processes

MDPI

Article

Wind Speed Generation Method of Desert—Gobi—Wasteland Renewable Energy Base Based on Physical-Informed Neural Networks

Xinping Gao [1], Yuanzhi Li [1], Ling Hao [2,3,*], Xinhua Lei [1], Guixia Han [1], Fei Xu [2,3], Xiangyu Yan [4] and Lei Chen [2,3]

[1] Xinjiang Huadian Tianshan Power Generation Co., Ltd., Hami 839000, China; gxp_26@163.com (X.G.); lyz_2604@163.com (Y.L.); lxh_2622@163.com (X.L.); hgx_26@163.com (G.H.)
[2] State Key Laboratory of Power System Operation and Control, Tsinghua University, Haidian District, Beijing 100084, China; xufei@tsinghua.edu.cn (F.X.); chenlei08@tsinghua.edu.cn (L.C.)
[3] Department of Electrical Engineering and Applied Electronics Technology, Tsinghua University, Haidian District, Beijing 100084, China
[4] Huadian Electric Power Research Institute Co., Ltd., Hangzhou 310030, China; yxy_2604@163.com
* Correspondence: haolg@foxmail.com

Abstract

High spatial resolution wind speed data is very important for wind farm planning, design, operation and maintenance. But due to cost, site and other factors, it is impossible to build a large number of anemometer towers to obtain high spatial resolution measured data. Therefore, this paper proposes a method for generating wind speed data in renewable energy bases based on physics-informed neural networks, which incorporates fluid mechanics control equations such as the Navier—Stokes equation as physical constraints into the model training process. The model's input includes the wind speed data and the wind direction data of the anemometer towers as input, as well as the geographical difference data between the input anemometer towers and the output point, enabling to learn the mapping relationship between geographical differences and wind speed differences at different locations, achieving the goal of generating high spatial resolution wind speed data. Using normalized root mean absolute error (NMAE) to measure the model error, the average wind speed error and the average wind direction error of the proposed wind speed data generation method on different test sets are 8.28% and 10.50%, which is lower than that of BP neural network and graph convolutional neural network, and can provide more refined data support for wind turbine layout planning and wind farm power prediction of renewable energy bases.

Keywords: physics-informed neural networks; desert—gobi—wasteland renewable energy base; wind speed data generation; deep learning

Check for updates

Academic Editor: Krzysztof Rogowski

Received: 22 May 2026
Revised: 14 June 2026
Accepted: 17 June 2026
Published: 25 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In the planning, design, operation and maintenance of wind farms, wind speed data that accurately reflects the wind resource spatiotemporal distribution characteristics of the wind farm is of vital importance for issues such as the assessment of wind resource potential [1–3], the site selection of wind farms [4–6], and the optimization of the layout of anemometer towers and wind turbines [7–9]. It is also closely related to wind speed and power prediction, as well as the dispatching strategies [10,11] and economic benefits [12] of wind farms. Before a wind farm is put into operation, wind speed data is usually only obtained from anemometer towers. Conventional wind farms have a small range and

 https://doi.org/10.3390/pr14132058

strong consistency in terrain and landforms. When the terrain is relatively simple, a few anemometer towers can accurately reflect the wind resource situation of the wind farm. However, the installed capacity of renewable energy bases is usually in the tens of millions of kilowatts. They have a large number of wind turbines, occupy a large area, and the terrain is difficult to maintain consistency. In some renewable energy bases, there is also a significant altitude difference, resulting in poor uniformity of wind speed distribution within large bases.

People often plant grass or trees in desert, gobi and wasteland to prevent sand-dust storm, resulting in the terrain within the desert−gobi−wasteland renewable energy base being more complex and wind speed distribution more inconsistent. Due to the uneven wind speed distribution within the desert−gobi−wasteland renewable energy base, a few anemometer towers are unable to accurately reflect the distribution of wind resources in the renewable energy base. The insufficient representativeness of the measured data from the anemometer towers will directly affect the rationality of the wind turbine site selection and indirectly affect the accuracy of power prediction and the rationality of the dispatching strategy formulation in subsequent operations, thereby reducing the economic benefits of the large base. Therefore, it is necessary to adopt reasonable methods to accurately obtain the spatiotemporal distribution of wind resources in renewable energy bases.

The main ways to obtain wind speed data include reanalysis datasets, numerical simulation and on-site measurement. The reanalysis dataset covers a wide spatiotemporal range, is easy to obtain and convenient to access, and is widely used in wind resource assessment [13,14] and wind farm site selection [15], but the spatiotemporal resolution and data accuracy of the data are relatively low. On-site measurement can obtain the most accurate wind speed data by deploying anemometer towers or lidars. But measurement equipment is usually only installed at a limited number of points in the wind farm due to cost and space; for the desert−gobi−wasteland renewable energy base with a wide coverage area and diverse terrain and landforms, data from a small number of measurement devices cannot accurately describe the distribution of wind resources. Numerical simulation, including mesoscale weather research and forecasting (WRF) and computational fluid dynamics (CFD) simulation, is widely used for wind farm siting and layout optimization [16–19]. However, it suffers from high computational complexity, high computational cost, high time consumption, and limited spatial coverage. Hu et al. [17] scaled down a 5.5 km × 14.5 km CFD domain to 28 m × 37 m, completed 16 CFD simulations of different wind directions and corrected the results using the measured data. Ma et al. [18] used a domain of only 5 km × 5 km. Cuevas-Figueroa et al. [19] built a model with five nested domains (78 × 78 grids each; resolution from 32.4 km to 400 m) and stated that a one-week simulation required 128 processors and about 36 h.

In addition to the above three data sources, measure−correlate−predict (MCP) methods can also be used to generate wind speed data for a target point by training a neural network with short-term measured data from that point and long-term data from nearby towers or stations, where the network uses the long-term data as input to predict the target's wind speed. For instance, Xing et al. [20] used measurement data of the same anemometer tower but different heights or different time and mesoscale data, Qin et al. [21] used data from nearby anemometer towers and meteorological stations, Carta et al. [22] used reanalysis data, and Jia et al. [23] used data from nearby wind turbines. Nevertheless, these methods can only generate wind speed for locations with historical measurements (e.g., existing towers or turbines) and cannot accurately reflect the wind resource distribution across the entire wind farm. A summary of all the above methods is provided in Table 1.

Table 1. Summary of representative methods for wind speed data acquisition and generation.

Category	Reference	Data Source/Method	Key Limitations
Reanalysis data	[2,13–15]	Reanalysis datasets	Low spatiotemporal resolution, limited accuracy
On-site measurement	[1–4]	Anemometer towers, lidars	Sparse coverage due to cost and space
Numerical simulation	[16–18]	WRF or CFD simulation	High time and computational cost, small coverage area for CFD
Neural network	[20–23]	Data from nearby towers, meteorological stations, or reanalysis datasets	Requires historical data at target location

Because of the limitations of the aforementioned methods, there is currently no effective way to generate wind speed data for points without historical wind measurement data. Therefore, this paper proposes a wind speed data generation model for wind farms based on a physics-informed neural network (PINN), providing a promising alternative for obtaining wind speed data. The model considers the terrain differences between input and output objects and the constraints of the actual physical system, incorporating the continuity equation, Navier−Stokes equation and other fluid mechanics control equations as physical constraints into the model training process. Meanwhile, the terrain differences between the input and output points are also used as model input data, enabling the model to learn the mapping rules between wind speed variations and terrain differences, and achieving the goal of generating wind speed data at any point in the wind farm using wind speed data from the anemometer tower.

2. Data Source and Preliminary Analysis

This paper selected a renewable energy base in a desert area in northwest China as the research object. The base covers an area of 989.55 km^2 and has twenty anemometer towers and 404 wind turbines. The distribution of anemometer towers and wind turbines is shown in Figure 1. Unlike conventional wind farms, large bases occupy a vast area and have a huge installed capacity. It is necessary to conduct a preliminary analysis of the terrain of the large base and the representativeness of the anemometer towers to determine whether the existing anemometer towers can accurately reflect the distribution of wind resources in the large base.

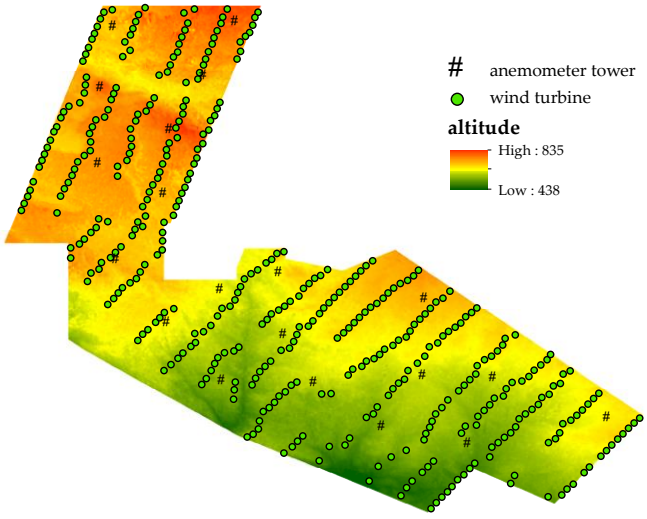
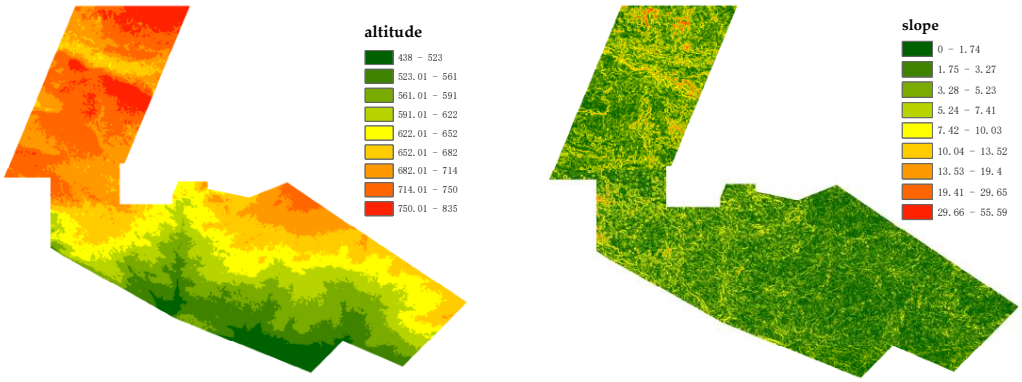


Figure 1. Distribution map of wind towers and wind turbines.

2.1. Analysis of the Renewable Energy Base Terrain

The altitude range of the large base is 438–835 m. The altitude and slope distribution map was drawn using the ArcMap software, as shown in Figure 2. It can be seen that the large base as a whole shows a trend of being higher in the north and lower in the south. The altitude in the east rises in a stepwise manner with a relatively gentle slope. The altitude change in the middle is still relatively regular, but the rate of change accelerates and the slope increases near the western boundary. The altitude change in the north is relatively complex, with uneven terrain and more obvious highlands and depressions within the area, and some areas have a relatively large slope.



(**a**) (**b**)

Figure 2. (**a**) Altitude distribution map of the bases; (**b**) slope distribution map of the bases.

Within the large base area, the altitude difference is 397 m. The differences in terrain among various regions lead to uneven wind speed distribution. Whether the existing anemometer towers can accurately reflect the wind speed changes in the large base directly affects the construction and operation effect of the large base wind farm.

2.2. Representative Analysis of Anemometer Towers

There are a total of twenty anemometer towers in the large base, among which the lowest altitude is 561 m and the highest is 780 m, which is quite different from the altitude range of the large base. The lowest-altitude anemometer tower is 123 m higher than the lowest point of the large base, and the highest-altitude anemometer tower is 55 m lower than the highest point of the large base. The altitude range of the 404 wind turbines in the large base is 476 to 810 m. The lowest altitude of the wind turbine is 85 m lower than that of the anemometer tower, and the highest altitude of the wind turbine is 30 m higher than that of the anemometer tower.

According to the requirements listed in the industry standard NB/T 31147-2018 Technical Specification for Wind Energy Resources Measurement and Evaluation of Wind Farm Engineering [24], the altitude difference between the anemometer tower and the wind turbine should not exceed 50 m. The altitude difference between each wind turbine and its nearest anemometer tower was calculated, and the results are shown in Figure 3. There are 12 anemometer towers with an altitude difference of more than 50 m from the surrounding 42 wind turbines. Among them, the height difference between 3 wind turbines and the nearest anemometer tower is greater than 100 m, and the height difference between 12 wind turbines and the nearest anemometer tower is between 70 and 100 m, far exceeding the 50 m range stipulated in the standard. From the perspective of altitude distribution, anemometer towers cannot accurately reflect the wind speed conditions at large bases, nor can they accurately reflect the wind speed conditions at the locations of wind turbines.

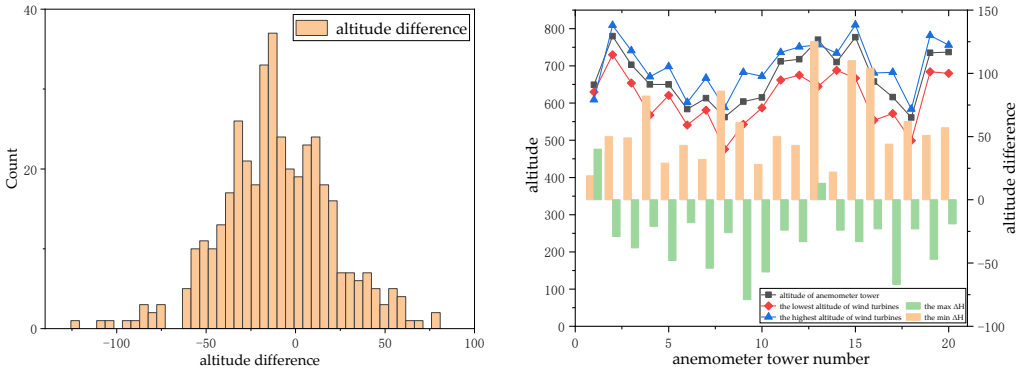


Figure 3. (**a**) Histogram of the distribution of altitude difference between the anemometer tower and the nearest wind turbine; (**b**) statistical results of the maximum and minimum height difference between the anemometer tower and the nearest wind turbine. ΔH is the altitude difference between the anemometer and the wind turbine.

Further divide the large base into regular grids, calculate the average slope and altitude standard deviation of all grids, and analyze from the terrain perspective whether the anemometer tower can accurately reflect the wind speed of the large base and the wind turbines. Slope can reflect the regular changes in terrain, and the standard deviation of altitude can reflect the intensity of terrain undulations. The combination of the two can comprehensively reflect the terrain conditions. The large base area was divided into 1 km grids, and a total of 1112 grids were obtained. The average slope and altitude standard deviation of each grid were calculated, and the results are shown in Figure 4.

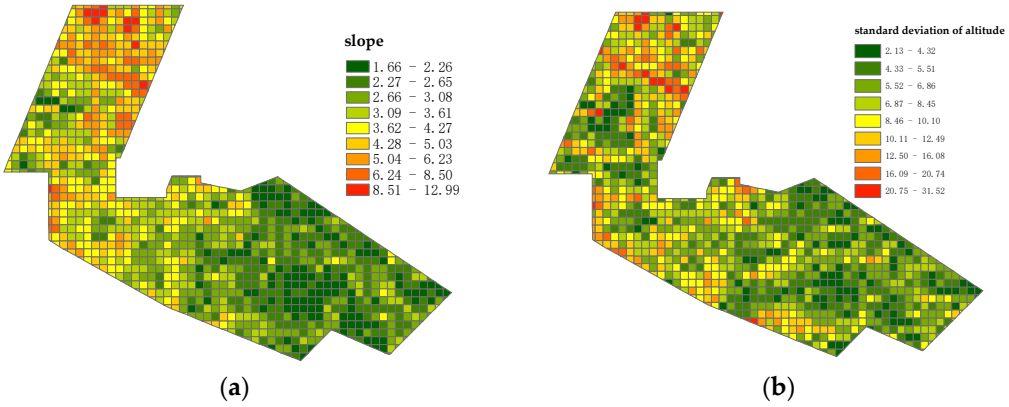


Figure 4. (**a**) Distribution map of 1 km grid elevation standard deviation of the base; (**b**) distribution map of 1 km grid average slope of the base.

The average slope of all grids ranges from 1.66 to 12.99°. The minimum average slope of the grid where the anemometer tower is located is 1.98°, which is basically the same as the minimum slopes of other grids. The maximum slope of the grid where the anemometer tower is located is 5.49°, which is much smaller than the maximum slopes of other grids. The standard deviation of altitude for all grids ranges from 2.13 to 31.52 m. The minimum standard deviation of altitude for the grid where the anemometer tower is located is 3.95 m, slightly larger than that of other grids. The maximum standard deviation of altitude for the grid where the anemometer tower is located is 16.30 m, much smaller than that of 31.52 m for other grids. It can be seen that the terrain of the area where the anemometer tower is located has certain differences from that of other areas of the large base. Therefore, the wind speed measured by the anemometer tower is difficult to represent the wind

resource situation of the large base, especially that of areas with more intense undulations and more complex terrain.

Based on the above analysis results, it can be seen that for large sand and desert bases with extensive coverage and complex and changeable terrain, a few anemometer towers are difficult to accurately reflect the distribution of wind resources in large bases. Using wind speed data measured by a few anemometer towers for the planning and prediction of wind farms will inevitably lead to inaccurate estimation of wind resources and deviations in wind power prediction, which will seriously affect the operational efficiency of wind farms. It is necessary to have a more accurate perception and assessment of the wind resource situation of wind farms.

Therefore, this paper proposed a wind speed generation model based on physics-informed neural networks. It uses a few anemometer towers to measure wind speed. The model considers the geographical difference data between input and output points, constructs a mapping model of wind speed at input points and output points, and uses the wind speed of the anemometer towers to generate wind speed at any point in the wind farm, achieving precise perception of the distribution of wind resources in the wind farm, helping wind farms plan the location of anemometer towers and wind turbines more reasonably, predicting power more accurately and carry out operation planning.

3. Model Building

The input of the model proposed in this paper includes the wind speed and direction data of the four anemometer towers that are closest to the output anemometer tower, as well as the geographical difference data between these four anemometer towers and the output anemometer tower. The geographical difference refers to the distance, height difference, slope, slope direction and the angle of wind direction between two points. The output includes the wind speed and direction data of the target anemometer tower, as shown in Equations (1)–(4).

$$X^t = \left[X_1^t, X_2^t, X_3^t, X_4^t\right] \tag{1}$$

$$X_i^t = \left[v_i^t, \theta_i^t, L_{i-j}, h_{i-j}, \alpha_{i-j}, \theta_{i-j}^t\right] \tag{2}$$

$$Y^t = \left[v_j^t, \theta_j^t\right] \tag{3}$$

$$Y^t = f(X^t) \tag{4}$$

where i is the index of the input anemometer tower; j is the index of the output anemometer tower. v_i^t and v_j^t are the wind speeds at time t for the input and output anemometer towers respectively. θ_i^t and θ_j^t are the wind directions at time t for the input and output anemometer towers respectively. L_{i-j}, h_{i-j} and α_{i-j} are the distances, altitude differences and slopes between the input and output anemometer towers respectively. θ_{i-j}^t is the difference between the wind speed at the input anemometer tower and the slope direction between the input anemometer tower and the output anemometer tower. For each output Y^t, there is a corresponding input X^t, which consists of four sets of data, respectively coming from the four anemometer towers that are closest to the output anemometer towers.

Through training, the model learns the mapping relationship between the input wind speed and direction, the geographical differences between the input and output points, and the differences in wind speed and direction among different points, thereby achieving the goal of generating wind speed and direction data for points without wind measurement using limited data from anemometer towers. However, since the training data only comes from a limited number of anemometer towers within the large base, the geographical difference data that can be provided is not much. Taking the twenty anemometer tow-

ers within the large base used in this paper as an example, if any two are paired, only 190 pairs of geographical difference data can be generated. For conventional neural network models, 190 pairs of geographical difference data cannot enable the model to learn well the correlation between wind speed differences and geographical differences, resulting in a poor training effect of the model and large errors in the generated data.

Physics-informed neural networks can solve this problem. The physics-informed neural network was proposed by M. Raissi et al. in 2019 [25]. The initial purpose of proposing this model was to use deep learning algorithms to solve partial differential equations, which was accomplished by constructing a neural network model and treating the partial differential equations to be solved as part of the model's loss function, thereby enabling to rapidly solve such equations. Most physical models are described by partial differential equations, and the proposal of physics-informed neural networks allows this prior physical knowledge—expressed through partial differential equations—to be integrated into the training process of neural network models.

In conventional data-driven models, the model relies entirely on training data to learn the mapping relationship from inputs to outputs. When there is too little training data, for example, in this paper, there are only 190 pairs of geographical difference data, traditional data-driven models are prone to overfitting or fail to learn all possible mapping relationships. However, in the PINN model, the prior physical knowledge is added to the model training process in the form of a part of the loss function, forcing the model to generate solutions that basically satisfy the physical constraints, reducing the degrees of freedom of the model output. Even when the training data is small, more reliable solutions can be generated relying on physical law constraints. With less training data, the same or even better training results than those of the original models can be achieved, and for this reason, the approach is widely applied across multiple fields, including biochemistry [26], medicine [27], architecture [28], and geographic information [29].

At present, there is already a relatively complete physical theory to describe wind. This theory is the governing equation of fluid mechanics and also the core theory of CFD simulation. In this paper, a physics-informed neural network is utilized to incorporate the fluid mechanics control equation as part of the loss function into the model training process. A mapping model between the geographical difference data and the wind speed difference data of the anemometer towers is trained using limited wind speed and geographic data of the anemometer towers, achieving accurate generation of wind speed at unknown points. The model structure is shown in Figure 5.

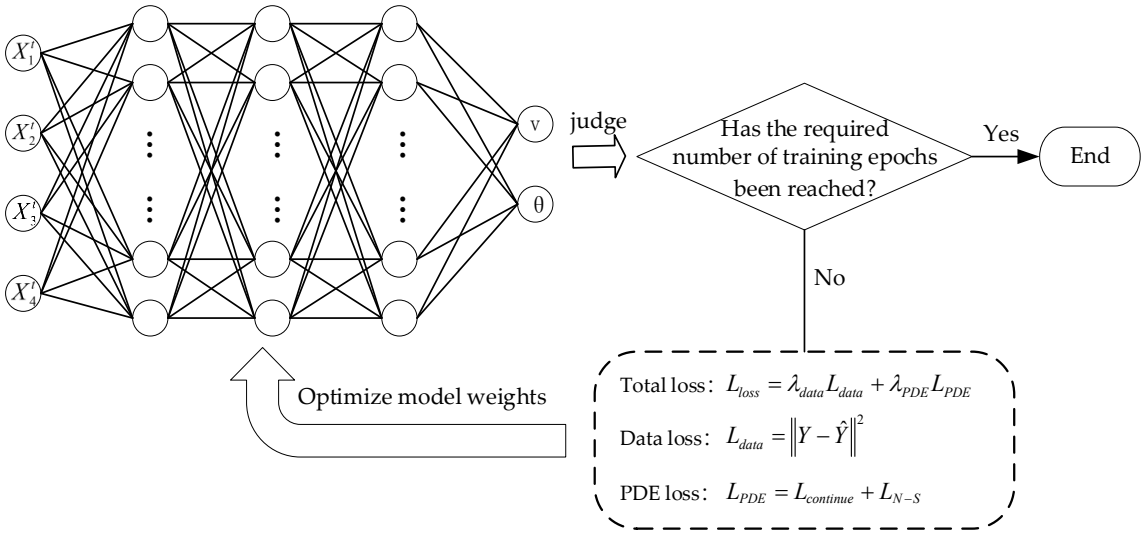


Figure 5. Wind speed generation model based on physics-informed neural network.

4. Construction of the Loss Function

The characteristic of PINN is integrating the constraints of the physical system into the training process of the neural network, making the training of the model no longer solely rely on known data. The core is the construction of the loss function. This section will elaborate on the form of the loss function adopted by the model, the derivation of physical constraints, and the calculation of the loss function.

4.1. Loss Function Form

The loss function of PINN generally consists of two parts: data loss and physical information loss, as shown in Equations (5)–(8).

$$L_{loss} = \lambda_{data} L_{data} + \lambda_{PDE} L_{PDE} \tag{5}$$

$$L_{data} = \left\| Y - \hat{Y} \right\|^2 \tag{6}$$

$$L_{PDE} = [f(y_1, y_2, \cdots, y_n, \partial u/\partial y_1, \cdots, \partial^{\alpha} u/(\partial^{\alpha_1} y_1 \partial^{\alpha_2} y_2 \cdots \partial^{\alpha_n} y_n))]^2 \tag{7}$$

$$f(y_1, y_2, \cdots, y_n, \partial u/\partial y_1, \cdots, \partial^{\alpha} u/(\partial^{\alpha_1} y_1 \partial^{\alpha_2} y_2 \cdots \partial^{\alpha_n} y_n)) = 0 \tag{8}$$

where L_{data} and L_{PDE} respectively represent the data loss and the physical information loss. λ_{data} and λ_{PDE} respectively represent the weights of the data loss and the physical information loss. Equation (8) is the general form of the partial differential equation. In this model, it should be correspondingly replaced with the core equation describing the change in wind speed, which is also an important step in integrating physical laws into the training of neural networks.

4.2. Physical Constraint Conditions

At present, there is a relatively complete fluid mechanics theory to describe the movement of wind [30]. The governing equations describing fluid motion include the mass conservation equation (also known as the continuity equation), the momentum conservation equation, and the energy conservation equation, as shown in Equations (9)–(11).

$$\partial \rho / \partial t + \nabla \cdot \left(\rho \vec{V} \right) = 0 \tag{9}$$

$$\rho D\vec{V}/Dt = \rho \vec{f} + \nabla \cdot \vec{P} \tag{10}$$

$$\rho D\left(e + \frac{v^2}{2} \right)/Dt = \rho q + \nabla \cdot (k \nabla T) + \rho \vec{f} \cdot \vec{V} + \nabla \cdot \left(\vec{P} \cdot \vec{V} \right) \tag{11}$$

It is generally believed that wind is a Newtonian fluid and its dynamic viscosity is constant. At this time, the momentum conservation equation can be transformed into the Navier−Stokes equation (referred to as the N-S equation), as shown in Equation (12).

$$\rho D\vec{V}/Dt = \rho \vec{f} + \mu \nabla^2 \vec{V} + \frac{\mu}{3} \nabla \left(\nabla \cdot \vec{V} \right) \tag{12}$$

Although strictly speaking, any fluid is compressible, to simplify the problem, fluids with very low compressibility are usually approximately regarded as incompressible fluids. Generally, low-speed gas flows with a velocity less than 100 m/s are regarded as incompressible flows. That is, for the movement of wind in a wind farm, it can be considered

that the wind is an incompressible fluid. Under these conditions, $\nabla \cdot \vec{V} = 0$ and the N-S equation can be simplified to the form shown in Equation (13).

$$\rho D\vec{V}/Dt = \rho\vec{f} - \nabla p + \mu\nabla^2\vec{V} \tag{13}$$

After expanding the N-S equation on the rectangular coordinate system, it is shown in Equations (14)–(16).

$$\rho Du/Dt = -\partial p/\partial x + \mu\left(\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 + \partial^2 u/\partial z^2\right) + \rho f_x \tag{14}$$

$$\rho Dv/Dt = -\partial p/\partial y + \mu\left(\partial^2 v/\partial x^2 + \partial^2 v/\partial y^2 + \partial^2 v/\partial z^2\right) + \rho f_y \tag{15}$$

$$\rho Dw/Dt = -\partial p/\partial z + \mu\left(\partial^2 w/\partial x^2 + \partial^2 w/\partial y^2 + \partial^2 w/\partial z^2\right) + \rho f_z \tag{16}$$

where f_x, f_y and f_z are the external forces in the x, y, and z directions respectively. For the scenario studied in this paper, the only external force is gravity, which acts on the Z-axis. Since the measured wind speed of the anemometer tower only shows the wind speed in the horizontal direction, after decomposition, only the wind speed in the X-axis and Y-axis can be obtained. Moreover, this paper only focuses on the wind speed variation in the horizontal direction at the same height. Therefore, the fluid mechanics control equation is further simplified and the Z-axis is ignored. At this time, there is no external force acting on the X-axis and Y-axis, that is, $f_x = 0$, $f_y = 0$.

Ultimately, the fluid mechanics control equations are simplified to the continuity equation and the N-S equation on a two-dimensional plane, as shown in Equations (17)–(19).

$$\partial u/\partial x + \partial v/\partial y = 0 \tag{17}$$

$$\rho(\partial u/\partial t + u\partial u/\partial x + v\partial u/\partial y) = -\partial p/\partial x + \mu\left(\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2\right) \tag{18}$$

$$\rho(\partial v/\partial t + u\partial v/\partial x + v\partial v/\partial y) = -\partial p/\partial y + \mu\left(\partial^2 v/\partial x^2 + \partial^2 v/\partial y^2\right) \tag{19}$$

Incorporate Equations (17)–(19) as physical constraints into the calculation of the loss function, which is shown in Equations (20)–(24).

$$L_{PDE} = L_{continue} + L_{N-S} \tag{20}$$

$$L_{continue} = (\partial u/\partial x + \partial v/\partial y)^2 \tag{21}$$

$$L_{N-S} = \left(L_x + L_y\right)^2 \tag{22}$$

$$L_x = \rho(\partial u/\partial t + u\partial u/\partial x + v\partial u/\partial y) + \partial p/\partial x - \mu\left(\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2\right) \tag{23}$$

$$L_y = \rho(\partial v/\partial t + u\partial v/\partial x + v\partial v/\partial y) + \partial p/\partial y - \mu\left(\partial^2 v/\partial x^2 + \partial^2 v/\partial y^2\right) \tag{24}$$

where $L_{continue}$ is the physical loss of the continuity equation and L_{N-S} is the physical loss of the N-S equation, including the x and y directions. u and v are the wind speed components in the two directions respectively, which can be calculated based on wind speed and wind direction. In addition, in the formula, ρ is the air density, taking $\rho = 1.225\ \mathrm{kg/m^3}$. μ is the dynamic viscosity, taking $\mu = 1.8 \times 10^{-5}$ Pa $\times$ s. p is the air pressure, which can be obtained using the Poisson equation based on the wind speed data. The specific formula is shown in Equation (25).

$$\nabla^2 p = \partial^2 p/\partial x^2 + \partial^2 p/\partial y^2 = -\rho\left[(\partial u/\partial x)^2 + 2(\partial u/\partial y)(\partial v/\partial x) + (\partial v/\partial y)^2\right] \tag{25}$$

4.3. Calculation of the Loss Function

When calculating physical losses, it is necessary to take the second-order partial derivatives for x and y and the first-order partial derivatives for t. However, in the model proposed in this paper, position and time information are not directly used as input information but are implicitly included in the input wind speed, direction and geographical difference information, making it difficult to use automatic differentiation to take the partial derivatives for x, y and t. Based on the above problems, the finite difference method is used to calculate the physical loss for the training data of the same batch.

The finite difference method is derived from the definition of derivatives, as shown in Equation (26). By using a very small h, Equation (26) can be approximately transformed into the form of Equation (27).

$$f'(x) = \lim_{h \to 0} [f(x+h) - f(x)]/h \tag{26}$$

$$f'(x) = [f(x+h) - f(x)]/h \tag{27}$$

During the model training process, data from adjacent time steps of multiple wind measurement towers will be processed in the same batch. Take a virtual point 30 m from each of the east, south, west and north directions of the wind measurement tower. Based on the wind speed and direction data of these four virtual points and the wind anemometer tower, calculate the derivatives in the X- and Y-axis using the finite difference method. Use the wind speed and direction data of the wind measurement tower in the same batch and adjacent time steps to calculate the derivative with reference to time.

5. Analysis of Examples

5.1. Data Declaration

This article takes the data of a certain wind farm in northwest China as an example for case study analysis. There are a total of twenty anemometer towers in this wind farm, numbered from 1 to 20, and their geographical locations are shown in Figure 6. T1 to T20 represent the serial numbers of the anemometer towers, and the pink line indicates the scope of the wind farm in Figure 6.

The No. 1 and No. 16 anemometer towers are retained as the test set, while the rest are used as the training set. The anemometer towers in the training set are alternately used as output target points, and the training data set is formed in the manner described above for model training. The Normalized Mean Absolute Error (NMAE) and normalized root mean square error (NRMSE) are used to measure the model performance. The definitions of NRMSE and NMAE for measuring wind speed are shown in Equations (28) and (29).

$$NRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2}/(y_{\max} - y_{\min}) \tag{28}$$

$$NMAE = \sum_{i=1}^{n}|y_i - \overline{y}|/[n(y_{\max} - y_{\min})] \tag{29}$$

Since the wind direction circulates within the range of 0 to 360°, when the predicted wind direction is 1° and the measured wind direction is 359°, the error is 2° instead of 358°. Therefore, the definitions of NRMSE and NMAE when calculating the wind direction error are shown in Equations (30) and (31).

$$NRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\{MIN[y_i - \overline{y}, 360 - (y_i - \overline{y})]\}^2}/(y_{\max} - y_{\min}) \tag{30}$$

$$NMAE = \sum_{i=1}^{n} |MIN[y_i - \overline{y}, 360 - (y_i - \overline{y})]| / [n(y_{\max} - y_{\min})] \quad (31)$$

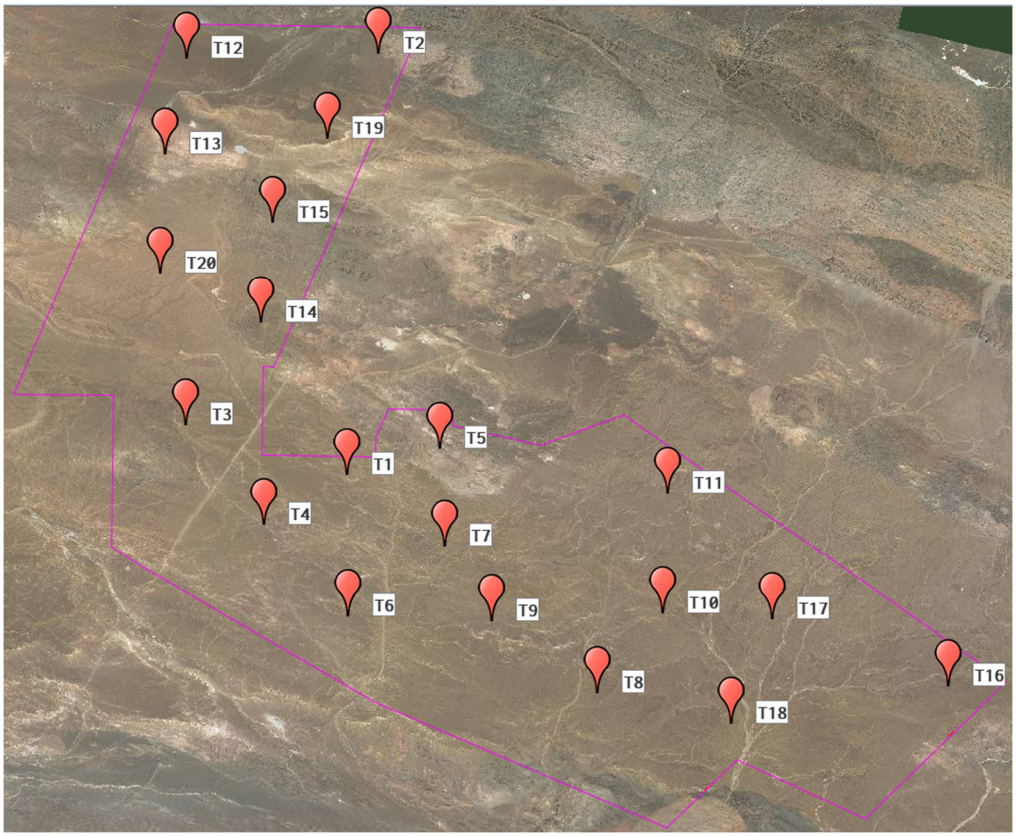


Figure 6. Schematic diagram of the location of the wind tower.

5.2. Interpretation of Result

The feasibility of the model proposed in this paper was verified by using anemometer towers No. 1 and No. 16 as test sets and the data of all anemometer towers except No. 1 and No. 16 as training sets. Meanwhile, the performance of the model proposed in this paper was verified by using the BP model and GCN model as controls.

The PINN model consists of three hidden layers with 128 neurons per layer. It uses the tanh activation function and Adam optimizer, with a batch size of 32 and a fixed learning rate of 0.001. The total number of training sessions is 80, taking the data loss weight as 1 and the physical loss weight as 0.1. The BP model and the GCN model have the same parameters as the PINN model, such as the number of hidden layers, the number of neurons and other parameters, but there is only data loss in the BP model and the GCN model. The BP model and GCN model have the same input and output as the PINN model proposed in this paper. The GCN model consists of five nodes, one of which is the target node and the other four are the wind measurement towers closest to the target node. The features of each node include the wind speed and direction of the wind measurement tower at the current moment and the geographical difference data between the wind measurement tower and the target node.

The input data include the wind speed and direction data of the four anemometer towers closest to the target anemometer tower and the geographical difference data between these four anemometer towers and the target anemometer tower. The output is the wind speed and direction data of the target anemometer tower. The normalized root mean square error (NRMSE) and Normalized Mean Absolute Error (NMAE) were used to measure the

errors of the model output. The convergence curves of the three models are shown in Figure 7.

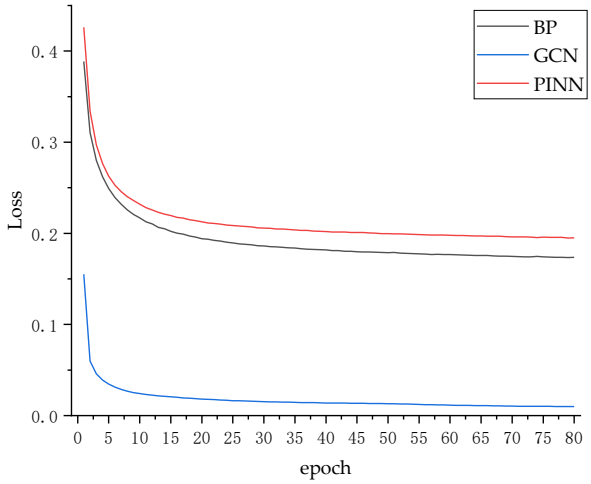


Figure 7. Convergence curves of BP, GCN, and PINN.

The training time for the BP model was 52 min and 20 s, for the GCN model, it was 42 min and 20 s, and for the PINN model, it was 81 min and 37 s. The errors of the three models are shown in Table 2 and Figure 8.

Table 2. Comparison of error results of the three models.

Wind Speed Error	No. 1		No. 16		Wind Direction Error	No. 1		No. 16	
	NRMSE	NMAE	NRMSE	NMAE		NRMSE	NMAE	NRMSE	NMAE
BP	21.05%	17.94%	20.18%	16.02%	BP	24.08%	19.26%	24.61%	21.97%
GCN	16.96%	13.15%	24.03%	18.80%	GCN	13.11%	8.44%	24.20%	18.85%
PINN	10.25%	7.41%	16.41%	12.69%	PINN	11.24%	6.85%	25.70%	20.34%

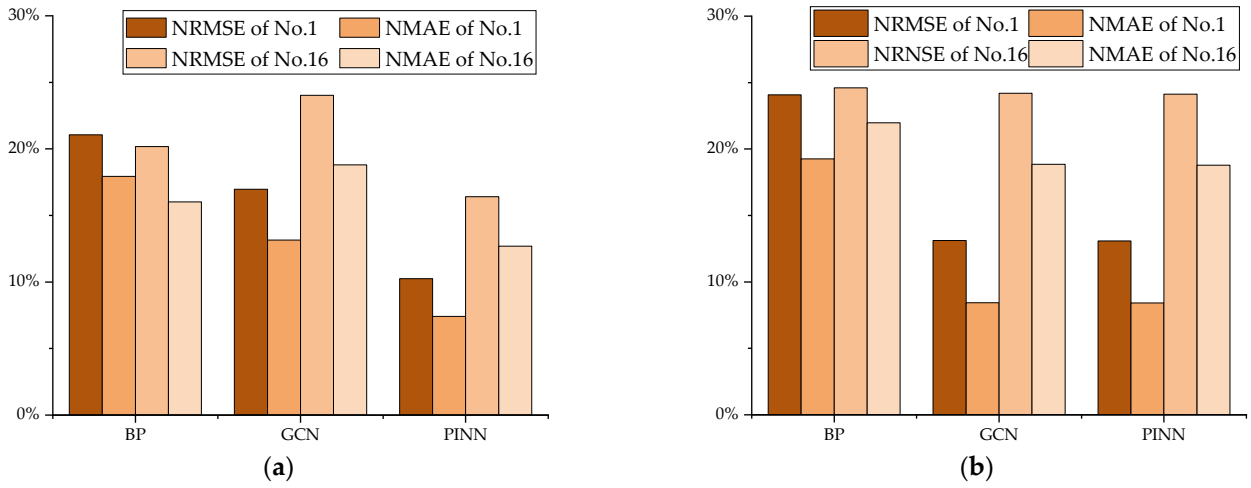


Figure 8. (**a**) Comparison of wind speed errors of the three models; (**b**) comparison of wind direction errors of the three models.

It can be seen from the results that the physics-informed neural network model integrating the control equations of fluid mechanics proposed in this paper has errors within an acceptable range in the data of anemometer towers No. 1 and No. 16 outside the training set. Moreover, the wind speed error of the model proposed in this paper is significantly smaller than that of the BP neural network and the GCN model, and the wind direction error is also slightly improved compared with the GCN model. The validity of the model proposed in this paper has been proved.

For the situation where the error of anemometer tower No. 16 is much greater than that of Tower No. 1, the training set and test set were changed. Anemometer towers No. 2, No. 4, No. 5 and No. 7 were respectively used as the test set, and the rest of the anemometer towers were used as the training set. The results obtained by changing the training set and the test set are shown in Table 3.

Table 3. Error results for different training sets and test sets.

BP	**Wind Speed Error**		**Wind Direction Error**			**Wind Speed Error**		**Wind Direction Error**	
	NRMSE	**NMAE**	**NRMSE**	**NMAE**		**NRMSE**	**NMAE**	**NRMSE**	**NMAE**
No. 1	21.05%	17.94%	24.08%	19.26%	No. 2	11.82%	9.01%	19.01%	14.62%
No. 4	16.40%	11.58%	17.86%	11.35%	No. 5	15.25%	11.88%	13.99%	10.44%
No. 7	9.00%	6.70%	12.39%	9.32%	No. 16	20.18%	16.02%	24.61%	21.97%
GCN	**Wind Speed Error**		**Wind Direction Error**			**Wind Speed Error**		**Wind Direction Error**	
	NRMSE	**NMAE**	**NRMSE**	**NMAE**		**NRMSE**	**NMAE**	**NRMSE**	**NMAE**
No. 1	16.96%	13.15%	13.11%	8.44%	No. 2	10.15%	9.11%	10.11%	9.54%
No. 4	11.21%	9.92%	18.01%	16.69%	No. 5	9.28%	7.93%	21.20%	20.33%
No. 7	15.80%	15.80%	11.57%	10.78%	No. 16	24.03%	18.80%	24.20%	18.85%
PINN	**Wind Speed Error**		**Wind Direction Error**			**Wind Speed Error**		**Wind Direction Error**	
	NRMSE	**NMAE**	**NRMSE**	**NMAE**		**NRMSE**	**NMAE**	**NRMSE**	**NMAE**
No. 1	10.25%	7.41%	11.24%	6.85%	No. 2	10.98%	8.33%	17.12%	11.33%
No. 4	15.18%	10.43%	17.84%	10.73%	No. 5	9.61%	7.24%	14.14%	6.83%
No. 7	8.95%	6.76%	12.39%	6.69%	No. 16	16.41%	12.69%	25.70%	20.34%

As shown in Table 3, after changing the test set and training set, the error of the PINN model is still smaller than that of the BP model and the GCN model. Then, analyze the relationship between the input data of the PINN model and the error. Calculate the correlation coefficient of sudden wind speed and direction between each target anemometer tower and the four input anemometer towers, and further analyze the factors affecting the model accuracy. Since the input data of the model comes from four anemometer towers, and as the target anemometer tower changes, the input anemometer towers will also change accordingly, the input anemometer tower numbers are not listed here. Instead, according to the distance between the input anemometer tower and the target anemometer tower, the input anemometer towers are respectively called Input 1, Input 2, Input 3 and Input 4 from near to far, and the distance is normalized. The correlation coefficients with wind speed and direction are plotted on the same graph, as shown in Figure 9. As can be seen from Figure 9, the correlation between the wind speed and direction data of the input anemometer tower and the output anemometer tower will directly affect the accuracy of the data generated by the model. In addition, the short distance between the input anemometer tower and the output anemometer tower and their distribution in different directions around the output anemometer tower are conducive to achieving higher output accuracy results of the model.

To evaluate the generalization ability of this model under different spatial conditions, the data from one of the 20 anemometer towers were used as the test set, and the data from the other remaining 19 anemometer towers were used as the training set. The model was repeatedly trained 20 times and the errors of the model under different combinations of training sets and test sets were tested. Figure 10 represented the results. Using MAE to measure the errors of each model, the error range for wind speed was 4.29% to 14.62%, with an average of 8.28%, and the error range for wind direction was 4.30% to 27.26%, with an average of 10.50%. The average distance between each anemometer tower is 6.05 km, among which the distance between anemometer tower No. 16 and its nearest neighbor is the largest, at 10.59 km, and the wind speed and direction error of anemometer tower No. 16 is also the largest. Based on these observations, we believe that when the input point is within a 10 km radius around the target point, the proposed PINN model can reliably generate wind speed and direction data. However, since this method was only tested and

trained on a dataset of 20 wind measurement towers with a specific spatial distribution, the proposed method cannot be universally applicable to any wind farm for the time being without recalibration or verification of other datasets. In the future, it is necessary to further verify how to reliably extend the proposed method to more application scenarios and study adaptive strategies to improve the generalization ability of this method.

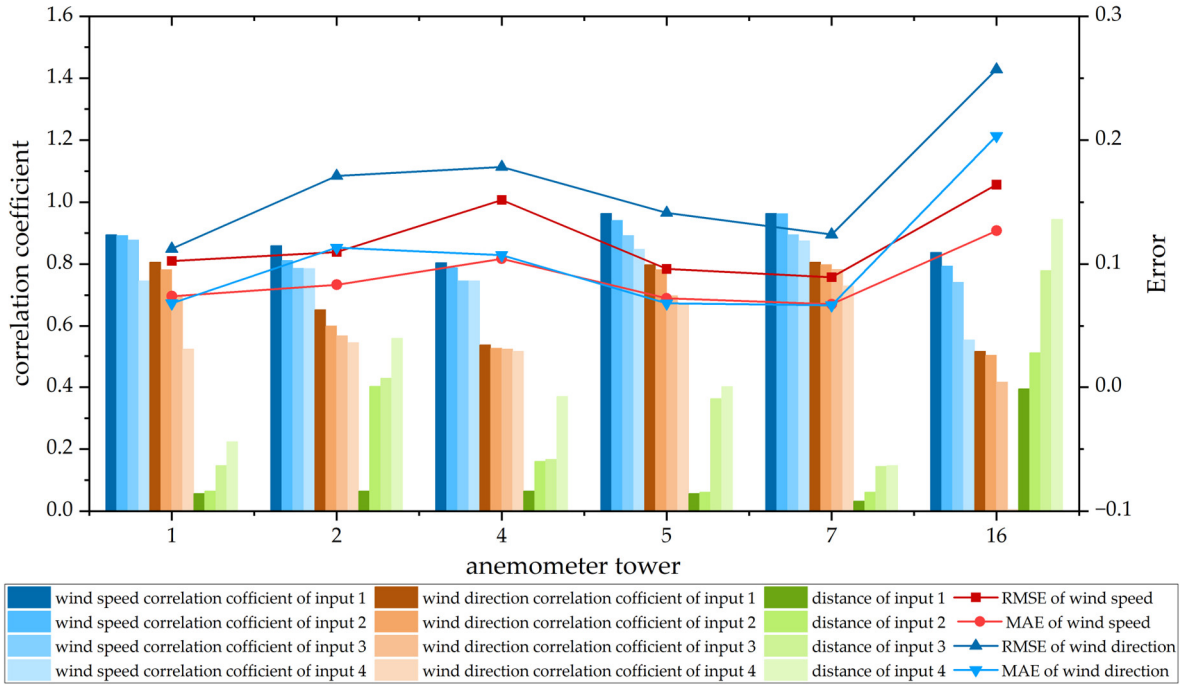


Figure 9. Comparison of output errors and input data of different training sets and test sets.

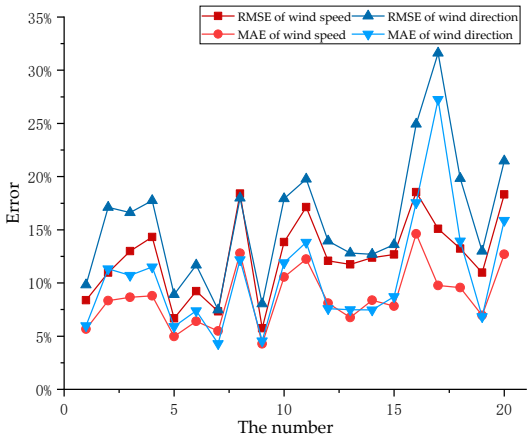


Figure 10. The results of twenty different test sets.

6. Conclusions

Aiming to solve the problem of the points which are difficult to measure wind speed data or model and generate wind speed data, this paper proposed a wind farm wind speed data generation method based on physics-informed neural networks, and conducted a case study analysis of the measured data of an anemometer tower in a certain wind farm in the northwest of China, and obtained the following conclusions:

1. Establishing a wind speed generation model that takes into account the geographical differences between input and output points and actual physical constraints. This model has achieved good results on the test set and can generate wind speeds at other points outside the training set with errors within an acceptable range.

2. The model proposed in this paper is based on a physics-informed neural network and incorporates the fluid mechanics control equation as part of the loss function into the model training process, which can achieve higher accuracy with less training data. Through comparative analysis, it can be seen that the performance of the model proposed in this paper is significantly better than that of the BP and GCN models.
3. There are lots of factors that affect the accuracy of the data generated by the model, including the correlation of wind speed between the input and output points, the correlation of wind direction between them, the distance between the two points, and the orientation of the output point relative to the input point. When the correlation of wind speed and direction between the input anemometer tower and the output anemometer tower is high, the distance is short, and the distribution of the input anemometer tower is around the output anemometer tower, the accuracy of the data generated by the model is the highest.

Author Contributions: Conceptualization, X.G. and L.H.; methodology, Y.L. and X.L.; validation, X.G., Y.L. and L.H.; formal analysis, G.H.; data curation, G.H.; writing—original draft preparation, Y.L., X.L. and L.H.; writing—review and editing, X.G., F.X. and X.Y.; visualization, L.C.; supervision, L.C. and X.Y.; project administration, F.X. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of Xinjiang Huadian Tianshan Power Generation Co., Ltd., grant number CHDKJ24-02-01-35.

Data Availability Statement: Data are available on request due to restrictions on privacy or ethics. The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Authors Xinping Gao, Yuanzhi Li, Xinhua Lei, Guixia Han were employed by the company Xinjiang Huadian Tianshan Power Generation Co., Ltd. Author Xiangyu Yan was employed by the company Huadian Electric Power Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The companies had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Jiang, L.; Ji, X.; Yang, S.; Zhang, Z. Assessment of wind resource potential at different heights in urban areas: A case study of Beijing. *Renew. Energy* **2025**, *248*, 123175. [CrossRef]
2. Ma, X.; Liu, Y.; Yan, J.; Han, S.; Li, L.; Meng, H.; Deveci, M.; Kolle, K.; Cali, U. Assessment Method of Offshore Wind Resource Based on Multi-dimensional Indexes System. *CSEE J. Power Energy Syst.* **2024**, *10*, 76–87. [CrossRef]
3. He, J.Y.; Chan, P.W.; Li, Q.S.; Huang, T.; Yim, S.H.L. Assessment of urban wind energy resource in Hong Kong based on multi-instrument observations. *Renew. Sustain. Energy Rev.* **2024**, *191*, 114123.
4. Gao, Y.; Ma, S.; Wang, T.; Miao, C.; Yang, F. Distributed onshore wind farm siting using intelligent optimization algorithm based on spatial and temporal variability of wind energy. *Energy* **2022**, *258*, 124816. [CrossRef]
5. Wimhurst, J.J.; Greene, J.S.; Koch, J. Predicting commercial wind farm site suitability in the conterminous United States using a logistic regression model. *Appl. Energy* **2023**, *352*, 121880. [CrossRef]
6. Uguz, Y.F.; Bilgili, A.; Uzar, M. Toward improved siting of wind–solar hybrid farms: A novel framework integrating multi-criteria decision making, machine learning–driven feature selection, and spatial clustering. *Appl. Energy* **2026**, *411*, 127609.
7. Brigada, D.J.; Ryvkina, J. Radar-Optimized Wind Turbine Siting. *IEEE Trans. Sustain. Energy* **2022**, *13*, 403–413. [CrossRef]
8. Dinçer, A.E.; Demir, A.; Yılmaz, K. Enhancing wind turbine site selection through a novel wake penalty criterion. *Energy* **2023**, *283*, 129096. [CrossRef]
9. Dinçer, A.E.; Demir, A.; Yılmaz, K. Multi-objective turbine allocation on a wind farm site. *Appl. Energy* **2024**, *355*, 122346. [CrossRef]
10. Wang, K.; Zhu, H.; Dang, J.; Ming, B.; Wu, X. Short-term optimal scheduling of wind-photovoltaic-hydropower-thermal-pumped hydro storage coupled system based on a novel multi-objective priority stratification method. *Energy* **2024**, *309*, 133190. [CrossRef]

11. Aghamohamadi, M.; Mahmoudi, A.; Haque, M.H. Adaptive Robust Recourse-Based Bidding Strategy and Capacity Allocation of PV-WT-BES Owning Prosumers Under Uncertainties. *IEEE Trans. Ind. Appl.* **2021**, *57*, 4170–4186. [CrossRef]
12. Yang, Q.; Peng, L.; Cai, S.; Huang, G.; Wei, G.; Jiang, Y. Uncertainty Study on Annual Power Generation of Wind Farms. *J. Southeast Univ. (Nat. Sci. Ed.)* **2025**, *55*, 1255–1264. (In Chinese)
13. Ayik, A.; Ijumba, N.; Kabiri, C.; Goffin, P. Preliminary wind resource assessment in South Sudan using reanalysis data and statistical methods. *Renew. Sustain. Energy Rev.* **2020**, *138*, 110621.
14. Gualtieri, G. Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review. *Renew. Sustain. Energy Rev.* **2022**, *167*, 112741. [CrossRef]
15. Majidi Nezhad, M.; Neshat, M.; Groppi, D.; Marzialetti, P.; Heydari, A.; Sylaios, G.; Garcia, D.A. A primary offshore wind farm site assessment using reanalysis data: A case study for Samothraki island. *Renew. Energy* **2021**, *172*, 667–679. [CrossRef]
16. Zhao, S.; Jin, J.; Han, P. Numerical Simulation of Offshore Wind Energy Resources Based on CFD. *Acta Energiae Solaris Sin.* **2025**, *46*, 378–385. (In Chinese)
17. Hu, W.; Yang, Q.; Yuan, Z.; Yang, F. Wind farm layout optimization in complex terrain based on CFD and IGA-PSO. *Energy* **2024**, *288*, 129745.
18. Ma, G.; Li, J.; Song, Y.; Tian, L.; Liu, X.; Zhao, N. Research on wind resource assessment of complex mountain wind farms based on medium-micro scale coupling technology. *Acta Energiae Solaris Sin.* **2025**, *46*, 642–654. (In Chinese)
19. Cuevas-Figueroa, G.; Stansby, P.K.; Stallard, T. Accuracy of WRF for prediction of operational wind farm data and assessment of influence of upwind farms on power production. *Energy* **2022**, *254*, 124362. [CrossRef]
20. Xing, Z.; Chou, J.; Guo, S.; Chen, M.; Chen, L.; Liu, Y. Research on Wind Measurement Data Interpolation Method Based on Hybrid Neural Network. *Acta Energiae Solaris Sin.* **2025**, *46*, 458–464. (In Chinese)
21. Qin, Q.; Liu, S.; Lai, X.; Liu, X. Wind Farm Wind Measurement Data Interpolation for GA Optimized ELM Neural Network. *Acta Energiae Solaris Sin.* **2018**, *39*, 2125–2132. (In Chinese)
22. Carta, J.A.; Cabrera, P. Transformation of reanalysis data for improved long-term estimation of wind speed and direction at a target site. *Renew. Energy* **2026**, *261*, 125280. [CrossRef]
23. Jia, X.; Shi, L.; Hao, Y. Comparison of artificial intelligence interpolation methods based on wind speed in wind farm area. *Acta Energiae Solaris Sin.* **2025**, *46*, 168–175. (In Chinese)
24. *NB/T 31147—2018*; Technical Code for Wind Energy Resource Measurement and Assessment of Wind Power Projects: NB/T 31147—2018. Benliu Electronic Audiovisual Publishing (Beijing) Co., Ltd.: Beijing, China, 2018. (In Chinese)
25. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707. [CrossRef]
26. Thirugnanasambandam, M.K.; Pinto, J.; Moskovkina, E.; Costa, R.S.; Oliveira, R. A Physics-Informed Neural Network (PINN) framework for generic bioreactor modelling. *Comput. Chem. Eng.* **2025**, *203*, 109354. [CrossRef]
27. Zhang, X.; Mao, B.; Che, Y.; Kang, J.; Luo, M.; Qiao, A.; Liu, Y.; Anzai, H.; Ohta, M.; Guo, Y.; et al. Physics-informed neural networks (PINNs) for 4D hemodynamics prediction: An investigation of optimal framework based on vascular morphology. *Comput. Biol. Med.* **2023**, *164*, 107287. [CrossRef] [PubMed]
28. Saeed, M.H.; Kazmi, H.; Deconinck, G. Dyna-PINN: Physics-informed deep dyna-q reinforcement learning for intelligent control of building heating system in low-diversity training data regimes. *Energy Build.* **2024**, *324*, 114879.
29. Shi, M.; Fang, H.; Xie, Y.; Du, H.; Liu, S.; Ndayiragije, J.M.; Liu, N. A flood forecasting method coupling the CE-QUAL-W2 and PINN models. *J. Hydrol.* **2025**, *66*, 133699. [CrossRef]
30. Lin, J. *Fluid Mechanics*, 2nd ed.; Tsinghua University Press: Beijing, China, 2013; pp. 94–110. (In Chinese)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.