Day-Ahead Nonparametric Probabilistic Forecasting of Photovoltaic Power Generation Based on the LSTM-QRA Ensemble Model

FEI MEI1, (Member, IEEE), JIAQI GU1, JIXIANG LU2, JINJUN LU2, JIATANG ZHANG1, YUHAN JIANG1, TIAN SHI3, AND JIANYONG ZHENG3

1College of Energy and Electrical Engineering, Hohai University, Nanjing 211100, China
2State Key Laboratory of Smart Grid Protection and Control, NARI Group Corporation, Nanjing 211000, China
3School of Electrical Engineering, Southeast University, Nanjing 210096, China

Corresponding author: Fei Mei (meifei@hhu.edu.cn)

This work was supported in part by the State Key Laboratory of Smart Grid Protection and Control, SKL of SGPC, and in part by the National Key Research and Development Program of China under Grant 2018YFB0905000 and Grant SGTJDK00DWJS1800232.

ABSTRACT With the rapid growth of photovoltaic (PV) power in recent years, the stability of system operation, the performance of system contingency analysis as well as the power quality of the power grid are threatened by the inherent uncertainty and fluctuation of PV output. It is necessary to have the knowledge of PV output characteristics for reliable power system dispatching. Day-ahead PV power forecasting is an effective support for achieving optimal dispatching. Probabilistic forecasting can describe the uncertainty that is difficult to depict by deterministic forecasting, and the forecasting results are more comprehensive. An ensemble nonparametric probabilistic forecasting model of PV output is proposed based on the traditional deterministic forecasting method. Quantile regression averaging (QRA) is used to ensemble a group of independent long short-term memory (LSTM) deterministic forecasting models for obtaining the probabilistic forecasting of PV output. Real measured data are used to verify the effectiveness of this nonparametric probabilistic forecasting model. Additionally, in comparison with the benchmark methods, LSTM-QRA has higher prediction performance due to the better forecasting accuracy of independent deterministic forecasts.

INDEX TERMS Probabilistic forecasting, photovoltaic output, quantile regression averaging (QRA), long short-term memory (LSTM), interval prediction, nonparametric forecasting.

I. INTRODUCTION

In recent years, new energy power generation technologies such as photovoltaic (PV) and wind power have developed rapidly. The PV installed capacity has been increasing continually worldwide. However, with the gradual increase in PV penetration in the power grid, the inherent uncertainty and fluctuation of PV power will lead to serious risks for the safety of the system operation and power quality of the power grid [1]. To securely consume as much grid-connected PV power as possible, operators need to fully understand the PV output characteristics. Short-term day-ahead PV output forecasting can provide the prediction results for the next day, which can be an important basis for optimized decision-making and day-ahead optimal dispatch of power systems. A forecasting with high accuracy can not only enhance the performance of power system contingency analysis, but also promote the system operating efficiency as well as reduce operational cost.

Deterministic forecasting methods are widely adopted in traditional PV output forecasting. These methods can be divided into four types: statistical models, machine learning models [2], numerical weather prediction (NWP)-based models, and sky or satellite cloud image-based models [3]. Statistical models mainly include multiple regression models and time series models. A weather classification multivariate adaptive regression spline (MARS) PV forecasting model is introduced for complex weather conditions in all seasons [4]. In [5] a MARS model based on past power measurement and NWP is proposed. Li et al. [6] have adopted an autoregressive integrated moving average model with exogenous variable (ARMAX) forecasting model, which allows
for meteorological inputs for time series model. In [7], a time series ensemble model is used for day-ahead PV power forecasting. Machine learning models mainly include neural network, support vector machine, random forest and long short-term memory (LSTM). In [8], a methodology based on artificial neural networks and an analog ensemble to generate 72 h-ahead forecasts. A multilayer perceptron-based ANN model is proposed for day-ahead forecasting of the PV power in [9]. Das et al. [10] have proposed a PV power forecasts based on support vector regression, historical PV power output, and corresponding meteorological data. In [11], a forecasting model based on ν-SVR is used to predict the PV energy for the next day with an interval of 15 min. In [12], an optimized hybrid method called PCA-K-means-HGWO-RF is proposed to forecast PV power. Random forests technique also be applied to forecast the output current of a photovoltaic grid-connected system [13]. A set of LSTM and deep neural network forecasting model combined with stationary wavelet transform is proposed in [14]. Zhou et al. [15] applies attention-based LSTM and achieves good performance in all four seasons. The NWP-based models use horizontal irradiance and other meteorological information from NWP, combined with the corresponding relationship between PV output and irradiance, to obtain PV output forecasting results [16]. Sky or satellite cloud image-based models provide ultra-short-term forecasting results through the distribution of cloud images [17].

Recently, some power system management algorithms have been applied that take parameter uncertainty into account, such as probabilistic load flow [18] and robust energy management [19]. However, traditional deterministic forecasting of PV output cannot give reasonable prediction intervals or probability densities. Probabilistic forecasting is a kind of uncertainty forecasting. Compared to traditional deterministic forecasting, probabilistic forecasting can provide both deterministic predicted values and probability distributions, which depicts the uncertainty of PV output. However, probabilistic forecasting will cost much more computing resource. With the rapid development of algorithms and hardware, more and more researchers focus on probabilistic forecasting. In reference [20], quantile regression (QR) and NWP are used to obtain a set of probabilistic forecasting results for PV output, and ensemble probabilistic forecasting is achieved based on the optimal weights calculated by minimizing the continuous ranked probability score (CRPS). In order to obtain nonparametric probability density distribution (PDF) rather than quantile estimation, reference [21] proposes a forecasting method for PV output based on improved Markov chains is proposed, and good results are achieved. Due to the strong learning ability of machine learning, a probabilistic forecasting method for PV output based on a Bayesian neural network (BNN) is proposed that represents neural network weights as probability distributions [22]. To combine the advantages of various algorithms, Wang et al. [23] use quantile regression neural network (QRNN), quantile regression random forests (QRRF) and quantile regression gradient boosting (QRGB) to obtain individual probabilistic load forecasts, then constrained quantile regression averaging (CQRA) and several averaging strategies are used to ensemble individual probabilistic load forecasts and reduce the pinball loss. However, the above methods mainly focus on probabilistic forecasting models that process deterministic forecasting and probabilistic forecasting separately and overlook the influence of deterministic forecasting on the probabilistic forecasting. To combine them and seek the connection between deterministic and probabilistic forecasting, an ensemble probabilistic forecasting model for PV output is proposed in this article. First, historical PV and NWP data are used to train the LSTM networks. Second, a set of prediction results are obtained from these LSTM networks. Finally, quantile regression averaging (QRA) is used to combine independent deterministic prediction results to achieve probabilistic forecasting. The performance of the proposed method is verified using real measured data.

The main contributions of this article are as follows:

- A new nonparametric probabilistic forecasting model of PV output is proposed. A set of independent LSTM deterministic forecasting models are established using historical PV output data and NWP data. Then, nonparametric probabilistic forecasting is realized by QRA. The feasibility of generating a nonparametric probabilistic forecasting model by integrating independent deterministic models with QRA is verified.

- The relationship between the pinball loss function of the QRA ensemble probabilistic forecasting model and the mean absolute error (MAE) of each independent deterministic model is determined. A set of deterministic models with smaller MAE also has smaller pinball loss functions.

The paper is structured as follows: Section 2 introduces the related theories and methods involved in the forecasting models. Section 3 provides the specific process of the nonparametric probabilistic forecasting model. Section 4 verifies the performance of the proposed model through modelling using real data and comparing the results with other forecasting models. Final conclusions are drawn in Section 5.

II. METHODOLOGY

A. LONG SHORT-TERM MEMORY

LSTM networks can store important events with relatively long intervals and delays in time series. Compared with common neural networks, LSTM is more adaptable in short-term and ultra-short-term forecasting and analysis [15]. LSTM is a variant of traditional recurrent neural networks (RNN). The structure of an RNN is shown in Figure 1.

\[
x_t \quad \text{is the input variable, and} \quad h_t \quad \text{is the output variable.}
\]

RNN can be regarded as multiple copies of the same neural network. Each neural network module passes the message to the next. Because of the correlation features of RNN, this model has advantages in processing time series data and text translation [24]. However, traditional RNN can only memorize short-term information and associate it with new...
input information, and not having the ability to remember long-distance information.

The performance of LSTM in controlling the storage state has been greatly improved compared with that of traditional RNN. The LSTM network consists of one or more LSTM units. The structure of each LSTM unit is shown in Figure 2.

There are three gates in the LSTM unit: the forget gate, input gate and output gate. The LSTM unit uses input \( x_t \) of the current series, output \( h_{t-1} \) of the last time, and memory state \( C_{t-1} \) as the input of the current time. \( h_t \) and \( C_t \) are preserved for the next time. The formula of the forget gate \( f_t \) can be expressed as:

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

where \( \sigma \) is a sigmoid function. \([\cdot, \cdot]\) is the combination of elements in a vector. \( W_f \) and \( b_f \) are the weight matrix and the bias of the forget gate, respectively. The formula of the input gate \( i_t \) can be expressed as:

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

\[
C_t = \tanh (W_C [h_{t-1}, x_t] + b_C)
\]

where \( C_t \) is the candidate value for updating the memory state at time \( t \). \( W_i \) and \( b_i \) are the weight matrix and the bias of the input gate, respectively, and \( W_C, b_C \) are the weight matrix and the bias of \( C_t \), respectively. \( \otimes \) is the elementwise multiplication of vectors. \( C_t \) is the combination of \( C_t \) and \( f_t \). The formula of the output gate \( o_t \) can be expressed as:

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

\[
h_t = o_t \otimes \tanh (C_t)
\]

where \( W_o \) and \( b_o \) are the weight matrix and the bias of the output gate, respectively. The LSTM network is trained by back propagation (BP), the same as the general neural network. Each network parameter of the LSTM network can be a number or a vector. Except for the input and output parameters, the dimensions of the other parameters depend on the number of hidden layer units in the LSTM network. Because the number of hidden layer units and output dimensions are usually different, a fully connected layer is usually added at the end of the output gate.

### B. QUANTILE REGRESSION AVERAGING

In most studies, nonparametric probabilistic forecasting models provide the probability density or prediction interval of PV output through QR or machine learning methods based on the pinball loss function. Nevertheless, deterministic and probabilistic forecasting models are rarely connected with each other. In references [25], [26], a modified QR algorithm is applied to ensemble multiple independent deterministic forecasting models for a probabilistic forecasting model. This method is named QRA. The formula of QRA can be expressed as:

\[
\hat{Q}(q, \hat{p}_t) = \hat{p}_t w_{q, t}
\]

where \( \hat{Q}(q, \hat{p}_t) \) is the quantile value under the \( q \)-th condition of the prediction results of the PV output probability distribution \((0 < q < 1)\). \( \hat{p}_t = [p_{1,t}, p_{2,t}, \ldots, p_{k,t}] \) are the vectors of the prediction values of the independent deterministic forecasting models (regression factors). \( k \) is the number of independent deterministic forecasting models. \( t \) is the sampling time. \( w_{q, t} \) is the weight vector. The objective function is:

\[
\min_{w_{q,t}} \left[ \sum_{p_t-\hat{p}_t w_{q,t} \geq 0} q |p_t - \hat{p}_t w_{q,t}| + \sum_{p_t-\hat{p}_t w_{q,t} < 0} (1-q) |p_t - \hat{p}_t w_{q,t}| \right]
\]

where \( p_t \) is the actual PV output at time \( t \), \( w_{q,t} \) is the optimization variable \((0 \leq w_{q,t} \leq 1)\). The weight vector \( w_{q,t} \) is obtained by minimizing the objective function (8) of the \( q \)-th quantile, and the prediction of the probability distribution can be obtained by formula (7). Because QRA is essentially a nonconvex optimization problem, it is difficult to give an accurate theoretical optimal solution. Usually, solvers or meta-heuristic algorithms are helpful for obtaining solutions.

### III. NONPARAMETRIC PROBABILISTIC FORECASTING MODEL BASED ON THE LSTM-QRA ENSEMBLE MODEL

#### A. NONPARAMETRIC PROBABILISTIC FORECASTING

Assume \( p_t \) is the measured value of the PV output at time \( t \), and \( F_t \) is the probability distribution function of \( p_t \). The quantile of the PV output measured value under the probability of \( q \) can be expressed as \( Q_{t,q} \) \((0 < q < 1)\). From the definition
of the quantile:

\[ F_t(p_t \leq Q_{t,q}) = q \]  

(9)

Assume that \( t_1 \) to \( t_n \) are historical moments, and \( t_{n+1} \) is the future moment. The probability distribution \( F_{t_{n+1}} \) at the future moment is obviously unknown. If the existing probability distribution such as the normal distribution or Beta distribution is used directly, the probability distribution at future moment \( F_{t_{n+1}} \) can be obtained by fitting the probability distribution at historical moment \( F_{t_1 \ldots t_n} \). These models are called parametric probabilistic forecasting models [27].

In the nonparametric probabilistic forecasting model, the probability distribution \( F_{t_{n+1}} \) at the future moment is not considered to be accurately expressed by the existing probability distribution function. However, quantiles with different probability \( q \) can be predicted. In other words, \( \hat{Q}_{t_{n+1},q} \) represents the predicted quantiles at probability \( q \) at the \( t_{n+1} \) moment, which can be regarded as an estimate of the actual value \( Q_{t_{n+1},q} \). The single predicted quantile can only provide limited uncertainty information [28]. However, when the resolution of the quantile is improved, the probability distribution in the future of PV output can be represented in detail. Compared with the parametric probabilistic forecasting model, non-parametric model is more flexible. Non-parametric model does not need to make assumptions about the model and just relies on the historical data, however the computation of which is usually larger.

Moreover, the quantiles under probability \( q \) and \( (1-q) \) can determine a prediction interval (PI) with \( (1-2q) \times 100\% \) nominal coverage (NC).

\[ I_{t_{n+1}}^{1-2q} = \left[ \hat{Q}_{t_{n+1},q}, \hat{Q}_{t_{n+1},1-q} \right] \]  

(10)

where \( I_{t_{n+1}}^{1-2q} \) is the PI at \( t_{n+1} \) moment with \( (1-2q) \times 100\% \) NC and \( 0 < q < (1-q) < 1 \).

**B. LSTM-QRA**

Unlike load forecasting, historical output data are often not the main factors affecting PV output forecasting due to the strong uncertainty. In reference [29], the influencing factors on PV output are analysed, and solar irradiance is regarded as the main factor. Moreover, under the same irradiance, the PV output decline with increasing temperature. In this article, historical PV output data from the day before the predicted day, global horizontal irradiances (GHI) and predicted temperatures on the predicted day are the inputs of the LSTMs. The output of the LSTMs is the predicted PV output on the predicted day. The predicted GHI and temperatures can be obtained from the NWP, such as the global forecast system (GFS) of the National Oceanic and Atmospheric Administration (NOAA), which can provide the predicted GHI and temperatures at one-hour intervals [16], [30]. The predicted data from NWP are updated at the zero hour of each day. The forecasting model in this article can be trained and predicted at midnight, when the PV output is usually zero. Therefore, the training process has no effect on the forecasting.

The number of hidden layer units in the LSTM network needs to be pre-set, as well as in the traditional neural networks. In Figure 2, the vector length of output \( h_t \) of the LSTM and memory state \( C_t \) at time \( t \) are determined by the number of hidden layer units. Fewer hidden units restricts the information storage ability of the LSTM. As a result, it may cause under fitting problem when training the forecasting model. On the contrary, more hidden layer units can increase the vector length of \( h_t \) and \( C_t \) to store more information. However, the overfitting risk of the model increases, which is not conducive to the generalization of training results. Therefore, artificial, grid or meta-heuristic algorithms are usually chosen to optimize the number of hidden layer units [31]. However, optimization can only guarantee better results for the development set, which cannot be strictly extended to the test set or real forecasting situations. Therefore, ensemble forecasting models are considered [32]. Through the weighting of multiple forecasting models with different hyperparameters, it is possible to avoid a one-sided understanding of the rule by a single model. Many studies have shown that the overall forecasting performance of ensemble forecasting models is better than that of single independent forecasting models [33]. In this article, a group of LSTM networks with different numbers of hidden layer units are selected as a set of independent deterministic forecasting models.

For a series of PV output data \( P \) and corresponding GHI data \( I \) and temperature data \( T \), the prediction process of the LSTM-QRA ensemble forecasting model is as follows:

1. **Step 1:** Min-max normalization is used to standardize \( P \), \( I \) and \( T \). The maximum and minimum values are saved for denormalization to restore the actual value of the predicted data.
2. **Step 2:** Datasets \( P \), \( I \) and \( T \) are combined together. For example, Figure 3 shows the structure of the dataset on the j-th day. \( p_1 \) to \( p_n \) are the PV output power at different sampling times. \( n \) is the sampling number of the PV output in a day. If the sampling interval is one hour, then \( n = 24 \). \( i_1 \) to \( i_{24} \) and \( T_1 \) to \( T_{24} \) are the GHI and temperature predictions per hour from NWP. Afterwards, these datasets are divided into training and development sets.
3. **Step 3:** A set of LSTM networks with different hidden layer units is constructed. The training set from step 2 is used for training, and the cross-validation method is used to adjust some hyperparameters. Afterwards, the development set is used for prediction. \( \hat{p}_j^{de} \) is defined

![FIGURE 3. Dataset on j-th day.](image-url)
as the prediction results of every independent LSTM model on the j-th day. $\hat{p}^{de}_{j,i} = \left[\hat{p}^{de}_{j,1,1}, \hat{p}^{de}_{j,1,2}, \ldots, \hat{p}^{de}_{j,1,k}\right]$ is the prediction result of every independent LSTM model at the first sampling time on the j-th day. $k$ is the number of LSTM models.

- Step 4: Based on the principle of QRA in Section 2.2 and the prediction results of the development set in Step 3, the objective function of the q-th quantile can be rewritten as:

$$
\min_{w_{q,t}} \sum_j \left[ \sum_{p_{j,i}^{de}} q \left| p_{j,i}^{de} - \hat{p}_{j,i}^{de} w_{q,t} \right| ight. \\
+ \left. \sum_{p_{j,i}^{de}} (1-q) \left| p_{j,i}^{de} - \hat{p}_{j,i}^{de} w_{q,t} \right| \right] (11)
$$

where $p_{j,i}^{de}$ is the real measured value of the development set on the t-th sampling time on the j-th day. The optimization of $w_{q,t}$ can be obtained by minimizing formula (11).

- Step 5: Day-ahead deterministic forecasting of the PV output is executed using a set of trained independent LSTM models in Step 3 combining NWP data and historical PV output data. The prediction value $\hat{p}_{i}$ and optimization of $w_{q,t}$ in Step 4 are input into formula (7), and quantile forecasting of the PV output can be obtained. Moreover, with the change in q, complete non-parametric probabilistic forecasting and PIs of different NCs can be obtained.

- Step 6: Based on the maximum and minimum saved in step 1, the quantile prediction results obtained in step 5 are denormalized, and the quantile prediction results can be gained to realize nonparametric probabilistic forecasting.

In step 2, the sampling number of the PV output in a day can be 96 or 288 etc. Although the NWP data is hourly, the LSTM and other machine learning based model can handle the problem that the NWP data interval does not correspond to the PV output interval easily. That is, when the sampling interval of the PV power varies from 5 min to 1 h, our deterministic forecasting model can give the intra-hour forecasting values without corresponding intra-hour NWP values.

The specific process of the nonparametric probabilistic forecasting model based on LSTM-QRA is shown in Figure 4.

IV. CASE STUDY

A. PV OUTPUT DATA
In this article, PV output power data are collected by a rooftop-mounted PV power generation system in eastern China. The installation capacity is 2.8 MW. The collected data include the historical PV output power, GHI and temperature. The data collection time ranges from September 14, 2017 to September 13, 2018. The sampling interval of the PV output power is 15 min, and the GHI and temperature are 1 h. These collected data are screened for abnormal data, and extreme outliers are eliminated. The following examples are executed using MATLAB 2018a and TensorFlow.

B. FORECASTING EVALUATION INDEXES
For deterministic forecasting, the mean absolute percentage error (MAPE) is often used to measure the forecasting performance. However, in PV power forecasting, there are many zero or minimum values in the PV data series. These values make the MAPE extremely large or infinite. Therefore, the normalized mean absolute error (MAE) and normalized root mean squared error (RMSE) are adopted as the forecasting evaluation indexes. The formulas of normalized MAE and RMSE are as follows:

$$\text{MAE(pu)} = \frac{1}{n} \sum_{i=1}^{n} |p_i - \hat{p}_i| \times 100\% \quad (12)$$

$$\text{RMSE(pu)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - \hat{p}_i)^2} \times 100\% \quad (13)$$

where $n$ is the number of samples and $p_i$, $\hat{p}_i$ are the real and predicted values of the PV output at the i-th time.

Because probabilistic forecasting can provide prediction curves for each quantile, evaluation indexes such as MAE and
RMSE used in traditional deterministic forecasting obviously cannot be used to measure the performance of probabilistic forecasting. In GEFcom2014 and GEFcom2017 [34], [35], the pinball loss function is used to measure the error in probabilistic forecasting. For $q$-th quantile forecasting, the pinball loss function is defined as:

$$L(q, p_t, \hat{p}_{t,q}) = \begin{cases} (1 - q) |p_t - \hat{p}_{t,q}|, & p_t < \hat{p}_{t,q} \\ q |p_t - \hat{p}_{t,q}|, & p_t \geq \hat{p}_{t,q} \end{cases}$$ (14)

where $p_t$ is the real value of the PV output at time $t$, $\hat{p}_{t,q}$ is the prediction value of the $q$-th quantile at time $t$. The smaller $L(\cdot)$ is, the more accurate the probabilistic forecasting.

In references [26] and [36], the Winkler score is proposed as an evaluation method for probabilistic forecasting. Different from the pinball loss function, the Winkler score focuses more on the evaluation of probability intervals. Moreover, both the width of the interval and the coverage of the actual value are evaluated simultaneously. For a $(1 - \alpha) \times 100\%$ prediction interval, the Winkler score is defined as:

$$\text{Winkler} = \begin{cases} \delta_l, & L_l \leq p_t \leq U_l \\ \delta_l + 2 (L_l - p_t) / \alpha, & p_t < L_l \\ \delta_l + 2 (p_t - U_l) / \alpha, & p_t > U_l \end{cases}$$ (15)

where $\delta_l = U_l - L_l$, $U_l$ and $L_l$ are the upper and lower bounds of the prediction interval, respectively, at time $t$. A smaller Winkler score indicates that the predicted interval has a higher actual value coverage and a smaller interval width.

### C. Prediction Results of the LSTM-QRA Model

In this article, the collected dataset includes a total of 364 days. The data from 1-300 days are selected as the training set, the data from 301-332 days are the development set, and the data from days 333-364 are the test set. We perform min-max normalization on the training and development set and record the scaling factors. Moreover, the test set is normalized by the maximum and minimum values (scaling factors) of training and development set to avoid data leakage. Ten different numbers of hidden units are used to construct a set of independent LSTM forecasting models. The numbers of hidden units are 0.25a, 0.5a, a, 2a, 3a, 4a, 5a, 6a, 7a and 8a, and a is the dimensions of LSTM input features. Because the data sampling interval is 15 min, $n = 96$, and $a = 144$. The hyperparameters of ten independent LSTM networks are shown in Table 1.

The training set is used to train these ten independent LSTM networks. These trained LSTM networks are used to forecast the deterministic prediction value for the development and test sets. The MAE (pu) and RMSE (pu) are used to measure the performance of these deterministic forecasts. The forecasting performance for the test set is shown in Table 2.

In Table 2, the numbers of hidden units of LSTM1-LSTM10 are 36, 72, 144, 288, 432, 576, 720, 864, 1008, and 1152. It is clear that LSTM1 has the minimum MAE and RMSE in these independent forecasting models. Obviously, the overall forecasting performance of LSTM1 on the test set is better than those of the other models. For further study on the forecasting performance, the data from September 10, 2018 are selected as the research object, and the prediction results of each model are shown in Figure 5.

In Figure 5, the solid black line is the real measured value of the PV output power, and the remaining solid lines represent the predicted values of each independent LSTM model. To more intuitively reflect the prediction performance of each model, the MAE and RMSE are calculated, which are shown in Table 3.

### Table 1. Hyperparameters of Independent LSTM Networks.

| Hyperparameter          | Value               |
|-------------------------|---------------------|
| Input unit              | 144                 |
| Hidden unit             | 36, 72, 144, 288, 432, 576, 720, 864, 1008, 1152 |
| Output unit             | 96                  |
| Learning rate           | 0.01                |
| Optimization algorithm  | Adam                |
| Maximum epochs          | 500                 |

### Table 2. Prediction Performance of LSTM.

| Forecasting Model | MAE (pu) | RMSE (pu) |
|-------------------|----------|-----------|
| LSTM1             | 31.8874  | 58.9834   |
| LSTM2             | 33.1326  | 63.2188   |
| LSTM3             | 36.8061  | 68.5956   |
| LSTM4             | 34.3431  | 64.2149   |
| LSTM5             | 37.6755  | 71.1089   |
| LSTM6             | 36.6594  | 67.8151   |
| LSTM7             | 36.9296  | 68.6016   |
| LSTM8             | 34.2934  | 64.3060   |
| LSTM9             | 34.2084  | 65.3501   |
| LSTM10            | 32.9212  | 61.6342   |

Table 3 shows that the LSTM4 model has better performance in the forecasting of PV output on September 10, 2018. According to MAE, the prediction accuracy of the LSTM1 model only ranks 7th out of ten independent models. Moreover, according to RMSE, the prediction accuracy of the LSTM1 model ranks 2nd. The results in Table 3 explain why ensemble forecasting models generally have better forecasting performance than independent models [33]. An independent model with a good overall forecasting performance may not have a high prediction accuracy on a certain day, but an ensemble model can use a proper ensemble algorithm to learn...
F. Mei et al.: Day-Ahead Nonparametric Probabilistic Forecasting of PV Power Generation

FIGURE 5. Prediction results of each LSTM model on September 10, 2018.

FIGURE 6. Day-ahead PV output nonparametric probabilistic forecasting based on LSTM-QRA August 25, sunny; (b) August 26, rainy; (c) August 30, sunny; (d) August 31, sunny; (e) September 7, cloudy and (f) September 12, cloudy.

To describe the predicted probability density, in this article, we select a probability $q$ from 0.05 to 0.95 every 0.05 and apply fmincon function of MATLAB with the default solver ‘interior-point’ to solve the QRA algorithm. Hence, there are 19 quantile predictions. Optimal weights $w_{q,t}$ are obtained through the QRA algorithm, the real measured PV output for the development set, and the prediction results of each independent LSTM model for the development set. Then, the nonparametric probabilistic forecasting on the test set is accomplished by integrating the prediction results of ten independent LSTM models using formula (7) and the optimal weights $w_{q,t}$.

The six-day probability prediction results of the LSTM-QRA ensemble model on the test set are shown in Figure 6, which include three different kinds of weather. The solid red line indicates the real measured PV output power, and the 19 grey dashed lines represent the prediction power for each quantile. Because interval prediction with high NC is critical for power system planning and economic dispatch [26], prediction intervals (PIs) with NC = 90% are given by two quantile prediction lines under $q = 0.05$ and $q = 0.95$ in Figure 7. It can be seen that the PIs with
NC = 90% cover most of the real measured PV output power. Moreover, PIs have higher reliability on sunny days, but the forecasting performance on rainy and cloudy days is relatively worse. To fully inspect the probability prediction performance of the LSTM-QRA ensemble model under various weather conditions, the sum of the pinball loss function at $q = 0.05 - 0.95, t = 0 - 24$ hour (sampling interval is 15 min) for each day in Figure 6 and the Winkler score of PI at NC = 90% in Figure 7 are calculated. The sum of the pinball loss function $L_{\text{Pinball}}$, the Winkler score, and the mean value of each independent deterministic forecasting model’s MAE of the above 6 days for the test set are shown in Table 4.

Table 4 shows that the results of $L_{\text{Pinball}}$ and the Winkler score are quite different. This is because $L_{\text{Pinball}}$ pays more attention to measuring the overall effect of probabilistic forecasting, and there is no strict requirement on the width of PI. However, the real measured values are mostly included in the PI when NC = 90%. The Winkler scores can be simplified to the superposition of the PI width, and the prediction with a smaller PI width has a greater advantage. Moreover, the trend of $L_{\text{Pinball}}$ is consistent with the MAE of the deterministic forecasting model. On sunny days, the deterministic forecasting model has a smaller MAE, and the probabilistic forecasting model also achieves a smaller $L_{\text{Pinball}}$. On cloudy and rainy days, the corresponding $L_{\text{Pinball}}$ and MAE increase to some extent. However, because the $L_{\text{Pinball}}$ and MAE in Table 4 are the results of different days, the relationship between the pinball loss function of the ensemble probabilistic forecasting model and the MAE of each independent deterministic forecasting model cannot be strictly determined.

D. COMPARISON WITH BENCHMARK MODELS

To further study the relationship between the QRA-based ensemble probabilistic forecasting model and each independent deterministic forecasting model, a set of independent LSTM networks are replaced with a set of artificial neural networks (ANNs) with different hidden units, a set of deep neural networks (DNNs) with different hidden units and a set of naive persistence models. The ensemble probabilistic forecasting models are generated by QRA in the same way. The proposed model also be compared with quantile regression neural network (QRNN), which is one of the states of art models of probability forecasting.

1) ANN-QRA

The deterministic forecast values of ANNs are ensembled by QRA to generate the probabilistic forecasting.
The hyperparameters of the ANNs are roughly the same as those of the LSTM network, as shown in Table 5:

**TABLE 5. Hyperparameters of independent ANNs and DNNs.**

| Hyperparameter     | Value          |
|--------------------|----------------|
| Input unit         | 144            |
| Hidden unit        | 36, 72, 144, 288, 432, 576, 720, 864, 1088, 1152 |
| Output unit        | 96             |
| Learning rate      | 0.01           |
| Optimization algo. | LMBP           |
| Maximum epochs     | 200            |

LMBP is used to solve the ANNs’ parameters [37]. Because LMBP has a faster convergence speed when training a simple ANN, lower maximum epochs are adopted in training. The division of the dataset is the same as in Section 4.3. The input elements of ANNs are the same as the input of the LSTM network.

2) DNN-QRA

DNNs are the neural networks with more than one hidden layer (usually greater than 3). In this article, we set the number of hidden layers to be 3, and the hidden units of each layer are the same. Other hyperparameters are the same as those of the ANNs in Table 5. The deterministic forecast values of DNNs are also ensembled by QRA to generate the probabilistic forecasting.

3) PERSISTENCE-QRA

We selected the PV output of the last 10 days before the predicted day as the forecast values of independent deterministic forecasts. To generate the probabilistic forecasting, the PV output of the last 10 days are ensembled by QRA. For instance, to obtain the PV power forecasts of August 30, the PV power from August 20 to August 29 are used as the forecast values of deterministic forecasts and ensembled by QRA.

4) QRNN

To make a more comprehensive comparison, we employ QRNN to compare with the QRA based models. The loss function of traditional neural network always be mean square error, which makes ANNs and DNNs suitable for fitting and forecasting. However, the pinball loss function is adopted as the loss function of QRNN. The replacement of loss function enables QRNN to make estimation of potentially nonlinear quantile models [23], [38].

In this article, we set the number of hidden layers of QRNN to 2, and the hidden units of each layer to be 288 through manual tuning. The Maximum epochs is set to 1000 to reduce the probability of underfitting. Early stopping and cross validation are also adopted to deal with overfitting. The input elements of QRNN are the same as the input of the LSTM network. The hyperparameters of QRNN are shown in Table 6:

**TABLE 6. Hyperparameters of QRNN.**

| Hyperparameter               | Value |
|------------------------------|-------|
| Input unit                   | 144   |
| Hidden unit                  | 288   |
| Output unit                  | 96    |
| Learning rate                | 0.01  |
| Optimization algo.           | Adam  |
| Maximum epochs               | 1000  |

5) COMPARISON RESULTS

A forecasting performance comparison of ANN-QRA, DNN-QRA, Persistence-QRA, QRNN and LSTM-QRA on the test set is provided, including a scatter plot of the daily $L_{\text{Pinball}}$ of each model on the test set in Figure 8. In Table 7, $L_{\text{Pinball}}-\text{m}$ is defined as the mean of the daily $L_{\text{Pinball}}$ on the test set, and the MAE of all the independent deterministic forecasting models on the entire test set are also provided. The $L_{\text{Pinball}}-\text{m}$ is used to measure the performance of the probabilistic forecasts. Meanwhile, The MAE is adopted to evaluate the performance of the independent deterministic forecasts (LSTM, ANN, DNN and persistence model) of LSTM-QRA, ANN-QRA, DNN-QRA and Persistence-QRA ensemble models.

**TABLE 7. Forecasting performance comparison of ANN-QRA, DNN-QRA, Persistence-QRA, QRNN and LSTM-QRA.**

| Model              | MAE (pu) | $L_{\text{Pinball}}$-m |
|--------------------|----------|------------------------|
| LSTM-QRA           | 34.857   | 49.6370                |
| DNN-QRA            | 41.6019  | 76.7559                |
| ANN-QRA            | 70.4060  | 89.4531                |
| Persistence-QRA    | 78.1367  | 122.0268               |
| QRNN               | -        | 59.9841                |

In Figure 8, all the daily $L_{\text{Pinball}}$ are normalized by the maximum value of the pinball loss. The abscissa of the picture is the daily $L_{\text{Pinball}}$ of benchmark model and the ordinate of the picture is the daily $L_{\text{Pinball}}$ of LSTM-QRA model. For example, for the green dot, the abscissa is the daily $L_{\text{Pinball}}$ of DNN-QRA model and the ordinate is the daily $L_{\text{Pinball}}$ of LSTM-QRA model. For each colour dot, there are 32 points in total. The point which is above $Y = X$ means that the daily $L_{\text{Pinball}}$ of the benchmark model is smaller than the daily $L_{\text{Pinball}}$ of LSTM-QRA model.
of LSTM-QRA on that day. This indicates that the benchmark model performs better than LSTM-QRA on that day. Overall, most points are below $Y = X$, which means LSTM-QRA has a better performance than the other benchmark models on most days.

Table 7 shows that compared with the ANN, DNN, and persistence model, LSTM forecasting model has the minimum MAE, indicating that the deterministic forecasting performance of LSTM is better than other models. Meanwhile, the LSTM-QRA has the minimum $L_{\text{Pinball}}^{m}$, indicating that the LSTM-QRA performs better than other benchmark models. Due to its characteristics, QRNN has no MAE.

In Table 7, the trend of the $L_{\text{Pinball}}^{m}$ is consistent with that of MAE of the deterministic forecasting model. Since $L_{\text{Pinball}}^{m}$ is defined as the mean of the daily $L_{\text{Pinball}}$ on the test set, $L_{\text{Pinball}}^{m}$ is proportional to the sum of the pinball loss function on the entire test set. Therefore, the trends in Table 7 is a validation of the extrapolation of the trend of the $L_{\text{Pinball}}$ and MAE of the deterministic forecasting models in Section 4.3. Therefore, when QRA is used to generate an ensemble probabilistic forecasting model, choosing a deterministic forecasting model with a smaller MAE can usually effectively improve the performance of the probabilistic forecasting model.

V. CONCLUSION

In this article, an LSTM-QRA-based nonparametric probabilistic forecasting model for PV output power is proposed. A set of independent LSTM deterministic forecasting models is trained using historical PV data and NWP data. Nonparametric probabilistic forecasting is generated by integrating the independent LSTM prediction models with QRA. Real measured data are used to verify the effectiveness and feasibility of the proposed model. The numerical results show that QRA can ensemble a set of deterministic forecasting models to generate corresponding nonparametric probabilistic forecasting, which means that to simply obtain nonparametric probabilistic forecasting, it is not necessary to build a brand-new probabilistic forecasting model. Moreover, this forecasting can be obtained by an ensemble of existing deterministic forecasting using QRA. To further analyse the relationship between the QRA ensemble probabilistic forecasting model and deterministic forecasting models, ANN-QRA, DNN-QRA and Persistence-QRA are used as benchmark models. A comparison of the pinball loss function of probabilistic forecasting with the MAE of deterministic forecasting shows that the pinball loss function has the same trend as MAE; that is, a group of deterministic forecasting models with smaller MAE can usually generate probabilistic forecasting with a smaller pinball loss function through QRA. It is believed that when the MAE of deterministic forecasting is further reduced, the performance of QRA based model will be much better than QRNN. The relationship between the pinball loss function and MAE provides a direction for further improving the forecasting performance of the QRA ensemble probabilistic forecasting model.

The probabilistic forecasting model proposed in this article is still simple and naive, and there are many areas that can be improved and optimized. For example, the selection of input features, the optimization of scaling factors of normalization to not only avoid data leakage, but also preserve the characteristics of raw data. The selection of deterministic models may not be limited to one model but can adopt a variety of deterministic forecasting models with different principles, as well as the optimization of the QRA integration algorithm, as shown in Section 4.3. Although the ensemble model has a smaller pinball loss function, this cannot guarantee a narrow PI (high percentage NC) width. Therefore, reduction of the pinball loss function and consideration of the width of PI (high percentage NC) are problems we need to solve in future work.

REFERENCES

[1] C. Voyant, G. Nutton, S. Kalegirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloux, “Machine learning methods for solar radiation forecasting: A review,” Renew. Energy, vol. 105, pp. 569–582, May 2017.

[2] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. M. Shah, “Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques,” IET Renew. Power Gener., vol. 13, no. 7, pp. 1009–1023, May 2019.

[3] Y. F. Gong, Z. X. Lu, Y. Qiao, and Q. Wang, “An overview of photovoltaic energy system output forecasting technology,” Autom. Electr. Power Syst., vol. 40, no. 4, pp. 140–151, 2016.

[4] X. Zhang, F. Fang, and J. Liu, “Weather-classification-MARS-based photovoltaic power forecasting for energy imbalance market,” IEEE Trans. Ind. Electron., vol. 66, no. 11, pp. 8692–8702, Nov. 2019.

[5] L. Massidda and M. Marroco, “Use of multilinear adaptive regression splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany,” Sol. Energy, vol. 146, pp. 141–149, Apr. 2017.

[6] Y. Li, Y. Su, and L. Shu, “An ARMAX model for forecasting the power output of a grid connected photovoltaic system,” Renew. Energy, vol. 66, pp. 78–89, Jun. 2014.

[7] D. Yang and Z. Dong, “Operational photovoltaics power forecasting using seasonal time series ensemble,” Sol. Energy, vol. 166, pp. 529–541, May 2018.

[8] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, “Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble,” Renew. Energy, vol. 108, pp. 274–286, Apr. 2017.

[9] R. Muhammad Ehsan, S. P. Simion, and P. R. Venkateswaran, “Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron,” Neural Comput. Appl., vol. 28, no. 12, pp. 3981–3992, Dec. 2017.

[10] U. Das, K. Tey, M. Seyedmahmoudian, M. I. Idris, S. Mekhilef, B. Horan, and A. Stojcevski, “SVR-based model to forecast PV power generation under different weather conditions,” Energies, vol. 10, no. 7, p. 876, Jun. 2017.

[11] R. De Leone, M. Pietrini, and A. Giovannelli, “Photovoltaic energy production forecast using support vector regression,” Neural Comput. Appl., vol. 26, no. 8, pp. 1955–1962, Nov. 2015.

[12] D. Liu and K. Sun, “Random forest solar power forecast based on classification optimization,” Energy, vol. 187, Nov. 2019, Art. no. 115940.

[13] J. Ospina, A. Alzate, and A. Mohandes, “Modeling of the output current of a photovoltaic grid-connected system using random forest technique,” Energy Explor. Exploitation, vol. 36, no. 1, pp. 132–148, Jan. 2018.

[14] I. A. Ibrahim, T. Khatib, A. Mohamed, and W. Elmenreich, “Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model,” IET Renew. Power Gener., vol. 13, no. 7, pp. 1087–1095, May 2019.

[15] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, “Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism,” IEEE Access, vol. 7, pp. 78063–78074, 2019.


[16] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, “Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators,” IEEE Trans. Smart Grid, vol. 6, no. 5, pp. 2253–2262, Sep. 2015.

[17] Y. Sun, V. Venugopal, and A. R. Brandt, “Short-term solar power forecast with deep learning: Exploring optimal input and output configuration,” Sol. Energy, vol. 158, pp. 730–741, Aug. 2019.

[18] P. Zhang and S. T. Lee, “Probabilistic load flow computation using the method of combined cumulants and gram-charlier expansion,” IEEE Trans. Power Syst., vol. 19, no. 1, pp. 676–682, Feb. 2004.

[19] Y. Zhang, N. Gatisis, and G. B. Giannakis, “Robust energy management for microgrids with high-penetration renewables,” IEEE Trans. Sustain. Energy, vol. 4, no. 4, pp. 944–953, Oct. 2013.

[20] J. Thorey, C. Chaussin, and V. Mallet, “Ensemble forecast of photovoltaic power with online CRPS learning,” Int. J. Forecasting, vol. 34, no. 4, pp. 762–773, Oct. 2018.

[21] M. J. Sanjari and H. B. Gooi, “Probabilistic forecast of PV power generation based on higher order Markov chain,” IEEE Trans. Power Syst., vol. 32, no. 4, pp. 2942–2952, Jul. 2017.

[22] A. Ciaramella, A. Staiano, G. Cervone, and S. Alessandrini, “A Bayesian-based neural network model for solar photovoltaic power forecasting,” in Proc. WIRN Vietri sul Mare, Italy, 2015, pp. 169–177.

[23] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, and C. Kang, “Combining probabilistic load forecasts,” IEEE Trans. Smart Grid, vol. 10, no. 4, pp. 3664–3674, Jul. 2019.

[24] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[25] J. Nowotarski and R. Weron, “Computing electricity spot price prediction intervals using quantile regression and forecast averaging,” Comput. Statist., vol. 30, no. 3, pp. 791–803, Sep. 2015.

[26] B. Liu, J. Nowotarski, T. Hong, and R. Weron, “Probabilistic load forecasting via quantile regression averaging on sister forecasts,” IEEE Trans. Smart Grid, vol. 8, no. 2, pp. 730–737, Mar. 2017.

[27] M. D. Tabone and D. S. Callaway, “Modeling variability and uncertainty of photovoltaic generation: A hidden state spatial statistical approach,” IEEE Trans. Power Syst., vol. 30, no. 6, pp. 2965–2973, Nov. 2015.

[28] F. Golestaneh, P. Pinson, and H. B. Gooi, “Very short-term nonparametric probabilistic forecasting of renewable energy Generation—With application to solar energy,” IEEE Trans. Power Syst., vol. 31, no. 5, pp. 3850–3863, Sep. 2016.

[29] J. J. Roberts, A. A. M. Zevallos, and A. M. Cassula, “Assessment of photovoltaic performance models for system simulation,” Renew. Sustain. Energy Rev., vol. 72, pp. 1104–1123, May 2017.

[30] P. Mathiesen and J. Kleissl, “Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states,” Sol. Energy, vol. 85, no. 5, pp. 967–977, May 2011.

[31] T. Shi, F. Mei, J. Lu, J. Lu, Y. Pan, C. Zhou, J. Wu, and J. Zheng, “Phase space reconstruction algorithm and deep learning-based very short-term bus load forecasting,” Energies, vol. 12, no. 22, p. 4349, Nov. 2019.

[32] A. J. Landgraf, “An ensemble approach to GEFCom2017 probabilistic load forecasting,” Int. J. Forecasting, vol. 35, no. 4, pp. 1432–1438, Oct. 2019.

[33] A. Khoshrou and E. J. Pauwels, “Short-term scenario-based probabilistic load forecasting: A data-driven approach,” Appl. Energy, vol. 238, pp. 1258–1268, Mar. 2019.

[34] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, “Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond,” Int. J. Forecasting, vol. 32, no. 3, pp. 896–913, Jul. 2016.

[35] T. Hong, J. Xie, and J. Black, “Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting,” Int. J. Forecasting, vol. 35, no. 4, pp. 1389–1399, Oct. 2019.

[36] R. L. Winkler, “A decision-theoretic approach to interval estimation,” J. Amer. Stat. Assoc., vol. 67, no. 337, pp. 187–191, Mar. 1972.

[37] F. Mei, Q. Wu, T. Shi, J. Lu, Y. Pan, and J. Zheng, “An ultrashort-term net load forecasting model based on phase space reconstruction and deep neural network,” Appl. Sci., vol. 9, no. 7, p. 1487, Apr. 2019.

[38] J. W. Taylor, “A quantile regression neural network approach to estimating the conditional density of multiperiod returns,” J. Forecasting, vol. 19, no. 4, pp. 299–311, Jul. 2000.
YUHAN JIANG received the B.S. degree in electrical engineering from the School of Electrical and Electronic Engineering, Hubei University of Technology (HPUT), Wuhan, China, in 2019. She is currently pursuing the M.S. degree with the College of Energy and Electrical Engineering, Hohai University. Her research interests include renewable energy power generation technology and integrated energy system optimal operation.

TIAN SHI received the B.S. degree from the School of Electric Power Engineering, Nanjing Institute of Technology (NIIT), Nanjing, China, in 2018. He is currently pursuing the M.S. degree with the School of Electrical Engineering, Southeast University (SEU). His research interest includes renewable energy power generation technology.

JIANYONG ZHENG was born in China, in 1966. He received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering, Southeast University (SEU), Nanjing, China, in 1988, 1991, and 1999, respectively. He is currently a Full Professor with the School of Electrical Engineering, Southeast University. His research interests include the application of power electronics in power systems, renewable energy technology, artificial intelligence, and big data.