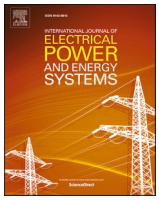
Electrical Power and Energy Systems 162 (2024) 110229

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes



Multi-step prediction of offshore wind power based on Transformer network and Huber loss

Haoyi Xiao [a], Xiaoxia He [a,b,*], Chunli Li [a,b]

[a] College of Science, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China
[b] Hubei Province Key Laboratory of Systems Science in Metallurgical Process, Wuhan University of Science and Technology, Wuhan 430081, China

ARTICLE INFO

Keywords:
Offshore wind power prediction
Transformer network
Huber loss function
Autoencoder
Slime mould optimization algorithm
Multi-step prediction

ABSTRACT

In the context of the burgeoning expansion of renewable energy sources, the precise prediction of offshore wind power assumes a pivotal role in safeguarding the reliability, economic viability, and sustainable progression of offshore wind farms. The present study introduces a novel methodology for offshore wind power prediction, predicated upon the synergy of the Transformer network and Huber loss function. Empirical validation is conducted utilizing authentic data from a European offshore wind farm. The resulting analyses delineate a discernible superiority of the Transformer network over classical LSTM and GRU models in capturing the intricate long-term dependencies intrinsic to the time series. Furthermore, the inclusion of the Huber loss function effectively mitigates the challenges posed by the high volatility often characteristic of offshore wind power data. The study also demonstrates the beneficial integration of autoencoder reconstruction for denoising and slime mould optimization algorithm to augment prediction performance. Distinctively diverging from traditional single-step prediction paradigms, the multi-step prediction model constructed within this research offers a more comprehensive and precise prediction of wind power. Such an innovative approach represents a valuable contribution to the field, with tangible implications for the dependable operation and future advancement of wind power.

1. Introduction

In light of the growing international urgency to find viable alternatives to conventional energy sources, as underscored by the United Nations' Sustainable Development Goal 7 (SDG 7), which aims to ensure access to affordable, reliable, sustainable, and modern energy for all, wind energy has garnered significant attention within scientific and technological domains due to its renewable and non-polluting characteristics. Offshore wind power, as a particularly noteworthy and rapidly advancing segment within this field, has been recognized for its substantial potential and competitiveness as a renewable energy technology [1]. Accurate wind power prediction is fundamentally intertwined with the critical elements of offshore wind farm development, including the reliability, economic viability, and environmental sustainability of these complex energy systems [2].

The nuanced operational and managerial dynamics of offshore wind farms are characterized by a multifaceted and inherently complex interplay of environmental factors. This complexity encompasses spatial and temporal fluctuations in wind velocities, the influence of marine meteorological phenomena, and the seasonally-dependent variability inherent in these systems, thereby rendering the task of offshore wind power prediction a formidable scientific challenge [3]. In light of these considerations, the strategic optimization of wind farm energy output, the enhancement of power grid stability, and the facilitation of a broader transition toward sustainable energy paradigms necessitate the ongoing development and refinement of precise and robust methodologies for offshore wind power prediction [2].

In the annals of scientific literature pertaining to wind power prediction, there has been a notable reliance on traditional statistical methodologies and physical models. Physical models, in this context, are primarily grounded in the utilization of numerical weather prediction (NWP) algorithms tailored for long-term forecasting paradigms [4]. The statistical approaches are anchored in time series forecasting methodologies, including but not limited to frameworks such as Autoregressive Integrated Moving Average (ARIMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and their respective extended models [5]. While these mechanisms have demonstrated appreciable proficiency in forecasting performance within certain domains, they are

* Corresponding author at: College of Science, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China.
E-mail address: hexiaoxia@wust.edu.cn (X. He).

https://doi.org/10.1016/j.ijepes.2024.110229
Received 28 January 2024; Received in revised form 17 August 2024; Accepted 7 September 2024
Available online 21 September 2024

not without limitations. The stringency of distributional assumptions and the necessity for smoothness tests on the data present inherent constraints, thereby circumscribing the generalizability and adaptability of these statistical models across varying contexts and conditions.

In the emergent epoch of technological innovation characterized by the expeditious advancement of machine learning and deep learning paradigms, the field of offshore wind power prediction has witnessed a transformative development. This has been exemplified by the development and application of cutting-edge models such as Support Vector Machine (SVM) [6], Random Forest (RF) [7], and XGboost [8], each tailored to the intricate task of wind speed or wind power prediction. The salient attributes of these machine learning methodologies lie in their robust data processing capacities, augmented predictive precision, and enhanced generalizability across disparate contexts [3]. Despite these strengths, traditional machine learning frameworks encounter inherent difficulties in discerning temporal characteristics and maintaining long-term dependencies, thereby imposing constraints on the realization of further enhancements in predictive accuracy. In contradistinction, the advent of deep learning, such as Long Short-Term Memory Neural Networks (LSTM) [9–11], Gated Unit Recurrent Neural Networks (GRU) [12,13], Extreme Learning Machines (ELM) [14], and Convolutional Neural Networks (CNN) [15,16]—has heralded a new frontier in the field. These models, characterized by their superior ability to learn complex features and navigate nonlinear challenges, have been extensively employed in short-term wind power prediction. Their innovative application has culminated in discernible advancements in forecasting performance, signifying a promising trajectory in the ongoing evolution of wind power prediction techniques.

While the preponderance of contemporary neural network models predicated on the recurrent neural network (RNN) architecture have manifested commendable efficacy in handling sequence data, certain intrinsic challenges persist. Specifically, the employment of RNN-based strategies to model sequence data in a recurrent fashion entails not merely a substantial computational expense in training but also a potential diminution in performance for extended sequence data. This limitation stems from the architectural constraint of RNNs that allows consideration only of the hidden state of the immediately preceding moment within the sequence [17]. In contrast, the emergence of the Transformer network, introduced by Google in 2017 [18], has constituted a seminal advancement in the domain of natural language processing, extending its profound influence into diverse spheres of deep learning applications. Distinctively characterized by its exclusive reliance on self-attention mechanisms, the Transformer network facilitates the establishment of global dependencies on sequence data, thereby enabling the extraction of intricate correlation information across various scales of sequences [19]. The application of the Transformer network within the context of wind power prediction has garnered increased scholarly attention [19–21], and its core multi-attention mechanism—when synergistically integrated with other models—has proven adept at efficiently capturing long-term dependencies and local correlations within time-series data [22–24].

Within the corpus of existing research, there prevails a predominant utilization of the classical sum-of-squares loss function in the context of neural network models. However, this approach has been identified as sensitive to anomalous data points, resulting in reduced stability. The studies delineated in references [25,26] employed the Quantile Regression Neural Network (QRNN) for wind power prediction, thereby amalgamating the merits of quantile regression (QR)—namely, the capability to appraise the conditional distributions of explanatory variables independent of the distributional characteristics of the random variables—with the potent nonlinear fitting faculties inherent in neural networks. Furthermore, reference [27] introduced the Expectile Regression Neural Network (ERNN), a novel neural network architecture equipped with an innovative loss function. This particular design has exhibited increased robustness and stability, particularly when contending with skewed distributions or data sets compromised by outliers [28]. In a parallel vein, the Huber loss function has demonstrated exemplary proficiency in managing outliers and extreme variations, a quality that lends itself advantageously to the inherently volatile nature of offshore wind data [29]. An additional virtue of the Huber loss function is continuity and derivability, a property that augments the stability of the optimization process and facilitates efficient resolution of model parameters through techniques such as gradient descent.

In the continually evolving landscape of offshore wind power research, there has been an observable gravitation towards the exploration of multistep prediction methodologies [20]. This approach, juxtaposed with traditional single-step prediction, extends the predictive scope to encompass a more expansive future time horizon. The resultant effect is an enhancement in the efficiency of power system operation and dispatch, a factor critical to the preservation of grid quality and reliability [25]. Within the context of this manuscript, the focus is deliberately directed towards the implementation of multistep prediction as it pertains to offshore wind power.

The predictive efficacy of a neural network is demonstrably influenced by the intricate interplay of model structure and hyperparameters [4]. To attain an elevated level of predictive precision, scholarly endeavors have commonly fused neural network models with either modal decomposition techniques [20,23,30,31] or optimization algorithms [32,33]. Acknowledging the inherently intermittent, stochastic, and uncontrollable characteristics of offshore wind data, the present study embarks on the introduction of an avant-garde approach to offshore wind power prediction. By integrating the Transformer network with the Huber loss function, the global fitting accuracy and robustness of the prediction can be enhanced. This is then combined with an optimization algorithm, Slime Mould Optimization Algorithm (SMA), to search for its optimal hyperparameters and model structure. An overall framework for multi-step prediction of offshore wind power is constructed, utilizing an open-source multi-variable wind power dataset from a certain European wind farm to validate the superiority of the proposed SMA-Transformer-Huber approach.

This manuscript distinguishes itself by advancing three principal contributions relative to preceding research in this domain:

(1) The Transformer network, which can capture the internal correlations and remote dependencies of longer sequence data, is migrated for wind power prediction, and the Huber loss function is used to incorporate the Transformer network, which shows excellent performance in handling time series data with outliers and outliers.
(2) Addressing the volatility and unpredictability of offshore wind data, the study employs an autoencoder for data reconstruction and denoising [34], and introduces the slime mould optimization algorithm for neural network optimization. This results in enhanced estimation accuracy through fine-tuned hyperparameters and model structure.
(3) Transitioning from single-step to multi-step prediction models, the research offers a more detailed and comprehensive analysis for offshore wind power operation and management, improving the efficiency and reliability of power system operation.

2. Related theories

2.1. Autoencoder

The autoencoder, an unsupervised learning paradigm, serves the intricate function of distilling an efficient low-dimensional representation of voluminous data sets. As depicted in Fig. 1, this intricate neural architecture is bifurcated into two essential constituents: the encoder and the decoder. The encoding segment embodies a mathematical transformation $f : R^d \rightarrow R^h$, responsible for transmuting the input data $x \in R^d$ into a concealed representation $h \in R^h$. Conversely, the decoding

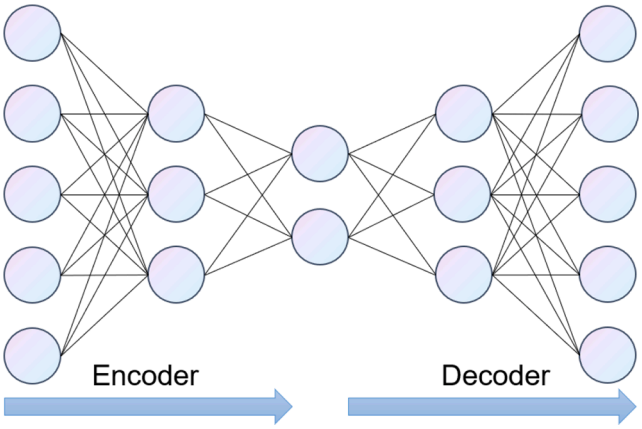


Fig. 1. Schematic diagram of the autoencoder.

segment, represented as the function $g: R^h \rightarrow R^d$, orchestrates the remapping of the concealed representation h back into the originating input space, culminating in the synthesis of a reconstructed input $\widehat{x} = g(f(x))$.

The training objective of the autoencoder is to optimize the parameters by minimizing a loss function $L(x, \widehat{x})$, where $L(x, \widehat{x})$ is the distance between the input x and the reconstructed input $\widehat{x}$. This is usually accomplished by the following mean square error loss function:

$$L(x, \widehat{x}) = (x - \widehat{x})^2 \tag{1}$$

The architecture delineated integrates a sophisticated learning mechanism enabling the autoencoder to assiduously minimize the reconstruction error through the discovery of an efficient representation of the data manifold. By controlling the reconstruction error within a set limit, the autoencoder can effectively identify and highlight the key features inherent in the input data. This imbues the autoencoder with substantial efficacy as an unsupervised learning paradigm, rendering it suitable for an array of complex tasks including, but not limited to, data compression, dimensionality contraction, and noise attenuation. Within the context of this research, a specifically tailored auto-encoder construct is applied to wind power data, an approach designed to faithfully retain inherent characteristics of the data whilst simultaneously implementing robust denoising techniques as far as practicable, thus improving prediction accuracy and reliability.

2.2. Transformer network

2.2.1. Attention mechanism

The Transformer network is a seminal transduction model, uniquely characterized by its exclusive dependence on a self-attention mechanism to synthesize both its input and output representations [20]. The core benefit of the attention mechanism is its ability to extract relevant information from a large amount of input data in the context of the current task. In self-attention, the input sequence $\boldsymbol{X} \in \mathbb{R}^{l \times d}$ is transformed by matrix operations into $\boldsymbol{Q}$(Query), $\boldsymbol{K}$(Key), $\boldsymbol{V}$(Value), l is the sequence length, d is the model dimension.

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_{\boldsymbol{Q}}, \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_{\boldsymbol{K}}, \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_{\boldsymbol{V}}, \tag{2}$$

where $\boldsymbol{W}_{\boldsymbol{Q}} \in \mathbb{R}^{d \times d_{qk}}$, $\boldsymbol{W}_{\boldsymbol{K}} \in \mathbb{R}^{d \times d_{qk}}$, $\boldsymbol{W}_{\boldsymbol{v}} \in \mathbb{R}^{d \times d_v}$ is the weight matrix parameter that the neural network is trained to through iterations, $\boldsymbol{Q} \in \mathbb{R}^{l \times d_{qk}}, \boldsymbol{K} \in \mathbb{R}^{l \times d_{qk}}, \boldsymbol{V} \in \mathbb{R}^{l \times d_v}$ is calculated by the following form to the output of the self-attentive mechanism.

$$\boldsymbol{A} = Softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_{qk}}}\right)\boldsymbol{V}. \tag{3}$$

It is evident that $\boldsymbol{Q}\boldsymbol{K}^T$ contain the information of different positions in the entire sequence, and after normalization, it represents the attention weights for each position. And matrix multiplication with $\boldsymbol{V}$ results in the output of attention $\boldsymbol{A} \in \mathbb{R}^{l \times d_v}$. Finally, the output is transformed through linear transformation as specified in the following form.

$$\boldsymbol{O} = \boldsymbol{A}\boldsymbol{W}_{\boldsymbol{O}}, \tag{4}$$

where $\boldsymbol{W}_{\boldsymbol{O}} \in \mathbb{R}^{d_v \times d_{out}}$ is the linear layer training weight matrix, and the final output is $\boldsymbol{O} \in \mathbb{R}^{l \times d_{out}}$.

Within the Transformer network, the self-attention mechanism is extended to a Multi-head attention mechanism, which is computed in a similar manner. The primary difference is that the input sequence $\boldsymbol{X}$ is divided into n subspaces, and parallel operations of the self-attention mechanism are performed on each subspace. The attention outputs obtained from each head, $\boldsymbol{A}^1, \boldsymbol{A}^2, \cdots\cdots \boldsymbol{A}^n$, are then concatenated, and the final output $\boldsymbol{O}$ is obtained through a linear transformation.

$$\boldsymbol{O} = \boldsymbol{Concat}\left(\boldsymbol{A}^1, \boldsymbol{A}^2, \cdots\cdots \boldsymbol{A}^n\right)\boldsymbol{W}_{\boldsymbol{O}}. \tag{5}$$

Despite the presence of multiple heads, the number of parameters and time complexity remains comparable to that of self-attention [20]. The utilization of multi-head attention allows it to attend to different representation subspaces across different positions, resulting in enhanced forecasting capabilities.

2.2.2. Position encoding

While self-attention considers information from all positions of the sequence data, it may not fully capture the influence of positional differences. To fully utilize the location information of sequence data, this paper incorporates position encoding information into the sequence data. The position encoding is calculated as follows:

$$PE(pos, 2i) = \sin\left(pos/10000^{2i/d}\right), \tag{6}$$

$$PE(pos, 2i+1) = \cos\left(pos/10000^{2i/d}\right), \tag{7}$$

where pos denotes the sequence length index, i respresents the dimensional index from 0 to $d/2$.

2.2.3. Transformer

The architectural schematic of the Transformer model employed in this study is depicted in Fig. 2. The classical Transformer design is bifurcated into two essential components: an encoder and a decoder. However, in the context of this investigation, we have leveraged only the Transformer's encoder structure, which is deemed sufficient to address the regression problem at hand. The encoder serves as a versatile module, ingesting a sequential input and metamorphosing it into an enriched and utilitarian feature representation. Given the Transformer's genesis in the domain of Natural Language Processing (NLP), certain modifications were deemed necessary for its application to current specific context. This included the utilization of a linear layer as a preprocessing step for the input data, supplanting the traditional word vector embedding layer. Additionally, the predictive outcomes are disseminated through a linear layer devoid of an activation function, as opposed to employing a Softmax layer typically reserved for probabilistic predictions. The remaining components, encompassing polytope attention, a pair of normalization layers, a single linear layer, and two residual connections, remain congruent with the original Transformer design.

2.3. Huber loss function

This paper employs Quantile Regression (QR) and Composite Quantile Regression (CQR) for comparative analysis, initially outlining the basic principles of each approach. Given the response variable $\boldsymbol{Y}$ and the covariate matrix $\boldsymbol{X}$, the observations are represented as $(y_i, \boldsymbol{x}_i)$, where i denotes the sample index with $i = 1, 2, \cdots, n$. And n is the total

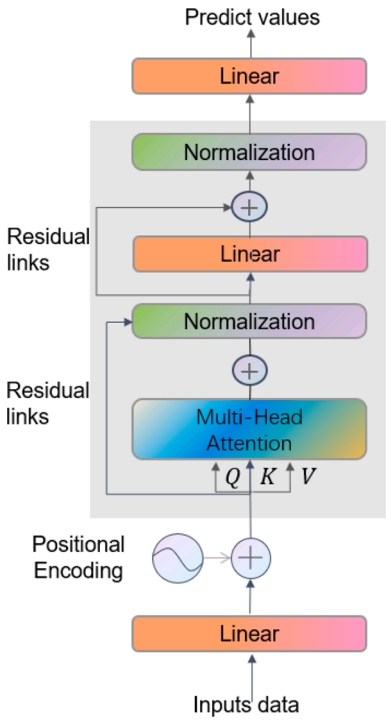


Fig. 2. Schematic representation of the proposed Transformer network.

number of samples. The response variable y_i at the τ-th quantile can be estimated using the classical QR model as follows.

$$\widehat{Q}_{y_i}(\tau|\boldsymbol{x_i}) = \boldsymbol{x_i}'\widehat{\beta}(\tau), i = 1, 2, \cdots, n \tag{8}$$

$$\widehat{\beta}(\tau) = \text{argmin}\sum_{i=1}^{n}\varphi_\tau(y_i - \boldsymbol{x_i}'\beta) \tag{9}$$

$$\varphi_\tau(u) = \begin{cases} \tau u, u \geq 0 \\ (\tau - 1)u, u < 0 \end{cases} \tag{10}$$

The quantile τ within the interval (0,1) represents the quantile of a specified weight level, indicating the degree of asymmetry in the loss function. The τ-th conditional quantile of the response variable y_i, denoted as $Q_{y_i}(\tau|\boldsymbol{x_i})$, and the regression coefficients at the given τ, $\widehat{\beta}(\tau)$ are derived from solving the optimization problem as outlined in Eq. (9). The asymmetric loss function $\varphi_\tau(u)$ is contingent on the quantile τ and is structured to assess the asymmetry of the model's loss function across different τ values.

However, the losses and corresponding parameters computed at different quantile levels τ are not identical, and consequently, the predictive outcomes vary. To address this, Composite Quantile Regression enhances the approach by calculating and averaging losses across a spectrum of distinct τ values. For a given positive integer K, set $\tau_k = \frac{k}{K+1}$, $k = 1, 2, \cdots K$. The loss function of the corresponding composite quantile is as follows:

$$\widehat{w}(K) = \text{argmin}\frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{n}\varphi_\tau(y_i - f(\boldsymbol{x_i}, w(\tau_k)\,)) \tag{11}$$

The loss function for the composite quantile neural network is thus formulated to average the losses under various τ_k, enhancing the robustness and stability of the parameter estimates.

The Huber loss function, frequently utilized in regression analyses, offers a robust solution particularly in scenarios characterized by the presence of an excessive number of outliers. This loss function amalgamates the properties of squared loss and absolute loss, rendering it sensitive to errors akin to the squared loss when the error is minimal and demonstrating insensitivity to outliers comparable to the absolute loss when the error is substantial. Consequently, Huber's loss enjoys a versatile range of practical applications. Mathematically, the Huber loss function is defined for an error between the predicted value $\widehat{y}$ and the actual value y as:

$$L_\delta(y, \widehat{y}) = \begin{cases} \frac{1}{2}(y - \widehat{y})^2 |y - \widehat{y}| \leq \delta \\ \delta|y - \widehat{y}| - \frac{1}{2}\delta^2 \text{otherwise} \end{cases} \tag{12}$$

Within this expression, δ represents a hyperparameter that demarcates the threshold for the transition between the squared loss and absolute loss regimes. In the context of this study, the value of δ is set to 1. This parameterization implies that when the prediction error is less than or equal to 1, the loss function assumes a squared loss form, which aids in minimizing the error magnitude. Conversely, when the error exceeds 1, the loss function transitions to a linear loss, which is advantageous for mitigating the impact of outliers. This value for δ is commonly adopted in prior research and is known to strike an optimal balance between accuracy and robustness [36].

A neural network can be conceptualized as a nonlinear function, symbolized by $f(\bullet)$, representing a generalized form of the nonlinear model. Given an input x_i, the corresponding model estimate is articulated as:

$$\widehat{y}_i = f(x_i, \boldsymbol{w}_\delta), i = 1, 2, \cdots, n \tag{13}$$

Within this context, $\boldsymbol{w}_\delta$ delineates the generalized model parameter requiring estimation. This encapsulates the architecture of the neural network as well as the weight and bias matrices spanning all layers. The estimation for this model is procured iteratively by minimizing the loss as described in Eq. (14):

$$\widehat{\boldsymbol{w}}_\delta = \text{argmin}\sum_{i=1}^{n} L_\delta(y_i, \widehat{y}_i), \tag{14}$$

Here, L_δ corresponds to the expression found in equation (12). Notably, the Huber loss function transitions between utilizing a squared loss for smaller errors and an absolute loss for larger errors. This dual nature equips the Huber loss function with enhanced resilience against outliers in comparison to the conventional squared loss, especially when engaging with datasets replete with outliers. Furthermore, the loss function maintains derivability across its entire domain, facilitating the estimation of model parameters. This enables the application of standard backpropagation techniques and gradient descent optimization algorithms within neural networks to proficiently ascertain the optimal parameters $\widehat{\boldsymbol{w}}_\delta$.

2.4. Slime mould algorithm

The Slime Mould Optimization Algorithm (SMA) represents a heuristic optimization approach, the inspiration for which is drawn from the foraging behavior exhibited by natural slime molds, as characterized in reference [35]. This unicellular biological entity engages in a search for nutrients through a combination of chemical diffusion and cellular movement, simultaneously avoiding harmful substances. By modeling this behavior, the SMA provides a means to navigate the solution space in search of optimal solutions with greater efficiency in comparison to traditional optimization algorithms such as genetic algorithms or particle swarm algorithms. The foraging behavior of slime molds can be methodically partitioned into three primary stages: 1) the approach to food sources, 2) the encirclement of these sources, and 3) the eventual acquisition of nourishment. Such predatory behavior is mathematically mirrored in the subsequent model:

$$X(t+1)=\begin{cases} rand\times(ub-lb)+lb, r<z \\ X_b(t)+vb(W\times X_A(t)-X_B(t)), r<p \\ vc\times X(t), r\geq p \end{cases} \quad (15)$$

$$p=tanh(|S(i)-DF|), \quad (16)$$

Let $X(t+1)$ and $X(t)$ denote the positions of a particular slime mold entity at iterations $t+1$ and t, respectively. $X_b(t)$ refers to the location of the slime mold individual corresponding to the highest food concentration at the t-th iteration. This position serves as a guide, enabling the searching individual to continually refine its location according to $X_b(t)$. Additionally, $X_A(t)$ and $X_B(t)$ symbolize the positions of two randomly chosen slime mold entities at the t-th iteration, while ub and lb define the upper and lower boundaries of the exploration space, respectively. Random value r is generated within the range $[0, 1]$, and z serves to switch between global and local search phases, signifies the proportion of randomly dispersed slime mold individuals relative to the overall population, and p modulates the method of updating the positions of slime molds. Furthermore, $S(i)$ represents the fitness value corresponding to the current individual i position, N is indicative of the population size, and DF is the optimal individual position value obtained from the best fitness values during the iterative process.

Slime molds utilize wave propagation to modify the velocity of cytoplasmic flow within their veins. The alterations in vein width and the oscillation frequency of the biological oscillator are simulated through parameters vb, vc and W. This modeling allows the slime molds to approach food at a slower rate when the concentration is low and accelerate when higher-quality nourishment is detected. The value of vb lies within the range $[-a, a]$, and vc diminishes linearly from 1 to 0. The parameter a is computed using the subsequent equation:

$$a=arctanh(1-(t/T)), \quad (17)$$

t symbolizes the current iteration number, and T stands for the maximum number of iterations permitted.

The variable W embodies the weight of the slime mold, mimicking the oscillation frequency of the organism at varying food concentrations. This parameterization aids the slime mold in approaching nourishment more rapidly when high-quality sustenance is detected, thereby enhancing the efficiency of the search process. The associated formula can be expressed as:

$$W(SortIndex(i))=\begin{cases} 1+r_1\times\log\left(\frac{bF-S(i)}{bF-wF}+1\right), condition \\ 1-r_1\times\log\left(\frac{bF-S(i)}{bF-wF}+1\right), others \end{cases} \quad (18)$$

$$SortIndex=sort(S), \quad (19)$$

Within this context, the term *condition* refers to an individual whose fitness value is ranked within the upper half of the population. This concept emulates the slime molds' behavior of adjusting their position in response to food concentration; if this concentration is low, the slime molds will redirect their search toward other regions. Further, bF and wF denote the optimal and worst fitness values in the current iteration, respectively, and *SortIndex* represents the ordering of the sorted fitness values. Additionally, r_1, a random value generated within the range $[0, 1]$, is employed to simulate the uncertainty inherent in the contraction pattern of Mucor's veins. The specific procedural details and flow of the Mucor algorithm, including the manner in which these variables and parameters interact to govern the slime mold's behavior, are graphically delineated in Fig. 3.

In the research presented within this manuscript, the hyperparameters of the neural network model are identified as the search parameters. The global architecture of the neural network model is employed as the objective function, with the goodness-of-fit metric pertaining to the model's predictive efficacy on the test set data functioning as the adaptive value. The process of optimization is directed towards the maximization of this adaptive value, constituting the primary objective in the search for the model's optimal hyperparameter configuration.

2.5. Multi-step prediction structure

In the context of this research, a multivariate dataset pertaining to

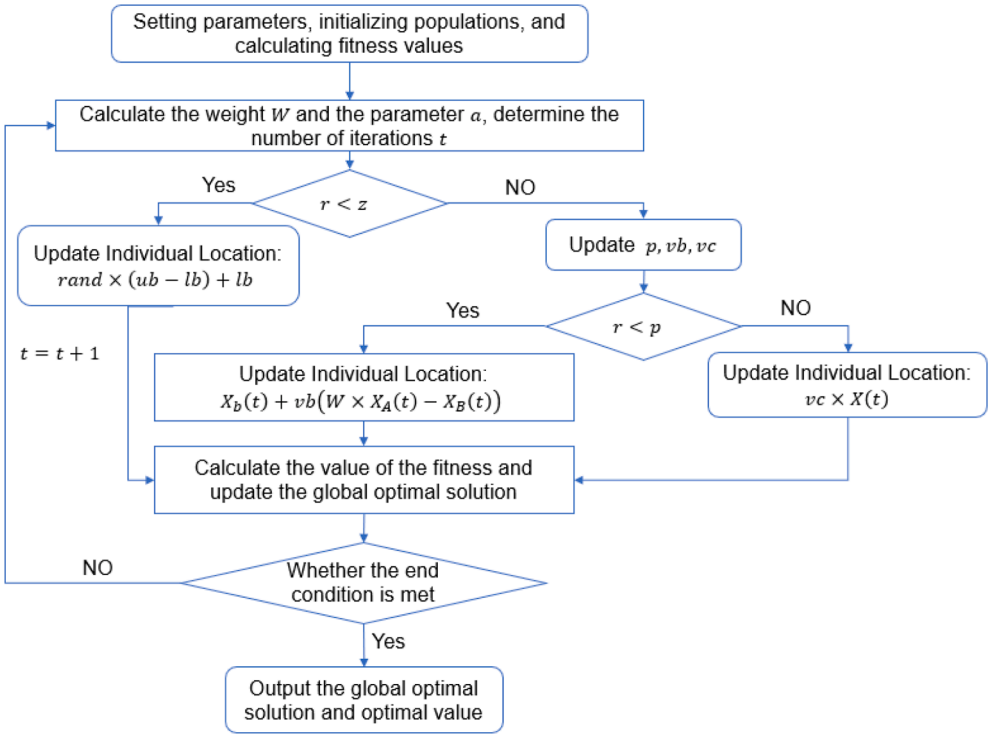


Fig. 3. Slime Mould algorithm overall process.

offshore wind power is employed, and the forecasting objective is specifically directed towards offshore wind power. The sliding window within the study is configured with a time step of 144, with a predictive time step of 24, signifying that the multivariate time series data from the preceding 144 intervals $(\boldsymbol{x}_{t-144}, \boldsymbol{x}_{t-143}, \cdots, \boldsymbol{x}_{t-1}, \boldsymbol{x}_t)$ are utilized to forecast the subsequent 24 periods of wind power series data $\left(x_{t+1}^{(8)}, x_{t+2}^{(8)}, \cdots, x_{t+24}^{(8)}\right)$. Within this construct, the vector $\boldsymbol{x}_t = \left[x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(8)}\right]$ represents eight characteristic variables at time t, with $x_t^{(8)}$ denoting the eighth characteristic variable at that moment, namely, wind power.

The aggregate feature variable$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n]$ manifests as a three-dimensional matrix with dimensions corresponding to the sample size (n), time step (w), and number of features (p). Fig. 4 illustrates the composition of the feature variable X_i(i = 1,2…n) for an individual sample and its corresponding response variable Y_i.

This vectorized output approach for multi-step prediction does not necessitate the integration of newly predicted values from the algorithm into the feature variables for subsequent predictions. Since all the feature variables consist of actual historical values, this methodology significantly mitigates error, resulting in predictions that are more congruent with true values. As a consequence, this mechanism can furnish precise and comprehensive predictions of offshore wind power for future intervals.

3. Methodology framework and evaluation metrics

3.1. Methodology framework

The comprehensive workflow developed in this study for the prediction of offshore wind power is illustrated in Fig. 5, consisting of several meticulously defined steps:

(1) Initial preprocessing of offshore wind power data involves reconstructing and denoising the data utilizing an autoencoder. Subsequently, the data are partitioned into training and test sets, normalized, and transformed into feature variables and response variables for multi-step prediction through the application of a time-series sliding window methodology.
(2) A total of five neural network models, namely MLP, RNN, LSTM, GRU, and Transformer, are paired with four distinct loss functions (MSE, QR, CQR, Huber) to create various models tailored for the single-step prediction of offshore wind sequences. Comparative analysis of model prediction performance is conducted using four prevalent evaluation metrics for point estimation, which include MAE, RMSE, MAPE, and R^2. The objective of this phase is to identify the optimal neural network model and associated loss function, which will be employed in the ensuing multi-step prediction.
(3) The multi-step prediction of offshore wind power is executed utilizing the structure and hyperparameters of the optimal neural network and loss function, as determined through the application of the SMA. An analytical comparison is performed on the prediction outcomes generated by different models across various time steps. SMA is employed to optimize the hyperparameters of the Transformer model, including parameters such as the number of neurons per layer, batch size, number of training epochs, and the number of multi-attention heads. These parameters correspond to the location parameters in the slime mold optimization algorithm. The discrepancy between the predicted and actual values from the neural network on the training set is used as the objective function. SMA is utilized to search for the optimal combination of hyperparameters in the solution space, aiming to minimize the loss and thereby maximize the prediction performance. The optimization process initiates with a set of randomly chosen hyperparameters, followed by simulating the slime mold's adaptive behavior to environmental resources through the position updating mechanism during the iterative process, thereby identifying the hyperparameter configuration that results in the minimal loss.

3.2. Evaluation metrics

In order to rigorously assess the prediction efficacy of the various models, this study employed four evaluation metrics that are commonly utilized and considered highly reliable in the context of regression analyses. These metrics include the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). The mathematical expressions defining these metrics are delineated in Eqs. (20) through (23).

$$MAE = \frac{1}{n}\sum_{t=1}^{n} \|y_i - \hat{y}_i\|, \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_i - \hat{y}_i)^2}, \tag{21}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left\|\frac{y_i - \hat{y}_i}{y_i}\right\|, \tag{22}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{t=1}^{n}(y_i - \bar{y})^2}, \tag{23}$$

where n represents the number of predicted samples, y_i denotes the ith true value of the response variable, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean value of the real data.

4. Empirical results

4.1. Data sources and pre-processing

This study utilizes a publicly accessible dataset originating from a European wind farm [9], available online at (https://zenodo.org/). Situated within Europe, the offshore wind farm from which the data is derived encompasses measurements from a meteorological mast and two offshore turbines. The data, representing the offshore wind, is bifurcated into two distinct datasets, labeled as WT5 and WT6,

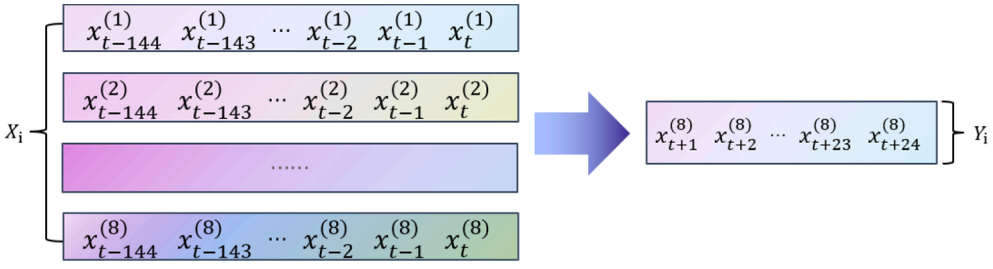


Fig. 4. Depicts the structure of both feature and response variables.

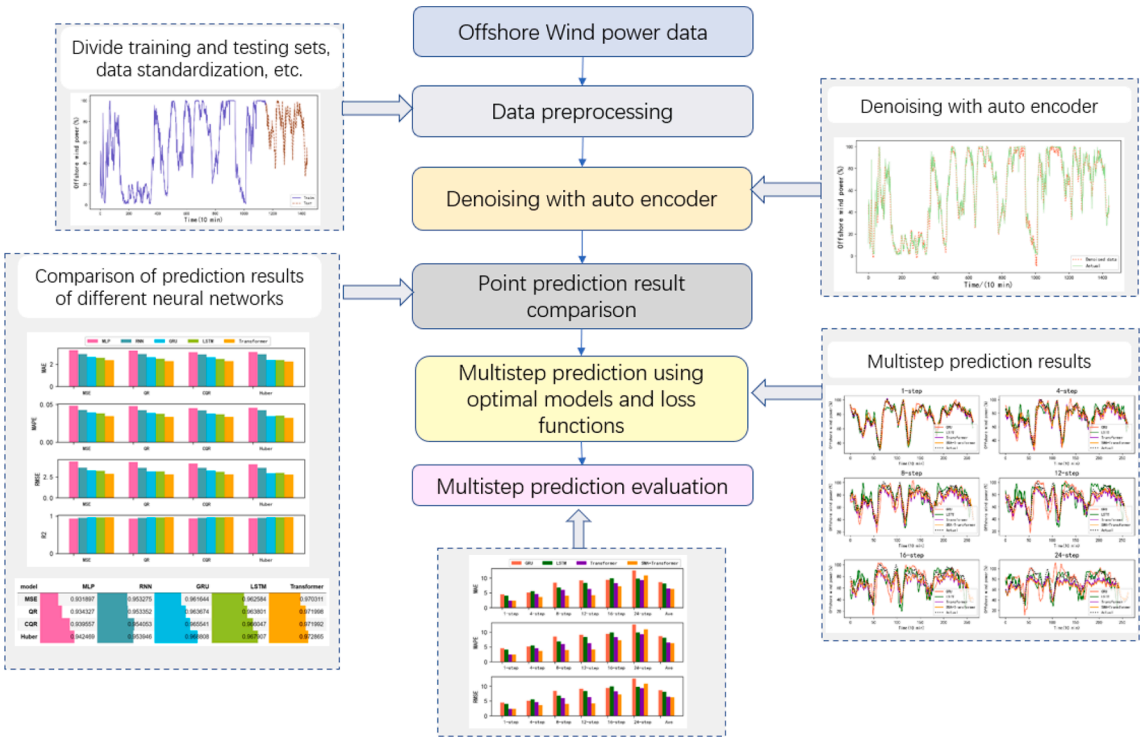


Fig. 5. The framework of the offshore wind power prediction.

encompassing observations collected from January 1, 2009, through December 31, 2009, at 10-minute intervals. Table 1 delineates the specific nomenclature and the corresponding interpretations of various variables included within the dataset.

We selected wind generation data from January 1 to February 28 for WT5 offshore turbines. The frequency window of the original data is 10 min with a total of 1440 sample points. The post-processed data after dividing the training and test data is shown in Fig. 6.

In order to observe outliers with large fluctuations in the data, we compare each data to the range of means and variances of its neighbors, and if a data is outside two times the variance of the mean of the data surrounding it (window size 48), we define it as an outlier and mark it as a yellow dot, as shown in Fig. 7.

An examination of the sequence in Fig. 6 and Fig. 7 elucidates the volatile and stochastic nature of the offshore wind power data. Consequently, a multi-step prediction of offshore wind power becomes imperative to articulate the uncertainty inherent in wind power output, thereby furnishing valuable insights for relevant decision-makers and stakeholders. The application of autoencoder in this study serves to denoise the offshore wind power data, with the resultant denoising effect illustrated in Fig. 8.

An inspection of Fig. 8 reveals that the denoised data aligns closely with the original values within the monotonic intervals, while demonstrating increased smoothness in regions characterized by volatility, thus achieving the intended denoising effect. For the purposes of this study, the dataset was partitioned into two segments: 80 % of the data, represented by the purple solid line, was designated as the training set, while the remaining 20 %, delineated by the brown dashed line, constituted the test set. The feature variables were constructed using a sliding window encompassing 144 periods (equivalent to one day), with the subsequent 24 periods (or 4 h) forming the response variables. Following the completion of this procedure, the resulting 3D tensor data from both training and test sets were normalized to facilitate the fitting of the neural network model. The principal parameters governing the neural network model are detailed in Table 2.

Table 1
Data set variables.

Variable	Range
Wind Speed(V)	
Wind Direction (D)	$[0^{\circ}, 360^{\circ}]$
Air Density (ρ)	\
Humidity (H)	$[0, 100\%]$
Turbulence Intensity (I)	
Wind Shear (S)	$S1, S2$
Relative power (P)	Ratio of actual to rated power

4.2. *Single-step prediction results*

In the assessment of point estimation prediction outcomes employing classical single-step prediction methodologies, the present study utilizes five conventional neural network models. These models are compared based on their predictions across four distinct loss functions and are evaluated using four metrics: MAE, RMSE, MAPE, and R^2.

An initial examination is conducted of the comparative effects of various models employing Huber loss function, as depicted in Fig. 9. Concurrently, Fig. 10 illustrates the prediction graph of the Transformer model utilizing different loss functions.

The predictions derived from the Transformer model, as represented in Figs. 9 and 10, align closely with the true values. The application of the Huber loss function appears to be particularly effective, with the models exhibiting highly accurate predictions during intervals of monotonically ascending or descending wind data, and demonstrating greater bias in predictions within regions characterized by oscillating and diverse wind power.

A comprehensive error metric analysis was performed for all models on the test set, the results of which are tabulated in Table 3, with corresponding bar charts presented in Fig. 11. For enhanced visual comparison of the model effects, the bar charts for R^2 are separately depicted in Fig. 12.

An analysis of the results delineated in Table 3 and Fig. 12 yields the following insights: (1) Of all the models under consideration, the

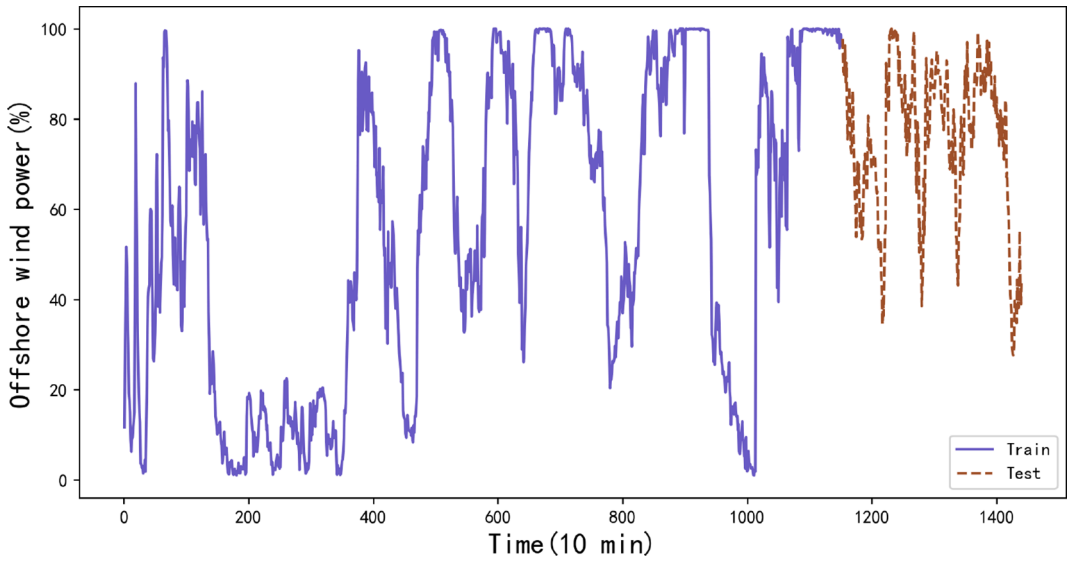


Fig. 6. Offshore wind power timing diagram.

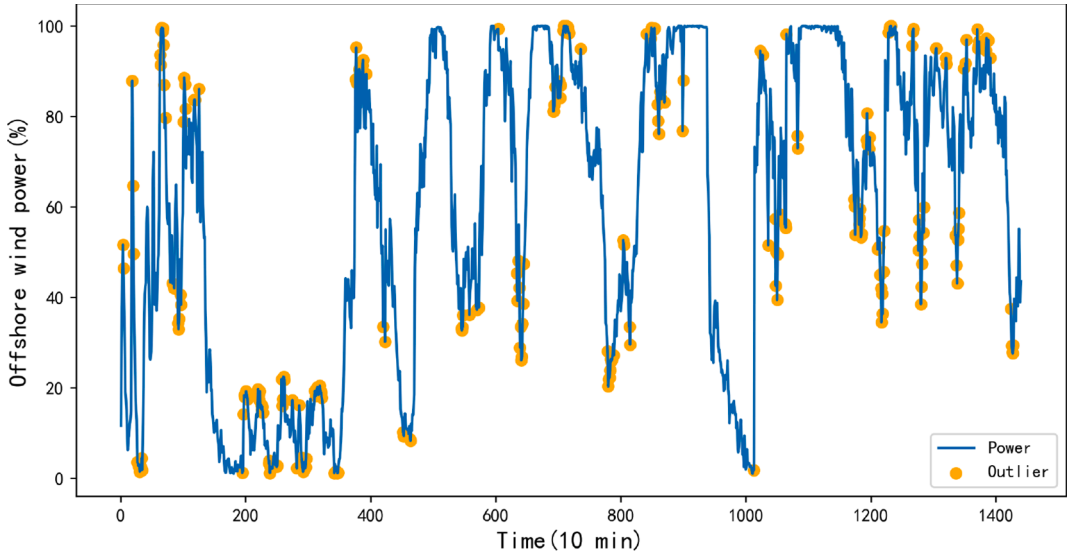


Fig. 7. Offshore wind power timing map with outliers.

Transformer model demonstrates superior prediction performance on wind power data. In terms of this particular sequence data, the Transformer's proficiency in discerning the internal correlation within longer sequential data outperforms that of the RNNs of networks. Based on established principles of deep learning, this effect is anticipated to become increasingly pronounced with the augmentation of training data volume. (2) Within this specific sequence modeling context, independent of the loss function and evaluative index employed, the Transformer consistently outperforms other neural networks. Subsequent rankings include LSTM and GRU, which display roughly equivalent efficacy, followed by RNN. MLP is found to be the least effective, owing to its failure to incorporate the temporal factor. (3) A comparative analysis of loss functions reveals that the Huber loss function exhibits better performance across most neural network models, succeeded by CQR. This illustrates its suitability for highly stochastic and volatile offshore wind data. Conversely, the MSE loss function demonstrates suboptimal performance in a majority of models, likely attributable to its heightened sensitivity to outliers.

4.3. Multi-step prediction of results

In the subsequent phase of multi-step prediction, a uniform application of the Huber loss function is implemented. Owing to their relatively inferior efficacy, the MLP and RNN models are excluded from further examination. Instead, a comparative analysis of multi-step predictions is conducted utilizing the LSTM, GRU, and Transformer models, including a variant of the Transformer model optimized via the SMA. The assessment specifically encompasses error metrics associated with 1-step, 4-step, 8-step, 12-step, 16-step, and 24-step predictions, the details of which are tabulated in Table 4. Accompanying bar charts are depicted in Fig. 13, and the predictive outcomes are graphically represented in Fig. 14.

An examination of Table 4, Fig. 13, and Fig. 14 reveals distinct insights into the comparative performance of various models (GRU, LSTM, Transformer, SMA+Transformer) across different time steps ($t + 1$, $t + 4$, $t + 8$, $t + 12$, $t + 16$, $t + 24$) for multi-step prediction. The ensuing analysis yields the following observations.

The comparative efficacy of different models, namely Transformer, LSTM, and GRU, in multi-step prediction and single-step prediction

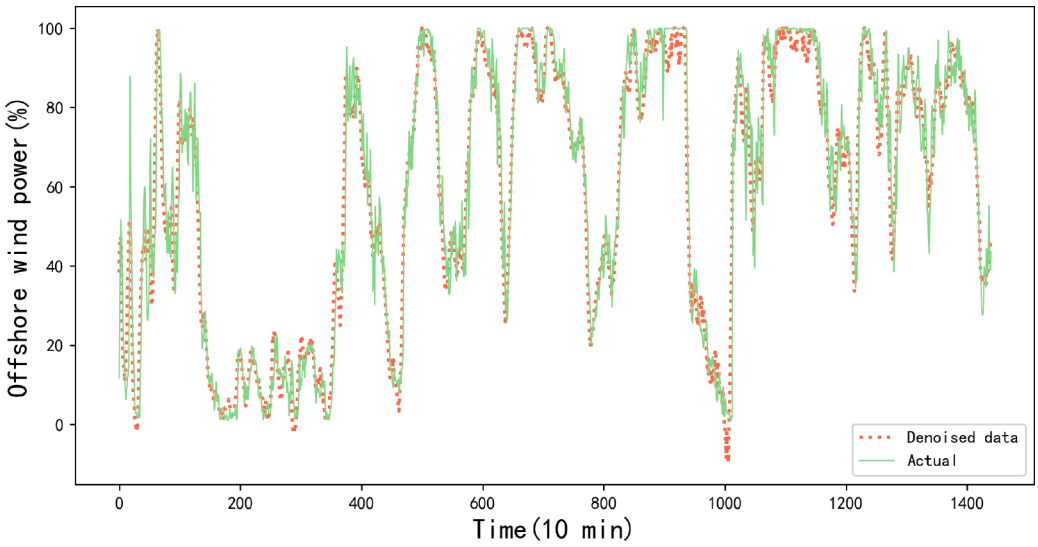


Fig. 8. Timing diagram of offshore wind power after denoising.

Table 2
The main parameters of the neural network models.

Parameter	Value
Number of hidden layers	2
Number of neurons in the hidden layer	[64, 32]
Batch size	32
Maximum number of iterations	50
Embedding layer dimension	32
Number of Multi-head attention heads	4
Parameters of Huber loss function δ	1

follows a consistent pattern. The Transformer model demonstrates the highest effectiveness, followed by the LSTM, while the GRU ranks last. Specifically, the GRU exhibits a pronounced weakness in the multi-step prediction task, characterized by a progressive increase in error commensurate with the increment in prediction steps. While the LSTM surpasses the GRU in terms of accuracy, it remains suboptimal compared to the Transformer. This relative inefficiency becomes more acute in long-term predictions, such as at t + 16 and t + 24 time steps, where the error exhibits a marked amplification. The Transformer model excels in both short- and medium-term forecasting (e.g., from t + 1 to t + 12), yet its effectiveness wanes in extended or forward forecasting scenarios (e.g., t + 16 and t + 24). A hybrid model, the SMA+Transformer, exhibits robust performance across all temporal stages, particularly distinguishing itself in the medium-term forecasting domain, where it records comparatively lower error.

The diminishing efficacy of all the considered models in terms of prediction accuracy becomes manifest as the temporal scope of the prediction enlarges. In the immediate prediction frame (e.g., at t + 1), both the SMA+Transformer and the standalone Transformer models yield the most satisfactory outcomes, as evidenced by their relatively low values of MAE, MAPE, and RMSE. Within the medium-term prediction span (e.g., from t + 4 to t + 16), the Transformer continues to demonstrate robust performance. However, the SMA+Transformer begins to reveal its superiority, particularly in its progressively declining RMSE. In the more distal or forward prediction phase (beyond t + 20), the divergence between the SMA+Transformer and Transformer models starts to contract. The SMA+Transformer's error exhibits a relatively higher growth rate, a phenomenon that might be attributed to a mild overfitting occurrence spurred by the hyperparameter optimization process executed through the SMA. An overarching assessment indicates the efficacy of the SMA optimization in the context of Transformer network hyperparameters. The amalgamated SMA+Transformer model

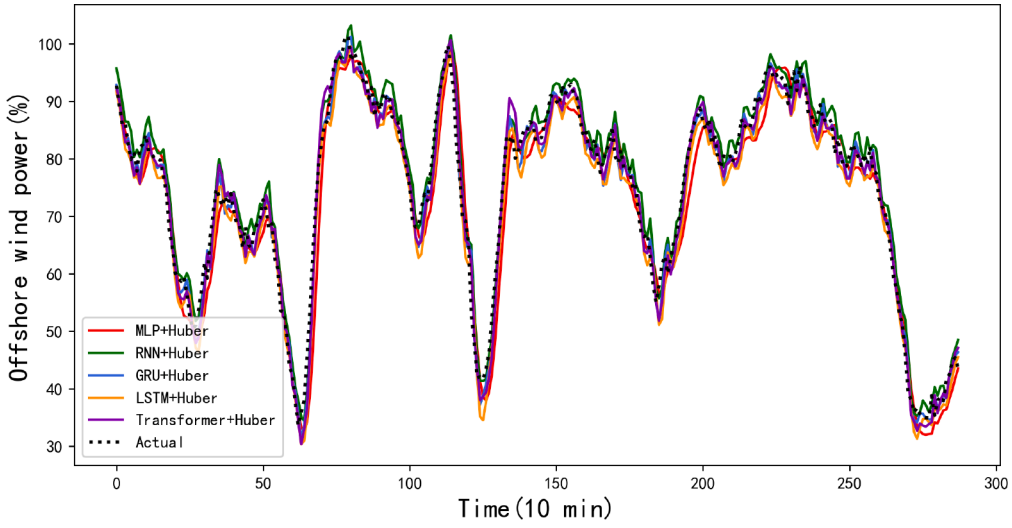


Fig. 9. Comparison of predictions of different models using Huber loss function.

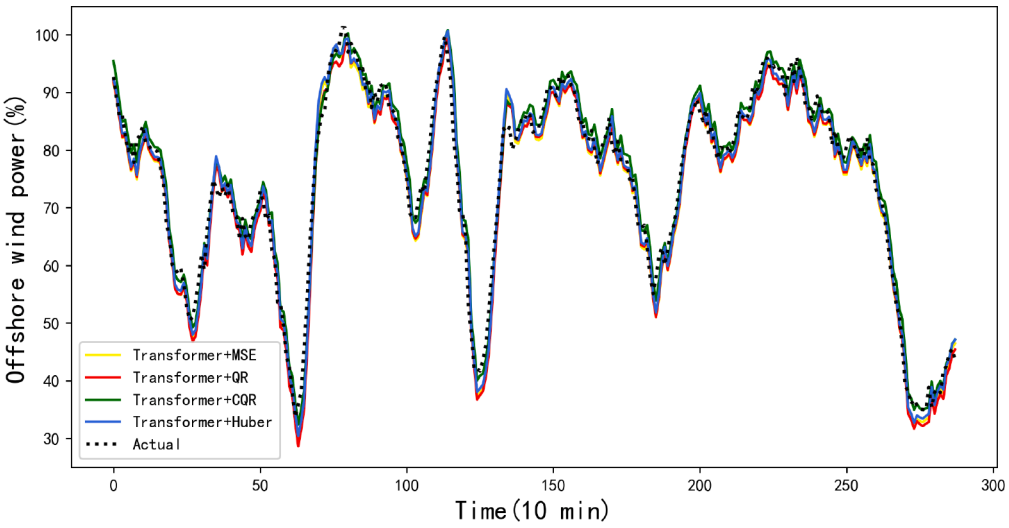


Fig. 10. Comparison of Transformer network predictions using different loss functions.

Table 3
Comparison of evaluation indicators for single-step prediction.

Index	Model	MSE	QR	CQR	Huber
MAE	MLP	3.283734	3.244126	3.100303	3.106675
	RNN	2.922517	2.920182	2.900095	2.901285
	GRU	2.659288	2.647061	2.642061	2.389519
	LSTM	2.589331	2.495637	2.489224	2.38035
	Transformer	2.366972	2.285821	2.285954	**2.248485**
MAPE	MLP	0.048181	0.04769	0.045012	0.045526
	RNN	0.042545	0.042361	0.042152	0.042245
	GRU	0.03943	0.038461	0.037855	0.034408
	LSTM	0.038094	0.037566	0.036936	0.035161
	Transformer	0.034396	0.033343	0.033345	**0.032216**
RMSE	MLP	4.364467	4.285902	4.111726	4.011451
	RNN	3.61515	3.591561	3.589046	3.589073
	GRU	3.275415	3.165892	3.256494	2.953752
	LSTM	3.235048	3.158697	3.131523	2.99608
	Transformer	2.881709	2.798627	2.798912	**2.776425**
R^2	MLP	0.931897	0.934327	0.939557	0.942469
	RNN	0.953275	0.953352	0.954053	0.953946
	GRU	0.961644	0.963674	0.965541	0.968808
	LSTM	0.962584	0.963801	0.966047	0.967907
	Transformer	0.970311	0.971998	0.971992	**0.972865**

records the least mean value of prediction errors across a 24-step prediction horizon.

4.4. Elia offshore wind power multi-step prediction case

In order to ensure that the proposed method can still accurately predict wind power generation in different situations such as different regions and different seasons, the same four sets of models, LSTM, GRU, Transformer and Transformer optimized using SMA, are used to compare the multistep prediction with the data from Elia offshore wind farms. The data was obtained from this website (https://www.elia.be/en/grid-data/generation-data/wind-power-generation) and originates from the Belgian operator of the high-voltage transmission grid. Elia Group, one of Europe's five largest transmission system operators, provides these data as part of its operations in Belgium and Germany through its subsidiaries, while also advancing the offshore grid development through initiatives like WindGrid. The data not only facilitate grid operation optimization but also provide empirical evidence for energy policy formulation and electricity market analysis, thereby supporting the broader goals of energy transition and grid modernization.

In this thesis, we specifically focus on the offshore wind farm dataset provided by Elia, utilizing data from January 1, 2024, to February 28, 2024, for our experiments. The training set consists of data up to February 15, 2024, while the test set includes data from February 16, 2024, onward. We present results for 1-step, 4-step, 8-step, 12-step, 16-step, and 24-step models. The error indices calculated for the 12-step, 16-step, and 24-step predictions are summarized in Table 5, with corresponding histograms displayed in Fig. 15, and the prediction results illustrated in Fig. 16.

From Table 5 and Figs. 15 and 16, a comparative analysis of the performance of various models (GRU, LSTM, Transformer, SMA+-Transformer) across different forecast horizons (t + 1, t + 4, t + 8, t + 12, t + 16, t + 24) is presented for multistep prediction. The ensuing analysis yields consistent conclusions as outlined below.

1. Overall Model Performance Analysis. The GRU model does not exhibit a competitive edge, with its predictive performance falling short of the LSTM. Meanwhile, the Transformer model generally surpasses both in certain metrics at specific time steps but does not match the SMA+Transformer in long-step predictions and overall performance. The SMA+Transformer model surpasses the other models in all three evaluation metrics (MAE, MAPE, and RMSE) in terms of average performance. This outcome implies that the integration of the SMA component may have enhanced the Transformer's ability to effectively manage the noise and fluctuations inherent in time series data, thereby elevating the precision of its forecasts.
2. Short-term and Long-term Forecasting Analysis. In the realm of short-term forecasting (t + 1, t + 4), the SMA+Transformer model demonstrates marked superiority, particularly in the MAE and RMSE metrics. This indicates the model's efficacy in processing immediate data, effectively capturing the essential elements influencing short-term outputs. In contrast, during long-term forecasts (t + 12, t + 16, t + 24), despite an overall rise in prediction errors across all models, the SMA+Transformer model sustains lower errors, most notably in MAE and RMSE.
3. Model Stability Analysis. The performance of GRU and LSTM exhibits considerable variability across different forecast horizons, particularly evident in long-term predictions. For instance, at the t + 24 time step, the RMSE values for both GRU and LSTM are higher than those for other models. This variability suggests that traditional recurrent neural networks may encounter difficulties when managing longer sequence predictions. In contrast, the SMA+Transformer model demonstrates more consistent performance across various time steps, indicating reduced volatility.

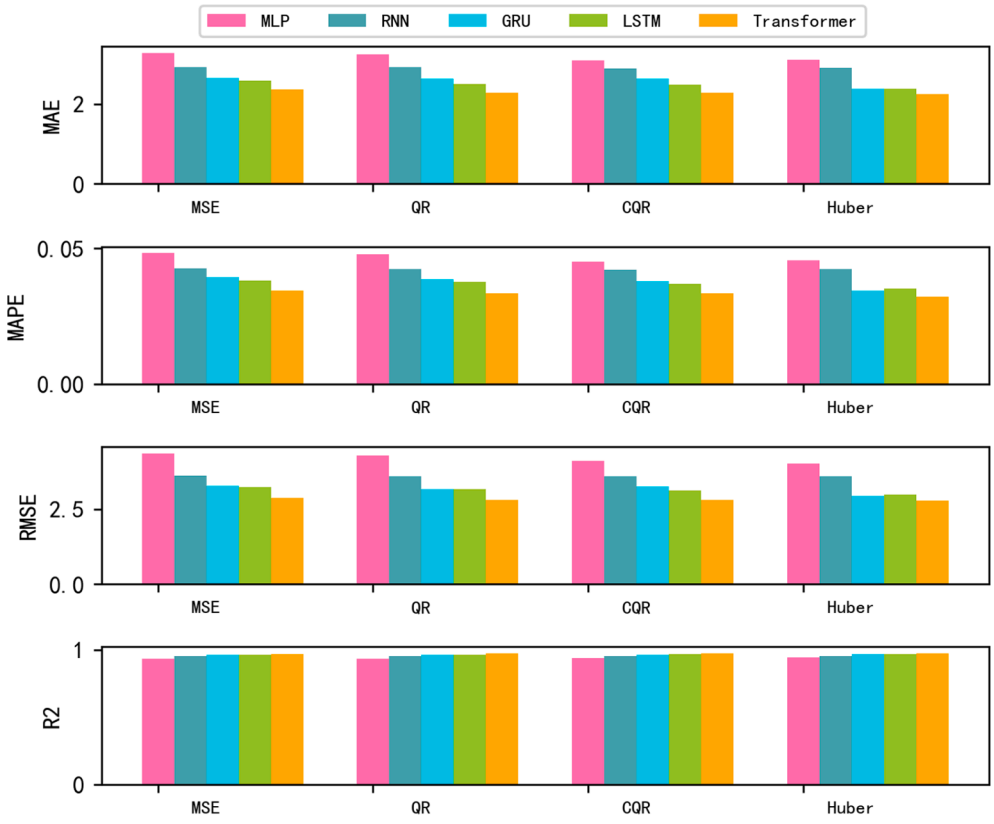


Fig. 11. Comparison of evaluation indicators for single-step prediction.

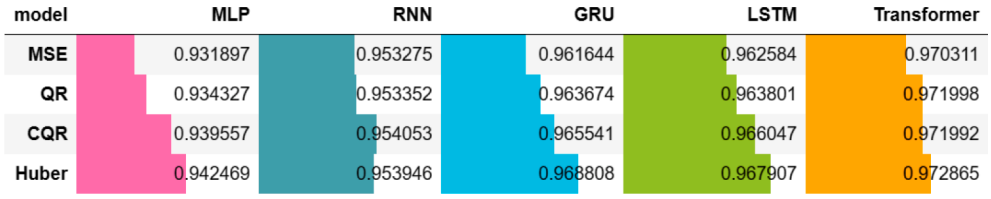


Fig. 12. R^2 for different models and different loss functions.

Table 4
Comparison of evaluation indicators for multi-step prediction.

Time	index	GRU	LSTM	Transformer	SMA+Transformer
t + 1	MAE	4.567153	4.052457	2.412578	2.415152
	MAPE	0.063697	0.055646	0.033268	0.033538
	RMSE	5.883398	4.992293	2.960626	3.030939
t + 4	MAE	5.121458	5.544084	4.619325	3.665646
	MAPE	0.072761	0.077095	0.063121	0.049933
	RMSE	6.480977	6.945328	5.684424	4.557756
t + 8	MAE	8.475906	6.811311	6.011629	3.990305
	MAPE	0.116736	0.095536	0.080425	0.053048
	RMSE	10.82858	8.793484	7.194181	4.849639
t + 12	MAE	9.15376	8.349009	6.313374	4.211127
	MAPE	0.127874	0.117809	0.082109	0.05565
	RMSE	12.27364	10.98287	7.292775	4.998936
t + 16	MAE	9.424327	9.890189	8.299239	7.219986
	MAPE	0.134764	0.14152	0.109366	0.093861
	RMSE	12.88688	12.9596	9.480932	8.220265
t + 24	MAE	12.57395	9.853952	9.319692	9.89215
	MAPE	0.191369	0.141354	0.121958	0.138719
	RMSE	17.06191	12.6531	10.63862	11.33011
Ave	MAE	8.652382	8.111007	6.47678	**6.279147**
	MAPE	0.125016	0.114672	0.085672	**0.081546**
	RMSE	11.56664	10.47268	7.572995	**7.255083**

In summary, the SMA+Transformer model excels in the multi-step forecasting of Elia offshore wind power, particularly notable for its robustness and accuracy in both short-term and long-term forecasts.

5. Conclusion

In the context of this research, a novel methodology for offshore wind power prediction is introduced, employing a Transformer network model, Huber loss function, and the optimization of hyperparameters through Slime Mould algorithm. Distinct from conventional single-step prediction techniques, the present work incorporates a multi-step prediction framework, computing distinct evaluation metrics to offer a comprehensive understanding of offshore wind power uncertainty. Data from a European offshore wind turbine and Elia offshore wind data were selected for validation and testing, yielding several noteworthy conclusions.

(1) Offshore wind power data is inherently characterized by randomness and volatility. The model proposed herein, referred to as SMA-Transformer-Huber, is adept at performing an encompassing and precise wind energy prediction, thus mitigating the risks associated with wind power generation and enhancing the overall stability of the energy system. (2) When juxtaposed with traditional models such as LSTM and GRU, the Transformer network exhibits a superior capability to discern

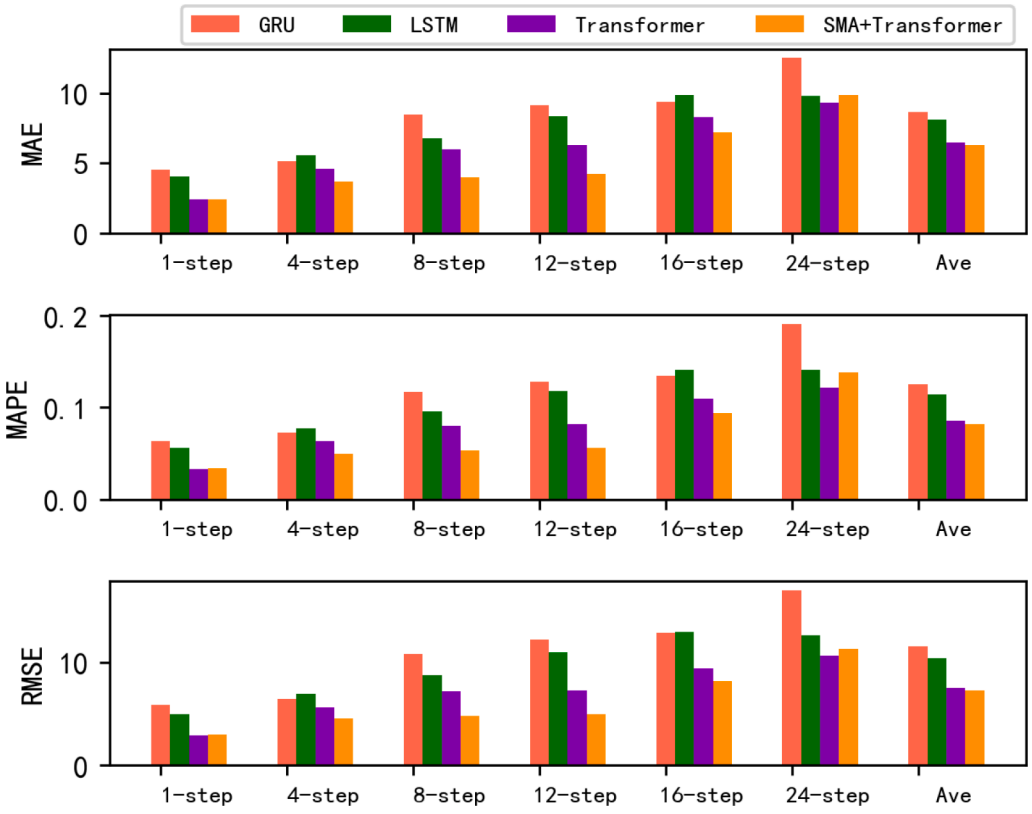


Fig. 13. Comparison of evaluation indicators for multi-step prediction.

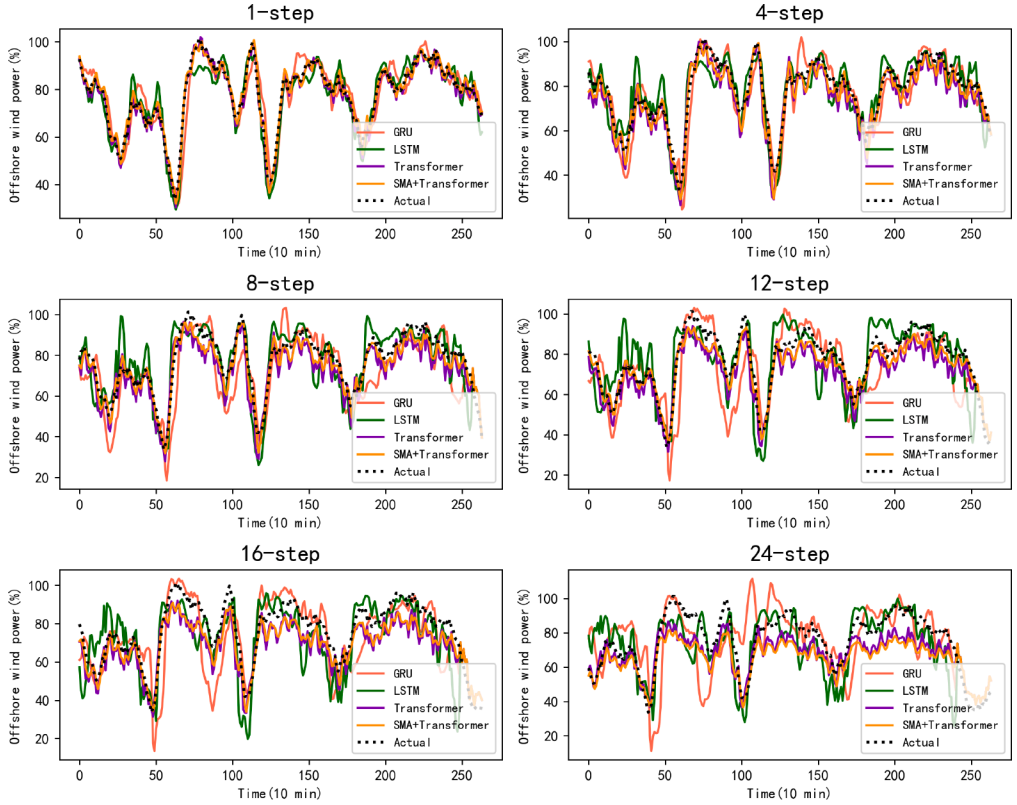


Fig. 14. Comparison of the effect of multi-step prediction.

Table 5
Comparison of evaluation metrics for multi-step prediction of offshore wind power in Elia.

Time	index	LSTM	GRU	Transformer	SMA+Transformer
t + 1	MAE	67.178759	57.941715	52.532329	53.126836
	MAPE	0.134225	0.197298	0.184221	0.017284
	RMSE	94.435220	89.228084	82.105326	63.311144
t + 4	MAE	115.363673	63.137025	58.658015	52.911338
	MAPE	2.011250	0.869681	0.707449	0.037096
	RMSE	141.778105	97.148963	91.801272	87.101771
t + 8	MAE	123.019398	82.051283	78.505737	76.247484
	MAPE	3.269040	1.687478	1.125109	0.844708
	RMSE	155.439041	112.227224	110.228224	121.763080
t + 12	MAE	132.029089	105.871941	117.491387	117.737930
	MAPE	2.948845	2.417936	2.019982	1.645524
	RMSE	178.568710	133.521633	142.681142	132.648686
t + 16	MAE	141.276193	137.526617	135.609218	118.674610
	MAPE	4.764176	3.816612	3.116751	2.407367
	RMSE	193.207619	166.327528	166.219210	144.279194
t + 24	MAE	165.952849	159.651665	157.366462	155.967684
	MAPE	7.387045	6.080582	5.920533	4.752830
	RMSE	205.031194	218.881855	207.875295	174.784882
Ave	MAE	126.095670	115.994472	115.363906	112.044595
	MAPE	3.805909	2.997307	2.599556	1.995791
	RMSE	167.947318	145.925488	145.098913	127.267031

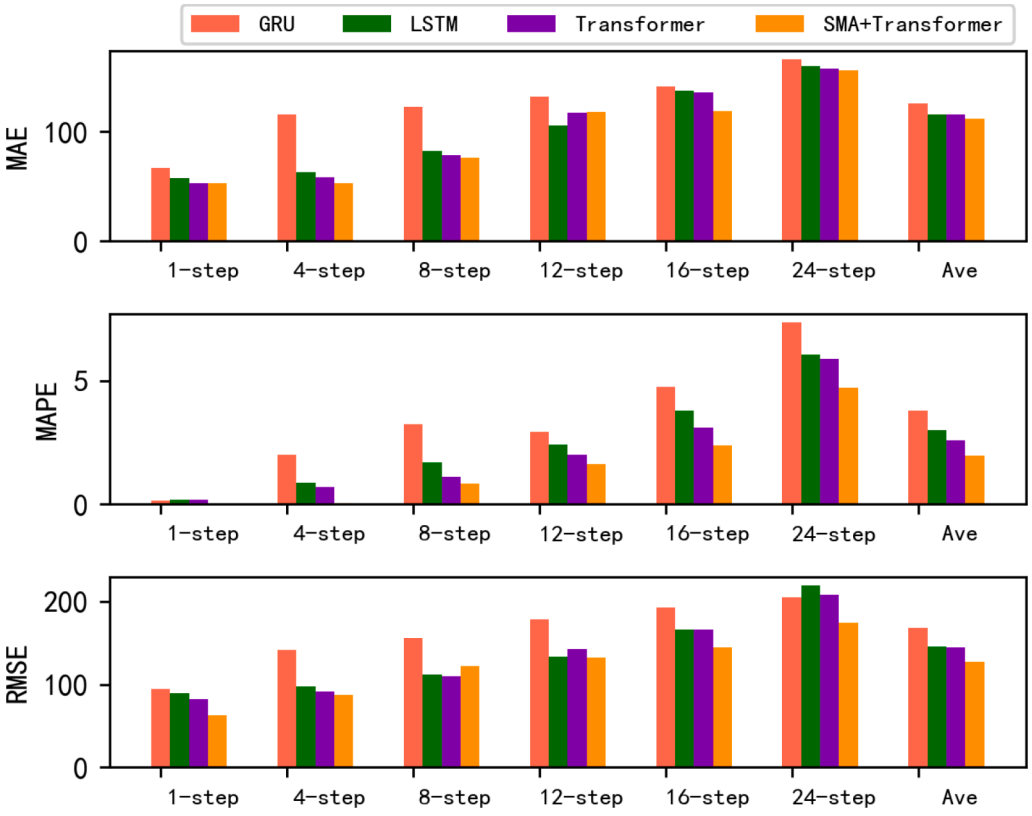


Fig. 15. Comparison of evaluation metrics for multi-step forecasting of offshore wind power in Elia.

the intrinsic correlation and distant dependencies within extended sequences, culminating in a heightened prediction accuracy. (3) The Huber loss function is more effective in predicting offshore wind data with high volatility and stochasticity due to its segmented treatment of losses. (4) The utilization of an autoencoder to reconstruct and denoise the original dataset effectively mitigates the effects induced by outliers, while the employment of the Slime Mould algorithm for the determination of optimal model structure and hyperparameters augments the model's predictive precision. (5) The construction of a multi-step prediction model for offshore wind power transcends traditional single-step predictions, offering more exhaustive and accurate predictive information—a facet vital to ensuring the steadfast operation of the power grid.

The reliable operation and development of wind energy systems that are feasible for practical application necessitate the testing of a broader spectrum and greater volume of data. This requirement marks a limitation of the present study. Consequently, it is our hope that future endeavors will allocate increased computational resources to train on a more expansive array of offshore wind turbine model data and to conduct a thorough search for optimal hyperparameters. In summation, the methodology delineated in this study represents an innovative approach in the domain of offshore wind power prediction, setting a

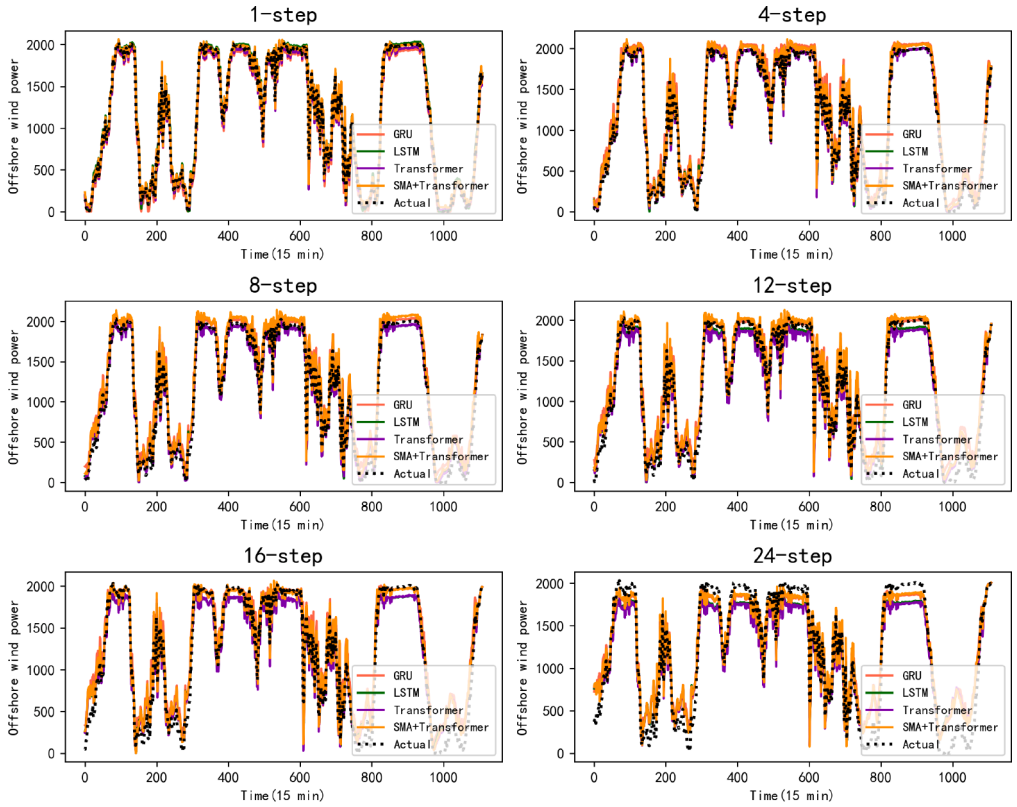


Fig. 16. Comparison of the effectiveness of multi-step forecasting for offshore wind power in Elia.

foundation for further advancements in this field.

CRediT authorship contribution statement

Haoyi Xiao: Software, Methodology, Data curation. **Xiaoxia He:** Validation, Supervision, Methodology, Funding acquisition. **Chunli Li:** Writing – review & editing, Writing – original draft, Formal analysis.

Declaration of competing interest

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

Data availability

The European offshore wind farms data presented in this study are available from the following website: https://zenodo.org/records/5525065. And the Elia offshore wind power data are available from the website: https://www.elia.be/en/grid-data/generation-data/wind-power-generation.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (NSFC) (No. 11201356) and Hubei Province Key Laboratory of Systems Science in Metallurgical Process (Wuhan University of Science and Technology) (No. Y202201).

References

[1] Ren Y, Li Z, Xu L, Yu J. The data-based adaptive graph learning network for analysis and prediction of offshore wind speed[J]. Energy 2023;267:126590.
[2] He Y, Zhang W. Probability density forecasting of wind power based on multi-core parallel quantile regression neural network[J]. Knowl-Based Syst 2020;209:106431.
[3] Niu D, Sun L, Yu M, Wang K. Point and interval forecasting of ultra-short-term wind power based on a data-driven method and hybrid deep learning model[J]. Energy 2022:124384.
[4] Wang Y, Zou R, Liu F, Zhang L, Liu Q. A review of wind speed and wind power forecasting with deep neural networks[J]. Appl Energy 2021;304:117766.
[5] Qiao B, Liu J, Wu P, Teng Y. Wind power forecasting based on variational mode decomposition and high-order fuzzy cognitive maps[J]. Appl Soft Comput 2022;129:109586.
[6] Ding M, Zhou H, Xie H, Wu W, Liu K, Nakanishi Y, et al. A time series model based on hybrid-kernel least-squares support vector machine for short-term wind power forecasting[J]. ISA Trans 2021;108:58–68.
[7] Richmond M, Sobey A, Pandit R, Kolios A. Stochastic assessment of aerodynamics within offshore wind farms based on machine-learning[J]. Renew Energy 2020;161:650–61.
[8] Zha W, Liu J, Li Y, Liang Y. Ultra-short-term power forecast method for the wind farm based on feature selection and temporal convolution network[J]. ISA transactions, 2022.
[9] Wang H, Ye J, Huang L, Wang Q, Zhang H. A multivariable hybrid prediction model of offshore wind power based on multi-stage optimization and reconstruction prediction[J]. Energy 2023;262:125428.
[10] Wei J, Wu X, Yang T, Jiao R. Ultra-short-term forecasting of wind power based on multi-task learning and LSTM[J]. Int J Electr Power Energy Syst 2023;149:109073.
[11] Cui Y, Chen Z, He Y, Xiong X, Li F. An algorithm for forecasting day-ahead wind power via novel long short-term memory and wind power ramp events[J]. Energy 2022:125888.
[12] Ahmad T, Zhang D. A data-driven deep sequence-to-sequence long-short memory method along with a gated recurrent neural network for wind power forecasting [J]. Energy 2022;239:122109.
[13] Niu Z, Yu Z, Tang W, Wu Q, Reformat M. Wind power forecasting using attention-based gated recurrent unit network[J]. Energy 2020;196:117081.
[14] Li LL, Liu ZF, Tseng ML, Jantarakolica K, Lim MK. Using enhanced crow search algorithm optimization-extreme learning machine model to forecast short-term wind power[J]. Expert Syst Appl 2021;184:115579.
[15] Jalali SMJ, Ahmadian S, Khodayar M, Khosravi A, Shafie KM, Nahavandi S, et al. An advanced short-term wind power forecasting framework based on the optimized deep neural network models[J]. Int J Electr Power Energy Syst 2022;141:108143.
[16] Chen X, Zhang X, Dong M, Huang L, Guo Y, He S. Deep learning-based prediction of wind power for multi-turbines in a wind farm[J]. Front Energy Res 2021;9:723775.
[17] Chen D, Hong W, Zhou X. Transformer Network for Remaining Useful Life Prediction of Lithium-Ion Batteries[J]. IEEE Access 2022;10:19621–8.

[18] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Aidan N, et al. Attention is all you need[J]. Adv Neural Inf Proces Syst 2017;30.
[19] Wang L, He Y, Liu X, Li L, Shao K. M2TNet: Multi-modal multi-task Transformer network for ultra-short-term wind power multi-step forecasting[J]. Energy Rep 2022;8:7628–42.
[20] Wu H, Meng K, Fan D, Zhang Z, Liu Q. Multistep short-term wind speed forecasting using transformer[J]. Energy 2022;261:125231.
[21] Li N, Dong J, Liu L, Yan J. A novel EMD and causal convolutional network integrated with Transformer for ultra short-term wind power forecasting[J]. Int J Electr Power Energy Syst 2023;154:109470.
[22] Xiong B, Lou L, Meng X, Wang X, Ma H, Wang Z. Short-term wind power forecasting based on Attention Mechanism and Deep Learning[J]. Electr Pow Syst Res 2022;206:107776.
[23] Zhou X, Liu C, Luo Y, Wua B, Donga N, Xiao T, et al. Wind power forecast based on variational mode decomposition and long short term memory attention network [J]. Energy Rep 2022;8:922–31.
[24] Tian C, Niu T, Wei W. Developing a wind power forecasting system based on deep learning with attention mechanism[J]. Energy 2022;257:124750.
[25] Zhang W, He Y, Yang S. A multi-step probability density prediction model based on gaussian approximation of quantiles for offshore wind power[J]. Renew Energy 2023;202:992–1011.
[26] He Y, Li H. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation[J]. Energ Conver Manage 2018;164:374–84.
[27] Jiang C, Jiang M, Xu Q, Huang X. Expectile regression neural network model with applications[J]. Neurocomputing 2017;247:73–86.
[28] Xiao H, He X, Li C. Probability density forecasting of wind power based on transformer network with expectile regression and Kernel Density Estimation[J]. Electronics 2023;12(5):1187.
[29] Tan H, Li H, Xiang X, Du J, Yao R, Zhou H. Graph Data Driven Power Flow Model for Offshore Wind Farm Considering Internal and External Characteristics[J]. Available at SSRN 4371072.
[30] Lu W, Duan J, Wang P, Ma W, Fang S. Short-term Wind Power Forecasting Using the Hybrid Model of Improved Variational Mode Decomposition and Maximum Mixture Correntropy Long Short-term Memory Neural Network[J]. Int J Electr Power Energy Syst 2023;144:108552.
[31] Chen H, Wu H, Kan T, Zhang J, Li H. Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction[J]. Int J Electr Power Energy Syst 2023;154:109420.
[32] Ewees AA, Al-qaness MAA, Abualigah L, Elaziz AM. HBO-LSTM: optimized long short term memory with heap-based optimizer for wind power forecasting[J]. Energ Conver Manage 2022;268:116022.
[33] Dong Y, Zhang H, Wang C, Zhou X. Wind power forecasting based on stacking ensemble model, decomposition and intelligent optimization algorithm[J]. Neurocomputing 2021;462:169–84.
[34] Zheng Z, Zhang Z, Wang L, Luo X. Denoising temporal convolutional recurrent autoencoders for time series classification[J]. Inf Sci 2022;588:159–73.
[35] Li S, Chen H, Wang M, Heidari AA, Mirjalili S. Slime mould algorithm: a new method for stochastic optimization[J]. Futur Gener Comput Syst 2020;111:300–23.
[36] Huber P J. Robust estimation of a location parameter[M]//Breakthroughs in statistics: Methodology and distribution. New York, NY: Springer New York; 1992: 492-518.