



Article An Adaptive Hybrid Model for Wind Power Prediction Based on the IVMD-FE-Ad-Informer

Yuqian Tian¹, Dazhi Wang^{1,*}, Guolin Zhou¹, Jiaxing Wang¹, Shuming Zhao¹ and Yongliang Ni²

- ¹ College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
- ² China North Vehicle Research Institute, Beijing 100072, China
- * Correspondence: prodzw@126.com

Abstract: Accurate wind power prediction can increase the utilization rate of wind power generation and maintain the stability of the power system. At present, a large number of wind power prediction studies are based on the mean square error (MSE) loss function, which generates many errors when predicting original data with random fluctuation and non-stationarity. Therefore, a hybrid model for wind power prediction named IVMD-FE-Ad-Informer, which is based on Informer with an adaptive loss function and combines improved variational mode decomposition (IVMD) and fuzzy entropy (FE), is proposed. Firstly, the original data are decomposed into *K* subsequences by IVMD, which possess distinct frequency domain characteristics. Secondly, the sub-series are reconstructed into new elements using FE. Then, the adaptive and robust Ad-Informer model predicts new elements and the predicted values of each element are superimposed to obtain the final results of wind power. Finally, the model is analyzed and evaluated on two real datasets collected from wind farms in China and Spain. The results demonstrate that the proposed model is superior to other models in the performance and accuracy on different datasets, and this model can effectively meet the demand for actual wind power prediction.

check for updates

Citation: Tian, Y.; Wang, D.; Zhou, G.; Wang, J.; Zhao, S.; Ni, Y. An Adaptive Hybrid Model for Wind Power Prediction Based on the IVMD-FE-Ad-Informer. *Entropy* **2023**, 25, 647. https://doi.org/10.3390/ e25040647

Academic Editors: Luis Hernández-Callejo, Sergio Nesmachnow and Sara Gallardo Saavedra

Received: 6 March 2023 Revised: 5 April 2023 Accepted: 11 April 2023 Published: 12 April 2023



Copyright: © 2023 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 (https:// creativecommons.org/licenses/by/ 4.0/). **Keywords:** wind power prediction; improved variational mode decomposition; fuzzy entropy; adaptive loss function; Informer

1. Introduction

The global energy-shortage problem is becoming more and more serious, and it is essential to accelerate the pace of energy structure transformation based on the increasing proportion of renewable energy. Wind power, as an economical and environmentally friendly emerging renewable energy source, has been vigorously developed by various countries, and its application prospects are promising [1,2]. However, with the random fluctuation of wind power, it has strong uncontrollability, resulting in a decrease in the dispatching efficiency of the power grid and an imbalance between energy supply and demand [3]. Therefore, achieving high-accuracy and high-reliability prediction of wind power in practical applications can minimize energy loss and make the power grid operate more stably and safely.

In recent years, a large number of scholars have studied wind power prediction models, which can be mainly divided into physical models [4], statistical models [5], artificial intelligence (AI) models [6], and hybrid models [7]. The physical models are based on the method of fluid mechanics, which uses numerical weather prediction data to calculate the wind turbine output curve and then calculate wind power from it [8]. However, the fluid mechanics method has the disadvantages of the high complexity of building the model and massive computational cost. The statistical models are based on the mapping relationship between historical data and future data [9,10]. Rajagopalan et al. [11] proposed an autoregressive moving average (ARMA) model for ultra-short-term wind power forecasting and achieved superior results. The autoregressive integrated moving

average (ARIMA) model is a widely used statistical model that is based on ARMA with the addition of difference operation [12,13]. However, this method requires a large-scale dataset, making it difficult to mine the nonlinear relationship of complex data.

AI models are current technical trends and are widely used in the field of largescale and multi-dimensional data prediction [14]. AI models are mainly divided into machine-learning models and deep-learning models [15]. For example, an echo state network (ESN) [16] was applied to wind speed forecasting and improved the prediction performance. Khan et al. [17] used the Naive Bayes Tree (NB) to extract the probabilities of each feature of wind power, successfully predicting wind power values from hours to years. Machine-learning methods are based on rigorous mathematical theories that enable rapid computation in high-dimensional spaces. Owing to weak generalization ability, machine-learning methods are prone to overfitting, and it is difficult to achieve good prediction effects. In contrast, deep-learning models unify feature-learning tasks and prediction tasks into one model, making them more suitable than shallow machine-learning models to solve wind power prediction problems in complicated uncertainty scenarios [18]. Tian et al. [19] used a model based on the attention mechanism and demonstrated its efficacy in wind power prediction. Liu et al. [20] presented a novel deep convolutional neural network (CNN) capable of automatically extracting hidden information from multidimensional data and efficiently implementing multi-step prediction. Hu et al. [21] applied a model integrated with a deep-learning framework and basic ESN network for energy prediction, which enhanced the model's memory capacity with a stacked hierarchy of reservoirs. Although these methods have achieved some success in wind power prediction, the fact that a single model cannot fully exploit the time series information leads to limited prediction performance [22].

Hybrid models are created by multiple intelligent algorithms or prediction models that combine the advantages of different models to achieve an improvement in prediction accuracy. Hybrid prediction models consist of combined multiple models and stacking models based on data processing [23,24]. Chen et al. [25] designed a weighted combination prediction composed of six long short-term memory networks (LSTM), and its prediction effect is better than that of a single prediction model. Xiong et al. [26] proposed a multiscale hybrid prediction model that combines attention mechanism, CNN, and LSTM to adequately capture the high-dimensional features in wind farm data. Zheng et al. [27] established a hybrid model combining bidirectional long-short-term memory (Bi-LSTM) and CNN, which adopted a unique feature extraction method of space and then time. Although the combined prediction method of multiple models exhibits high prediction accuracy, it suffers low computational efficiency and narrow application scenarios [28]. Considering the nonlinear implicit relationship in the time series of wind power data, the stacking model based on data processing is proposed to improve the prediction accuracy by mining deep features through data decomposition. For example, Wu et al. [29] proposed a multi-step prediction method using variational modal decomposition (VMD) and chain ESN, which achieved multi-steps prediction at multiple time scales. Yang et al. [30] employed the VMD method to decompose the wind speed data, which was then utilized as input for an optimized LSTM network to perform predictions. Ren et al. [31] proposed a hybrid model of empirical-mode decomposition (EMD) and support-vector regression (SVR) for wind power prediction. Lv et al. [32] decomposed wind speed data into 3-dimensional input features using singular spectrum analysis (SSA) and fed them into a convolutional longshort-term memory (ConvLSTM) network, which effectively enhanced the local correlation between multivariate data. Khazaei and Ehsan [33] used prediction methods combining wavelet transform (WT) decomposition with the AI model, and the results showed that the model has high accuracy. Hybrid models based on data processing have a simple structure and strong feasibility, but the accuracy of the prediction greatly depends on the effectiveness of data decomposition. Over-decomposition of data can result in redundant components and reduce the efficiency of calculation, while the insufficient decomposition of data can lead to mode mixing, which fails to meet the needs of high-precision prediction [34].

Most prediction models use mean square error (MSE) as a loss function, which needs to meet the condition that prediction errors obey a Gaussian distribution. However, the use of MSE as a loss function in models that are insensitive to outliers in wind power data with high randomness may result in large errors [35]. In response to this issue, some researchers have improved the loss function of the model to minimize the impact of errors on the prediction results. Hu et al. [36] proposed a loss function without fixed distribution, which effectively solves the problem of prediction-gradient descent at wind power intervals. Duan et al. [35] designed a loss function with non-Gaussian distributed errors and combined it with an LSTM model to predict wind power. The loss function is significantly important in the wind power prediction process as it determines the training direction and accuracy of the model [37]. Although these improved loss functions have positive effects on wind power prediction, the models still exhibit limitations in terms of their adaptability and robustness.

Based on the above analysis, a robust and adaptive IVMD-FE-Ad-Informer hybrid model for wind power prediction is proposed in this paper, which aims to improve the precision of data decomposition and the predictive performance of non-stationary wind power data. The main contributions of this paper are outlined as follows: (1) Considering the difficulty in selecting the number of VMD, the IVMD algorithm improved by the maximum information coefficient (MIC) decomposes the original wind power data into K optimal sub-series, which effectively reduces the difficulty of wind power prediction. (2) fuzzy entropy (FE) is used to reconstruct sub-series into new elements of similar complexity together, alleviating the burden of model operation. (3) An adaptive loss function is innovatively introduced into the Informer network to solve the problem of traditional MSE's insensitivity to randomly fluctuating wind power data. This novel model can reduce the impact of outliers in non-smooth wind power data. (4) Ablation experiments and comparative experiments are performed on datasets collected from both different wind farms to verify the effectiveness and stability of the model. The prediction results show that the proposed model framework is reasonable, and it exhibits significantly better prediction performance and accuracy compared to other models.

The specific contents of this paper are as follows: Section 2 specifically describes the basic methodologies of hybrid models; Section 3 presents the construction framework and evaluation indicators of the IVMD-FE-Ad-Informer model; Section 4 constructs four experiments to verify the accuracy and validity of the proposed model; and finally, this paper is summarized in Section 5.

2. Methodologies

2.1. Variational Mode Decomposition

VMD [38] is a commonly used data decomposition method that converts wind power sequences from the time domain to the frequency domain and subsequently decomposes them into *K* intrinsic mode functions (IMFs). Firstly, build the variational constraint equation:

$$\begin{cases} \min\left\{\sum_{K=1}^{K} \left\|\partial_t \left[\left(\delta(t) + \frac{i}{\pi t}\right) * u_K(t)\right] e^{-iw_K t}\right\|_2^2\right\} \\ \text{s.t.} \sum_{K=1}^{K} u_K = x(t) \end{cases}$$
(1)

where * is the convolution calculation symbol, x(t) is the wind power sequence, w_K and u_K are the central frequency and band components of the *k*th IMF value, $\delta(t)$ is the impulse function, and ∂_t is used to denote the derivative of the function.

To simplify the variational constraint equation to a simple unconstrained problem, the Lagrange function $\lambda(t)$ and the penalty factor α are introduced:

$$L(\{u_{K}\},\{w_{K}\},\lambda) = \alpha \sum_{K=1}^{K} \left\| \partial_{t} \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_{K}(t) \right] e^{-iw_{K}t} \right\|_{2}^{2} + \left\| x(t) - \sum_{K=1}^{K} u_{K}(t) \right\|_{2}^{2} + \left\langle \lambda(t), x(t) - \sum_{K=1}^{K} u_{K}(t) \right\rangle$$
(2)

Then the optimal solution of the unconstrained problem is solved using the alternating direction method of multiplication with the following iterative procedure:

$$\hat{u}_{K}^{n+1}(w) = \frac{\hat{x}(w) - \sum_{i \neq K} \hat{u}_{j}(w) + (\hat{\lambda}(w)/2)}{1 + 2\alpha(w - w_{K})^{2}}$$
(3)

$$w_{K}^{n+1} = \frac{\int_{0}^{\infty} w \left| \hat{u}_{K}^{n+1}(w) \right|^{2} dw}{\int_{0}^{\infty} \left| \hat{u}_{K}^{n+1}(w) \right|^{2} dw}$$
(4)

Finally, after applying the above process, the original wind power series is decomposed into the *K* sub-series.

2.2. Fuzzy Entropy

Fuzzy entropy (FE) [39] is a dynamical method for analyzing the complexity of time series. The FE value changes smoothly with changes in the set parameters, which makes it more robust to noise and more resistant to interference. Firstly, for time series with the length of n, the FE algorithm is introduced into the fuzzy membership function, and the specific formula is as follows:

$$D(x) = \exp\left[-\ln(2)\left(\frac{x}{r}\right)^2\right]$$
(5)

where *r* is the similarity tolerance, $x = d_{ij}^m$, d_{ij}^m is the distance between vectors that reconstruct the time series into m-dimensional phase space, and *i*, *j* = 1,2..., *n* – *m* + 1, *i* \neq *j*.

Averaging over each *i* in D_{ij}^m yields, the average similarity function is as follows:

$$\phi^{m}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \left(\frac{1}{N-m} \sum_{j=1, j \neq i}^{N-m+1} D_{ij}^{m} \right)$$
(6)

Therefore, the FE expression is as follows:

$$FuzzyEn(m, r, n) = \ln \phi^m(r) - \ln \phi^{m+1}(r)$$
(7)

2.3. Informer

The Informer network is a variant of the Transformer that effectively addresses the long-sequence prediction problem [40]. Improvements of the Informer include: using a probsparse self-attention mechanism to reduce the complexity of matrix computation; introducing a self-attention distillation mechanism to extract the main features of time series, which effectively reduces memory usage; using a decoder to directly output the predicted values generatively to achieve the purpose of long-series prediction. The structure of the Informer model is shown in Figure 1.

The traditional self-attention mechanism consists of query, key, and value, and the expression is as follows:

$$f_A(Q, K, V) = \operatorname{softmax}\left(\frac{QK^{\top}}{\sqrt{d}}\right)V$$
 (8)

where $Q \in \mathbb{R}^{L_Q \times d}$, $K \in \mathbb{R}^{L_K \times d}$, $V \in \mathbb{R}^{L_V \times d}$, *d* is the input dimension.



Figure 1. The structure of the Informer.

As the matrix multiplication involved in Equation (8) is computationally huge, the probsparse self-attention mechanism is introduced to select the important elements in Q to calculate the attention values.

$$f_A(Q, K, V) = \operatorname{softmax}\left(\frac{\overline{Q}K^{\top}}{\sqrt{d}}\right)V$$
(9)

where \overline{Q} is obtained through probabilistic sparsity of Q and controlled by a constant sampling factor c and the number of \overline{Q} is c * ln L_K .

Therefore, the similarity and importance between query and key are measured by Kullback–Leibler divergence, as follows:

$$k(q_i, k_j) = \ln \sum_{l=1}^{L_K} e^{\frac{q_i k_j^{\perp}}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^{\top}}{\sqrt{d}} - \ln L_K$$
(10)

where the relevance of q_i to k_j is proportional to the magnitude of $k(q_i, k_j)$. If $p(k_j|q_i)$ is close to a uniform distribution, i.e., $p(k_j|q_i) = 1/L_K$, indicating that q_i has the same similarity to all k_j , then q_i is deemed as a redundant vector and can be dropped.

Based on this, the sparsity evaluation formula that defines the *i*-th query is:

$$M(q_i, K) = \ln \sum_{l=1}^{L_K} e^{\frac{q_i k_l^\top}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^\top}{\sqrt{d}}$$
(11)

The self-attentive distillation mechanism is introduced in the encoder. The width of the feature map is reduced to half its length after the distillation layer, which can reduce the overall memory usage and effectively solve the problem of long input. The concrete representation is as follows:

$$X_{j+1}^{t} = f_{\text{MP}} \left(\text{ELU} \left(f_{\text{Conv}} \left(\left[X_{j}^{t} \right]_{AB} \right) \right) \right)$$
(12)

where f_{MP} represents the maximum pooling layer function, f_{Conv} denotes the convolutional layer function, and $[\cdot]_{AB}$ is the attention unit.

The input of the decoder uses the time shield technique, and its input vector is as follows:

$$X_{\text{dec}}^{\text{in}} = f_{\text{Concat}} \left(X_{\text{token}}, X_0 \right) \in \mathbb{R}^{(L_{\text{token}} + L_p) \times a_{\text{model}}}$$
(13)

where $X_{\text{token}} \in \mathbb{R}^{L_{\text{token}} \times d_{\text{model}}}$ is the input start token, L_{token} is the length of the start token, $X_0 \in \mathbb{R}^{L_p \times d_{\text{model}}}$ is the 0-value matrix, and L_p is the length of the part to be predicted.

2.4. Adaptive Loss Function

The adaptive loss function [41] obtains a generalized loss function by introducing robustness as a continuous parameter. During the training process, the adaptive loss function automatically adjusts the robustness parameters around the minimization loss algorithm, thereby enhancing the prediction accuracy. The generalized loss function formula is as follows:

$$f(z,\beta,c) = \frac{|\beta - 2|}{\beta} \left(\left(\frac{(z/c)^2}{|\beta - 2|} + 1 \right)^{\beta/2} - 1 \right)$$
(14)

where *z* is the difference between the true value and the predicted value, c > 0 serves as a scale factor that controls the curvature of the quadratic function at x = 0, and β is a variable parameter that controls the robustness.

By analyzing Equation (14), the adaptive loss function changes with the change of β . For different β , the adaptive loss function formula is as follows:

$$L(z,\beta,c) = \begin{cases} \frac{1}{2}(z/c)^{2} & \text{if } \beta = 2\\ \log\left(\frac{1}{2}(z/c)^{2} + 1\right) & \text{if } \beta = 0\\ \sqrt{(z/c)^{2} + 1} - 1 & \text{if } \beta = 1\\ 1 - \exp\left(-\frac{1}{2}(z/c)^{2}\right) & \text{if } \beta = -\infty\\ \frac{|\beta - 2|}{\beta} \left(\left(\frac{(z/c)^{2}}{|\beta - 2|} + 1\right)^{\alpha/2} - 1\right) & \text{otherwise} \end{cases}$$
(15)

It can be seen that the adaptive loss function can be a variety of loss functions, such as the MSE, Cauchy, Charbonnier, and Welsch loss functions, by adjusting the value of the variable parameter β .

3. Proposed Model

3.1. Improved VMD

With a solid mathematical theoretical foundation, VMD can effectively separate the components of complex signals and greatly suppress mode mixing. However, the decomposition parameter of VMD is given in advance, which limits the performance of data decomposition. In order to overcome the shortcomings of VMD in a parameter setting, this paper proposes the incorporation of the decomposition method and MIC [42] to determine the most suitable number of decompositions *K*. The degree of decomposition is determined by calculating the MIC value between the original sequence y and the reconstructed sequence y', and the MICyy' value is positively correlated with the number of decomposition numbers *K*. The closer MICyy' value is to 1, the less information is lost during VMD decomposition, indicating a more adequate decomposition.

3.2. IVMD-FE-Ad-Informer Model Framework

In consideration of the high volatility of wind power data, this paper introduces the improved VMD and FE methods to the Informer network with adaptive loss function, and the framework is shown in Figure 2. In the data processing stage, the original data are decomposed into *K* IMFs by IVMD. Next, FE is used to calculate the complexity of each IMFs, and the IMFs with similar values are reconstructed into new elements. In

the model-building stage, the input variables for the model are obtained through feature selection using the MIC algorithm and then input into a robust Ad-Informer prediction model. In the results analysis stage, the wind power forecasting results are obtained by linearly superposing the predicted values of each element, followed by visualizing the forecasting curve.



Figure 2. The framework of IVMD-FE-Ad-Informer model.

3.3. Evaluation Indexes

The mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) are used as evaluation indicators for the prediction performance of IVMD-FE-Informer and other benchmark models. The mathematical formula is as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| q_{\text{true}}(t) - q_{\text{pred}}(t) \right|$$
(16)

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left(q_{\text{true}} \left(t \right) - q_{\text{pred}} \left(t \right) \right)^2}$$
(17)

$$R^{2} = 1 - \frac{\sum_{t=1}^{N} (q_{\text{true}}(t) - q_{\text{pred}}(t))^{2}}{\sum_{t=1}^{N} (q_{\text{true}}(t) - \overline{q})^{2}}$$
(18)

where $q_{\text{true}}(t)$ and $q_{\text{pred}}(t)$ denote the true and predicted values of wind power at time *t*, respectively, \overline{q} is the mean value of q_{true} , and *N* is the number of samples in the dataset.

4. Experiment and Analysis

In this section, four sets of experiments are conducted on datasets with different sampling intervals, capacities, and regions. Experiment 1 aims to describe the specific details of the data processing. Experiment 2 is designed as an ablation experiment to verify the prediction performance of the hybrid model. Experiment 3 mainly aims to design a comparative experiment to verify the viability and superiority of each module. Experiment 4 aims to verify the applicability and stability of the proposed model on different datasets. All experiments are run in Python 3.7 and Pytorch environment with Intel(R) Core (TM) i5-12500H CPU @ 4.50 GHz, 12 Cores, NVIDIA GeForce RTX 3050 GPU, a memory capacity of 16 Gb, and Windows 11 operating system.

4.1. Data Description

The experiments are mainly conducted on two complete datasets without missing values in this paper. Dataset A is based on a wind farm in Gansu, China, which was selected from 1 July to 30 September 2019, with a sampling interval of 15 min. Dataset A contains wind power, wind speeds at different heights (10 m, 30 m, 50 m, 70 m, and hub height), air temperature, air pressure, and humidity features. Dataset B was collected from the Sotavento Galicia wind farm in Spain from 18 January to 12 March 2020, with a sampling interval of 10 min. Dataset B contains only wind power, wind speed, and wind direction features. The wind power curves from different datasets are show in Figure 3.

The prediction process is the same for different datasets; in fact, dataset A is used for Experiments 1 to 3, and dataset B is used for Experiment 4. The datasets are divided into the training set, validation set, and test set in a ratio of 7:2:1, and the results of each experiment are obtained by taking the average of 10 iterations. The characteristics of the datasets, including number, maximum value (Max), minimum value (Min), mean, standard deviation (Std), and coefficient of variation (COV), are shown in Table 1.

Dataset	Number	Max (MW)	Min (MW)	Mean (MW)	Std (MW)	COV
Dataset A	8832	120.43	0	23.51	24.74	1.0523
Dataset B	7920	2.803	0	0.8976	0.7885	0.8784

4.2. Experiment 1: The Specific Details of Data Processing

The data processing part mainly includes data decomposition, new elements reconstruction, and feature selection, and in this part, the operation process and the selection of parameters for data processing will be specifically discussed.



Figure 3. The curve of datasets.

4.2.1. Data Decomposition

The IVMD algorithm solves the traditional VMD problem of *K* selection by calculating the MICyy' value. The original wind power data are fed into the IVMD model, which is decomposed into *K* IMFs. Based on the results of MICyy' corresponding to different values of *K* as indicated in Figure 4, it can be observed that the value of MICyy' remains stable and constant for K = 16. The IMFs curve after IVMD decomposition and its corresponding spectrum diagram are shown in Figure 5. By observing the principal frequencies of different IMFs from Figure 5, it can be concluded that the IVMD algorithm proposed in this paper can effectively separate each IMF accurately.



Figure 4. The curve of MICyy' vs. *K*.



Figure 5. The results of the IVMD algorithm. The IMFs curves are shown (**left**), and the spectral densities corresponding to the IMFs are shown (**right**).

4.2.2. New Elements Reconstruction

The original wind power data are decomposed into 16 IMFs, and if all the sub-series are directly fed into the prediction model, it will increase the operational burden of the prediction model. Therefore, the complexity of these IMFs will be evaluated by FE, and then the IMFs with similar complexity will be reconstructed into new elements. After conducting extensive experiments, the values of m = 2 and r = 0.25std are found to be the optimal settings for achieving the best accuracy and running time of the model. The FE values of each IMF are shown in Figure 6, and the reconstructed new elements based on these FE values are shown in Table 2.



Figure 6. The FE value of each IMF.

Tal	ole	2.	New	elements	reconstruction.
-----	-----	----	-----	----------	-----------------

Reconstruction Elements	IMFs
Element 1	IMF1, IMF2
Element 2	IMF3, IMF4, IMF16
Element 3	IMF5, IMF6, IMF7, IMF8, IMF15
Element 4	IMF9, IMF10, IMF13, IMF14
Element 5	IMF11, IMF12

4.2.3. Feature Selection

The computational efficiency and generalization ability of the model can be improved by removing some irrelevant or redundant features from the original dataset. Therefore, MIC is used to analyze the correlation between meteorological features and each element and extract typical features reflecting each element through MIC value. The confusion matrix of MIC is given in Figure 7. It can be found from Figure 7 that the influence characteristics of each element are different, reflecting the overall correlation and local characteristics, respectively. In order to select the features with the highest relevance to build the input variables, the MIC thresholds of each element are set to 0.5. The input feature selection results are shown in Table 3.



Figure 7. The confusion matrix of MIC.

Table 3. The feature selection results.

Element	Input Variables
Element 1	Element 1, 10 m, 30 m, 50 m, wheel height wind speed
Element 2	Element 2, 10 m, 50 m, wheel height wind speed
Element 3	Element 3, 30 m, 50 m, wheel height wind speed
Element 4	Element 4
Element 5	Element 5

4.3. Experiment 2: Ablation Experiment

The purpose of conducting ablation experiments is to verify whether the complex hybrid model has improved the prediction accuracy as compared to simple combinatorial models and single models. The selected benchmark models are IVMD-FE-Ad-Informer, Ad-Informer, and Informer, of which the Informer model uses the MSE loss function. The parameters of AD-Informer are obtained using the grid search method, where the robustness parameter β is adaptively adjusted using the Adam optimizer. The specific parameters are shown in Table 4. The input size of the encoder and decoder is equal to the number of input variables of the model. The prediction curve results of each sub-mode after the training of the IVMD-FE-Ad-Informer model are shown in Figure 8, and the wind power prediction results can be obtained by superimposing them. The final forecasting curves of the ablation experiment are shown in Figure 9, and forecasting errors are shown in Table 5. Figure 9 not only portrays the overall trend of the test set but also amplifies the values from position 300 to 480 in order to offer an in-depth analysis of the predicted results. The main reason is that the wind power data within the test set from the 300th to the 480th position displays more sudden changes and a wider range of variation, thus providing a more comprehensive evaluation of the predictive performance of the proposed model.

Parameters	Values
Input sequence length	96
Start token length	24–96
Prediction sequence length	24–96
Num of encoder layers	3
Num of decoder layers	2
Input size of encoder	5-1
Input size of decoder	5-1
Decoder output	1
Num of heads	8
Dimension of model	512
Probsparse attention factor	5
Early stopping patience	5
Learning rate	0.0001
Dropout	0.05
Epochs	100
Scale factor	1.2
Optimizer	Adam
Ĝpu	Cuda0

Table 4. Parameter setting of the Ad-Informer.

From Figure 9, it can be seen that the Ad-Informer model is significantly closer to the true value than the Informer model at the inflection point, indicating that the proposed adaptive function can effectively mitigate the impact of errors at the abrupt points. Compared with the Ad-Informer model, the IVMD-FE-Ad-Informer model is closer to the real value, indicating that the data processing method can reduce the time delay in the process of prediction. According to Table 5, the Ad-Informer model requires less time than the Informer model, mainly attributed to the automatic adjustment of the adaptive loss function, which enables the model to obtain the optimal loss during the training process and enhance its robustness. The hybrid model proposed in this paper shows significant improvements over the AD-Informer and Informer models, with a decrease of 45.09% and 59.67% in MAE, 44.4% and 55.44% in RMSE, and an increase of 11.42% and 22.72% in R2, respectively. By comparing the considered models, it can be seen that IVMD-FE-Ad-Informer decomposes the original wind power data into finer granularity, which can better explore the internal features of wind power, resulting in a significant improvement in both prediction accuracy and performance.

Table 5. Forecasting errors of ablation experiment.

Model	MAE (MW)	RMSE (MW)	R ²	Time (s)
IVMD-FE-Ad- Informer	3.19	4.67	0.956	1633.21
Ad-Informer	5.81	8.40	0.858	177.13
Informer	7.91	10.48	0.779	165.559

4.4. Experiment 3: Comparative Experiment

To verify the superiority of each module, EMD-FE-Ad-Informer, IVMD-FE-Informer, IVMD-FE-LSTM, LSTM, and ANN are used as benchmark models in the comparison experiments, and the parameter settings of ANN and LSTM are the same as [19,43]. EMD decomposes wind power data into 11 IMFs by trail-and-error method, and then these IMFs are reconstructed into three new components (IMF1~IMF3, IMF4~IMF6, and IMF7~IMF11) using FE. The forecasting curves of different models are shown in Figure 10. The forecasting errors are shown in Table 6. The boxplots of the forecasting errors for each model are given in Figure 11.



Figure 8. The prediction curve results of each sub-mode. (a) Element 1 prediction curve; (b) Element 2 prediction curve; (c) Element 3 prediction curve; (d) Element 4 prediction curve; (e) Element 5 prediction curve.



Figure 9. The forecasting curves of the ablation experiment. The overall forecasting trends are shown at the (**top**), and the local enlargement is shown at the (**bottom**).

Table 6.	Forecasting	errors of	different	models.
----------	-------------	-----------	-----------	---------

Model	MAE (MW)	RMSE (MW)	R ²	Time (s)
IVMD-FE-Ad-Informer	3.19	4.67	0.956	1633.21
EMD-FE-Ad-Informer	4.96	7.31	0.905	1362.47
IVMD-FE-Informer	5.81	8.40	0.889	1878.63
IVMD-FE-LSTM	6.63	9.79	0.808	1732.57
LSTM	7.86	10.95	0.759	305.11
ANN	8.04	11.58	0.731	81.02

Based on the results in Table 6, the IVMD-FE-Ad-Informer model outperforms the single prediction model and the other hybrid models across all evaluation metrics. It can be concluded that MAE decreased by about 35.68–60.32%, RMSE decreased by about 36.11–59.67%, and R^2 increased by about 5.64–30.78%. According to Table 6 and Figure 10, it can be inferred that the IVMD algorithm has superior data decomposition ability compared to the traditional EMD algorithm under similar data processing. This improved ability

enables the IVMD algorithm to more effectively reduce non-smooth features in the original data, resulting in smoother data and improved wind power prediction accuracy. Furthermore, the prediction accuracy of Ad-Informer is much higher than that of Informer and LSTM for the same data processing method, with R² of 0.925, 0.889, and 0.808, respectively. While IVMD-FE-Ad-Informer is relatively time-consuming due to the implementation of the Ad-Informer prediction module five times after the IVMD-FE data preprocessing, it demonstrates a closer resemblance to the actual curve and produces the smallest forecasting errors. It can be indicated that the model proposed in this paper is an optimal combined model with high prediction performance.



Figure 10. The forecasting curves of different models. The overall forecasting trends are shown at the (**top**), and the local enlargement is shown at the (**bottom**).



Figure 11. The boxplots of different models.

4.5. Experiment 4: The Stability of IVMD-FE-Ad-Informer Forecasting

The experimental results demonstrate that IVMD-FE-Ad-Informer outperforms other benchmark models on dataset A and exhibits considerable wind power prediction ability. However, the statistical distributions of wind power data vary across different time intervals, regions, and capacities, which may lead to the phenomenon of unstable forecasting. Therefore, the stability and applicability of the model still need further discussion. In this section, EMD-FE-Ad-Informer, Ad-Informer, LSTM, and ANN are used as benchmark models on dataset B, which are collected from the Sotavento Galicia wind farm in Spain at 10 min sampling intervals. The parameter-setting method of this experiment is the same as Experiment 3, and the specific parameter settings of each algorithm are shown in Table A1, Appendix A. The forecasting curves of dataset B are shown in Figure 12, and the forecasting errors are shown in Table 7.

Table 7. Forecasting errors of different datasets.

Model	MAE (kW)	RMSE (kW)	R ²	Time (s)
IVMD-FE-Ad-Informer	83.01	60.43	0.962	1076.34
EMD-FE-Ad-Informer	115.89	70.75	0.914	671.47
Ad-Informer	144.51	105.46	0.866	156.91
LSTM	186.63	131.04	0.762	228.88
ANN	197.16	140.62	0.746	62.15

According to Figure 12 and Table 7, the results obtained from dataset B are comparable to those from dataset A, indicating that the model proposed in this paper has high stability and generalization ability on different datasets. From Figure 12, IVMD-FE-Ad-Informer exhibited the closest fit to the true values among all the considered models, with the EMD-FE-Ad-Informer following closely behind. It can also be seen from Table 7 that the IVMD-FE-Ad-Informer has the best prediction performance with regard to MAE, RMSE, and R², which are 83.01 kW, 60.43 kW, and 0.962, respectively. The results further confirm that the IVMD algorithm is a superior and effective method for wind power data decomposition.

Based on the experimental results of the two different datasets, it is apparent that the IVMD-FE-Ad-Informer outperforms other benchmark models in terms of all evaluation metrics and has the closest fit of prediction curves to the true values. Meanwhile, the COV value is introduced for further analysis of the influence of prediction accuracy on different datasets. This value is a typical indicator of the degree of data fluctuation, with more volatile data having a higher COV value [44]. It also can be concluded that the accuracy

of the proposed model prediction is inversely related to the degree of fluctuation in the original data. For example, when using Ad-Informer to forecast wind power on dataset A, the R² is 0.866, whereas, on dataset B with a higher COV, R² is slightly lower at 0.858. Furthermore, the superiority of the proposed model in terms of prediction performance becomes more prominent as the original wind power sequence contains more nonlinear features. The outstanding contribution is the development of an adaptive loss function, which can accurately identify and predict violent changes in wind power, thereby effectively mitigating the impact of outliers.



Figure 12. The forecasting curves of different datasets. The overall forecasting trends are shown at the (**top**), and the local enlargement is shown at the (**bottom**).

5. Conclusions

The actual operation of wind farms is influenced by various factors such as weather conditions, season variation, and atmospheric circulation, which can lead to numerous outliers and non-smooth features in the wind power data. The presence of such factors brings many obstacles to achieving the further improvement of accuracy and performance of wind power prediction. Thus, an adaptive hybrid model for wind power prediction based on improved VMD, FE, and Informer in conjunction with adaptive loss function is proposed in this paper. The IVMD-FE-Ad-Informer model is a promising hybrid model that enables adaptive forecasting of stochastically fluctuating wind power data, and its main advantages are summarized as follows:

- 1. The IVMD-FE-Ad-Informer is a hybrid model that demonstrates high accuracy and better robustness by integrating the advantages of multiple technologies, outperforming the basic EMD- FE-Ad-Informer, Ad-Informer, LSTM, and ANN. The results of the proposed model obtained from the Spanish and Chinese datasets demonstrate a significant improvement compared to benchmark models, with a maximum reduction of 57.89% in MAE, 57.03% in RMSE, and a maximum increase of 30.78% in R²;
- Compared with traditional data decomposition methods, VMD improved by MIC can better mine the nonlinear features of the original data, which effectively improves the data quality and reduces the difficulty of prediction;
- 3. Based on a comprehensive analysis of experimental results, the adaptive loss function has a rapid response to non-Gaussian distributed wind power data, which can react quickly to outliers and predict variation trends;
- 4. By prediction experiments on wind farm datasets with different sampling intervals, capacities, and regions, the proposed model shows the best prediction results and closest proximity to the true value. It can be demonstrated that IVMD-FE-Ad-Informer has remarkable generalization ability and broad prospects in wind power prediction.

As can be seen from the above, the hybrid wind power prediction model that combines the advantages of several algorithms has higher prediction accuracy and better robustness. However, there are some problems in this study that need to be improved in the future. Firstly, in this paper, only the correlation factor is considered in feature selection, while F-score and sensitivity factors are not taken into account. In future work, the analysis of the relationship between other variables and wind power using F-score and sensitivity will be conducted to reduce the redundancy of massive data. Then, the parameter selection in this paper may not be precise enough, and to address this issue, the optimization algorithm will be introduced to overcome the sensitive defect of deep-learning networkparameter selection.

Author Contributions: Y.T. and D.W. described the proposed framework and wrote the whole manuscript; Y.T. implemented the simulation experiments; G.Z. and J.W. collected data; Y.N. and S.Z. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported in part by the Natural Science Foundation of China under Grant 52077027 Study and in part by the Liaoning Province Science and Technology Major Project No. 2022021000014.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Acknowledgments: The authors thank the chief editor and the reviewers for their valuable comments on how to improve the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. The parameter settings of each algorithm on dataset B.

Algorithms	Parameters	Values
IVMD	К	17
EE	m	2
T.E.	r	0.25std
IVMD-FE	New mode	4(mode1 = IMF1, IMF2, IMF17; mode2 = IMF3, IMF4, IMF15, IMF16; mode3 = IMF5, IMF12~IMF14; mode4 = IMF6~IMF11)

Table	A1.	Cont.
Table	A1.	Cont.

Algorithms	Parameters	Values
EMD	K	11
		3(mode1 = IMF1, IMF6, IMF7;
EMD-FE	New mode	mode2 = IMF2~IMF5;
		mode3 = IMF8~IMF11)

References

- Lin, Z.; Liu, X. Wind Power Forecasting of an Offshore Wind Turbine Based on High-Frequency Scada Data and Deep Learning Neural Network. *Energy* 2020, 201, 117693. [CrossRef]
- Wang, K.; Zhang, Y.; Lin, F.; Wang, J.; Zhu, M. Nonparametric Probabilistic Forecasting for Wind Power Generation Using Quadratic Spline Quantile Function and Autoregressive Recurrent Neural Network. *IEEE Trans. Sustain. Energy* 2022, 13, 1930–1943. [CrossRef]
- Naik, J.; Dash, P.K.; Dhar, S. A Multi-Objective Wind Speed and Wind Power Prediction Interval Forecasting Using Variational Modes Decomposition Based Multi-Kernel Robust Ridge Regression. *Renew. Energy* 2019, 136, 701–731. [CrossRef]
- Ahmed, A.; Khalid, M. A Review on the Selected Applications of Forecasting Models in Renewable Power Systems. *Renew. Sustain. Energy Rev.* 2019, 100, 9–21. [CrossRef]
- Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A Review of Wind Speed and Wind Power Forecasting with Deep Neural Networks. *Appl. Energy* 2021, 304, 117766. [CrossRef]
- Jiang, Z.; Che, J.; He, M.; Yuan, F. A Cgru Multi-Step Wind Speed Forecasting Model Based on Multi-Label Specific Xgboost Feature Selection and Secondary Decomposition. *Renew. Energy* 2023, 203, 802–827. [CrossRef]
- 7. Ye, J.; Xie, L.; Ma, L.; Bian, Y.; Xu, X. A Novel Hybrid Model Based on Laguerre Polynomial and Multi-Objective Runge-Kutta Algorithm for Wind Power Forecasting. *Int. J. Electr. Power Energy Syst.* **2023**, *146*, 108726. [CrossRef]
- Kisvari, A.; Lin, Z.; Liu, X. Wind Power Forecasting—A Data-Driven Method Along with Gated Recurrent Neural Network. *Renew. Energy* 2021, 163, 1895–1909. [CrossRef]
- 9. Jin, H.; Li, Y.; Wang, B.; Yang, B.; Jin, H.; Cao, Y. Adaptive Forecasting of Wind Power Based on Selective Ensemble of Offline Global and Online Local Learning. *Energy Convers. Manag.* **2022**, *271*, 116296. [CrossRef]
- 10. Zheng, H.; Hu, Z.; Wang, X.; Ni, J.; Cui, M. Vmd-Cat: A Hybrid Model for Short-Term Wind Power Prediction. *Energy Rep.* 2023, *9*, 199–211. [CrossRef]
- Rajagopalan, S.; Santoso, S. Wind Power Forecasting and Error Analysis Using the Autoregressive Moving Average Modeling. In Proceedings of the Paper presented at the General Meeting of the IEEE-Power-and-Energy-Society, Calgary, AB, Canada, 26–30 July 2009.
- Kaytez, F. A Hybrid Approach Based on Autoregressive Integrated Moving Average and Least-Square Support Vector Machine for Long-Term Forecasting of Net Electricity Consumption. *Energy* 2020, 197, 117200. [CrossRef]
- 13. Yang, Y.; Lu, J. Foreformer: An Enhanced Transformer-Based Framework for Multivariate Time Series Forecasting. *Appl. Intell.* 2022. [CrossRef]
- 14. Xiang, L.; Li, J.; Hu, A.; Zhang, Y. Deterministic and Probabilistic Multi-Step Forecasting for Short-Term Wind Speed Based on Secondary Decomposition and a Deep Learning Method. *Energy Convers. Manag.* **2020**, 220, 113098. [CrossRef]
- 15. Rangel-Martinez, D.; Nigam, K.D.P.; Ricardez-Sandoval, L.A. Machine Learning on Sustainable Energy: A Review and Outlook on Renewable Energy Systems, Catalysis, Smart Grid and Energy Storage. *Chem. Eng. Res. Des.* **2021**, *174*, 414–441. [CrossRef]
- 16. Hu, H.; Wang, L.; Tao, R. Wind Speed Forecasting Based on Variational Mode Decomposition and Improved Echo State Network. *Renew. Energy* **2021**, *164*, 729–751. [CrossRef]
- 17. Khan, M.; He, C.; Liu, T.; Ullah, F. A New Hybrid Approach of Clustering Based Probabilistic Decision Tree to Forecast Wind Power on Large Scales. *J. Electr. Eng. Technol.* **2021**, *16*, 697–710. [CrossRef]
- Yin, H.; Ou, Z.; Huang, S.; Meng, A. A Cascaded Deep Learning Wind Power Prediction Approach Based on a Two-Layer of Mode Decomposition. *Energy* 2019, 189, 116316. [CrossRef]
- 19. Tian, C.; Niu, T.; Wei, W. Developing a Wind Power Forecasting System Based on Deep Learning with Attention Mechanism. *Energy* **2022**, 257, 18. [CrossRef]
- Liu, X.; Yang, L.; Zhang, Z. Short-Term Multi-Step Ahead Wind Power Predictions Based on a Novel Deep Convolutional Recurrent Network Method. *IEEE Trans. Sustain. Energy* 2021, 12, 1820–1833. [CrossRef]
- 21. Hu, H.; Wang, L.; Lv, S.X. Forecasting Energy Consumption and Wind Power Generation Using Deep Echo State Network. *Renew. Energy* **2020**, *154*, 598–613. [CrossRef]
- 22. Lu, P.; Ye, L.; Zhao, Y.; Dai, B.; Pei, M.; Tang, Y. Review of Meta-Heuristic Algorithms for Wind Power Prediction: Methodologies, Applications and Challenges. *Appl. Energy* **2021**, *301*, 117446. [CrossRef]
- Del Ser, J.; Casillas-Perez, D.; Cornejo-Bueno, L.; Prieto-Godino, L.; Sanz-Justo, J.; Casanova-Mateo, C.; Salcedo-Sanz, S. Randomization-Based Machine Learning in Renewable Energy Prediction Problems: Critical Literature Review, New Results and Perspectives. *Appl. Soft Comput.* 2022, 118, 108526. [CrossRef]

- 24. Meng, A.; Chen, S.; Ou, Z.; Ding, W.; Zhou, H.; Fan, J.; Yin, H. A Hybrid Deep Learning Architecture for Wind Power Prediction Based on Bi-Attention Mechanism and Crisscross Optimization. *Energy* **2022**, *238*, 121795. [CrossRef]
- Chen, J.; Zeng, G.Q.; Zhou, W.; Du, W.; Lu, K.D. Wind Speed Forecasting Using Nonlinear-Learning Ensemble of Deep Learning Time Series Prediction and Extremal Optimization. *Energy Convers. Manag.* 2018, 165, 681–695. [CrossRef]
- Xiong, B.; Lou, L.; Meng, X.; Wang, X.; Ma, H.; Wang, Z. Short-Term Wind Power Forecasting Based on Attention Mechanism and Deep Learning. *Electr. Power Syst. Res.* 2022, 206, 107776. [CrossRef]
- Zhen, H.; Niu, D.; Yu, M.; Wang, K.; Liang, Y.; Xu, X. A Hybrid Deep Learning Model and Comparison for Wind Power Forecasting Considering Temporal-Spatial Feature Extraction. *Sustainability* 2022, *12*, 9490. [CrossRef]
- Tascikaraoglu, A.; Uzunoglu, M. A Review of Combined Approaches for Prediction of Short-Term Wind Speed and Power. *Renew. Sustain. Energy Rev.* 2014, 34, 243–254. [CrossRef]
- 29. Wu, Z.; Zeng, S.; Jiang, R.; Zhang, H.; Yang, Z. Explainable Temporal Dependence in Multi-Step Wind Power Forecast Via Decomposition Based Chain Echo State Networks. *Energy* **2023**, *270*, 126906. [CrossRef]
- 30. Yang, S.; Yang, H.; Li, N.; Ding, Z. Short-Term Prediction of 80-88 Km Wind Speed in near Space Based on Vmd-Pso-Lstm. *Atmosphere* **2023**, *14*, 315. [CrossRef]
- Ren, Y.; Suganthan, P.N.; Srikanth, N. A Novel Empirical Mode Decomposition with Support Vector Regression for Wind Speed Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* 2016, 27, 1793–1798. [CrossRef]
- Lv, S.X.; Wang, L. Multivariate Wind Speed Forecasting Based on Multi-Objective Feature Selection Approach and Hybrid Deep Learning Model. *Energy* 2023, 263, 126100. [CrossRef]
- Khazaei, S.; Ehsan, M.; Soleymani, S.; Mohammadnezhad-Shourkaei, H. A High-Accuracy Hybrid Method for Short-Term Wind Power Forecasting. *Energy* 2022, 238, 122020. [CrossRef]
- Duan, J.; Wang, P.; Ma, W.; Tian, X.; Fang, S.; Cheng, Y.; Chang, Y.; Liu, H. Short-Term Wind Power Forecasting Using the Hybrid Model of Improved Variational Mode Decomposition and Correntropy Long Short-Term Memory Neural Network. *Energy* 2021, 214, 118980. [CrossRef]
- Luo, X.; Sun, J.; Wang, L.; Wang, W.; Zhao, W.; Wu, J.; Wang, J.H.; Zhang, Z. Short-Term Wind Speed Forecasting Via Stacked Extreme Learning Machine with Generalized Correntropy. *IEEE Trans. Ind. Inform.* 2018, 14, 4963–4971. [CrossRef]
- Hu, J.; Lin, Y.; Tang, J.; Zhao, J. A New Wind Power Interval Prediction Approach Based on Reservoir Computing and a Quality-Driven Loss Function. *Appl. Soft Comput.* 2020, 92, 106327. [CrossRef]
- Yin, H.; Ou, Z.; Zhu, Z.; Xu, X.; Fan, J.; Meng, A. A Novel Asexual-Reproduction Evolutionary Neural Network for Wind Power Prediction Based on Generative Adversarial Networks. *Energy Convers. Manag.* 2021, 247, 114714. [CrossRef]
- 38. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. IEEE Trans. Signal Process. 2014, 62, 531–544. [CrossRef]
- 39. Chen, W.; Zhuang, J.; Yu, W.; Wang, Z. Measuring Complexity Using Fuzzyen, Apen, and Sampen. *Med. Eng. Phys.* 2009, 31, 61–68. [CrossRef]
- 40. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *AAAI* 2021, *35*, 11106–11115. [CrossRef]
- Barron, J.T. A General and Adaptive Robust Loss Function. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
- Lin, G.; Lin, A.; Gu, D. Using Support Vector Regression and K-Nearest Neighbors for Short-Term Traffic Flow Prediction Based on Maximal Information Coefficient. *Inf. Sci.* 2022, 608, 517–531. [CrossRef]
- Memarzadeh, G.; Keynia, F. A New Short-Term Wind Speed Forecasting Method Based on Fine-Tuned Lstm Neural Network and Optimal Input Sets. *Energy Convers. Manag.* 2020, 213, 15. [CrossRef]
- Zhang, Y.; Tao, P.; Wu, X.; Yang, C.; Han, G.; Zhou, H.; Hu, Y. Hourly Electricity Price Prediction for Electricity Market with High Proportion of Wind and Solar Power. *Energies* 2022, 15, 1345. [CrossRef]

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.