirates [10]. Renno et al. [11] developed two ANNs to forecast daily global radiation and hourly direct normal irradiance for the University of Salerno. Notwithstanding its growing popularity, ANNs have their drawbacks. For example, the backpropagation neural network (BPNN) exhibits difficulty in training the input data owing to the iterative tuning of parameters required for hidden neurons because it does not always produce a unique global solution with different models; moreover, BPNN exhibits a slow response owing to its gradient-based learning algorithm and requires a longer training time [12, 13].

It is well known that the support vector machine (SVM) developed by Vapnik [14] is a powerful tool for exploring appropriate SVM regressors to establish a prediction model with sufficiently high accuracy. The main goal of SVM regressors is to construct the optimum prediction model that boosts the prediction accuracy efficiently. It is a regularization network and exhibits an advantage over ANN model. To reduce the high computation cost, least-squares SVM (LSSVM) uses a least-squares method via a set of linear equations based on structural risk minimization; therefore, it is able to prevent overfitting of the training data and does not require iterative tuning of model parameters [15–17]. Wu and Liu investigated an SVM model to estimate monthly mean daily solar radiation using measured minimum, mean, and maximum air temperatures in 24 sites in China. Five SVMs based on different input covariates were developed [18]. Chen et al. used seven SVMs and five empirical sunshine-based models to forecast daily global solar radiation at three sites in Liaoning, China. The experimental results demonstrated that all SVMs exhibited higher performance than empirical models [19]. Ekici proposed a LSSVM-based intelligent model to forecast the daily solar insolation for Turkey [20]. It was noted that the temporal behavior of global solar radiation and its corresponding atmospheric input covariates are complex and chaotic [21]. Although SVM is flexible in handling nonlinear input covariates, it encounters fluctuations and difficulties in training input datasets with wide frequencies [22]. Some researchers have focused on Gaussian process regression

(GPR) methodology, which shows promising results with lower computational costs. Rohani et al. [23] presented GPR with a K -fold cross-validation (CV) model to forecast daily and monthly solar radiation at Mashhad city in Iran based on several meteorological parameters. Considering the effect of climatic factors such as sunshine duration, relative humidity, and air temperature, Guermoui et al. [24] used the GPR to predict the daily global solar radiation in the Saharan climate. Guermoui et al. [25] proposed a novel method called weighted Gaussian process regression (WGPR) to predict daily global and direct horizontal solar radiation in Saharan climate. The experimental results illustrate that these developed models exhibit good forecasting performance.

It is challenging for a single forecasting model to provide an optimum solution to a complex problem. Each model has advantages and drawbacks. Thus, it is worthwhile to investigate appropriate combinations of their advantages for implementing the forecasting task. As solar radiation simultaneously exhibits nonstationary and nonlinear characteristics at different time scales, the analysis of these fluctuations at all scales can be performed using multi-scale decomposition methods. Wang et al. [26] decomposed solar radiation into a quasi-stationary process at different frequency scales using wavelet transform (WT) and employed BPNN to predict solar radiation based on temperature, clearness index, and phase space reconstruction radiation. Monjoly et al. [27] applied three multiscale decomposition methods, empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and WT, to analyze the nonstationary and nonlinear features of the clear sky index, and then developed a hybrid model AR-NN combined autoregressive process (AR) and neural network (NN) to forecast global horizontal solar irradiation (GHI). The experimental results showed that the WT-hybrid model has the best forecasting accuracy. To improve the forecasting performance, Hussain and Alalili [28] investigated a new hybrid technique based on WT and four different architectures of ANN to forecast GHI in the Abu Dhabi region.

Ensemble methods are straightforward and highly effective techniques that average the approximate predictions of a few weak learners to present accurate estimates rather than find a single remarkable sophisticated learner [29–31]. The advent of ensemble methods has gained prominence. The most commonly used ensemble methods include random forest, gradient boosting, and bagging. Although these methods have wide applications in other scientific fields, they have only rarely been adopted in forecasting global solar radiation [6]. Gala et al. [32] used support vector regression, random forest regression, and gradient boosted regression as a hybrid model to improve 3-hour accumulated radiation forecasts for seven sites in Spain. Sun et al. [33] proposed different random forest models with and without considering the air pollution index to estimate solar radiation in three sites in China through comparisons with a few empirical models; their experimental results demonstrated that the performance of random forest techniques with the air pollution index is higher. Aler et al. [34] applied gradient boosting machine learning to improve the performance of diffuse and direct separation models. Ensembles of linear and nonlinear weak learners have both

been adopted. The results indicated that the gradient boosting method can reduce large random errors of separation models efficiently.

The computational cost is mainly determined by the complexity of the forecasting model. It is time-consuming if several covariates are employed to establish the model. Thus, it is essential to investigate an efficient method to simplify the model structure. Covariate selection is an efficient method to extract the important covariates. There are several penalized covariate selection methods including least absolute shrinkage and selection operator (LASSO) [35], least angle regression [36], elastic net [37], and adaptive LASSO [38]. LASSO is a popular covariate selection method applying an ℓ_1 -type penalty. Yang et al. [39] selected the important time-lagged covariate using LASSO and revealed its advantages over the exponential smoothing (ES) model and the autoregressive integrated moving average (ARIMA) model. LARS is closely associated with LASSO, and their connection is established in Efron et al. [36]. Zou proposed adaptive LASSO involving the addition of weights to the penalty parameter of each feature [38]. It outperforms LASSO in terms of selection consistency. The elastic net was first proposed by Zou, and it considers ℓ_2 -type penalty in addition to the ℓ_1 -type penalty [37]. Thus, the elastic net applied $\ell_1 + \ell_2$ -type penalty, which provides the following advantages: the ℓ_1 part focuses on feature selection and the ℓ_2 part can be applied to boost forecasting accuracy. However, Zhang determined that the abovementioned feature selection methods apply convex penalties, which makes it difficult to detect the important covariates when the model coherence is high [40]. To solve this problem, the nonconvex penalized covariate selection approaches are advocated. Knight and Fu proposed bridge regression, which focuses on the penalty function of the ℓ_r ($0 < r < 2$) form [41]. Fan and Li [42] proposed the smoothly clipped absolute deviation (SCAD) method and demonstrated its advantage over LASSO. However, to the best of the present authors' knowledge, there is no research work on feature selection using both square root loss function and SCAD penalty in the literature. The main contribution of this study is the advocacy of a novel global solar radiation forecasting approach that combines ensemble learning, square root smoothly clipped absolute deviation (SRSCAD), and the firefly algorithm (FA). Specifically, the contributions of this study are given as follows:

- (i) An SRSCAD method is proposed to extract the important covariates efficiently. Moreover, the optimal regularization parameter is selected by a 10-fold CV; this CV completely utilizes the data through resampling, which is particularly useful for data with a small sample size.
- (ii) The ensemble learning is the main structure of the proposed method. The diversity created by covariate selection of the ensemble system substantially aids in boosting the forecasting accuracy. The FA, which is demonstrated to obtain the global optimal solution successfully, is applied as a weight policy in the forecasting model.
- (iii) A simple-to-implement and efficient algorithm is designed based on thresholding rules and an accelerated gradient method is used to further speed up the convergence of the algorithm.

The rest of this paper is organized as follows. Section 2 presents a preliminary discussion of related methodologies. Section 3 shows the proposed method and the design of the algorithm used in the proposed method is also revealed. The empirical study and corresponding results analysis are presented in Section 4. Finally, Section 5 summarizes this study and provides concluding remarks. The nomenclature is provided in Table 1.

2. Methodology and Materials

2.1. Forecasting and Covariate Selection in the Interactive Model. Several nonlinear models including SVMs and ANNs have been considered to forecast global solar radiation because researchers have gradually realized that the outcomes given by a linear regression model are not accurate. Interactive models are one of nonlinear models and have wide applications in many scientific fields. They are favored because they take the connections between covariates into account to capture the nonlinear relationship between the global solar radiation and the other meteorological and geographical covariates. In this study, an interactive model is considered to build the forecasting model in a global solar radiation problem and study the following nonlinear additive setup:

$$y = \beta_0 + \sum_j \beta_j x_j + \sum_{j,k} \theta_{jk} x_j \odot x_k + \varepsilon, \quad (1)$$

where y denotes the global solar radiation to be studied, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_p] \in \mathbb{R}^{n \times p}$ is the data matrix with n samples and p meteorological and geographical covariates, ε represents the noise term, which follows a Gaussian distribution $N(\mathbf{0}, \sigma^2 \mathbf{I})$ with σ^2 representing the noise level, and let $\mathbf{I}$ be an identity matrix. The distribution of ε can also be a binomial distribution if a classification problem is studied. Use $\mathbf{x}_j \odot \mathbf{x}_k$ to denote the two-way covariates. Here β_0 is the intercept term, and $\{\beta_j\}_{j=1}^p$ and $\{\theta_{jk}\}_{1 \leq j, k \leq p^2}$ are the coefficients of covariates to be estimated. Note that several covariates are included in this model. Furthermore, the large computation time is one challenge caused by the excessive number of covariates.

In the global solar radiation forecasting domain, a proper metric that measures the fitness of global solar radiation and other meteorological and geographical features needs to be defined. This metric is called a loss function, denoted by L , in the machine learning community. Because ε follows a Gaussian distribution in the interactive model (1), a simple way to define this metric is to consider the negative loglikelihood of the distribution of y (after centering), that is, $f(y | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta})$

$$L := \log(f(y | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\theta}))$$

$$= \frac{1}{2\sigma^2} \left\| y - \sum_j \beta_j x_j - \sum_{j,k} \theta_{jk} x_j \odot x_k \right\|_2^2, \quad (2)$$

TABLE 1: Nomenclature.

Abbreviations	
AGM	Accelerated gradient method
AIC	Akaike information criterion
BPNN	Back propagation neural network
CC (seconds)	Computation cost
CS	Cuckoo search
CV	Cross-validation
ENN	Elman neural network
FA	Firefly algorithm
GCV	Generalized cross-validation
LASSO	Least absolute shrinkage and selection operator and selection operator
LOOCV	Leave-one-out cross-validation
LSSVM	Least-squares support vector machine
MAPE	Mean absolute percentage error
PSO	Particle swarm optimization
QRBF	Quadratic radial basis function
RS	Random subspace
RMSE (W/m ²)	Root-mean-square error
SCAD	Smoothly clipped absolute deviation
SRPQVSP	Square root progressive quantile variable screening procedure
SVM	Support vector machine
SRL	Square root LASSO
SRSCAD	Square root smoothly clipped absolute deviation
TIC	Theil inequality coefficient
Notation	
$\mathbf{X}$	Design matrix
$\mathbf{y}$	Response variable
β_0	Intercept term
β_j	Coefficient for main effects
θ_{jk}	Coefficients for covariates
ε	Error term
$\mathcal{S}$	Threshold rule
η	Grid values
G	Cumulative distribution function of Gaussian distribution
N	Population size
γ	Light absorption coefficient
λ	Regularization parameter
τ	Step size
M	Maximum number of iterations
K	Number of folds in CV
ℓ	Location latitude
D_n (days)	Number of days in the year
Sd (hours)	Possible maximum monthly mean of daily sunshine duration
N_h	Number of hidden neurons
B	Number of blocks generated by RS
b	Number of features in each block
g_0	Attractiveness
r	Cartesian distance

where $\|\cdot\|_2$ represents the Euclidean norm. A common procedure is to restrict the number of covariates to avoid the overfitting problem. Namely, a few of the covariates may not be useful to the model and they are considered to be nuisance

covariates. Therefore, how to extract the important covariates that make great contributions to the model is worth studying. This problem is well known as covariate selection. Covariate selection is essential because the dimensions are high in the interactive model. LASSO is a famous covariate selection method that minimizes the square error loss function and ℓ_1 -type penalty to enforce sparsity into the model. It has been shown to be accurate and can be seen as a Bayesian covariate selection using a Laplace prior distribution [43, 44]. However, LASSO often selects nuisance variables when the model coherence is high. Namely, if two covariates play a similar role in global solar radiation forecasting, it is not easy to distinguish between them. To overcome the disadvantages of LASSO, a nonconvex penalty SCAD is proposed and is shown to have better statistical properties (unbiased, continuity, and sparsity) than LASSO [42]. SCAD considers the following optimization:

$$\min_{\beta_j, \theta_{jk}} \left\{ L + \sum_j \mathcal{P}_{\text{SCAD}}(\beta_j; \lambda) + \sum_{j,k} \mathcal{P}_{\text{SCAD}}(\theta_{jk}; \lambda) \right\} \quad (3)$$

with the SCAD penalty expressed by

$$\mathcal{P}_{\text{SCAD}}(t; \lambda) = \int_0^{|t|} \left\{ \lambda 1_{\{z \leq \lambda\}} + \frac{a\lambda - z}{a-1} 1_{\{z > \lambda\}} \right\} dz \quad (4)$$

for $a = 3.7$, which is determined using generalized cross-validation.

To further increase the forecasting accuracy and facilitate the parameter tuning work, square root regularization approaches are used in this paper. Square root regularization has attracted the interest of researchers in the statistics and machine learning community for years because it makes the parameter tuning work easier. Specifically, in LASSO or SCAD problems, the tuning parameter λ depends on the value of the noise level σ , i.e., $\lambda \sim \mathcal{O}(\sigma\sqrt{2(\log p)})$. However, the noise level σ is difficult to estimate. To avoid this problem, Belloni et al. (2009) proposed square root LASSO (SRL), which applied a square root loss function $\sqrt{L}$ and considered the following optimization problem:

$$\min_{\beta_j, \theta_{jk}} \left\{ \sqrt{L} + \lambda \left(\sum_j |\beta_j| + \sum_{j,k} |\theta_{jk}| \right) \right\}. \quad (5)$$

Different than LASSO, the regularization parameter λ in SRL depends on $\sqrt{2(\log p)}$, which is independent of the noise level σ . As is known, parameter tuning is highly significant in determining the forecasting accuracy. Therefore, square root regularization may obtain accurate outcomes. To this end, this paper studies a novel covariate selection method that borrows strengths from both the square root loss function $\sqrt{L}$ and the SCAD penalty, which is given as follows:

$$\min_{\beta_j, \theta_{jk}} \left\{ \sqrt{L} + \sum_j \mathcal{P}_{\text{SCAD}}(\beta_j; \lambda) + \sum_{j,k} \mathcal{P}_{\text{SCAD}}(\theta_{jk}; \lambda) \right\}. \quad (6)$$

In the following, we focus on (6).

2.2. Ensemble Learning. Famous statistician George E.P. Box said “all the models are wrong but some of them are useful” [45]. This indicates that a single forecasting model cannot always perform very well under all circumstances. The basic idea of ensemble learning is to create a finite number of individual learners and combine them through assigning different weights and summing them up [46]. Ensemble learning is a supervised learning algorithm because it can be used to make predictions and classifications. A single model that is going to be trained can be seen as one hypothesis. However, it is difficult to decide which hypothesis is the best. Thus, the ensemble learning system combines all the hypotheses and creates a better hypothesis although the computational time is increased. Therefore, ensemble learning can represent more flexible functions than a single model. Note that an overfitting problem can be caused by the ensemble techniques. In practice, regularization methods are applied to reduce the problem related to overfitting of the training data.

Empirically, ensemble learning tends to provide better results than a single model using the significant diversity among the models. Therefore, many ensemble methods seek to promote diversity. Specifically, the diversity can be generated from the sample space (boosting and bagging), covariate space (RS [47] and random forest [48]), and parameter space. To gain better performance, the method used to combine single models is very important. The simplest way is just to calculate the average of all the results given by the single models. However, these single models may play different roles in the ensemble system. Thus, linear and nonlinear weighting strategies are applied to evaluate their contributions. In this paper, the FA, which is introduced in the following section, is used as a weighting strategy for single models.

2.3. Firefly Algorithm. Xin-She Yang proposed a nature-inspired metaheuristic in 2008 called FA, which was derived from the flashing behavior of fireflies [49]. The main purpose of the FA is to use a signal flash to attract other fireflies. There are three hypotheses in this algorithm: (i) no consideration of their sex, each firefly will move towards all other fireflies; (ii) attractiveness is associated with brightness, the fireflies with low light intensities will be attracted by the brighter ones in the vicinity; nevertheless, the brightness will increase as their distance decreases; (iii) if no fireflies are brighter than the given one, it will move randomly [50, 51].

Let F_i be the i th firefly in the population, where each firefly represents a candidate solution in the search space and N is the population size. Fireflies move towards the optimal solution positions. The attractiveness is related to the light intensity; thus we can define the attractiveness between two fireflies F_i and F_j as follows:

$$g(r_{ij}) = g_0 e^{-\gamma r_{ij}^2}, \quad (7)$$

and

$$r_{ij} = \|F_i - F_j\|_2 = \sqrt{\sum_{d=1}^D (f_{id} - f_{jd})^2}, \quad (8)$$

for $d = 1, \dots, D$, where $\|\cdot\|_2$ denotes the Euclidean norm, D is the number of dimensions, r_{ij} is the Cartesian distance between F_i and F_j , and f_{id} and f_{jd} are the d th components of F_i and F_j , respectively. At the distance $r = 0$, the parameter g_0 represents the attractiveness and γ denotes the light absorption coefficient. According to [52], γ is used as a typical initial value, $\gamma = 1/\Gamma^2$, where Γ is the length scale of the optimization problem.

The movement of the firefly F_i towards another brighter firefly F_j can be given by

$$f_{id}(t+1) = f_{id}(t) + g_0 e^{-\gamma r_{ij}^2} [f_{jd}(t) - f_{id}(t)] + \alpha \operatorname{sgn}\left(\operatorname{rand} - \frac{1}{2}\right) \oplus \text{Lévy}, \quad (9)$$

$$\text{Lévy} : u = t^{-a} \quad (1 < a \leq 3), \quad (10)$$

where $\alpha \in [0, 1]$ is the step size of the random walk and $\operatorname{sgn}(\operatorname{rand} - 1/2)$ ($\operatorname{rand} \in [0, 1]$) is a random sign whereas the random step length is obtained from a Lévy distribution in (10). In this paper, the minimization problems are considered. Let V be the fitness function. If $V(F_i) < V(F_j)$, this means that F_i is brighter than the firefly F_j . The main steps of the FA are described as follows:

- (i) Step 1. Initialize the positions of fireflies (or N solutions) $F_i = [f_{i1}, f_{i2}, \dots, f_{id}]$ randomly based on the following:

$$f_{id} = \operatorname{low} + \operatorname{rand}(0, 1)(\operatorname{up} - \operatorname{low}) \quad (i = 1, 2, \dots, N, \quad d = 1, 2, \dots, D) \quad (11)$$

where low and up denote the lower and upper bounds for the dimension D , respectively. The fitness function can be used to evaluate the position of these fireflies in the initial population.

- (ii) Step 2. Compare the positions of F_i and F_j ($i \neq j$); if $V(F_i) < V(F_j)$, F_i will be attracted by F_j and updates its position based on (9). Then the updated position of each firefly can also be evaluated.
- (iii) Step 3. If the iteration process satisfies the stopping criteria, the algorithm stops; otherwise, continue with step 2.

3. The RS-SRSCAD-FA Method

In this paper, a novel global solar radiation forecasting method is invented based on RS, SRSCAD, and the FA and is called RS-SRSCAD-FA for short. The main procedure of RS-SRSCAD-FA is described as follows:

- (i) Step 1: Build the interactive model as described in (1).
- (ii) Step 2: The RS method is applied to the interactive model. In particular, randomly select b covariates without replacement for B times, which results in B blocks denoted by $\{G_1, \dots, G_B\}$ with L covariates in each block.

FIGURE 1: Flowchart of the RS-SRSCAD-FA model.

(iii) Step 3: SRSCAD covariate selection method is used in each block $\{G_i\}_{i=1}^B$.

(iv) Step 4: Assign weights to B blocks and concatenate the estimates using the weights determined by FA.

Remark. In step 1, an interactive model is established so that all the two-way covariates $\mathbf{x}_j \odot \mathbf{x}_k$ that describe the possible relationships between covariates $\mathbf{x}_j$ and $\mathbf{x}_k$ are considered. Note that the dimension of the model has increased from p to $p^2 + p$. In step 2, similar to random forest, the interactive model is divided into a number of blocks using the RS method, which randomly extracts a number of covariates. This procedure is the same as the divide procedure in the “divide and conquer” strategy of the ensemble learning. The complex model is divided into some simple models that are easy to handle separately. After this procedure, all the solutions are combined. In step 3, the proposed dimension reduction method SRSCAD is used to reduce the number of covariates. This step is essential because it can reduce the number of covariates effectively. This delivers two benefits: (1) the computation issue is solved because only

a few covariates are included; (2) the collinearity between covariates is decreased. In step 4, the SRSCAD estimates obtained in step 3 are concatenated using the FA instead of linear combination. This procedure aggregates the diversity generated by steps 2 and 3. The flowchart of the proposed method is shown in Figure 1.

To solve the penalized optimization problems efficiently, the thresholding rule (generally referred to as the $\mathcal{S}$ function [53]), rather than the penalty function, is applied as the main tool throughout this study. The threshold function is defined rigorously as follows.

Definition (threshold rule). A threshold rule is a real-valued function $\mathcal{S}(t; \lambda)$ defined for $-\infty < t < \infty$ and $0 \leq \lambda < \infty$ such that

- (1) $\mathcal{S}(-t; \lambda) = -\mathcal{S}(t; \lambda)$;
- (2) $\mathcal{S}(t; \lambda) \leq \mathcal{S}(t'; \lambda)$ for $t \leq t'$;
- (3) $\lim_{t \rightarrow +\infty} \mathcal{S}(t; \lambda) = \infty$; and
- (4) $0 \leq \mathcal{S}(t; \lambda) \leq t$ for $0 \leq t \leq \infty$.

Inputs: $\mathbf{Z} = (\mathbf{X}, \mathbf{y}) = \{(\tilde{\mathbf{x}}_1, \mathbf{y}_1), \dots, (\tilde{\mathbf{x}}_N, \mathbf{y}_N)\}$: $\tilde{\mathbf{x}}_k \in \mathbb{R}^p$, the dataset
 M : the number of blocks generated by Random Splitting
 m : maximum number of iterations in the SRSCAD algorithm
 tol : the error tolerance used in the SRSCAD algorithm
 K : number of folds in cross-validation
Output: The ensemble test error

- 1: Randomly divide the original data $\mathbf{Z}$ into the training dataset $\mathbf{T}$ and test dataset $\mathbf{F}$.
- 2: Establish interactive models $\mathbf{QT}$ and $\mathbf{QF}$, respectively
- 3: Randomly divide $\mathbf{TIT}$ into B blocks $\{\mathcal{D}_i\}_{i=1}^B$ with b features in each model
- 4: Initialize FA parameters including the maximum number of iterations $Itermax$
- 5: **for** $i = 1$ to B **do**
- 6: Divide the training data $\mathcal{D}_i$ into K folds
- 7: **for** $j = 1$ to K
- 8: Use the j th fold as subset data $\mathcal{D}_{i,j}^{train} = (\mathbf{X}_{i,j}^{train}, \mathbf{y}_{i,j}^{train})$ and the remaining folds are regarded as subset data $\mathcal{D}_{i,j}^{test}$
- 9: Generate grid values of $\lambda^G = \{\lambda_u\}_{u=1}^s$
- 10: **for** $u = 1$ to s
- 11: Initialization: $t \leftarrow 0, \Omega^{(t)} \leftarrow \mathbf{0}$
- 12: Scaling: $\mathbf{X}_{i,j}^{train} \leftarrow \mathbf{X}_{i,j}^{train} / \tau, \mathbf{y}_{i,j}^{train} \leftarrow \mathbf{y}_{i,j}^{train} / \tau$
- 13: **while** $\|\Omega^{(t+1)} - \Omega^{(t)}\| < tol$ or $j > m$ **do**
- 14: Step 1. $\xi^{(t)} \leftarrow \Omega^{(t)} + ((\omega^{(t-1)} - 1) / \omega^{(t)}) (\Omega^{(t)} - \Omega^{(t-1)})$
- 15: Step 2. $\zeta^{(t)} \leftarrow \xi^{(t)} + \mathbf{X}_{i,j}^{trainT} (\mathbf{y}_{i,j}^{train} - \mathbf{X}_{i,j}^{train} \xi^{(t)})$
- 16: Step 3. $\Omega^{(t+1)} \leftarrow \mathcal{S}_{SCAD}(\zeta^{(t)}; \lambda_u \|\mathbf{y} - \mathbf{X} \Omega^{(t)}\|_2 / \tau_2^2)$
- 17: Step 4. $\omega^{(t+1)} = (1 + \sqrt{1 + 4\omega^{(t)2}}) / 2$
- 18: **end while**
- 19: **end for**
- 20: Obtain the solution path $\mathcal{B} = \{b_u\}_{u=1}^s$ and the corresponding sparsity patterns $\mathcal{G} = \{g_u\}_{u=1}^s$.
- 21: **end for**
- 22: Calculate CV errors using $\mathcal{B}$ and $\mathcal{G}$. Determine the optimal tuning parameters λ_{opt} w.r.t. obtain the optimal SRSCAD estimator $\hat{\Omega}_{SRSCAD-\mathcal{D}_i}$
- 23: **endfor**
- 24: Concatenate the optimal SRSCAD estimates $\{\hat{\Omega}_{SRSCAD-\mathcal{D}_i}\}_{i=1}^K$ using $\mathbf{w}$ and calculate the fitness using the fitness function
- 25: Obtain the optimal weight vector by the FA algorithm, and concatenate the optimal SRSCAD estimates using $\mathbf{w}_{opt}$
- 26: Calculate the forecasting error using the test dataset $\mathbf{QF}$
- 27: Output the ensemble test error

ALGORITHM 1: The RS-SRSCAD-FA algorithm.

From the definition, it is evident that $\mathcal{S}(\cdot; \lambda)$ is an odd monotone unbounded shrinkage rule for t , at any λ . The LASSO and SCAD rules are defined as follows:

$$\mathcal{S}_{SOFT}(t; \lambda) = \text{sgn}(t) (|t| - \lambda)_+; \quad (12)$$

$$\mathcal{S}_{SCAD}(t; \lambda) = \begin{cases} \lambda & \text{if } t \leq \lambda \\ \frac{(3.7\lambda - t)}{(2.7)} & \text{if } \lambda \leq t \leq a\lambda; \\ 0 & \text{if } t \geq a\lambda \end{cases} \quad (13)$$

and sgn is the sign function. The vector versions of $\mathcal{S}_{SOFT}(t; \lambda)$ and $\mathcal{S}_{SCAD}(t; \lambda)$ are denoted by $\mathcal{S}_{SOFT}(\mathbf{t}'; \lambda)$ and $\mathcal{S}_{SCAD}(\mathbf{t}'; \lambda)$ for any vector $\mathbf{t}'$.

To describe the proposed RS-SRSCAD-FA algorithm in a simple way, we first define a matrix Ω with its j th column given by $\Omega_j = [\beta_j, \theta_{j1}, \dots, \theta_{jp}]$, which indicates the j th covariate and all of its associated two-way covariates. The

details of the proposed algorithm are given in Algorithm 1 after initialization with $\omega^{(-1)} = 0$, $\omega^{(0)} = 1$ and $\Omega^{(0)} = \Omega^{(-1)}$. When designing the algorithm, the following points should be noted:

- (i) *Data splitting scheme.* A total of 75% of the original data are going to be applied as training data and the remaining part (25%) is considered to be the test data. The training data is applied to train a forecasting model and the test data is used to evaluate the model performances.
- (ii) *Convergence.*

Step size. The step size τ is used to ensure the convergence of the SRSCAD algorithm (cf. lines 11–19 in Algorithm 1). Usually, it can be determined based on a theoretical analysis with the surrogate function defined. Another way to

determine the step size is to apply a line search method.

Stopping criteria. The error tolerance between two successive iterates and maximum iteration number are determined by trial and error. In particular, the error tolerance is chosen from the set $\{1e-4, 1e-5, 1e-6\}$ and the maximum number of iterations is selected from $\{100, 500, 1000\}$. The optimal combination of the error tolerance and maximum iteration number is determined using the trade-off between forecasting accuracy and computational efficiency.

- (iii) *Acceleration.* Accelerated gradient method (AGM) is applied to increase the convergence speed. The design of AGM was first proposed by [54] and the computation complexity can be reduced from $\mathcal{O}(1/M)$ to $\mathcal{O}(1/M^2)$, where M represents the number of iterations of the algorithm.
- (iv) *Parameter tuning.* In a regularization-based problem, the choice of regularization parameter is critical. In this work, the regularization parameters are selected based on 10-fold CV. The main procedure can be described as follows: the data is divided into K parts with roughly equal sizes. The $K-1$ parts will be used to train the model and the left one part is applied to calculate the test error. CV repeats this procedure for K times for each candidate regularization parameter and the final CV error is the average of K test errors. The optimal parameter is selected based on the smallest CV error. When $K = n$, this procedure is known as leave-one-out cross-validation (LOOCV), which is often used in complex problems.

4. Forecasting Global Solar Radiation in Xinjiang Province of China

In this section, to demonstrate the forecasting accuracy of the proposed RS-SRSCAD-FA algorithm, comparisons are made between this and other forecasting methods including two benchmark models, the Elman neural network (ENN) and LSSVM, the Angstrom–Prescott empirical model (Angstrom–Prescott) [55], popular feature selection models including LASSO, SCAD, and SRL and existing hybrid models including the combination of cuckoo search and square root progressive quantile variable selection procedure in sparse quadratic radial basis function (CS-SRPQVSP-QRBF) [56], particle swarm optimization in a backpropagation neural network (PSO-BPNN) [57], and the firefly algorithm in a support vector machine (SVM-FA) [58]. The experimental results are evaluated using forecasting accuracy and computation efficiency. Furthermore, the analysis of multihour-ahead cases including 24 h ahead and 48 h ahead is also provided in the experiments.

4.1. Data Collection and Description. The Xinjiang area is rich in sunshine and far from the population centers so

that it is appropriate for large-scale solar thermal power industry installations. Therefore, it is worthwhile to investigate the global solar radiation by conducting research in eight sites in this region; see Figure 2. This dataset is collected from the National Renewable Energy Laboratory (NREL) and is available at the following website <http://www.nrel.gov/gis/solar.html>. In addition to the global solar radiation (W/m^2), this data consists of seven meteorological covariates, which are the solar zenith angle (degrees), precipitation (cm), temperature ($^{\circ}\text{C}$), wind direction (degrees), wind speed (m/s), relative humidity (%), and air pressure (mbar). It was collected between 11:00 and 20:00 from 1/1/2014 to 12/31/2014 as the sunshine is poor during other periods. As an illustrative example, the global solar radiation signal and other meteorological covariates from 1/1/2014 to 1/31/2014 in Site 1 are shown in Figure 3. The main purpose of this paper is to forecast the hourly global solar radiation using these meteorological covariates accurately and efficiently, which could play a dominant role in the design of solar power plants.

4.2. Individual Models. ENN and LSSVM are implemented using the MATLAB neural network toolbox and LSSVMlab 1.5 toolbox. Specifically, a three-layer ENN is established with the number of input neurons $N_i = 56$, which is the number of covariates in the interactive model. The optimal number of hidden neurons is 5, which is selected from the set $N_h = \{5, 10, 15, 20\}$ using 10-fold CV. The number of output neurons $N_o = 1$. Thus, ENN uses a $56 \times 5 \times 1$ network architecture. Furthermore, to estimate the weights between layers more accurately, weight regularization with ℓ_2 penalty function is applied during the weight estimation procedure. The ℓ_2 regularization parameter is selected from a grid of values denoted by $\eta = \{1e-5, 1e-4, 1e-3, 1e-2, 1e-1\}$. The optimal regularization parameter η^o is determined as $1e-5$ using the Akaike information criterion (AIC) [59]. LSSVM is established by applying an RBF with two kernel parameters selected from the interval $[2^{-6}, 2^{-5}, \dots, 2^5, 2^6]$ using 10-fold CV. SRL is implemented based on the algorithm COORD, the MATLAB code of which may be downloaded from Belloni's website. Theoretical selection is applied for the selection of regularization parameter λ in COORD: $\lambda = 1.1G^{-1}(1 - 0.05/(2(p^2 + p)))$ in the optimization problem (5) recommended in [60], with G representing the cumulative distribution function of a Gaussian distribution. For the Angstrom–Prescott empirical model, we apply the empirical model in the form of a second polynomial, which has been proved to be superior to other forms of model by [55]. The possible maximum monthly mean of the daily sunshine duration is given by $Sd = (2/15)\cos^{-1}(-\tan \ell \tan \delta)$ with ℓ representing the location latitude and $\delta = 23.45 \sin(360(284 + D_n)/365)$, where D_n is the number of days in the year. CS-SRPQVSP-QBF, PSO-BPNN, and SVM-FA are implemented using MATLAB and all the model parameters are tuned to give good performances. The proposed RS-SRSCAD-FA is implemented based on Algorithm 1.

4.3. Statistical Measures of Forecasting Performance. In this paper, four common criteria, the mean absolute percentage error (MAPE), root-mean-square error (RMSE), Theil

FIGURE 2: Details of the eight sites in the Xinjiang area for measuring global solar radiation.

inequality coefficient (TIC), and correlation coefficient (R), are used to evaluate the prediction accuracy. The definitions of the criteria are as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (15)$$

$$\text{TIC} = \frac{\sqrt{(1/n) \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{(1/n) \sum_{i=1}^n \hat{y}_i^2} + \sqrt{(1/n) \sum_{i=1}^n y_i^2}}, \quad (16)$$

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}, \quad (17)$$

where n is the sample size of the test data, y_i and $\hat{y}_i$ represent the true value and estimated value, respectively, and $\bar{y}$ and $\bar{\hat{y}}$ denote the mean value of y and $\hat{y}$. The data is divided into the training dataset (75%) and test dataset (25%). Specifically, the training data is used to establish the forecasting models, and the out-of-sample forecasting performances are evaluated based on the test data. The division procedure is repeated 30 times, and the median of the MAPEs, RMSEs, TICs, and R values is reported. Furthermore, the convergence cost

(in seconds) is denoted by CC, and the experiments are implemented using a PC with an Intel Core i7 CPU at 3.60 GHz.

4.4. Results Analysis. Table 2 summarizes the forecasting results and computation cost of all the compared approaches at Sites 1–8. As the median is more robust to the outliers than the mean, we focus on the median when making comparisons. Apparently, the performance of the proposed RS-SRSCAD-FA is remarkable because it provides the lowest RMSE and TIC values, whereas its MAPEs are marginally higher than those of other forecasting methods. For instance, RMSEs of RS-SRSCAD-FA are 20.21, 21.93, 21.20, and 20.29 W/m² at sites 1–4, which are approximately 20.37%, 13.22%, 18.99%, and 22.65% higher than these of ENN. The Angstrom–Prescott model performs better than both CS-SRPQVSP-QRBF and SVM-FA in terms of accuracy, except for Site 8. For instance, RMSEs of the Angstrom–Prescott model are given as 23.13, 24.13, 24.63, 25.14, 24.74, 22.32, and 22.22 W/m², which are lower than those of CS-SRPQVSP-QRBF at 25.78, 25.16, 25.62, 22.34, 27.41, 22.65, and 22.89 W/m². CS-SRPQVSP-QRBF delivers better results than SVM-FA except for Sites 1 and 2. Both CS-SRPQVSP-QRBF and SVM-FA outperform PSO-BPNN in terms of MAPE, RMSE, and TIC. The reasons for this are probably that CS and FA are superior over PSO in parameter estimation. However, PSO-BPNN still produces better outcomes than LSSVM,

FIGURE 3: Example of a global solar radiation signal and other meteorological covariates including solar zenith angle (degrees), precipitation (cm), temperature ($^{\circ}\text{C}$), wind direction (degrees), wind speed (m/s), relative humidity (%), and air pressure (mbar).

LASSO, SCAD, and SRL, which demonstrates that the hybrid forecasting methods have proved the model performance to some extent.

LASSO, SCAD, and SRL provide comparable forecasting results and training times that are much shorter than other methods. Although LSSVM is computationally rapid, its performance is not as good as ENN and provides the largest forecasting errors at Sites 2–7. The Angstrom–Prescott

model is more computationally efficient than CS-SRPQVSP-QRBF, PSO-BPNN, and SVM-FA, which take more training time. For instance, in Site 4, the computational time of the Angstrom–Prescott model is 4.06 s, and those of CS-SRPQVSP-QRBF, PSO-BPNN, and SVM-FA are 415.35, 502.35, and 408.56 s, respectively. The reason why CS-SRPQVSP-QRBF takes more computation time is that it considers the two-way interaction terms and the selection

TABLE 2: Statistical performances of the proposed RS-SRSCAD-FA and other competitors. The statistical performances are evaluated by mean absolute percentage error (MAPE), root-mean-square error (RMSE, W/m²), Theil inequality coefficient (TIC), computation cost (CC, in seconds), and correlation coefficient (R) at Sites 1–8.

	MAPE	RMSE	TIC	CC	R	MAPE	RMSE	TIC	CC	R
<i>Sites 1 and 2</i>										
ENN	0.062	25.38	0.021	4.46	0.82	0.083	25.27	0.021	2.84	0.84
LSSVM	0.086	21.67	0.018	0.08	0.97	0.175	30.95	0.026	0.06	0.82
LASSO	0.057	28.84	0.024	1.25	0.76	0.078	28.06	0.023	0.80	0.79
SCAD	0.073	29.29	0.024	0.93	0.75	0.077	28.47	0.024	0.95	0.78
SRL	0.073	29.30	0.024	0.61	0.83	0.077	28.49	0.024	0.06	0.85
Angstrom–Prescott	0.078	23.13	0.019	2.40	0.94	0.087	24.13	0.020	4.08	0.93
CS-SRPQVSP-QRBF	0.073	25.78	0.018	585.03	0.81	0.095	25.16	0.021	425.13	0.91
PSO-BPNN	0.082	28.16	0.223	438.12	0.80	0.092	26.65	0.022	473.15	0.88
SVM-FA	0.084	24.13	0.020	312.56	0.88	0.091	26.34	0.022	387.13	0.88
RS-SRSCAD-FA	0.066	20.21	0.016	3.40	0.98	0.10	21.93	0.018	3.06	0.96
<i>Sites 3 and 4</i>										
ENN	0.082	26.17	0.022	2.64	0.85	0.079	26.23	0.022	2.87	0.83
LSSVM	0.186	31.47	0.026	0.06	0.82	0.153	32.07	0.027	0.07	0.72
LASSO	0.080	29.01	0.024	0.78	0.86	0.078	29.44	0.025	0.84	0.76
SCAD	0.081	29.42	0.024	0.90	0.88	0.079	29.91	0.025	0.90	0.79
SRL	0.082	29.75	0.025	0.56	0.87	0.079	29.79	0.025	0.61	0.78
Angstrom–Prescott	0.086	24.63	0.020	3.12	0.91	0.09	25.14	0.021	4.06	0.92
CS-SRPQVSP-QRBF	0.092	25.62	0.021	321.45	0.90	0.078	22.34	0.018	415.35	0.89
PSO-BPNN	0.100	27.65	0.023	501.18	0.87	0.110	28.35	0.023	502.23	0.79
SVM-FA	0.093	26.85	0.022	392.15	0.89	0.097	27.05	0.022	408.56	0.82
RS-SRSCAD-FA	0.091	21.20	0.017	3.29	0.92	0.078	20.29	0.017	2.96	0.94
<i>Sites 5 and 6</i>										
ENN	0.077	25.83	0.021	4.03	0.89	0.063	25.67	0.021	2.68	0.85
LSSVM	0.148	31.76	0.027	0.06	0.79	0.071	28.52	0.024	0.78	0.82
LASSO	0.079	29.03	0.024	0.79	0.80	0.069	29.03	0.024	0.91	0.79
SCAD	0.078	29.30	0.024	0.87	0.80	0.069	29.03	0.024	0.91	0.79
SRL	0.076	29.35	0.024	0.56	0.82	0.072	28.92	0.024	0.60	0.83
Angstrom–Prescott	0.092	24.74	0.021	4.40	0.91	0.084	22.32	0.019	3.78	0.90
CS-SRPQVSP-QRBF	0.085	27.41	0.022	450.56	0.89	0.067	22.65	0.018	403.12	0.89
PSO-BPNN	0.096	28.16	0.024	492.56	0.87	0.090	27.96	0.023	592.87	0.86
SVM-FA	0.094	26.97	0.022	413.95	0.88	0.089	27.35	0.023	459.78	0.86
RS-SRSCAD-FA	0.075	20.99	0.017	3.02	0.94	0.066	20.74	0.017	3.15	0.91
<i>Sites 7 and 8</i>										
ENN	0.059	25.03	0.021	2.69	0.88	0.042	27.07	0.019	2.69	0.79
LSSVM	0.119	31.41	0.026	0.06	0.74	0.059	32.05	0.023	0.06	0.72
LASSO	0.069	27.79	0.023	0.80	0.71	0.060	33.96	0.024	0.79	0.75
SCAD	0.067	28.33	0.023	0.89	0.68	0.050	30.77	0.022	0.91	0.82
SRL	0.069	28.17	0.023	0.56	0.75	0.052	31.57	0.022	0.58	0.83
Angstrom–Prescott	0.081	22.22	0.018	4.40	0.92	0.079	21.92	0.016	4.09	0.91
CS-SRPQVSP-QRBF	0.065	22.89	0.018	357.15	0.91	0.033	21.32	0.015	442.18	0.90
PSO-BPNN	0.083	26.16	0.022	398.92	0.85	0.082	27.96	0.020	402.15	0.86
SVM-FA	0.083	25.94	0.021	456.13	0.87	0.079	25.74	0.018	502.89	0.89
RS-SRSCAD-FA	0.059	19.46	0.016	3.01	0.95	0.033	21.15	0.015	4.03	0.93

of parameters in the RBF also takes some time. The the low training speed of PSO-BPNN and SVM-FA is that the models are retrained when the parameters changed. Furthermore, both ENN and RS-SRSCAD-FA compute slowly as their model structures are more complex than those of other

models. However, this is an acceptable trade-off between forecasting accuracy and computational efficiency because the accuracy is improved at the cost of more computation time. The correlation coefficients of the compared methods are given in the last column of the table. RS-SRSCAD-FA

TABLE 3: Statistical performances of the proposed RS-SRSCAD-FA and other competitors. The statistical performances are evaluated by mean absolute percentage error (MAPE), root-mean-square error (RMSE, W/m²), Theil inequality coefficient (TIC), computation cost (CC, in seconds), and correlation coefficient (R) at Sites 1–8 for 24-hour ahead forecasting.

	MAPE	RMSE	TIC	CC	R	MAPE	RMSE	TIC	CC	R
<i>Sites 1 and 2</i>										
ENN	0.044	17.94	0.014	0.23	0.82	0.057	19.62	0.015	1.16	0.85
LSSVM	0.061	16.40	0.014	0.62	0.94	0.104	22.74	0.020	0.02	0.80
LASSO	0.047	20.05	0.019	0.50	0.75	0.055	21.70	0.017	0.42	0.81
SCAD	0.036	20.43	0.017	0.37	0.73	0.053	22.34	0.020	0.59	0.82
SRL	0.041	20.25	0.016	0.30	0.81	0.053	22.31	0.020	0.03	0.84
Angstrom–Prescott	0.040	15.49	0.012	0.98	0.93	0.060	18.05	0.016	1.88	0.91
CS-SRPQVSP-QRBF	0.047	16.95	0.012	216.31	0.85	0.067	19.82	0.018	255.17	0.90
PSO-BPNN	0.053	19.75	0.015	200.01	0.84	0.065	19.98	0.018	211.24	0.87
SVM-FA	0.052	16.24	0.013	150.34	0.89	0.064	20.45	0.018	198.57	0.87
RS-SRSCAD-FA	0.044	13.11	0.010	1.47	0.96	0.069	16.90	0.014	1.90	0.95
<i>Sites 3 and 4</i>										
ENN	0.057	18.92	0.014	1.39	0.88	0.060	17.87	0.013	1.38	0.82
LSSVM	0.129	21.34	0.017	0.02	0.79	0.118	21.87	0.016	0.04	0.78
LASSO	0.055	20.53	0.016	0.38	0.8	0.059	21.25	0.016	0.49	0.74
SCAD	0.056	20.83	0.016	0.55	0.89	0.060	20.82	0.015	0.56	0.8
SRL	0.057	20.78	0.017	0.34	0.87	0.059	21.11	0.016	0.42	0.77
Angstrom–Prescott	0.059	17.18	0.012	2.06	0.89	0.067	18.49	0.014	2.82	0.89
CS-SRPQVSP-QRBF	0.064	18.05	0.014	177.89	0.89	0.060	15.73	0.012	233.37	0.90
PSO-BPNN	0.066	19.70	0.016	281.79	0.88	0.074	20.30	0.012	308.98	0.79
SVM-FA	0.064	18.64	0.014	202.40	0.92	0.073	19.55	0.014	182.97	0.85
RS-SRSCAD-FA	0.062	13.72	0.010	1.81	0.94	0.058	14.49	0.010	1.40	0.97
<i>Sites 5 and 6</i>										
ENN	0.057	17.52	0.014	2.22	0.88	0.033	18.38	0.013	1.35	0.84
LSSVM	0.110	21.47	0.016	0.03	0.78	0.047	19.74	0.014	0.32	0.83
LASSO	0.056	20.29	0.015	0.53	0.79	0.039	20.54	0.014	0.56	0.77
SCAD	0.059	19.63	0.015	0.47	0.81	0.038	20.54	0.014	0.48	0.76
SRL	0.055	20.12	0.014	0.33	0.8	0.054	18.59	0.015	0.42	0.77
Angstrom–Prescott	0.069	16.67	0.013	2.81	0.83	0.064	16.27	0.012	2.30	0.83
CS-SRPQVSP-QRBF	0.064	18.63	0.014	229.50	0.91	0.048	15.90	0.012	235.35	0.88
PSO-BPNN	0.070	18.97	0.015	289.92	0.89	0.065	19.94	0.015	350.75	0.86
SVM-FA	0.070	17.15	0.012	278.75	0.88	0.067	19.14	0.015	286.67	0.86
RS-SRSCAD-FA	0.056	13.16	0.011	1.84	0.92	0.048	14.71	0.011	1.90	0.89
<i>Sites 7 and 8</i>										
ENN	0.038	17.58	0.012	1.84	0.84	0.031	19.31	0.012	1.28	0.78
LSSVM	0.705	22.66	0.014	0.02	0.73	0.038	21.59	0.014	0.02	0.72
LASSO	0.051	19.04	0.014	0.48	0.68	0.045	22.76	0.016	0.35	0.73
SCAD	0.051	19.37	0.014	0.52	0.69	0.037	21.13	0.014	0.59	0.81
SRL	0.051	19.23	0.014	0.35	0.74	0.040	21.12	0.013	0.34	0.81
Angstrom–Prescott	0.062	16.12	0.011	2.38	0.93	0.060	15.35	0.010	2.80	0.90
CS-SRPQVSP-QRBF	0.051	15.46	0.011	217.58	0.89	0.024	14.79	0.009	272.25	0.89
PSO-BPNN	0.064	18.40	0.014	214.68	0.83	0.063	19.37	0.012	256.45	0.86
SVM-FA	0.065	17.58	0.014	282.36	0.85	0.060	17.38	0.010	342.09	0.88
RS-SRSCAD-FA	0.044	13.70	0.010	1.49	0.94	0.025	14.53	0.009	1.50	0.92

is still the best by giving the highest R values. For instance, the R values of RS-SRSCAD-FA are 0.98, 0.96, 0.92, 0.94, 0.94, 0.91, 0.95, and 0.93 for Sites 1–8, respectively. The Angstrom–Prescott model also provides comparable R values, which are also higher than 0.90. LSSVM only produces

R values of 0.97 for Site 1 and around 0.80 for the other sites. The performances of the compared methods on cases 24-hour ahead forecasting and 48-hour ahead forecasting are listed in Tables 3 and 4, respectively. The results are consistent with what we have discussed above. It is observed

TABLE 4: Statistical performances of the proposed RS-SRSCAD-FA and other competitors. The statistical performances are evaluated by mean absolute percentage error (MAPE), root-mean-square error (RMSE, W/m^2), Theil inequality coefficient (TIC), computation cost (CC, in seconds), and correlation coefficient (R) at Sites 1–8 for 48-hour ahead forecasting.

	MAPE	RMSE	TIC	CC	R	MAPE	RMSE	TIC	CC	R
<i>Sites 1 and 2</i>										
ENN	0.059	24.91	0.019	0.45	0.81	0.078	24.53	0.018	1.79	0.82
LSSVM	0.081	22.78	0.02	1.21	0.92	0.142	28.43	0.024	0.03	0.78
LASSO	0.063	27.85	0.026	0.98	0.74	0.075	27.12	0.02	0.64	0.76
SCAD	0.048	28.37	0.023	0.72	0.72	0.073	27.92	0.023	0.91	0.81
SRL	0.054	28.12	0.022	0.58	0.80	0.073	27.89	0.023	0.04	0.82
Angstrom–Prescott	0.053	21.52	0.016	1.92	0.92	0.082	22.56	0.019	2.89	0.90
CS-SRPQVSP-QRBF	0.063	23.54	0.017	424.13	0.84	0.092	24.78	0.021	392.57	0.89
PSO-BPNN	0.071	27.43	0.021	392.18	0.83	0.089	24.98	0.021	324.98	0.86
SVM-FA	0.069	22.56	0.018	294.78	0.87	0.087	25.56	0.021	305.49	0.86
RS-SRSCAD-FA	0.059	18.21	0.014	2.89	0.95	0.094	21.12	0.017	2.92	0.94
<i>Sites 3 and 4</i>										
ENN	0.079	25.92	0.021	1.98	0.87	0.076	24.15	0.019	1.92	0.81
LSSVM	0.179	29.23	0.024	0.03	0.78	0.149	29.56	0.024	0.05	0.74
LASSO	0.077	28.12	0.023	0.54	0.76	0.075	28.72	0.023	0.68	0.72
SCAD	0.078	28.54	0.023	0.78	0.87	0.076	28.14	0.022	0.78	0.78
SRL	0.079	28.46	0.024	0.49	0.85	0.075	28.53	0.024	0.59	0.76
Angstrom–Prescott	0.082	23.54	0.017	2.94	0.88	0.085	24.98	0.02	3.92	0.88
CS-SRPQVSP-QRBF	0.089	24.72	0.02	254.13	0.88	0.076	21.26	0.017	324.13	0.87
PSO-BPNN	0.092	26.98	0.023	402.56	0.87	0.094	27.43	0.018	429.14	0.78
SVM-FA	0.089	25.54	0.02	289.14	0.9	0.092	26.42	0.02	254.12	0.84
RS-SRSCAD-FA	0.086	18.79	0.014	2.59	0.91	0.074	19.58	0.015	1.94	0.96
<i>Sites 5 and 6</i>										
ENN	0.072	23.68	0.02	3.09	0.87	0.042	24.84	0.019	1.87	0.82
LSSVM	0.139	29.01	0.024	0.04	0.76	0.059	26.67	0.021	0.45	0.80
LASSO	0.071	27.42	0.022	0.74	0.78	0.049	27.76	0.021	0.78	0.76
SCAD	0.075	26.53	0.022	0.65	0.79	0.048	27.76	0.021	0.67	0.74
SRL	0.069	27.19	0.021	0.46	0.78	0.068	25.12	0.022	0.59	0.76
Angstrom–Prescott	0.087	22.53	0.019	0.39	0.80	0.081	21.98	0.018	3.19	0.82
CS-SRPQVSP-QRBF	0.081	25.18	0.02	318.75	0.89	0.061	21.49	0.017	326.87	0.87
PSO-BPNN	0.089	25.63	0.022	402.67	0.86	0.082	26.94	0.022	487.15	0.85
SVM-FA	0.088	23.18	0.018	387.15	0.85	0.085	25.87	0.022	398.15	0.85
RS-SRSCAD-FA	0.071	17.79	0.016	2.56	0.90	0.061	19.88	0.016	2.64	0.88
<i>Sites 7 and 8</i>										
ENN	0.048	23.76	0.018	2.56	0.83	0.039	26.09	0.017	1.78	0.76
LSSVM	0.892	30.62	0.02	0.03	0.72	0.048	29.17	0.021	0.03	0.68
LASSO	0.065	25.73	0.021	0.67	0.68	0.057	30.76	0.023	0.48	0.71
SCAD	0.065	26.18	0.021	0.72	0.66	0.047	28.56	0.02	0.82	0.79
SRL	0.064	25.98	0.02	0.48	0.72	0.05	28.54	0.019	0.47	0.79
Angstrom–Prescott	0.079	21.79	0.016	3.3	0.91	0.076	20.74	0.015	3.89	0.90
CS-SRPQVSP-QRBF	0.064	20.89	0.016	302.19	0.87	0.031	19.98	0.013	378.12	0.88
PSO-BPNN	0.081	24.87	0.021	298.17	0.82	0.08	26.18	0.018	356.18	0.84
SVM-FA	0.082	23.75	0.02	392.17	0.84	0.076	23.48	0.015	475.13	0.86
RS-SRSCAD-FA	0.056	18.51	0.014	2.07	0.93	0.032	19.63	0.013	2.09	0.91

that RS-SRSCAD-FA provides remarkable results followed by the Angstrom–Prescott model and CS-SRPQVSP-QRBF. The RMSEs of RS-SRSCAD-FA for 24 hours ahead are 13.11, 16.90, 13.72, 14.49, 13.16, 14.71, 13.70, and 14.53 W/m^2 for

Sites 1–8, respectively. The correlation coefficients of RS-SRSCAD-FA are given around 0.90, which are also higher than other methods. The RMSEs of LASSO, SCAD, and SRL are approximately 20–21 W/m^2 , which are comparable to but

FIGURE 4: Boxplots of MAPE, RMSE, and TIC at Sites 1–8. The compared models ENN, LSSVM, LASSO, SCAD, SRL, Angstrom-PreScott, CS-SRPQVSP-QRBF, PSO-BPNN, SVM-FA, and RS-SRSCAD-FA are indicated by numbers 1–10, respectively (the y-axis is truncated to more effectively display the remaining boxplots).

worse than those of CS-SRPQVSP-QRBF, PSO-BPNN, and SVM-FA (approximately 14–18 W/m²). Similar phenomena can also be seen in Table 4 for the case of 48-hour ahead forecasting.

The boxplots shown in Figure 4 clearly demonstrate that the forecasting values of RS-SRSCAD-FA are critically lower and a few outliers are observed (the red lines represent the median values). Figures 5 and 6 exhibit the relationship

FIGURE 5: Plot of forecasting values and actual values of global solar radiation at Sites 1–4.

between the measured data and the estimated values in the solar radiation datasets. Apparently, the points of RS-SRSCAD-FA are closer to the straight line, which demonstrates its satisfactory performances. ENN demonstrates

better forecasting results than the other penalized feature selection methods. Figure 7 exhibits the errors between actual values and forecast values provided by the proposed RS-SRSCAD-FA. It is obvious that the RS-SRSCAD-FA values

FIGURE 6: Plot of forecasting values and actual values of global solar radiation at Sites 5–8.

closely match the actual data in almost all time periods. That is, RS-SRSCAD-FA consistently displays the best forecasting performance over a vast majority of the time period.

4.5. Statistical Comparisons of Results over Multiple Real Datasets. Statistical tests for comparison of the proposed

algorithm with other existing algorithm are more reliable and trustworthy on multiple datasets than on a single dataset [61]. To make a fair comparison, a nonparametric Friedman test is applied to evaluate the forecasting and selection performances of the algorithms in this global solar radiation datasets. The Friedman test examines whether the measured

FIGURE 7: Plot of errors between actual global solar radiation and forecast values of RS-SRSCAD-FA at Sites 1–8.

average rank is significantly different from the mean rank that is likely under the null hypothesis. Table 5 reveals the ranks of the MAPEs, RMSEs, and TICs of the proposed approach and of the other competitors. It is convenient to calculate the χ_F^2 and F_F values as shown in the table. Using the six approaches and eight datasets, F_F is distributed according to $8 - 1 = 7$ and $(10 - 1) * (6 - 1) = 45$ degrees of freedom. The critical value for the statistic $F(7, 45)$ for $\alpha = 0.1$ is 1.85; therefore, the null hypothesis is rejected, which indicates that there are differences between these algorithms and more tests are needed.

To confirm the advantage, the newly proposed RS-SRSCAD-FA method is regarded as the control algorithm, and the other competitors are tested against it. The most straightforward method is to calculate critical difference (CD) with the Bonferroni–Dunn test [62]. According to [61], the critical value for $q_{0.1}$ for the 10 methods is 2.326. Thus, $CD = 2.326 \times \sqrt{(10 \times 11)/(6 \times 8)} = 3.52$. It is evident that the forecasting models except LSSVM, PSO-BPNN, and SVM-FA yield comparable MAPEs as the differences between their average ranks are less than the CD value 3.52. For example, the difference between SRL and RS-SRSCAD-FA is less than 3.52 ($4.31 - 3.25 = 1.06 < 3.52$). In terms of RMSE, RS-SRSCAD-FA performs better than LSSVM, LASSO, and SCAD, and SRL and is comparable with ENN, Angstrom–Prescott, CS-SRPQVSP-QRBF, and SVM-FA based on the differences between the average ranks. For instance, the RMSE differences between LASSO and RS-SRSCAD-FA are larger than 3.52 ($7.81 - 1.00 = 6.81 > 3.52$). The differences between Angstrom–Prescott and RS-SRSCAD-FA are less than 3.52 ($2.38 - 1.00 = 1.38 < 3.52$). Similar phenomena can be observed from the aspect of TIC and R values: most of the results of the nonparametric Friedman test are consistent with our observations. Unlike ENN and SVM-FA, RS-SRSCAD-FA provides an interpretable model that

does not employ all the covariates. Therefore, RS-SRSCAD-FA exhibits a higher prediction capacity relative to other forecasting methods.

5. Conclusions

Forecasting solar radiation is fundamental to solar power technology. For the utilization and conversion of solar power, accurate and continuous radiation data is essential. Therefore, establishing an accurate and efficient forecasting model plays a vital role in global solar radiation forecasting. To overcome the drawbacks of a single model, which yields low forecasting accuracy, a novel global solar radiation forecasting method called RS-SRSCAD-FA has been proposed. The main structure of the novel method is ensemble learning for enhancing the forecasting accuracy and model stability. An efficient covariate selection method SRSCAD is used in the input space; simultaneously, a weight vector, which best represents the importance of each individual model in the ensemble system, is determined by the FA. To illustrate the validity and superiority of the proposed method, the datasets from eight locations in the Xinjiang area of China have been applied as real data examples. The results demonstrate that the proposed RS-SRSCAD-FA method is superior to other forecasting approaches.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that the received funding did not lead to any conflicts of interest regarding the publication of this

TABLE 5: Ranks of forecasting models for Sites 1–8 of global solar radiation data.

Methods	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Average rank	χ_F^2	F_F
<i>MAPE</i>											
ENN	2	4	3.5	5	3	1	1.5	3	2.88	46.23	12.56
LSSVM	10	10	10	10	10	6	10	6	9.00		
LASSO	1	3	1	2	5	4.5	5.5	7	3.63		
SCAD	5	1.5	2	5	4	4.5	4	4	3.75		
SRL	5	1.5	3.5	5	2	7	5.5	5	4.31		
Angstrom–Prescott	7	5	5	7	7	8	7	8.5	6.81		
CS-SRPQVSP-QRBF	5	8	7	2	6	3	3	1.5	4.44		
PSO-BPNN	8	7	9	9	9	10	8.5	10	8.81		
SVM-FA	9	6	8	8	8	9	8.5	8.5	8.13		
RS-SRSCAD-FA	3	9	6	2	1	2	1.5	1.5	3.25		
<i>RMSE</i>											
ENN	5	4	4	4	3	4	4	5	4.13	61.63	41.60
LSSVM	2	10	10	10	10	7	10	9	8.50		
LASSO	8	7	7	7	7	9.5	7	10	7.81		
SCAD	9	8	8	9	8	9.5	9	7	8.44		
SRL	10	9	9	8	9	8	8	8	8.63		
Angstrom–Prescott	3	2	2	3	2	2	2	3	2.38		
CS-SRPQVSP-QRBF	6	3	3	2	5	3	3	2	3.38		
PSO-BPNN	7	6	6	6	6	6	6	6	6.13		
SVM-FA	4	5	5	5	4	5	5	4	4.63		
RS-SRSCAD-FA	1	1	1	1	1	1	1	1	1.00		
<i>TIC</i>											
ENN	6	3.5	4.5	4.5	2.5	4	4.5	5	4.31	59.57	33.56
LSSVM	2.5	10	10	10	10	8.5	10	9	8.75		
LASSO	8	7	7.5	8	7.5	8.5	8	10	8.06		
SCAD	8	8.5	7.5	8	7.5	8.5	8	7.5	7.94		
SRL	8	8.5	9	8	7.5	8.5	8	7.5	8.13		
Angstrom–Prescott	4	2	2	3	2.5	3	2.5	3	2.75		
CS-SRPQVSP-QRBF	2.5	3.5	3	2	4.5	2	2.5	1.5	2.69		
PSO-BPNN	10	5.5	6	6	7.5	5.5	6	6	6.56		
SVM-FA	5	5.5	4.5	4.5	4.5	5.5	4.5	4	4.75		
RS-SRSCAD-FA	1	1	1	1	1	1	1	1.5	1.06		
<i>R</i>											
ENN	6	7	9	4	3.5	6	4	8	5.94	56.57	25.67
LSSVM	2	8	10	10	10	8	8	10	8.25		
LASSO	9	9	8	9	8.5	9.5	9	9	8.88		
SCAD	10	10	5	6.5	8.5	9.5	10	7	8.31		
SRL	5	6	6.5	8	7	7	7	6	6.56		
Angstrom–Prescott	3	2	2	2	2	2	2	2	2.13		
CS-SRPQVSP-QRBF	7	3	3	3	3.5	3	3	3	3.56		
PSO-BPNN	8	4.5	6.5	6.5	6	4.5	6	5	5.88		
SVM-FA	4	4.5	4	5	5	4.5	5	4	4.50		
RS-SRSCAD-FA	1	1	1	1	1	1	1	1	1.00		

manuscript and there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 71761016 and

71861012), the Natural Science Foundation of Jiangxi, China (Grant Nos. 20181BAB211020 and 20171BAA218001), the China Postdoctoral Science Foundation (Grant Nos. 2017M620277 and 2018T110654), and Scientific Research Fund of Jiangxi Provincial Education Department (Grant No. GJJ180287).

References

- [1] H. Jiang, Y. Dong, J. Wang, and Y. Li, "Intelligent optimization models based on hard-ridge penalty and RBF for forecasting global solar radiation," *Energy Conversion and Management*, vol. 95, pp. 42–58, 2015.
- [2] R. Calif, F. G. Schmitt, Y. Huang, and T. Soubdhan, "Intermittency study of high frequency global solar radiation sequences under a tropical climate," *Solar Energy*, vol. 98, pp. 349–365, 2013.
- [3] R. Calif, F. G. Schmitt, and O. Durán, Medina, *5/3 Kolmogorov Turbulent Behaviour and Intermittent Sustainable Energies*, 2017.
- [4] Z. Zeng, H. Yang, R. Zhao, and J. Meng, "Nonlinear characteristics of observed solar radiation data," *Solar Energy*, vol. 87, no. 1, pp. 204–218, 2013.
- [5] S. Hussain and A. A. Alili, "A pruning approach to optimize synaptic connections and select relevant input parameters for neural network modelling of solar radiation," *Applied Soft Computing*, vol. 52, pp. 898–908, 2017.
- [6] C. Voyant, G. Notton, S. Kalogirou et al., "Machine learning methods for solar radiation forecasting: a review," *Journal of Renewable Energy*, vol. 105, pp. 569–582, 2017.
- [7] Y. Kashyap, A. Bansal, and A. K. Sao, "Solar radiation forecasting with multiple parameters neural networks," *Renewable & Sustainable Energy Reviews*, vol. 49, pp. 825–835, 2015.
- [8] K. Gairaa, A. Khellaf, Y. Messlem, and F. Chellali, "Estimation of the daily global solar radiation based on Box-Jenkins and ANN models: a combined approach," *Renewable & Sustainable Energy Reviews*, vol. 57, pp. 238–249, 2016.
- [9] M. Ozgoren, M. Bilgili, and B. Sahin, "Estimation of global solar radiation using ANN over Turkey," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5043–5051, 2012.
- [10] H. A. N. Hejase, M. H. Al-Shamisi, and A. H. Assi, "Modeling of global horizontal irradiance in the United Arab Emirates with artificial neural networks," *Energy*, vol. 77, pp. 542–552, 2014.
- [11] C. Renno, F. Petito, and A. Gatto, "ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building," *Journal of Cleaner Production*, vol. 135, pp. 1298–1316, 2016.
- [12] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Applied Energy*, vol. 168, pp. 568–593, 2016.
- [13] P. Coulibaly and N. D. Evora, "Comparison of neural network methods for infilling missing daily weather records," *Journal of Hydrology*, vol. 341, no. 1–2, pp. 27–41, 2007.
- [14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
- [15] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 130, no. 4, pp. 18–28, 2002.
- [16] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3–4, pp. 704–716, 2006.
- [17] B. Zhu and Y. Wei, "Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology," *OMEGA - The International Journal of Management Science*, vol. 41, no. 3, pp. 517–524, 2012.
- [18] W. Wu and H.-B. Liu, "Assessment of monthly solar radiation estimates using support vector machines and air temperatures," *International Journal of Climatology*, vol. 32, no. 2, pp. 274–285, 2012.
- [19] J.-L. Chen, G.-S. Li, and S.-J. Wu, "Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration," *Energy Conversion and Management*, vol. 75, pp. 311–318, 2013.
- [20] B. B. Ekici, "A least squares support vector machine model for prediction of the next day solar insolation for effective use of PV systems," *Measurement*, vol. 50, no. 1, pp. 255–262, 2014.
- [21] S. Cao and J. Cao, "Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis," *Applied Thermal Engineering*, vol. 25, no. 2–3, pp. 161–172, 2005.
- [22] H. Jiang and Y. Dong, "Global horizontal radiation forecast using forward regression on a quadratic kernel support vector machine: case study of the Tibet autonomous region in China," *Energy*, vol. 133, pp. 270–283, 2017.
- [23] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)," *Journal of Renewable Energy*, vol. 115, pp. 411–422, 2018.
- [24] M. Guermoui, K. Gairaa, A. Rabehi, D. Djafer, and S. Benkacali, "Estimation of the daily global solar radiation based on the Gaussian process regression methodology in the Saharan climate," *The European Physical Journal Plus*, vol. 133, no. 6, p. 211, 2018.
- [25] M. Guermoui, F. Melgani, and C. Danilo, "Multi-step ahead forecasting of daily global and direct solar radiation: a review and case study of Ghardaia region," *Journal of Cleaner Production*, vol. 201, pp. 716–734, 2018.
- [26] J. Wang, Y. Xie, C. Zhu, and X. Xu, "Solar radiation prediction based on phase space reconstruction of wavelet neural network," *Procedia Engineering*, vol. 150, no. 9, pp. 4603–4607, 2011.
- [27] S. Monjoly, M. André, R. Calif, and T. Soubdhan, "Hourly forecasting of global solar radiation based on multiscale decomposition methods: a hybrid approach," *Energy*, vol. 119, pp. 288–298, 2017.
- [28] S. Hussain and A. Alilili, "A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks," *Applied Energy*, vol. 208, pp. 540–550, 2017.
- [29] M. A. Hassan, A. Khalil, S. Kaseb, and M. A. Kassem, "Exploring the potential of tree-based ensemble methods in solar radiation modeling," *Applied Energy*, vol. 203, pp. 897–916, 2017.
- [30] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, 2012.
- [31] W. G. Zhao, J. Z. Wang, and H. Y. Lu, "Combining forecasts of electricity consumption in China with time-varying weights updated by a high-order Markov chain model," *Omega*, vol. 450, no. 45, pp. 80–91, 2014.
- [32] Y. Gala, Á. Fernández, J. Díaz, and J. R. Dorronsoro, "Hybrid machine learning forecasting of solar radiation values," *Neurocomputing*, vol. 176, pp. 48–59, 2016.
- [33] H. Sun, D. Gui, B. Yan et al., "Assessing the potential of random forest method for estimating solar radiation using air pollution index," *Energy Conversion and Management*, vol. 119, pp. 121–129, 2016.
- [34] R. Aler, I. M. Galván, J. A. Ruiz-Arias, and C. A. Gueymard, "Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting," *Solar Energy*, vol. 150, pp. 558–569, 2017.

- [35] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996.
- [36] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [37] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.
- [38] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [39] D. Yang, Z. Ye, L. H. I. Lim, and Z. Dong, "Very short term irradiance forecasting using the lasso," *Solar Energy*, vol. 114, pp. 314–326, 2015.
- [40] T. Zhang, "Multi-stage convex relaxation for feature selection," *Bernoulli*, vol. 19, no. 5, pp. 2277–2293, 2013.
- [41] K. Knight and W. Fu, "Asymptotics for lasso-type estimators," *The Annals of Statistics*, vol. 28, no. 5, pp. 1356–1378, 2000.
- [42] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [43] P. Zhao and B. Yu, "On model selection consistency of Lasso," *Journal of Machine Learning Research*, vol. 7, no. 12, pp. 2541–2563, 2006.
- [44] K. Lounici, "Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators," *Electronic Journal of Statistics*, vol. 2, pp. 90–102, 2008.
- [45] G. E. P. Box, "Science and statistics," *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 791–799, 1976.
- [46] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Operational Research Quarterly*, vol. 20, no. 4, pp. 451–468, 1969.
- [47] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [48] T. K. Ho, "Random decision forests," in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, pp. 278–282, Montreal, Canada, 1995.
- [49] X. She Yang, "Firefly algorithm, lévy flights and global optimization. Research and development in intelligent systems," *Research and Development in Intelligent Systems XXVI*, vol. 26, pp. 209–218, 2010.
- [50] O. P. Verma, D. Aggarwal, and T. Patodi, "Opposition and dimensional based modified firefly algorithm," *Expert Systems with Applications*, vol. 44, pp. 168–176, 2016.
- [51] H. Wang, W. Wang, X. Zhou et al., "Firefly algorithm with neighborhood attraction," *Information Sciences*, vol. 382–383, pp. 374–387, 2017.
- [52] X. She Yang, *Nature-Inspired Metaheuristic Algorithms*, Luniver Press, 2008.
- [53] Y. She, "Thresholding-based iterative selection procedures for model selection and shrinkage," *Electronic Journal of Statistics*, vol. 3, pp. 384–415, 2009.
- [54] Yu. Nesterov, "Gradient methods for minimizing composite objective function," Technical Report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2007.
- [55] L. Feng, A. Lin, L. Wang, W. Qin, and W. Gong, "Evaluation of sunshine-based models for predicting diffuse solar radiation in China," *Renewable & Sustainable Energy Reviews*, vol. 94, pp. 168–182, 2018.
- [56] H. Jiang, "A novel approach for forecasting global horizontal irradiance based on sparse quadratic RBF neural network," *Energy Conversion and Management*, vol. 152, pp. 266–280, 2017.
- [57] M. A. Mohandes, "Modeling global solar radiation using Particle Swarm Optimization (PSO)," *Solar Energy*, vol. 86, no. 11, pp. 3137–3145, 2012.
- [58] L. Olatomiwa, S. Mekhilef, S. Shamshirband, K. Mohammadi, D. Petković, and C. Sudheer, "A support vector machine-firefly algorithm-based model for global solar radiation prediction," *Solar Energy*, vol. 115, pp. 632–644, 2015.
- [59] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [60] A. Belloni, V. Chernozhukov, and L. Wang, "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.
- [61] A. Janecz, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 70, no. 1, pp. 1–30, 2006.
- [62] O. J. Dunn, "Multiple comparisons among means," *Journal of the American Statistical Association*, vol. 56, no. 293, pp. 52–64, 1961.

Submit your manuscripts at
www.hindawi.com