Heliyon 9 (2023) e23071

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

A widely applicable and robust LightGBM - Artificial neural network forecasting model for short-term wind power density

Xiangrui Zeng [a], Nibras Abdullah [a,b,*], Baixue Liang [a]

[a] School of Computer Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia
[b] Faculty of Computer Science and Engineering, Hodeidah University, Hodeidah, Yemen

ARTICLE INFO

Keywords:
Wind power density
Short-term
Next hour forecasting
LightGBM
Random forest
Artificial neural network

ABSTRACT

Wind energy is a clean and renewable source that reduces greenhouse gas emissions. To smooth the impact of wind energy fluctuations on the power grid and power supply, much research has predicted the wind speed of wind farms to estimate power generation. However, most studies overlook the nonlinear relationship between wind speed and power generation, and data sources are usually limited to one or two wind farms. This study constructs a wind power density prediction model based on the LightGBM and artificial neural network to solve the above problems. Its data collection process does not require meteorological measurement equipment and has good universality, stability, and robustness. LightGBM is used to extract feature information and then train it using an artificial neural network, which has a high tolerance for data loss. The model performance was validated using data from six terrains from 2020 to 2022. While results showed that the average prediction error was 71.68 % less than 2 % and 82.188 % less than 6 %, with an average R^2 of 0.9755 and an average correlation coefficient of 0.9875, proving the practical significance of the model that can be used to guide electricity trade.

1. Introduction

1.1. Background

As the population and economy grow, the energy demand follows. Simultaneously, the problems of environmental pollution and resource depletion have become increasingly severe, and the conventional energy crisis has become increasingly prominent. Within the LCA of a power station, coal-fired power plants emit the highest amount of greenhouse gases, followed by oil-fired and gas-fired. Instead, the intensity of greenhouse gas emissions from renewable energy power generation is generally low, with solar energy, biomass, hydropower, and wind energy ranking from high to low, indicating the development, utilization, and popularization of new energy have become inevitable. Wind energy is obtained from the wind by converting the kinetic energy into electrical energy through electromagnetic induction. For unit power generation, the greenhouse gas emissions from wind power are only about one-seventieth of that from coal power [1]. From the perspective of LCA, wind power has the lowest carbon and best environmental protection compared to other energy types. Its advantages include renewable, a small footprint for power generation equipment, a short energy payback period [2], less carbon dioxide emissions [3], and more.

Power generation is directly related to wind, but the wind will change with time. If the grid connects large-scale wind power and

* Corresponding author. School of Computer Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia.
E-mail address: nibras@usm.my (N. Abdullah).

https://doi.org/10.1016/j.heliyon.2023.e23071
Received 28 July 2023; Received in revised form 9 November 2023; Accepted 25 November 2023
Available online 1 December 2023

then encounters wind speed fluctuation violently, that will be harmful. Adding electricity storage to smooth out fluctuations is a choice but will incur additional operating costs, which lead to higher terminal electricity prices [4]. One solution to overcome this problem is wind forecasting, which sells or buys electricity to the market by knowing the peaks and valleys of electricity production ahead. Many deregulated markets liquidate every 5 min and settle every 30 min [5] so that wind forecasts over 30 min help to develop appropriate hedging and coordination measures. Therefore, this study attempts to predict the change in wind energy data in the next 1 h.

1.2. Related works

Numerous researchers have done a lot of research on forecasting in the field of wind energy. The main forecasting methods in such research in the past five years majority are four types: physics, statistics, AI, and hybrid.

Physical models such as NWP and its improved methods [6–9] could use meteorological data to predict the probability density of wind speed or wind energy change in the extensive range in the medium and long term (similar to weather forecast) with less performance fluctuation. They do not need to use historical data for training so that no cold start problem. However, physical models usually have a mass demand for computing resources and do not accurate for predicting small-scale and short-term wind energy-related information.

Both statistical and AI models use historical data to make predictions. Among them, statistical methods such as SARIMA [10] have already been used to predict the wind speed of the next hour. However, this type of linear model is tough to perform a good enough prediction (Max R^2 only 0.6619 [10]) for the wind speed information with randomness and nonlinearity [11].

Single AI models have been developed for short-term prediction yet, including the use of ELM [12], DBNGA [13], ANN [14,15], and Functional Networks [16], among others. Some scholars believe that a single model has weaknesses that not easy to extract complex relationships in sequences [17]. However, in the prediction records of WPD, M. Hossain et al. [18] and F. Rodríguez et al. [19] constructed a single prediction model using ANFIS and RNN algorithms, proving that a single model could perform well. Insufficiently, the two teams only divided the training and test set in the neural network algorithm cause the test set was used for validation many times during the training process. Simultaneously, the comparative analysis of pseudo-regression was also lacking.

In recent years, research hotspots have focused on hybrid models, but the performance of the filter merge statistical method [20], the filter merge AI method [21–25], the statistical method merge the AI method [26], and the multiple AI stacking methods [27–29] does not have obvious advantages compared with the single AI model. The best performing and most popular is to use the EMD-like [30–42] or SSA-like [43–45] or a combination of both data decomposition methods [46] merge AI methods to constitute a hybrid model. In these time-series decomposition models, researchers first decompose all data to obtain multiple components and then train a prediction model for each separately. Finally, the predictions of the original data are obtained by somehow stacking the results of all the sub-models. This method first decomposes the entire data and then divides the data set. When decomposing the original sequence, the decomposed values need to be calculated by adding future data that will cause data leakage even if one divides the dataset first.

1.3. Contributions

Based on the review and analysis of the above literature, the time-series decomposition type model lacks the potential for practical application due to the risk of data leakage. The single-AI and multi-AI model stacking are well-behaved and easy to use, so they deserve further study. Most previous studies have focused on wind speed or power generation in wind farms, but computing Equation (1) for wind turbine power generation shows a nonlinear relationship between wind speed and power generation, which means wind speed is less suitable as an indicator to represent power generation potential.

$$\begin{aligned} P &= D \cdot A\omega \\ D &= 1/T \cdot \int 0.5\rho v^3 dt \end{aligned} \tag{1}$$

Where D is the average WPD in the period in W/m^2, ρ is the average air density in the period (unit is kg/m^3), and v is the wind speed in T at time m/s. Further, A and ω are the swept area, and the efficiency, respectively.

In Equation (1), the efficiency of the generator ω and the swept area A are both constant, so the WPD (D) can linearly reflect the power generation capacity. It is more practical to predict such an indicator that is linearly related to power generation and not affected by random factors. However, there is little research on predicting objects by WPD [18,19], and past forecasts have not adequately addressed the model's performance in different regions. Therefore, this study will explore a suitable prediction model and attempt to improve its stability, universality, and robustness.

This study proposes a combined AI model named LG-ANN that exhibits high accuracy in predicting WPD, which combines LightGBM and ANN. This model utilizes the noise and multi-dimensional data handling ability of LightGBM, along with the learning ability of neural networks for complex nonlinear data. The specific steps of the study are to first collect historical meteorological data of any location in the public database as input, then use preprocessing methods that do not pose a risk of data leakage, and finally, use the data extracted by LightGBM to train the ANN model. Six metrics evaluated the performance showed it was superior to other single-AI models.

To summarize the contributions of this study.

(1) Improved the process and content of data preprocessing, avoiding data leakage issues caused by decomposition-related methods in previous studies.

(2) A novel method of utilizing historical data has been proposed, which significantly improves the performance of the predictive model.
(3) Improved feature extraction method, utilizing the ability of LightGBM to extract non-noise information from high dimensional data.
(4) The proposed LG-ANN combined model has robustness and universality in the practical application under the high accuracy premise.

2. Methodology

This section describes step-by-step the construction process of the predictive model, including the description and analysis of the data set used, the processing of the data, the introduction of algorithms used, and the method for performance measurement. The methodology of the entire model is built on the Python 3.9.

2.1. Data collection and analysis

Past research is often based on datasets in one or two regions and lacks concern about the scope of application of predictive models since this study collected data sets from six places for model training and validation to demonstrate that the model can be applied widely. The data were free collected from the NASA LaRC POWER Project, funded through the NASA Earth Science/Applied Science Program [47]. The six datasets come from regions with various terrain (Fig. 1), namely Demyansk (Hills), Flinders Island (Offshore), Baird (Plain), Caracas (Hilly), Zhegu Town (Plateau), and Hader Hudson (Basin). The WPD map is from the GWA 3.0 [48], a free, web-based application developed, owned, and operated by the DTU. The GWA 3.0 is released online in partnership with the World Bank Group. It utilizes data provided by Vortex and uses funding provided by the ESMAP.

All datasets contain hourly weather information such as wind speed, wind direction, and temperature in the region, from June 1, 2012 to June 1, 2022, with 6 × 87,672 pieces for each set (Table 1 displays their feature descriptions and statistics). The wind speed of all datasets did not exceed 30 m/s (cut-off speed) with no sudden changes, so no items were judged as outliers or noise. The crucial WPD information required is not contained in the metadata derivation is required.

2.2. Data preparation

2.2.1. Wind power density generation

Although the WPD feature is non in the meta-dataset, this information can be derived from wind speed and air density by Equation (1). This study uses the value of the "WS10 M″ feature as v, which is the predicted WPD at 10 m above the ground. Nevertheless, the dataset only contains surface pressure and 2-m-height temperature data, which should be converted using the vertical temperature lapse rate. However, the lapse rates vary widely by geography, season, day-night, and altitude [49]. Based on the consideration of not expanding the error, this study ignores the vertical temperature error and directly uses the surface pressure and 2-m-height temperature data to participate in the calculation.

Air density ρ could refer to Equation (2) to calculate.

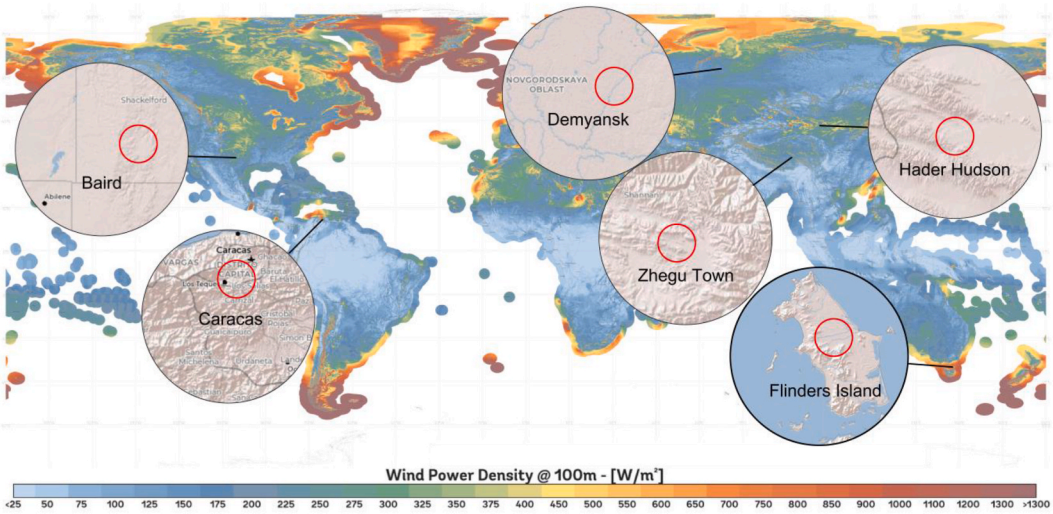

Fig. 1. Average annual wind power density potential.

Table 1
Feature description and statistics of six datasets.

Feature Name	DESC	Demyansk	Flinders Island	Baird	Caracas	Zhegu Town	Hader Hudson
		Scope	Scope	Scope	Scope	Scope	Scope
YEAR	–	2012–2022					
MO	Month	1–12					
DAY	–	1–31					
HR	Hour	1–24					
PS	Surface Pressure (kPa)	94.68–103.13	98.2–103.7	93.59–98.4	94.1–95.34	54.76–57.2	61.79–64.72
WS10 M	Wind Speed at 10 Meters (m/s)	0–8.07	0.01–24.47	0.02–15.18	0.02–5.59	0.01–16.64	0.01–19.8
WD10 M	Wind Direction at 10 Meters (°)	0–359.88	0–359.96	0–359.96	0–359.8	0–359.91	0–359.92
WS50 M	Wind Speed at 50 Meters (m/s)	0.02–13.7	0.04–28.75	0.02–19.1	0.02–8.81	0.02–19.22	0.02–23.83
WD50 M	Wind Direction at 50 Meters (°)	0–359.93	0–359.95	0–359.96	0–369.74	0–359.94	0–359.94
QV2M	Specific Humidity at 2 Meters (g/kg)	0.12–19.41	3.6–15.69	0.67–21.36	9.34–19.9	0.24–11.29	0.06–10.86
RH2M	Relative Humidity at 2 Meters (%)	34.31–100	37.5–100	4.19–100	30.88–100	3.75–100	1.31–100
T2M	Temperature at 2 Meters (°C)	−35.1–34.15	5.8–28.9	−16.81–43.87	17.58–33.76	−23.52–21.46	−24.89–27.18
T2MDEW	Dew/Frost Point at 2 Meters (°C)	−35.44–24.15	−0.35–21.18	−20.33–25.34	11.96–23.78	−33.71–7.2	−43.77–8.5
T2MWET	Wet Bulb Temperature at 2 Meters (°C)	−35.24–27.83	3.3–23.07	−18.57–29.73	16.9–26.27	−27.19–10.62	−32.44–13.63

$$\rho = \frac{1+d}{V_h} \tag{2}$$

Where d represents the moisture content in the air (how many grams of water vapor per kilogram of moist air) in kg and V_h represents the moisture-specific volume (i.e., the space occupied by 1 kg of dry gas together with the moisture it carries) in m^3. They could be calculated by Equation (3) and Equation (4).

$$d = \frac{P_v}{P - P_v} \tag{3}$$

$$V_h = \frac{R\widehat{T}}{P}\left(\frac{1}{M_g} \cdot \frac{d}{M_v}\right) \tag{4}$$

Where P_v is the partial pressure of water vapor (Equation (5)), P is the air pressure ("PS" feature in the dataset), both in KPa. R stands for the Molar gas constant, which is $8.314\ J/(mol \cdot K)$ in this study. $\widehat{T}$ is the Kelvin temperature in K ("T2M″ feature in the dataset). M_g represents the relative molecular mass of dry air, and M_v is the relative molecular mass of water vapor, which is 29 and 18 without units, respectively.

$$P_v = P_{ws} \cdot \varphi \tag{5}$$

Where P_{ws} is water vapor saturation pressure and φ represents the relative humidity of moist air ("RH2M″ feature in the dataset). P_{ws} was calculated in this study using the Hyland-Wexler formula [50], which includes the sub-zero (273.15 K) part of Equation (6) and the above-zero part of Equation (7).

$$\ln P_{ws} = C_1/T + C_2 + C_3 T + C_4 T^2 + C_5 T^3 + C_6 T^4 + C_7 \ln T \tag{6}$$

$$\ln P_{ws} = C_8/T + C_9 + C_{10} T + C_{11} T^2 + C_{12} T^3 + C_{13} \ln T \tag{7}$$

Table 2lists the specific values of $C_1 - C_{13}$.

Fig. 2 shows the distribution of WPD in the six datasets. The ideal cut-in speed for wind turbines is generally 3 m/s [51], and if estimated with an air density of 1.15 kg/m^3, the corresponding WPD is about 15 W/m^2. All datasets contain varying degrees of WPD below 15 W/m^2, but the main objective is to explore the predictive power. Hence all datasets still retain the data below this value. The Kruskal-Wallis H test [52] is suitable to determine whether there is a statistical difference between the data sets. The Null Hypothesis (H0) of the Kruskal-Wallis H test indicates no differences between datasets, while the Alternative Hypothesis (H1) indicates statistical

Table 2
Hyland-Wexler formula parameters.

C_1	C_2	C_3	C_4	C_5	C_6	C_7
−5.675E+03	6.392E+00	−9.678E-03	6.222E-07	2.075E-09	−9.484E-13	4.164E+00
C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	
−5.800E+03	1.392E+00	−4.864E-02	4.177E-05	−1.445E-08	6.546E+00	

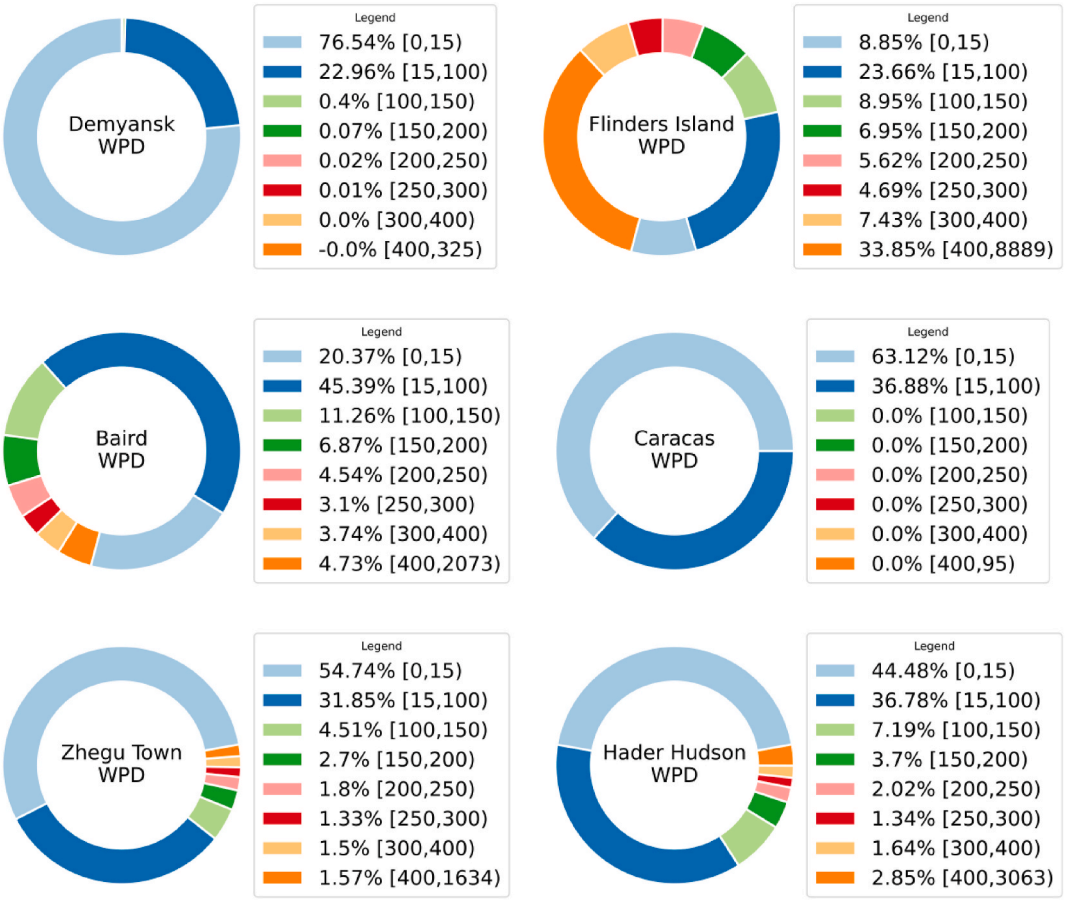

Fig. 2. Distribution of WPD in six points.

differences between datasets. The highest test P-value after comparing in pairs is 0, thus rejecting the H0.

2.2.2. *Target feature*

The purpose of this study is to forecast the next hour of WPD. Table 3 shows an example of target feature generation. The dataset will remove the last row After the feature offset (the last row has no future WPD).

2.2.3. *Historical data*

The prediction for a time point may rely on the information before the time point to improve the model performance, so the dataset needs the past WPD information as new features. Table 4 shows an example after adding the WPD of the first 3 h as a memory. This study will explore the appropriate memory time for each model in the 0–6 h range by cross-validation (CV = 5), and algorithm parameters will be combined to perform a Grid Search. After adding, the dataset needs to remove the first H rows since they do have no

Table 3
Example of target feature generation.

Original					
Feature Name	MONTH	DAY	HOUR		ENERGY (W/m^2)
Value	6	1	1		100
	6	1	2		110
	6	1	3		105
After one step shift					
Feature Name	**MONTH**	**DAY**	**HOUR**	**ENERGY (W/m^2)**	**TARGET (W/m^2)**
Value	6	1	1	100	**110**
	6	1	2	110	**105**
	6	1	3	105	–

Table 4
Example of adding 3 h of historical data.

After one step shift								
Feature Name	MONTH	DAY	HOUR			ENERGY (W/m^2)		TARGET (W/m^2)
Value	6	1	1			100		110
	6	1	2			110		105
	6	1	3			105		130
After one step shift with 3 h memory								
Feature Name	**MONTH**	**DAY**	**HOUR**	**ENERGY (W/m^2)**	**ENERGY-1 (W/m^2)**	**ENERGY-2 (W/m^2)**	**ENERGY-3 (W/m^2)**	**TARGET (W/m^2)**
Value	6	1	1	**100**	–	–	–	110
	6	1	2	**110**	–	–	–	105
	6	1	3	**105**	–	–	–	130
	6	1	4	130	**105**	**110**	**100**	200

enough previous WPD.

2.2.4. Feature analysis

Each dataset contains fourteen features (Table 1) that the Pearson coefficient is suited to judge the correlation with WPD, and the air density ρ generated before also joins the judgment process. The resulting range of the correlation coefficient is [−1,1], and the farther its value is from 0, the higher the correlation between the two features. Fig. 3 shows the analysis results of the six datasets.

The analysis results show that the importance of all features in different datasets is not consistent, but almost all have correlations greater than 0.01. All methods would use all attributes to retain as much information as possible without being discarded. Due to the target feature being a WPD at a height of 10 m, the WS50 M and WD50 M in the dataset will not be included even if they are highly correlated.

2.2.5. Model validation

The model will use the first 72 % of the data as a training set, the middle 8 % as a validation set, and the last 20 % (2020–2022) as the test set. The validation set aims to find suitable memory hours and splice parameters, while the test set has been used once only in the performance measurement process.

2.3. Forecasting methods

Simple algorithms may sometimes achieve adequate results [14,16]. This study uses the RFR to construct a single AI predictor, and the LSTM algorithm commonly used in previous studies has been added to the experiment process for comparative analysis. One combined model called LG-ANN is proposed as original content.

2.3.1. Random forest regression

RFR belongs to the Bagging algorithm of ensemble learning in AI. It uses the Bootstrap [53] strategy (that is, with replacement sampling) to divide the training set into multiple sub-datasets of the same size for training separately. And the average of all the results is the final prediction output. Each sub-dataset uses a CART regression tree to partition nodes and derive predictions. For example,

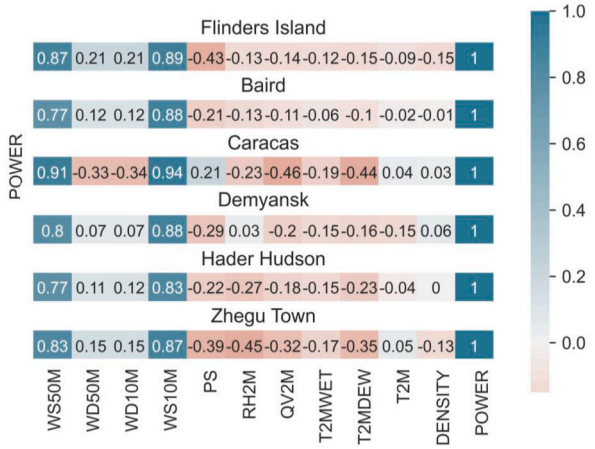

Fig. 3. Pearson correlation between non-temporal features.

when each sub-data set has M attributes, randomly select m attributes from the M attributes, satisfying the condition m « M. Then calculate all MSE (Equation (8)) values obtained by taking each feature as the dividing point, select the feature with the lowest MSE as the splitting feature of the node, and repeat this process until the entire sub-data set can no longer be a meaningful split [54]. The mean of all final nodes (leaf nodes) is the predicted value of the regression tree.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}|h(x_i) - y_i|^2 \tag{8}$$

Where $h(x_i)$ and y_i represent a single predicted and actual value, respectively.

This model adopts a two-step prediction strategy. The first step takes the WPD of the training set as the target feature, and the second step takes the residual of the WPD as the target. Both procedures take the same input, and then Equation (9) sums the results of each step as the final output.

$$result = first\_step + second\_step \tag{9}$$

This study only adjusted the number of regression trees for the RFR, which were searched along with memory hours through a Grid (CV = 5). The remaining parts use the default settings of scikit learn 1.3.0.

2.3.2. Long short-term memory neural network

The data used in this study are of the time series-related type, with significant contextual relationships between each record. RNN [55] is a feasible choice to take full advantage of this relationship, in which the mathematical expression of the traditional neural network is improved tiny from Equation (10) to Equation (11). However, RNN has a long-term dependency problem [56].

$$h_T = \sigma(x_T \times w_{x_T} + b) \tag{10}$$

$$h_T = \sigma(x_T \times w_{x_T} + h_{T-1} \times w_{h_T} + b) \tag{11}$$

LSTM is an improved version of RNN, first proposed by Hochreiter & Schmidhuber [57], which introduces a gate mechanism to control the circulation and loss of features. The motivation is to solve the long-term dependency problem.

LSTM is free to have any number of hidden layers and neurons. This study used the same neuronal structure and optimizer (Metric = MSE (Equation (8)) as in LG-ANN, with layers and memory hours searched for in Grid (CV = 5). Formula (9) is used to stack the first and second training results.

2.3.3. LightGBM-artificial neural network

The feature analysis section has determined that all features can be used to train the model, but the information importance in different datasets varies. Furthermore, the attributes used for model training may contain significant noise information, which may cause fewer benefits than the confusion provided. As a solution, this study uses nonlinear LightGBM to extract information and then ANN to train the model.

LightGBM recursively filters the best features of the dataset (CV = 5) and then flows the filtered features into ANN. The number of neurons in the input layer is feature number +1 (to reduce the influence of bias), and the output layer is 1. Neurons in the hidden layer are twice that of the input layer [58] (considering the small number of neurons in the input layer). MSE (Equation (8)) and Adam are combined as the optimizer, and all activation functions use ReLu. The Grid will be used to search the number of hidden layers and memory hours (CV = 5).

2.4. Performance metrics

For judging the fitting degree of the prediction model, this study uses MAE, NMAE, RMSE, NRMSE, R^2, and Correlation coefficient as performance indicators. RMSE (Equation (14)) can measure the deviation between predicted values and actual, especially sensitive to outliers in predicted values. The NRMSE (Equation (15)) could meter the errors between observed and predicted values, representing the normalized sample standard deviation of the difference between them. MAE (Equation (12)) can reflect the factual error between the forecasted value and the actual value and is a more natural mean error measure than the RMSE [59] indicator commonly used in climate and environmental literature. As an improvement, NMAE (Equation (13)) can avoid the situation of index soar growth caused by little errors caused by small individual target values. The smaller the value of the above, the better the model performance (minimum is 0). However, there is no upper limit for these indicators, and due to different scales, it is hard to use them for horizontal comparison between other research. R^2 (Equation (16)) and correlation coefficient (Equation (17)) reduce the regression performance evaluation results to [0,1], which provides conditions for researchers to compare the prediction results of different models to a certain extent.

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|h(x_i) - y_i| \tag{12}$$

$$NMAE = \frac{\sum_{i=1}^{m} |h(x_i) - y_i|}{\sum_{i=1}^{m} |y_i|} \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (h(x_i) - y_i)^2} \tag{14}$$

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} (h(x_i) - y_i)^2}}{y_{\max} - y_{\min}} \tag{15}$$

$$R^2 = 1 - \frac{\Sigma_{i=1}^{m} (y_i - h(x_i))^2}{\Sigma_{i=1}^{m} (y_i - \bar{y})^2} \tag{16}$$

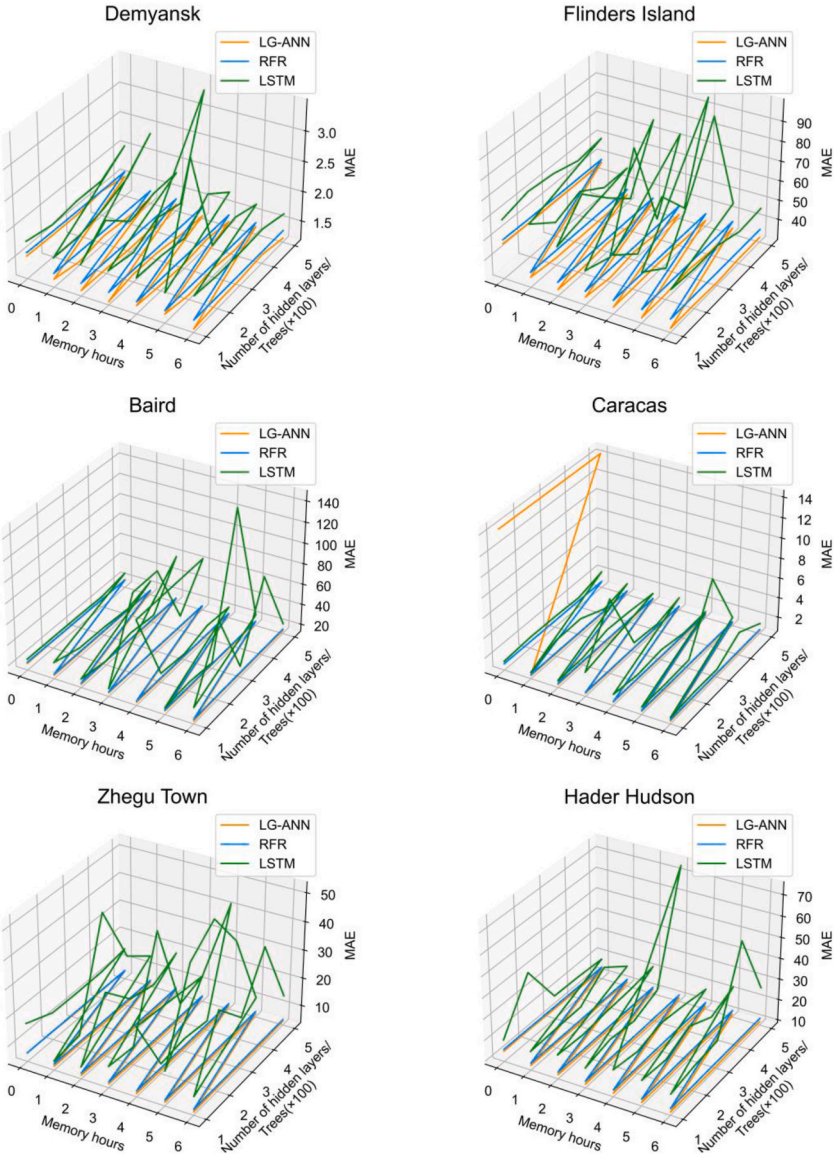

Fig. 4. Parameter search results of three algorithms on MAE

$$R = \frac{Cov(h(x), y)}{\sqrt{Var[h(x)]Var[y]}} \tag{17}$$

In the above equations, $h(x)$ and y represent the prediction set and actual value set, $Cov()$ and $Var()$ means covariance and variance.

3. Result analysis and discussion

Cross-validation and Grid search were executed to find the optimal memory hours, the number of regression trees in RFR, and the hidden layers number in LSTM and LG-ANN. Fig. 4 shows the search results on MAE loss, indicating that LSTM performance is unstable and the number of hidden layers or regression trees has no significant impact on performance. The crucial parameter is the number of memory hours since RFR and LG-ANN show remarkable performance improvement when greater than 1. However, the ascending is no longer significant when the memory hours are further increased.

Table 5 shows the best results obtained by three methods, while Table 6 details the performance of LG-ANN running ten times (Hidden layers = 1, Memory hours = 1). Overall, LG-ANN has the best performance and stability.

Pseudo-regression refers to omitting the prediction process of WPD, directly setting a time as the predicted value of the next hour. In this way, various indicators will perform well (Table 5 also shows the indicator data of Pseudo-regression), but the model is mean-ingless. Pseudo-regression can be distinguished well by visualizing the predicted-true curve. Fig. 5 illustrates the comparison between LG-ANN prediction and pseudo regression from May 25th to 31st (the last week of the dataset).

This study visualized the actual and predicted WPD values of LG-ANN from May 25 to 31 (the last week in the dataset) (Fig. 6). It shows that the forecast results have more errors at the trend change point of WPD, which one reason is the historical data may not reflect the trend change signal. However, the prediction performance is excellent in the stage where the WPD trend is unchanged, and the curve of the predicted value fits the actual curve well. In most cases, the prediction error is not significant, which also corresponds to the situation in Fig. 7, i.e., most deviations are at the point where the trend changes. In theory, the later date of the data in the test set, the higher the corresponding prediction errors should be. The predicted-true curves of Flinders Island in the first and last week of the test set in Figs. 5 and 6 confirm this conjecture.

The MAE values shown in Table 5 can reflect the average error between predicted and actual values but have the limitations of reference significance caused by different WPD ranges in various datasets. Fig. 7 further divided the error level (average error rate) into six hierarchies for statistical purposes, indicating that most of the errors in all data sets are within 2 % (an average of 71.68 % of the prediction errors are less than 2 %), and the predictions with an error of less than 6 % account for an average of 82.188 % of the total.

WPD data contains a considerable degree of random information, i.e., noise, which is almost unpredictable. Noise is also hard to extract from the data, but it can be well distinguished whether the data belongs to noise through some characteristics. The residual series could be fine-obtained by subtracting the model prediction from the actual WPD, and the residual series statistics for each

Table 5
Comparison of predictive performance.

Dataset	Model	Metrics					
		R^2	MAE	NMAE	NRMSE	RMSE	R
Demyansk	**RFR**	0.9524	1.3763	0.1167	0.0013	3.3544	0.976
Flinders Island		0.9792	35.96	0.083	0.00087	83.6888	0.9897
Baird		0.9259	17.366	0.1538	0.00166	42.24	0.9625
Caracas		0.976	0.955	0.0665	0.0006	1.474	0.9883
Zhegu Town		0.9714	5.76	0.1282	0.00112	13.96	0.9859
Hader Hudson		0.8847	11.8355	0.18777	0.00228	41.823	0.9414
AVERAGE		0.9483	12.2088	0.12266	0.0013	31.09	0.974
Demyansk	**LSTM**	0.9629	1.6636	0.14438	0.001186	3.1286	0.9813
Flinders Island		0.98	39.867	0.08588	0.00087	92.716	0.9887
Baird		0.954	17.335	0.1534	0.0013	33.255	0.9605
Caracas		0.973	1.1185	0.0843	0.00075	1.66595	0.9866
Zhegu Town		0.9745	6.9894	0.1517	0.00107	14.389	0.9876
Hader Hudson		0.923	17.633	0.2615	0.00187	37.394	0.9608
AVERAGE		0.9612	14.101	0.14686	0.00117	30.425	0.9776
Demyansk	**LG-ANN**	0.9786	1.2254	0.1039	0.00088	2.24848	0.9894
Flinders Island		0.9863	30.709	0.0709	0.0007	68.1137	0.9927
Baird		0.9547	15.83	0.14	0.0013	33.02	0.9776
Caracas		0.9844	0.838	0.058	0.0005	1.2	0.9915
Zhegu Town		0.9878	4.753	0.1058	0.0007	9.14	0.9935
Hader Hudson		0.961	9.977	0.158	0.0013	24.4341	0.9802
AVERAGE		0.9755	10.555	0.1061	0.000897	23.0267	0.9875
Demyansk	**Pseudo**	0.964	1.62	0.13734	0.00113	2.9075	0.11
Flinders Island		0.971	51/686	0.11935	0.00103	98.92	−0.0072
Baird		0.9089	23.204	0.2053	0.00184	46.813	0.1358
Caracas		0.937	1.674	0.1166	0.00105	2.415	0.2484
Zhegu Town		0.888	12.873	0.2864	0.00222	27.6	0.0077
Hader Hudson		0.8958	18.89	0.2997	0.00217	39.752	0.0384

Table 6
LG-ANN results from ten runs on each dataset.

Times	Metrics	Dataset					
		Demyansk	Flinders Island	Baird	Caracas	Zhegu Town	Hader Hudson
1	**RMSE**	2.25	70.85	33.32	1.26	9.757	25.675
	R^2	0.9786	0.985	0.9539	0.9829	0.985	0.9565
	MAE	1.2477	35.2	16.67	0.887	5.358	10.882
	R	0.9895	0.993	0.9768	0.9915	0.993	0.9787
2		2.2956	70.45	34.04	1.24	9.6714	24.4341
		0.9778	0.985	0.9519	0.9834	0.986	0.961
		1.2285	33.5	16.87	0.88	5.247	10.77
		0.9892	0.993	0.9756	0.9918	0.9932	0.9802
3		2.245	70.8	33.811	1.2578	9.55	26.21
		0.9787	0.985	0.9525	0.983	0.9866	0.955
		1.2558	34.88	16.798	0.879	5.189	10.789
		0.9893	0.993	0.9765	0.9915	0.9935	0.978
4		2.2357	70.19	34.587	1.27	9.604	26.034
		0.9765	0.985	0.95	0.98266	0.9865	−.9553
		1.377	33.46	16.858	0.9034	5.1063	11.1277
		0.9889	0.993	0.9749	0.9913	0.9934	0.9784
5		2.2456	69.87	33.56	1.257	0.976	25.8269
		0.9787	0.985	0.9532	0.983	0.986	0.956
		1.287	34.03	16.35	0.88	5.18	10.73
		0.9893	0.993	0.9764	0.9913	0.993	0.9791
6		2.279	70.73	32.72	1.28	9.5	24.828
		0.978	0.985	0.9555	0.9824	0.9868	0.9594
		1.283	34.92	16.016	0.9015	5.24	10.479
		0.9892	0.993	0.9776	0.9912	0.9934	0.9795
7		2.266	70.73	33.722	1.27	9.624	26.33
		0.9783	0.985	0.95275	0.9827	0.9864	0.9543
		1.29	34.92	16.348	0.8954	5.147	10.876
		0.9891	0.993	0.9763	0.9913	0.9932	0.978
8		2.313	70.624	33.281	1.269	9.45	25.429
		0.9774	0.985	0.95398	0.9827	0.9869	0.95737
		1.296	33.066	16.3451	0.88996	4.9688	10.7589
		0.9893	0.993	0.9772	0.9915	0.9934	0.9786
Times	Metrics	Dataset					
		Demyansk	Flinders Island	Baird	Caracas	Zhegu Town	Hader Hudson
9	**RMSE**	2.265	69.48	33.784	1.2878	9.599	25.34
	R^2	0.9783	0.986	0.95258	0.9822	0.986	0.958
	MAE	1.28	32.428	17.22	0.905	5.146	10.6599
	R	0.9894	0.993	0.9761	0.9913	0.9933	0.9791
10		2.286	70.33	33.43	1.288	9.68	24.6
		0.978	0.9853	0.953576	0.9822	0.986	0.96
		1.334	33.257	16.217	0.905	5.2	10.3
		0.989	0.993	0.9767	0.9913	0.9932	0.98

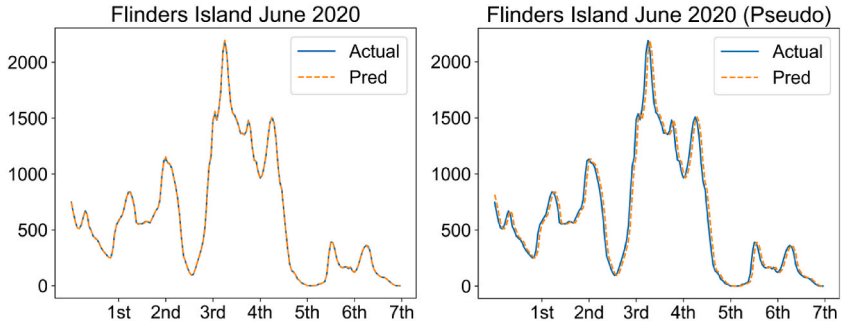

Fig. 5. Comparison of true prediction (left) and pseudo-regression (right).

dataset are shown in Fig. 8. The closer the residual sequence is to a normal distribution, the closer the mean is to 0, which indicates the more random the residual is, i.e., the better the model performance (the model could extract non-noise information well). According to Fig. 8, the mean of the absolute value of the errors calculated for all the data sets is 0.323, with no region having an average residual

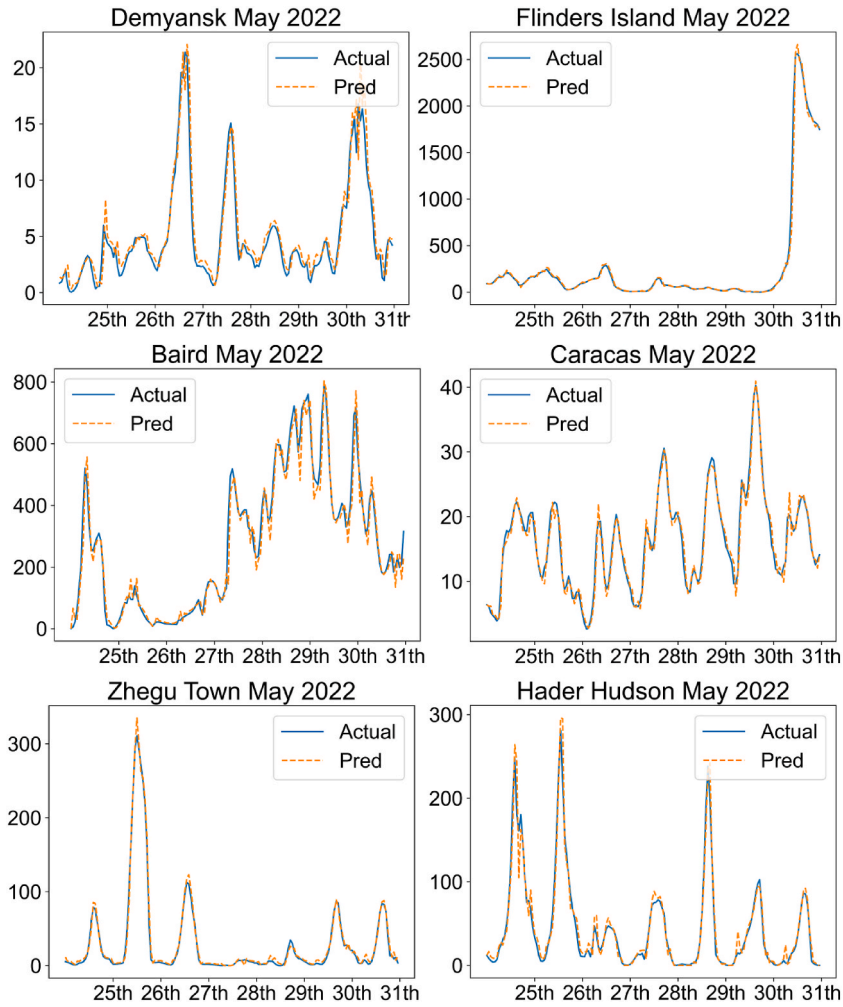

Fig. 6. Visualization of predicted and actual values for the last week of May 2020.

value greater than 1 W/m^2. Furthermore, all the residual series in the figure are in a normal distribution state, which indicates that the residual series is close to the noise state and the model has splendidly extracted meaningful data.

Compared with similar studies of the short-term type WPD prediction, the fitting degree of the curve in Fig. 6 is significantly better than that of F. Rodríguez et al. [19], and it also shows fewer pseudo-regression features. In their study, the prediction error bigger than 10 % (8 %) is smaller than in this study (12.35 %), but the errors less than 2 % (40 %→71.68 %) and 6 % (81 %→82.188 %) in this study have improved by 79.2 % and 1.47 %, respectively. A reduction in medium error indicates a remarkable improvement in low error prediction.

The model used in this study has more errors at trend change points and is best applicable to climate trends. Furthermore, the model usage requires retraining in different locations, and methods such as transfer training cannot be used to save time. The algorithm relies heavily on historical data, making it difficult to ensure accuracy without it.

4. Robust and external validation

Theoretically, the model is not sensitive to missing data as it does not require time series data. Randomly deleting a portion of the training set (Hidden layers = 1, Memory hours = 1 of LG-ANN) can verify this assumption. Fig. 9 shows the R^2 change of three datasets after random loss of training data, with 75 % representing the removal of the number of items in the training set to be equal to the test set. The experimental results show that when the number of training sets is three times larger than the test set, the lack of data has almost no effect on performance (i.e., deleting less than 25 % of the training set), and a ratio smaller than this may lead to a decrease in performance.

External validation is the utilization of not-used data to evaluate performance. Generally, it is divided into temporal validation (using data from the same source as the model development but different periods for validation) and Geographic validation (verifying the model's performance in other data sources), focusing on transportability and generalizability. This study used data on the location

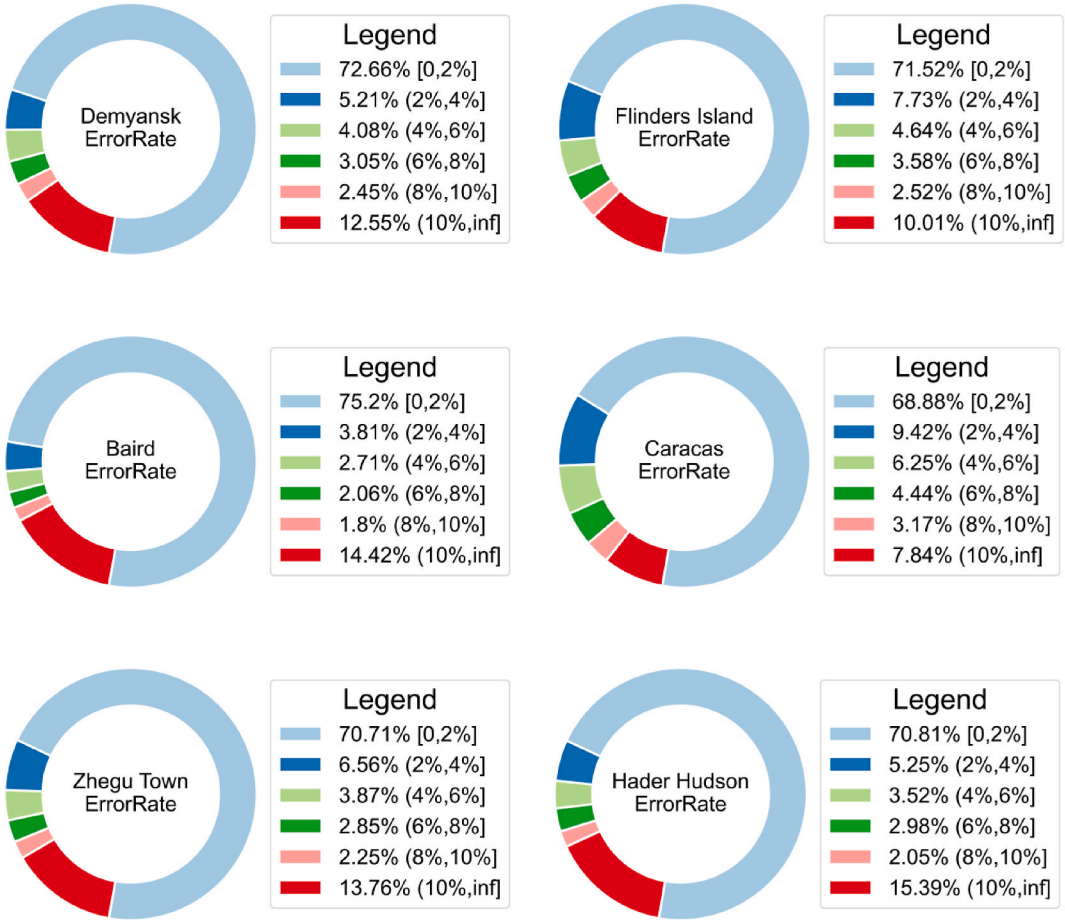

Fig. 7. Prediction error statistics for WPD.

of meteorological towers from Flatirons Campus for geometric validation. The instrument is on 39° 54′ 38.34″ N and 105° 14′ 5.28″ W [59]. The performance shown in Fig. 10 indicates that the model performs equally well on different datasets.

5. Conclusion

This study aims to discover and establish a model for predicting the average WPD at a height of 10 m in the next hour using meteorological data such as wind speed, pressure, and temperature. The proposed model constructed using the LighGBM and ANN algorithms achieved the research objectives, which can successfully predict any magnitude of WPD. This model will shrink accuracy in the case of sudden changes in WPD, which is the most common reason for error. However, the performance is excellent when WPD is in a trend state, with errors mostly less than 2 % at this time. The model does not need to be trained on the immediately preceding data, but the closer the training set is to the predicted time, the better its performance will be. Experiments in six regions, including Demyansk (Hills), Flinders Island (Offshore), Baird (Plain), Caracas (Hills), Zhegu Town (Plateau), and Hader Hudson (Basin), have shown that the model has a wide range of applications and can pass stability verification. Simultaneous, it exhibits robustness when the training set data is more than three times larger than the test set. Overall, the model found in this study extends the field of WPD prediction and helps power companies gain a better position in power dispatch and the electricity market.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

The data used to support the findings of this study are publicly available at https://power.larc.nasa.gov/data-access-viewer/and https://midcdmz.nrel.gov/apps/sitehome.pl?siteNWTC.

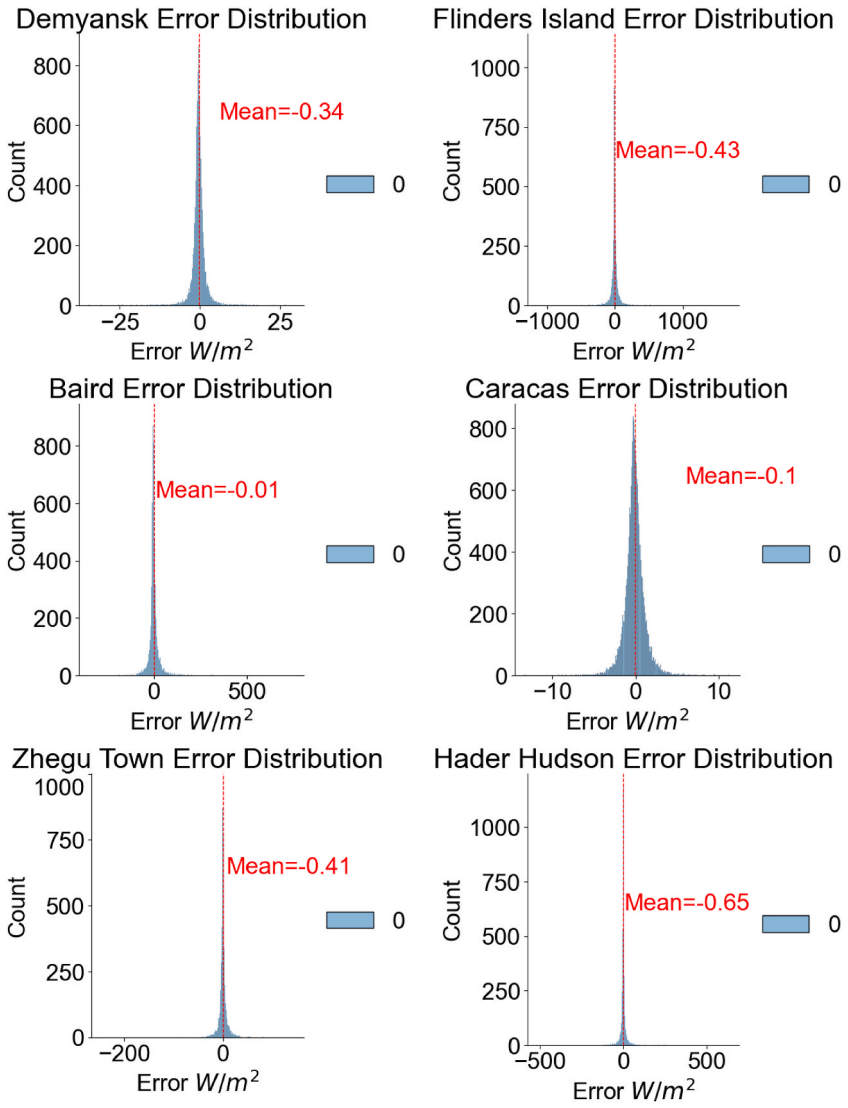

Fig. 8. Model residuals visualization.

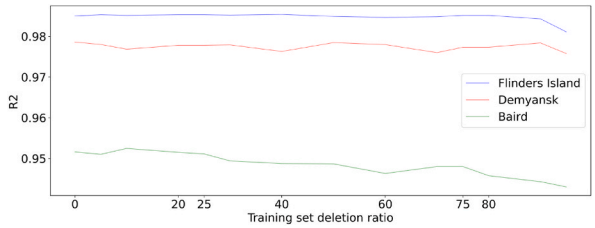

Fig. 9. Robustness testing through training set deletion.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Xiangrui Zeng: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project

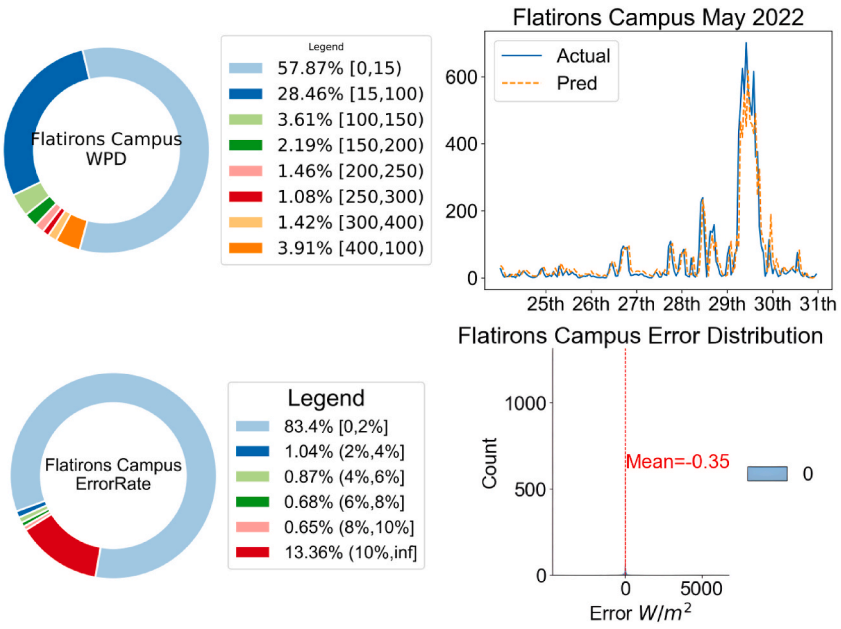

Fig. 10. External validation performance.

administration, Resources, Software, Validation, Visualization, Writing – original draft. **Nibras Abdullah:** Supervision. **Baixue Liang:** Writing – review & editing.

Declaration of competing interest

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

References

[1] United Nations Economic Commission for Europe, Carbon Neutrality in the UNECE Region: Integrated Life-Cycle Assessment of Electricity Sources, United Nations, 2022, https://doi.org/10.18356/9789210014854.
[2] B. Guezuraga, R. Zauner, W. Pölz, Life cycle assessment of two different 2 MW class wind turbines, Renew. Energy 37 (2012) 37–44, https://doi.org/10.1016/j.renene.2011.05.008.
[3] Y. Wang, T. Sun, Life cycle assessment of CO2 emissions from wind power plants: methodology and case studies, Renew. Energy 43 (2012) 30–36, https://doi.org/10.1016/j.renene.2011.12.017.
[4] A. Fabbri, T.G.S. Roman, J.R. Abbad, V.H.M. Quezada, Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market, IEEE Trans. Power Syst. 20 (2005) 1440–1446, https://doi.org/10.1109/TPWRS.2005.852148.
[5] S.S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium 2010, 2010, pp. 1–8, https://doi.org/10.1109/NAPS.2010.5619586.
[6] J.B. Olson, J.S. Kenyon, I. Djalalova, L. Bianco, D.D. Turner, Y. Pichugina, A. Choukulkar, M.D. Toy, J.M. Brown, W.M. Angevine, E. Akish, J.-W. Bao, P. Jimenez, B. Kosovic, K.A. Lundquist, C. Draxl, J.K. Lundquist, J. McCaa, K. McCaffrey, K. Lantz, C. Long, J. Wilczak, R. Banta, M. Marquis, S. Redfern, L. K. Berg, W. Shaw, J. Cline, Improving wind energy forecasting through numerical weather prediction model development, Bull. Am. Meteorol. Soc. 100 (2019) 2201–2220, https://doi.org/10.1175/BAMS-D-18-0040.1.
[7] X. Zhao, J. Liu, D. Yu, J. Chang, One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data, Energy Convers. Manag. 164 (2018) 560–569, https://doi.org/10.1016/j.enconman.2018.03.030.
[8] V. Hoolohan, A.S. Tomlin, T. Cockerill, Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data, Renew. Energy 126 (2018) 1043–1054, https://doi.org/10.1016/j.renene.2018.04.019.
[9] J. Zhao, J. Wang, Z. Guo, Y. Guo, W. Lin, Y. Lin, Multi-step wind speed forecasting based on numerical simulations and an optimized stochastic ensemble method, Appl. Energy 255 (2019), 113833, https://doi.org/10.1016/j.apenergy.2019.113833.
[10] X. Liu, Z. Lin, Z. Feng, Short-term offshore wind speed forecast by seasonal ARIMA - a comparison against GRU and LSTM, Energy 227 (2021), 120492, https://doi.org/10.1016/j.energy.2021.120492.
[11] X. Ma, Z. Liu, Y. Wang, Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China, J. Comput. Appl. Math. 347 (2019) 84–94, https://doi.org/10.1016/j.cam.2018.07.044.
[12] M. Qolipour, A. Mostafaeipour, M. Saidi-Mehrabad, H.R. Arabnia, Prediction of Wind Speed Using a New Grey-Extreme Learning Machine Hybrid Algorithm: A Case Study, vol. 30, Energy & Environment, 2019, pp. 44–62, https://doi.org/10.1177/0958305X18787258.
[13] K.-P. Lin, P.-F. Pai, Y.-J. Ting, Deep Belief networks with genetic algorithms in forecasting wind speed, IEEE Access 7 (2019) 99244–99253, https://doi.org/10.1109/ACCESS.2019.2929542.
[14] G. Nazaré, R. Castro, L.R.A. Gabriel Filho, Wind power forecast using neural networks: tuning with optimization techniques and error analysis, Wind Energy 23 (2020) 810–824, https://doi.org/10.1002/we.2460.
[15] D.T. Viet, V.V. Phuong, M.Q. Duong, Q.T. Tran, Models for short-term wind power forecasting based on improved artificial neural network using particle swarm optimization and genetic algorithms, Energies 13 (2020) 2873, https://doi.org/10.3390/en13112873.

[16] A. Ahmed, M. Khalid, An intelligent framework for short-term multi-step wind speed forecasting based on Functional Networks, Appl. Energy 225 (2018) 902–911, https://doi.org/10.1016/j.apenergy.2018.04.101.
[17] Y. Chen, Z. He, Z. Shang, C. Li, L. Li, M. Xu, A novel combined model based on echo state network for multi-step ahead wind speed forecasting: a case study of NREL, Energy Convers. Manag. 179 (2019) 13–29, https://doi.org/10.1016/j.enconman.2018.10.068.
[18] M. Hossain, S. Mekhilef, F. Afifi, L.M. Halabi, L. Olatomiwa, M. Seyedmahmoudian, B. Horan, A. Stojcevski, Application of the hybrid ANFIS models for long term wind power density prediction with extrapolation capability, PLoS One 13 (2018), e0193772, https://doi.org/10.1371/journal.pone.0193772.
[19] F. Rodríguez, A.M. Florez-Tapia, L. Fontán, A. Galarza, Very short-term wind power density forecasting through artificial neural networks for microgrid control, Renew. Energy 145 (2020) 1517–1527, https://doi.org/10.1016/j.renene.2019.07.067.
[20] S. Aasim, N. Singh, A. Mohapatra, Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting, Renew. Energy 136 (2019) 758–768, https://doi.org/10.1016/j.renene.2019.01.031.
[21] D.Y. Hong, T.Y. Ji, M.S. Li, Q.H. Wu, Ultra-short-term forecast of wind speed and wind power based on morphological high frequency filter and double similarity search algorithm, Int. J. Electr. Power Energy Syst. 104 (2019) 868–879, https://doi.org/10.1016/j.ijepes.2018.07.061.
[22] S. Pei, H. Qin, Z. Zhang, L. Yao, Y. Wang, C. Wang, Y. Liu, Z. Jiang, J. Zhou, T. Yi, Wind speed prediction method based on empirical wavelet transform and new cell update long short-term memory network, Energy Convers. Manag. 196 (2019) 779–792, https://doi.org/10.1016/j.enconman.2019.06.041.
[23] L. Xiang, Z. Deng, A. Hu, Forecasting short-term wind speed based on IEWT-LSSVM model optimized by bird swarm algorithm, IEEE Access 7 (2019) 59333–59345, https://doi.org/10.1109/ACCESS.2019.2914251.
[24] C. Yu, Y. Li, H. Xiang, M. Zhang, Data mining-assisted short-term wind speed forecasting by wavelet packet decomposition and Elman neural network, J. Wind Eng. Ind. Aerod. 175 (2018) 136–143, https://doi.org/10.1016/j.jweia.2018.01.020.
[25] Q. Zhou, C. Wang, G. Zhang, Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems, Appl. Energy 250 (2019) 1559–1580, https://doi.org/10.1016/j.apenergy.2019.05.016.
[26] D.B. Alencar, C.M. Affonso, R.C.L. Oliveira, J.C.R. Filho, Hybrid approach combining SARIMA and neural networks for multi-step ahead wind speed forecasting in Brazil, IEEE Access 6 (2018) 55986–55994, https://doi.org/10.1109/ACCESS.2018.2872720.
[27] E.H. Houssein, Particle swarm optimization-enhanced twin support vector regression for wind speed forecasting, J. Intell. Syst. 28 (2019) 905–914, https://doi.org/10.1515/jisys-2017-0378.
[28] M.-R. Chen, G.-Q. Zeng, K.-D. Lu, J. Weng, A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM, IEEE Internet Things J. 6 (2019) 6997–7010, https://doi.org/10.1109/JIOT.2019.2913176.
[29] Y. Chen, Y. Wang, Z. Dong, J. Su, Z. Han, D. Zhou, Y. Zhao, Y. Bao, 2-D regional short-term wind speed forecast based on CNN-LSTM deep learning model, Energy Convers. Manag. 244 (2021), 114451, https://doi.org/10.1016/j.enconman.2021.114451.
[30] Z. Tian, H. Li, F. Li, A combination forecasting model of wind speed based on decomposition, Energy Rep. 7 (2021) 1217–1233, https://doi.org/10.1016/j.egyr.2021.02.002.
[31] J. Wang, S. Wang, W. Yang, A novel non-linear combination system for short-term wind speed forecast, Renew. Energy 143 (2019) 1172–1192, https://doi.org/10.1016/j.renene.2019.04.154.
[32] J. Wang, Z. Yang, Ultra-short-term wind speed forecasting using an optimized artificial intelligence algorithm, Renew. Energy 171 (2021) 1418–1435, https://doi.org/10.1016/j.renene.2021.03.020.
[33] R. Li, Y. Jin, A wind speed interval prediction system based on multi-objective optimization for machine learning method, Appl. Energy 228 (2018) 2207–2220, https://doi.org/10.1016/j.apenergy.2018.07.032.
[34] C. Wu, J. Wang, X. Chen, P. Du, W. Yang, A novel hybrid system based on multi-objective optimization for wind speed forecasting, Renew. Energy 146 (2020) 149–165, https://doi.org/10.1016/j.renene.2019.04.157.
[35] L. Han, H. Jing, R. Zhang, Z. Gao, Wind power forecast based on improved Long Short Term Memory network, Energy 189 (2019), 116300, https://doi.org/10.1016/j.energy.2019.116300.
[36] H.H. Çevik, M. Çunkaş, K. Polat, A new multistage short-term wind power forecast model using decomposition and artificial intelligence methods, Phys. Stat. Mech. Appl. 534 (2019), 122177, https://doi.org/10.1016/j.physa.2019.122177.
[37] H. Hu, L. Wang, R. Tao, Wind speed forecasting based on variational mode decomposition and improved echo state network, Renew. Energy 164 (2021) 729–751, https://doi.org/10.1016/j.renene.2020.09.109.
[38] Z. Liu, P. Jiang, L. Zhang, X. Niu, A combined forecasting model for time series: application to short-term wind speed forecasting, Appl. Energy 259 (2020), 114137, https://doi.org/10.1016/j.apenergy.2019.114137.
[39] Z. Liu, R. Hara, H. Kita, Hybrid forecasting system based on data area division and deep learning neural network for short-term wind speed forecasting, Energy Convers. Manag. 238 (2021), 114136, https://doi.org/10.1016/j.enconman.2021.114136.
[40] W. Yang, J. Wang, H. Lu, T. Niu, P. Du, Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: a case study in China, J. Clean. Prod. 222 (2019) 942–959, https://doi.org/10.1016/j.jclepro.2019.03.036.
[41] W. Fu, K. Wang, J. Tan, K. Zhang, A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting, Energy Convers. Manag. 205 (2020), 112461, https://doi.org/10.1016/j.enconman.2019.112461.
[42] X.-J. Chen, J. Zhao, X.-Z. Jia, Z.-L. Li, Multi-step wind speed forecast based on sample clustering and an optimized hybrid system, Renew. Energy 165 (2021) 595–611, https://doi.org/10.1016/j.renene.2020.11.038.
[43] X. Yan, Y. Liu, Y. Xu, M. Jia, Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition, Energy Convers. Manag. 225 (2020), 113456, https://doi.org/10.1016/j.enconman.2020.113456.
[44] S.R. Moreno, L. dos Santos Coelho, Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro Fuzzy inference system, Renew. Energy 126 (2018) 736–754, https://doi.org/10.1016/j.renene.2017.11.089.
[45] X. Zhao, C. Wang, J. Su, J. Wang, Research and application based on the swarm intelligence algorithm and artificial intelligence for wind farm decision system, Renew. Energy 134 (2019) 681–697, https://doi.org/10.1016/j.renene.2018.11.061.
[46] S. Rodrigues Moreno, R. Gomes da Silva, V. Cocco Mariani, L. dos Santos Coelho, Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network, Energy Convers. Manag. 213 (2020), 112869, https://doi.org/10.1016/j.enconman.2020.112869.
[47] The Prediction of Worldwide Energy Resources (POWER) Project, 2022. https://power.larc.nasa.gov/docs/. (Accessed 23 August 2022).
[48] Global Wind Atlas. https://globalwindatlas.info/, 2022. (Accessed 23 August 2022).
[49] J.D. Lundquist, D.R. Cayan, Surface temperature patterns in complex terrain: daily variations and long-term change in the central Sierra Nevada, California, J. Geophys. Res. Atmos. 112 (2007), https://doi.org/10.1029/2006JD007561.
[50] Ashrae, ASHRAE Handbook – Fundamentals, 2021, Ashrae, 2021.
[51] P.R. Ebert, D.H. Wood, Observations of the starting behaviour of a small horizontalaxis wind turbine, Renew. Energy 12 (1997) 245–257, https://doi.org/10.1016/S0960-1481(97)00035-9.
[52] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis, Null 47 (1952) 583–621, https://doi.org/10.1080/01621459.1952.10483441.
[53] B. Efron, Bootstrap methods: another look at the jackknife, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer, New York, NY, 1992, pp. 569–593, https://doi.org/10.1007/978-1-4612-4380-9_41.
[54] J. Korstanje, The Decision Tree Model, Advanced Forecasting with Python, 2021, pp. 159–168, https://doi.org/10.1007/978-1-4842-7150-6_12.
[55] W. Zaremba, I. Sutskever, O. Vinyals, Recurrent Neural Network Regularization, 2015, https://doi.org/10.48550/arXiv.1409.2329.
[56] Y. Bengio, P. Frasconi, P. Simard, The Problem of Learning Long-Term Dependencies in Recurrent Networks, vol. 3, IEEE International Conference on Neural Networks, 1993, pp. 1183–1188, https://doi.org/10.1109/ICNN.1993.298725.

[57] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[58] M.J.A. Berry, G.S. Linoff, Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2004.
[59] D. Jager, A. Andreas, NREL National Wind Technology Center (NWTC): M2 Tower; Boulder, Colorado (Data), National Renewable Energy Lab. (NREL), Golden, CO (United States), 1996, https://doi.org/10.7799/1052222.