

Article

Hybrid Decomposition-Reconfiguration Models for Long-Term Solar Radiation Prediction Only Using Historical Radiation Records

Si-Ya Wang¹, Jun Qiu^{2,3,*} and Fang-Fang Li^{1,3,*} 

¹ College of Water Resources & Civil Engineering, China Agricultural University; Beijing 100083, China; wangsy0405@163.com

² State Key Laboratory of Hydrosience & Engineering, Tsinghua University, Beijing 100084, China

³ State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China

* Correspondence: aeroengine@mail.tsinghua.edu.cn (J.Q.); liff@cau.edu.cn (F.-F.L.);
Tel.: +86-138-1066-7527 (J.Q.) +86-138-1120-4521 (F.-F.L.)

Received: 8 April 2018; Accepted: 22 May 2018; Published: 29 May 2018



Abstract: Solar radiation prediction is significant for solar energy utilization. This paper presents hybrid methods following the decomposition-prediction-reconfiguration paradigm using only historical radiation records with different combination of decomposition methods, Ensemble Empirical Mode Decomposition (EEMD) and Wavelet Analysis (WA), and the reconfiguration methods, regression model (RE) and Artificial Neural Network (ANN). The application in west China indicates that these hybrid decomposition-reconfiguration models perform well for monthly prediction, while the comparisons of the daily prediction show that the hybrid EEMD-RE model has a higher degree of fitting and a better prediction effect in long-term prediction of solar radiation intensity, which verifies (1) decomposition of original solar radiation data results in components with regular characteristics; (2) the relationship between the original solar radiation sequence and the derived intrinsic mode functions (IMFs) is linear; and (3) EEMD has strong adaptivity for non-linear and non-stationary series. The proposed hybrid decomposition-reconfiguration models have great application prospect for monthly long-term prediction of solar radiation intensity, especially in the areas where complex climate data is difficult to obtain, and the EEMD-RE model is recommended for the daily long-term prediction.

Keywords: long-term prediction; solar radiation; hybrid model; decomposition-reconfiguration

1. Introduction

Solar energy is one of the most favorable renewable energy sources, it has been continuously explored in recent years. Solar radiation data is the fundamental input for solar energy applications, and its reliability appears important to designing, developing and evaluating solar technologies [1]. Optimal design of solar power systems needs the expected long-term solar radiation on the horizontal plane. For example, sizing the projects is related to solar collector and PV systems [2]. Moreover, when solar energy is produced on large-scale and grid-connected, an accurate knowledge of long-term solar radiation makes a lot of sense for balancing the energy supply and demand [3].

Various researches on solar radiation forecasting methods have been reported, classified into physical models and statistical methods. Physical models are based on the physical state and dynamic motions of the atmosphere, also known as Numerical Weather Prediction (NWP) models [4], which was believed the most appropriate for day-ahead and “multi-day forecast horizons [5]. However, the NWP models are greatly affected by weather factors, such as cloudiness, cloud evolution and optical properties in the forecast area [6]. Generally, such models result in good predictions in clear

sky conditions, while with the effect of clouds, the prediction results become worse [7]. Besides, the application of such physical models [8] on long-term daily solar radiation prediction is also limited by their computational complexity. There are two types of statistical models: mathematical statistics and machine learning algorithms. Mathematical statistics mainly includes regression analysis [9], time series analysis [10], fuzzy theory [11], wavelet analysis [12] and Kalman filtering [13]. Regression analysis determines the best combination of the independent variables to predict the dependent variable, but the selection procedure is not always easy [14]. Nourani found that Auto Regressive Integrated Moving Average (ARIMA) model had a limited ability to capture non-stationarities and non-linearities [15]. In practice, the prediction accuracy of the statistical methods is not as high as the NWP models, as the parameters change over time due to various factors [16]. Typical machine learning algorithms include: Artificial Neural Networks (ANN) [17], Support Vector Machines (SVM) [18] and heuristic intelligent optimization algorithms [19]. Gala et al. believed that hybrid artificial intelligence systems are quite effective for solar energy prediction [20]. Lauret et al. found that the improvement of the machine learning techniques for hour ahead solar forecasting appears to be more pronounced in case of unstable sky conditions [18].

As for long-term solar radiation prediction, a limited number of related publications can be found, most of which focus on the characteristic analysis rather than prediction. The complexity of the relationship between solar radiation and meteorological, terrestrial, and extra-terrestrial variables makes it difficult to make long-term solar radiation prediction [21]. Coelho and Boaventura-Cunha [22] found even their proposed method combining support vector regression and Markov chains performed poorly when the prediction was sixty-step ahead after comparing linear autoregressive, nonlinear autoregressive, and support vector models on long-term solar prediction.

With the development of the big data-mining technology in recent decades, the machine learning algorithms have drawn much attention. As one of the most commonly used methods, ANN have been successfully applied to solar radiation prediction and solar systems design [3] since it has strong ability to solve non-linear function estimation, pattern detection and data sorting. Cao [16,23] predicted solar radiation in Shanghai and Baoshan by using a BP (back propagation) neural network after preprocessing the data with wavelet analysis, and found that the recursive BP network combined with wavelet analysis improves in both speed and accuracy. Paoli et al. [24] used mixed models to predict total daily solar radiation in three sites in France. They first used the seasonal index adjustment method to preprocess the original solar radiation sequence, and then applied daily multi-layer perceptive neural network (MLPNN) on daily solar radiation prediction. Amrouche and Le Pivert [25] used spatial models and ANNs to predict the daily solar radiation intensity at four US sites. Pedro and Coimbra [26] compared ARIMA, k-nearest neighbors (kNNs), ANN and Genetic Algorithm (GA) optimized neural networks (GA/ANN). It was found that the neural network optimized by GA is superior to other algorithms in hourly prediction. Khatib et al. [27] compared existing methods including linear, nonlinear and ANN models and pointed out that compared with linear and nonlinear models, ANNs are more accurate to predict solar energy. At the same time, it was found that the sunshine ratio, the ambient temperature and the relative humidity are the most relevant coefficients for predicting solar radiation. Yadav and Chandel [28] chose different ANN models based on different geographical locations for prediction, and found that the reasonable choice of model parameters had a great influence on the prediction results. Voyant et al. [29] found that the predictive effects of these methods were affected by the weather and seasonal factors by comparing the ARIMA model, an ANN using only endogenous inputs (univariate) of pretreatment and an ANN using both endogenous and exogenous inputs for pretreatment. Ozgoren et al. [30] used the ANN model based on Multiple Nonlinear Regression (MNL) to predict the monthly average solar radiation in Turkey. The method requires the input of latitude, longitude, altitude, monthly temperature and monthly minimum temperature, maximum temperature, average temperature, soil temperature, relative humidity, wind speed, rainfall, barometric pressure, vapor pressure, cloud cover and sunshine duration and other variables, and the MNL method is used to determine the most appropriate independent input

variables. Koca et al. [31] applied ANN model to the prediction of the monthly mean solar radiation in the Mediterranean region of Anatolia in Turkey by inputting different parameters, and found that the number of the input parameters was the most effective parameter. Generally, the existing ANN model needs a lot of meteorological parameters when applied to radiation prediction to make the results more accurate [3]. The input parameters are basically a certain combination of meteorological and topographical data, which include day of the year, wind speed, rainfall, relative humidity, temperature, latitude, longitude, altitude and so on. [32–34]. Thus there exists great limitation when applying ANNs in some areas where meteorological data is hard to obtain.

As to the long-term time sequence itself, the information in historical data needs to be explored fully. The Wavelet Analysis (WA) and the Ensemble Empirical Mode Decomposition (EEMD) are two typical decomposition methods to extract the regular components from a fluctuant time series.

WA was developed on the basis of the Fourier Transform (FT) in the early 1980s, overcoming the shortcomings of traditional spectral analysis methods and satisfying the local variation requirements in the time and frequency domains by a variable window [35]. Almasria et al. [36] applied WA to the empirical study of Swedish temperature data from 1850 to 1999. Kisi [37] predicted monthly runoff using wavelet regression instead of ANN. Nourani et al. [15] combined WA and ANN to predict the runoff in the Ligvanchai valley of Tabriz, Iran. Partal [38] conducted a reference evapotranspiration estimation using the wavelet transform and the feedforward neural network methods to evaluate climate data (temperature, solar radiation, wind speed, relative humidity) at two stations in the United States.

The EEMD is an improved version of the empirical mode decomposition (EMD) [39]. EEMD overcomes the essential defect of EMD modalities and is an adaptive data processing method adapted to nonlinear and nonstationary time series. EMD and EEMD have been widely used in some complex system models. Monjoly et al. [40] compared the data processing methods, EMD, EEMD and WA, using classical prediction model (Auto-Regression, AR) and nonlinear method to predict solar radiation intensity and found that the multi-step prediction hybrid approach led to additional improvements.

In this study, an attempt to rollingly predict long-term solar radiation by only using historical radiation data is carried out. WA and EEMD are firstly used to decompose the historical daily sequence of solar radiation into regular and predictable sub sequences, and then the relationships between these sub components and the original sequence are established by Regression Equation or ANN model. Different combination of the decomposition methods and the relational models are tested, including EEMD-RE, EEMD-ANN, WA-RE, WA-ANN. The Autoregressive Integrated Moving Average model (ARIMA) is also compared. The results show that the EEMD-RE model performs superior to the other ones, which is capable of capturing the main characteristics of solar radiation in the next year. With daily data of ten years, the monthly means prediction almost has the same accuracy as the published studies using diverse meteorological and topographic data. The method can be employed for the study and design of solar projects, particularly in underdeveloped areas where it is difficult to obtain complex data.

The rest of the paper is organized as follows: the method used in this study is explained in Section 2. Simulative experiments with different methods are presented in Section 3. Section 4 contains the comparison results. Sections 5 and 6 present the discussion and conclusion, respectively.

2. Methodology

2.1. Empirical Mode Decomposition (EMD)

The EMD is efficient to analyze non-linear and non-stationary signal sequences with high signal-to-noise ratio, which decomposes a complex signal into a finite number of intrinsic mode functions (IMF) with local characteristics of different time scales.

Each IMF needs to meet the following two requirements: (1) Throughout the data sequence, the number of extremums and zero values across the entire sample dataset must be equal or differ

by one; (2) the mean of the envelope formed by the local maximum and the local minimum is zero at any point of the sequence. Taking signal $s(t)$ as an example, the process of screening programs is summarized as follows:

Step 1: Find all the local maxima and local minima in $s(t)$, and connect all local maxima by a cubic spline line to configure the upper envelope; This process is repeated with a local minima to produce the lower envelope.

Step 2: Construct the mean envelope $m_1(t)$ with the average of the upper and lower envelopes.

Step 3: The average envelope is subtracted from the original signal $s(t)$ to derive the first component $h_1(t)$:

$$h_1(t) = s(t) - m_1(t) \quad (1)$$

Step 4: Check if $h_1(t)$ meets the IMF's conditions. If not, go back to Step 1 and use $h_1(t)$ as the original signal for the second screening:

$$h_2(t) = h_1(t) - m_2(t) \quad (2)$$

Repeat screening for k times, until $h_k(t)$ meets the IMF's conditions, when the first IMF component $c_1(t)$ is derived:

$$c_1(t) = h_k(t) \quad (3)$$

Step 5: Subtract $c_1(t)$ from the original signal $s(t)$ to get the residual $r_1(t)$:

$$r_1(t) = s(t) - c_1(t) \quad (4)$$

Step 6: Take $r_1(t)$ as the new original signal, and perform step 1 to step 5 to obtain a new residual $r_2(t)$. Repeat the steps above for n times. When the n th residual $r_n(t)$ becomes a monotonic function, the IMF cannot be decomposed anymore and the entire EMD is completed. The original signal $s(t)$ can be expressed as a combination of n IMF components and an average trend component $r_n(t)$, as shown in Equation (5):

$$s(t) = \sum_{k=1}^n c_k(t) + r_n(t) \quad (5)$$

With the Hilbert transform, the IMFs yield instantaneous frequencies as functions of time that give sharp identifications of imbedded structures. Each IMF can be either linear or non-linear with corresponding physical background.

2.2. Ensemble Empirical Mode Decomposition (EEMD)

Although the EMD shows great superiority in the analysis of non-linear and non-stationary signals, the mode mixing problem resulting from the intermittency of signals still exists. The EEMD adds white Gaussian noise to the EMD to solve such problem. The basic idea is to eliminate the intermittency of the original signal in the frequency domain by using the statistical characteristics of uniformly distributed Gauss white noise, so that the mode mixing can be avoided.

The specific decomposition steps of EEMD are as follows:

Step 1: A series of random Gauss white noise signals $w_i(t)$ are added to the original signal $s(t)$ to get a total signal $X(t)$:

$$X_i(t) = s(t) + kw_i(t) \quad (6)$$

where $w_i(t)$ indicates the total signal after the i th time adding noise. k is the amplitude coefficient of $w_i(t)$, usually $0.05 < k < 0.5$.

Step 2: Decompose $X_i(t)$ in accordance with step 1 to 6 in Section 2.1. However, it's necessary to replace spline interpolation with piecewise cubic Hermite interpolation in the first step of 2.1 section to obtain the maximum and minimum envelope of signal:

$$X_i(t) = \sum_{j=1}^n c_{ij}(t) + r_i(t) \quad (7)$$

where $c_{ij}(t)$ represents the j th IMF after the noise is added i times.

Step 3: To obtain the average values of all IMF and residuals obtained by the above steps, as in Equations (8) and (9):

$$c_j(t) = \sum_{i=1}^M c_{ij}(t) / M \quad (8)$$

$$r_j(t) = \sum_{i=1}^M r_i(t) / M \quad (9)$$

where $c_j(t)$ and $r_j(t)$ stand for the j th IMF and j th residual component obtained by EMD technique. M denotes the number of the Gaussian white noise, usually $M = 100$.

2.3. Wavelet Analysis (WA)

The WA is a time-frequency localization analysis method with fixed time-frequency window area but changing time window and frequency window. Through the wavelet transforming of the original data sequence and mapping it to a different time-frequency domain, the inverse transforming of each frequency-domain component can be obtained. The separate analysis of these components helps understand their variation law in different frequency domains. Select the mother wavelet $Y(t)$, where t stands for time, and the wavelet sequence $Y_{j,k}(t)$ can be obtained by expanding and transforming $Y(t)$. In computation and practical application, a discrete wavelet sequence is usually used, which can be obtained by Equation (10):

$$Y_{j,k(t)} = A_0^{-j/2} Y \left[\frac{t - kA_0^j B_0}{A_0^j} \right] = A_0^{-j/2} (A_0^{-j} - kB_0) \quad j, k = 0, \pm 1, \pm 2, \dots \quad (10)$$

where A_0^{-j} is a scale factor, $kA_0^j B_0$ is a shift factor. When $A_0 = 2, B_0 = 1$, the above formula is a binary wavelet sequence. Let $\varphi(t)$ be the scaling function corresponding to the mother wavelet $Y(t)$; then the sequence of the binary functions $\varphi_{j,k(t)}$ is:

$$\varphi_{j,k(t)} = 2^{-\frac{j}{2}} \varphi(2^{-j}t - k) \quad j, k = 0, \pm 1, \pm 2, \dots \quad (11)$$

After the decomposition of the original data $f(m)$, the corresponding low-frequency series a_N and the high-frequency series d_1, d_2, \dots, d_N can be obtained. The specific relationship is as follows:

$$f(m) = d_1 + d_2 + \dots + d_N + a_N \quad (12)$$

The results of the wavelet decomposition vary according to the chosen mother wavelet; the resulting frequency domain alias also has different degrees. The more severe the alias in the frequency domain, the less obvious the variation of the components in the frequency domain. Therefore, the selection of the mother wavelet should be excluded from the frequency domain caused by the phenomenon of alias serious mother wavelet.

2.4. Back Propagation-Artificial Neural Network (BP-ANN)

Due to its strong ability of non-linear mapping, learning as well as fault tolerance, ANNs have been widely applied to nonlinear forecasting problems. The Kolmogorov continuity theorem guarantees the feasibility and validity of using neural networks for time series prediction mathematically. BP-ANN back propagation includes input layer, hidden layer and output layer. The existence theorem of Kolmogorov three-layer neural network has proved that any continuous function can be mapped to a three-layer BP network.

The output E_n of a neuron j on the BP-ANN hidden layer and output layer is given by Equation (13):

$$E_n = f_j(Net_j) = f_j\left(\sum w_{ij}e_i + \theta_j\right) \quad (13)$$

where f_j is the activation function corresponding to the neuron j , and usually Sigmoid function $f(x) = 1/(1 + \exp(-x))$ is adopted; θ_j represents the threshold of the neuron j ; e_i is the input of neuron j ; w_{ij} indicates the connection weight of the corresponding input and the neuron.

2.5. Regression Model (RE)

Generalized linear models (GLM) are a unified class of regression methods for discrete and continuous response variables. There are some special cases, such as Logistic regression for binary responses, linear regression for continuous responses, log-linear models for counts, and some survival analysis methods. The systematic component and the random component compose a GLM. For the systematic component, one relates Y to x by assuming the average among individuals with a common value of x , $\eta = \lambda(Y)$, satisfying:

$$g(\eta) = x_1\alpha_1 + x_2\alpha_2 + \dots + x_r\alpha_r \quad (14)$$

where g is a prespecified function known as the 'link function'. α are regression coefficients.

In this study, the linear regression model is selected, the coefficients of which are determined by the least square method.

2.6. Hybrid EEMD/WA-RE Model

The basic idea of the hybrid EEMD/WA-RE method proposed in this study follows the decomposition-prediction-reconfiguration paradigm. The main purpose of EEMD or WA decomposition is to better extract valid information from the data, simplify the original goal, and decompose it to more regular components for predictable sub-goals. First, the EEMD or the WA is used to decompose the long disordered sequence into several sub-sequences (IMFs for EEMD, and sequences with different frequencies for WA). Theoretically, EEMD can be applied to any time series without the requirements for stationarity, and does not require the default basic functionality. For WA, the key technique to alleviate the aliasing phenomenon is the selection of the mother wavelet. In this study, db7 is set as the mother wavelet according to the previous study and testing results [41].

The daily average solar radiation sequence of 1~T year is decomposed into sub-sequences by EEMD or WA, which are used as the independent variables in RE, and the data of 2~(T + 1) year is taken as the dependent variable. The regression equation g is then established to predict the daily average solar radiation.

2.7. Hybrid EEMD/WA-ANN Model

The sub-sequences obtained by EEMD or WA using radiation data of 1~T years can also be used as the input to the ANN model, and the data of 2~(T + 1) year is the output. After training the ANN model, it can be used for prediction in the future.

The decomposition-prediction-reconfiguration idea derives four different combination of hybrid models in this study: EEMD-RE, WA-RE, EEMD-ANN, and WA-ANN, which are compared and

evaluated. Figure 1 shows the flowchart of these four models. The step 1 and 2 with black circles aim at training the model, and establishing the relationship between $X_{1\sim T}$ and $X_{2\sim(T+1)}$; while step 3 to 5 with blue circle use such model to predict X_{T+2} . The part with blue background indicates predicting process in Figure 1.

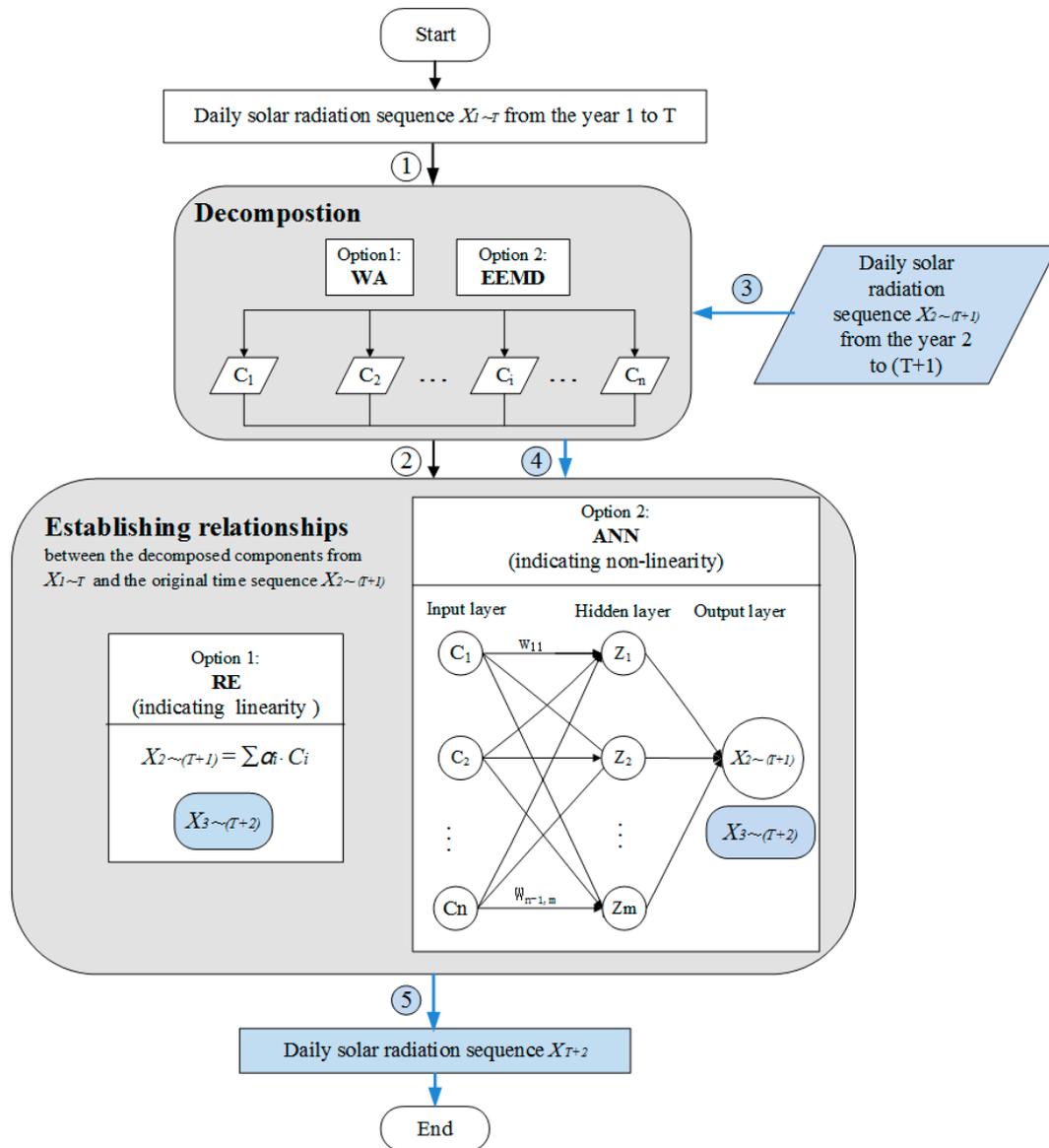


Figure 1. Flowchart of hybrid EEMD-RE, WA-RE, EEMD-ANN, and WA-ANN models.

3. Case Study

3.1. Study Case

The Qinghai province is located in west China with an average elevation of above 3000 m. It has good atmospheric transparency, high sunlight transmittance, long sunshine duration and abundant solar energy resources. The annual sunshine hours in eastern Qinghai Province are 3000 to 3200 h, and the annual solar radiation is 5860 to 6700 MJ/m² [42], ranking the second in the country. The whole Qinghai province has about 200,000 km² unused desert, which is suitable for the large-scale solar energy exploration [43].

In the past 10 years, the solar energy industry in Qinghai province has been developing vigorously, with a speed of 'one watt per watt'. By the end of 2017, the installation capacity of the photovoltaic (PV) power in the Qinghai Province had reached 7910 MW [44], and more projects are planned for construction.

The design of a PV power station needs accurate long-term radiation prediction. Gonghe County in Qinghai province, where a large scale PV power station is planned, is taken as the research area in this study. Solar radiation intensity data used in this experiment was obtained from NASA. The sample data is from 1st January 1984 to 31st December 1995.

3.2. Implementation

Using EEMD, daily average solar radiation intensity data of the area from 1 January of the year 1984 to 31 December of the year 1993 are decomposed to obtain 12 IMFs. The 12 IMFs are taken as independent variables, and the data from 1 January of the year 1985 to 31 December of the year 1994 is taken as dependent variables to establish the regression equation, as in Equation (15):

$$Y = \sum_{i=1}^n \zeta_i C_i \quad (15)$$

where ζ_i are the regression coefficients and C_i is the IMFs. Equation (15) is then used to predict the solar radiation of the year 1995.

The 12 IMFs derived from EEMD using the data from 1984 to 1993 can also be taken as the input to train an ANN model, and the data of the year 1985 to 1994 is the output. The number of hidden layer neurons of the ANN model in this study is 10 and the output layer neurons is 1. After training the BP-ANN model, it is used to predict the daily radiation sequence of the year 1995 with the data of the IMFs from 1985 to 1994.

Taking db7 as the mother wavelet, the three-scale Mallat pyramid wavelet decomposition of the solar radiation data series is carried out to obtain the low frequency sequence a_3 and the high frequency series $d1$, $d2$ and $d3$ of the solar radiation. Then the similar process is carried out as the EEMD-RE model and the EEMD-ANN model to establish the regression equation and the ANN model to predict the radiation for 1995.

Another typical prediction method for the ARIMA time series is also tested for comparison. The ARIMA model $(3,0,4) \times (0,0,1)$ is chosen after auto regression, partial regression and unit root test for daily data, while the ARIMA model $(5,0,5)$ is chosen for monthly data.

To compare the predictive effect in different time scales, the daily, ten-day, and monthly results are calculated with the daily prediction. On the other hand, to verify the data mining effect by decomposition methods, the monthly data is also used for the four hybrid models to derive the monthly prediction, which is compared with the monthly statistics from daily predictions.

3.3. Model Evaluation Criteria

The standard root mean square error (RMSE), the mean absolute percentage error (MAPE), the correlation coefficient (r) and the coefficient of determination (R^2) are chosen as the evaluation criteria of the predictive value, as defined in Equations (16)–(19). RMSE reflects the extent to which the predicted data deviates from the true value. The smaller the RMSE value, the better the prediction. MAPE can be used to measure the quality of a model prediction; the smaller the MAPE value, the better the prediction. r and R^2 reflect the fitting degree of the model; the closer the r and R^2 to 1, the better the fitting degree of the model:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_{\text{hist},i} - X_{\text{pred},i})^2} \quad (16)$$

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^n \left| \frac{X_{hist,i} - X_{pred,i}}{X_{hist,i}} \right| \right) \times 100\% \tag{17}$$

$$r = \frac{\sum_{i=1}^n (X_{hist,i} - \bar{X}_{hist})(X_{pred,i} - \bar{X}_{pred})}{\sqrt{\sum_{i=1}^n (X_{hist,i} - \bar{X}_{hist})^2 (X_{pred,i} - \bar{X}_{pred})^2}} \tag{18}$$

$$R^2 = \frac{\sum_{i=1}^n (X_{pred,i} - X_{hist,i})^2}{\sum_{i=1}^n (X_{hist,i} - \bar{X}_{hist})^2} \tag{19}$$

where the subscript *hist* represents historical data, and the subscript *pred* represents the predictive results.

4. Results

Figure 2 shows the 12 IMFs obtained from the EEMD decomposition and the subsequences with different frequencies obtained from the WA using daily solar radiation intensity from the year 1984 to 1993, and 1985 to 1994, respectively; Figure 3 shows the 7 IMFs and sequences with different frequencies derived from monthly data from 1984 to 1993, and from 1985 to 1994, respectively.

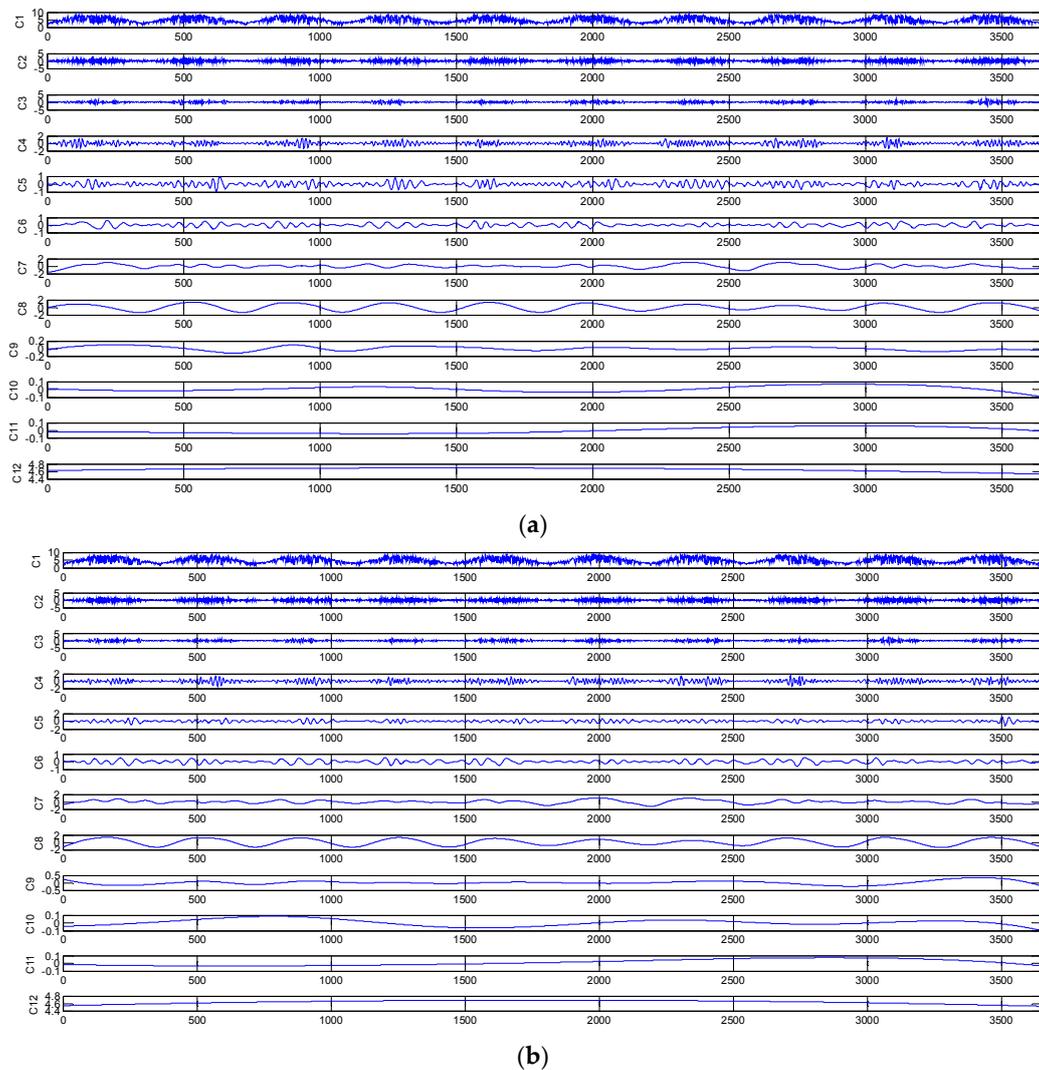
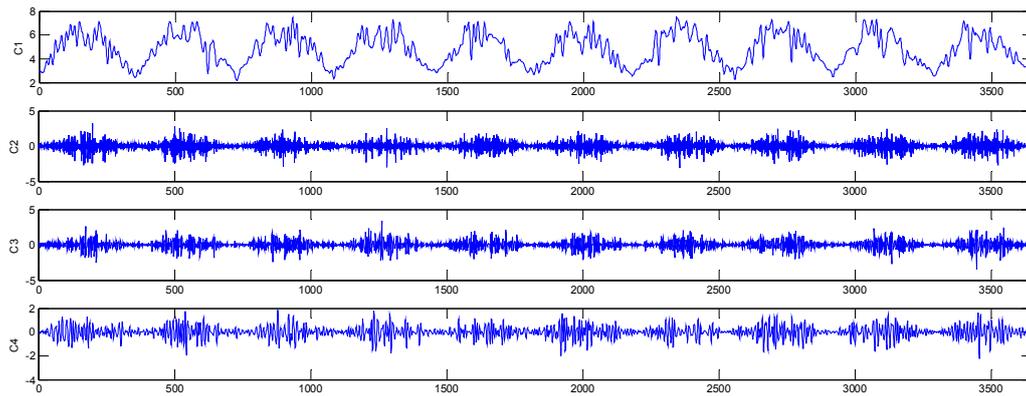
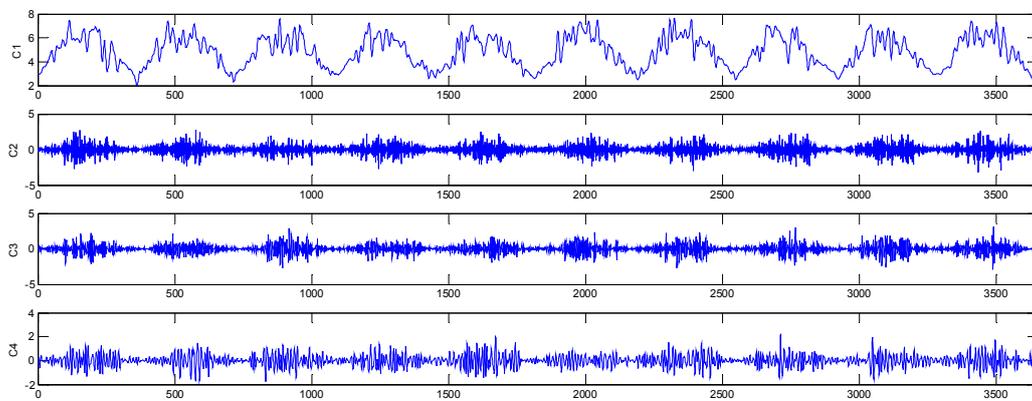


Figure 2. Cont.

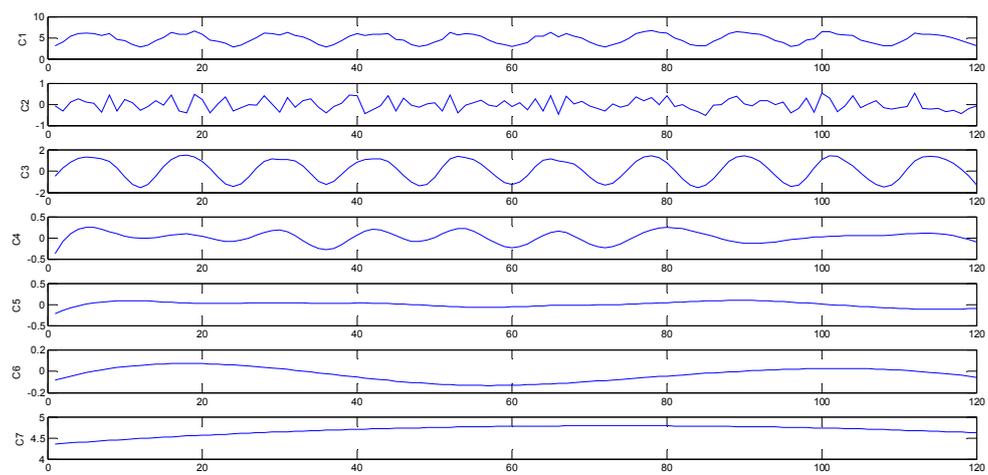


(c)



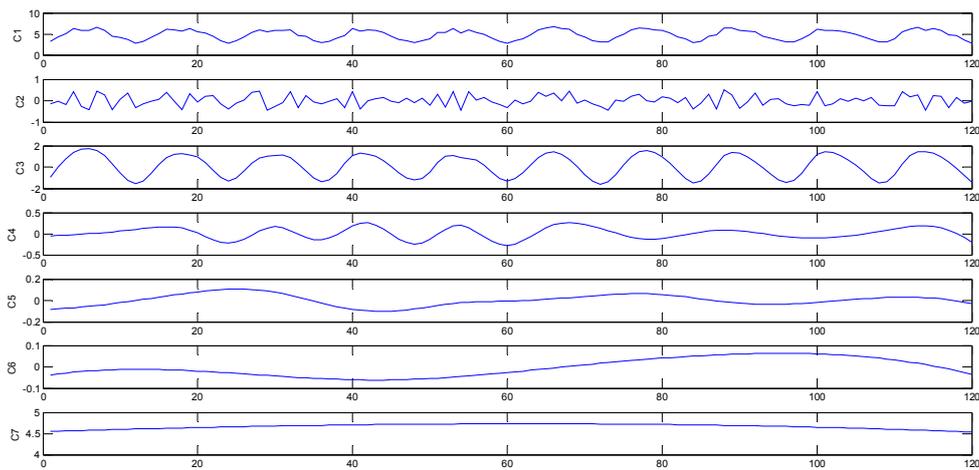
(d)

Figure 2. IMFs obtained from EEMD using daily solar radiation data of (a) 1984–1993, (b) 1985–1994; and sub sequences with different frequencies derived from WA using daily solar radiation data of (c) 1984–1993, (d) 1985–1994.

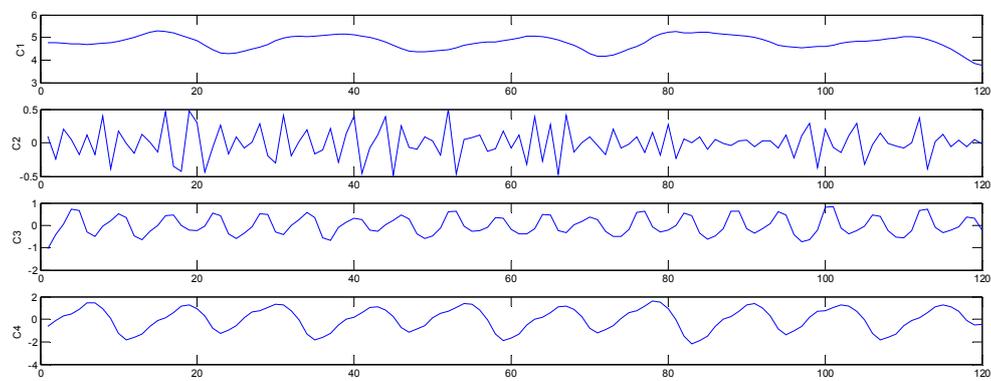


(a)

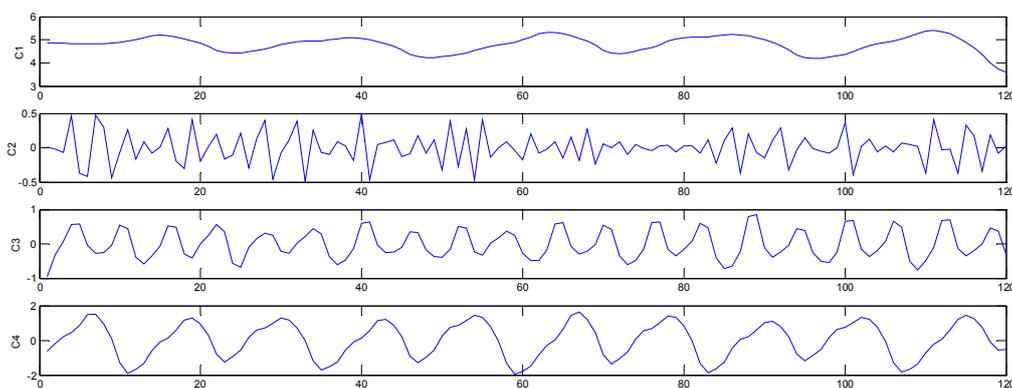
Figure 3. Cont.



(b)



(c)



(d)

Figure 3. IMFs obtained from EEMD using monthly solar radiation data of (a) 1984–1993, (b) 1985–1994, sequences with different frequencies obtained from WA using monthly solar radiation data of (c) 1984–1993, (d) 1985–1994.

Figure 4 shows the daily prediction results by different prediction models, as well as the its statistical results with a time step of 10 days and 1 month.

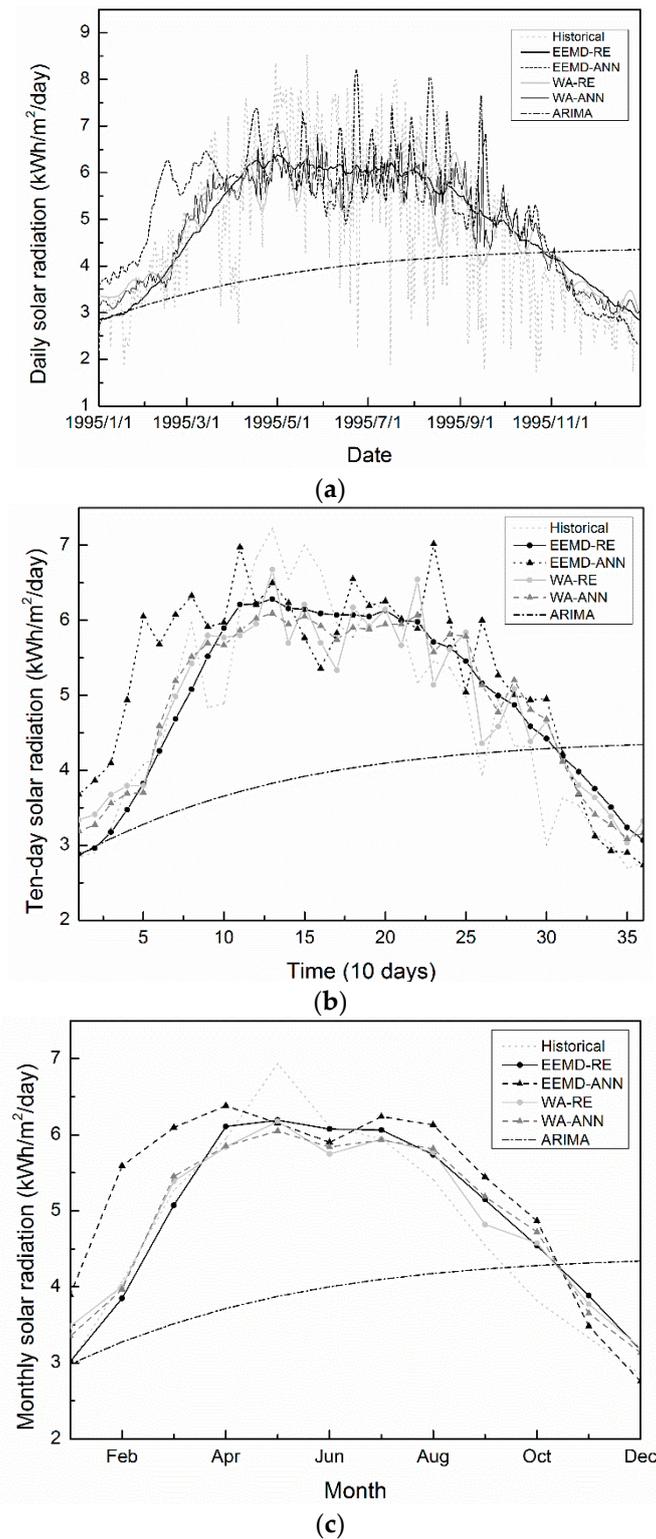


Figure 4. (a) Daily prediction results of the year 1995, and its statistical results with a time step of (b) 10 days and (c) 1 month.

The predictive accuracy of the different models and time scales are shown in Tables 1–4. Important cases where the computational definition of R^2 can yield negative values, depending on the definition used, arise where the predictions that are being compared to the corresponding outcomes have not been

derived from a model-fitting procedure using those data, and where linear regression is conducted without including an intercept [43]. Additionally, negative values of R^2 may occur when fitting non-linear functions to data [44]. In cases where negative values arise, the mean of the data provides a better fit to the outcomes than do the fitted function values, according to this particular criterion [45,46].

Table 1. Predictive accuracy of different models with daily data.

Methods	RMSE	MAPE (%)	r	R^2
EEMD-RE	1.135	22.11	0.748	0.5484
EEMD-ANN	1.474	29.47	0.590	0.2387
WA-RE	1.181	22.58	0.723	0.5116
WA-ANN	1.188	22.56	0.725	0.5052
ARIMA	1.948	33.21	0.035	-0.3097

Table 2. Predictive accuracy of statistical results using daily data with 10 days interval.

Methods	RMSE	MAPE (%)	r	R^2
EEMD-RE	0.571	10.23	0.918	0.8247
EEMD-ANN	0.935	16.92	0.812	0.5297
WA-RE	0.637	12.03	0.897	0.7810
WA-ANN	0.619	11.34	0.904	0.7913
ARIMA	1.666	27.60	0.060	-0.4945

Table 3. Predictive accuracy of statistical results using daily data with 1 month interval.

Methods	RMSE	MAPE (%)	r	R^2
EEMD-RE	0.417	4.25	0.957	0.8973
EEMD-ANN	0.775	14.87	0.883	0.6454
WA-RE	0.416	8.07	0.970	0.8979
WA-ANN	0.467	8.49	0.950	0.8712
ARIMA	1.607	27.01	0.040	0.5256

Table 4. Predictive accuracy of different models with monthly data.

Methods	RMSE	MAPE (%)	r	R^2
EEMD-RE	0.339	6.00	0.979	0.9319
EEMD-ANN	0.362	5.66	0.966	0.9226
WA-RE	0.305	5.23	0.980	0.9450
WA-ANN	0.377	6.83	0.959	0.9161
ARIMA	0.368	6.79	0.967	0.9199

5. Discussion

The decomposition results shown in Figure 2 show that there are 12 IMFs derived from EEMD, while four sub-sequences derived from WA as a comparison. It can be inferred that EEMD has stronger ability in mining more sufficient information with regularity. The comparison of the decomposed results by EEMD and WA with monthly data in Figure 3 also shows that the IMFs from EEMD for different time series seems more regular than the sub sequences from WA in Figure 3c,d, which vary distinctly for different years. It can be seen from Figure 4 that the predictive result by EEMD-RE model is the most stable one compared with other models, which indicates that EEMD-RE model can capture the stable information in the data instead of paying attention to the uncertainties. The statistical results in Tables 1–3 also confirm that the EEMD-RE model has the highest predictive accuracy with daily data compared to other models, with a smaller RMSE and higher model fitting degree. The comparison between the EEMD-RE and EEMD-ANN implies that the relationship between the original

solar radiation sequence and the derived IMFs is linear. Thus the superiority of ANNs for complex non-linear problem does not work for solar radiation data. The comparison between EEMD and WA verifies the strong adaptivity of EEMD for non-linear and non-stationary series; while the WA relies greatly on the mother wavelet, while it may lead to virtual fluctuations. Especially in Figure 4c, which shows that the four kinds of hybrid models, including EEMD-RE, EEMD-ANN, WA-RE, and WA-ANN all perform well when predicting monthly solar radiation for the next year with historical daily data. Such results verify the validity and effectiveness of the idea that decomposing time series into sub sequences with more regularity is helpful for long-term prediction. Interestingly, the monthly predictive effects with monthly data in Figure 5 and Table 4 are better than those with daily data. Although we thought more information could be explored for data with smaller time intervals, it seems that more randomness and uncertainty were introduced for daily data compared to monthly data, and some errors in the prediction with shorter time interval might be smoothed in the statistical process for the longer time interval. The model fitting degree of the ARIMA model is low using daily data, indicating that it is not suitable for long-term prediction with large amount of data. The proposed EEMD-RE model is thus recommended for long-term solar radiation predictions.

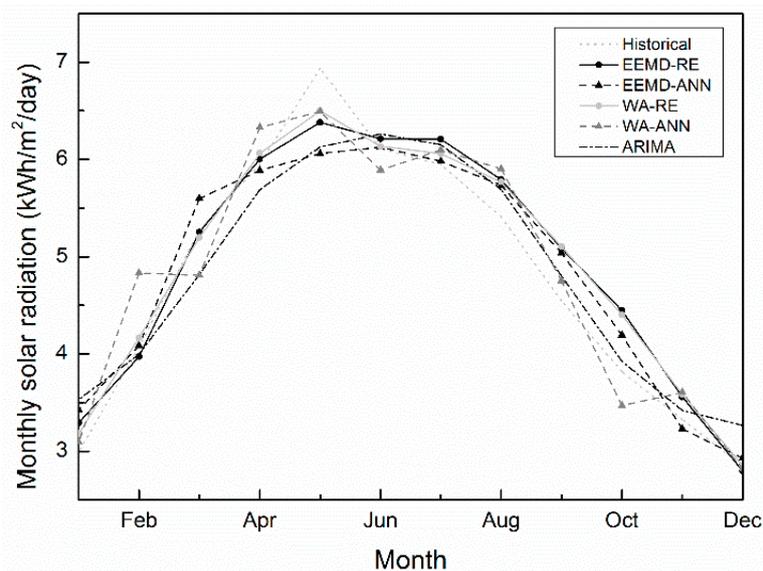


Figure 5. Predictive solar radiation results of the year 1995 using monthly data.

6. Conclusions

The solar radiation forecast is important for solar energy utilization. The causes of variations in solar radiation are various. There exists a complicated coupling relationship between the solar radiation intensity and the meteorological elements and terrain factors, but the data of complicated climate conditions is often difficult to obtain.

In this paper, hybrid methods following the decomposition-prediction-reconfiguration paradigm are proposed with different combination of EEMD, WA, RE, and ANN, which is only based on historical solar radiation data. The application on the west of China shows that basically these hybrid decomposition-reconfiguration models perform well for monthly prediction using monthly historical data; while for the daily prediction, the EEMD-RE model outperforms other models, since (1) the decomposition results in components with regular characteristics; (2) the relationship between the original solar radiation sequence and the derived IMFs is linear; and (3) the EEMD has strong adaptivity for non-linear and non-stationary series. The proposed hybrid decomposition-reconfiguration models only relying on the historical radiation records have great practical value for long-term prediction of solar radiation intensity, especially in the areas where complex climate data is difficult to obtain.

Author Contributions: F.-F.L. and J.Q. conceived and designed the experiments; S.-Y.W. performed the experiments; S.-Y.W. and F.-F.L. analyzed the data; J.Q. contributed reagents/materials/analysis tools; S.-Y.W. and F.-F.L. wrote the paper.

Acknowledgments: This research was supported by National Key R&D Program of China (2017YFC0403600, 2017YFC0403602), the Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University.(Grant No. 2016-KF-03), and the Open Research Fund Program of State key Laboratory of Hydrosience and Engineering (Grant No. sklhse-2016-B-03).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bulut, H.; Buyukalaca, O. Simple model for the generation of daily global solar-radiation data in Turkey. *Appl. Energy* **2007**, *84*, 477–491. [[CrossRef](#)]
2. Kaplanis, S.; Kumar, J.; Kaplani, E. On a universal model for the prediction of the daily global solar radiation. *Renew. Energy* **2016**, *91*, 178–188. [[CrossRef](#)]
3. 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](#)]
4. Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76. [[CrossRef](#)]
5. Perez, R.; Lorenz, E.; Pelland, S.; Beauharnois, M.; Van Knowe, G.; Hemker, K.; Heinemann, D.; Remund, J.; Muller, S.C.; Traunmuller, W.; et al. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Sol. Energy* **2013**, *94*, 305–326. [[CrossRef](#)]
6. Mathiesen, P.; Collier, C.; Kleissl, J. A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting. *Sol. Energy* **2013**, *92*, 47–61. [[CrossRef](#)]
7. Zamora, R.J.; Dutton, E.G.; Trainer, M.; McKeen, S.A.; Wilczak, J.M.; Hou, Y.T. The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction. *Mon. Weather Rev.* **2005**, *133*, 783–792. [[CrossRef](#)]
8. Wang, F.; Mi, Z.Q.; Su, S.; Zhao, H.S. Short-term solar irradiance forecasting model based on Artificial Neural Network using statistical feature parameters. *Energies* **2012**, *5*, 1355–1370. [[CrossRef](#)]
9. Trapero, J.R.; Kourentzes, N.; Martin, A. Short-term solar irradiation forecasting based on Dynamic Harmonic Regression. *Energy* **2015**, *84*, 289–295. [[CrossRef](#)]
10. Voyant, C.; Paoli, C.; Muselli, M.; Nivet, M.L. Multi-horizon solar radiation forecasting for Mediterranean locations using time series models. *Renew. Sustain. Energy Rev.* **2013**, *28*, 44–52. [[CrossRef](#)]
11. Chen, S.X.; Gooi, H.B.; Wang, M.Q. Solar radiation forecast based on fuzzy logic and neural networks. *Renew. Energy* **2013**, *60*, 195–201. [[CrossRef](#)]
12. Mellit, A.; Benghanem, M.; Kalogirou, S.A. An adaptive wavelet-network model for forecasting daily total solar-radiation. *Appl. Energy* **2006**, *83*, 705–722. [[CrossRef](#)]
13. Akarlan, E.; Hocaoglu, F.O.; Edizkan, R. A novel M-D (multi-dimensional) linear prediction filter approach for hourly solar radiation forecasting. *Energy* **2014**, *73*, 978–986. [[CrossRef](#)]
14. Cevik, A. Unified formulation for web crippling strength of cold-formed steel sheeting using stepwise regression. *J. Constr. Steel Res.* **2007**, *63*, 1305–1316. [[CrossRef](#)]
15. Nourani, V.; Komasi, M.; Mano, A. A Multivariate ANN-Wavelet Approach for Rainfall-Runoff Modeling. *Water Resour. Manag.* **2009**, *23*, 2877–2894. [[CrossRef](#)]
16. Cao, S.H.; Cao, J.C. Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. *Appl. Ther. Eng.* **2005**, *25*, 161–172. [[CrossRef](#)]
17. Sharifi, S.S.; Rezaverdinejad, V.; Nourani, V. Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperature-based approaches. *J. Atmos. Sol. Terr. Phys.* **2016**, *149*, 131–145. [[CrossRef](#)]
18. Lauret, P.; Voyant, C.; Soubdhan, T.; David, M.; Poggi, P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Sol. Energy* **2015**, *112*, 446–457. [[CrossRef](#)]
19. Wang, J.Z.; Jiang, H.; Wu, Y.J.; Dong, Y. Forecasting solar radiation using an optimized hybrid model by Cuckoo Search algorithm. *Energy* **2015**, *81*, 627–644. [[CrossRef](#)]

20. Gala, Y.; Fernandez, A.; Diaz, J.; Dorronsoró, J.R. Support Vector Forecasting of Solar Radiation Values. *Hybrid Artif. Intell. Syst.* **2013**, *8073*, 51–60.
21. Baser, F.; Demirhan, H. A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. *Energy* **2017**, *123*, 229–240. [[CrossRef](#)]
22. Coelho, J.P.; Boaventura-Cunha, J. Long term solar radiation forecast using computational intelligence methods. *Appl. Comput. Intell. Soft Comput.* **2015**, *2014*, 21. [[CrossRef](#)]
23. Cao, J.C.; Cao, S.H. Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. *Energy* **2006**, *31*, 3435–3445. [[CrossRef](#)]
24. Paoli, C.; Voyant, C.; Muselli, M.; Nivet, M.L. Forecasting of preprocessed daily solar radiation time series using neural networks. *Sol. Energy* **2010**, *84*, 2146–2160. [[CrossRef](#)]
25. Amrouche, B.; Le Pivert, X. Artificial neural network based daily local forecasting for global solar radiation. *Appl. Energy* **2014**, *130*, 333–341. [[CrossRef](#)]
26. Pedro, H.T.C.; Coimbra, C.F.M. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Sol. Energy* **2012**, *86*, 2017–2028. [[CrossRef](#)]
27. Khatib, T.; Mohamed, A.; Sopian, K. A review of solar energy modeling techniques. *Renew. Sust. Energy Rev.* **2012**, *16*, 2864–2869. [[CrossRef](#)]
28. Yadav, A.K.; Chandel, S.S. Solar radiation prediction using Artificial Neural Network techniques: A review. *Renew. Sust. Energy Rev.* **2014**, *33*, 772–781. [[CrossRef](#)]
29. Voyant, C.; Muselli, M.; Paoli, C.; Nivet, M.L. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy* **2011**, *36*, 348–359. [[CrossRef](#)]
30. Ozgoren, M.; Bilgili, M.; Sahin, B. Estimation of global solar radiation using ANN over Turkey. *Expert Syst. Appl.* **2012**, *39*, 5043–5051. [[CrossRef](#)]
31. Koca, A.; Oztop, H.F.; Varol, Y.; Koca, G.O. Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. *Expert Syst. Appl.* **2011**, *38*, 8756–8762. [[CrossRef](#)]
32. Celik, O.; Teke, A.; Yildirim, H.B. The optimized artificial neural network model with Levenberg-Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Turkey. *J. Clean. Prod.* **2016**, *116*, 1–12. [[CrossRef](#)]
33. Chiteka, K.; Enweremadu, C.C. Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks. *J. Clean. Prod.* **2016**, *135*, 701–711. [[CrossRef](#)]
34. Renno, C.; Petito, E.; Gatto, A. ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building. *J. Clean. Prod.* **2016**, *135*, 1298–1316. [[CrossRef](#)]
35. Foufoula-Georgiou, E.; Kumar, P. (Eds.) *Wavelets in Geophysics*; Academic Press: San Diego, CA, USA, 1994.
36. Almasri, A.; Locking, H.; Shukur, G. Testing for climate warming in Sweden during 1850–1999, using wavelets analysis. *J. Appl. Stat.* **2008**, *35*, 431–443. [[CrossRef](#)]
37. Kisi, O. Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. *Hydrol. Process.* **2009**, *23*, 3583–3597. [[CrossRef](#)]
38. Partal, T. Modelling evapotranspiration using discrete wavelet transform and neural networks. *Hydrol. Process.* **2009**, *23*, 3545–3555. [[CrossRef](#)]
39. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
40. Monjoly, S.; Andre, M.; Calif, R.; Soubdhan, T. Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach. *Energy* **2017**, *119*, 288–298. [[CrossRef](#)]
41. Shuanghua, C.; Jiacong, C.; Fengqiang, L. Study of application of wavelet analysis to neural networks for the forecast of solar irradiance. *J. Donghua Univ.* **2004**, *6*, 18–22.
42. Meng, K.F.; Li, C.L.; Zhang, H.N.; Yang, L.B.; Zhang, S.C.; Wang, X.; Yang, J.; Li, Z.X. Analysis and discussion on related problems of solar energy resources development in Qinghai province. In Proceedings of the CSEE (Chinese Society for Electrical Engineering) Annual Meeting, Chengdu, China, 30 September 2015; pp. 1–3.
43. Yan, L.; Zhou, X.; Zhang, C.; Fei, W.Y.; Xia, X.C.; Zhou, F.Q.; Hu, X.H.; Xu, H.H.; Huang, C.G.; Liu, F.S.; et al. A proposal for planning and constructing a national integrated energy base combined with large-scale photo-voltaic power and hydropower in Qinghai province. *Adv. Technol. Electr. Eng. Energy* **2011**, *29*, 1–9.

44. Photovoltaic leads the development of new energy and new materials industry in Qinghai Province Available online: Available online: http://www.cpicorp.com.cn/zhxx/201803/t20180326_287170.htm (accessed on 26 March 2018).
45. Cameron, C.A.; Windmeijer, F.A.G. An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econ.* **1997**, *77*, 1790–1792. [[CrossRef](#)]
46. Muhammad, I. Coefficient of Determination. Available online: Itfeature.com (accessed on 29 April 2012).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).