

## Research Article

# Comparison of the Applicability of Two Reanalysis Products in Estimating Tall Tower Wind Based on Multiple Linear Regression and Artificial Neural Network in South China

Xiangxiang Li <sup>1</sup>, Xiaochen Qin <sup>2</sup>, Jun Yang <sup>1</sup>, and Weiming Xu <sup>1</sup>

<sup>1</sup>Meteorological Science Institute of Jiangxi Province, Nanchang, Jiangxi, China

<sup>2</sup>Jiangxi Provincial Climate Center, Nanchang, Jiangxi, China

Correspondence should be addressed to Weiming Xu; 653437554@qq.com

Received 29 April 2022; Revised 13 August 2022; Accepted 7 September 2022; Published 11 October 2022

Academic Editor: Stefano Federico

Copyright © 2022 Xiangxiang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Climate reanalysis products have been widely used to overcome the absence of high-quality and long-term observational records for wind energy users. In this study, the applicability of two popular reanalysis datasets (ERA5 and MERRA2) in estimating wind characteristics for four tall tower observatories (TTOs) in South China was assessed. For each TTO, linear and nonlinear downscaling techniques, namely, multiple linear regression (MLR) and an artificial neural network (ANN), respectively, were adopted for the downscaling of the scalar wind speed and the corresponding U/V components. The downscaled wind speed and U/V components were subsequently compared with the TTO observations by correlation coefficient (Pearson's  $r$ ), the root mean square error (RMSE), the uncertainty analysis (U95), and the reliability analysis (RE). According to the results, ERA5 had a better applicability (higher Pearson's  $r$  and RE, but lower RMSE and U95) in estimating TTO wind speed than MERRA2 when using both the MLR and ANN downscaling method. The average Pearson's  $r$ , RE, RMSE, and U95 of the downscaled wind from ERA5 by the MLR (ANN) method were 0.66 (0.69), 40.8% (41.8%), 2.20 m/s (2.11 m/s), 0.181 m/s (0.179 m/s), respectively, and 0.60 (0.63), 38.0% (39.7%), 2.32 m/s (2.25 m/s), 0.189 m/s (0.187 m/s), respectively, for MERRA2. The wind components analysis showed that the better performance of ERA5 was attributed to its smaller error in estimating V component than MERRA2. For the wind direction, the two reanalysis datasets did not display distinct differences. Additionally, the misalignment of the wind direction between the reanalysis products and the TTOs was higher for the secondary predominant wind direction (SPWD) than for the predominant wind direction (PWD). Furthermore, we found that the reanalysis U wind had a higher RMSE but a lower RE and Pearson's  $r$  than the V wind, which indicates that the misalignment in the wind direction was mainly associated with the deviation in the U component.

## 1. Introduction

High-quality wind observations are essential for wind resource assessment prior to the construction of wind farms [1]. Traditionally, the hub-height wind speed is measured at tall tower observatories (TTOs) at the representative location of a wind farm, and observations should be recorded for at least one year. However, factors such as tower icing and equipment failure may lead to the absence of wind records. Thus, TTO wind measurements must be inspected before being used for wind resource assessment.

Traditionally, missing wind records were interpolated with data from nearby meteorological stations. However, the station data must meet certain prerequisites, such as the climate conditions and the time resolution should be similar to the TTO. A low-density distribution of meteorological stations may limit the effectiveness of using station data for interpretation, especially in developing countries. In China, most stations are located near urban areas and are usually far from areas suitable for wind generation and development. This could create distinct microclimates between TTO and meteorological station.

Reanalysis datasets are the outputs of numerical models based on reprocessed and assimilated meteorological observations [2, 3]. Such data have global coverage and are more resolved in space and time than meteorological stations. Thus, reanalysis has gained popularity among many wind power researchers [4, 5]. In recent years, MERRA and MERRA2—with hourly temporal resolutions and an adequate wind speed height (50 m)—have been widely used in wind power modeling [6, 7], and the results have displayed relatively low errors when compared to measurements. In recent years, a new-generation reanalysis product named ERA5 has been released, and many studies have demonstrated that ERA5 has better accuracy than MERRA2 for wind resource assessment [1, 8, 9]. In certain applications, the variability of surface wind is often represented by the wind error in space and time. However, few studies have compared the differences in applicability of different reanalysis wind speeds at the hub height.

Since wind resource assessments may be required at various scales that differ from the pixel scale for reanalysis, downscaling is necessary for various applications. Downscaling techniques are classified into statistical and dynamic methods [10]. Dynamic downscaling uses numerical weather models to produce high-resolution variables for selected regions or in situ sites by processing the physical mechanism of a land-atmosphere system; notably, large amounts of computational resources and detailed knowledge of land-surface properties are required [11]. In contrast, statistical downscaling relies on the statistical relationships between in situ data and large-scale variables, thus providing an attractive alternative to dynamic downscaling due to its simplicity and lower computational demands [12]. The most widely used approach for establishing statistical relationships is the multiple linear regression (MLR), which has been proven effective in estimating regional climate variables such as precipitation [13] and temperature [14]. However, since the relationships among climate variables are highly nonlinear, some high-potential tools, such as artificial neural networks (ANNs), have risen in popularity for use in complex nonlinear models. With respect to ANN, the distinctive structure of the network and the nonlinear transfer function related to each hidden and output node allow ANNs to estimate highly nonlinear relationships [12]. Moreover, while other regression techniques assume a functional form, ANNs allow the data to define the functional form [15, 16]. Comparisons have been performed between MLR and ANNs in estimating regional temperature [17] and precipitation [12]. However, comparisons of different downscaling methods have rarely been conducted for wind variables.

China has been experiencing a high development speed in wind energy development in recent years. MERRA or MERRA2 has been widely used in characterizing wind resource in China [18, 19]. However, few studies have analyzed the reliability of the new ERA5 in China. Thus, the purpose of this paper was to reveal the reliability of ERA5 and MERRA2 in estimating tall tower wind characteristics and to explore the applicability of MLR and ANN downscaling methods.

## 2. Datasets and Methods

**2.1. Tall Tower Observations.** The observed wind speed data used in this study came from four TTOs located in mountainous areas of South China (Figure 1). Two towers were 80 m high, and the other two were 90 m. The data period for each tower was listed in Table 1. Data were recorded at hourly resolution.

**2.2. Reanalysis Datasets.** MERRA2 from the NASA Center for Climate Simulations was released in 2016 to replace the first version of MERRA. MERRA2 uses an upgraded data assimilation system, namely, the Goddard Earth observing system model, version 5 (GEOS-5) data assimilation system. ERA5 is the most recent climate reanalysis product of the European Center for Medium-range Weather Forecasts. The major improvement of ERA5 over the former product, ERA-Interim, is that it has higher spatial and temporal resolutions and provides wind variables at 100 m, which makes it particularly suitable for wind resource assessment.

The characteristics of MERRA2 and ERA5 were presented in Table 2. Both reanalysis products had the same temporal resolution of 1 hour, but MERRA2 had a coarser spatial resolution.

**2.3. Statistical Downscaling.** MLR and ANN, representing linear and nonlinear methods, respectively, were used to downscale the reanalysis data to the TTOs. For MLR, the overall relationship can be written as follows:

$$y = a_1 \times x_1 + a_2 \times x_2 + a_3 \times x_3 + \dots + a_n \times x_n + b, \quad (1)$$

where  $y$  represents the simulated results;  $x_i$  represents the independent variables;  $a_i$  is the estimated regression coefficient for each independent variable; and  $b$  is the deviation value.

For ANN, we selected the back-propagation (BP) network for analysis. A BP network contained three layers: the input layer, one or more hidden layers, and the output layer. Each layer consisted of processing elements, called neurons, which were fully connected to the neurons of the preceding layers by weight ( $W$ ) and bias ( $b$ ). The training of a BP network involved determining the best combination of  $W$  and  $b$  for each neuron that produces the smallest difference between the output and actual value. For the input and output layers, the number of neurons depended on the input and output variables of the problem. However, the number of neurons in the hidden layer depended on the complexity of the problem, which was specifically determined during the training of the network. Typically, a BP network with one hidden layer can predict any nonlinear input-output relationship. Figure 2 illustrates the BP network considered in this study.

In the hidden layer, each neuron was passed an input value and exported the corresponding outcome based on the following processing:

$$a_n = \sum W_{in}x_i + b_n, \quad (2)$$

$$O_n = \varphi(a_n), \quad (3)$$

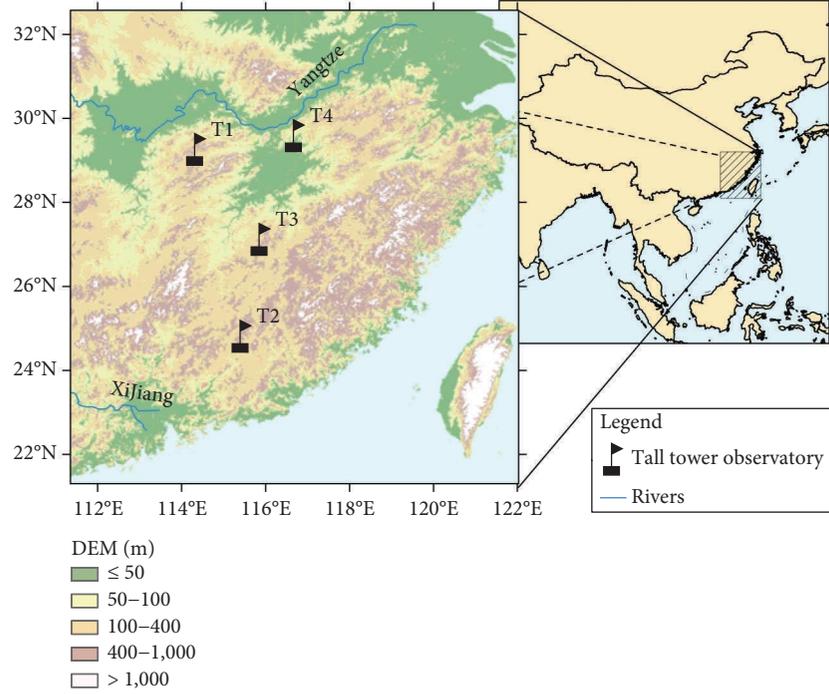


FIGURE 1: Geographical information for the tall tower observatories.

TABLE 1: General information for the tall tower observatories.

Tall tower observatory	Elevation (m)	Longitude	Latitude	Data length (days)	Measurement height (m)	Time resolution (h)
T1	996	114.28	29.24	349	90	1
T2	1009	115.37	24.78	410	80	1
T3	918	115.82	27.09	337	90	1
T4	372	116.64	29.57	362	80	1

TABLE 2: Characteristics of MERRA2 and ERA5.

	MERRA2	ERA5
Temporal resolution	1 h	1 h
Spatial resolution	$0.5^\circ \times 0.625^\circ$	31 km ( $\approx 0.28^\circ$ )
Wind speed height	50 m	100 m

where  $O_n$  is the outcome of the  $n$  th neuron;  $a_n$  is the weighted sum at the  $n$  th neuron;  $W_{in}$  is the weight for the connection of the  $i$  th input neuron and the  $n$  th hidden layer neuron;  $x_i$  is the  $i$  th input variable;  $b_n$  is the bias of the  $n$  th hidden layer neuron; and  $\varphi$  is the nonlinear activation function for the hidden layer.

The output layer neuron implemented the same processing operations as the hidden layer neuron. The output of the network ( $S$ ) can be written as follows:

$$S = \tau\left(\sum O_n W_n\right) + b, \quad (4)$$

where  $\tau$  is the nonlinear activation function for the output layer, and  $b$  is the bias of the output layer neuron.

**2.4. Data Process.** The flowchart of this study was presented in Figure 3. The wind records for each TTO and reanalysis data were randomly divided into three groups: a training group, a validation group, and a testing group, with a

proportion of 7:1:2. The training group was used for training the MLR and ANN models. During the training process, five variables of the reanalysis grid which was nearest to the TTO were selected as model inputs: the reanalysis wind speed ( $W$ ), 2-meter air temperature ( $T$ ), hourly variation in the 2-meter air temperature ( $DT$ ), surface pressure ( $P$ ), and hourly variation in surface pressure ( $DP$ ), as these variables have been shown to have a strong physical relationship with the wind speed parameter and have been used to train neural network models in other studies [20, 21]. That is, the ANN models had five neurons in the input layer. For the MLR method, the variables with insignificant regression coefficients were eliminated from the regression equation until all the coefficients reached a significant level [22]. While training the model, the validation group was established to ensure that ANN training did not result in overfitting. After the MLR and ANN model were trained, variables from the testing groups in MERRA2 and ERA5 were applied to estimate the wind values, respectively. The accuracy of the estimated results were then determined by comparing them with the actual values of the four TTOs using the following evaluation metrics:

**Correlation coefficient (Pearson's  $r$ ):** Pearson's  $r$  depicts the degree of agreement between the simulated and the

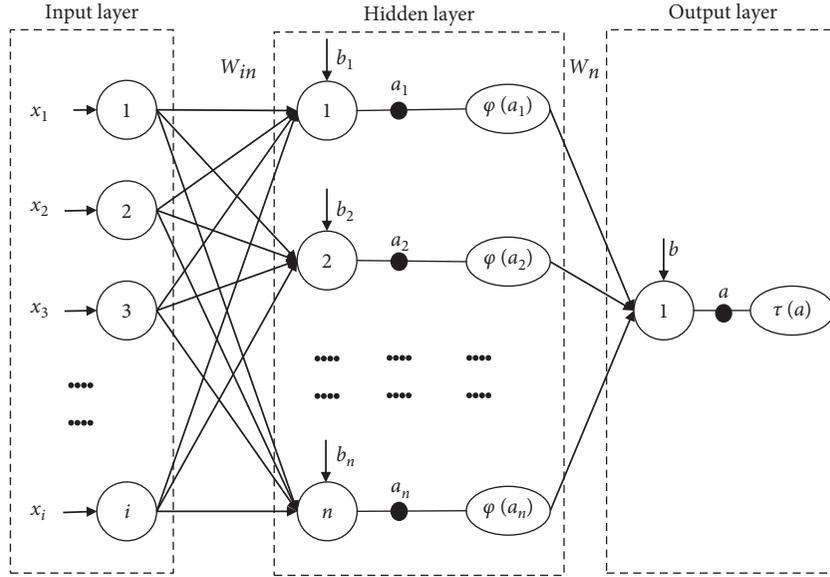


FIGURE 2: Architecture of a three-layer feed-forward back-propagation network.

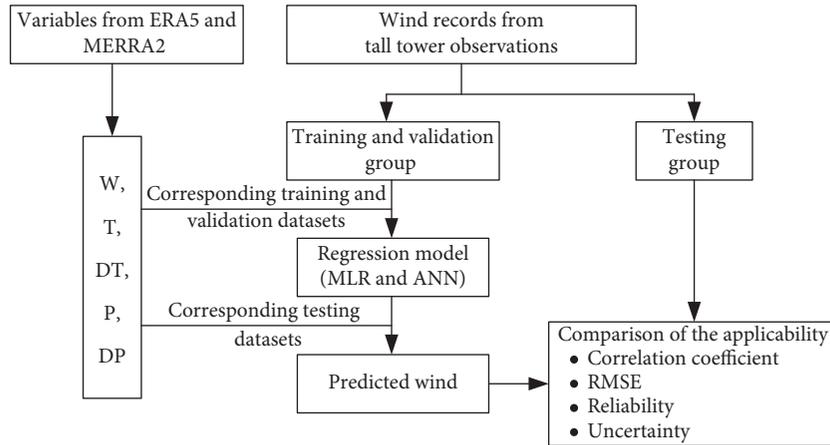


FIGURE 3: Research flowchart. Five variables of the reanalysis were selected as model inputs, including  $W$  the reanalysis wind speed;  $T$  2-meter air temperature;  $DT$ : hourly variation in the 2-meter air temperature;  $P$  surface pressure;  $DP$ : hourly variation in surface pressure.

observed wind, with values range from 0 to 1, where a larger value of Pearson's  $r$  means higher agreements.

*Root mean squared error (RMSE)*: the RMSE represents the magnitude of deviation, and a lower RMSE value indicates a smaller magnitude of the bias.

*Uncertainty analysis (U95)*: the major goal of the uncertainty analysis is to restrict the expected range in which the true value of the outcome of an experiment lies. Here, following Saberi-Movahed et al. [23], the U95 was used to compute the uncertainty interval, which can be interpreted as follows: performing the given experiment repeatedly over and over again, the true value of the outcome of that experiment will lie in the offered uncertainty interval for approximately 95 times out of each 100 trials.

*Reliability analysis (RE)*: the reliability analysis is a statistical method for measuring the overall consistency

of a model. To be more specific, it determines if a suggested model achieves a permissible level of performance.

These evaluation metrics were given by

$$r = \frac{\sum_{i=1}^m (y_i - \bar{y}_i)(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^m (y_i - \bar{y}_i)^2} \sqrt{\sum_{i=1}^m (\hat{y}_i - \bar{\hat{y}}_i)^2}}, \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}, \quad (6)$$

$$U95 = \left(\frac{1.96}{m}\right) \sqrt{\sum_{i=1}^m (y_i - \bar{y}_i)^2 + \sum_{i=1}^m (\hat{y}_i - \bar{\hat{y}}_i)^2}, \quad (7)$$

TABLE 3: Network properties.

Input	Output	Network type	Activation function in hidden layer	Activation function in output layer	Performance function	Number of neurons in hidden layer
$W, T, DT, P, DP$	Wind speed and direction	Back-propagation	Tansig	Purelin	Mean square error (MSE)	10

$$RE = \left( \frac{100\%}{m} \right) \sum_{i=1}^m k_i, \quad (8)$$

where  $m$  is the length of testing group;  $y_i$  represents the wind speed of the tall tower observatory;  $\hat{y}_i$  is the wind speed simulated by the MLR or ANN. In Formula (8),  $k_i$  is obtained through two steps. First, the relative average error (RAE) is defined as a vector whose  $i$  th component is

$$RAE_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \quad (9)$$

Next, if  $RAE_i \leq \Delta$ , then  $k_i = 1$ ; otherwise,  $k_i = 0$ , where  $\Delta$  is the threshold value of reasonable wind deviation. In other words,  $k_i$  is defined as the number of times the value of RAE is less than or equal to that of  $\Delta$ . Following Saberi-Movahed et al. [23], the optimum value of  $\Delta$  is 0.2 or equivalently 20%.

Additionally, before training the ANN model, the input and output data were normalized using min-max normalization to avoid network nonconvergence due to the magnitude difference among different variables. In this study, a total of 24 ANNs were trained based on observations from 4 TTOs and 2 reanalysis products for 3 targets (the scalar wind speed, U component of wind, and V component of wind). Notably, the wind direction was calculated based on the U/V components. The properties of the network used in this study can be seen in Table 3. The training and testing procedures were implemented with the AMORE package in the R language.

### 3. Results and Analysis

**3.1. The Interpolation of Wind Speed.** The scatter plots of the observed wind speed and the wind speed simulated by the MLR and ANN interpolation methods were presented in Figures 4 and 5, respectively. The upper row of each plot was simulated from ERA5, and the second row was simulated from MERRA2. Among the four TTOs, T4 had the best correlation and the lowest RMSE for both the MLR and ANN methods; taking MLR as an example, the correlation coefficients were 0.80 and 0.70, and the RMSEs were 1.70 m/s and 1.94 m/s for ERA5 and MERRA2 (Figures 4(d) and 4(h)), respectively. In contrast, the correlation was the poorest for T3, and the RMSEs were 2.71 m/s and 2.79 m/s for ERA5 and MERRA2 based on the MLR method (Figures 4(c) and 4(g)), respectively.

For all the tall towers, the wind speed simulated from ERA5 yielded higher correlations and lower root mean square error (RMSE) than MERRA2 for both the MLR and ANN downscaling methods. For the MLR method, the average RMSE of ERA5 was 2.20 m/s, which was 0.12 m/s

lower than MERRA2. For the ANN method, the average RMSE of ERA5 was 0.14 m/s lower than MERRA2.

In addition, the interpolation results based on ANN were better than MLR for all four tall tower observatories. However, the difference in RMSE between different downscaling methods was smaller than that between different reanalyses. For ERA5, the average RMSE of ANN was 0.09 m/s lower than the MLR method. For MERRA, the average RMSE of the BP network was 0.07 m/s lower than the MLR method.

The results of uncertainty and reliability evaluation were presented in Table 4. Table 4 shows that ERA5 had a lower value of U95 for either the MLR or ANN downscaling method when compared to MERRA2. For example, for the ANN method, the average U95 of ERA5 was 0.179 m/s, which is lower than MERRA2 (U95 = 0.187 m/s). Table 4 also shows that the reliability of wind speed downscaled from ERA5 by both the MLR (RE = 40.8%) and ANN (RE = 41.8%) method was higher than MERRA2 (The RE value was 38.0% and 39.7% for the MLR and ANN, respectively). Additionally, comparing different downscaling methods, the ANN provided lower uncertainty but higher reliability than the MLR method. Take ERA5 as an example, the average uncertainty value was 0.179 m/s for the ANN method, which was lower than that for the MLR method (U95 = 0.181 m/s). Also, the reliability value was 41.8% for the ANN method, which was higher than that for the MLR method (RE = 41.8%). Moreover, it should be noted that, like the RMSE, the difference in uncertainty and reliability between different downscaling methods was smaller than that between different reanalyses. For example, the difference in U95 between different downscaling methods was 0.002 m/s for both ERA5 and MERRA2, while they were 0.008 m/s between different reanalyses for both the MLR and ANN methods.

**3.2. The Interpolation of Wind Direction.** Figures 6 and 7 show the wind rose plots drawn from the TTOs and reanalysis downscaled by the MLR and ANN methods, respectively. As shown, the wind rose plots drawn from the reanalysis were similar to those based on observations, and the difference in the wind rose plots based on ERA5 and MERRA2 was no significant difference in visual appearance, as was that between the plots based on the MLR and ANN interpretation methods.

Compared to TTOs, both the reanalysis products had better results for the predominant wind direction (PWD) than for the secondary predominant wind direction (SPWD). For example, for T2, the PWD was NNE (Figure 6(d)), which was perfectly estimated by EAR5 and MERRA2 by either the MLR method (Figures 6(e) and 6(f))

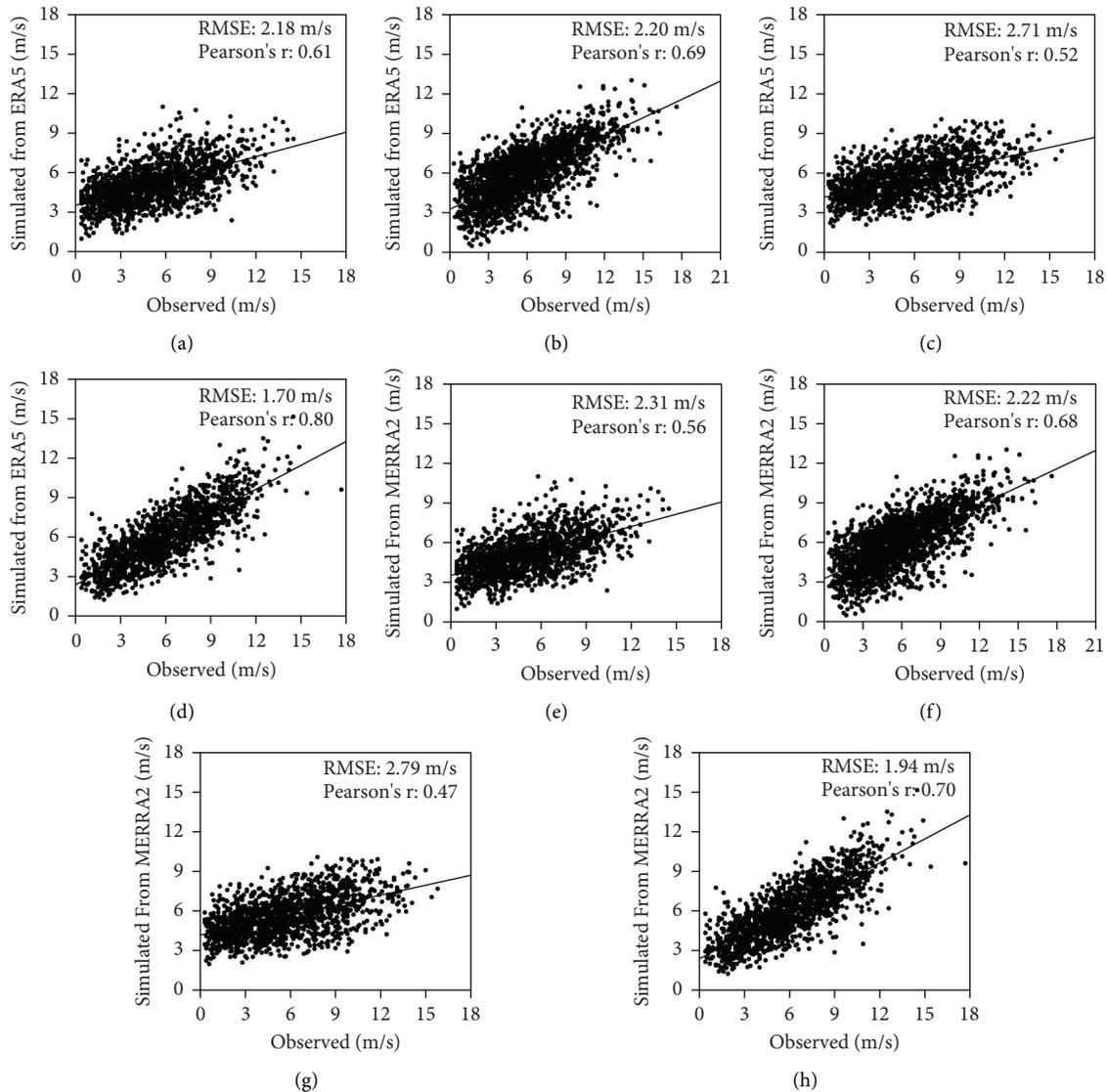


FIGURE 4: Scatterplots for the multiple linear regression (MLR) downscaling method. The upper row shows the results for ERA5, and the second row shows the results for MERRA2. T1, T2, T3 and T4 represent the four tall towers. (a) T1. (b) T2. (c) T3. (d) T4. (e) T1. (f) T2. (g) T3. (h) T4.

or the ANN method (Figures 7(e) and 7(f)). However, for the SPWD, direction S was obtained based on ERA5 and MERRA2 for both MLR and ANN downscaling method, which was 2 wind direction angles away from the TTO (SW direction). Similarly, for T3, the SPWD obtained from the downscaling of the reanalysis data was not clear (Figures 6(h) and 7(h) and Figures 6(i) and 7(i)), which was unlike the TTO which has a clear subdominant wind direction (NE direction). However, both the two reanalyses had clear PWD (SW direction, Figures 6(h) and 7(h) and Figures 6(i) and 7(i)), and the differences with the actual PWD measured by TTO were within 1 wind angle for both the MLR and ANN downscaling methods.

**3.3. The Interpolation of U/V Components of Wind.** Figure 8 presents the RMSE and Pearson's r for the downscaled U/V components of wind based on the MLR

(Figure 8(a)) and ANN (Figure 8(b)) methods. As shown, the figure showed an obviously lower RMSE but higher Pearson's r for the V wind than the U wind for both ERA5 and MERRA2. For example, for ERA5 downscaled by the MLR method (Figure 8(a)), the RMSE and Pearson's r were 2.80 m/s and 0.80 for the V wind, respectively, and 3.18 m/s and 0.55, respectively, for the U wind. Similarly, for MERRA2 downscaled by the MLR method (Figure 8(a)), the average RMSE of the U/V wind was 3.14/2.95 m/s, and Pearson's r was 0.55/0.80. The results were the same for the ANN downscaling method (Figure 8(b)).

For different interpolation methods, the ANN method displayed slightly better performance than MLR in simulating both U/V winds. The average RMSE of the U wind predicted by the ANN method was 3.07 m/s and 2.98 m/s for ERA5 and MERRA2, respectively (Figure 8(b)), and the corresponding values obtained for the MLR method were higher (3.18 m/s and 3.14 m/s for ERA5 and MERRA2,

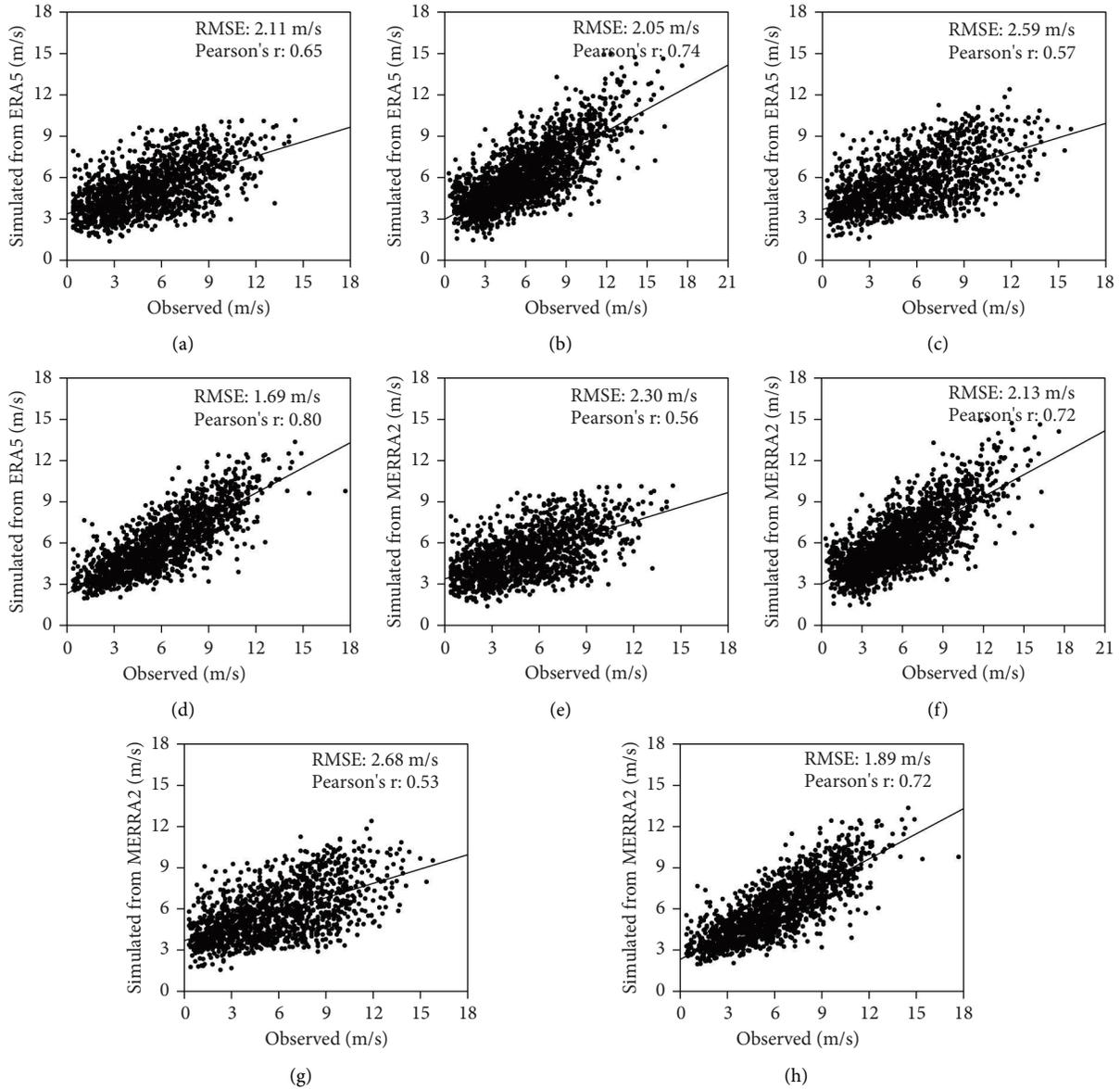


FIGURE 5: Scatterplots for the ANN technique. The upper row shows the results for ERA5, and the second row shows the results for MERRA2. T1, T2, T3 and T4 represent the four tall towers. (a) T1. (b) T2. (c) T3. (d) T4. (e) T1. (f) T2. (g) T3. (h) T4.

TABLE 4: Comparisons of the performance results for the uncertainty and reliability.

	ERA5				MERRA2			
	MLR		ANN		MLR		ANN	
	U95 (m/s)	RE (%)						
T1	0.178	38.3	0.175	38.7	0.183	32.5	0.182	34.1
T2	0.167	42.1	0.163	45.4	0.168	40.4	0.165	44.7
T3	0.218	30.8	0.214	30.5	0.221	29.2	0.217	30.0
T4	0.162	52.2	0.162	52.3	0.186	50.1	0.184	50.0
Average	0.181	40.8	0.179	41.8	0.189	38.0	0.187	39.7

respectively) (Figure 8(a)). Similarly, for the V wind, the average RMSE predicted by the ANN method was 2.68 m/s and 2.78 m/s for ERA5 and MERRA2, respectively; the RMSE values were higher for the MLR method (the corresponding RMSE was 2.80 m/s and 2.98 m/s, respectively).

Table 5 shows the uncertainty and reliability analysis for the U/V components of wind. From the results of uncertainty (U95), the V wind had higher U95 than the U wind for both ERA5 and MERRA2. The average U95 of V wind downscaled by the MLR method was 0.282 m/s and 0.296 m/

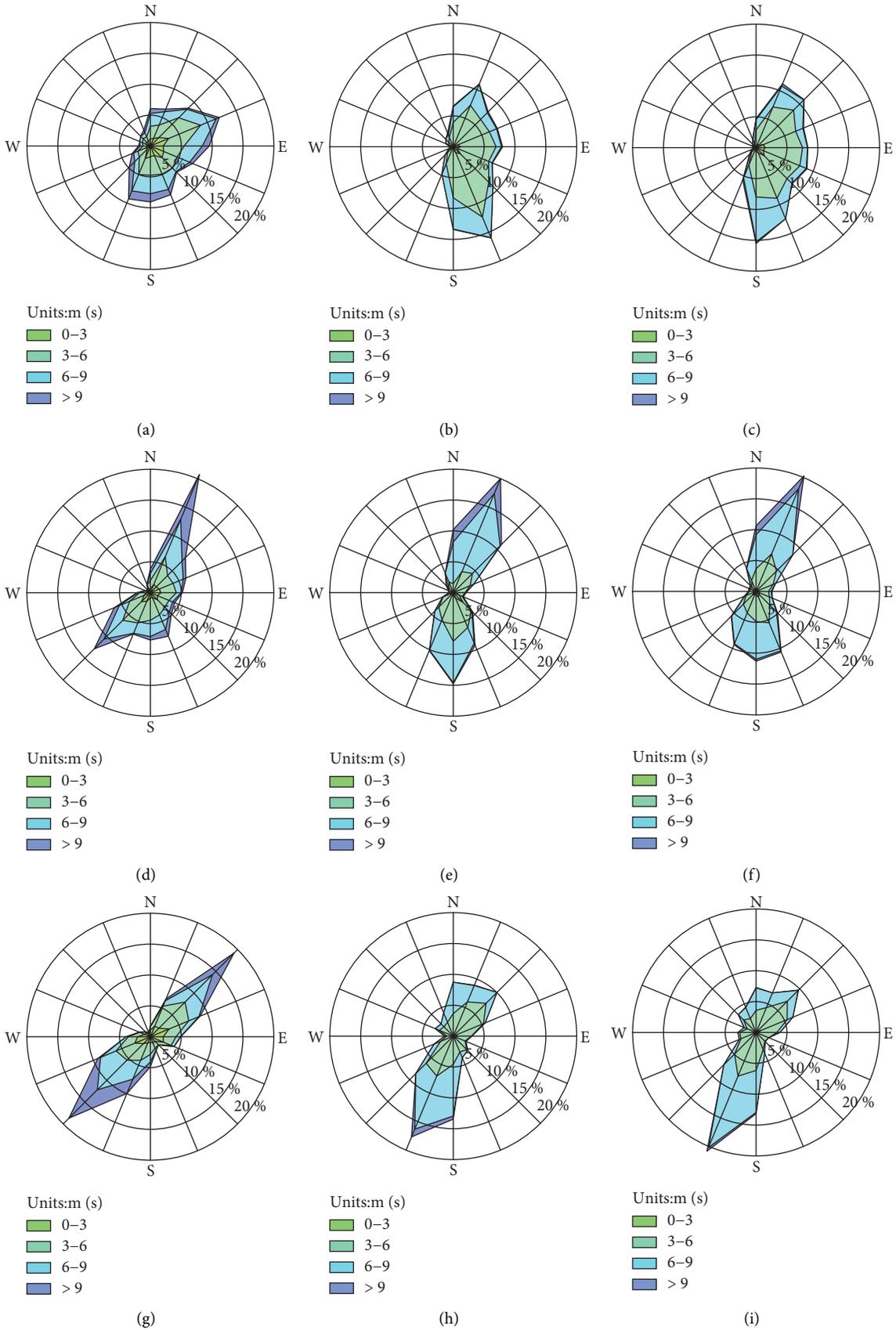


FIGURE 6: Continued.

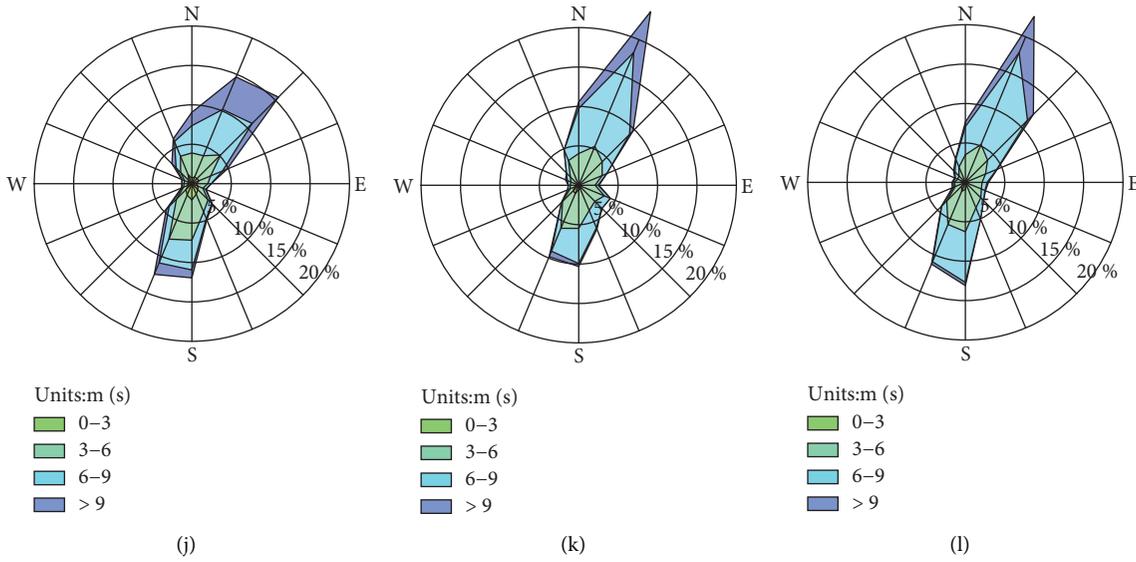


FIGURE 6: The wind rose plots for the observations, the ERA5 predictions and the MERRA2 predictions based on the MLR downscaling method. (a) T1-Obs. (b) T1-ERA5. (c) T1-MERRA2. (d) T2-Obs. (e) T2-ERA5. (f) T2-MERRA2. (g) T3-Obs. (h) T3-ERA5. (i) T3-MERRA2. (j) T4-Obs. (k) T4-ERA5. (l) T4-MERRA2.

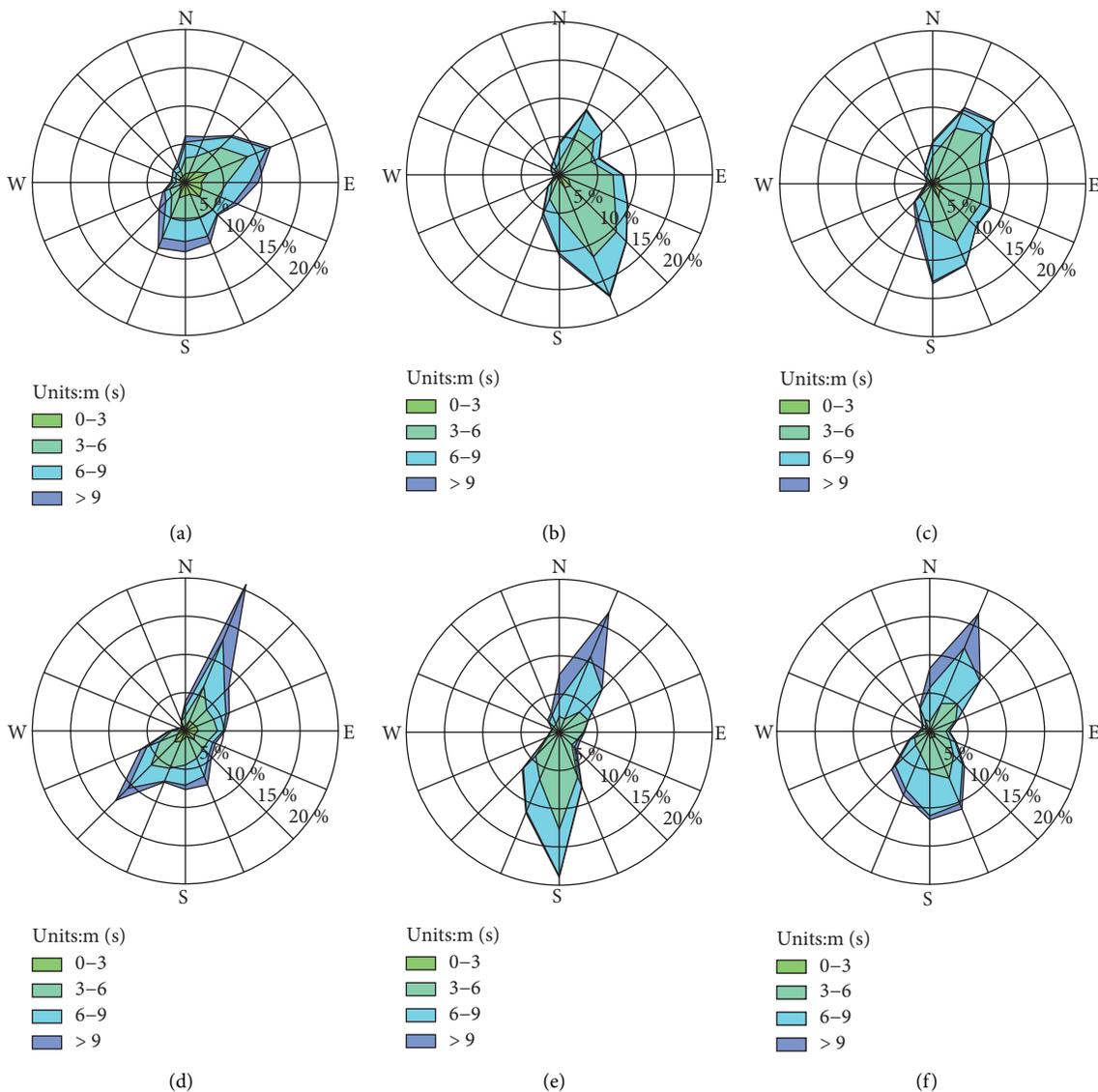


FIGURE 7: Continued.

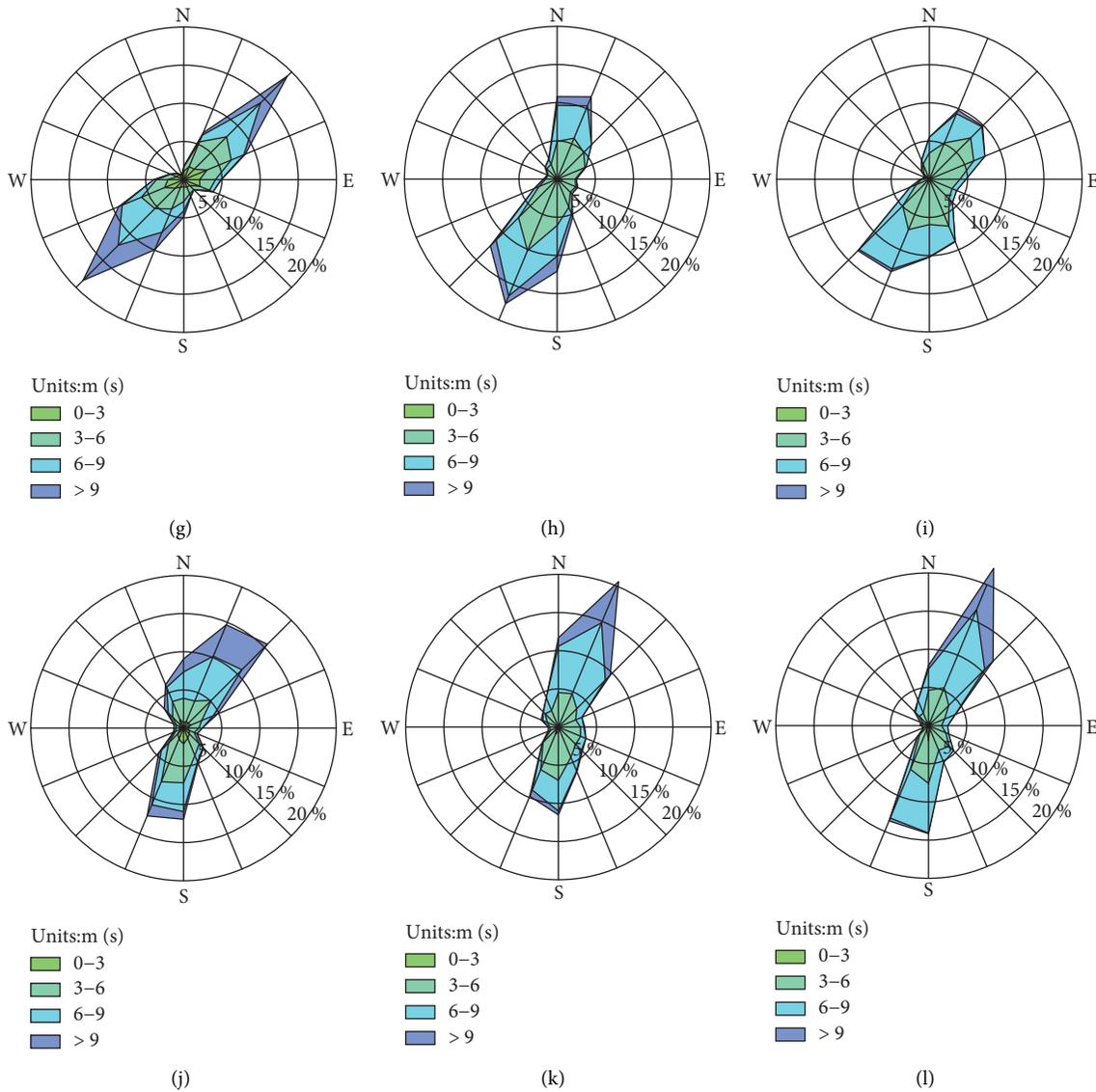


FIGURE 7: The wind rose plots for the observations, the ERA5 predictions and the MERRA2 predictions based on the ANN downscaling method. (a) T1-Obs. (b) T1-ERA5. (c) T1-MERRA2. (d) T2-Obs. (e) T2-ERA5. (f) T2-MERRA2. (g) T3-Obs. (h) T3-ERA5. (i) T3-MERRA2. (j) T4-Obs. (k) T4-ERA5. (l) T4-MERRA2.

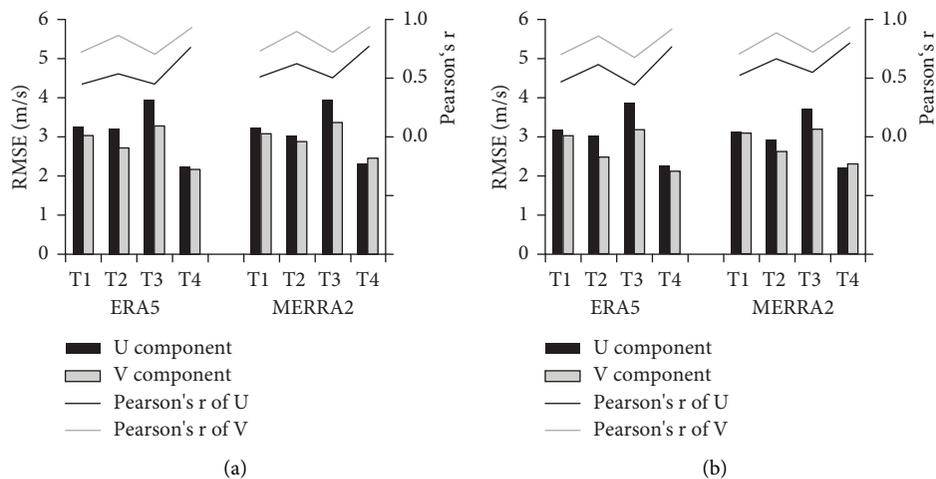


FIGURE 8: RMSE and Pearson's r for the U/V components of wind predicted by ERA5 and MERRA2 based on the MLR (a) and ANN (b) downscaling methods.

TABLE 5: Comparisons of the performance results for the U/V components of wind.

	ERA5				MERRA2			
	MLR		ANN		MLR		ANN	
	U95 (m/s)	RE (%)						
<i>U component of wind</i>								
T1	0.250	9.9	0.246	11.6	0.247	11.3	0.243	12.1
T2	0.223	11.5	0.217	13.4	0.219	12.2	0.216	16.3
T3	0.311	6.9	0.307	8.4	0.310	5.1	0.301	9.4
T4	0.204	18.9	0.204	18.6	0.234	18.2	0.230	19.1
Average	0.247	11.8	0.244	13.0	0.252	11.7	0.247	14.2
<i>V component of wind</i>								
T1	0.269	13.1	0.269	14.3	0.272	14.0	0.271	13.9
T2	0.271	26.1	0.266	28.8	0.275	24.7	0.270	28.8
T3	0.296	13.8	0.292	17.6	0.295	12.4	0.289	17.0
T4	0.293	39.8	0.292	41.1	0.341	35.0	0.338	37.8
Average	0.282	23.2	0.280	25.4	0.296	21.5	0.292	24.3

s for ERA5 and MERRA2, respectively, and 0.247 m/s and 0.252 m/s, respectively, for the U wind. It was the same for the ANN method. This is inconsistent with the results of RMSE and Pearson's  $r$ . According to Formula (7), the U95 is not a normalized value. It positively corresponds to the magnitude of wind. As Figures 6 and 7 shown, the most frequency of wind direction showed a south-north pattern, indicating that the wind component in the latitudinal direction (i.e., the V wind) was larger than that in the longitudinal direction (i.e., the U wind). This leads to a larger U95 for the V wind. As a normalized value, RE is more suitable for comparing the applicability of U/V wind. Table 5 shows that the RE of V wind was higher than U wind, which was consistent with the results of RMSE and Pearson's  $r$ . For example, the average RE of V wind downscaled from ERA5 by the MLR method was 23.2%, and that was 11.8% for the U wind. Furthermore, in the estimation of V wind, ERA5 showed better uncertainty and reliability than MERRA2. For example, the average U95 of the MLR method was 0.282 m/s for ERA5, which is lower than MERRA2 (U95 = 0.296 m/s). However, with respect to U wind, the applicability of the two reanalyses in different TTO had its own advantages and disadvantages. Additionally, by comparing different downscaling methods, we found that the ANN method produced better performance in the uncertainty and reliability analysis. Take V wind of ERA5 as an example, the value of U95 for the ANN method was 0.280 m/s, which was lower than the MLR method (U95 = 0.282 m/s). And the value of reliability for the ANN method was 25.4%, while it was 23.2% for the MLR method.

#### 4. Discussion

Climate reanalysis has been widely used to overcome the absence of high-quality and long-term observational records for wind energy users. The new ERA5 dataset has been proven superior to the traditional MERRA2 product in predicting the regional distributions of wind resources in the South China Sea [24], Europe [25], America [26], and other areas worldwide [9]. Here, we analyzed the applicability of these two reanalyses in estimating tall tower wind

characteristics using different statistical downscaling methods in South China.

On average, ERA5 displayed improvements (higher Pearson's  $r$  and RE, but lower RMSE and U95) over MERRA2 in the ability to estimate wind speed based on both the MLR and ANN downscaling methods. Similar results were obtained in Europe [1, 27, 28] and the Arctic [8]. For example, Ramon et al. [28] compared five global reanalysis data sets (including ERA5, MERRA2, ERA-Interim, JRA55, and NCEP-R1) with in situ observations worldwide, and the results showed that the ERA5 surface winds exhibited the best agreement with the in situ winds at a daily time scale. As for the reasons, it may include the various representations of land-surface roughness and varying data density and quality and the output properties in different reanalysis models. For example, the output wind of ERA5 was 100 m high, which was closer to the tall tower height in this study than MERRA2 [1]. Additionally, the accuracy of wind speed components for different reanalyses also contributes to the reasons. In this study, regardless of the downscaling method, the V component of ERA5 had a smaller error compared to MERRA2. For example, when using the ANN downscaling method, ERA5 produced a lower RMSE (2.68 m/s) than MERRA2 (2.78 m/s) when estimating the V component of wind (Figure 8(b)). However, with respect to U wind, the applicability of the two reanalyses in different TTO had its own advantages and disadvantages. Therefore, the better ability of ERA5 in estimating the V component of wind could be another reason for its better applicability over MERRA2.

For the wind direction, the downscaling simulations driven by ERA5 and MERRA2 did not display distinct differences in the reproduction of wind direction. This is consistent with the comparison among other reanalysis products [29]. They tested the reconstruction skill of three reanalysis products (i.e., ERA-Interim, CFSR, and JRA55) in reproducing the China Yellow Sea coastal winds and demonstrated that these products were consistent with each other in the reproduction of local wind direction. However, comparisons of the reconstruction skill of wind direction between ERA5 and MERRA2 are few studied in other

locations. Here, we found that both ERA5 and MERRA2 could reproduce the PWD better than the SPWD. The misalignment was typically within 1 wind direction angle for the PWD. However, the SPWD was biased by at least 2 wind direction angles. We reveal the reasons for wind direction differences by analyzing the bias of wind speed components, and the results showed that the V wind had a higher Pearson's  $r$  and RE but lower RMSE than the U wind for all four TTOs, indicating that the misalignment in wind direction was mainly associated with the deviation in the U component of wind.

Additionally, our results showed that the ANN method yielded better estimations of wind speed than the traditional MLR method. The average RMSEs of the ANN/MLR methods were 2.11/2.20 m/s and 2.25/2.32 m/s for ERA5 and MERRA2, respectively. The superiority of the ANN method over other downscaling methods has also been verified for other climate factors, such as precipitation and evapotranspiration [12, 22]. Our findings suggest that the ANN is an attractive alternative to traditional methods for the interpolation of wind records. However, we found that difference applicability between different downscaling methods was smaller than that between different reanalyses. Thus, for the wind resources assessment, a comparison of the applicability between different reanalysis products is a priority issue to be addressed.

## 5. Conclusion

In this study, wind characteristics downscaled from two reanalysis products by two downscaling methods were compared with the observations from four tall towers in South China. Various evaluation metrics (including the correlation coefficient, the root mean square error, the uncertainty analysis, and the reliability analysis) were applied to reveal the differences in applicability between different reanalysis products and different downscaling methods. The following specific conclusions were drawn from this study:

- (1) ERA5 had a better performance over MERRA2 in the ability to estimate wind speed based on both the MLR and ANN downscaling methods. Smaller error of ERA5 in estimating the V component could be one of the reasons for its superiority over MERRA2.
- (2) For the wind direction, the wind rose plots produced by ERA5 and MERRA2 were generally consistent. Additionally, both the reanalysis products were more accurate for the estimation of the predominant wind direction (PWD) than the secondary predominant wind direction (SPWD).
- (3) The ANN downscaling method yielded better metrics than the MLR method. However, the difference in capability between different downscaling methods was smaller than the difference between different reanalyses. Therefore, for the interpolation of wind records, the choice of reanalysis products is more important than downscaling methods.

## Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the data also form part of an ongoing study but are available from the corresponding author on reasonable request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Meteorological Science and Technology Project of Jiangxi Province (Grant number 20190108). The ERA5 was obtained from the Climate data store (<https://cds.climate.copernicus.eu/>). The MERRA2 was obtained from the NASA Center for Climate Simulation/Goddard Space Flight Center (<https://esgf.nccs.nasa.gov/>).

## References

- [1] J. Olauson, "ERA5: the new champion of wind power modelling?" *Renewable Energy*, vol. 126, pp. 322–331, 2018.
- [2] L. Bengtsson, P. Arkin, P. Berrisford et al., "The need for a dynamical climate reanalysis," *Bulletin of the American Meteorological Society*, vol. 88, no. 4, pp. 495–502, 2007.
- [3] L. Bengtsson, S. Hagemann, and K. I. Hodges, "Can climate trends be calculated from reanalysis data?" *Journal of Geophysical Research*, vol. 109, no. D11, p. D11111, 2004.
- [4] D. J. Cannon, D. J. Brayshaw, J. Methven, P. J. Coker, and D. Lenaghan, "Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in great Britain," *Renewable Energy*, vol. 75, pp. 767–778, 2015.
- [5] I. Staffell and S. Pfenniger, "Using bias-corrected reanalysis to simulate current and future wind power output," *Energy*, vol. 114, pp. 1224–1239, 2016.
- [6] J. Olauson and M. Bergkvist, "Modelling the Swedish wind power production using MERRA reanalysis data," *Renewable Energy*, vol. 76, pp. 717–725, 2015.
- [7] H. Zhang, Y. Cao, Y. Zhang, and V. Terzija, "Quantitative synergy assessment of regional wind-solar energy resources based on MERRA reanalysis data," *Applied Energy*, vol. 216, pp. 172–182, 2018.
- [8] R. M. Graham, S. R. Hudson, and M. Maturilli, "Improved performance of ERA5 in arctic gateway relative to four global atmospheric reanalyses," *Geophysical Research Letters*, vol. 46, no. 11, pp. 6138–6147, 2019.
- [9] G. Gualtieri, "Reliability of ERA5 reanalysis data for wind resource assessment: a comparison against tall towers," *Energies*, vol. 14, no. 14, p. 4169, 2021.
- [10] A. W. Wood, L. R. Leung, V. Sridhar, and D. P. Lettenmaier, "Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs," *Climatic Change*, vol. 62, no. 1–3, pp. 189–216, 2004.
- [11] J. Murphy, "An evaluation of statistical and dynamical techniques for downscaling local climate," *Journal of Climate*, vol. 12, no. 8, pp. 2256–2284, 1999.
- [12] E. Sharifi, B. Saghafian, and R. Steinacker, "Downscaling satellite precipitation estimates with multiple linear regression, artificial neural networks, and spline interpolation techniques," *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 2, pp. 789–805, 2019.

- [13] D. I. Jeong, A. St-Hilaire, T. B. M. J. Ouarda, and P. Gachon, "Multisite statistical downscaling model for daily precipitation combined by multivariate multiple linear regression and stochastic weather generator," *Climatic Change*, vol. 114, no. 3-4, pp. 567-591, 2012.
- [14] D. Joshi, A. St-Hilaire, T. Ouarda, and A. Daigle, "Statistical downscaling of precipitation and temperature using sparse Bayesian learning, multiple linear regression and genetic programming frameworks," *Canadian Water Resources Journal/Revue Canadienne des Ressources Hydriques*, vol. 40, no. 4, pp. 392-408, 2015.
- [15] P. Coulibaly, Y. B. Dibike, and F. Anctil, "Downscaling precipitation and temperature with temporal neural networks," *Journal of Hydrometeorology*, vol. 6, no. 4, pp. 483-496, 2005.
- [16] A. Pasini, "Artificial neural networks for small dataset analysis," *Journal of Thoracic Disease*, vol. 7, no. 5, pp. 953-960, 2015.
- [17] C. Wang, R. M. Graham, K. Wang, S. Gerland, and M. A. Granskog, "Comparison of ERA5 and ERA-interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution," *The Cryosphere*, vol. 13, no. 6, pp. 1661-1679, 2019.
- [18] G. Ren, J. Wan, J. Liu, and D. Yu, "Characterization of wind resource in China from a new perspective," *Energy*, vol. 167, pp. 994-1010, 2019.
- [19] G. Ren, J. Wan, J. Liu, and D. Yu, "Spatial and temporal assessments of complementarity for renewable energy resources in China," *Energy*, vol. 177, pp. 262-275, 2019.
- [20] T. R. Karl, W.-C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman, "A method of relating general circulation model simulated climate to the observed local climate. part I: seasonal statistics," *Journal of Climate*, vol. 3, no. 10, pp. 1053-1079, 1990.
- [21] D. J. Sailor, T. Hu, X. Li, and J. N. Rosen, "A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change," *Renewable Energy*, vol. 19, no. 3, pp. 359-378, 2000.
- [22] X. Dai, H. Shi, Y. Li, Z. Ouyang, and Z. Huo, "Artificial neural network models for estimating regional reference evapotranspiration based on climate factors," *Hydrological Processes*, vol. 23, no. 3, pp. 442-450, 2009.
- [23] F. Saberi-Movahed, M. Najafzadeh, and A. Mehrpooya, "Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions," *Water Resources Management*, vol. 34, no. 2, pp. 529-561, 2020.
- [24] Y. Jiang, S. Han, C. Shi, T. Gao, H. Zhen, and X. Liu, "Evaluation of HRCLDAS and ERA5 datasets for near-surface wind over hainan island and south China sea," *Atmosphere*, vol. 12, no. 6, p. 766, 2021.
- [25] M. O. Molina, C. Gutierrez, and E. Sanchez, "Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset," *International Journal of Climatology*, vol. 41, no. 10, pp. 4864-4878, 2021.
- [26] R. A. Merino and M. I. Gassmann, "Wind trends analysis in southern south America from weather station and reanalysis data," *International Journal of Climatology*, vol. 42, no. 4, pp. 2117-2134, 2021.
- [27] I. Gonzalez-Aparicio, F. Monforti, P. Volker et al., "Simulating European wind power generation applying statistical downscaling to reanalysis data," *Applied Energy*, vol. 199, pp. 155-168, 2017.
- [28] J. Ramon, L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, "What global reanalysis best represents near-surface winds?" *Quarterly Journal of the Royal Meteorological Society*, vol. 145, no. 724, pp. 3236-3251, 2019.
- [29] D. Li, H. V. Storch, and B. Geyer, "Testing reanalyses in constraining dynamical downscaling," *Journal of the Meteorological Society of Japan*, vol. 94A, pp. 47-68, 2016.