

Article

A Hybrid Prediction Model for Solar Radiation Based on Long Short-Term Memory, Empirical Mode Decomposition, and Solar Profiles for Energy Harvesting Wireless Sensor Networks

Yujia Ge ^{1,2,*}, Yurong Nan ¹ and Lijun Bai ²¹ College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China; nyr@zjut.edu.cn² School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China; blj@mail.zjgsu.edu.cn* Correspondence: yge@mail.zjgsu.edu.cn

Received: 26 October 2019; Accepted: 11 December 2019; Published: 13 December 2019



Abstract: For power management in the energy harvesting wireless sensor networks (EH-WSNs), it is necessary to know in advance the collectable solar energy data of each node in the network. Our work aims to improve the accuracy of solar energy predictions. Therefore, several existing prediction algorithms in the literature are surveyed, and then this paper proposes a solar radiance prediction model based on a long short-term memory (LSTM) neural network in combination with the signal processing algorithm empirical mode decomposition (EMD). The EMD method is used to decompose the time sequence data into a series of relatively stable component sequences. For improving the prediction accuracy further by utilizing the current day solar radiation profile in one-hour-ahead predictions, similar solar radiation profile data were selected for training LSTM neural networks. Simulation results show that the hybrid model achieves better prediction performance than traditional prediction methods, such as the exponentially-weighted moving average (EWMA), weather conditioned moving average (WCMA), and only LSTM models.

Keywords: solar radiance; machine learning; time series prediction

1. Introduction

The energy harvesting technique is a promising approach for widening applications of wireless sensor networks (WSNs) in the Internet of Things (IoT) fields by breaking the power limitations and extending the lifetime of the whole network. Among the available energies that could be harvested, such as wind, solar power, thermoelectric, and piezoelectric, solar power is the most efficient and widely used form [1]. Because solar energy is not controllable but predictable, the average energy that can be obtained from solar energy varies periodically with the season and time, as shown in Figure 1a. The efficiency of solar energy is affected by factors such as geographical location, sun illumination time, and lighting trend. Figure 1b indicates that the daily light intensity fluctuates greatly due to the weather, and the patterns of the solar radiation curve in two adjacent days are completely different. On the morning of 8 January 2010, the weather pattern turned to overcast from cloudy, and in the afternoon, it changed to cloudy in contrast to the typical sunny day of 9 January 2010. The wireless sensor network for energy harvesting (EH-WSN) constantly collects these environmental energies and the remaining usable energy changes in regularity with time, which is not like the traditional WSNs where the node energy model is of continuously decreasing energy. Therefore, accurate energy prediction methods for each node have significant importance in EH-WSNs [2]. Time series prediction methods play a very important role in these practical engineering fields, such as energy and information technology [3].

Accurate prediction results could be further used to optimize energy utilization, such as making routing decisions and adjusting duty cycles [4].

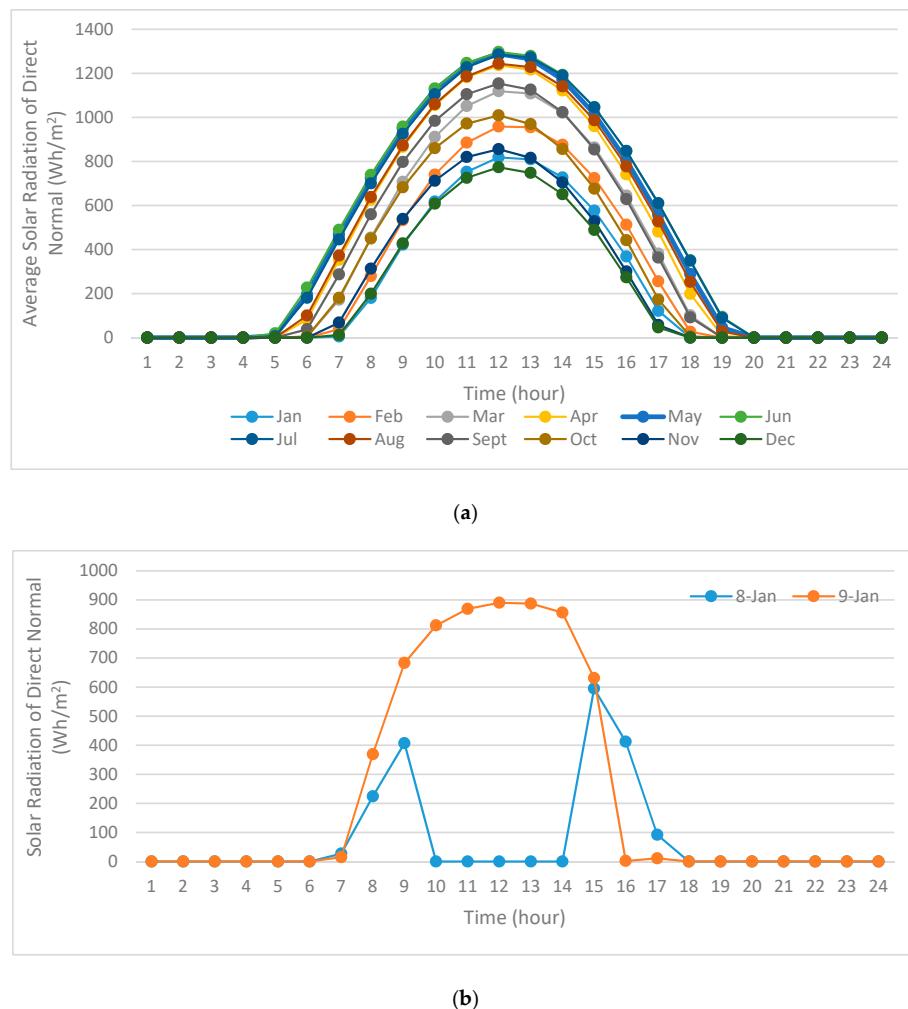


Figure 1. (a) Average solar radiation varied in months; (b) Different solar radiation on different days.

As a result, researches have been carried on the studies of solar radiation prediction algorithms. Prediction algorithms can make predictions in the presence of weather forecasting information or not. Since the weather forecasting information is not always available, our research focuses on prediction approaches without weather information. Under this category, the solar radiation prediction models are categorized into three major classes: statistical, stochastic, and machine learning methods [5]. Statistical models are based on statistical information, such as standard deviation, variance, mean, and moving average, which includes the classic exponential weighted moving average (EWMA) [6], the weather conditioned moving average (WCMA) [7] and their improvements. Stochastic models use various stochastic processes to represent signals, such as Markov chains. Machine learning prediction uses machine learning-based techniques, such as neural networks (NN) [8] and fuzzy logic (FL) [9], to build models to handle time series prediction. Machine learning prediction schemes are shown to outperform the traditional models by achieving increased accuracy but with a more substantial computational burden [5]. However, neural networks have two obvious weaknesses, i.e., slow convergence and the presence of local optima. The prediction error could be large if using a single neural network model, for example, long short-term memory (LSTM) [10]. To improve prediction accuracy, this paper takes advantage of empirical mode decomposition (EMD) to decompose the original signal into more stabilized components. Although theories of the EMD method are still under research,

such as end-effects, over-envelopes, under-envelopes, and modal confusion, the method has been widely used in the seismic signal analysis, marine signal analysis, mechanical fault diagnosis, and other fields [11–13]. This paper attempts to fuse these algorithms in solar radiation prediction field and build a hybrid model to improve the prediction accuracy. In an EH-WSN, there are different requirements of prediction horizon, from short-term prediction such as several-minute-ahead to one-hour-ahead, to medium-term such as one-hour-ahead to one-day-ahead, and long-term prediction such as several-day-ahead to one-year-ahead. Especially for short-term and medium-term prediction when the solar profile of that day is available, this profile can be utilized for improving the prediction accuracy further. Therefore, a solar profile selection method based on K-means clustering [14] is performed for better data training in LSTMs. The experimental simulation shows that this joint model has better prediction accuracy than other single models.

Our contributions in this paper can be summarized as follows:

- (1) A hybrid algorithm of prediction algorithms based on EMD and LSTM is proposed for improving the accuracy of prediction results by stabilizing elements of data through the EMD method.
- (2) On short-term and medium-term prediction, when the current day solar radiation profiles are available, solar profiles are classified by the K-means clustering method, and similar solar profiles can be retrieved to improve the prediction accuracy more efficiently.
- (3) Designed experiments and simulations are conducted to compare the proposed algorithm with existing popular algorithms, i.e., EWMA, WCMA, and single LSTM model on performance. Parameters for different models are tuned carefully. The prediction error rate is analyzed for different time slots in a day, as well.

The remaining paper is organized as follows. Section 2 reviews the related work on state-of-the-art prediction models and discusses their advantages and limitations. Section 3 introduces our proposed prediction method based on LSTM neural networks, EMD method, and solar profiles. Then Section 4 presents the designed simulation and comparison results of our method to three other methods. Finally, conclusion and future work are given in Section 5.

2. Related Work

In this section, we summarize the state-of-the-art prediction models. As we have mentioned earlier, solar radiation prediction models have statistical, stochastic, and machine learning methods [5]. Statistical models include the classic exponential weighted moving average (EWMA) [6], the weather conditioned moving average (WCMA) [7], and the profile-energy (Pro-Energy) model [15]. The autoregressive integrated moving average (ARIMA) and linear regression (LR), which fall into this category, were also used for solar prediction in [16]. A multivariate linear regression (MLR) analysis model was proposed to generate solar energy prediction with probabilities [17]. Stochastic models use stochastic processes, such as Markov chains, to represent signals. A first-order Markov chain model was developed in [18] for classifying global solar irradiation and generating predictions for photovoltaic systems. The accurate solar irradiance prediction model (ASIM) [19] uses increasing order Markov chains to predict solar energy in a long term prediction horizon. Although there are quite a lot of prediction methods on time series in general, we focus more heavily on typical prediction models on solar radiation in the wireless sensor network area, i.e., traditional EWMA, WCMA, Pro-Energy, and machine learning approaches, and review their advantages and limitations in detail.

2.1. Exponentially-Weighted Moving Average

The EWMA and its improved algorithms [6,20] are the most popular and commonly used algorithms for solar energy prediction. The EWMA algorithm divides one day into N fixed-length (usually 30 min) time slots. Its underlying principle is that the energy collected at a time duration on a certain day is assumed to be similar to the energy collected at that time duration on the previous day.

Therefore, in the EWMA, the predicted energy is the weighted average of the energy from the previous days, and the closer to the day, the greater the energy coefficient given in Equation (1).

$$E(d, n) = \alpha E(d - 1, n) + (1 - \alpha)H(d - 1, n) \quad (1)$$

where d represents the current date, and n represents the time slot number. The EWMA adds up the last harvested energy H and the estimated energy E according to the weighting factor α ($0 < \alpha < 1$).

The advantage of the EWMA is that it makes full use of the solar cycle and adapt to seasonal changes. When the weather has been in a stable state, such as continuous sunny days and cloudy days, the prediction error of the algorithm is extremely small. At the same time, the main disadvantage of the EWMA is its vulnerability to rapidly changing weather conditions. In particular, the EWMA produces significant prediction errors during mixed sunny and cloudy days. To reduce the error rate under unstable weather conditions, the current solar conditions should be integrated into the energy estimate.

2.2. Weather Conditioned Moving Average

The WCMA model [7,21] is a statistical-based algorithm designed to consider the current and past weather conditions. It collects the energy values of the past D days and stores them in the matrix $E(D, N)$, where N is the time slot in D day. The WCMA does not maintain a weighted average like the EWMA but instead incorporates the energy collected in the previous time slot into the prediction formula. The average of energy values in the certain time slot of previous days also contributes to the prediction equation. Therefore, the prediction equation for a particular time slot is related to the energy in the previous time slot, the average of the corresponding time slots and the current solar conditions are given in Equation (2).

$$E(d, n + 1) = \alpha E(d, n) + (1 - \alpha)M(d, n + 1)GAP(d, n, K) \quad (2)$$

where α is the weighted factor, and M represents the average of the $(n + 1)$ th slots in previous D days, $E(d, n)$ is the actual harvested energy in the last slot, and $GAP(d, n, k)$ is the value that reflects the current solar condition related to previous days, defined as Equation (3).

$$GAP = \frac{V \cdot P}{\sum P} \quad (3)$$

where the vector V is the ratio of the value to the average value of previous values, the vector P indicates the distance which means the closer the sample is, the bigger the weight it will be given. The UD-WCMA [22] is an improvement of the WCMA by adaptive tuning the weighting factor depending on the changes.

The peak of the error of the WCMA algorithm appears at the sunset and sunrise times and is more obvious when $\alpha > 0.5$. This is due to the fact that the WCMA takes into account the preceding time slots when predicting solar radiation. There are always dramatic changes in sunshine conditions during sunrise and sunset, so high errors are caused.

2.3. Profile-Energy Model

The principle of Pro-Energy [15] is to use a representative full-day energy harvesting profile to represent available energy. Each day is divided into N time slots. The vector with length N stores the energy collected on the day. Pro-Energy estimates the available energy for the next time slot by looking for the profile in the profile pool that is most similar to the weather of the day. The similarity of two different profiles is determined by calculating the Euclidean distance of the two vectors. The available energy for the next time slot is calculated from this most similar profile. Therefore, the combination of the energy observed in the previous slot and the energy of the most similar day helps predict the current energy, as shown in Equation (4).

$$E(d, n) = \alpha H + (1 - \alpha)E_{MS} \quad (4)$$

where H represents the amount of energy collected in the previous time slot and E_{MS} is the energy observed in time slot n on the most similar day. To determine the level of similarity from the previous few days to the day of D , the mean absolute error (MAE) of each stored day of the K previous time slots up to the current time slot is calculated in Pro-Energy as in Equation (5). Calculate the mean absolute error (MAE) with each stored profile and select the profile with the smallest MAE.

$$E^d = \min_{E^d \in E} \sum_{i=t-K}^t \frac{1}{K} |C_i - E_i^d| \quad (5)$$

where K is the number of previous time slots we use, C_i is the solar energy in time slot i of the current day C . When the MAE is above a set threshold, a new profile is stored into the database.

Pro-Energy tracks a typical set of previous profiles, each representing different solar conditions. The stored profile is dynamically updated to accommodate predictions for changing seasonal patterns. To further improve the accuracy of the forecast, Pro-Energy recommends combining multiple profiles instead of extracting values from the most similar days. In addition, as an analytical method, Pro-Energy can outperform the EWMA and WCMA for utilizing solar radiation profiles and overcoming their poor performance in dramatic weather changes. An improvement energy prediction model for lowering memory and energy usage in Pro-Energy is proposed as Pro-Energy-VLT (Profile energy prediction model with variable-length timeslots) [23], with varying length of timeslots according to the harvested energy.

2.4. Machine Learning Methods

Machine learning methods, such as neural networks (NN), fuzzy logic (FL), and reinforcement learning (RL), are introduced in prediction solar energy in the works of literature. A neural network model in [8] was proposed to predict solar radiance over a half-day time, which outperforms the autoregressive and fuzzy logic models. A hybrid model based on the generalized fuzzy model (GFM), which incorporated a Gaussian mixture model (GMM), was proposed in [24] for long-term prediction in solar energy. Deep learning methods are also incorporated in some research, such as an autoencoder-LSTM based model to predict solar energy [25]. The experiment results showed that deep learning algorithms outperform other artificial neural networks. Studies have been undertaken to compare LSTM with other machine-learning models in one-day-ahead prediction in solar radiance, and LSTM achieves the best performance overall [26].

Reinforcement learning has also been tried in solar power prediction research. A Q-learning based solar energy prediction algorithm (QL-SEP) is proposed as Equation (6) and is compared with other algorithms, such as EWMA, Pro-Energy, which shows QL-SEP outperforms other algorithms [27].

$$Q_{t+1}(s) = Q_t(s) + \gamma(r - Q_t(s)) \quad (6)$$

where s is the time slot, r is the reward calculated to be -1 or $+1$ according to the reliability of the prediction to the actual value, and γ is the learning rate. As shown in Equation (6), basically the Q-learning in this research uses value iteration to update the reliability of prediction accuracy in previous time slots. It does not show the suitableness by using the reinforcement learning method in a time series prediction problem.

Table 1 shows the solar prediction related methods in the literature review overall. More specifically, the EWMA and WCMA methods are fundamental solar radiation prediction methods in the wireless sensor network area. Pro-Energy and its improvement utilize the solar profiles to improve the approaches. Machine learning methods have been proved to achieve the best performance over traditional methods. Based on these facts, a hybrid solar radiation prediction method is proposed in the next section.

Table 1. Literature review of solar radiation prediction methods.

Category	Prediction Models	Description
Statistical	EWMA [6,20]	EWMA computes the predicted energy as the weighted average energy from the previous days, and the closer to the day, the greater the energy coefficient given.
	WCMA [7,21] and UD-WCMA [22]	WCMA predicts energy in a particular time slot being related to the energy in the previous time slot, the average of the corresponding time slots, and the current solar conditions. UD-WCMA is adaptive, tuning the weighting factor in WCMA.
	Pro-Energy [15] and Pro-Energy-VLT [23]	Pro-Energy uses energy in the previous time slot and the most similar profile selected from the profile pool for prediction. Pro-Energy-VLT is an improvement prediction model for lowering memory and energy usage in Pro-Energy with varying lengths of timeslots.
	ARIMA [16]	ARIMA is a general statistical time series prediction model also used as solar irradiation prediction method.
Stochastic	MLR [17]	A multivariate linear regression (MLR) analysis model is proposed to generate predictions of solar energy prediction with probabilities.
	ASIM [19]	ASIM uses increasing order Markov chains to predict the solar energy in a long term prediction horizon.
	First-order Markov chain approach [18]	A first-order Markov chain approach for classifying global solar irradiation and generating predictions for photovoltaic systems.
Machine learning	Generalized fuzzy model (GFM) + Gaussian mixture model (GMM) [13]	A hybrid model based on GFM incorporated with a GMM is proposed for long-term prediction in solar energy.
	Autoencoder-LSTM [25]	Autoencoder with LSTM based model is proposed to predict solar energy.
	QL-SEP [26]	QL-SEP is a Q-learning prediction model based on the prediction reliability of different time slots.

3. Hybrid Solar Radiation Prediction Method

To increase the accuracy of the solar prediction method, we propose the EMD model to stabilize the time-series information before the LSTM structure. In addition, having learned from the idea of Pro-Energy that using the solar energy profile of the current day could help to improve the accuracy of the prediction, we also utilize the profiles to increase the accuracy of the prediction and the similar solar radiation datasets as the training dataset for the LSTM model to improve the convergence and reduce the computation time.

Figure 2 shows the overall structure of our proposed method. The original signal is compared with stored solar radiation categories and selects the most similar category. These data in this category will later be trained for the LSTMs. After that, the EMD module is then applied to decompose the original file into the different components, and each component goes into the LSTM neural network, and the results are finally summed and reconstructed into the prediction result.

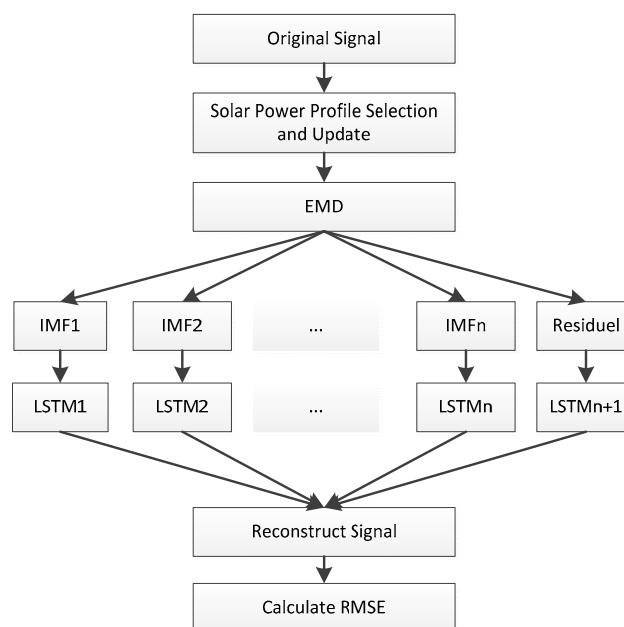


Figure 2. Structure of our proposed hybrid method.

3.1. Empirical Mode Decomposition

Solar radiance sequences are non-stationary time series with certain periodicity and randomness. Empirical mode decomposition (EMD) [28,29] is a method to process time series, which may be non-linear or non-stationary and decomposes the signal into some sequences which overcome the difficulty of selecting wavelet basis function in other transforming methods. The idea underlying EMD is that the time series need to be transformed when the number of minima or maxima is greater than the upper zero crossings or the number of zero crossings is two or more. The original data are then decomposed into several sub-sequences from a sifting process, and the sub-sequences are called the intrinsic mode function (IMF) component. The process of the EMD algorithm is shown in Figure 3. For any original signal $x(t)$, maximum and minimum points are identified. Then the upper and lower envelopes of the signal $u(t)$ and $l(t)$ are recognized, and the average of the envelopes are calculated as $m(t)$. The candidate IMF component from the envelope mean is obtained from subtracting $x(t)$ from $m(t)$, and whether it meets the criteria of an IMF is determined. If it meets the criteria of an IMF, it is considered as an IMF component, and the original signal is subtracted from this IMF component as the new $x(t)$ and continues from the beginning. Otherwise, the candidate IMF is considered as the new signal and repeats the loop from the beginning. Finally, the EMD decomposes the time sequence original data into a series of relatively stable IMF components and a residual.

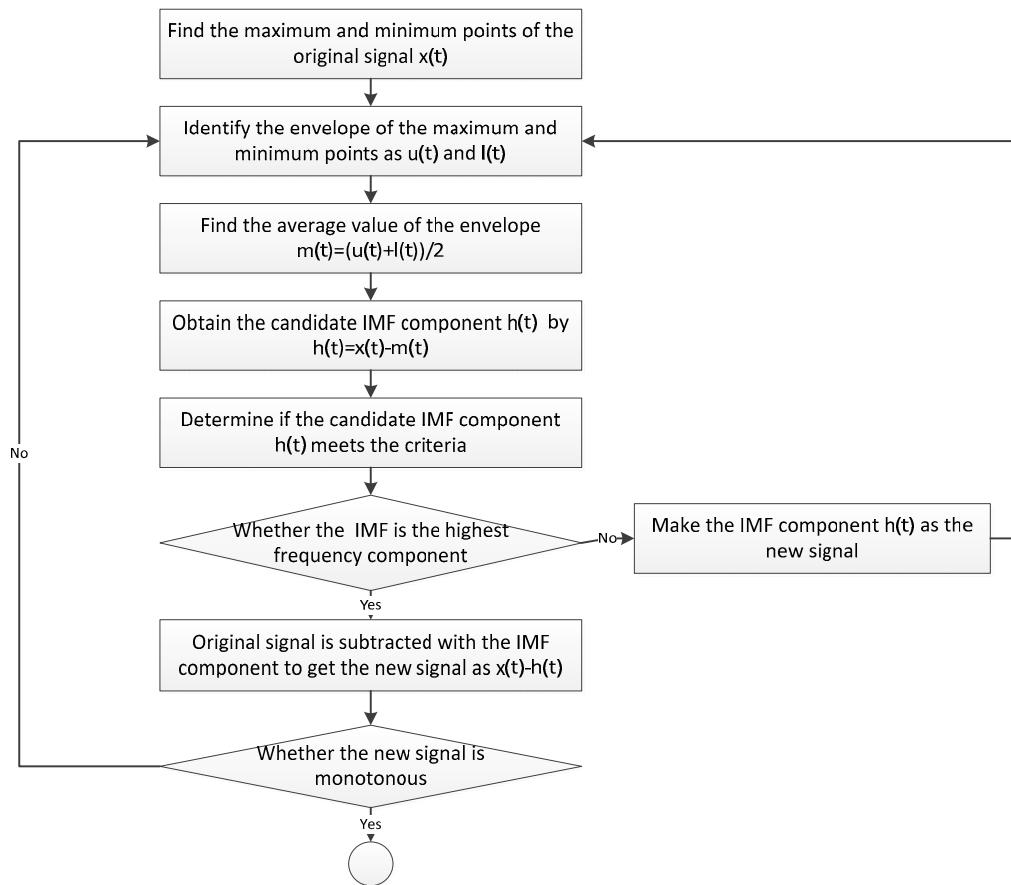


Figure 3. Process of the empirical mode decomposition (EMD) algorithm.

Figure 4 is an example of daily global horizontal solar radiation data from 1 January to 31 December 2008 in Alabama, which depicts the total amount of modeled direct and diffuse solar radiation received on a horizontal surface. The data are retrieved from the United States national solar radiation database [30]. Figure 5 shows the corresponding original, hourly data and the 10 extracted IMF components decomposed by the empirical mode decomposition method from high frequencies to low frequencies in order.

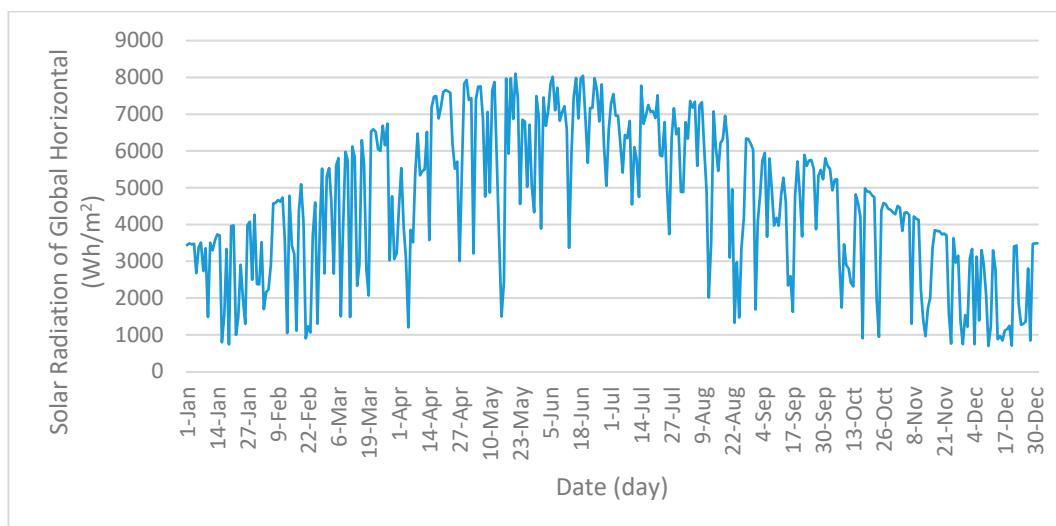


Figure 4. Typical daily solar radiation data from 1 January to 31 December 2008 in Alabama.

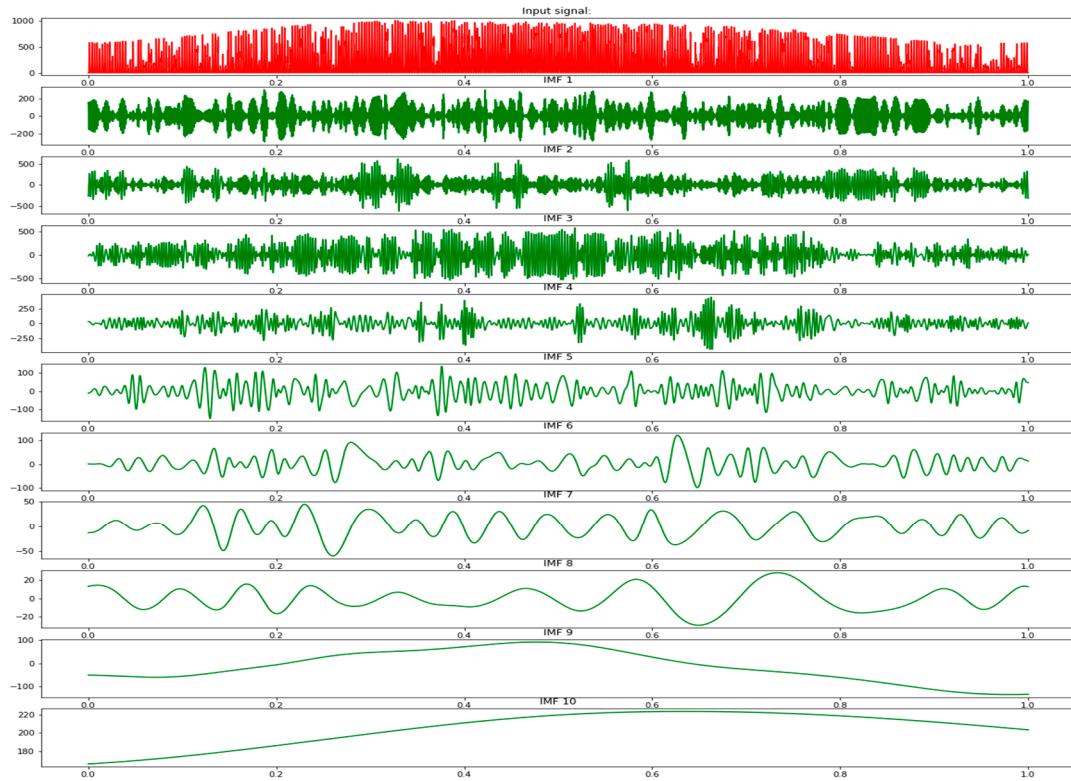


Figure 5. An original solar radiation signal and the results decomposed by the EMD method.

3.2. Long Short-Term Memory Networks

The LSTM networks proposed by Hochreiter et al. in 1997 [31,32] are based on recurrent neural networks (RNN) architecture. LSTM was mainly motivated and designed to mitigate the vanishing gradient problem of the standard RNN when dealing with long term dependencies and have been extensively applied in various fields. Moreover, LSTM is a popular time series forecasting model and can expertly deal with long-term dependencies data.

The LSTM model has a special structure called a memory cell, which includes the input gate, output gate, and forget gate. As shown in Figure 6, the gates control whether the information can go through or be got rid of. The activation functions of the gates are described in Equations (7)–(12).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (7)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (8)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (9)$$

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t, \quad (10)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (11)$$

$$h_t = o_t \times \tanh(C_t), \quad (12)$$

where f_t represents the forget gate, i_t represents the input gate, C_{t-1} and C_t represent the last cell state and the current cell state, respectively, o_t represents the output gate, h_{t-1} and h_t represent the output of the previous cell and current cell, respectively.

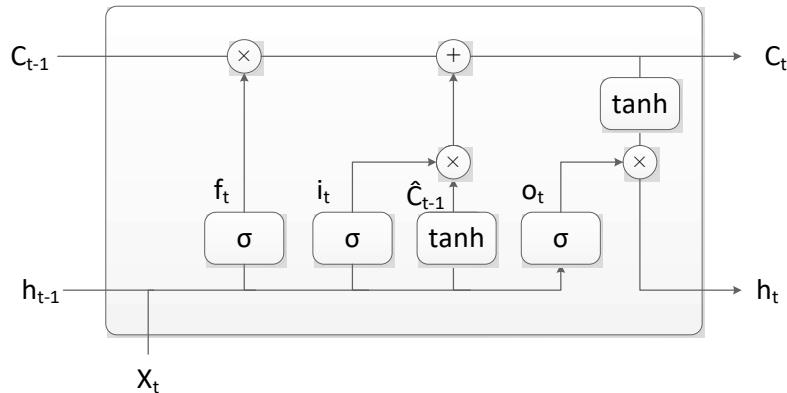


Figure 6. Long short-term memory (LSTM) cell structure.

Our LSTM network uses a multilayer architecture that consists of two LSTM layers and one fully connected layer. A series of IMFs decomposed by the EMD use this model to predict for each subsequence. The model finally combines the prediction results of each subsequence to obtain the final predicted value.

3.3. Energy Profile Selections by K-Means Clustering

Due to the different clarity situations of the sky and other weather conditions, solar radiation for each day obviously has different energy profiles. Figure 7 shows the values of global horizontal solar radiation in a continuous 5 days at the site of Alabama in 2008 from the United States national solar radiation database [30]. It shows the different energy profiles with changes in the amount of hourly solar radiation received on a horizontal surface due to the different weather conditions. From the research of Pro-Energy [15], it stores typical solar radiation profiles in the database which cover the clear sky and cloudy sky and dramatic changes in weather condition. The profile analyzer selects the most similar stored profile with the smallest mean absolute error (MAE) compared to the current day with each stored profile. The whole process can improve prediction accuracy based on EWMA. With similar consideration as in Pro-Energy, in medium-term prediction horizon, such as the one-hour-ahead prediction, the already occurred radiation profile of the current day could be utilized for improving our LSTM model. So for any particular dataset, we first use the K-means clustering method to classify the radiation profiles into specific N clusters where the K-means model is one of the most popular clustering algorithms [14]. The average of all the data in this cluster is called the centroid. The distance between each data and its centroid is calculated using a proximity measure, such as the Euclidean distance. Each data is then assigned to the closest centroid. The centroid of each cluster is updated based on the mean of data in that cluster. The assignment of data points to the closest cluster, and the updating of the centroids is repeated until no data points change their cluster, and the centroids remain the same.

Equation (13) shows the objective function F of K-means where N is the number of clusters, C_k is the centroid of k th cluster, and n is the number of data in one cluster, x_i is the i th data in one cluster.

$$F = \sum_{k=1}^N \sum_{x_i \in C_k}^n \|x_i - C_k\|^2 \quad (13)$$

For any historical solar radiation dataset, solar radiation data are divided by 24 h, so one sample of each day has 24 attributes. The number of N clusters is specified by users and need to be optimized during the simulation. When the current solar radiation profile is partially available, the current solar data are classified into these N categories.

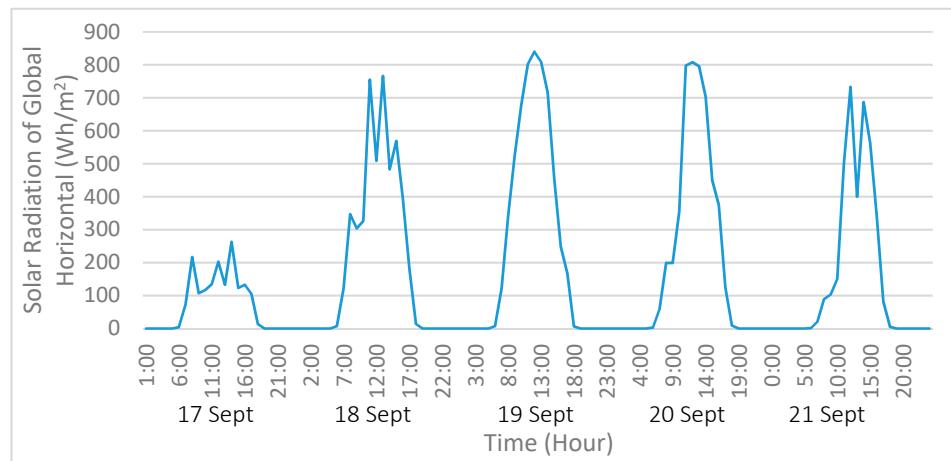


Figure 7. Different energy radiation profiles of 2008 in Alabama.

4. Performance Evaluation and Discussion

To show the performance of our proposed approach in predicting solar radiation, we designed experiments to evaluate these prediction models, including EWMA, WCMA, LSTM, and our model. We first start by explaining the chosen dataset and then show the steps of our experiments, including the tuning parameters, the performance results of different algorithms, and how well the algorithm performs when adding clustering solar profile data. All the algorithms are written in Python.

4.1. Datasets

Datasets are normally researched for validating and evaluating our model. The data from the United States national solar radiation database [30] contains comprehensive solar and meteorological related data in more than 1000 locations of the United States for the years 1999–2010. This solar radiation historical data have a one-hour sample rate over a whole year period, and normal statistics results are accessible. Lately, datasets are also available, which can be retrieved with smaller time duration and weather condition information. Data from three different locations, one from Michigan, one from Alabama, and the other from Nevada, were used in the experiments. Although the simulation can still not be exhaustive, selected various locations provide sufficient coverage of different solar radiation conditions.

4.2. Performance Metrics

For measuring prediction accuracy, since root mean square error (RMSE) and mean absolute percentage error (MAPE) are the most commonly used and have typical application areas, both metrics were chosen to evaluate the experimental results. The RMSE as Equation (14) was chosen as the metric to evaluate the performance between these prediction models in our experiments.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (14)$$

where y_i is the actual value, \hat{y}_i is the prediction value, and N is the number of tested data. In the RMSE, where errors are squared before averaging, it may give a relatively high weight to abnormal points. Mean absolute percentage error (MAPE) was also used as a metric defined as in Equation (15) when analyzing the prediction accuracy in different time slots.

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (15)$$

4.3. Tuning Parameters in LSTM

The performance of LSTM models usually relies heavily on several hyperparameters. Our model consists of two LSTM layers and one fully connected layer. By tuning these hyperparameters listed in Table 2, some experiments were conducted to get a relatively good performance. The epoch number was set as 200, which can achieve relatively low RMSE, and the iteration number was set to be 300. The data set was normalized between 0 and 1 using the MinMax normalization. When applying our LSTM models, four steps were taken: clean data, do the normalization of the data, split data, and construct network structure. The split rate in the simulation was set as 0.8, which means around 80% of the dataset (produced by 8760 or 8764 original records) was processed as training data to build the LSTM model and 20% as testing data. According to different prediction horizons, i.e., one-hour-ahead, several-hour-ahead, or one-day-ahead, all the original data were loaded and retrieved into the training dataset. For example, for one-hour-ahead prediction, each data record was composed of one specific hourly solar radiation data and 24 previous slots data. About 6988 records were set to be training data, and 1747 were set to be testing data. For six-hour-ahead prediction, each data record was composed of certain data and the previous 30th to 7th slot solar radiation data. The optimizer of the neural network training was RMSprop, an implementation of a mini-batch stochastic gradient descent algorithm [33].

Table 2. Hyperparameter setting in long short-term memory (LSTM).

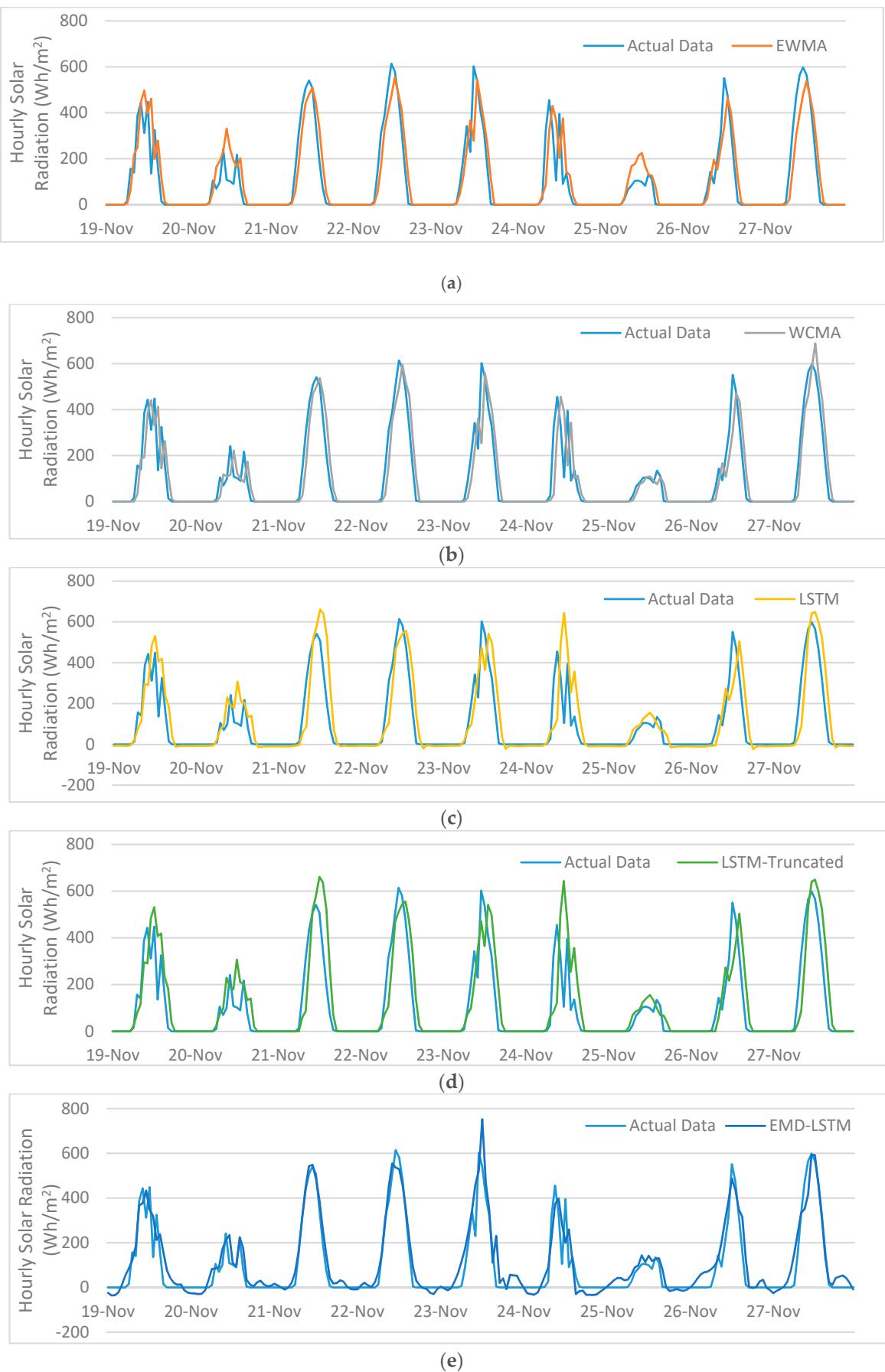
Epoch Number	Iteration	Normalization	Split Rate	Optimizer
200	300	MinMax	0.8	RMSProp

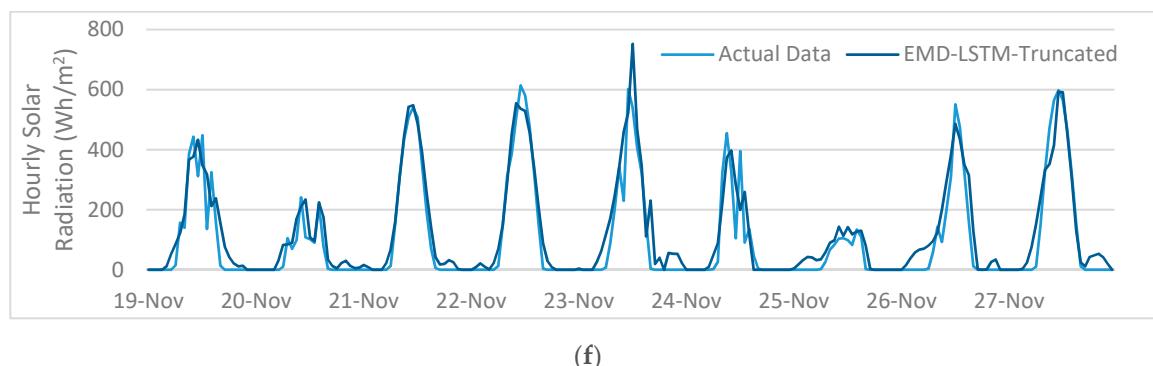
When dealing with the solar radiation data (value = 0) during night hours, it is common to remove these night hours during the data cleaning preprocess. These night hour data were kept for our simulation results for considering more general situations when night hours could be different in some areas. For fitting the general situation, the data preprocessing procedure will be complicated, but the LSTM neural network still produces negative results when the solar radiation values are close to zero. During the design procedure, different activation functions of LSTM layers that limit the negative prediction output were tried, but these options affected the prediction accuracy. Therefore, in our solution, the final data processing step was added to set these negative values to zero, which also led to an increase in prediction accuracy and lowered the prediction error. The details of the performance improvement in the simulation are shown in Section 4.

4.4. Experiment Results

Three different areas' solar radiation data in three selected years, i.e., 1999, 2004, and 2008, were used for comparing the performance of EWMA, WMCA, LSTM, and EMD-LSTM methods. In both the EWMA and WMCA models, the parameters D , K , and α were set as 4, 3, and 0.7. Figure 8 presents the example results achieved by these four methods in 9 days that have different weather conditions. The total daily solar radiation in these 9 days of the year 1999 in Alabama varied from 815 Wh/m^2 to 3642 Wh/m^2 . Table 3 shows the mean varied from 46.83 Wh/m^2 to 151.75 Wh/m^2 , and the standard deviation from 48.04 to 222.82.

Figure 8 shows that all the four methods can follow the radiation trends no matter how dramatically the weather changes. Since EWMA and WCMA are basically weighted average algorithms, they are very accurate under stable weather conditions, as well as during the evening when the amount of solar radiation is continuously 0. The data line of the LSTM algorithm had an obvious depression where night time starts and later stays stably in 0 while the prediction values from EMD-LSTM fluctuated around 0 in a small amount of variation during the night time. Considering the solar radiation is not possible to be below 0, a final processing method was utilized to truncate all the data below zero as 0. Figure 8 also shows the results from the truncated version of the LSTM model and the truncated version of the EMD-LSTM model.

**Figure 8. Cont.**



(f)

Figure 8. Prediction results of the different methods compared to actual data: (a) exponentially-weighted moving average (EWMA); (b) weather conditioned moving average (WCMA); (c) long short-term memory (LSTM); (d) LSTM-Truncated; (e) EMD-LSTM; (f) EMD-LSTM-Truncated.

Table 3. Mean and standard deviation of solar radiation data in 9 days.

	19 Nov	20 Nov	21 Nov	22 Nov	23 Nov	24 Nov	25 Nov	26 Nov	27 Nov
Daily Total (Wh/m ²)	105.58	46.83	128.79	146.58	121.08	79.67	34.00	97.13	151.75
Standard Deviation (Wh/m ²)	159.24	70.59	196.23	216.72	188.28	141.85	48.04	161.84	222.82

To evaluate the accuracy of these prediction methods, which is one of the most important metrics, experiments for different situations were designed separately for comparisons: (1) one-hour-ahead prediction, (2) from two-hour-ahead to one-day-ahead prediction, (3) different time period accuracy, (4) daily solar radiation prediction, and (5) daily profile for prediction.

(1) One-hour-ahead prediction

Tables 4–6 show the RMSE of different methods achieved for one-hour-ahead prediction in Alabama, Michigan, and Nevada, respectively, using previously stated datasets. Yearly results varied according to different solar conditions. The diversity of weather conditions can be observed between these states. The total amount of energy received by the site in Alabama, Michigan, and Nevada varied at 1.40×10^6 Wh/m², 1.68×10^6 Wh/m², and 2.11×10^6 Wh/m². The solar radiation intensity, total sunny and clear days in these areas were significantly different. The EWMA and WCMA models have similar performance, although the WCMA is supposed to be an improved algorithm. The LSTM model alone can achieve good results when the parameters are set appropriately. The performance of truncated versions of LSTM, which set all the negative values as zero, improved by 1.2%–2.2%. The performance of truncated versions of EMD-LSTM improved by 1.75%–2.5%. Compared to the LSTM-Truncated model alone, the prediction accuracy of EMD-LSTM-Truncated improved 5.0%–15.7%. Compared to the EWMA and WCMA models, the truncated version of EMD-LSTM improved 25.0%–44.3% and 29.0%–48.7%, respectively. Compared with other models, our hybrid model has the lowest prediction errors in one-hour-ahead prediction in all the cases, which indicates that our model enhances solar radiation prediction accuracy by retrieving stabilized elements of data through the EMD method.

Table 4. Root mean square error (RMSE) (absolute) of one-hour-ahead prediction results of the exponentially-weighted moving average (EWMA), weather conditioned moving average (WCMA), long short-term memory (LSTM)-Truncated, and empirical mode decomposition (EMD)-LSTM (Alabama).

Year	EWMA	WCMA	LSTM	LSTM-Truncated	EMD-LSTM	EMD-LSTM-Truncated
1999	82.64	80.93	50.34	49.24	46.85	45.89
2004	67.45	55.24	38.15	37.70	35.23	34.61
2008	79.45	78.23	51.34	50.45	44.68	43.78

Table 5. RMSE (absolute) of one-hour-ahead prediction of EWMA, WCMA, LSTM-truncated, and EMD-LSTM-truncated in one-hour-ahead prediction (Michigan).

Year	EWMA	WCMA	LSTM	LSTM-Truncated	EMD-LSTM	EMD-LSTM-Truncated
1999	93.26	92.38	55.02	54.17	52.76	51.44
2004	82.72	85.82	69.25	67.93	58.19	57.07
2008	69.16	76.44	51.94	51.25	47.86	46.82

Table 6. RMSE (absolute) of one-hour-ahead prediction results of EWMA, WCMA, LSTM-Truncated, and EMD-LSTM-Truncated (Nevada).

Year	EWMA	WCMA	LSTM	LSTM-Truncated	EMD-LSTM	EMD-LSTM-Truncated
1999	72.64	70.93	51.34	50.56	48.85	47.83
2004	63.86	66.37	49.14	48.31	46.17	45.30
2008	57.79	55.23	48.72	47.96	42.36	41.45

(2) From two-hour-ahead to one-day-ahead prediction

We compared all four models and the truncated versions in the one-hour-ahead prediction comparison. But since EWMA and WCMA are designed for only one-timeslot-ahead prediction, our model was compared with the single LSTM model for several-hour-ahead and one-day-ahead prediction. Figure 9 shows the RMSE results from LSTM-Truncated and EMD-LSTM-Truncated models in two-hour-ahead, six-hour-ahead, twelve-hour-ahead, and one-day-ahead prediction from the solar radiation data of Alabama, Michigan, and Nevada in 2008. There are some observable tendencies. RMSE increased discernibly when the prediction horizon increases, which is understandable since the longer the horizon, the lower the accuracy. In the two-hour-ahead prediction results of Alabama, predictions of LSTM-Truncated and EMD-LSTM-Truncated models had the absolute value of RMSE as 60.93 and 55.61, respectively, which was 21.0% and 27.0% above one-hour-prediction accuracy. The other two locations shared the same trends. In six-hour-ahead, twelve-hour-ahead, and one-day-ahead predictions, comparably more obvious higher RMSE was observed. Twelve-hour-ahead prediction and one-day-ahead prediction showed similar prediction accuracy in both models. Overall EMD-LSTM-truncated model works better than LSTM alone in all different prediction horizons by 5.8–12.5%.

(3) Different time period prediction

As we all know that the solar radiation around the sunrise time can have a dramatic change and cause high prediction error, we compared the average prediction percentage error according to different time slots of the day, as shown in Table 7. Since the MAPE did not fit the value, which equals 0, the solar radiation of night-time slots was not considered in this particular experiment. Only the time slots from 7 a.m. to 5 p.m. were considered. In the timeslots from 8 a.m. to 5 p.m., the EMD-LSTM model outperformed EWMA, WCMA, and LSTM except that in the timeslot 11 a.m., the prediction accuracy of EMD-LSTM-Truncated (1.80%) was slightly above WCMA (1.78%). The result also verifies the LSTM with the EMD model achieves the lowest average error percentage. Under the MAPE metrics, LSTM-Truncated does not show good performance and has a higher percentage error than other models in this case. Among all the models, we can see the trend that the maximum prediction error occurs during the sunrise and sunset. Our proposed EMD-LSTM model and its truncated version have not solved the problem. Ensemble methods were proposed recently for time-series prediction [34–36], which could be a possible solution by combining different models using adaptive weighting schemes.

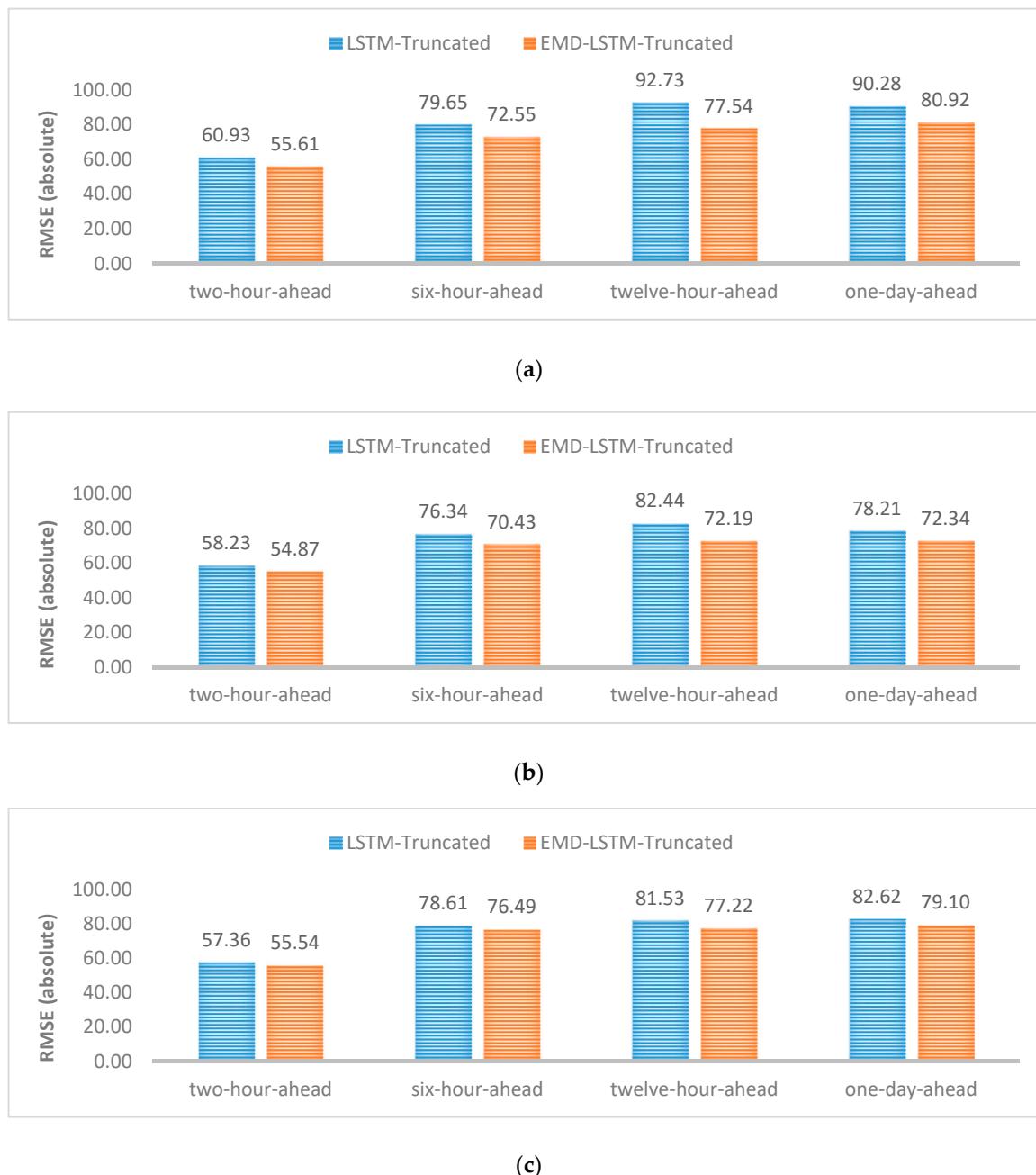


Figure 9. Results of LSTM-Truncated and EMD-LSTM-Truncated in two-hour-ahead, six-hour-ahead, twelve-hour-ahead, one-day-ahead prediction: (a) Alabama; (b) Michigan; (c) Nevada in 2008.

(4) Daily solar radiation prediction

For a longer-term prediction, we also designed the experiments for the accumulated daily solar radiation data for one whole year. The data included the 365 or 366 records of each day summed from 24 h of solar radiation values. The daily solar radiation density and weather conditions in different locations varied dramatically. For example, in Alabama, in the year 2008, the minimum and maximum daily solar radiation were 698 Wh/m^2 and 8099 Wh/m^2 . The minimum and maximum in Michigan were 661 Wh/m^2 and 3279 Wh/m^2 , and those in Nevada were 485 Wh/m^2 and 8133 Wh/m^2 , respectively.

Table 7. Mean absolute percentage error (MAPE) (%) of prediction results of the EWMA, WCMA, LSTM, and EMD-LSTM in different time slots.

Hour.	EWMA	WCMA	LSTM-Truncated	EMD-LSTM-Truncated
7	8.72%	14.29%	36.10%	134.15%
8	7.42%	9.88%	7.25%	4.60%
9	4.96%	4.69%	6.53%	1.45%
10	3.32%	2.51%	3.55%	1.41%
11	2.64%	1.78%	2.43%	1.80%
12	2.98%	2.10%	3.08%	2.03%
13	1.80%	1.23%	2.27%	1.02%
14	2.97%	2.37%	3.90%	1.53%
15	1.92%	2.17%	4.23%	0.79%
16	5.04%	4.72%	8.23%	2.34%
17	88.98%	88.78%	145.29%	71.57%

The hyperparameters of LSTM-Truncated and EMD-LSTM-Truncated were set as the same in Section 4.3. In all the situations, the EMD-LSTM-Truncated model outperformed the LSTM-Truncated model by 30.1% to 40.2%, shown in Table 8, which has more performance improvement compared to 5%–15.7% in one-hour-ahead prediction. The results show that in daily solar prediction, the EMD-LSTM method has more obvious advantages by training LSTM with stabilized subsequences of data when datasets are not large enough for training.

(5) Daily profile for prediction

To test the performance of utilizing the solar radiation profile of the current day, preliminary experiments were also done for three different years in Alabama. Each whole year data set was clustered into N clusters which was set to be 10 during whole experiments. Randomly select 30 days as test data from the dataset and supposed solar radiation of the previous 16 h in the current day was known and profiles in the particular cluster most similar to the current day were chosen to train LSTMs. As a result, the Pro-EMD-LSTM-Truncated model achieved better prediction results than EMD-LSTM-Truncated for one-hour-ahead prediction shown in Table 9. It shows the LSTM model with the radiation profile method had smaller RMSE results compared to the EMD-LSTM-Truncated model in three different datasets by 3.7%–10.4%. The model can enhance solar radiation prediction by adopting more suitable training data to avoid local optima. The main disadvantage of the K-means algorithm is that the initial clustering centroids are randomly selected which could lead to different cluster formations. A poor cluster initialization may cause bad results in clustering [37].

These designed experiments demonstrate the following results:

- (1) Overall the proposed LSTM method based on EMD and solar profiles can improve the accuracy of prediction and achieve better performance than traditional solar prediction methods, such as WCMA, EWMA. The RMSE values indicate the proposed hybrid model has the lowest prediction error among all the models in one-ahead-hour prediction.
- (2) The LSTM model based on the EMD method is slightly better than solely using the LSTM neural networks model in the medium prediction horizon, that is, from several-hour-ahead prediction to one-day-ahead prediction. In the daily solar radiation prediction, using the EMD method has obvious advantages over the LSTMs alone. The data decomposition method divides time series data into more stabilized separated IMFs and makes the LSTM more easily to be trained, which improves the performance of the model.
- (3) Using similar day profiles to train data in LSTM neural networks helps to improve the prediction accuracy noticeably by preventing LSTM from the local optima in one-hour-ahead prediction.
- (4) The MAPE metric also shows the hybrid model achieves the best performance in different time durations of a day among all the models. One thing that needs to be improved is to decrease the error rate of sunrise and sunset periods in LSTM based models in future work.

Table 8. RMSE (absolute) of daily prediction results of LSTM-truncated and EMD-LSTM-truncated.

Year	Alabama		Michigan		Nevada	
	LSTM-Truncated	EMD-LSTM-Truncated	LSTM-Truncated	EMD-LSTM-Truncated	LSTM-Truncated	EMD-LSTM-Truncated
1999	1134.23	786.42	1258.23	875.93	1182.48	793.24
2004	1258.23	843.36	1200.91	812.35	1279.23	765.39
2008	1302.72	793.25	1283.01	849.72	1241.23	836.57

Table 9. RMSE (absolute) of EMD-LSTM-Truncated and Pro-EMD-LSTM-truncated for the years 1999, 2004, and 2008.

Year	EMD-LSTM-Truncated	Pro-EMD-LSTM-Truncated
1999	49.24	47.34
2004	37.70	36.29
2008	50.45	45.18

5. Conclusions and Future Work

In this paper, an LSTM neural network model with EMD methods was proposed. For a one-hour-ahead prediction algorithm, the solar radiation profile of that day was also utilized to select training data from LSTMs. Similar profiles were selected by calculating distances from the typical profile cluster in the database. EMD was employed to decompose the data into stabilized components and make LSTM neural networks predict more accurately in the one-day-ahead horizon. Experiments were done to compare the proposed model with the LSTM, EWMA, WCMA models from the United States national solar radiation dataset for the one-hour-ahead prediction. Comparison results demonstrated that the proposed model can approve prediction accuracy.

In future work, we will improve the model in several ways. First, problems with the high error rate of sunrise and sunset periods in LSTM based models need to be improved. Second, parameters of K-means clustering for time series data are currently optimized by preliminary tests where we will do more extensive experiments to learn in more depth the effects on the prediction accuracy. Third, using the EMD method to decompose data before training as one option, other signal processing methods would also be tested for better performance. In addition, considering the sole model may not achieve the highest performance, a combined model combining the state-of-art algorithms may be considered, for example, proposing a method to weight the combined model in prediction solar radiation.

Author Contributions: Conceptualization, Y.G.; methodology, Y.G., Y.N.; software, Y.G.; resources, L.B.; data curation, Y.G., L.B.; supervision, Y.N.; validation, Y.G., L.B.; writing—original draft, Y.G.; writing—review & editing, Y.G., Y.N. and L.B.

Funding: This work was partly funded by the Scientific Research Fund of Zhejiang Provincial Education Department [No. JG20180128].

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. Sudevalayam, S.; Kulkarni, P. Energy harvesting sensor nodes: Survey and implications. *IEEE Commun. Surv. Tutor.* **2001**, *13*, 443–461. [[CrossRef](#)]
2. Adu-Manu, K.; Adam, N.; Tapparello, C.; Ayatollahi, H.; Heinzelman, W. Energy-harvesting wireless sensor networks (EH-WSNs): A review. *ACM Trans. Sens. Netw.* **2018**, *14*, 1–50. [[CrossRef](#)]
3. Utgikar, V.P.; Scott, J.P. Energy forecasting: Predictions, reality and analysis of causes of error. *Energy Policy* **2006**, *34*, 3087–3092. [[CrossRef](#)]
4. Vigorito, C.M.; Ganesan, D.; Barto, A.G. Adaptive Control of Duty Cycling in Energy-Harvesting Wireless Sensor Networks. In Proceedings of the 2007 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, San Diego, CA, USA, 18–21 June 2007; pp. 21–30.
5. Muhammad, M.; Qureshi, H.K.; Saleem, U.; Saleem, M.; Pitsillides, A.; Lestas, M. Harvested Energy Prediction Schemes for Wireless Sensor Networks: Performance Evaluation and Enhancements. *Wirel. Commun. Mob. Comput.* **2017**, *2017*, 6928325. [[CrossRef](#)]
6. Kansal, A.; Hsu, J.; Srivastava, M.; Raghunathan, V. Harvesting aware power management for sensor networks. In Proceedings of the 43rd ACM Design Automation Conference (DAC), San Francisco, CA, USA, 24–28 July 2006; pp. 651–656.
7. Piorno, J.R.; Bergonzini, C.; Atienza, D.; Rosing, T.S. Prediction and management in energy harvested wireless sensor nodes. In Proceedings of the 2009 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, Aalborg, Denmark, 17–20 May 2009; pp. 6–10.
8. Martín, L.; Zarzalejo, L.F.; Polo, J.; Navarro, A.; Marchante, R.; Cony, M. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Sol. Energy* **2010**, *84*, 1772–1781. [[CrossRef](#)]

9. Suganthi, L.; Iniya, S.; Samuel, A.A. Applications of fuzzy logic in renewable energy systems—A review. *Renew. Sustain. Energy Rev.* **2015**, *48*, 585–607. [[CrossRef](#)]
10. Lee, W.; Kim, K.; Park, J.; Kim, J.; Kim, Y. Forecasting Solar Power Using Long-Short Term Memory and Convolutional Neural Networks. *IEEE Access* **2018**, *6*, 73068–73080. [[CrossRef](#)]
11. Moghtaderi, A.; Flandrin, P.; Borgnat, P. Trend filtering via empirical mode decompositions. *Comput. Stat. Data Anal.* **2013**, *58*, 114–126. [[CrossRef](#)]
12. An, X.; Jiang, D.; Zhao, M.; Liu, C. Short-term prediction of wind power using emd and chaotic theory. *Commun. Nonlinear Sci. Numer. Simul.* **2012**, *17*, 1036–1042. [[CrossRef](#)]
13. Dong, Y.; Ma, X.; Ma, C.; Wang, J. Research and application of a hybrid forecasting model based on data decomposition for electrical load forecasting. *Energies* **2016**, *9*, 1050. [[CrossRef](#)]
14. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [[CrossRef](#)]
15. Cammarano, A.; Petrioli, C.; Spenza, D. Pro-Energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks. In Proceedings of the 2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012), Las Vegas, NV, USA, 8–11 October 2012; pp. 75–83.
16. Ji, W.; Chan, C.K.; Loh, J.W.; Choo, F.H.; Chen, L.H. Solar radiation prediction using statistical approaches. In Proceedings of the 2009 7th International Conference on Information, Communications and Signal (ICICS), Macau, China, 8–10 December 2009; pp. 1–5.
17. Abuella, M.; Chowdhury, B. Solar power probabilistic forecasting by using multiple linear regression analysis. In Proceedings of the 2015 IEEE SoutheastCon, Fort Lauderdale, FL, USA, 9–12 April 2015; pp. 1–5.
18. Muselli, M.; Poggi, P.; Notton, G.; Alain, L. First order Markov chain model for generating synthetic typical days series of global irradiation in order to design PV stand alone systems. *Energy Convers. Manag.* **2001**, *42*, 675–687. [[CrossRef](#)]
19. Ghuman, M.; Iqbal, A.; Qureshi, H.; Lestas, M. ASIM: Solar Energy Availability Model for Wireless Sensor Networks. In Proceedings of the ACM 3rd International Workshop on Energy Harvesting & Energy Neutral Sensing Systems, Seoul, Korea, 1 November 2015; pp. 21–26.
20. Hsu, J.; Zahedi, S.; Kansal, A.; Srivastava, M.; Raghunathan, V. Adaptive duty cycling for energy harvesting systems. In Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED), Tegernsee, Germany, 4–6 October 2006; pp. 180–185.
21. Bergonzini, C.; Brunelli, D.; Benini, L. Comparison of energy intake prediction algorithms for systems powered by photovoltaic harvesters. *Microelectron. J.* **2010**, *41*, 766–777. [[CrossRef](#)]
22. Dehwah, A.; Elmetennani, S.; Claudel, C. UD-WCMA: An energy estimation and forecast scheme for solar powered wireless sensor networks. *J. Netw. Comput. Appl.* **2017**, *90*, 17–25. [[CrossRef](#)]
23. Cammarano, A.; Petrioli, C.; Spenza, D. Improving Energy Predictions in EH-WSNS with Pro-Energy-VLT. In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, Roma, Italy, 11–13 November 2013. Article No. 41.
24. Srivastava, S.; Bhardwaj, S.; Sastri, O.S. A novel hybrid model for solar radiation prediction. In Proceedings of the 2012 International Conference on Emerging Trends in Electrical Engineering and Energy Management (ICETEEEM), Chennai, India, 13–15 December 2012; pp. 243–248.
25. Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting-An approach using AutoEncoder and LSTM Neural Networks. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 002858–002865.
26. Srivastava, S.; Lessmann, S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Sol. Energy* **2018**, *162*, 232–247. [[CrossRef](#)]
27. Kosunalp, S. A new energy prediction algorithm for energy-harvesting wireless sensor networks with Q-learning. *IEEE Access* **2016**, *4*, 5755–5763. [[CrossRef](#)]
28. 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. *R. Soc. Lond. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
29. Huang, B.; Kunoth, A. An optimization based empirical mode decomposition scheme. *J. Comput. Appl. Math.* **2013**, *240*, 174–183. [[CrossRef](#)]
30. US National Solar Data Base. Available online: http://rredc.nrel.gov/solar/old_data/nsrdb/ (accessed on 20 October 2019).
31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]

32. Graves, A. Supervised sequence labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin, Germany, 2012; pp. 5–13.
33. Hinton, G.; Srivastava, N.; Swersky, K. Lecture 6a Overview of Mini-Batch Gradient Descent. Lecture Notes Distributed in CSC321 of University of Toronto. Available online: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf (accessed on 18 November 2019).
34. Adhikari, R. A neural network based linear ensemble framework for time series forecasting. *Neurocomputing* **2015**, *157*, 231–242. [[CrossRef](#)]
35. Kourentzes, N.; Barrow, D.; Crone, S. Neural Network Ensemble Operators for Time Series Forecasting. *Expert Syst. Appl.* **2014**, *41*, 4235–4244. [[CrossRef](#)]
36. Choi, J.Y.; Lee, B. Combining LSTM Network Ensemble via Adaptive Weighting for Improved Time Series Forecasting. *Math. Probl. Eng.* **2018**, *2018*, 1–8.
37. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Philip, S.Y. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]



© 2019 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/>).