A Comprehensive Application of Machine Learning Techniques for Short-Term Solar Radiation Prediction

Linhua Wang and Jiarong Shi *

School of Science, Xi’an University of Architecture and Technology, Xi’an 710055, China; fhy9704251522@163.com
* Correspondence: shijiarong@xauat.edu.cn; Tel.: +86-29-8220-5670

Abstract: Forecasting the output power of solar PV systems is required for the good operation of the power grid and the optimal management of energy fluxes occurring in the solar system. Before forecasting the solar system’s output, it is essential to focus on the prediction of solar irradiance. In this paper, the solar radiation data collected for two years in a certain place in Jiangsu in China are investigated. The objective of this paper is to improve the ability of short-term solar radiation prediction. Firstly, missing data are recovered through the means of matrix completion. Then the completed data are denoised via robust principal component analysis. To reduce the influence of weather types on solar radiation, spectral clustering is adopted by fusing sparse subspace representation and k-nearest-neighbor to partition the data into three clusters. Next, for each cluster, four neural networks are established to predict the short-term solar radiation. The experimental results show that the proposed method can enhance the solar radiation accuracy.

Keywords: solar radiation; matrix completion; robust principal component analysis; sparse subspace representation; k-nearest-neighbor; artificial neural network

1. Introduction

In recent years, the scale of renewable energy power generation has expanded rapidly. Many countries are considering incorporating renewable energy into the grid [1]. Solar energy has become one of the main sources of renewable energy [2]. The narrow definition of solar energy is solar radiation [3]. Broadly speaking, solar energy also includes other forms of energy converted from the solar radiation, such as coal, oil, natural gas, hydropower, wind energy, biological energy, etc. Solar radiation is affected by the seasons and geography, and has obvious discontinuities and uncertainties [4,5]. These characteristics are the reason that the focus of prediction must be on solar radiation prior to predicting the output of a solar system.

Photovoltaic (PV) power generation is typically divided into two forms: off-grid and grid-connected. With the maturity and development of grid-connected PV technology, grid-connected PV power generation has become a mainstream trend [6]. The capacity of large-sale centralized grid-connected PV power generation systems is rapidly increasing. However, the output power of grid-connected PV power generation systems is inherently intermittent and uncontrollable. These intrinsic characteristics cause an adverse impact on the grid and seriously restrict grid-connected PV power generation [7].

At present, research on solar radiation prediction has become more and more extensive and in depth. Among various prediction methods, the simplest is the persistence method which assumes that the future solar radiation is equal to the current solar radiation. Other solar radiation prediction methods can be classified into four categories: physical methods, statistical methods, machine learning methods and hybrid methods [8–11]. Figure 1 briefly summarizes four types of prediction methods on solar radiation.
Among the four categories in Figure 1, the physical methods establish the solar power generation forecast model according to the geographical environment and weather data (such as temperature, humidity, pressure, etc.) [8]. These methods can be further grouped into two subcategories: numerical weather prediction (NWP) methods [12] and spatial correlation methods [13]. NWP methods use numerical simulation to predict, that is, mathematical and physical models are applied on analyzing atmospheric conditions, and high-speed computers are utilized to forecast solar radiation [14]. Under normal conditions, NWP methods probably take a long time to predict [15]. Moreover, the meteorological and environmental factors in NWP methods are the most complicated and difficult to make accurate decisions [8,16]. In current research, it has always been difficult to improve forecast accuracy. The spatial correlation methods harness the spatial correlation of solar radiation to predict solar energy of several places. It should be noted that spatial correlation methods require rich historical data to simulate complex temporal and spatial changes. In summary, NWP methods and other physical models are not suitable for use in short-term cases and in small areas, owing to long runtimes. Meanwhile, they have high demands on computing resources.

The forecasting of solar radiation intensity and solar energy based on historical experimental data is more suitable for short-term prediction [17]. Statistical methods can be mainly classified into moving average (MA), autoregressive (AR) and autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), autoregressive conditional heteroscedasticity (ARCH) and Kalman filtering [18,19]. The above models have fast calculation speeds, a strong interpretation ability and simple structures. However, statistical methods establish rigorous mathematical relationships between the inputs and outputs, which means that they cannot learn and change prediction strategies. In addition, a large amount of historical recording is required. As a result, it is almost impossible to capture nonlinear behavior in a time series, so the prediction accuracy may be decreased as time goes by.

With the booming development of artificial intelligence, the application of machine learning technology on predicting PV generation is becoming more popular. These advanced techniques include artificial neural networks (ANN), fuzzy logic (FL), support vector machines (SVM), random forest (RF) and the naive Bayesian algorithm [20–25]. The main principle of machine learning methods is as follows. Among them, artificial neural networks are a frequently used method, mainly containing black-propagation (BP) neural networks [26], radial basis function (RBF) neural networks [27], extreme learning machine
(ELM) networks [28], and long short-term memory (LSTM) neural networks [29]. Several types of elements affecting solar radiation are determined firstly as the input features, then a nonlinear and highly complex mapping relationship is constructed. Finally, the model parameters are learned according to historical data. Traditional statistical methods cannot attain the above complex representation in most situations. In contrast, machine learning methods can overcome this deficiency.

Hybrid methods of solar radiation prediction mainly consist of weight-based methods and prediction-assisted methods. The former type is a combined model composed of multiple single models with the same structure. Each model gives a unique prediction, and the weighted average of the prediction results of all models is regarded as the final prediction result [30, 31]. Unlike weight-based methods, prediction assistance methods usually include two models, one for power prediction and the other for auxiliary processes, such as data filtering, data decomposition, optimal parameter selection, and residual evaluation. According to the auxiliary technology, the forecast methods can be further divided into three groups: data preprocessing techniques, parameter optimization techniques, and residual calibration techniques. Among them, data preprocessing techniques are the commonly used methods, and they mainly include principal component analysis (PCA) and cluster analysis [31, 32], the wavelet transform (WT) [33], empirical mode decomposition (EMD) [34] and variational mode decomposition (VMD) [35], etc. Reasonable selections of preprocessing methods can reduce the negative impact of the systematic error on prediction accuracy to a certain extent. In summary, each single model has its advantages and disadvantages, and the hybrid model combines advantages of different methods to obtain a better prediction performance.

This paper aims to predict short-term solar radiation through a comprehensive application of machine learning techniques. Firstly, the missing values are recovered via the means of matrix completion with low-rank structure. Robust principal component analysis, a method of strong robustness to large sparse noise, is employed to denoise the recovered data. Next, solar radiation data after denoising is clustered by fusing sparse subspace representation and k-nearest-neighbor. Subsequently, four artificial neural network models are used to forecast, and thus a kind of hybrid model for short-term solar radiation prediction is proposed.

The main structure of the paper is organized as follows. Section 2 describes the experimental dataset and methods. Machine learning techniques for data preprocessing are introduced in Section 3. Section 4 presents several machine learning techniques for forecasting solar radiation. In Section 5, the experiments are carried out, and a comparison of experimental results is provided. Section 6 draws conclusions.

2. Materials and Methods

2.1. Dataset

The global horizontal irradiation data was collected at a PV power plant in Jiangsu in China. The installed capacity of the power plant was nearly 1.1 megawatts-peak (MWP). The acquired data were recorded every 5 min and the period was from 2018 to 2019, but the dates of 25 and 26 September in 2019 were not considered, due to an error in collection. There were in totally 288 recordings each day. A small amount of abnormal data was generated due to equipment or operation failures, and these unreasonable recordings were regarded as missing entries in this paper. Figure 2 illustrates the solar radiation in 2018 and 2019, respectively, where the blue trend represents lower solar radiation intensity and the yellow trend represents higher solar radiation intensity. It can be seen from two the subfigures that solar radiation is mainly concentrated from 7 a.m. to 6 p.m. each day, and the solar radiation intensity is generally stronger from 12 a.m. to 3 p.m.
We separated all solar radiation data according to the four seasons: spring, summer, autumn and winter. In total, spring owns 180 days including January, February and March. In summer, there are 182 days in April, May and June. There are 182 days in July, August and September in autumn. Winter has 184 days in October, November and December. Figure 3 further illustrates the solar radiation intensity of the four seasons in 2018. In that year, the average daily maximum value of solar radiation intensity of spring is 592.64 Wh/m². In summer and autumn, the average daily maximum solar radiation intensity is 879.88 Wh/m² and 949.67 Wh/m², respectively. Compared with the other seasons, the average daily maximum solar radiation intensity in winter is only 549.33 Wh/m².

Figure 2. Visualization of the global horizontal irradiation. (a) 2018; (b) 2019.

Figure 3. Solar radiation in four seasons in 2018 and 2019. (a) Spring; (b) Summer; (c) Autumn; (d) Winter.
By observing Figures 2 and 3, we can see that these data show no obvious trend or seasonality. What is more, the data are relatively complicated and there are many missing elements. If the original data were directly utilized to perform forecasting, the prediction results would probably have a large error, which would affect the normal operation of the PV power grid. According to the above data characteristics, it is essential to choose the appropriate data processing method. For obtaining higher data quality, we need to recover the missing entries and denoise the completed data. Subsequently, cluster analysis is adopted to reduce the complexity of data. The data of each cluster are chosen to make predictions separately, which can improve the prediction efficiency and accuracy to some extent.

### 2.2. Construction of the Hybrid Model

The hybrid model presented in this paper can be divided into the following two parts. The first part adopts machine learning methods including matrix completion, RPCA (robust principal component analysis) and cluster analysis to preprocess the original data. In the second part, the prediction is carried out by making use of the neural network in a machine learning method such as the BP neural network, radial basis function, extreme learning machine and long short-term memory.

Figure 4 shows the flow chart of the proposed hybrid model. The main process of data preprocessing includes the recovery of missing data through matrix completion, denoising of the completed dataset, and spectral clustering based on the combination of sparse subspace representation and k-nearest-neighbor. By integrating the neural network models, a short-term solar radiation prediction model is accomplished.

![Figure 4. Framework of the hybrid solar radiation forecast model.](image-url)
3. Machine Learning Techniques for Data Preprocessing

This section will introduce three unsupervised machine learning methods to preprocess solar radiation data. Firstly, a matrix completion method is utilized to recover the missing data. Then, robust principal component analysis (RPCA) is employed to denoise the completed data. Ultimately, a spectral clustering method based on the fusion of k-nearest-neighbor and sparse subspace representation is proposed to perform cluster analysis on the denoised data.

3.1. Data Completion and Denoising

3.1.1. Matrix Completion

Let \( z_i \in \mathbb{R}^{m \times 1} \) be an irradiation measure vector in the \( i \)-th day, \( i = 1, 2, \ldots, n \). These \( n \) samples can be expressed as the matrix \( Z = (z_1, z_2, \ldots, z_n) \). For any real matrix \( Z = (z_{ij})_{m \times n} \), its nuclear norm is defined as \( \|Z\|_* = \sum_{j=1}^{\min(m,n)} \sigma_j \), where \( \sigma_j \) is the \( j \)-th largest singular value of \( Z \). To indicate all missing entries of \( Z \), an index set \( \Omega \subset \{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\} \) is firstly introduced. Subsequently, a projection operator of \( P_\Omega(\cdot) : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n} \) is defined as follows. If \((i, j) \in \Omega\), then \( P_\Omega(z_{ij}) = z_{ij} \); otherwise, \( P_\Omega(z_{ij}) = 0 \).

With regard to solar radiation, \( n \) samples can be roughly divided into several groups. Therefore, \( Z \) is approximately low-rank when \( n \) is relatively large. In the presence of missing elements, the recovery technique by the aid of the low-rank structure is called matrix completion [36,37]. Matrix completion is initially described as an affine rank minimization problem. However, due to the non-convexity and discontinuity of the rank function, it is intractable to address this problem. To this end, the aforementioned optimization model can be convexly relaxed into the matrix nuclear norm minimization [38,39]. Thus, the mathematical model of matrix completion is formulated as follows:

\[
\min_{\tilde{Z}} \|\tilde{Z}\|_* \quad \text{s.t.} \quad P_\Omega(\tilde{Z}) = P_\Omega(Z)
\]  

where \( \tilde{Z} \) is the completed matrix. The optimal solution of the above minimization is also denoted as \( \tilde{Z} \) to avoid abuse of symbols.

3.1.2. Robust Principal Component Analysis

Observing Figures 1 and 2, we find that the solar radiation data is contaminated by some large sparse noise, which is detrimental for forecasting the future trend. For a low-rank data matrix corrupted by small dense noise, principal component analysis (PCA) can effectively perform dimension reduction, noise elimination, and feature extraction [39]. However, generally speaking, PCA does not work well when the studied dataset is superposed by large sparse noise or outliers. Hence, research on the robustness of PCA has always been the main focus of attention.

The emerging robust principal component analysis (RPCA) decomposes a matrix into the sum of a low-rank matrix and a sparse noise matrix, and the principal component pursuit is proposed to obtain the optimal decomposition [40]. This robust version of PCA can accurately recover the low rank component and the sparse noise under some conditions [41,42]. Formally, RPCA is modeled as follows:

\[
\min_{A,E} \|A\|_* + \lambda \|E\|_1, \quad \text{s.t.} \quad \tilde{Z} = A + E
\]  

where \( A \) is the low-rank component, \( E \) is the sparse noise matrix, \( \lambda > 0 \) balances the low rankness and the sparsity, \( \|\cdot\|_1 \) is the \( l_1 \)-norm of a matrix (i.e., the sum of absolute values of all elements). The alternating direction method of multipliers is frequently used to solve the nuclear norm minimization (1) and (2). The optimal solution of the above minimization is denoted as \( \tilde{A} \).
3.2. Data Cluster Analysis

3.2.1. k-Nearest-Neighbor

Cluster analysis refers to the process of dividing a given data set into several groups based on the similarity or the distance between two samples without prior information, and it is beneficial to further explore and mine the essence of the data [43]. Spectral clustering is a popular clustering method based on graph theory. The cluster task is achieved by clustering the eigenvectors of the Laplacian matrix for the sample set [44]. It can be explained that the core of spectral clustering is to map the points from high-dimensional space to low-dimensional space, and some clustering algorithms are used in a low-dimensional space. A similarity matrix is an important index in spectral clustering, and it is constructed according to the distance metric among all points [45]. In the spectral clustering algorithm, the similarity between two points with a short distance is relatively high, otherwise the similarity is relatively low. The similarity matrix can be built up through three ways: ε-neighborhood graph, k-nearest-neighbor graph and fully connected graph [46]. Among these three manners, ε-neighborhood graph and fully connected graph probably lose more information, which leads to less accurate results. In contrast, the k-nearest-neighbor graph generally has a precise calculation result. Meanwhile, it is simple and easy to realize [47].

The first step of spectral clustering is to establish a weighted graph $G = (V, E)$, where $V$ is the set of $n$ nodes and $E$ is a collection of edges among nodes. Let $\tilde{A} = (\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$. The $i$-th node corresponds to the processed observation vector $\tilde{a}_i$. Constructing the similarity graph is the most crucial task for spectral clustering. Among the existing approaches, the k-nearest-neighbor graph is generally recommended as the first choice. Next, we introduce the construction process of the similarity matrix based on the k-nearest-neighbor graph.

In a manifold space, there exists an approximate liner relationship among the adjacent points. Under this circumstance, the distance between $\tilde{a}_i$ and $\tilde{a}_j$ is calculated by $d_{ij} = \|\tilde{a}_i - \tilde{a}_j\|_2^1$, where $\|\cdot\|_2$ is the $l_2$-norm of a vector. Given $\tilde{a}_i$, we first compute $n - 1$ distances $\{d_{i1}, d_{i2}, \ldots, d_{ij-1}, d_{ij+1}, \ldots, d_{in}\}$ and then sort them by the increasing order. Therefore, k nearest neighbors of $\tilde{a}_i$ can be found according to k smallest distances. On this basis, we build up a similarity graph $G$: if $\tilde{a}_i$ is one of k nearest neighbors for $\tilde{a}_j$ or $\tilde{a}_j$ is one of k-nearest-neighbors for $\tilde{a}_i$, then an edge is added between the $i$-th and the $j$-th nodes and the weight $s_{ij}$ is set to 1; otherwise, no edge exists and $s_{ij} = 0$. Eventually, the similarity matrix $S = (s_{ij})_{n \times n}$ is obtained and it is symmetrized by $s_{ij} \leftarrow \max\{s_{ij}, s_{ji}\}$.

3.2.2. Sparse Subspace Representation

As a sparse learning method, sparse subspace representation is another manner to construct the similarity matrix in spectral clustering. Generally speaking, a high-dimensional dataset is distributed in the union of several low-dimensional subspaces. Hence, the representation of high-dimensional data is characterized by the sparseness [48–50].

It is supposed that the dataset locates in the union of several disjoint linear subspaces. The purpose of cluster analysis is to recognize and separate these subspaces. If $\tilde{A}$ is noise-free and there exist sufficient samples or points in each subspace, then the $i$-th sample can be expressed by the linear combination of the remainder, $i = 1, 2, 3, \ldots, n$. Thus, it holds that

$$\tilde{a}_i = v_1\tilde{a}_1 + v_2\tilde{a}_2 + \ldots + v_{i-1}\tilde{a}_{i-1} + v_{i+1}\tilde{a}_{i+1} + \ldots + v_n\tilde{a}_n \quad (3)$$

where $v_{ji} (j \neq i)$ is the linear representation coefficient of the $j$-th sample. Let $v_i = (v_{i1}, v_{i2}, \ldots, v_{in})^T$, where $v_{ij} = 0$. If $v_{ij}$ is large in the sense of absolute value, the $i$-th and the $j$-th samples probably have strong similarity.

Equation (3) can be written as $\tilde{a}_i = \tilde{A}v_i$. In detailed implementation, the optimal coefficient vector $v_i$ can be obtained by solving the following $l_p$-norm optimization problem:

$$\min_{v_i} \|v_i\|_p, \text{ s.t. } \tilde{a}_i = \tilde{A}v_i$$

(4)
The value of $p$ is commonly set to 1 or 2. The samples matrix $\tilde{A}$ is frequently corrupted by the superposition of small dense noise and large sparse noise. Under this circumstance, the vector $\tilde{a}_i$ is rewritten as:

$$\tilde{a}_i = \tilde{A}v_i + e_i + o_i$$

where $e_i$ is a large sparse noise vector and the noise vector $o_i$ is dense. We assume that $e_i$ follows a Gaussian distribution, both $e_i$ and $v_i$ obey two multivariate Laplacian distributions. By applying the maximum likelihood estimation, we formulate an optimization problem:

$$\min_{v_i,e_i,o_i} \|v_i\|_1 + \lambda_1\|e_i\|_1 + \lambda_2\|\tilde{a}_i\|_2^2 / 2, \text{ s.t. } \tilde{a}_i = \tilde{A}v_i + e_i + o_i$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are two regularization parameters. Hence, the minimization problem (6), also named as sparse subspace representation, is more robust than problem (4) in dealing with dense and large sparse noise.

3.2.3. Spectral Clustering via Fusing k-Nearest-Neighbor and Sparse Subspace Representation

The main disadvantage of k-nearest-neighbor is that it does not use global information, which possibly leads to the less robustness to data noise. As an exceedingly robust method, the sparse subspace representation not only transmits valuable information in the classification task, but also makes use of the overall context information to provide a data adaptive neighborhood [46]. However, a drawback of the sparse subspace representation is that two samples with a large Euclidean distance may be classified into the same cluster. For this propose, this paper provides a spectral clustering fusing the advantages of k-nearest-neighbor and sparse subspace representation, and applies it to the cluster analysis on solar radiation.

Let $W_k$ be the similarity matrix constructed by the k-nearest-neighbor and $W_s$ be the similarity matrix obtained by sparse subspace representation. In consideration of their construction principles, both $W_k$ and $W_s$ are sparse and symmetric. We propose a weighted similarity matrix defined by the convex combination between $W_k$ and $W_s$:

$$W = \gamma W_k + (1 - \gamma) W_s$$

where $\gamma \in [0,1]$ is the trade-off parameter. When $\gamma = 0$, the similarity matrix is constructed by sparse subspace representation. In addition, $\gamma = 1$ means that the similarity matrix is formed by the k-nearest-neighbor.

Based on the resulting similarity matrix $W$, we divide the dataset matrix $\tilde{A}$ into $s$ clusters via the spectral clustering method. The following lists the implementation procedure of spectral clustering. Firstly, a diagonal matrix $D$ is calculated, whose diagonal elements are constructed by the sum of each row in $W$. Then the Laplacian matrix is computed as $L = D^{-1/2}WD^{-1/2}$. Next, the eigen decomposition is performed on $L$ and $s$ mutually orthogonal unit eigenvectors $\{a_i^o \in \mathbb{R}^{m \times 1}\}_{i=1}^s$ corresponding to $s$ largest eigenvalues are acquired. Denote $A^o = (a_1^o, \ldots, a_s^o) \in \mathbb{R}^{m \times s}$. Each row of $A^o$ is further transformed into unit vectors in the sense of $l_2$-norm and the normalized matrix is indicated by $\tilde{A}^o$. Finally, each row of $\tilde{A}$ is regarded as a sample and $m$ samples are partitioned into $s$ clusters by k-means clustering. Compared with only using k-nearest-neighbor or sparse subspace representation, the proposed spectral clustering can maintain a stronger stability and robustness.

4. Machine Learning Techniques for Forecasting

Given input–output paired training samples $\{(x_i,y_j)\}_{i=1}^N$, we consider the supervised learning task of seeking for an approximate function $y = f(x)$, where $x_i \in \mathbb{R}^{D_1 \times 1}$ is the input vector and $y_j \in \mathbb{R}^{D_2 \times 1}$ is the output vector. To learn the function relationship between the input and the output, this section will introduce four supervised machine learning methods, namely, BP neural networks, radial basis function networks, extreme learning machines and long-short term memory models.
4.1. BP Neural Networks

Neural networks construct the functional form of $y = f(x)$ from the viewpoint of a network model [26,51]. For an input vector $x = \left( x^{(1)}, x^{(2)}, \ldots, x^{(D_1)} \right)^T$, a feedforward network with $K$-1 hidden layers can be expressed by:

$$y \approx f_{\text{BP}}(x) = s^{(K)} \left( W^{(K)} s^{(K-1)} \left( \ldots s^{(1)} \left( W^{(1)} x + b^{(1)} \right) \ldots \right) + b^{(K-1)} \right) + b^{(K)}$$

(8)

where $W^{(k)} \in \mathbb{R}^{d_k \times d_{k-1}}$ is the weights matrix in the $k$-th hidden layer, $b^{(k)} \in \mathbb{R}^{d_k \times 1}$ is the corresponding bias vector, $s^{(k)}(\cdot)$ is the nonlinear activation function adopted in the $k$-th hidden layer, $d_0 = D_1$ and $d_K = D_2$.

Denote the model parameters set by $\theta = \{ W^{(k)}, b^{(k)} \}_{k=1}^{K}$. By training the network according to all training samples, the optimal network parameters $\theta$ can be obtained. For this purpose, we minimize the following error function:

$$E(\theta) = \frac{1}{2} \sum_{i=1}^{N} \| f_{\text{BP}}(x_i) - y_i \|_2^2$$

(9)

The simplest and the most effective approach is the gradient descent, and the update formulation is

$$\theta \leftarrow \theta - \eta \nabla E(\theta)$$

(10)

where $\nabla E(\theta)$ is the gradient of $E(\theta)$ with respect to $\theta$, and the step size $\eta$ is called the learning rate.

Each parameter updating step consists of two stages. The first stage evaluates the derivatives of the error function with respect to the weight matrices and the bias vectors. The backpropagation technique propagates errors backwards through the network and it has become a computationally efficient method for evaluating the derivatives. The derivatives are employed to adjust all parameters in the second stage. Hence, the multilayer perceptron is also called a back-propagation (BP) neural network. Figure 5 depicts the topological structure of a BP neural network with one hidden layer. In detailed implementation, mini-batch gradient descent is usually utilized to update parameters to reduce the computation burden.

![Diagram of a BP neural network with a single hidden layer.](image-url)
4.2. RBF Neural Networks

As a two-layer feedforward network, a radial basis function (RBF) neural network is composed of an input layer, a hidden layer and an output layer [27]. An RBF network is a special case of BP network, and the major difference lies in that the former uses a radial basis function as activation function instead of other functions, such as a sigmoid activation function. The sigmoid activation function forces the neurons to have a large input visible area [52]. In contrast, the activation function in an RBF network has a small input space region. Consequently, an RBF network needs more radial basis neurons. Moreover, an RBF network is mainly applied to the one-dimensional output case. Figure 6 plots the RBF neural network with the case $D_2 = 1$.

![Diagram of an RBF neural network.](image)

The commonly used radial basis function in an RBF neural network is the Gaussian function. Under this circumstance, the activation function for a given input feature $x$ can be expressed as

$$\varphi(x; u, \sigma) = \exp\left(-\frac{1}{2\sigma^2}||x - u||^2\right)$$  \hspace{1cm} (11)

where $u \in \mathbb{R}^{D_1 \times 1}$ and $\sigma$ are the center and the standard deviation of the Gaussian function, respectively. The mathematical model of the RBF network with $L$ hidden units can be written as

$$y \approx f_{\text{RBF}}(x; \theta, w) = \sum_{j=1}^{L} w_j \varphi(x; u_j, \sigma_j)$$  \hspace{1cm} (12)

where $w = (w_1, w_2, \ldots, w_L)^T$ is the weights vector connecting the hidden layer to the output layer, $\theta = \{(u_j, \sigma_j)\}_{j=1}^{L}$ is a set composed by $L$ center vectors and $L$ standard deviations.

Formally, the parameters of the RBF neural network can be obtained by minimizing the following errors:

$$\min_{\theta, w} \frac{1}{2} \sum_{i=1}^{N} (y_i - f_{\text{RBF}}(x_i; \theta, w))^2$$  \hspace{1cm} (13)

If $\theta$ is fixed, the optimal weights vector $w$ is calculated as

$$w \leftarrow \Phi^+ Y$$  \hspace{1cm} (14)

where $Y = (y_1, y_2, \ldots, y_N)^T$, the notation $^+$ is the generalized inverse of a matrix, $\Phi = (\varphi_{ij})_{N \times L}$ is the design matrix with $\varphi_{ij} = \varphi(x_i; u_j, \sigma_j)$. The parameters set can be deter-
mined by the gradient descent or cross-validation method. In practice, \( \theta \) and \( w \) can be updated alternately.

4.3. ELM Neural Networks

ELM generalizes single hidden layer feedforward networks [53–55]. For an input sample \( x \in \mathbb{R}^{D_1 \times 1} \), ELM constructs a hidden layer with \( L \) nodes and the output of the \( i \)-th node is denoted by \( h_i(x) \), where \( h_i(x) \) is a nonlinear feature mapping. We can choose the output of all hidden layer nodes as follows:

\[
h(x; W, b) = (h_1(x), \ldots, h_L(x))^T = g(Wx + b)
\]

where \( W \in \mathbb{R}^{L \times D_1} \) and \( b \in \mathbb{R}^{L \times 1} \) are the weight matrix and the bias vector in the hidden layer, respectively, and \( g(\cdot) \) is the mapping function. Subsequently, the linear combination of \( \{h_i(x)\}_{i=1}^L \) is used as the resulting output of the prediction

\[
y \approx f_{\text{ELM}}(x) = \left(h(x; W, b)^T \beta\right)^T
\]

where \( \beta \in \mathbb{R}^{L \times D_2} \) is the output weight matrix. Figure 7 illustrates the diagram of an ELM neural network with one single hidden layer.

When all parameters \( \{W, b, \beta\} \) are unknown, the above prediction function can be regarded as the combination of RBF networks and BP neural networks with only one hidden layer. To simplify the network model, extreme learning machines generate randomly the hidden node parameters \( \{W, b\} \) according to some probability distributions. In other words, \( W \) and \( b \) do not need to be trained explicitly, resulting in a remarkable efficiency.

Let \( H = (h(x_1; W, b), \ldots, h(x_N; W, b))^T \), \( Y = (y_1, \ldots, y_N)^T \). The weights matrix \( \beta \) connecting the hidden layer and the output layer can be solved by minimizing the squared error loss:

\[
E(\beta) = \frac{1}{2} \sum_{i=1}^{N} \|f_{\text{ELM}}(x_i) - y_i\|^2 = \frac{1}{2} \|H\beta - Y\|_F^2
\]

where \( \| \cdot \|_F \) is the Frobenius norm of one matrix.

**Figure 7. Diagram of an ELM neural network with a single hidden layer.**
4.4. LSTM Neural Networks

As a special recurrent neural network (RNN), long short-term memory (LSTM) is suitable for processing and predicting important events with relatively long intervals and delays in the time series [56,57]. LSTM can alleviate the phenomenon of gradient disappearance in the structure of RNN [58]. As the result of a powerful representation ability, LSTM utilizes a complex nonlinear unit to construct larger deep neural networks.

LSTM controls long and short-term memory through gates and cell states [10]. As shown in Figure 8, the neurons in LSTM include input gate \(i\), forget gate \(f\), cell state \(c\), output gate \(y\). Among them, three gates are calculated as follows:

\[
\begin{align*}
    i_t &= \sigma(W_i^t[h_{t-1}, x_t] + b_i) \\
    f_t &= \sigma(W_f^t[h_{t-1}, x_t] + b_f) \\
    y_t &= \sigma(W_y^t[h_{t-1}, x_t] + b_y)
\end{align*}
\]

(18) \hspace{2cm} (19) \hspace{2cm} (20)

where \(b_i\), \(b_f\), \(b_y\) are bias terms, \(W_i\), \(W_f\), \(W_y\) are respectively the weight matrices of three gates, and \(\sigma\) is the sigmoid activation function. In Equation (18), \(W_i^t[h_{t-1}, x_t]\) indicates \(W_i^t[h_{t-1}] + W_{i2} x_t\) where \(W_i = (W_{i1}, W_{i2})\). At time \(t\), the update formula of cell state is:

\[
c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t
\]

(21)

where \(\odot\) is the Hadamard product and \(\tilde{c}_t\) is the candidate cell state. Let \(b_c\) and \(W\) be respectively the bias vector and the weight matrix of the candidate cell gate. Then \(\tilde{c}_t\) is computed as:

\[
\tilde{c}_t = \varphi(W_c^t[h_{t-1}, x_t] + b_c)
\]

(22)

where the activation function \(\varphi\) is usually chosen as the hyperbolic tangent. At last, the hidden vector is updated:

\[
h_t = y_t \odot \varphi(c_t)
\]

(23)

Figure 8. Diagram of LSTM neural networks.

For the input information \(x_t\), Equation (22) calculates the candidate cell state \(\tilde{c}_t\) at time \(t\) by \(h_{t-1}\) and \(x_t\). Equation (21) combines the input gate and the forgetting gate to update the cell state \(c_t\) at time \(t\). Equation (23) calculates the hidden layer information at time \(t\). Through the combination with gate control units, the LSTM network achieves the purpose of memorizing long- and short-term information of time series data by continuously updating the cell state at each moment.

5. Experimental Results
5.1. Model Implementation

The solar radiation observation dataset was collected from a certain part of Jiangsu Province in 2018 and 2019. It can be formed into the matrix $Z = (z_{ij})_{m \times n}$, where $m = 288$ is the total number of recordings per day, and $n = 728$ is the number of days for that selected two years.

First of all, in view of the incompleteness of these solar radiation data, the matrix completion method is utilized to infer the missing elements. This procedure can further refine and calibrate these data. In 2018, 363 days are considered and there are 92 pieces of missing data in total. There are 365 days in 2019, and 28 pieces of data are missing. Taking 11 January 2018 as an example, there are 279 recordings and 9 missing values, and Figure 9 illustrates the completed data in that day. In this figure, the blue stars represent the observed values and the red filled circles are the recovered values by matrix completion. As we can see, the matrix completion has a good recovery performance in that day. In summary, 104,544 and 105,120 recording values, respectively, are obtained after completion.

![Figure 9. Recovered solar radiation data by matrix completion for one day.](image)

The recovered solar radiation data are further divided into four parts according to the four seasons, and RPCA is employed to denoise the completed data of each season. Figure 10 shows the solar radiation waveforms of 20 non-repeated days before and after denoising for each season, and different colors indicate different days. It can be seen that the solar radiation is zero before 6 a.m. and after 7 p.m. in most cases. It is especially important that the denoised data are convenient for us to grasp the real trend of the variation in solar radiation data, which is conducive to a better prediction of solar radiation.
As can be seen from Figures 2 and 3, the differences of solar radiation intensity in the four seasons are particularly striking, and the sub-dataset of each season is disorganized without any seasonal characteristics and periodicity. For the solar radiation of each season, we utilize spectral clustering based on the fusion of k-nearest-neighbor and sparse subspace representation to divide all the days in each season into three clusters from the solar radiation intensity. In Figure 11, we cluster the radiation data of all days in each season.
into Cluster 1, Cluster 2 and Cluster 3 according to the solar radiation intensity from low to high. At the upper right of each subplot in Figure 11, the red asterisks, the blue hollow triangles and the green circles stand for Cluster 1, Cluster 2 and Cluster 3, respectively. Springs in Clusters 1, 2 and 3 have 60, 71 and 49 days, respectively, and summers have 55, 65 and 62 days. In autumn, there are 63, 68 and 51 days in the three clusters respectively, winters last 60, 71 and 49 days, respectively.

Figure 11. Cluster results of solar radiation in four seasons. (a) Spring; (b) Summer; (c) Autumn; (d) Winter.

When neural networks are used for predictions, we choose the solar radiation between 7 a.m. and 6 p.m. every day as the effective input data in order to improve the calculation speed and ensure the validity of the data. To enhance the short-term prediction ability of the proposed model, the solar radiation every two consecutive hours is selected as the training sample to predict that of the next moment. For each season, the last five days of each cluster are employed to construct the test set for final evaluation, and the remaining days are harnessed to train the neural networks. Attention should be paid as over-fitting may occur in the training process of the neural network, that is, the training error is small, but the generalization error is large. Therefore, in the experiment, we use the regularization technique for BP, RBF and ELM, and the dropout strategy for LSTM to prevent overfitting.

5.2. Performance Analysis

In order to verify the effectiveness of the proposed models, this subsection compares four commonly used neural networks, i.e., BP neural networks, RBF neural networks, ELM neural networks, and LSTM neural networks. Two commonly used statistical indices, the root mean square error (RMSE) and the mean absolute error (MAE), are adopted for model validation to quantitatively evaluate the prediction performance. Their formulations are as follows:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \| \hat{y}_i - y_i \|^2_2}
\]  

(24)
where $y_i$ is the actual value of the output data, and $\hat{y}_i$ is the corresponding predicted result.

Tables 1–8 list the prediction results of BP, RBF, ELM and LSTM in the situations of with or without clustering. For the case of without clustering, the RMSE/MAE values of 15 days in the test set of each season are reported. In the case of clustering, the prediction errors are recorded for the three clusters, respectively. The last columns in these tables give the average of the forecast results of the three clusters. Compared with the case of without clustering, the RMSE value of the BP network with clustering in the four seasons goes down respectively by 3.38, 4.51, 3.15 and 4.87, while MAE rises by 0.72, 1.96, 0.69 and 0.91, respectively. As for RBF, RMSE is respectively decreased by 40.64, 62.07, 65.83 and 28.96 and MAE is decreased by 21.07, 42.90, 19.83 and 25.18, respectively. When using ELM, RMSE is respectively reduced by 3.89, 12.51, 13.75 and 10.68 and MAE is reduced by 1.31, 5.04, 1.31 and 5.09. At last, RMSE of LSTM is improved respectively by 133.56, 41.38, 104.40 and 115.75, and MAE is improved by 145.04, 111.64, 80.6 and 120.29. The experimental results demonstrate that the performance of the four neural network models are enhanced via spectral clustering, which indicates that machine learning models are significative to improving the prediction results of short-time solar radiation.

Table 1. RMSE of solar radiation forecast errors in spring.

| without Clustering | with Clustering |
|-------------------|----------------|
|                  | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 53.84    | 58.21     | 58.96     | 34.21    | 50.46    |
| RBF               | 98.42    | 65.09     | 58.53     | 49.71    | 57.78    |
| ELM               | 58.91    | 56.15     | 63.30     | 45.61    | 55.02    |
| LSTM              | 226.26   | 76.78     | 56.08     | 145.23   | 92.70    |

Table 2. MAE of solar radiation forecast errors in spring.

| without Clustering | with Clustering |
|-------------------|----------------|
|                  | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 35.93    | 41.81     | 45.87     | 32.28    | 36.65    |
| RBF               | 65.44    | 50.90     | 45.58     | 36.64    | 44.37    |
| ELM               | 42.01    | 39.80     | 46.53     | 35.77    | 40.70    |
| LSTM              | 215.07   | 62.18     | 39.38     | 108.53   | 70.03    |

Table 3. RMSE of solar radiation forecast errors in summer.

| without Clustering | with Clustering |
|-------------------|----------------|
|                  | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 51.96    | 62.82     | 11.14     | 68.39    | 47.45    |
| RBF               | 119.66   | 76.71     | 12.47     | 83.58    | 57.59    |
| ELM               | 62.84    | 64.65     | 11.65     | 74.68    | 50.33    |
| LSTM              | 182.40   | 105.48    | 198.70    | 118.88   | 141.02   |

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \| \hat{y}_i - y_i \|_1$$
Table 4. MAE of solar radiation forecast errors in summer.

| without Clustering | with Clustering |
|-------------------|-----------------|
|                   | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 34.54     | 50.83     | 9.20      | 49.46   | 36.50   |
| RBF               | 88.71     | 61.76     | 9.81      | 65.85   | 45.81   |
| ELM               | 43.84     | 50.91     | 9.11      | 56.38   | 38.80   |
| LSTM              | 182.42    | 60.71     | 57.80     | 93.82   | 70.78   |

Table 5. RMSE of solar radiation forecast errors in autumn.

| without Clustering | with Clustering |
|-------------------|-----------------|
|                   | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 47.11     | 39.61     | 64.62     | 27.65   | 43.96   |
| RBF               | 105.10    | 45.05     | 62.96     | 9.81    | 39.27   |
| ELM               | 64.77     | 66.09     | 13.92     | 73.06   | 51.02   |
| LSTM              | 195.22    | 73.35     | 76.66     | 122.46  | 90.82   |

Table 6. MAE of solar radiation forecast errors in autumn.

| without Clustering | with Clustering |
|-------------------|-----------------|
|                   | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 28.33     | 27.95     | 43.09     | 16.02   | 29.02   |
| RBF               | 64.20     | 50.90     | 45.58     | 36.63   | 44.37   |
| ELM               | 42.01     | 39.80     | 46.53     | 35.77   | 40.70   |
| LSTM              | 158.67    | 39.38     | 77.85     | 116.81  | 78.01   |

Table 7. RMSE of solar radiation forecast errors in winter.

| without Clustering | with Clustering |
|-------------------|-----------------|
|                   | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 34.66     | 23.95     | 54.90     | 10.51   | 29.79   |
| RBF               | 62.16     | 28.14     | 54.28     | 17.18   | 33.20   |
| ELM               | 46.89     | 31.66     | 62.30     | 14.68   | 36.21   |
| LSTM              | 155.57    | 40.44     | 43.50     | 35.50   | 39.81   |

Table 8. MAE of solar radiation forecast errors in winter.

| without Clustering | with Clustering |
|-------------------|-----------------|
|                   | Cluster 1 | Cluster 2 | Cluster 3 | Average |
| BP                | 18.22     | 14.73     | 34.41     | 8.24    | 19.13   |
| RBF               | 46.93     | 12.78     | 34.68     | 16.28   | 21.25   |
| ELM               | 30.53     | 21.16     | 43.30     | 11.85   | 25.44   |
| LSTM              | 147.43    | 26.23     | 32.21     | 26.00   | 27.14   |
Tables 9–12 show the determination coefficient $R^2$ of solar radiation prediction by neural networks with and without clustering for four seasons. $R^2$ is a measure of how well the regression line represents the data, and the prediction models are more effective as $R^2$ approaches one. In contrast, in the case of without clustering, the $R^2$ of BP in the four seasons is respectively increased by 0.0946, 0.0206, $-0.0305$ and $-0.0365$ in the sense of average values. When applying RBF, $R^2$ is raised by 0.0642, $-0.0210$, $-0.0240$ and 0.0053. As for ELM, $R^2$ is improved by 0.0441, 0.0065, $-0.0165$ and 0.0031. With regard to LSTM, $R^2$ respectively went up by 0.0139, $-0.0094$, $-0.0011$ and 0.0098. The experimental results in these tables indicate that the proposed forecasting methods can significantly improve the prediction performance of short-term solar radiation in most cases.

Table 9. $R^2$ of solar radiation in spring.

|                | without Clustering | with Clustering |
|----------------|--------------------|-----------------|
|                | Cluster 1          | Cluster 2       | Cluster 3       | Average |
| BP             | 0.8491             | 0.8732          | 0.9879          | 0.9699   | 0.9437  |
| RBF            | 0.8318             | 0.8326          | 0.8763          | 0.9791   | 0.8960  |
| ELM            | 0.8640             | 0.9294          | 0.8169          | 0.9779   | 0.9081  |
| LSTM           | 0.8247             | 0.8129          | 0.8970          | 0.8060   | 0.8386  |

Table 10. $R^2$ of solar radiation in summer.

|                | without Clustering | with Clustering |
|----------------|--------------------|-----------------|
|                | Cluster 1          | Cluster 2       | Cluster 3       | Average |
| BP             | 0.9500             | 0.9590          | 0.9981          | 0.9548   | 0.9706  |
| RBF            | 0.9357             | 0.9294          | 0.9977          | 0.8169   | 0.9147  |
| ELM            | 0.9538             | 0.9436          | 0.9405          | 0.9968   | 0.9603  |
| LSTM           | 0.8576             | 0.8742          | 0.8176          | 0.8514   | 0.8477  |

Table 11. $R^2$ of solar radiation in autumn.

|                | without Clustering | with Clustering |
|----------------|--------------------|-----------------|
|                | Cluster 1          | Cluster 2       | Cluster 3       | Average |
| BP             | 0.9733             | 0.9916          | 0.9055          | 0.9313   | 0.9428  |
| RBF            | 0.9366             | 0.8777          | 0.8713          | 0.9888   | 0.9126  |
| ELM            | 0.9437             | 0.9249          | 0.9916          | 0.8651   | 0.9272  |
| LSTM           | 0.8554             | 0.8454          | 0.9196          | 0.7981   | 0.8543  |

Table 12. $R^2$ of solar radiation in winter.

|                | without Clustering | with Clustering |
|----------------|--------------------|-----------------|
|                | Cluster 1          | Cluster 2       | Cluster 3       | Average |
| BP             | 0.9692             | 0.9364          | 0.8636          | 0.9982   | 0.9327  |
| RBF            | 0.8769             | 0.9230          | 0.8687          | 0.9973   | 0.9299  |
| ELM            | 0.9155             | 0.9558          | 0.8030          | 0.9970   | 0.9186  |
| LSTM           | 0.8193             | 0.8309          | 0.8289          | 0.8877   | 0.8291  |
Due to the added cluster analysis, the four data sets of spring, summer, autumn and winter are divided into three clusters with different irradiation intensities. At the same time, the similarity of samples in each cluster is relatively high in general. It can be seen from the aforementioned experimental results that the clustering strategy does improve the prediction accuracy. This observation can be explained by the reasoning that data preprocessing and sample partitions have a favorable impact on short-term solar radiation prediction. Ultimately, through analyzing the prediction results of various artificial neural networks, the proposed methods have indeed improved the prediction accuracy on the whole. These experimental results mean that the hybrid models of machine learning have advantages to some extent.

6. Conclusions and Outlook

This paper proposes a comprehensive application of machine learning techniques for short-term solar radiation prediction. Firstly, aiming at the missing entries in solar radiation data, a matrix completion method is used to recover them. Then we denoise the completed data by robust principal component analysis. The denoised data is clustered into low, medium and high intensity types via fusing sparse subspace representation and k-nearest-neighbor. Subsequently, four commonly used neural networks (BP, RBF, ELM and LSTM) are adopted to predict the solar radiation. In order to quantitatively verify the performance of the prediction model, the RMSE and MAE indicators are applied for model evaluation. The experimental results show that the hybrid model can improve the solar radiation predication accuracy.

In future research work, we will try to improve the model in the following respects to enhance its prediction ability. A multi-step forward prediction is necessary in practice, and it is urgent to develop the corresponding forecasting models by an ensemble of machine learning techniques and signal decomposition methods. In the procedure of establishing the prediction model, the input meteorological element used in this paper is only global horizontal irradiance. In fact, there are many other elements that affect solar radiation, such as the variation of daily temperature and precipitation. The influence of multiple elements on solar radiation will be considered and analyzed so as to improve the prediction ability of solar radiation. Furthermore, this paper only merges a few machine learning techniques into the forecast of solar radiation. In particular, deep learning models have a powerful representative ability and their further application in forecasting solar radiation will be very prospective.

Author Contributions: Conception and design of the experiments: J.S., L.W.; Performance of the experiments: L.W.; Writing—original draft preparation: L.W.; Writing—review and editing: J.S., L.W.; Supervision: J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to their confidentiality.

Acknowledgments: This work is partially supported by the National Key R&D Program of China (2018YFB1502902) and the Natural Science Basic Research Plan in Shaanxi Province of China (2021JM-378).

Conflicts of Interest: The authors declare that they have no conflict of interest.
References

1. Duffie, J.A.; Beckman, W.A.; Blair, N. Solar Engineering of Thermal Processes; John Wiley & Sons: Hoboken, NJ, USA, 2013; pp. 569–582.

2. Qazi, A.; Fayaz, H.; Wadi, A.; Raj, R.G.; Rahim, N.A.; Khan, W.A. The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review. J. Clean. Prod. 2015, 104, 1–12. [CrossRef]

3. Yagüi, G.M.; Yang, D.; Srinivasan, D. Automatic hourly solar forecasting using machine learning models. Renew. Sustain. Energy Rev. 2019, 105, 487–498. [CrossRef]

4. Kleniewska, M.; Mitrowska, D.; Wasilewicz, M. Estimating daily global solar radiation with no meteorological data in Poland. Appl. Sci. 2020, 10, 778. [CrossRef]

5. Blal, M.; Seyfallah Khelifi, S.; Rachid Dabou, R. A prediction models for estimating global solar radiation and evaluation meteorological effect on solar radiation potential under several weather conditions at the surface of Adrar environment. Measurement 2020, 152, 107348. [CrossRef]

6. Ogliari, E.; Dolara, A.; Manzolini, G.; Leva, S. Physical and hybrid methods comparison for the day ahead PV output power forecast. Renew. Energy 2017, 113, 11–21. [CrossRef]

7. Başaran, K.; Bozyiğit, F.; Siano, P.; Taer, P.Y.; Kılınç, D. Systematic literature review of photovoltaic output power forecasting. IET Renew. Power Gener. 2021, 14, 3961–3973. [CrossRef]

8. Arif, B.M.; Hanafi, L.M. Physical reviews of solar radiation models for estimating global solar radiation in Indonesia. Energy Rep. 2020, 6, 1206–1211.

9. Paulescu, M.; Paulescu, E. Short-term forecasting of solar irradiance. Renew. Energy 2019, 143, 985–994. [CrossRef]

10. Huang, X.Q.; Li, Q.; Tai, Y.H.; Chen, Z.Q.; Zhang, J.; Shi, J.S.; Gao, B.X.; Liu, W.M. Hybrid deep neural model for hourly solar irradiance forecasting. Renew. Energy 2021, 171, 1041–1060. [CrossRef]

11. Nam, S.B.; Hur, J. A hybrid spatio-temporal forecasting of solar generating resources for grid integration. Energy 2019, 177, 503–510. [CrossRef]

12. Zhang, Y.; Li, Y.T.; Zhang, G.Y. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. Energy 2020, 213, 118371. [CrossRef]

13. Zang, H.X.; Liu, L.; Sun, L.; Cheng, L.L.; Wei, Z.N.; Sun, G.Q. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. Renew. Energy 2020, 160, 26–41. [CrossRef]

14. Schulz, B.; Ayari, M.E.; Lerch, S.; Baran, S. Post-processing numerical weather prediction ensembles for probabilistic solar irradiance forecasting. Sol. Energy 2021, 220, 1016–1031. [CrossRef]

15. Bakker, K.; Whan, K.; Knap, W.; Schmeits, M. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. Sol. Energy 2019, 191, 138–150. [CrossRef]

16. Verbois, H.; Huva, R.; Rusydi, A.; Walsh, W. Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. Sol. Energy 2018, 162, 265–277. [CrossRef]

17. Chen, J.L.; He, L.; Yang, H.; Ma, M.H.; Chen, Q.; Wu, S.J.; Xiao, Z.L. Empirical models for estimating monthly global solar radiation: A most comprehensive review and comparative case study in China. Renew. Sustain. Energy Rev. 2019, 108, 91–111. [CrossRef]

18. Zheng, J.Q.; Zhang, H.R.; Dai, Y.H.; Wang, B.H.; Zheng, T.C.; Liao, Q.; Li, J.; Song, X. Time series prediction for output of multi-region solar power plants. Appl. Energy 2020, 257, 114001. [CrossRef]

19. David, M.; Ramahatana, F.; Trombe, P.J.; Lauret, P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. Sol. Energy 2016, 133, 55–72. [CrossRef]

20. Lee, J.H.; Wang, W.; Harrou, F.; Sun, Y. Reliable solar irradiance prediction using ensemble learning-based models: A comparative study. Energy Convers. Manag. 2020, 208, 112582. [CrossRef]

21. Voyant, C.; Nottton, G.; Kalogirou, S. Machine learning methods for solar radiation forecasting: A review. Renew. Energy 2017, 105, 569–582. [CrossRef]

22. Gabriel, N.; Felipe, L.G.; Bressan Michael, B.; Andres, P. Machine learning for site-adaptation and solar radiation forecasting. Renew. Energy 2020, 167, 333–342. [CrossRef]

23. Pang, Z.H.; Niu, F.X.; O’Neill, Z. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. Renew. Energy 2020, 156, 279–289. [CrossRef]

24. Ayodele, T.R.; Ogunjuyigbe, A.S.O.; Amedu, A.; Munda, J.L. Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms. Renew. Energy Focus 2019, 29, 78–93. [CrossRef]

25. Panamtash, H.; Zhou, Q.; Hong, T.; Qu, Z.H.; Davis, K.O. A copula-based Bayesian method for probabilistic solar power forecasting. Sol. Energy 2020, 196, 336–345. [CrossRef]

26. Xue, X.H. Prediction of daily diffuse solar radiation using artificial neural networks. Int. J. Hydrog. Energy 2017, 42, 28214–28221. [CrossRef]

27. Al-Dahidi, S.; Ayadi, O.; Adeeb, J.; Alrbai, M.; Qawasmeh, B.R. Extreme learning machines for solar photovoltaic power predictions. Energies 2018, 11, 2725. [CrossRef]
29. Huynh, A.N.L.; Deo, R.C.; An-Vo, D.A.; Ali, M. Near real-time global solar radiation forecasting at multiple time-step horizons using the long short-term memory network. *Energies* 2020, 13, 3517. [CrossRef]
30. Sharma, A.; Kakkar, A. Forecasting daily global solar irradiance generation using machine learning. *Renew. Sustain. Energy Rev.* 2018, 82, 2254–2269. [CrossRef]
31. Lan, H.; Zhang, C.; Hong, H.H.; He, Y.; Wen, S.L. Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. *Appl. Energy* 2019, 247, 389–402. [CrossRef]
32. Hamid Mehdipour, S.; Tenreiro Machado, J.A. Cluster analysis of the large natural satellites in the solar system. *Appl. Math. Model.* 2021, 89, 1268–1278. [CrossRef]
33. Wang, K.J.; Qi, X.X.; Liu, H.D. A comparison of day-ahead photovoltaic power models based on deep learning neural network. *Appl. Energy* 2019, 251, 113315. [CrossRef]
34. Sun, S.L.; Wang, S.Y.; Zhang, G.W.; Zheng, J.L. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. *Sol. Energy* 2018, 163, 189–199. [CrossRef]
35. Majumder, I.; Dash, P.K.; Bisoi, R. Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting. *Energy Convers. Manag.* 2018, 171, 787–806. [CrossRef]
36. Mazumder, R.; Saldana, D.; Weng, H.L. Matrix completion with nonconvex regularization: Spectral operators and scalable algorithms. *Stat. Comput.* 2020, 30, 1113–1138. [CrossRef]
37. Shi, J.R.; Zheng, X.Y.; Zhou, S.S. Research progress in matrix completion algorithms. *Comput. Sci.* 2014, 41, 13–20.
38. Hu, Z.X.; Nie, F.P.; Wang, R.; Li, X.L. Low rank regularization: A review. *Neural Netw.* 2021, 136, 218–232. [CrossRef]
39. Shi, J.R.; Li, X.X. Meteorological data estimation based on matrix completion. *Meteorol. Sci. Technol.* 2019, 47, 420–425.
40. Shi, J.R.; Yang, W.; Zheng, X.Y. Robust generalized low rank approximations of matrices. *Entropy* 2015, 10, e0137028. [CrossRef]
41. Zhao, Q.; Meng, D.; Xu, Z. Robust principal component analysis with complex noise. In Proceedings of the 31st International Conference on Machine Learning ICML, Beijing, China, 21–26 June 2014; Volume 32, pp. 55–63.
42. Liu, L.; Gao, X.B.; Gao, Q.X.; Shao, L.; Han, J.G. Adaptive robust principal component analysis. *Neural Netw.* 2019, 119, 85–92. [CrossRef]
43. Dong, L.; Wang, L.J.; Khahro, S.F.; Gao, S.; Liao, X.Z. Wind power day-ahead prediction with cluster analysis of NWP. *Renew. Sustain. Energy Rev.* 2016, 60, 1206–1212. [CrossRef]
44. Luxburg, U.V. A tutorial on spectral clustering. *Stat. Comput.* 2004, 17, 395–416. [CrossRef]
45. Chen, W.F.; Feng, G.C. Spectral clustering with discriminant cuts. *Knowl. Based Syst.* 2012, 28, 27–37. [CrossRef]
46. Shi, J.R.; Yang, L. A climate classification of China through k-knearest-neighbor and sparse subspace representation. *J. Clim.* 2020, 33, 243–262. [CrossRef]
47. Filippone, M.; Camstra, F.; Masulli, F.; Rovetta, S. A survey of kernel and spectral methods for clustering. *Pattern Recognit.* 2008, 41, 176–190. [CrossRef]
48. Wang, W.W.; Xiao-Ping, L.I.; Feng, X.C. A survey on sparse subspace clustering. *Acta Autom. Sin.* 2019, 113, 1373–1384.
49. Elhamifar, E.; Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 2013, 35, 2765–2781. [CrossRef]
50. Zhou, Z.H.; Tian, B. Research on community detection of online social network members based on the sparse subspace clustering approach. *Future Internet* 2019, 11, 254. [CrossRef]
51. Wang, Z.; Wang, F.; Su, S. Solar Irradiance Short-Term Prediction Model Based on BP Neural Network. *Energy Procedia* 2011, 12, 488–494. [CrossRef]
52. Elsheikh, A.H.; Sharshir, S.W.; Elaziz, M.A.; Kabeel, A.E.; Wang, G.L.; Zhang, H.O. Modeling of solar energy systems using artificial neural network: A comprehensive review. *Sol. Energy* 2019, 180, 622–639. [CrossRef]
53. Huang, G.; Huang, G.B.; Song, S.J.; You, K.Y. Trends in extreme learning machines: A review. *Neural Netw.* 2015, 61, 32–48. [CrossRef]
54. Aybar-Ruiz, A.; Jiménez-Fernández, S.; Cornejo-Bueno, L.; Casanova-Mateo, C.; Sanz-Justo, J.; Salvador-González, P.; Salcedo-Sanz, S. A novel grouping genetic algorithm–extreme learning machine approach for global solar radiation prediction from numerical weather models inputs. *Sol. Energy* 2016, 132, 129–142. [CrossRef]
55. Jiang, X.W.; Yan, T.H.; Zhu, J.J.; He, B.; Li, W.H.; Du, H.P.; Sun, S.S. Densely connected deep extreme learning machine algorithm. *Cogn. Comput.* 2020, 12, 979–990. [CrossRef]
56. Naylani, H.W.; Maria, K.; Charalambides, A.G.; Angèle, R. Training and testing of a single-layer LSTM network for near-future solar forecasting. *Appl. Sci.* 2020, 10, 5873.
57. Qing, X.Y.; Niu, Y.G. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* 2018, 148, 461–468. [CrossRef]
58. Gao, B.X.; Huang, X.Q.; Shi, J.S.; Tai, Y.H.; Zhang, J. Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. *Renew. Energy* 2020, 162, 1665–1683. [CrossRef]