Applied Energy 257 (2020) 114001

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Time series prediction for output of multi-region solar power plants

Jianqin Zheng[a], Haoran Zhang[b,*], Yuanhao Dai[a], Bohong Wang[a], Taicheng Zheng[a], Qi Liao[a], Yongtu Liang[a], Fengwei Zhang[c], Xuan Song[b]

[a] National Engineering Laboratory for Pipeline Safety/MOE Key Laboratory of Petroleum Engineering/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Fuxue Road No. 18, Changping District, Beijing 102249, PR China
[b] Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan
[c] Department of Computer Science, Computer and Systems Security Laboratory, Wayne State University, Detroit, MI, USA

HIGHLIGHTS

- Long short-term memory is used to predict the multi-region solar power output.
- Particle swarm optimization is used to optimize the long short-term memory model.
- Sensitivity analysis is done to verify the stability of the method.
- Different long short-term memory structures are compared.
- A real case is studied to illustrate the effectiveness of the proposed method.

ARTICLE INFO

Keywords:
Solar power output prediction
Time series
Multi-region
Long short-term memory
Particle swarm optimization algorithm
Sensitivity analysis

ABSTRACT

Solar energy, as a renewable and clean energy source, has developed rapidly and has attracted considerable attention. The integration of solar energy into a power grid requires precise prediction of the power output of solar plants. Accurate solar power output prediction can promote power dispatch, maintaining the normal operation of power systems. However, research on multi-region solar power is still rare. In this study, long short-term memory and a particle swarm optimization algorithm contribute to solar power prediction considering time series. In order to improve the prediction accuracy, particle swarm optimization is used to optimize the parameters of the long short-term memory model. In addition, different long short-term memory structures are illustrated to determine the final prediction model with sensitivity analysis. Experiments are carried out to verify the effectiveness of the proposed method. The mean absolute error and root mean square error of the proposed method is the smallest among the prediction methods in four cases containing different seasons. In terms of prediction accuracy, results indicate that the proposed prediction model outperforms basic long short-term memory, artificial neural network, and extreme gradient boosting.

1. Introduction

1.1. Background

Rapid socio-economic development has increased energy demands, creating a huge impact on the environment [1]. In response, many countries have introduced policies promoting renewable energy [2], with solar energy being an integral part. In recent years, solar energy has developed rapidly and been recognized as one of the greatest potential energy sources [3]. The number of large-scale solar power plants is also increasing year by year [4]. If only 0.1% of the solar energy on earth is converted to electricity with an efficiency rate of 10%, 3000 GW of power will be generated, which is four times the annual world energy consumption [5]. However, an accurate and stable prediction of solar power output (SPO) is a big challenge for power systems due to the randomness and spacing of SPO generation [6]. Reliable solar power forecasting not only improves the operational coordination of solar and power systems, ensuring energy supply, but it can also promote power dispatch for energy companies, optimizing the power resource scheduling and increasing economic benefits [7]. Incorrect

Abbreviations: ANN, Artificial neural network; ARMA, Auto-regressive moving average; LSTM, Long short-term memory; MC, Markov chain; PSO, Particle swarm optimization; RNN, Recurrent neural network; SPO, Solar power output; SVM, Support vector machine

* Corresponding author.
E-mail address: zhang_ronan@csis.u-tokyo.ac.jp (H. Zhang).

https://doi.org/10.1016/j.apenergy.2019.114001
Received 7 June 2019; Received in revised form 22 September 2019; Accepted 13 October 2019
Available online 20 October 2019
0306-2619/

Nomenclature

Parameters

c_t	The cell state at time t
$\tilde{c_t}$	The new candidate information at time t
f_t	The forget gate at time t
h_t	The hidden state at time t
i_t	The input gate at time t
o_t	The output gate at time t
x_t	The input information at time t

predictions may result in insufficient energy supply, thus affecting society's normal operations [8]. Therefore, in order to prevent negative impacts, accurate SPO prediction is always required for power systems.

However, most previous work in this area is concerned with SPO in a certain region, and few studies take into account the power generation of multiple regions. From the perspective of a power system, the sum of the solar power generation of its plants is the key to balancing energy demand. Additionally, due to the high cost of solar irradiance meters [9], it is unlikely that a meter can be located in each region. Therefore, having solar power plants in different regions, and a lack of solar irradiance meters, poses a challenge for predicting power output accurately. This study aims to predict the sum of SPO for a power system with solar power plants in multiple regions. The complexity of SPO prediction is not simply related to the intensity of sunshine, as it also depends on weather conditions, which have a strong time-series relationship [10]. At the same time, the characteristics of the SPO time series are non-stationary in nature. The significance of SPO prediction is that the short stochastic characteristics and the dependence of the observation time series of solar radiation must be considered. The difficulty of SPO prediction in this work can be concluded as follows:

(1) Solar power plants are located in multiple regions with limited solar irradiance meters.
(2) The dependence of the time series of solar radiation needs to be taken into account.
(3) The material of the time series for the SPO is unstable.

To address the first difficulty, considering spatial and temporal distribution, the data of limited solar irradiance meters and weather conditions of all regions are collected for the prediction model based on a data-driven perspective. Considering the second difficulty, the LSTM model is used in this paper to capture the dependence of time series predicting for SPO. For the last difficulty, different LSTM structures are compared, and sensitivity analysis is conducted for the divisions of datasets, the parameters of the PSO algorithm, and the input weather variables to ensure the stability of the proposed method.

The study is stated as follows. By using photovoltaic power generation systems, solar energy is converted into electricity, and it is connected to an electrical grid for utilization. As a renewable energy, solar power is critical as a power supply source. As for power systems, knowing the SPO in advance is beneficial for promoting power dispatch, improving the stability of power supply and optimizing the power resource scheduling, increasing economic benefits. For this reason, precise prediction of solar power generation is urgent for dealing with dispatchers. However, solar power generation is a fluctuating power source that is heavily reliant on weather conditions, resulting in uncertainty and intermittency of solar energy. Put simply: when the weather is fine and there are lengthy periods of sunshine, more solar power is generated. These conditions create more uncertainties in predictions. Therefore, this work aims to predict the power output by plants rather than from solar irradiation. In addition, the capacity, azimuth angle and inclination angle also contribute to the output of each power plant, but in general, these three factors will not change in the short term.

For one power system, there are many plants in multiple regions, and the power generation from all plants is collected into the same system. As shown in Fig. 1, the orange dots represent the locations of the solar power plants, and the properties of these power plants are known and unchangeable. According to the location, the weather condition of each solar power plant can be determined. The red squares in Fig. 1 represent the measured locations. The temperature and global solar radiation values are measured precisely by instruments in these locations. If the solar radiation values of each orange dot are also known, then the amount of generated power can be predicted accurately. But how does the information from the red squares help to predict the power output of the orange dots? It is difficult to establish the relationship between the two from the physical mechanism, but using a data-driven perspective, with the weather condition and the measured data, finding a suitable prediction model will provide us with the sum of SPO, as shown in Fig. 2. But how can we find a suitable prediction model to ensure that the predicted result is precise? By analyzing these basic data, we see that the data is a time series. Considering the time series, the past data is used to train the prediction model, and with the numerical weather prediction, the prediction of the sum of SPO can be obtained for the power system.

The prediction of the SPO for power systems is described in this paper as follows.

Given:

(1) The number of solar power plants and location for each one.
(2) For each solar power plant, the weather condition, capacity, azimuth angle, inclination angle and the actual power output in the past.
(3) For measurement locations, the measured temperature and the global solar radiation values.

Determine:

(1) The sum of the power output from the set of specified solar power plants of a power system.

In order to predict SPO effectively, assumptions are as follows:

(1) The capacity, azimuth angle and inclination angle are not changed in the research.

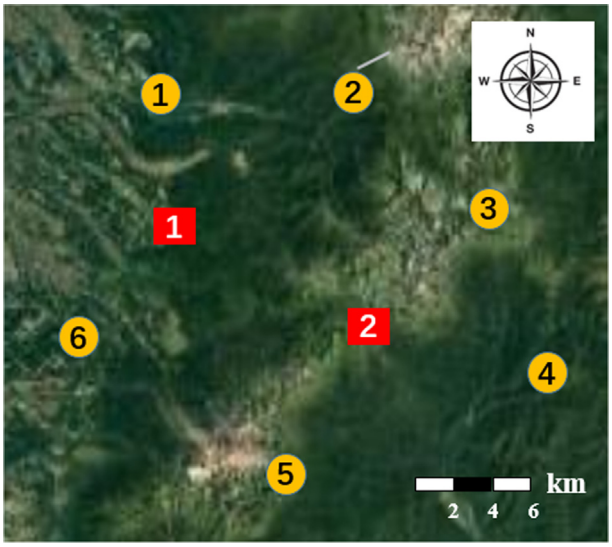

Fig. 1. The locations of solar power plants and measured points – orange dots represent solar power plants and red squares represent measured locations.

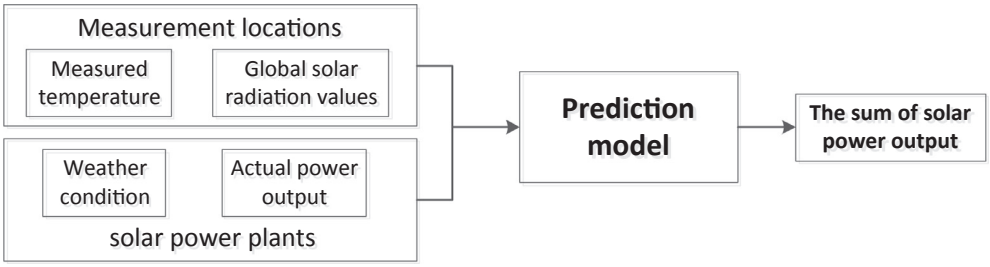

Fig. 2. The basic data for predicting the sum of SPO.

(2) The instrument measurement error is within the acceptable range and the basic data does not affect the prediction model.

1.2. Literature review

Many researchers have worked to develop effective prediction methods for SPO forecasting [11], especially through data-driven machine learning [12]. Yang et al. [13] used only off-the-shelf implementations of 68 machine learning models for hourly solar forecasting for two years, and results showed that despite no universal model being found, tree-based methods performed well in terms of two-year overall results not standing out during daily evaluation. Long et al. [14] used four well-known data-driven approaches to forecast daily SPO after a procedure of selecting the input parameters, and they inferred that some algorithms can outperform others in different considered scenarios. Different prediction models may vary widely in different situations, and since there are very few studies in time series predicting for the output of multi-region solar power plants, there is no reference, and it is more difficult to propose an effective prediction method if not all regions have solar irradiance meters.

Because solar irradiance meters are expensive, some scholars indirectly predict SPO by predicting the solar irradiance [15]. Kamadinata et al. [16] predicted minute-level solar irradiance by utilizing artificial neural networks based on sky photo images which capture the trends of fluctuating solar irradiance with only minor discrepancies. Li et al. [17] proposed an improved bootstrap method to construct the prediction interval for solar power, due to its nature of variability, in order to tackle the problem of invalid assumption about forecast errors. Izgi et al. [18] used an artificial neural network (ANN) to predict the electricity generated by photovoltaic energy, and they determined a representative time horizon according to the months for small-scale solar power systems. However, the prediction of solar irradiance does not directly reflect the generation of solar energy, because the SPO is not only related to the solar irradiance, but also with weather conditions such as temperature and wind speed. Therefore, these prediction models are not sufficiently accurate for SPO. The introduction of weather conditions brings more instability to the prediction model due to their variability. More emphasis should be given to feature engineering and determining the input weather variable for the prediction model.

Some scholars have made improvements to the traditional machine learning algorithm [19], represented by a support vector machine (SVM). Zeng and Qiao [20] forecasted short-term SPO based on a least-square SVM model, of which the output is the atmospheric transmissivity converted to solar energy, relying on the latitude of the site and the time of day. Li et al. [21] developed a hybrid improved multi-verse optimizer algorithm to optimize the SVM for predicting photovoltaic output due to its performance and stability. A common method for improving the accuracy of SPO prediction is by transforming traditional methods or adding combined models. Combined models play an increasingly important role in this field, and they will be used in this work.

In addition, the generation of solar power is related to the time series, and spatiotemporal data may improve the accuracy of prediction [22]. A time series is a set of statistical data, usually collected at regular intervals. The aims of time series analysis are to explain and conclude time series data and fit models, as well as making forecasts [23]. SPO is heavily dependent on the weather and the time of day, which have a time series relationship, and many time series prediction methods have emerged [24]. As statistical methods, algorithms based on an auto-regressive moving average (ARMA) [25] and a Markov chain (MC) [26] were employed to build the time series prediction model for SPO. Prema and Uma Rao [24] proposed statistical time series models for daily solar irradiance prediction to predict SPO, and the prediction accuracy can be improved when decomposing into seasonal and trend patterns. Yang and Dong [27] considered a total of 142 models from six families to applicate in day-head operational photovoltaic power output forecasting using time series ensembles, and experiments verified that the expanded information set provided by the numerical weather prediction model can further improve the ensemble-forecast accuracy. These studies have shown that considering the time series of solar irradiance and the weather can help improve the accuracy of the SPO prediction. However, these traditional time series prediction methods can only be applied to situations where the amount of data is relatively small.

With the rapid development of artificial intelligence, deep learning has attracted increasing attention for its excellent performance in machine translation [28], speech emotion recognition [29], and image recognition [30]. As deep learning algorithms, recurrent neural networks (RNNs) have also been used to forecast time series problems [31]. Proposed by Hochreiter and Schmidhuber [32], long short-term memory (LSTM) can analyze a time series with long spans for prediction and solve the problem of vanishing gradient encountered with RNN [33]. Taking the dependence between the consecutive hour of the same day, Qing and Niu [9] developed LSTM networks for hourly day-ahead solar irradiance by using the weather forecasting data provided by local meteorological organizations. Srivastava and Lessmann [34] concentrated on LSTM for forecasting day-ahead global horizontal irradiance and utilized remote-sensing data for testing predicting accuracy at 21 locations to assess the efficiency of the method comprehensively and reliably. Based on data-driven approaches, the LSTM model stands out in time series predicting, with better prediction results. After the aforementioned analysis, there are three main reasons why LSTM is novel in this work: (1) LSTM can handle the correlation of time series; (2) compared to other time series algorithms, it can train a lot of data caused by short intervals; (3) there is no gradient vanishing, which is found with traditional recurrent neural networks.

However, there is still room to improve the prediction accuracy by optimizing the LSTM model. Unfortunately, there is currently no such research on SPO prediction. In similar fields, some researchers have developed the LSTM model to improve its prediction effect. For example, Yu et al. [35] enhanced the forget-gate network of the LSTM model, filtered the data of the turbine groups and optimized the forecasting effect by clustering based on sequential correlation features for improving the accuracy of predicting wind power. Peng et al. [36] aimed to predict the electricity for balancing electricity generation and

consumption, so a novel evolutionary algorithm, differential evolution, has been designed to obtain suitable hyper-parameters for LSTM. Furthermore, experiments were conducted to verify that the proposed method is better in terms of prediction accuracy. The combined model reflects its powerful and accurate predictive ability.

LSTM, as a neural network, is sensitive to its structure network, which contributes to the prediction effectiveness of the LSTM model [37]. However, few studies have examined the LSTM prediction model based on an intelligent algorithm employed to optimize its structure for SPO forecasting, and any discussion of the intelligent algorithm must include particle swarm optimization (PSO), the genetic algorithm and the ant colony algorithm. The genetic algorithm is suitable for solving the discrete problem, while PSO is suitable for solving the continuous problem, to which the problem of optimizing the LSTM parameters belongs. The ant colony algorithm generally takes a long time to search and is prone to stagnation, while PSO has the advantages of fast search speed, fewer parameters to be adjusted, a simple structure and easy engineering implementation [38]. Also, its combination with the neural network has made significant progress in optimization [39]and decision systems [40]. It plays an important role in engineering optimization, such as optimizing an ANN structure to minimize the fuel consumption for cruise ships [41] and coupling with other intelligent algorithms to select a route for hazardous liquid railway networks [42]. Zhang et al. [43] used a back-propagation neural network optimized by PSO to predict daily global solar radiation in arid northwest China. For one-day-ahead forecasting of solar power generation, the PSO is used to optimize the parameters of the SVM to reach a higher accuracy in Ref. [44]. The idea of using the PSO algorithm to optimize the parameters of the LSTM model has not been found in the literature. The LSTM model with optimized parameters contributes to higher prediction accuracy for SPO.

The research gap of SPO prediction in this work can be concluded as follows:

(1) There are few studies in time series predicting for the output of multi-region solar power plants.
(2) Some studies indirectly predict the SPO by predicting solar irradiance, but they ignore the effects of weather conditions.
(3) Although some studies take into account time series, they only use traditional time series prediction methods, which can only be applied to small-scale data.
(4) The idea of using the intelligent algorithm to optimize the parameters of the LSTM model for SPO prediction has not been found in the literature.

The proposed method in this work is shown in Fig. 3. Firstly, the time-series data of multi-region solar power plants and measurement locations are collected into a database. Then, combining the advantages of the PSO algorithm and the LSTM network, the PSO algorithm is used to optimize the parameters of the LSTM model (abbreviated as PSO-LSTM), including the learning rate, time step, batch size and number of neural units. After obtaining the optimal parameters, the prediction effects of different LSTM structures are compared, and the suitable LSTM network is selected according to the training loss. Finally, the LSTM network needs to be trained to improve its accuracy and generalization ability. Thus, a model appropriate for multi-region SPO is established.

1.3. Contributions of this work

(1) Taking into account the correlation of time series, the LSTM network is applied to predict the multi-region SPO.
(2) This study is the first work to use the PSO algorithm to optimize the parameters of LSTM in SPO prediction.
(3) To illustrate the stability of the method, sensitivity analysis is conducted for the divisions of datasets, the parameters of the PSO algorithm, and the input weather variables.
(4) Different LSTM structures are compared to determine the final prediction model for SPO according to training loss.
(5) Real multi-region solar power plants of a power system are used as an example to verify the effectiveness and practicality of the proposed method.

The remainder of this paper is organized as follows. Data preparation and basic knowledge of the LSTM model and PSO algorithm are explained in Section 2. In Section 3, taking a power system with solar power plants located in multiple regions as a case study, the results are demonstrated and discussed in detail. Section 4 gives the conclusion and suggests future work.

2. Methodology

2.1. Data preparation

Data preparation includes a series of processes, such as data collection, data division, and data normalization. In this study, the basic data can be obtained from the power system, which contains the global solar radiation values and SPO, and the weather bureau, which contains the weather conditions. All of the collected data need to be divided into two parts. One is a small proportion of data for training the parameters for the PSO-LSTM model, and the other is used for training the LSTM prediction model for predicting the SPO. Additionally, the latter is also divided into three subsets: training set, validation set, and testing set. The training set is used to train the weight and bias of the LSTM model. The validation set is used to test the accuracy of the current model after each epoch is completed. The testing set tests the final accuracy of the trained prediction model, which is a measure of its performance.

Data normalization can accelerate the convergence rate of gradient

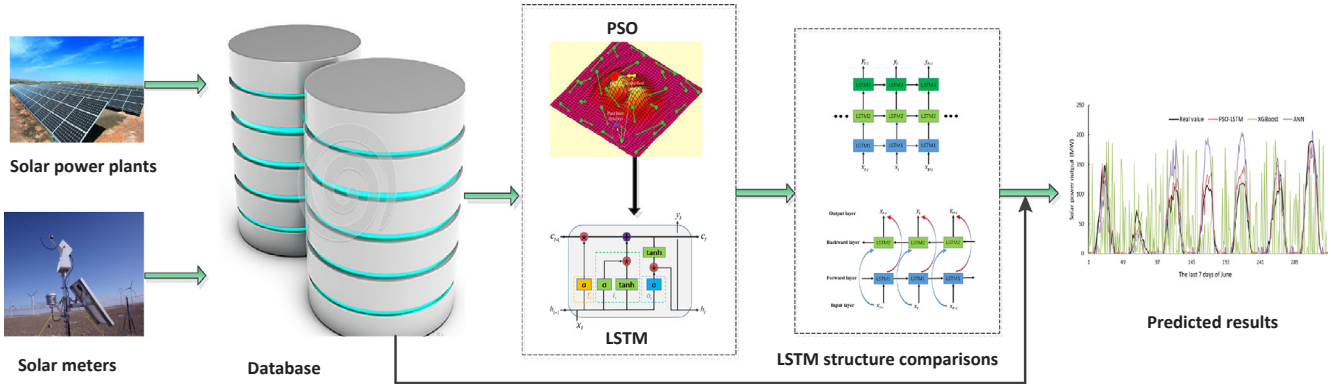

Fig. 3. The framework of SPO prediction.

descent and improve accuracy. In many cases, if the data is not standardized, the gradient descent becomes complex and the model does not work well. Empirically, normalization makes the features of different dimensions comparable with each other as numerical values, which can greatly improve the accuracy of prediction.

Each dimension of the data sets is normalized to values in the range 0–1 as follows:

$$\bar{X}_{\mathrm{i}} = \frac{X_{\mathrm{i}} - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where $\bar{X}_{\mathrm{i}}$ is the normalized value, X_{i} is the original value, and $X_{\max}$ and $X_{\min}$ are the maximum and minimum original values, respectively, of the data sets. The solar power-dependent data and numerical weather data in each sample are normalized.

2.2. LSTM model

A recurrent neural network (RNN) is a type of neural network for processing sequence data. The word "recurrent" highlights the core feature of the RNN – that is, the output will remain in the network, together with the input of the next moment to determine the output of the next moment. This is why it stands out from processing sequence data. Currently, the RNN has great applications in speech recognition, machine translation, and text generation, but one of its major drawbacks is the problem of gradient vanishing. Similar to the neural network, when parameters are updated by back-propagation, they are also optimized along the negative gradient direction. As the sequence accumulates, the gradient will gradually become smaller and approach zero. This will cause the gradient to disappear, and the network will not be updated. The LSTM network combines short-term with long-term memory through subtle gate control and solves the problem of gradient vanishing to some extent.

In this section, the basic LSTM cell is introduced, then the single-layer and multi-layer LSTM structures are described, and the bi-directional LSTM RNN is explained.

(1) Basic LSTM cell

The basic LSTM cell is shown in Fig. 4. At time t, the input of the unit network is the cell state c_{t-1} at t − 1, the hidden state h_{t-1} at t − 1, and the new information x_t at t. When the unit is active, gates allow information through optionally. The forget gate, input gate and output gate are influenced by x_t and h_{t-1}. These three gates all have an output coefficient value between 0 and 1 by the activation function, which is the sigmoid function defined in Eq. (2), determining the propagation of the information.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

The LSTM unit is composed of the following steps:

Step 1: The LSTM unit decides what stored information to throw away from the last cell state c_{t-1}. The forget gate f_t looks at x_t and h_{t-1}, and outputs a number between 0 and 1 for the last cell state c_{t-1}, with 1 representing keeping the stored information totally, and 0 representing getting rid of it completely. The f_t is calculated by Eq. (3).

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f) \tag{3}$$

Step 2: The unit decides what new information to the store in the cell. It consists of two parts. First, in order to ensure the normalization of the new information, the activation function – the tanh function defined in Eq. (4) – converts x_t and h_{t-1} into the new candidate information $\tilde{c}_t$ defined in Eq. (5), which is between −1 and 1. Next, similar to the forget gate, input gate i_t outputs a number between 0 and 1 to determine the proportion of the new information. The i_t is defined in Eq. (6). In the next step, these two parts are combined to update the cell state.

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{4}$$

$$\tilde{c}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c) \tag{5}$$

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i) \tag{6}$$

Step 3: It is time to update the cell state. The previous steps are prepared for the new cell state c_t. The old cell state is multiplied by f_t, forgetting the information decided upon in step 1. The processed information $\tilde{c}_t$ is then multiplied by i_t. Finally, these two components are added to determine c_t, defined in Eq. (7).

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \tag{7}$$

Step 4: This part determines what is going to be output. Similarly, output gate o_t outputs a number to decide what parts of the cell state are going to be output, which is defined in Eq. (8). The hidden state h_t is based on c_t converted by the tanh function and then multiplied by o_t to ensure that only valuable information is stored. The h_t is defined in Eq. (9). Ultimately, the output of LSTM is calculated in Eq. (10).

$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o) \tag{8}$$

$$h_t = o_t * \tanh(c_t) \tag{9}$$

$$y_t = \sigma(W_{hy} \cdot h_t + b_y) \tag{10}$$

In Eqs. (3)–(10), W_{xf}, W_{xi}, W_{xo} and W_{xc} are the input weight matrices; W_{hf}, W_{hi}, W_{ho} and W_{hc} are the recurrent weight matrices; and W_{hy} is the hidden output weight matrix. Vectors b_f, b_i, b_o, b_c and b_y are the corresponding bias vectors. For the LSTM network, the training aims to continuously optimize these weights and vectors until the network converges. In this work, X is the given data mentioned in Section 1.2, and y is SPO.

(2) Multi-layer LSTM

Like an artificial neural network, a single-layer LSTM represents only one hidden layer, and a multi-layer LSTM has multiple hidden layers. Adhering to the concept of deep learning, by increasing the number of network layers, the efficiency of training is improved and higher accuracy is obtained. As shown in Fig. 5, there are three hidden layers in the LSTM model. For the same moment, the output of the previous layer is the input of the next layer, and for the same hidden layer, the output from the last moment is the input at the next moment.

(3) Bi-directional LSTM

The aforementioned LSTM model has only one direction of transmission for the sequence. As shown in Fig. 6, the basic idea of the bi-directional LSTM is that the forward and backward training are both LSTM models for the sequence, and both are connected to an output layer. With this

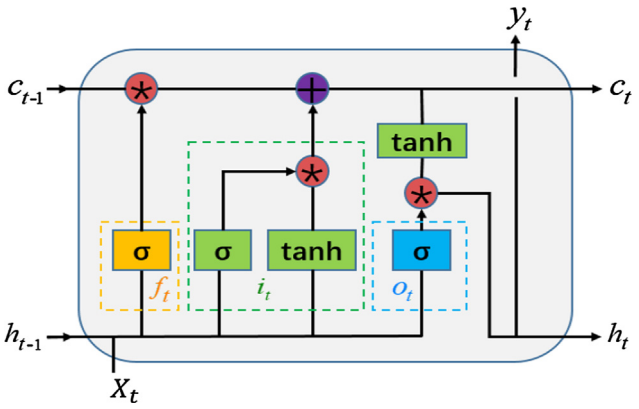

Fig. 4. The basic LSTM cell.

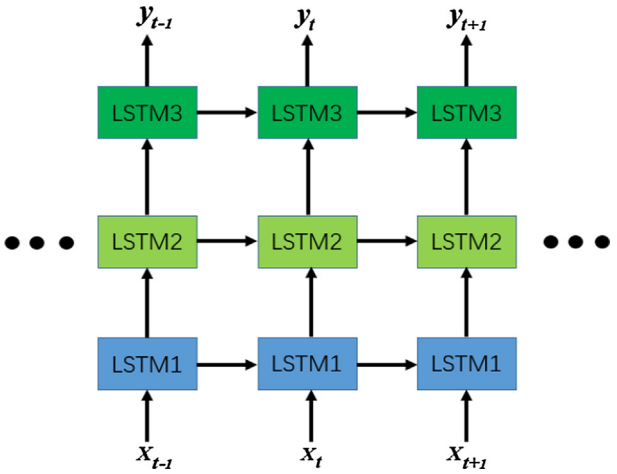

Fig. 5. The multi-layer LSTM model.

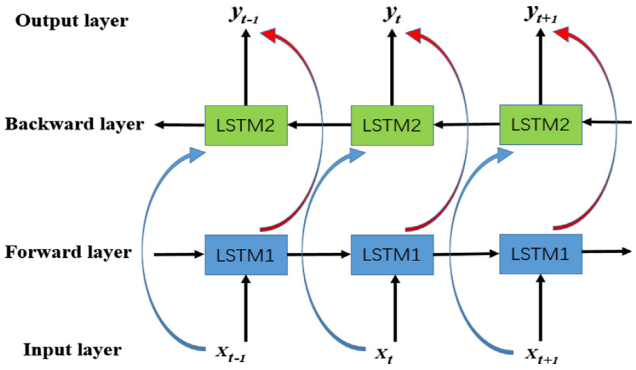

Fig. 6. The bi-directional LSTM structure.

structure, complete past and future information for each point in the input sequence can be provided to the output layer. It works well on text recognition and improves accuracy by inputting the context information [45].

2.3. PSO algorithm

Proposed by Eberhart and Kennedy [46] in 1995, PSO has great application in the field of optimization. In this study, PSO is used to select the optimal parameters for the LSTM structure. The process of the algorithm is detailed as follows.

(1) Initialization of particle swarm

The size of the population is determined, and then several particles are generated to form the first generation. In this study, each individual has four-dimensional information: learning rate, time step, batch size and number of units. At the same time, these particles contain corresponding velocity and position to update.

(2) Fitness evaluation

In the process of PSO-LSTM, the training loss is the fitness of each particle. Based on the principle that the fitness function is the criteria of the continuous evolutionary search of the algorithm, the training loss of the particle-based LSTM model is calculated.

(3) Two "extreme values"

In each iteration, by finding the two "extreme values" – namely, the optimal individual value and the optimal population value – the particles are updated. The less the training loss, the better the particle. So, the least training loss corresponds to the optimal value of the group.

(4) Particles update

By tracking the two "extreme values," the velocity and position of each particle are updated to form the next generation, and the principles during the iterations are according to (11) and (12):

$$v_i(t+1) = w \times v_i + c_1 \times r_1 \times (p_i(t) - x_i(t)) + c_2 \times r_2 \times (p_g(t) - x_i(t)) \quad (11)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (12)$$

where v_i and x_i are the velocity and position of particle i, respectively, w is the inertia weight, c_1 and c_2 are the respective acceleration parameters, representing two random numbers in the range [0, 1], p_i is the best position for particle i, and p_g denotes the best position in the group at the t^{th} iteration.

(5) Convergence

With the increasing number of iterations, the best particle of each iteration is better, and the best position will converge within the maximum iteration. Finally, the best individual of the group is found, and the parameters of the LSTM model are determined by it.

A detailed flow chart of the methodology is shown in Fig. 7. The method consists of four parts: data acquisition and organization, data preprocessing and division, training the PSO-LSTM to optimize the optimal parameters, and training the prediction model for solar power generation. Firstly, the data – namely, the measured temperature and global solar radiation values from measurement locations, as well as the weather conditions, capacity and actual power output from the solar power plants – are collected into the database. Secondly, by the processing mentioned in Section 2.1, the data is divided into two parts. A small part of the data is for the PSO algorithm to train the LSTM parameters, and the rest of the data is for the optimal LSTM model to predict the SPO. Thirdly, the PSO algorithm is executed with a small part of the data. It includes initializing the particle swarm, evaluating training loss to find the "extreme values", and ascertaining whether the convergence condition is met – if yes, the suitable parameters of the LSTM model are found; otherwise, it returns to evaluating the training loss and follows again. Finally, after finding the optimal LSTM structure, it uses the other data to train the prediction model. By training, validating and testing, the prediction model has enough accuracy to predict the SPO. Finally, with the help of the real-time numerical weather prediction, the sum of the output of the solar plants is obtained. This is the entire flow of the methodology.

3. Experiments and analysis

3.1. Input data selection and evaluation indexes

This study aims to predict the total power generation of solar power plants in different regions for a power system. Taking area A of Asia as an example, the solar data adopted in the experiment is from its power system, and the weather condition is from the corresponding weather bureau. As shown in Fig. 8, there are eight solar plants in area A, located in five regions and with two measurement locations. It is important to note that there are three solar plants in region 5 and two solar plants in region 4. Historical data of 2017 is selected for the experiment, and the time interval is 30 min. So, how do we choose a suitable input based on the data? The input of the prediction model plays an important role in improving performance and reducing the computational cost. Too many features might overfit the model, while too few might not be enough, resulting in poor effectiveness of the

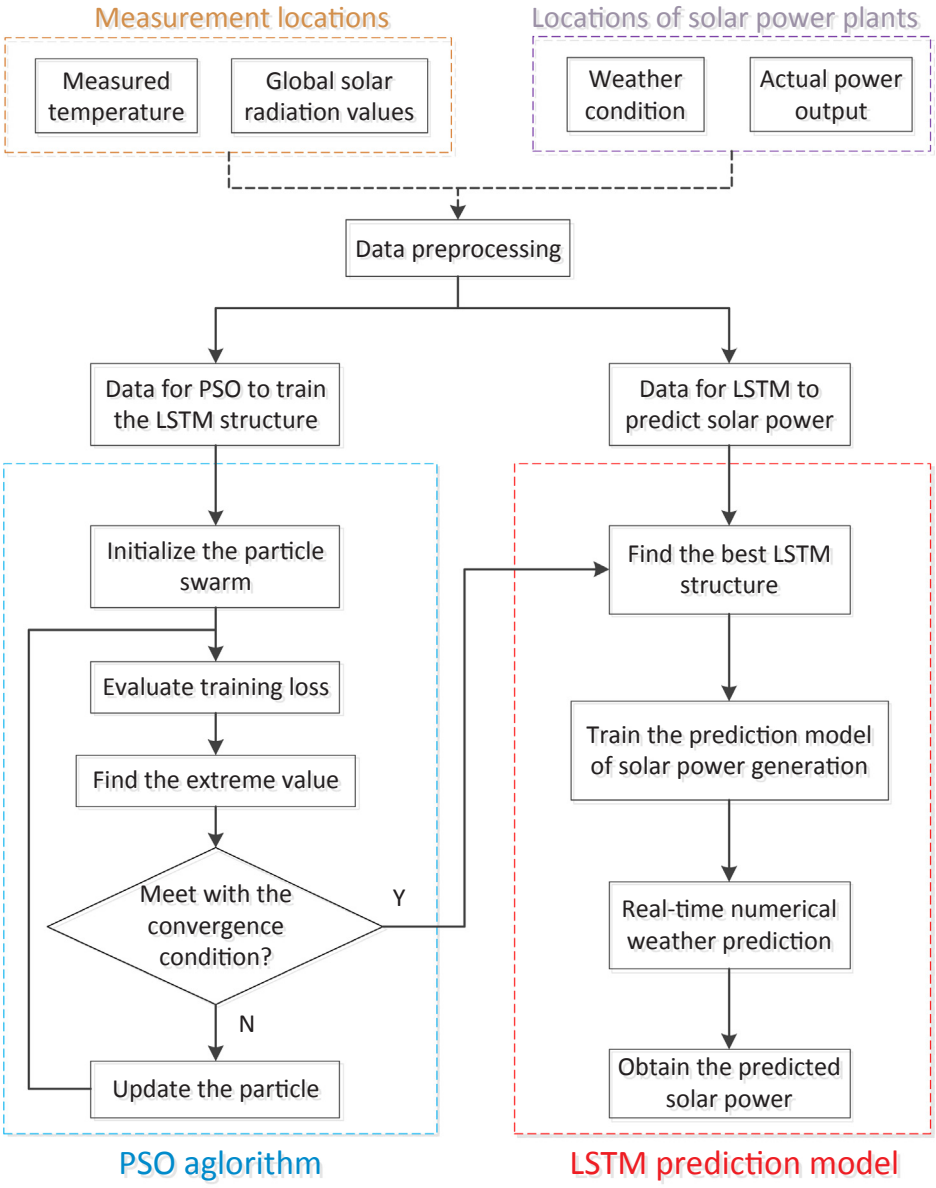

Fig. 7. The flowchart of the methodology.

prediction model.

The selection of the input variables for a prediction model is particularly critical. The input variables depend on data availability and their correlation [47]. For the two measurement locations, global solar radiation values and temperature are input features. For solar plant locations, statistical analysis is carried out to check the correlation of each available weather variable with SPO. Fig. 9 shows the correlation coefficients of all the six weather variables with the SPO by using the historical data. As shown in Fig. 9, the sunshine variable has the greatest positive correlation with solar power, followed by temperature, wind speed, and wind direction, while the humidity variable and dewpoint are negatively correlated with the power output. Considering the limitation of the number of input variables and the number of regions of this power system, the input variables of the weather data will be determined in the sensitivity analysis in Section 3.2.

In this work, mean absolute error (MAE) and root mean squared error (RMSE) are selected as evaluation indexes of the LSTM model. The formulae of these two evaluation indexes are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | p_i - r_i | \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (p_i - r_i)^2} \tag{14}$$

In Eqs. (13) and (14), n represents the total number of the data set, p_i is the predicted value, and r_i is the real value.

3.2. Sensitivity analysis

In this study, the proposed method is implemented on Tensorflow in Python, which is open-source software for deep learning proposed by Google. January's data is used for training PSO-LSTM, while the other data is used for training the prediction model of solar power. This section discusses the sensitivity of the different divisions of datasets, the setting of parameters of the PSO algorithm, and the input weather variables. The different divisions of datasets will affect the training of the model, which may influence the training effect of the model. In addition, different parameters of PSO will affect the optimization of the

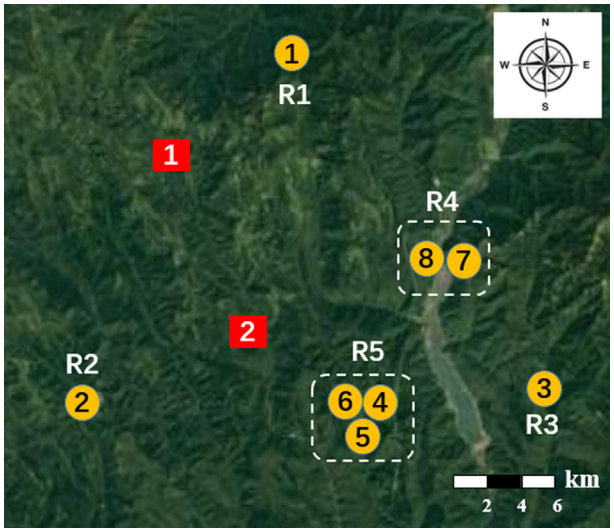

Fig. 8. Area A of Asia, with the dashed boxes R4 and R5 each representing a region.

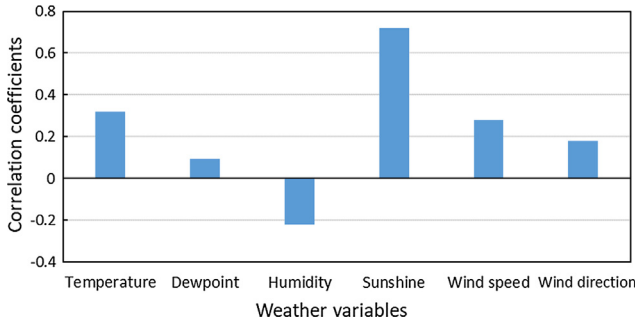

Fig. 9. The correlation coefficients of weather variables and SPO.

Table 1
Predicted results with different divisions of datasets.

	The training sets (%)	The validation sets (%)	The testing sets (%)	MAE	RMSE
Case 1	70	10	20	8.14	19.41
Case 2	65	15	20	8.32	19.78
Case 3	60	20	20	8.25	20.03

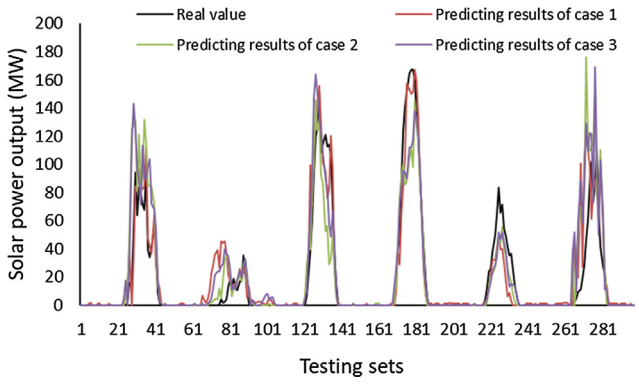

Fig. 10. The predicted results with different divisions of datasets.

population during the iterative process, which may result in finding different suitable parameters for the LSTM model during the process of PSO-LSTM. Different input variables will be discussed in this part to determine the input variables of weather data.

Table 2
Predicted results with different sizes of the group and different inertia weights.

	Size of group	Inertia weight	MAE	RMSE
Case 1	10	0.3	9.35	21.34
Case 2	10	0.5	9.69	21.42
Case 3	10	0.7	9.55	21.97
Case 4	20	0.3	8.37	19.56
Case 5	20	0.5	8.14	19.41
Case 6	20	0.7	8.25	19.89
Case 7	30	0.3	8.28	20.23
Case 8	30	0.5	8.65	19.68
Case 9	30	0.7	8.24	19.45

Table 3
Predicted results with different input weather variables.

	The input variables	MAE	RMSE
Case 1	Sunshine and temperature	12.66	28.41
Case 2	Sunshine, temperature, and wind speed	8.14	19.41
Case 3	Sunshine, temperature, wind speed, and humidity	8.17	19.67

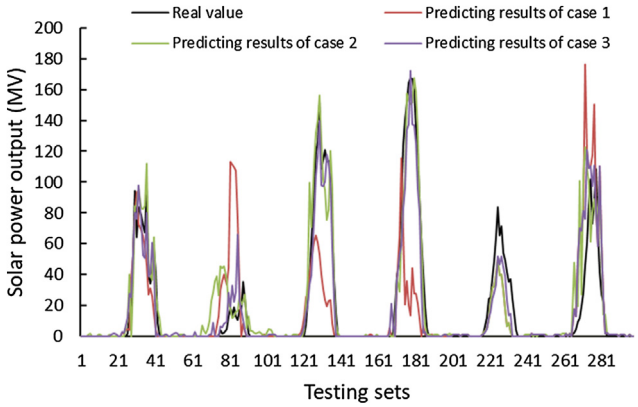

Fig. 11. The predicted results with different input weather variables.

The dataset is divided into three different cases, and the corresponding predicted results are presented in Table 1. The differences in MAE and RMSE values in the three cases are small. The actual SPO and predicting values generated in the three different cases are shown in Fig. 10. The results show that the different divisions of datasets have a negligible influence on the prediction performance.

The parameters, including the size of the group and the inertia weight mentioned in Section 2.3, of the PSO algorithm are used to test predicting accuracy. The errors are presented in Table 2. Comparing case 1 and case 4, or case 2 and case 5, or case 3 and case 6, shows that the values of MAE and RMSE are largest when the size of the group is 10. As the size of the group is increased, and the search space is also larger, finding better fitness is possible. However, the error values of case 4 and case 7, or case 5 and case 8, or case 6 and case 9, are not much different. This is because when the group size is larger than 20, increasing it further does not continue leading to better solutions. When the size of the group is the same, comparing cases 1, 2 and 3 or cases 4, 5 and 6, or cases 7, 8 and 9 show that there is almost no difference in the error values. This demonstrates that the inertia weight does not affect the training process of PSO-LSTM.

The input weather variables are composed of three different cases, and the corresponding predicted results are presented in Table 3. MAE and RMSE values in case 1 are the largest among the cases, while those for case 2 and case 3 are almost identical. This indicates that when sunshine and temperature are the only input weather variables, the prediction model cannot make full use of the weather conditions, but when the input data are sunshine, temperature and wind speed,

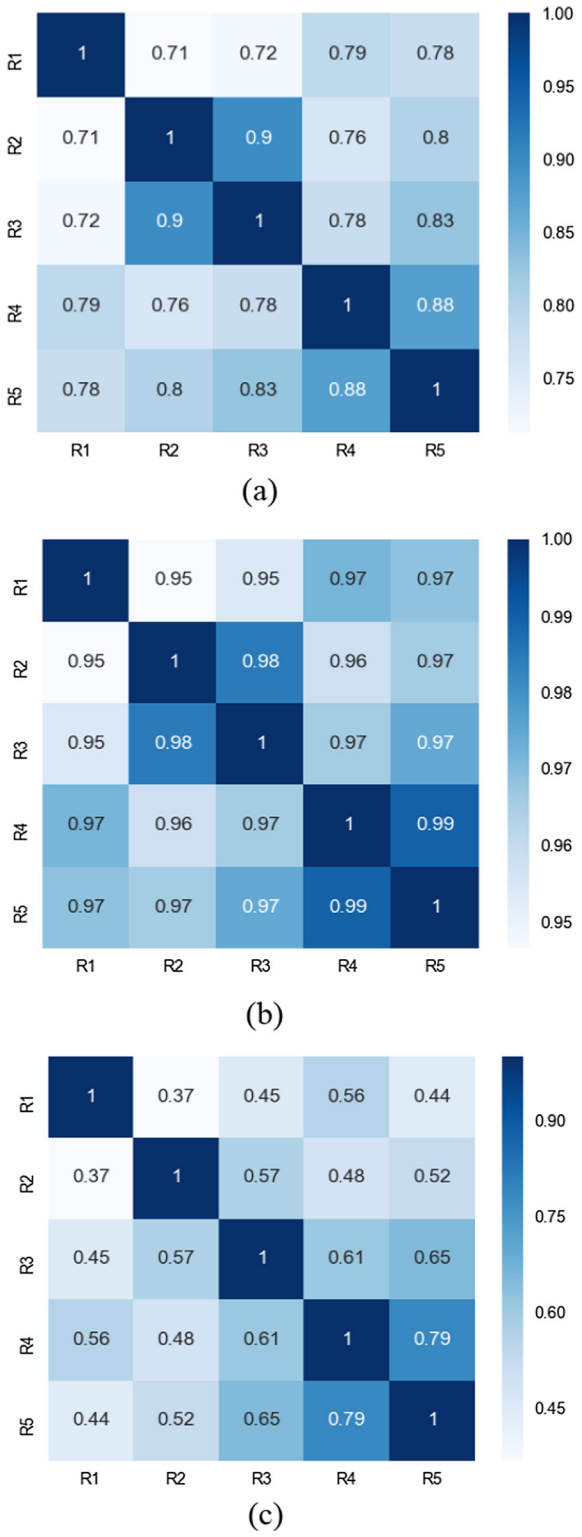

Fig. 12. The correlation coefficient of (a) sunshine, (b) temperature and (c) wind speed in these five regions.

Table 4
The parameters of the LSTM network.

	Neural units	Learning rate	Time step	Batch size
Range	5–40	0.0001–0.1	2–30	10–120
Result	27	0.01	5	80

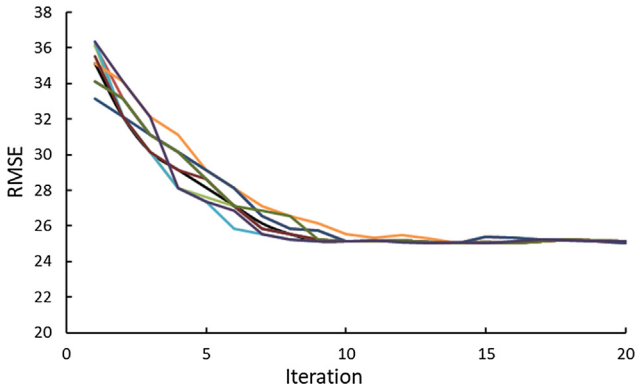

Fig. 13. RMSE of the PSO-LSTM.

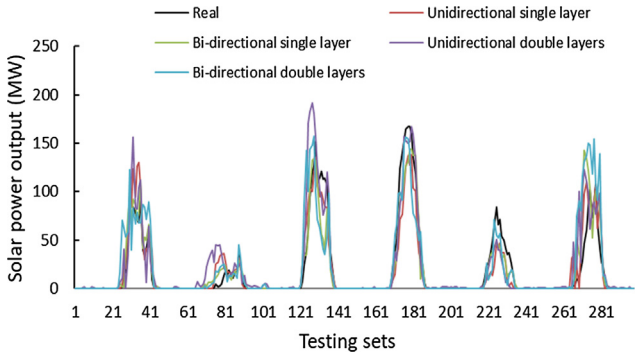

Fig. 14. The predicted results of testing sets.

Table 5
Comparison of different LSTM structures.

	Unidirectional single layer	Unidirectional double layers	Bi-directional single layer	Bi-directional double layers
MAE	8.32	10.66	8.14	11.76
RMSE	19.68	23.72	19.41	25.54

Table 6
The predicted errors of the different models in different cases.

		PSO-LSTM	Basic LSTM	Qing and Niu (2018)	XGBoost	ANN
MAE	Case 1	6.68	8.54	9.68	16.35	11.26
	Case 2	7.59	8.58	9.43	17.38	13.73
	Case 3	8.38	8.74	9.54	17.54	11.78
	Case 4	8.19	9.12	9.14	16.35	8.98
RMSE	Case 1	15.49	19.15	20.56	32.47	22.79
	Case 2	18.52	19.63	20.34	34.89	23.47
	Case 3	19.28	20.23	20.15	33.21	21.53
	Case 4	19.56	20.04	20.63	34.61	20.79

regardless of whether humidity is added or not, weather conditions can be fully considered in the prediction model, which is also illustrated in Section 3.1. The actual SPO and predicted results in the three different cases are shown in Fig. 11. Therefore, the input variables of weather data are represented by sunshine, temperature and wind speed.

The correlation of solar power outputs for all regions is then explored by analyzing their weather data correlations. The correlation coefficients of sunshine, temperature and wind speed in these five regions are shown in Fig. 12(a), (b) and (c), respectively. These three figures show that the temperatures of these five regions are highly correlated, followed by the correlations of sunshine and wind speed. In general, the weather conditions in these five regions have a certain

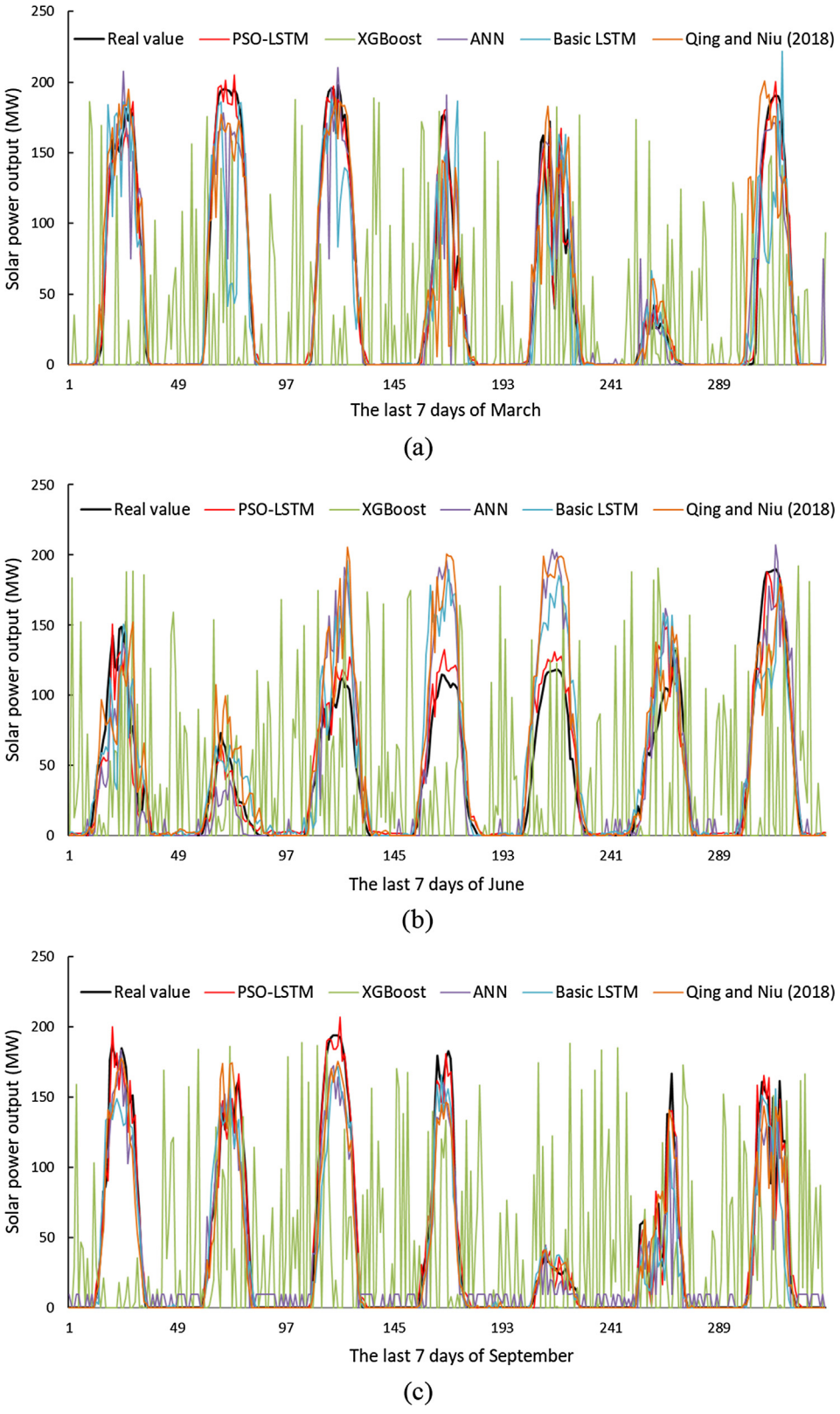

Fig. 15. The predicted results of different prediction models for (a) case 1, (b) case 2, (c) case 3 and (d) case 4.

correlation, but there is no high correlation or proportional relationship.

3.3. LSTM structure determination

In the process of the PSO-LSTM, the training sets account for 70%, the validation sets account for 10%, and the testing sets account for 20%. The RMSE of the testing data is considered as the fitness value of

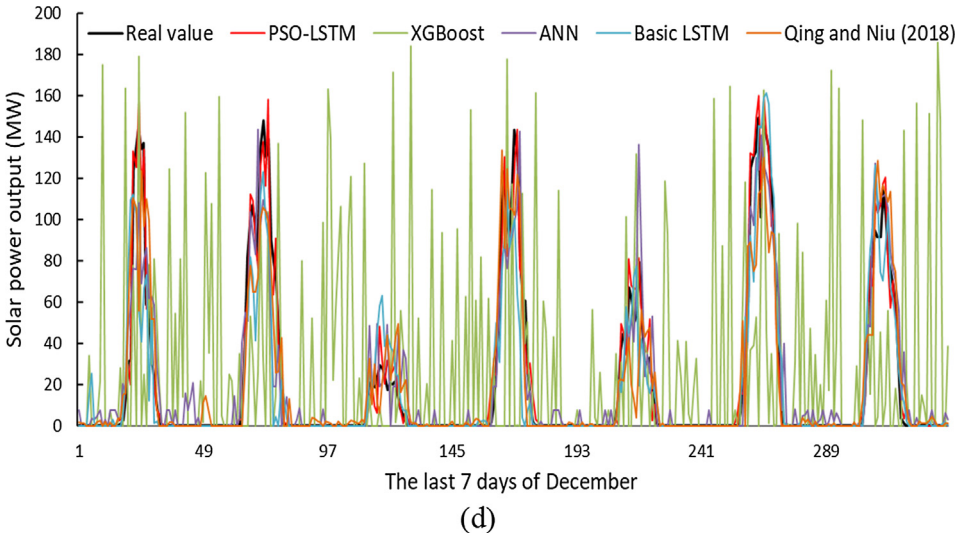

Fig. 15. (*continued*)

the PSO. The parameters of the LSTM model include neural units, learning rate, time step, batch size. As shown in Table 4, the range is 5–40, 0.0001–0.1, 2–30, and 10–120 for neural units, learning rate, time step, and batch size, respectively. After the training of the PSO-LSTM, the suitable parameters are determined, as shown in Table 4. The RMSE convergence process is presented in Fig. 13. Several tests are undertaken to ensure the stability of the PSO-LSTM. Finally, the neural units, learning rate, time step, and batch size are 27, 0.01, 5, and 80, respectively. After determining the parameters, different hidden layers, and whether the LTSM is unidirectional or bi-directional, are compared. The purpose of this part is to compare the training results of single-layer and multi-layer LSTM models. There are two main reasons for using double layers to represent multiple layers: (1) to some extent, the increase in the number of layer units can replace the increase in the number of layers; (2) the increase in the number of layers will make the network structure more complex and require longer training time. The real SPO and the predicted results of different LSTM structures are presented in Fig. 14. As shown in Table 5, for a bi-directional single layer, MAE is 8.14 and RMSE is 19.41, and they are the least in each of these four comparisons.

Through the above analysis and comparison, the best LSTM structure has the bi-directional single hidden layers, with parameters of 27, 0.01, 5, and 80 for neural units, learning rate, time step, and batch size, respectively.

3.4. Results and comparison

After determining the suitable LSTM structure, the SPO prediction model is established. In order to compare the prediction effects in different seasons, datasets of four different time ranges are used: January to March, January to June, January to September, January to and December. With the aim of verifying the superiority of the proposed model, the comparison is made with existing prediction models, such as basic LSTM, ANN [48], and XGBoost [49], and a previously proposed method [9]. 20% of the data for each case is used to test the accuracy of the prediction models. And the predicted errors of these prediction models for these four cases are presented in Table 6, which shows that the values of MAE and RMSE of PSO-LSTM are the smallest among the four prediction models for each case. In order to visually display the predicted results, the SPO prediction values of the last seven days of each case are shown in Fig. 15(a), (b), (c) and (d). Obviously, the predicted values of XGBoost have the largest deviation and are not suitable for predicting SPO. ANN performs best in case 4, which indicates that it requires more data to be trained. The method proposed by Qing and Niu [9] performs better and is more stable in these cases, but the prediction effect is not optimal. Compared to basic LSTM, PSO-LSTM performs better in each case, which shows that the LSTM network with optimized parameters is more suitable for SPO prediction. As shown in Fig. 15, the prediction accuracy of PSO-LSTM is better in (a), (b), (c) and (d). In terms of these curves, PSO-LSTM outperforms the other three prediction models.

In particular, PSO-LSTM performs better in the first case, then worsens and converges, most likely due to greater uncertainty in the weather data due to seasonal transitions. Therefore, establishing a SPO prediction model with recent weather data can improve the prediction accuracy. In addition, the complexity of multi-region prediction impedes accuracy. This complexity, combined with the uncertainty of weather data, poses more obstacles to the prediction model and is the difficulty of current research.

From the above analysis – i.e., determination of the input variables and the LSTM structure, sensitivity analysis, and prediction model comparison – it can be concluded that the PSO-LSTM model is effective for short-term SPO prediction. Table 6 and Fig. 15 indicate that the predicted results of the method have smaller errors and are better fitting. For the SPO prediction problem for solar power plants located in multiple regions, the sum of the output is available with the help of the PSO-LSTM model, which is input with temperature and global solar radiation values of measurement locations as well as the weather conditions of each solar power plant.

4. Conclusion

Solar power output prediction is extremely important because it can be a decision-making tool in power system operations. This paper focuses on designing the prediction model of Solar power output when solar plants are located in multiple regions, which has received minimal attention in the previous literature. In this work, the particle swarm optimization algorithm is applied to optimize the parameters of the long short-term memory model, and the sensitivity of the different divisions of datasets and the setting of parameters of the particle swarm optimization algorithm are discussed. Different long short-term memory structures are compared to determine the final prediction model. In order to demonstrate the effectiveness of the proposed method, a real power system with eight solar plants in five regions is taken as an example. Four cases containing different seasons then demonstrate the superiority of the proposed method over other prediction models.

In future work, other intelligent algorithms, such as differential

evolution, the simulated annealing algorithm, and ant colony optimization, can also be applied to select suitable parameters for long short-term memory. Besides, many improved long short-term memory models are emerging day by day. Models examining solar power output of multiple regions need to be built, increasing the predicting accuracy. At the same time, it is also important to study solar and wind energy access to power systems [50].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (51874325) and the Grant-in-Aid for Early-Career Scientists (19K15260) from the Japan Ministry of Education, Culture, Sports, Science and Technology. Additionally, thanks to the solar power output data provided by the “PV in HOKKAIDO” contest hosted by Tokyo Electric Power Company Holdings (TEPCO) and Hokkaido Electric Power Company (HEPCO).

References

[1] Wang Y. The analysis of the impacts of energy consumption on environment and public health in China. Energy 2010;35:4473–9.
[2] Sayigh A. Renewable energy—the way forward. Appl Energy 1999;64:15–30.
[3] Brunet C, Savadogo O, Baptiste P, Bouchard MA. Shedding some light on photovoltaic solar energy in Africa – A literature review. Renew Sustain Energy Rev 2018;96:325–42.
[4] Gorjian S, Zadeh BN, Eltrop L, Shamshiri RR, Amanlou Y. Solar photovoltaic power generation in Iran: Development, policies, and barriers. Renew Sustain Energy Rev 2019;106:110–23.
[5] Alamdari P, Nematollahi O, Alemrajabi AA. Solar energy potentials in Iran: A review. Renew Sustain Energy Rev 2013;21:778–88.
[6] Tascikaraoglu A, Sanandaji BM, Chicco G, Cocina V, Spertino F, Erdinc O, et al. Compressive spatio-temporal forecasting of meteorological quantities and photovoltaic power. IEEE Trans Sustainable Energy 2016;7:1295–305.
[7] Lin S, Li C, Xu F, Liu D, Liu J. Risk identification and analysis for new energy power system in China based on D numbers and decision-making trial and evaluation laboratory (DEMATEL). J Cleaner Prod 2018;180:81–96.
[8] Ehsan RM, Simon SP, Venkateswaran PR. Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. Neural Comput Appl 2017;28:3981–92.
[9] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 2018;148:461–8.
[10] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39:535–76.
[11] Sobri S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: A review. Energy Convers Manage 2018;156:459–97.
[12] Qin W, Wang L, Lin A, Zhang M, Xia X, Hu B, et al. Comparison of deterministic and data-driven models for solar radiation estimation in China. Renew Sustain Energy Rev 2018;81:579–94.
[13] Yagli GM, Yang DZ, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98.
[14] Long H, Zhang Z, Su Y. Analysis of daily solar power prediction with data-driven approaches. Appl Energy 2014;126:29–37.
[15] Amrouche B, Le Pivert X. Artificial neural network based daily local forecasting for global solar radiation. Appl Energy 2014;130:333–41.
[16] Kamadinata JO, Ken TL, Suwa T. Sky image-based solar irradiance prediction methodologies using artificial neural networks. Renewable Energy 2019;134:837–45.
[17] Li K, Wang R, Lei H, Zhang T, Liu Y, Zheng X. Interval prediction of solar power using an improved bootstrap method. Sol Energy 2018;159:97–112.
[18] İzgi E, Öztopal A, Yerli B, Kaymak MK, Şahin AD. Short–mid-term solar power prediction by using artificial neural networks. Sol Energy 2012;86:725–33.
[19] Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, et al. Machine learning methods for solar radiation forecasting: A review. Renewable Energy 2017;105:569–82.
[20] Zeng J, Qiao W. Short-term solar power prediction using a support vector machine. Renewable Energy 2013;52:118–27.
[21] Li L-L, Wen S-Y, Tseng M-L, Wang C-S. Renewable energy prediction: A novel short-term prediction model of photovoltaic output power. J Cleaner Prod 2019;228:359–75.
[22] Lan H, Zhang C, Hong Y-Y, He Y, Wen S. Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. Appl Energy 2019;247:389–402.
[23] Brockwell PJ, Davis RA, Calder MV. Introduction to time series and forecasting. Springer; 2002.
[24] Prema V, Rao KU. Development of statistical time series models for solar power prediction. Renewable Energy 2015;83:100–9.
[25] Reikard G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Sol Energy 2009;83:342–9.
[26] Amato U, Andretta A, Bartoli B, Coluzzi B, Cuomo V, Fontana F, et al. Markov processes and Fourier analysis as a tool to describe and simulate daily solar irradiance. Sol Energy 1986;37:179–94.
[27] Yang D, Dong Z. Operational photovoltaics power forecasting using seasonal time series ensemble. Sol Energy 2018;166:529–41.
[28] Costa-jussà MR, Allauzen A, Barrault L, Cho K, Schwenk H. Introduction to the special issue on deep learning approaches for machine translation. Comput Speech Lang 2017;46:367–73.
[29] Fayek HM, Lech M, Cavedon L. Evaluating deep learning architectures for speech emotion recognition. Neural Networks. 2017;92:60–8.
[30] Xing J, Li K, Hu W, Yuan C, Ling H. Diagnosing deep learning models for high accuracy age estimation from a single image. Pattern Recogn 2017;66:106–16.
[31] Zhou T, Han G, Xu X, Lin Z, Han C, Huang Y, et al. δ-agree AdaBoost stacked autoencoder for short-term traffic flow forecasting. Neurocomputing 2017;247:31–8.
[32] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.
[33] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. International conference on machine learning; 2013. p. 1310–8.
[34] Srivastava S, Lessmann S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Sol Energy 2018;162:232–47.
[35] Yu R, Gao J, Yu M, Lu W, Xu T, Zhao M, et al. LSTM-EFG for wind power forecasting based on sequential correlation features. Future Generation Comput Syst 2019;93:33–42.
[36] Peng L, Liu S, Liu R, Wang L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 2018;162:1301–14.
[37] Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: A review. Neural Networks 2019;113:54–71.
[38] Eberhart RC, Shi YH, Ieee I. Particle swarm optimization: Developments, applications and resources; 2001.
[39] Zhang CL, Shao HH, Li Y, Ieee I. Particle swarm optimisation for evolving artificial neural network. In: Smc 2000 conference proceedings: 2000 Ieee international conference on systems, man & cybernetics, vol. 1-52000. p. 2487–90.
[40] Kuo RJ, Hong SY, Huang YC. Integration of particle swarm optimization-based fuzzy neural network and artificial neural network for supplier selection. Appl Math Model 2010;34:3976–90.
[41] Zheng J, Zhang H, Yin L, Liang Y, Wang B, Li Z, et al. A voyage with minimal fuel consumption for cruise ships. J Cleaner Prod 2019;215:144–53.
[42] Zhang H, Yuan M, Liang Y, Wang B, Zhang W, Zheng J. A risk assessment based optimization method for route selection of hazardous liquid railway network. Saf Sci 2018;110:217–29.
[43] Zhang Y, Cui N, Feng Y, Gong D, Hu X. Comparison of BP, PSO-BP and statistical models for predicting daily global solar radiation in arid Northwest China. Comput Electron Agric 2019;164:104905.
[44] Eseye AT, Zhang J, Zheng D. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. Renewable Energy 2018;118:357–67.
[45] Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing. 2019;337:325–38.
[46] Kennedy J, Eberhart R. Particle swarm optimization. IEEE international conference on neural networks, 1995 proceedings, vol. 4; 2002. p. 1942–8.
[47] Ahmad A, Anderson TN, Lie TT. Hourly global solar irradiation forecasting for New Zealand. Sol Energy 2015;122:1398–408.
[48] Elsheikh AH, Sharshir SW, Abd Elaziz M, Kabeel AE, Guilan W, Haiou Z. Modeling of solar energy systems using artificial neural network: A comprehensive review. Sol Energy 2019;180:622–39.
[49] Fan J, Wang X, Wu L, Zhou H, Zhang F, Yu X, et al. Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Convers Manage 2018;164:102–11.
[50] Heydari A, Astiaso Garcia D, Keynia F, Bisegna F, De Santoli L. A novel composite neural network based method for wind and solar power forecasting in microgrids. Appl Energy 2019;251:113353.