Probabilistic Wind Speed Forecasting based on Minimal Gated Unit and Quantile Regression

Tianyang Chen\textsuperscript{1,*}, Zheng Qian\textsuperscript{1}, Bo Jing\textsuperscript{1}, Jiangwen Wan\textsuperscript{1} and Fanghong Zhang\textsuperscript{2}

\textsuperscript{1}School of Instrumentation and Optoelectronic Engineering, Beihang University, Haidian Dist., Beijing, China
\textsuperscript{2}CSIC HaiZhuang Wind power Co., Ltd, Yubei Dist., Chongqing, China

*Email: chenty@buaa.edu.cn

Abstract. High-quality wind speed forecasting (WSF) is of great significance to the power system planning and operation. In this paper, a hybrid method based on Minimal Gated Unit (MGU) and Quantile Regression (QR) is proposed for 2-hour WSF. Firstly, abnormal data is filtered by using the operating mode of SCADA system and Linear Interpolation algorithm is used for missing data imputation. Secondly, this paper embeds conditional quantile as an internal unit of MGU network. We estimate the parameters of MGU network under different quantile conditions, and calculate output under each quantile condition to obtain probabilistic WSF. At last, SCADA data collected from three wind turbines is applied to test the model performance. Both point and interval evaluation criteria are applied to evaluate the performance of models. The results show that the proposed model can obtain multi-step WSF with both point and interval prediction, compared with typical methods, it has higher accuracy in interval predictions and lower computation cost.

1. Introduction
The volatility of wind speed brings challenges to the development of wind energy industry. Accurate wind speed forecasting (WSF) is important for power dispatching and risk control. Therefore, WSF is essential for the operation of wind farms.

According to the previous studies, WSF methods can be divided into deterministic models and probabilistic models. Deterministic models include numerical weather prediction, persistence method, time series analysis, neural networks, etc. [1-4]. In recent years, deep learning methods, such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been wildly applied in WSF. Compared with traditional machine learning methods, deep learning models can extract the deep inherent features of wind speed and have good prediction performance due to its strong learning ability [5-8].

However, deterministic models are not able to provide risk or uncertainty of predicted wind speed. Therefore, researchers proposed probabilistic forecasting methods by adding the uncertainty analysis on the basis of deterministic models. In order to achieve probabilistic forecasting, [9] proposed a novel prediction framework, which directly generate the prediction intervals by calculating the confidence levels of coverage and sharpness. In [10], Variational Bayesian inference is applied to build an approximated posterior parameter distribution of the wind speed series. [11-12] established probabilistic WSF models by using Gaussian process. In recent years, Quantile Regression (QR) is widely used in probabilistic WSF for it can provide stable prediction information without prior
distribution assumption [13]. Several studies combine deep learning methods and QR to enhance the nonlinear mapping ability of WSF [14]. But this type of hybrid methods usually has complex model structure and large number of model parameters. In this paper, a novel probabilistic WSF model is proposed based on Minimal Gated Unit (MGU) and QR. The proposed method reduces the training time with less optimization variables and high prediction accuracy. We use measured data from the real wind farms to evaluate the performance of the proposed model. The results demonstrate that, the proposed model can make a multi-step predictions of wind speed in the next two hours. At the same time, it can provide an accurate and reliable uncertainty estimation.

2. Methodology

2.1. Quantile Regression

Quantile Regression is developed to study the effect of the independent variable $x$ on the conditional quantile of the dependent variable $y$ [15], which can be expressed as:

$$Q_\tau(x) = f(x\beta + \beta \tau X) + \beta \tau X$$

(1)

Where $Q_\tau(x)$ is the conditional $\tau$-th quantile of the dependent variable $y$ in the independent variable $X[1, x_2, ..., x_k]^T$, and $\tau \in (0, 1)$; $\beta(\tau) = [\beta_0(\tau), \beta_1(\tau), ..., \beta_k(\tau)]^T$ is the regression coefficient vector, which varies with the percentile $\tau$.

The estimated value $\hat{\beta}(\tau)$ of $\beta(\tau)$ can be obtained by equation (2):

$$\hat{\beta}(\tau) = \arg \min \sum_{i=1}^{N} \rho_\tau(y_i - X^\tau \beta)$$

(2)

Where the formula of $\rho_\tau(u)$ is defined as:

$$\rho_\tau(u) = \begin{cases} \tau u & u \geq 0 \\ (\tau - 1)u & u < 0 \end{cases}$$

(3)

2.2. Minimal Gated Unit

As a new Recurrent Neural Network (RNN) model, MGU further couples the input (reset) gate to the forget (update) gate, and the structure is illustrated in Figure 1.

According to [16], MGU has the minimal design with only one gate (the forget gate), which reduce the training time with simplified structure and high prediction accuracy. The implementation process of MGU and the calculation of corresponding gate are as follows:

1) Calculate the output of forget gates $f_i$:

$$f_i = \sigma(W_f [h_{i-1}, x_i] + b_f)$$

(4)

2) Calculate the information state $\tilde{h}_i$:

$$\tilde{h}_i = \tanh(W_h [f_i, e \ h_{i-1}, x_i] + b_h)$$

(5)

3) Calculate the hidden layer $h_i(\tau)$:
\[ h_t = (1 - f_t) e^{h_{t-1}} + f_t e^{f_t} \]  \hfill (6)

Where \( x_t \) represents input vector at time \( t \); \( h_t \) represents hidden state; \( \sigma \) represents the logistic sigmoid function; \( W_f \) and \( W_h \) are the weight matrix; \( b_f \) and \( b_h \) are the bias vectors; \( e \) means component-wise product.

### 2.3. The proposed model

Due to the nonlinear characteristics of wind speed series, single linear regression model is difficult to achieve interval prediction with high-precision. However, deep learning methods have powerful learning ability in nonlinear mapping, from which can learn the features from their multi-layer structure. Therefore, this paper proposes a hybrid model QRMGU based on MGU and QR. We embed quantile conditional as an internal unit of the MGU network. Then we estimate the parameters of MGU network under different quantile conditions, and calculate output under each quantile condition. Among them, L2 regularization and dropout are used to avoid over fitting, while the weight and bias are updated by Adam optimization. The detailed processes are as follows:

1) Calculate the output of forget gates \( f_t(\tau) \):
\[
f_t(\tau) = \sigma(W_f(\tau)[h_{t-1}(\tau), x_t] + b_f(\tau)) \]  \hfill (7)

2) Calculate the current information state \( \hat{h}_t(\tau) \):
\[
\hat{h}_t(\tau) = \tanh(W_h(\tau)[f_t(\tau) \odot h_{t-1}(\tau), x_t] + b_h(\tau)) \]  \hfill (8)

3) Calculate the hidden layer \( h_t(\tau) \):
\[
h_t(\tau) = (1 - f_t(\tau)) \odot h_{t-1}(\tau) + f_t(\tau) \odot \hat{h}_t(\tau) \]  \hfill (9)

4) Calculate the output:
\[
Q_t(\tau|x) = f(x, \beta(\tau)) = f(x, W(\tau), b(\tau)) = \sigma(W_f(\tau) \cdot h_t(\tau) + b_f(\tau)) \]  \hfill (10)

### 2.4. data preprocessing

At first, abnormal data is filtered according to the operating mode of SCADA system. As an RNN based method, MGU requires continuous time series for model training. In order to enhance the model performance, Linear Interpolation algorithm is used for missing data imputation, which is expressed as follows:
\[
v'_t = \frac{(v_{t-1} + v_{t+1})}{2} \]  \hfill (11)

Where \( v'_t \) is missing data point, \( v_{t-1} \) is the data point before the missing data points, \( v_{t+1} \) is the data from the moment after the missing data points.

Finally, we reconstruct time series of wind speed into a form of phase space by using the theory of chaos and phase space, the formula is as follows:
\[
X = \begin{bmatrix}
X_1 & X_2 & \ldots & X_k \\
X_2 & X_3 & \ldots & X_{k+1} \\
& \ldots & \ldots & \ldots \\
X_{N-k+1} & X_{N-k+1} & \ldots & X_N \\
\end{bmatrix} \]  \hfill (12)

Where \( N \) is the number of data, and \( k \) is the dimension of phase space. In this paper, \( k \) is 4 calculated by cross-validation.

### 3. Evaluation Criteria

#### 3.1. Point prediction

Root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the performance of models. The formulas are as follows:
RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \quad (13)

MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (14)

Where \( y_t \) represents the actual wind speed, \( \hat{y}_t \) is the forecasted values, and \( N \) represents the number of the wind speed series.

3.2. Interval prediction

As an evaluation function of probabilistic model, continuous ranked probability score (CRPS) is widely used in the field of probabilistic WSF. CRPS considers both forecasting reliability and interval sharpness. The smaller the value of CRPS, the better the comprehensive performance. The formula of CRPS is as follows:

\[
CRPS = \frac{1}{N} \sum_{t=1}^{N} \int_{-\infty}^{\infty} \left[ F(y) - H(y - y_t) \right]^2 dy
\]

\[
F(y) = \int_{-\infty}^{y} p(x) dx
\]

\[
H(y - y_t) = \begin{cases} 
0 & y < y_t \\
1 & y \geq y_t
\end{cases}
\]

where \( p(y) \) produces a probability distribution for \( y \), \( F(y) \) is the value of the corresponding cumulative distribution function, and \( H(y-y_t) \) is the step function.

4. Case study

In this study, the proposed model has been comprehensively tested on three wind speed datasets at two wind farms in Jiangsu Province and Yunnan Province, China. Data from the former wind farm is collected from 2018/07/10 to 2018/07/20 and from 2018/08/10 to 2018/09/01. Data from the latter wind farm is collected from 2018/04/20 to 2018/05/10. 70% of data for model training, and 30% for testing. We use three datasets and take the average as the final result. All the programs in this study are implemented on a computer with an Intel i7-9750H CPU (main frequency is 2.6-GHz) and 16.0 GB RAM.

To fully validate the effectiveness of the proposed model, the results are compared with the Quantile Regression Neural Network (QRNN), Quantile Regression Long Short-Term Memory (QRLSTM) and Quantile Regression Gated Recurrent Unit (QRGRU).

For each dataset, four different models are adopted. The average values of RMSE, MAPE, and CRPS obtained by the four forecasting models are shown in Figure 2-4, and the detail is shown in Table 1.

**Figure 2.** MAPE of four models.

**Figure 3.** RMSE of four models.
In Figure 2-4, with the increase of the number of predicted steps, the advantage of QRMGU in the accuracy of point prediction is remarkable. Compared with QRNN, QRLSTM and QRGRU, the average values of RMSE of eight steps decreases by 14.49%, 7.01% and 5.55% respectively, and the average values of MAPE decreases by 25.06%, 7.08% and 5.42%, respectively. The average values of training time of QRNN, QRLSTM, QRGRU and QRMGU in three datasets is 91s, 173s, 141s and 96s. Compared with QRNN, QRLSTM, and QRGRU, the CRPS of QRMGU model decreased by 22.11%, 10.81% and 9.23%.

Table 1. Comparison between comparison models and the proposed model

| Predict step | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|
| RMSE         |       |       |       |       |       |       |       |       |
| QRNN         | 0.71  | 1.03  | 1.30  | 1.54  | 1.75  | 1.95  | 2.14  | 2.33  |
| QRLSTM       | 0.70  | 0.99  | 1.22  | 1.40  | 1.56  | 1.72  | 1.87  | 2.02  |
| QRGRU        | 0.69  | 0.98  | 1.20  | 1.39  | 1.54  | 1.69  | 1.83  | 1.97  |
| QRMGU        | 0.69  | 0.96  | 1.16  | 1.30  | 1.45  | 1.55  | 1.67  | 1.79  |
| MAPE         |       |       |       |       |       |       |       |       |
| QRNN         | 7.75  | 11.54 | 14.51 | 17.09 | 19.32 | 21.32 | 23.18 | 25.06 |
| QRLSTM       | 6.74  | 9.69  | 11.90 | 13.60 | 15.14 | 16.55 | 18.06 | 19.55 |
| QRGRU        | 6.68  | 9.56  | 11.65 | 13.43 | 14.90 | 16.20 | 17.65 | 19.03 |
| QRMGU        | 6.60  | 9.29  | 11.16 | 12.67 | 13.97 | 15.06 | 16.24 | 17.47 |
| CRPS         |       |       |       |       |       |       |       |       |
| QRNN         | 0.38  | 0.57  | 0.74  | 0.88  | 0.97  | 1.05  | 1.11  | 1.17  |
| QRLSTM       | 0.38  | 0.53  | 0.66  | 0.75  | 0.81  | 0.87  | 0.92  | 0.97  |
| QRGRU        | 0.38  | 0.53  | 0.65  | 0.74  | 0.79  | 0.85  | 0.90  | 0.93  |
| QRMGU        | 0.36  | 0.50  | 0.59  | 0.66  | 0.70  | 0.74  | 0.79  | 0.83  |

Compared with benchmark models (QRNN, QRLSTM and QRGRU), the average MAPE, RMSE and CRPS of QRMGU decrease by 9.01%, 12.52% and 14.05% respectively. The results show that the proposed model reduces the computation while maintaining the prediction accuracy, so the model has the best performance of probabilistic prediction. Therefore, QRMGU has the best comprehensive performance.

5. Conclusion
This paper proposes a hybrid model based on MGU and QR for multi-step WSF. MGU has the minimal design with only one gate (the forget gate), which can reduce the parameters while maintaining the prediction accuracy. QR can provide the uncertainty value of prediction. In order to verify the performance of the model, the wind speed data of three wind turbines from different wind farms are used to make comparative analysis. Compared with other three forecasting models, the
average MAPE, RMSE and CRPS of QRMGU decreased by 9.01%, 12.52% and 14.05% respectively. The experimental results show that the proposed model has the best performance of probability prediction with high precision and small operation.

Acknowledgement
This work is supported by the National Natural Science Foundation of China (No. 61573046) and Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1203).

References:
[1] Cassola F and Burlando M 2012 Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output Appl. Energy 99 154-166
[2] Wang H, Han S, Liu Y, Yan J and Li L 2019 Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system Appl. Energy 237 1-10
[3] Aasim, Singh SN and Mohapatra A 2019 Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. Renew. Energy 136 758-768
[4] Yuan D, Qian Z, Jing B and Pei Y 2018 Short-term wind speed forecasting using STLSSVM hybrid model 2018 International Conference on Power System Technology (POWERCON) (China: Guangzhou)
[5] Chen Y, Zhang S, Zhang W, Peng J and Cai Y 2019 Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting Energy Convers. Manage. 185 783-799
[6] Liu H, Mi X and Li Y 2018 Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network Energy Convers. Manage. 166 120-131
[7] Ding M, Zhou H, Xie H, Wu M, Nakanishi Y and Yokoyama R 2019 A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting Neurocomputing 365 54-61
[8] Zhang J, Wei Y and Tan Z 2020 An adaptive hybrid model for short term wind speed forecasting Energy 190
[9] Wang H, Xue W, Liu Y, Peng J and Jiang H 2020 Probabilistic wind power forecasting based on spiking neural network Energy 196
[10] Liu Y, Qin H, Zhang Z, Pei S, Jiang Z, Feng Z and Zhou J 2020 Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model Appl. Energy 260
[11] Zhu S, Yuan X, Xu Z, Luo X, Zhang H 2019 Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast Energy Convers. Manage. 198
[12] Zhang Z, Ye L, Qin H, Liu Y, Wang C, Yu X, Yin X and Li J 2019 Wind speed prediction method using Shared Weight Long Short-Term Memory Network and Gaussian Process Regression Appl. Energy 247 270-284
[13] Haque A.U, Nehrir M.H and Mandal P 2014 A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting IEEE Trans. Power Syst. 29(4): 1663-1672
[14] Zhao X, Jiang N, Liu J, Yu D ang Chang J 2020 Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework Energy Convers. Manage. 203
[15] Koenker R, Hallock KF 2001 Quantile regression J. Econ. Perspect. 15(4) 143-156
[16] Zhou G, Wu J, Zhang C and Zhou Z 2016 Minimal Gated Unit for Recurrent Neural Networks. Int. J. of Automation and Computing 13(03) 226-234