IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 23, 2019, accepted June 10, 2019, date of publication June 14, 2019, date of current version June 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923006

Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism

HANGXIA ZHOU[1], YUJIN ZHANG[1], LINGFAN YANG[1], QIAN LIU[1], KE YAN[1,2], (Member, IEEE), AND YANG DU[3], (Member, IEEE)

[1]Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, Hangzhou 310018, China
[2]Department of Building, School of Design and Environment, National University of Singapore, Singapore 117566
[3]College of Science and Engineering, James Cook University, Cairns, QLD 4870, Australia

Corresponding author: Ke Yan (e-mail: yanke@cjlu.edu.cn)

This work was supported in part by the Public Welfare Research Project of Zhejiang Province, China, under Grant LGF18F020017, and in part by the National Natural Science Foundation of China under Grant 61850410531 and Grant 61803315.

ABSTRACT Photovoltaic power generation forecasting is an important topic in the field of sustainable power system design, energy conversion management, and smart grid construction. Difficulties arise while the generated PV power is usually unstable due to the variability of solar irradiance, temperature, and other meteorological factors. In this paper, a hybrid ensemble deep learning framework is proposed to forecast short-term photovoltaic power generation in a time series manner. Two LSTM neural networks are employed working on temperature and power outputs forecasting, respectively. The forecasting results are flattened and combined with a fully connected layer to enhance forecasting accuracy. Moreover, we adopted the attention mechanism for the two LSTM neural networks to adaptively focus on input features that are more significant in forecasting. Comprehensive experiments are conducted with recently collected real-world photovoltaic power generation datasets. Three error metrics were adopted to compare the forecasting results produced by attention LSTM model with state-of-art methods, including the persistent model, the auto-regressive integrated moving average model with exogenous variable (ARIMAX), multi-layer perceptron (MLP), and the traditional LSTM model in all four seasons and various forecasting horizons to show the effectiveness and robustness of the proposed method.

INDEX TERMS PV power generation, short-term forecasting, long short term memory, attention mechanism.

I. INTRODUCTION

Photovoltaic power is known as a clean, safe, sustainable and renewable energy, which is widely applied to replace fossil fuel power resources in the near future. The PV market system has a rapid development pace since 2007; and PV power generation, transmission, maintenance and consumption have become important components in the construction process of smart power grids [1], [2]. However, the energy generated by PV systems is always influenced by various factors, such as sunshine, cloud cover, temperature and relative humidity. When large-scale PV power plants are connected to the grid, the stability and safety of the power grid will be seriously threatened. Short-term PV forecasting is therefore one of the key technologies towards the solution stabilizing the PV power generation [3].

PV power generation forecasting is a long existed problem, which attracts a broad range of attentions. Various methods have been proposed in the literature, including the numerical weather prediction (NWP) model, statistical methods, machine learning methods and hybrid methods. The NWP model focuses on forecasting meteorological parameters, such as irradiance, atmospheric temperature, and wind speed. The forecasted meteorological parameters are then inserted into a physical model to forecast PV power generation [4]. Pierro *et al.* [5] employed the NWP model to forecast 1 day-ahead PV power generation. An ensemble model was proposed after comparing a variety of

The associate editor coordinating the review of this manuscript and approving it for publication was Lin Zhang.

TABLE 1. Abbreviation and full name.

Abbreviation	Full name
ALSTM	Attention Long Short Term Memory neural network
ANN	Artificial Neural Network
ARIMAX	AutoRegressive Integrated Moving Average Model-X
BRT	Boosted Regression Trees
ELM	Extreme Learning Machine
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short Term Memory neural network
MLP	Multi-Layer Perceptron
NWP	Numerical Weather Prediction
PM	Persistent Model
PSO	Particle Swarm Optimization
PV	Photovoltaic
SVM	Support Vector Machine

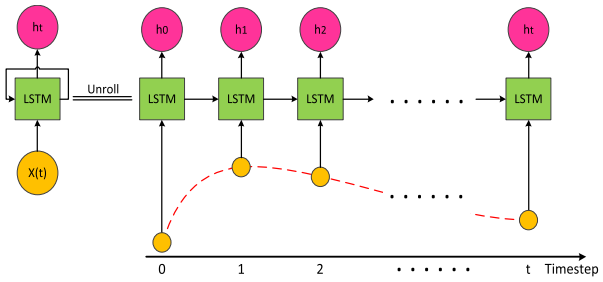

FIGURE 1. Unrolled LSTM uses time series data as input.

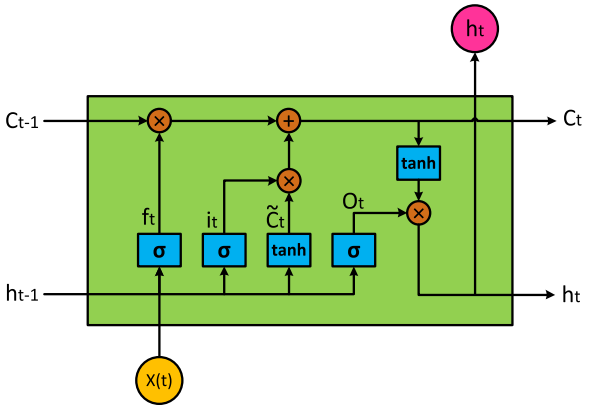

FIGURE 2. The internal structure of a LSTM unit, The symbols ⊗ and ⊕ represent pointwise scalar multiplication and the sum function, respectively.

forecasting models. Chen *et al.* [6] proposed an ultra-short time PV power forecasting method using the ground-based sensors; and a ramp rate control method is proposed to smooth the PV power output based on the forecasting results.

Statistical methods and machine learning methods are data-driven methods that use historical records of power generation, solar irradiance, atmospheric temperature, humidity, etc. to build forecasting models [7], [8]. Massidda and Marrocu [9] used the historical meteorological data and historical PV power data to establish the multilinear adaptive regression splines model; this model used weather forecast as input to forecast the power of a PV plant in Borkum, Germany. Machine learning is a popular forecasting method for time series data forecasting, such as the PV power outputs and other similar applications [10]–[12]. Mellit *et al.* [13] divided the data into three types: sunny, partly cloudy and overcast, according to the mean value of the solar irradiance. An artificial neural network (ANN) model was constructed using different types of data to carry out short-term PV power forecasting. Behera *et al.* [14] used the extreme learning machine (ELM) to forecast PV power in 15 min, 30 min and 60 min horizons. The particle swarm optimization (PSO) method is employed to optimize the ELM. Experimental results showed that the performance of ELM was better than that of ANN Huang *et al.* [15] proposed a data-driven framework based on fusing spatial and temporal information of the target PV station and its neighboring PV stations. Four data-driven models, including boosted regression tree (BRT), ANN, support vector machine (SVM), and least absolute shrinkage and selection operator (LASSO) regression model were employed for multi-step forward forecasting of solar irradiance. Experiments show that information from neighbor PV stations is useful for improving the forecasting accuracy.

Hybrid model methods that combine multiple cross-discipline methods are proposed to deal with various forecasting problems. Eseye *et al.* [16] proposed a hybrid forecasting model that combines wavelet transform, SVM and PSO. The wavelet transform was used to decompose NWP meteorological data and photovoltaic power time series into multiple subsequences. The subsequences obtained by wavelet transform were used to train SVM; and the PSO was used to optimize SVM parameters. Yao *et al.* [17] proposed a forecast model based on echo state network. This model uses the restricted Boltzmann machine to extract the relative feature of input data, and uses the principal component analysis to extract the main feature. The Davidon–Fletcher–Powell quasi-Newton algorithm is used to optimize the reservoir parameters of echo state network.

Deep learning is currently one of the hottest research areas of machine learning and artificial intelligence. The biggest difference between deep learning and traditional machine learning methods is that deep learning can automatically learn useful features from data instead of using traditional feature selection methods. Deep learning has been applied in the field of PV power forecasting and achieved outstanding results [18], [19]. LSTM, as one of the most important deep learning technique, is frequently applied in the related works [20].

Targeting at the real-world short-term photovoltaic power forecasting problem, in this study, we proposed a hybrid ensemble deep learning framework that combines the attention mechanism with multiple LSTM (ALSTM) models.

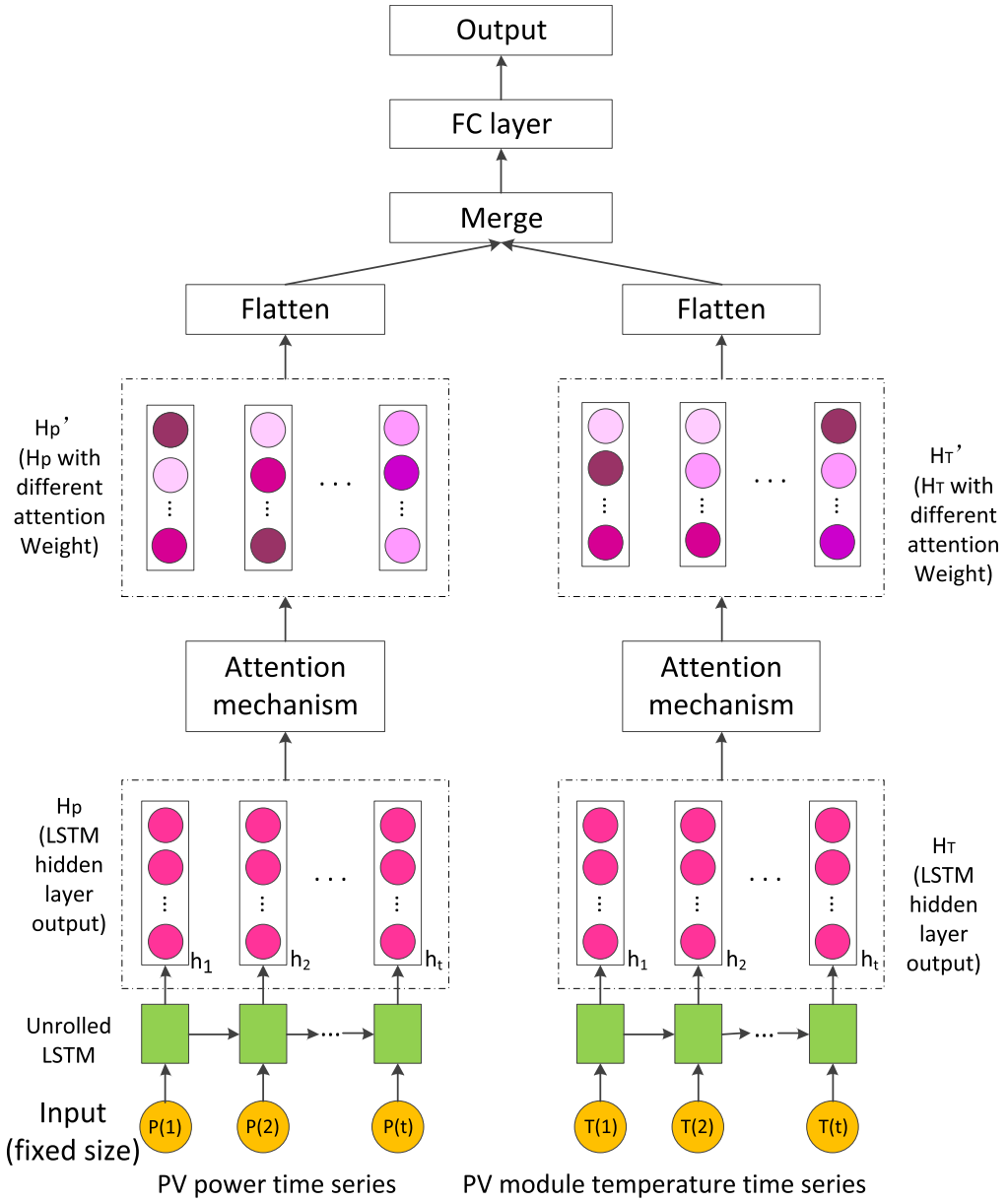

FIGURE 3. The proposed ALSTM model structure.

The ALSTM model uses the attention mechanism to give greater weight to the values associated with the model output in the LSTM hidden layer output vector. The PV power time series and the PV module temperature time series were used as ALSTM model input to forecast the PV power at the next moment.

In conclusion, this study has the following main contributions to the literature:

- **An ensemble deep learning framework.**We proposed an ensemble deep learning forecasting method that applies two LSTM neural networks on temperature and power output time series data forecasting, respectively. The two prediction results are flattened and merged through a fully connected (FC) layer to enhance the prediction accuracy.
- **A novel hybrid deep learning method.**The ensemble deep learning framework is further integrated with the attention mechanism to forecast photovoltaic power outputs. The attention mechanism allows the two LSTM neural networks to adaptively focus on input features that are more significant in forecasting. To our best knowledge, the entire proposed method is novel on the topic of solar energy power generation forecasting.
- **A comprehensive comparative study with state-of-art methods tested on real-world datasets and different time horizons.**The performance of ALSTM model is justified using real-world solar energy generation data that is collected in recent years. A comprehensive comparative study is conducted comparing the proposed method with state-of-art methods, including

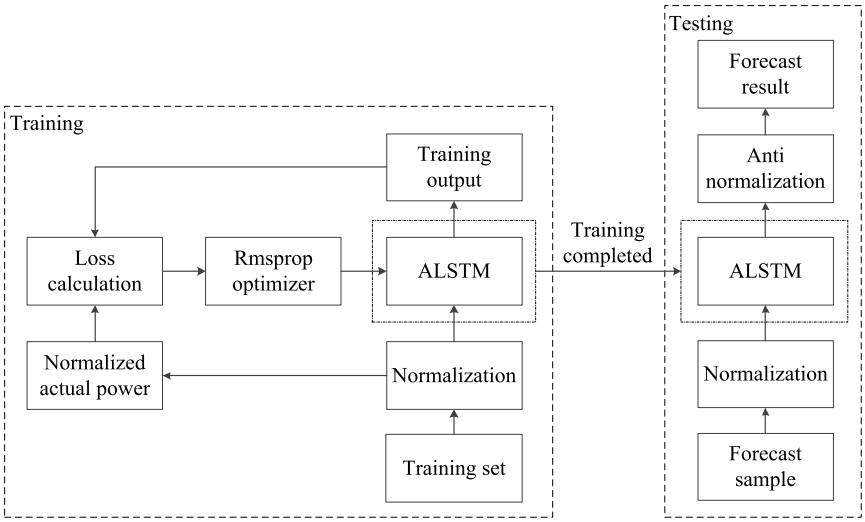

FIGURE 4. Detailed training and testing procedures for the proposed framework.

the persistence model, ARIMAX, MLP model and a single LSTM model on all four seasons and different time horizons of 7.5 min-ahead, 15 min-ahead, 30 min-ahead and 60 min-ahead. The experimental results show that the proposed ALSTM model has obvious advantage on forecasting accuracy.

II. METHODOLOGY DESCRIPTION

A. LSTM NEURAL NETWORK

Recurrent neural network introduces the concept of timing to the traditional neural network structure to make it adaptive to time horizon dependencies. For traditional RNN, there is always a problem of gradient disappearance or gradient explosion, resulting in training failures [22]. Aiming at the gradient vanishing problem, Hochreiter *et al.* proposed LSTM, which enables the recurrent neural network to process time series data effectively [23]. Fig.1 illustrates the underlining working mechanism of a LSTM neural network. A LSTM neural network consists of a series of LSTM cells. $h_0, h_1, \ldots, h_t$ are time series data samples that are feed into different LSTM cells simultaneously as inputs.

Fig.2 shows the internal structure of an LSTM cell. The internal cell states are denoted by C_t. LSTM updates, maintains, or deletes cell state information using forget gate f_t, input gate i_t and output gate O_t, respectively. At time t, the input is the sequence vector $X(t)$, hidden layer output h_{t-1} and cell state C_{t-1}. The outputs are LSTM hidden layer output h_t and cell state C_t. Forget gate, input gate, and output gate are calculated according to Equations (1), (2), and (3):

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

$$i_t = \sigma(W_i \bullet [h_{t-1}, x_t] + b_i) \quad (2)$$

$$O_t = \sigma(W_O \bullet [h_{t-1}, x_t] + b_o) \quad (3)$$

The current candidate cell state $\tilde{C}$ is calculated by Equation (4):

$$\tilde{C}_t = \tanh(W_C \bullet [h_{t-1}, x_t] + b_C) \quad (4)$$

TABLE 2. Mean and variance of daily power.

	2014	2015	2016	2017	2018
mean	5.33	4.77	4.73	4.99	5.21
variance	15.41	15.14	15.37	14.98	16.18

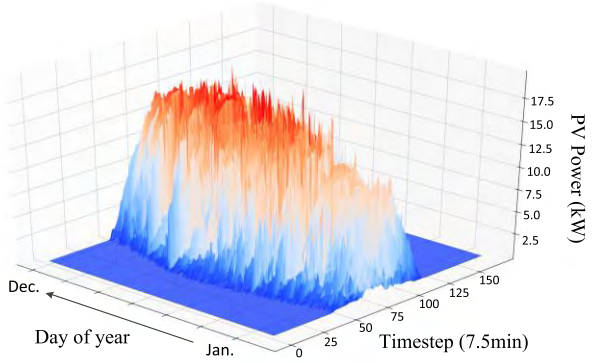

FIGURE 5. PV power generation data in three dimension.

The forget gate and the input gate determine the proportions of information occupied by the C_{t-1} and the $\tilde{C}$ in the current cell state C_t, respectively. The state C_t was updated by the Equation (5):

$$C_t = f_t \bullet C_{t-1} + i_t \bullet \tilde{C}_t \quad (5)$$

The current output of the hidden layer is calculated by Equation (6):

$$h_t = O_t \bullet \tanh(C_t) \quad (6)$$

where W_f, W_i and W_o are forget gate, input gate, and output gate weight matrix, respectively; and b_f, b_i and b_o are forget gate, input gate, and output gate bias, respectively.σ stands for the sigmoid activation function.

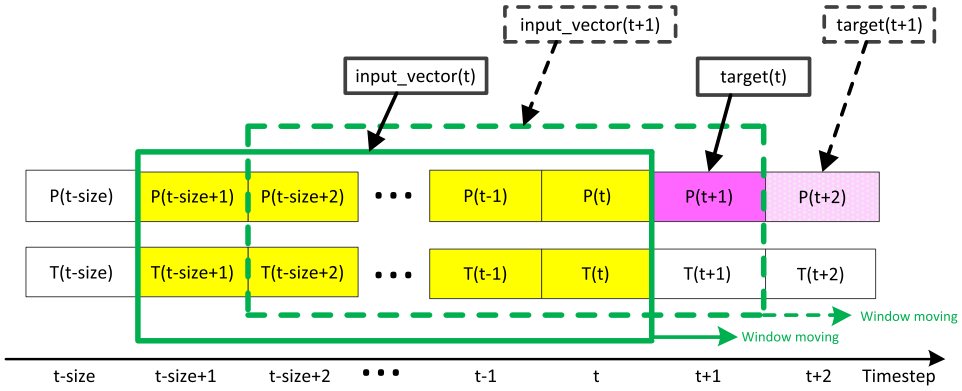

FIGURE 6. Window moves over time series.

B. ATTENTION MECHANISM

Attention in the biological vision system allows the animal to focus on the specific objects for observations. The attention mechanism is a neural network that simulates the attentions of brain. It has been successfully applied to machine translation [24], video analysis [25] and other related fields. Applying the attention mechanism to the deep neural network allows the neural network to adaptively focus on input features that are more important to the current output, and mitigate the interference of other features. Using the LSTM hidden layer output vector$H = \{h_1, h_2, \ldots, h_t\}$ as the input of the attention mechanism, the attention mechanism will look for the attention weight α_i of h_i, which can be calculated as shown in Equations (7) and (8):

$$e_i = \tanh(W_h h_i + b_h),\ e_i \in [-1, 1] \tag{7}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^{t} \exp(e_i)},\ \sum_{i=1}^{t} \alpha_i = 1 \tag{8}$$

where W_h is the weight matrix of h_i; and b_h is the bias. The values of W_h and b_h vary during the ALSTM training process. The attention vector $H^{'} = \{h_1^{'}, h_2^{'}, \ldots h_t^{'}\}$ can be obtained by multiplying attention weight α_i and h_i:

$$h_i^{'} = \alpha_i \bullet h_i \tag{9}$$

The attention mechanism is implemented as a custom layer where the parameters are optimized using RMSProp back-propagation [26].

C. PROPOSED MODEL

The proposed ALSTM model structure is shown in Figure 3. The ALSTM model uses the fixed size PV power time series and PV module temperature time series as inputs to forecast the next moment power. The PV power and PV module temperatures contain different information. Therefore, feature extractions of these two quantities are performed separately. LSTM is used to extract the time series features of PV power; and the attention mechanism is used to process the LSTM hidden layer output H_p. Each element in H_p is given different attention weights to obtain $H_p^{'}$. The same process was used for PV module temperature to get $H_T^{'}$. The proposed model was implemented using Keras (version 2.1.6). For both ALSTM neural networks, the two outputs $H_p^{'}$ and $H_T^{'}$ are 3-dimensional arrays, including batch size, time steps and the number of units. Then, the two 3-dimensional arrays are flattened into two dimensional vectors using flatten layers and merged into a single 2-dimensional array to meet the input requirement of the fully connected (FC) layer.

D. TRAINING AND TESTING

The PV power and the PV module temperature data from 2014 to 2016 is taken as the training dataset; and the data from 2017 to 2018 is used for testing.

The training and testing process of the ALSTM model is shown in Figure 4. During the model training, the MSE loss is calculated according to the training output and the normalized actual power. The RMSProp optimizer is used to optimize the entire network to minimize the MSE loss. The ALSTM model predicts the forecast samples and outputs the forecast results in the testing phase.

III. RESULTS

A. DATA DESCRIPTION AND PROCESSING

The data used in the experiments is collected from a 20kW rooftop PV power station located at Shaoxing city, Zhejiang Province, China (120°23′E, 29°72′N). The PV power and the PV module temperature are monitored from October 2014 to September 2018, in a time interval of 7.5 min. The mean and variance of daily PV power from 2014 to 2018 were shown in Table II.

From the data, power generation distribution in year 2016 is shown in Figure 5. The peak and the start/stop time of power generation in different seasons were different. The peak power in January and February (winter) is low; and the daily generation time is short. The peak power in summer is high; and the daily generation time is longer. Due to the seasonal characteristics of photovoltaic power generation, the

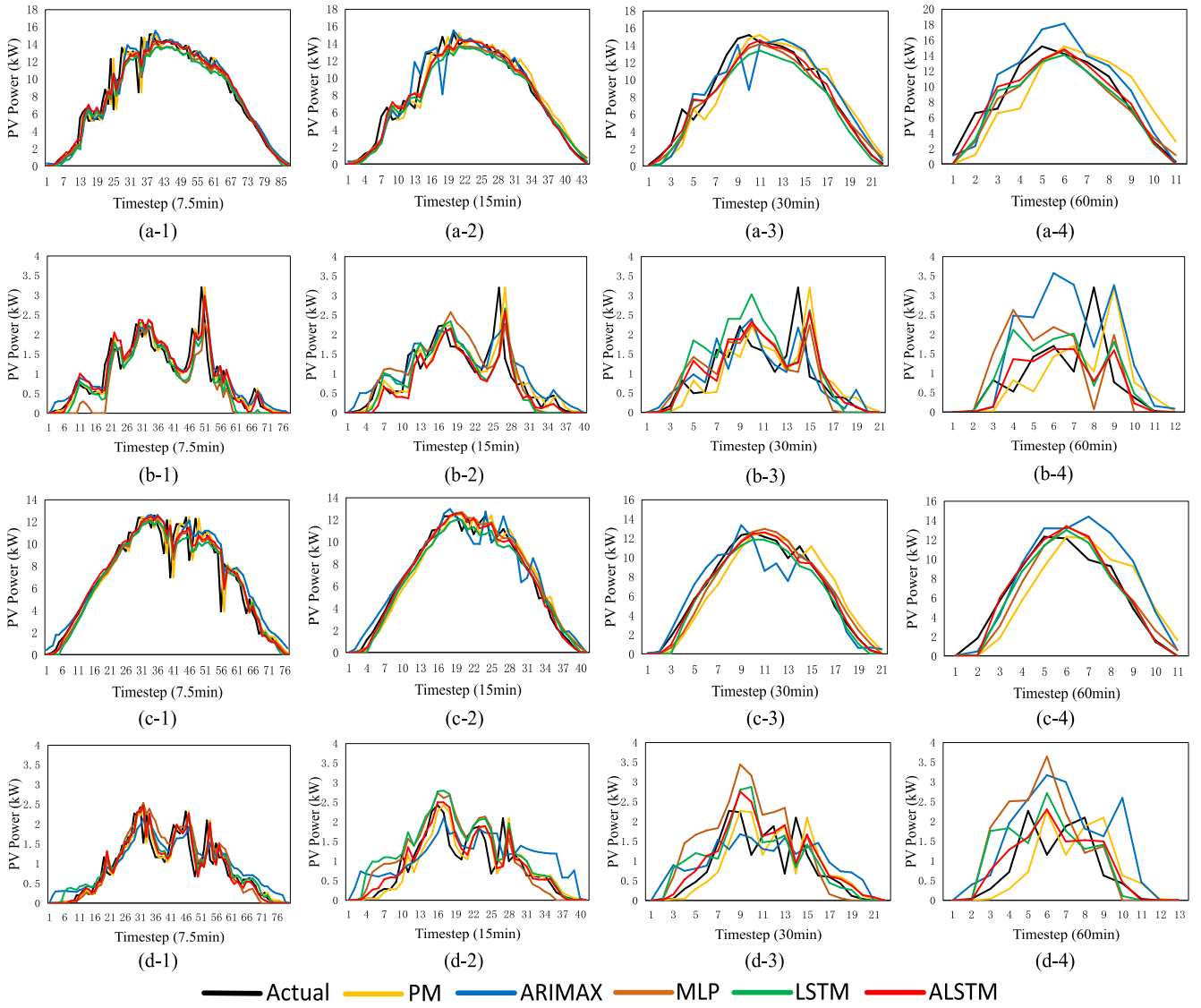

FIGURE 7. PV power forecast result in 7.5 min-ahead, 15 min-ahead, 30 min-ahead, and 60 min-ahead forecasting horizons. (a) forecast results for spring days in 2018, (b) forecasts result for summer days in 2018, (c) forecasts result for autumn days in 2017, (d) forecasts result for winter days in 2018.

forecast model is demanded to be distinguished over different seasons.

In the data pre-processing phase, normalization is performed according to Equation 10:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where x_i is the original data and x_i^* is the normalized data.

The training sample is obtained by the sliding window method as shown in Figure 6. A fixed-size window slides on the PV power time series and the PV module temperature time series, the value in the window is taken as the training input vector, and the PV power at the next moment is used as the training target. During the test, the power value and the temperature value in the current sliding window are used as model inputs to forecast the PV power at the next moment.

B. METRICS

Three evaluation metrics, MAPE, RMSE and MAE are used to evaluate the performance of the forecast model.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_{model,i} - x_{actual,i}|}{x_{actual,i}} \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{model,i} - x_{actual,i})^2} \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_{model,i} - x_{actual,i}| \tag{13}$$

where x_{model} is the forecasted value of the model and x_{actual} is the actual value.

TABLE 3. Comparison of MAPE for four forecast horizons in each month of a year.

		2017.10	2017.11	2017.12	2018.01	2018.02	2018.03	2018.04	2018.05	2018.06	2018.07	2018.08	2018.09	overall
7.5min	PM	23.22	28.11	17.90	23.92	21.96	22.15	18.43	26.20	22.74	20.77	22.02	25.88	22.78
	ARIMAX	27.69	28.73	24.60	20.78	23.72	23.42	21.69	26.74	28.42	24.72	24.93	27.72	25.26
	MLP	29.78	42.52	22.48	28.36	24.90	27.06	17.99	30.64	30.02	27.04	32.09	31.66	28.71
	LSTM	28.16	31.23	23.52	32.66	29.08	30.65	26.13	32.73	25.77	29.13	25.93	29.81	28.73
	ALSTM	21.98	30.33	18.99	27.72	25.87	27.73	18.13	28.12	26.09	23.85	21.15	25.85	24.65
15min	PM	28.48	32.69	25.49	30.65	27.57	28.76	25.14	32.79	29.25	26.84	29.68	32.19	29.13
	ARIMAX	34.24	35.20	30.53	26.93	30.96	29.18	26.61	32.56	33.67	32.75	31.31	31.38	31.28
	MLP	30.41	37.82	26.77	34.41	30.94	29.29	25.20	38.08	33.65	33.03	30.80	38.83	32.44
	LSTM	28.90	30.87	31.50	31.14	42.45	33.70	22.73	33.78	29.18	29.52	24.96	27.09	30.49
	ALSTM	27.49	31.35	25.49	30.87	35.84	29.04	24.44	30.98	28.13	26.48	24.70	30.86	28.81
30min	PM	36.84	39.53	34.18	38.91	38.21	36.43	31.80	41.09	37.01	33.13	37.37	42.70	37.27
	ARIMAX	40.77	40.72	36.51	33.28	38.00	37.51	33.16	37.60	40.91	35.06	38.11	39.45	37.59
	MLP	37.12	35.59	34.30	41.84	40.68	33.96	30.65	43.46	38.99	33.15	34.92	43.84	37.38
	LSTM	32.47	35.14	32.73	39.10	38.56	31.74	28.40	35.70	33.46	28.61	34.84	32.19	33.58
	ALSTM	32.06	33.86	30.95	39.65	36.72	31.48	26.03	36.29	30.70	28.08	27.22	33.17	32.18
60min	PM	50.70	50.50	46.94	50.12	46.68	49.93	43.16	48.61	45.31	43.46	46.93	48.00	47.53
	ARIMAX	48.93	55.13	53.75	52.11	48.79	48.49	45.45	52.63	50.60	46.29	41.73	45.44	49.11
	MLP	41.89	41.26	37.92	55.38	44.64	40.96	32.18	45.99	40.06	36.79	35.70	46.85	41.64
	LSTM	36.96	40.16	37.68	46.90	42.25	40.74	34.42	42.10	40.24	39.49	40.12	40.32	40.12
	ALSTM	36.56	37.29	35.95	50.98	36.21	34.21	32.49	43.97	37.91	34.74	33.81	39.76	37.82

C. BENCHMARK MODEL DESCRIPTION

Persistence model (PM): PM is widely used benchmark model, which is effective for short-term forecasting problems [27]. The persistence model suggests that the PV power at the next moment will be the same as the current PV power, it can be expressed as Equation 14:

$$\hat{P}(t+1) = P(t) \tag{14}$$

where $\hat{P}(t+1)$ is the forecasted value of PV power at the next moment and $P(t)$ is the current PV power.

Auto-Regressive Integrated Moving Average Model With Exogenous Variable (ARIMAX): ARIMAX is a time series forecasting algorithm in statistical model. ARIMAX requires data to be stable or stable after data difference, and is often used for short-term photovoltaic power forecasting. It adds a data differential operation to make the prediction more stable. This model uses PV power as endogenous variable, and use PV module temperature as exogenous variables.

MLP Model: MLP is a shallow neural network with good predictive ability, it has many researches and applications in the field of PV power forecasting [28]. This model takes the PV power time series and the PV module temperature time series as input, and uses two hidden layers to output the predicted PV power.

LSTM Model: The LSTM model uses two LSTM layers to extract features from the PV power time series and PV module temperature time series respectively; The hidden layer output of LSTM is expanded into a one-dimensional vector and merged; The last layer uses a fully connected layer to output the forecasted PV power outputs. This single LSTM model is used to compare with the ALSTM model to investigate the impact of attention mechanism on LSTM.

D. FORECASTING RESULT

This paper uses Keras version 2.1.6 to implement the ALSTM model. The LSTM layer uses 32 units and the full connection layer uses 128 hidden units. We compared the forecasting results of the proposed ALSTM model with PM, ARIMAX, MLP and LSTM models in 7.5 min-ahead, 15 min-ahead,

TABLE 4. Comparison of RMSE for four forecast horizons in each month of a year.

		2017.10	2017.11	2017.12	2018.01	2018.02	2018.03	2018.04	2018.05	2018.06	2018.07	2018.08	2018.09	over all
	PM	1.56	1.59	0.86	0.92	1.11	1.32	1.45	1.66	1.47	1.74	2.21	2.19	1.51
	ARIMAX	1.55	1.67	1.26	0.80	1.11	1.29	1.46	1.69	1.71	1.72	2.08	2.25	1.55
7.5min	MLP	1.59	1.73	0.85	0.94	1.08	1.38	1.37	1.66	1.51	1.77	2.31	2.20	1.53
	LSTM	1.36	1.55	0.79	0.86	1.09	1.24	1.43	1.56	1.41	1.66	2.03	2.04	1.42
	ALSTM	1.47	1.45	0.78	0.86	1.07	1.27	1.28	1.56	1.37	1.55	2.07	1.94	1.39
	PM	1.84	1.73	1.02	1.11	1.32	1.55	1.53	2.04	1.81	1.99	2.55	2.55	1.75
	ARIMAX	1.87	1.93	1.57	1.07	1.43	1.57	1.65	1.96	2.26	2.92	3.00	2.60	1.99
15min	MLP	1.73	1.58	1.00	1.06	1.22	1.56	1.48	2.00	1.83	2.06	2.48	2.52	1.71
	LSTM	1.62	1.53	0.92	1.00	1.23	1.48	1.43	1.90	1.78	1.88	2.44	2.16	1.61
	ALSTM	1.64	1.51	0.83	1.06	1.24	1.42	1.53	1.89	1.73	1.95	2.29	2.14	1.60
	PM	1.99	1.77	1.36	1.40	1.69	1.94	2.06	2.47	2.28	2.35	2.89	3.28	2.12
	ARIMAX	2.73	3.09	2.15	1.56	1.94	2.38	2.45	2.46	2.86	2.81	3.87	3.30	2.63
30min	MLP	1.90	1.64	1.12	1.18	1.40	1.72	1.75	2.34	2.15	2.17	2.73	3.00	1.93
	LSTM	1.88	1.49	1.08	1.14	1.32	1.69	1.71	2.30	2.06	2.05	2.71	2.70	1.84
	ALSTM	1.74	1.47	1.16	1.16	1.40	1.61	1.68	2.32	2.05	2.07	2.34	2.74	1.81
	PM	2.84	2.28	2.25	2.17	2.51	2.84	3.03	2.95	2.76	3.08	3.38	4.01	2.84
	ARIMAX	2.49	3.03	2.95	2.28	2.31	2.46	2.73	2.96	2.94	3.00	2.96	3.49	2.80
60min	MLP	2.29	1.80	1.70	1.67	1.90	2.05	2.15	2.62	2.36	2.62	2.77	3.19	2.26
	LSTM	1.89	1.56	1.40	1.45	1.75	2.10	1.90	2.31	2.58	2.88	2.84	2.70	2.11
	ALSTM	1.99	1.53	1.17	1.79	1.75	2.09	1.93	2.46	2.19	2.38	2.55	3.25	2.09

30 min-ahead, and 60 min-ahead horizons in spring, summer, autumn and winter (Figure 7). All experiments were repeated ten times by rebuilding the two LSTM models with random initialization. And all results stated in this section are averaged over the ten repeated runs. The differences between the five models are small for 7.5 min time interval. With the expansion of the forecasting time horizon, the forecasting curves of each model gradually deviate from the actual curve. The PM has advantages in the 15min and 7.5min forecasting. Especially in the case of sunny days, the forecast curve of persistence model is close to the actual curve, but the persistence model forecast curve showed significant hysteresis at 30 min and 60 min time interval experiments. ARIMAX has a good prediction result with 7.5 min forecasting. When the PV power fluctuates, the forecasting curve of ARIMAX fluctuates greatly, which results in larger prediction error. MLP and LSTM outperform PM and ARIMAX for most of the cases in 30 min interval, but still have a gap compared to ALSTM. ALSTM model effectively forecasts PV power in all four seasons; and the forecasting curves of the ALSTM model are very close to the actual curves in all different time intervals.

In order to investigate the performance of ALSTM model in different time horizons, experiments were carried out at 7.5 min, 15 min, 30 min and 60 min horizons to forecast the PV power over one year, i.e., from October 2017 to September 2018. Tables III – V list the MAPE, RMSE and MAE values of the five compared models, in each month, over the whole year, respectively. ARIMAX is suitable for forecast below 15 minutes, but has higher error rates in other horizons. LSTM has the ability to extract time series features, and its three error metric values are lower than MLP. The attention mechanism can further improve the forecast accuracy of LSTM. The proposed ALSTM model has the best performance over all compared methods in general.

Figure 8 shows the MAPE, RMSE, and MAE curves of the five compared forecasting models over different time horizons. PM and ARIMAX performed better at 7.5 min and 15 min time intervals forecasting. However, while the time horizon expands, the MAPE, RMSE and MAE values of

TABLE 5. Comparison of MAE for four forecast horizons in each month of a year.

		2017.10	2017.11	2017.12	2018.01	2018.02	2018.03	2018.04	2018.05	2018.06	2018.07	2018.08	2018.09	over all
7.5min	PM	0.81	0.96	0.44	0.47	0.58	0.69	0.79	0.97	0.81	0.86	1.08	1.32	0.82
	ARIMAX	0.90	1.00	0.68	0.41	0.60	0.72	0.84	1.02	1.03	1.00	1.21	1.38	0.90
	MLP	0.94	1.28	0.45	0.52	0.61	0.79	0.81	1.05	0.91	1.04	1.36	1.47	0.94
	LSTM	0.75	1.14	0.48	0.44	0.67	0.69	0.91	0.96	0.84	1.02	1.06	1.37	0.86
	ALSTM	0.80	0.90	0.40	0.42	0.64	0.71	0.75	0.94	0.78	0.85	1.08	1.28	0.80
15min	PM	1.03	1.06	0.68	0.67	0.81	0.93	1.02	1.26	1.08	1.12	1.43	1.67	1.06
	ARIMAX	1.20	1.23	0.92	0.58	0.84	0.97	1.07	1.30	1.40	1.73	1.87	1.75	1.24
	MLP	1.00	1.07	0.73	0.65	0.73	0.93	1.00	1.29	1.15	1.29	1.46	1.72	1.09
	LSTM	0.91	0.97	0.63	0.57	0.75	0.90	0.96	1.21	1.13	1.10	1.41	1.47	1.00
	ALSTM	0.91	0.99	0.57	0.68	0.80	0.82	1.06	1.17	1.10	1.16	1.30	1.49	1.00
30min	PM	1.36	1.29	1.07	0.96	1.22	1.33	1.54	1.68	1.50	1.56	1.94	2.35	1.48
	ARIMAX	1.87	2.18	1.32	0.92	1.22	1.50	1.57	1.66	1.90	1.93	2.54	2.42	1.75
	MLP	1.23	1.23	0.80	0.74	0.88	1.12	1.28	1.57	1.47	1.41	1.77	2.15	1.30
	LSTM	1.23	1.09	0.82	0.72	0.84	1.17	1.25	1.52	1.32	1.27	1.86	1.86	1.25
	ALSTM	1.12	1.09	0.94	0.76	0.94	1.05	1.21	1.53	1.35	1.24	1.46	1.92	1.22
60min	PM	2.04	1.68	1.83	1.47	1.83	2.09	2.38	2.19	1.97	2.35	2.62	3.02	2.12
	ARIMAX	1.80	2.31	2.08	1.56	1.57	1.71	1.86	2.08	2.03	2.16	2.08	2.58	1.98
	MLP	1.55	1.31	1.43	1.12	1.30	1.41	1.67	1.89	1.67	1.86	1.88	2.42	1.63
	LSTM	1.21	1.15	1.05	1.02	1.18	1.46	1.46	1.64	1.74	2.03	1.89	1.95	1.48
	ALSTM	1.36	0.99	0.81	1.26	1.21	1.52	1.48	1.79	1.48	1.61	1.65	2.42	1.47

TABLE 6. Comparison of t statistic and p value of LSTM and ALSTM for four forecast horizons in each month of a Year.

		2017.10	2017.11	2017.12	2018.01	2018.02	2018.03	2018.04	2018.05	2018.06	2018.07	2018.08	2018.09	over all
7.5min	T	1.1404	2.3062	-0.5379	-1.1986	-1.8938	1.3620	-2.3535	-1.5766	-0.1288	-0.1968	-0.2562	-1.6370	-1.5503
	p	0.2542	0.0214	0.5907	0.2308	0.0584	0.1733	0.0187	0.1150	0.8975	0.8440	0.7978	0.1020	0.1211
15min	T	1.1156	0.4711	-1.8631	1.0997	-3.2737	-1.6463	3.3015	-0.7247	1.9957	3.0190	2.0939	0.2367	1.9190
	p	0.2648	0.6379	0.0627	0.2717	0.0011	0.0999	0.0010	0.4687	0.0461	0.0026	0.0364	0.8130	0.0550
30min	T	-1.4395	0.2940	-1.1197	-0.4712	0.4454	-0.8117	-1.3874	-0.2199	-1.5002	0.7737	-4.2741	0.2905	-2.9027
	p	0.1504	0.7692	0.2633	0.6377	0.6562	0.4172	0.1658	0.8260	0.1338	0.4393	0.00002	0.7717	0.0037
60min	T	-1.6657	0.4304	1.1048	-0.1221	-0.5671	-0.4993	-0.1830	1.6143	-1.3532	-2.0816	-1.6573	-0.2136	-1.7760
	p	0.0966	0.6687	0.2703	0.9029	0.5709	0.6178	0.8549	0.1071	0.1765	0.0378	0.0983	0.8315	0.0758

the PM and ARIMAX increased monotonically. The MAPE, RMSE and MAE values of MLP are higher than those of LSTM and ALSTM. LSTM and ALSTM have the obvious advantage over the other methods in a time horizon for more than 15 min. The ALSTM model significantly reduces the MAPE, RMSE and MAE values of LSTM. In summary,

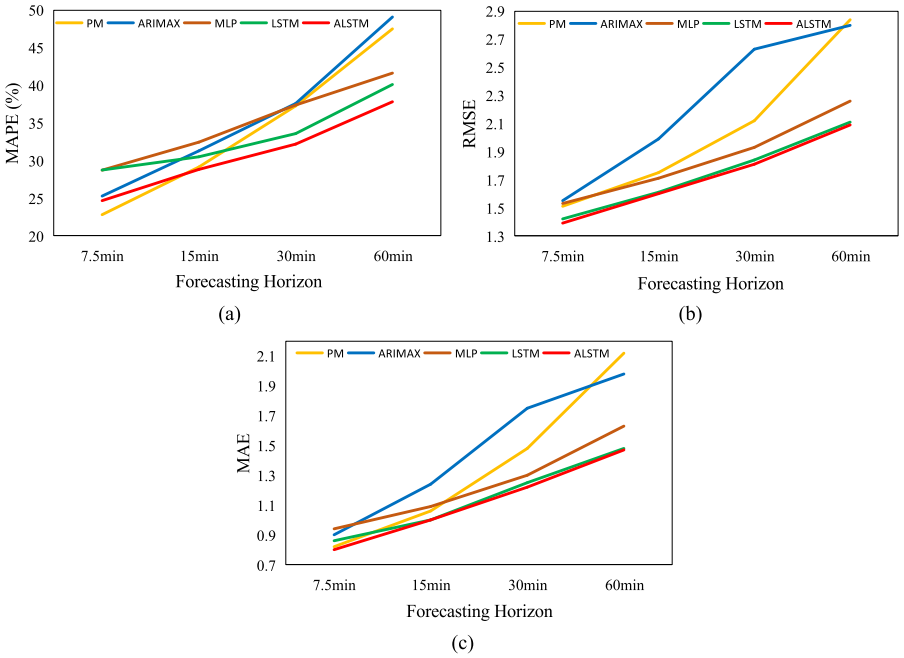

FIGURE 8. Comparison of different forecasting horizons: (a) MAPE (B) RMSE (c) MAE.

compared with four benchmark models, ALSTM has obvious advantages using all three error measurement metrics, including MAPE, RMSE and MAE.

The independent two-sample t-test was used to infer whether there is a difference between the population mean of forecast results of the LSTM and ALSTM. The t-statistic is calculated following Equations 15 and 16:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{15}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}} \tag{16}$$

where $\overline{X}_1$ and $\overline{X}_2$ are the sample mean of LSTM and ALSTM forecast results, respectively; S_{X1} and S_{X2} are the standard errors of LSTM and ALSTM forecast results, respectively; and n_1 and n_2 are the sizes of LSTM and ALSTM forecast results, respectively.

Null hypothesis H_0 and alternative hypothesis H_1 are defined as follows:

$$H_0 : \mu_1 = \mu_2 \tag{17}$$

$$H_1 : \mu_1 \neq \mu_2 \tag{18}$$

where μ_1 and μ_2 are the population mean of LSTM and ALSTM forecast results, respectively. The p value ($P \in [0,1]$) determines whether or not to reject null hypothesis. The smaller the p value, the more confident to reject null hypothesis.

T-statistics and p values of LSTM and ALSTM in different forecast horizons are shown in Table VI. From the monthly forecast results, there are differences between LSTM and ALSTM, and there are significant differences in some months. From the results of annual forecast, the p value of 15–60 min is small, and it can be concluded that there is a significant difference in the population mean of LSTM and ALSTM in annual forecast. In general, the forecast error of ALSTM is lower than that of LSTM, and there are differences between ALSTM and LSTM. From Table VI, it is shown that the attention mechanism significantly improves the LSTM forecasting results.

IV. CONCLUSION AND FUTURE WORK

PV power forecasting is of great significance for maintaining grid security and coordinating resource utilization. This paper proposed a new hybrid model based on LSTM and attention mechanism for short-term photovoltaic power forecasting. Different from traditional forecasting models, the proposed ALSTM model employs LSTM to extract features from the time series photovoltaic power data and learn long-term dependency information in sequence. We applied the trained attention mechanism to the LSTM neural networks to focus on important extracted features, which potentially enhance the original forecasting power of LSTM neural networks.

In the experiment section, a real-world dataset, which was collected from a 20kW PV power station located in eastern China, in 2017 and 2018, is employed. A comprehensive comparative study is conduced to compare the proposed ALSTM method with available state-of-art methods, including PM, ARIMAX, MLP and LSTM models in 7.5 min-ahead, 15 min-ahead, 30 min-ahead, and 60 min-ahead horizons in all four seasons with three different error metrics, namely, MAE, RMSE and MAPE, comprehensively.

The results show that LSTM and ALSTM have the obvious advantage over the other methods in a time horizon for more than 15 min. We further show the improvement of ALSTM over traditional LSTM using t-statistics and p values.

Future work of this study includes extending the current framework to longer term forecasting problems using multi-step forecasting strategies [15], [20].

REFERENCES

[1] E. Scolari, L. Reyes-Chamorro, F. Sossan, and M. Paolone, "A comprehensive assessment of the short-term uncertainty of grid-connected PV systems," *IEEE Trans. Sustain. Energy*, vol. 9, no. 3, pp. 1458–1467, Jul. 2018.

[2] W. Wang, H. Chen, B. Lou, N. Jin, X. Lou, and K. Yan, "Data-driven intelligent maintenance planning of smart meter reparations for large-scale smart electric power grid," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv.*, Oct. 2018, pp. 1929–1935.

[3] A. Saez-de-Ibarra, A. Milo, H. Gaztañaga, V. Debusschere, and S. Bacha, "Co-optimization of storage system sizing and control strategy for intelligent photovoltaic power plants market integration," *IEEE Trans. Sustain. Energy*, vol. 7, no. 4, pp. 1749–1761, Oct. 2016.

[4] Y. M. Saint-Drenan, G. H. Good, and M. Braun, "A probabilistic approach to the estimation of regional photovoltaic power production," *Solar Energy*, vol. 147, pp. 257–276, May 2017.

[5] M. Pierro, F. Bucci, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, and C. Cornaroa, "Multi-model ensemble for day ahead prediction of photovoltaic power generation," *Solar Energy*, vol. 134, pp. 132–146, Sep. 2016.

[6] X. Chen, Y. Du, H. Wen, L. Jiang, and W. Xiao, "Forecasting-based power ramp-rate control strategies for utility-scale PV systems," *IEEE Trans. Ind. Electron.*. vol. 66, no. 3, pp. 1862–1871, Mar. 2018. doi: 10.1109/TIE.2018.2840490.

[7] G. Wang, Y. Su, and L. Shu, "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models," *Renew. Energy*, vol. 96, pp. 469–478, Oct. 2016.

[8] M. Paulescu, M. Brabec, R. Boata, and V. Badescu, "Structured, physically inspired (gray box) models versus black box modeling for forecasting the output power of photovoltaic plants," *Energy*, vol. 121, pp. 792–802, Feb. 2017.

[9] L. Massidda and M. Marrocu, "Use of multilinear adaptive regression splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany," *Solar Energy*, vol. 146, pp. 141–149, Apr. 2017.

[10] M. Hu, W. Li, K. Yan, Z. Ji, and H. Hu, "Modern machine learning techniques for univariate tunnel settlement forecasting: A comparative study," *Math. Problems Eng.*, vol. 2019, Apr. 2019, Art. no. 7057612.

[11] K. Yan, Y. Du, and Z. Ren, "MPPT perturbation optimization of photovoltaic power systems based on solar irradiance data classification," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 514–521, Apr. 2018. doi: 10.1109/TSTE.2018.2834415.

[12] Y. Du, K. Yan, Z. Ren, and W. Xiao, "Designing Localized MPPT for PV Systems Using Fuzzy-Weighted Extreme Learning Machine," *Energies*, vol. 11(10), p. 2615, 2018.

[13] A. Mellit, A. M. Pavan, and V. Lughi, "Short-term forecasting of power production in a large-scale photovoltaic plant," *Solar Energy*, vol. 105, pp. 401–413, Jul. 2014.

[14] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Eng. Sci. Technol., Int. J.*, vol. 21, no. 3, pp. 428–438, Jun. 2018.

[15] C. Huang, L. Wang, and L. L. Lai, "Data-driven Short-term Solar Irradiance Forecasting Based on Information of Neighboring Sites," *IEEE Trans. Ind. Electron.*, to be published. doi: 10.1109/TIE.2018.2856199.

[16] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and meteorological information," *Renew. Energy*, vol. 118, pp. 357–367, Apr. 2018.

[17] X. Yao, Z. Wang, and H. Zhang, "A novel photovoltaic power forecasting model based on echo state network," *Neurocomputing*, vol. 325, pp. 182–189, Jan. 2019.

[18] S. Srivastava and S. Lessmann, "A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data," *Solar Energy*, vol. 162, pp. 232–247, Mar. 2018.

[19] H. Wang, H. Yi, J. Peng, G. Wang, Y. Liu, H. Jiang, and W. Liu, "Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network," *Energy Convers. Manage.*, vol. 153, pp. 409–422, Dec. 2017.

[20] K. Yan, X. Wang, Y. Du, N. Jin, H. Huang, and H. Zhou, "Multi-step short-term power consumption forecasting with a hybrid deep learning strategy," *Energies*, vol. 11, no. 11, p. 3089, Nov. 2018.

[21] P. Tang, H. Wang, and S. Kwong, "Deep sequential fusion LSTM network for image description," *Neurocomputing*, vol. 312, pp. 154–164, 2018.

[22] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[24] H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," *Neurocomputing*, vol. 284, pp. 171–176, Apr. 2018.

[25] W. Li, D. Guo, and X. Fang, "Multimodal architecture for video captioning with memory networks and an attention mechanism," *Pattern Recognit. Lett.*, vol. 105, pp. 23–29, Apr. 2018.

[26] J. Campbell, H. B. Amor, H. M. Ang, and G. Fainekos, "Traffic light status detection using movement patterns of vehicles," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 283–288.

[27] Y. Jiang, H. Long, Z. Zhang, and Z. Song, "Day-Ahead Prediction of Bihourly Solar Radiance With a Markov Switch Approach," *IEEE Trans. Sustain. Energy*, vol. 8, no. 4, pp. 1536–1547, Jul. 2017.

[28] D. Yang and Z. Dong, "Operational photovoltaics power forecasting using seasonal time series ensemble," *Solar Energy*, vol. 166, pp. 529–541, May 2018.

HANGXIA ZHOU received the master's degree from Zhejiang University, Hangzhou, China. She is currently a Professor with the College of Information Engineering, China Jiliang University, Hangzhou. Her research interests include big data analysis, machine learning, pattern recognition, and sustainable energy technologies. She is a Committee Member of the China Computer Federation (CCF).

YUJIN ZHANG received the master's degree from China Jiliang University, in 2019. His current research interests include deep learning, data mining, and smart grid.

LINGFAN YANG is currently pursuing the master's degree in computer applications with China Jiliang University. His research interests include machine learning and Apache spark.

QIAN LIU received the master's degree in computer application technology from China Jiliang University, in 2014. Her research interests include machine learning, data mining, and pattern recognition.

KE YAN received the bachelor's and Ph.D. degrees in computer science from the National University of Singapore, Singapore, in 2012, under the supervision of Dr. H.-L. Cheng. From 2013 to 2014, he was a Postdoctoral Researcher with the Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates. He is currently an Assistant Professor with the National University of Singapore, Singapore, and an Associate Professor with China Jiliang University, Hangzhou, China. His research interests include energy solutions for buildings, data mining, and machine learning.

YANG DU (S'09–M'13) received the Ph.D. degree in electrical engineering from The University of Sydney, Australia, in 2013. From 2013 to 2014, he was with the Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates, as a Postdoctoral Research Fellow. From 2014 to 2018, he was a Lecturer with Xi'an Jiaotong-Liverpool University, Suzhou, China. He is currently an Assistant Professor with James Cook University, Cairns, Australia. His research interests include photovoltaic power systems, power electronics, and smart grid.