Article

MoCoformer: Quantifying Temporal Irregularities in Solar Wind for Long-Term Sequence Prediction

Zheng Wang *, Jiaodi Zhang [ID] **and Meijun Sun**

College of Intelligence and Computing, Tianjin University, Tianjin 300354, China; 2021244066@tju.edu.cn (J.Z.); sunmeijun@tju.edu.cn (M.S.)

* Correspondence: wzheng@tju.edu.cn

Abstract: Long-term solar wind sequence forecasting is essential for understanding the influence of the solar wind on celestial settings, predicting variations in solar wind parameters, and identifying patterns of solar activity. The intrinsic erratic temporal features of solar wind datasets present significant challenges to the development of solar wind factor estimate techniques. In response to these challenges, we present MoCoformer, a novel model based on the Transformer model in deep learning that integrates the Multi-Mode Decomp Block and Mode Independence Attention. The Multi-Mode Decomp Block employs an optimized version of variational mode decomposition technology to flexibly handle irregular features by adaptively decomposing and modeling the impact of sudden events on the temporal dynamics, enhancing its ability to manage non-stationary and irregular features effectively. Meanwhile, the Mode Independence Attention module computes attention independently for each mode, capturing the correlation between sequences and mitigating the negative impact of irregular features on time series prediction. The experimental results on solar wind datasets demonstrate that MoCoformer significantly outperforms current state-of-the-art methods in time series forecasting, showcasing superior predictive performance. This underscores the resilience of MoCoformer in handling the intricate, irregular temporal characteristics of solar wind data, rendering it a valuable instrument for enhancing the understanding and forecasting of solar wind dynamics.

Keywords: long-term prediction; solar wind forecasting; deep learning; machine learning

1. Introduction

Solar wind is a continuous, high-speed stream of charged particles emanating from the Sun that travels through the solar system and causes significant and widespread effects on Earth. In addition to endangering astronaut safety, these effects include disruptions in power grids, satellite communications, and navigation systems. Therefore, it is essential to forecast near-Earth solar wind conditions precisely in order to mitigate the adverse impact of space weather on various human activities and daily life. Reliable solar wind forecasts play a vital role in establishing a comprehensive space weather prediction system, enhancing the ability to provide timely alerts and effective responses to sudden changes in the space environment. This capability is essential for safeguarding technological infrastructure and ensuring the smooth functioning of modern society.

Extensive research has been conducted by scientists over the years with the objective of developing predictive models of solar wind variations. The following elements represent the principal components of contemporary predictive modeling methodologies: (1) Physics-based modeling based on magnetohydrodynamics (MHD)—Zhou et al. [1] used a three-dimensional MHD model to explore the propagation characteristics of coronal mass ejections (CMEs); Shen et al. [2] improved the total variation diminishing MHD model for the solar corona–interplanetary medium by enhancing new boundary conditions; Guo et al. [3] conducted numerical MHD simulations to study the relationship between

interstellar shocks and solar wind events; The National Oceanic and Atmospheric Administration (NOAA) of the United States of America introduced the ENLIL model [4], while the Space Environment Modeling Centre (SEMC) released the Space Weather Modeling Framework (SWMF) [5]; Wu et al. proposed the 3-D MHD model [6], while the Hybrid Heliospheric Modeling System (HAMS) combines the 3-D MHD model [7] with the Hakamada–Akasofu–Fry (HAF) model. The COrona-INplanetary (COIN) model developed by the SIGMA group at the Chinese Academy of Sciences (CAS) [8] represents an improvement on the Interplanetary Total Variation Diminishing (IN-TVD) model. (2) Empirical or semi-empirical modeling based on statistics, for example, Bussy et al. [9], established probability distribution functions linking the current solar wind and slope with the solar wind one solar rotation period in the future; The Wang–Sheeley–Arge (WSA) model proposed by Wang et al. [10] is based on the negative correlation between the observed solar wind speed and the coronal magnetic field expansion factor at the source surface; Arge et al. [11] improved the WSA model by introducing the correlation of the continuum function. Riley et al. [12] further refined the model by introducing an angular distance from the coronal hole boundary. In addition, the Potential Field Source Surface (PFSS) model [13] is another semi-empirical model that can be used to compute the coronal magnetic field and derive the fs and b parameters. Owens et al. [14] proposed a simple method based on the Sun's 27-day rotational period for solar wind parameter prediction. In addition, Innocenti et al. [15] used a Kalman filter for data assimilation, which significantly improved the model performance compared to the baseline model, and Liu et al. [16] used a support vector machine algorithm to predict solar wind speed.

However, the irregular and intense solar wind features are shaped by irregular and vigorous solar activities and geomagnetic storms [17]. These pronounced irregularities lead to unpredictable oscillations and intermittent gaps, making the temporal structure complex and challenging to model using conventional methods. The introduction of deep learning models facilitates the identification of patterns in irregular solar wind properties and the production of precise forecasts.

The development of machine learning techniques has led to further optimization of solar wind predictions. Machine learning models can adaptively learn and make inferences about new or future data and play a role in long-term prediction. Machine learning algorithms such as regression algorithms, Support Vector Machines (SVM) [18], Random Forest (RF) [19], Bayesian Additive Regression Trees (BART) [20], and K-Nearest Neighbor(KNN) [21] are widely used in the solar wind output forecasting and wind speed prediction fields. Support vector machines are widely used in solar wind speed [22] prediction, and researchers have improved them to adapt to the characteristics of solar wind data. Parallel SVM (PSVM) [23] and Least Square SVM (LSSVM) [24] is a model that improves the robustness of SVM and the accuracy of solar wind output prediction. Lahouar et al. [25] used RF for hour-ahead solar wind output prediction, which does not require a lot of tuning, and has the advantage of being able to do so. Shi et al. [26] proposed a two-stage feature selection and decision tree restructuring method to improve the prediction accuracy, efficiency, and robustness of the RF model. Comparisons were made with the decision tree [27]. Wang et al. [28] used the RF algorithm for wind speed input feature selection, which simplified the model structure, reduced the training time, and improved the accuracy and generalization ability.

Common deep learning prediction models include autoregressive (AR) [29], moving average (MA) [30], autoregressive moving average (ARMA) [31], and autoregressive integrated moving average (ARIMA) [32]. Poggi et al. [33] used the AR model in wind speed prediction. The study by Liu et al. [34] used ARMA for wind speed prediction, while Magadum et al. [35] used a calibrated ARMA model to improve the accuracy of short- and medium-term wind power prediction. However, these methods are mainly applicable to linear and smooth time series, making it difficult to deal with non-linear and non-smooth series. To overcome this problem, recurrent neural networks (RNNs) [36], especially popular ones such as long short-term memory (LSTM) [37] and gated recurrent unit (GRU) [38], have been introduced in deep

learning variants, performing excellently in time series forecasting, Backhus et al. [39] predicted the power output of different wind turbine data, which were calculated using techniques such as LSTM. Additionally, RNN models based on attention mechanisms are widely used in time series forecasting tasks.

As a classic deep learning model, Transformer plays an indispensable role in long-term sequence prediction. Self-attention blocks, a cornerstone in Transformer [40], hold substantial innovative significance in natural language processing. Recent models, including Sparseformer [41] for sparse attention, Switchformer [42] for multimodal data, Compressive Transformer [43] based on Hadamard transform [44], and Linformer [45] with low-rank approximation, contribute to handling diverse data efficiently. These models advance natural language processing and broaden practical applications. Self-attention mechanisms [46] play a crucial role in time series forecasting. Transformer excels in various prediction tasks. However, its limitations in handling long sequences led to the development of enhanced models. LogSparse Transformer [47] employs sparse attention mechanisms, Pyraformer [48] reduces complexity with a pyramid attention approach, Informer [49] uses a hierarchical structure and masking mechanism, Autoformer [50] introduces adaptive feed-forward mechanisms and reversible embeddings, FEDformer [51] innovates with a federated learning framework, and InParformer [52] incorporates interactive attention for enhanced temporal pattern extraction. These models advance time series forecasting, offering diverse solutions for practical applications.

To address the challenge of significant temporal irregularities in solar wind sequences, this study proposes an innovative Transformer-based approach, focusing on two pivotal components: Multi-Mode Decomp Block and Mode Independence Attention. In the first step, the Multi-Mode Decomp Block adapts to the dynamic features of solar wind data, effectively extracting inherent patterns for adaptive decomposition and modeling. The subsequent application of Mode Independence Attention further fortifies predictions by capturing relationships between time series. This novel approach establishes a resilient long-term forecasting model, mitigating the impact of irregular features and providing a precise solution for solar wind prediction. This study introduces novel methods to tackle the challenges of irregular features, particularly in solar wind forecasting, with considerable academic and practical value.

In summary, this study presents the following contributions:

- **Multi-Mode Decomp Block:** Introducing a block named Multi-Mode Decomp for Multi-Mode decomposition of time series. This innovative method enhances the extraction of in-depth correlated information within the sequence, revealing latent patterns and structures in the solar wind data.
- **Mode Independence Attention:** A self-attention module is proposed, the Mode Independence Block, which computes attention independently for each subsequence. This module is designed to focus on capturing the correlations in time series, enhancing the influence of valid features on prediction, and reducing the adverse effects of irregular features on overall forecasting accuracy.
- **Experimental Evaluations:** Conducting extensive experiments on the solar wind dataset, the results showcase MoCoformer's state-of-the-art performance in the realm of time series solar wind prediction.

2. Prior Technologies

2.1. Variational Mode Decomposition

The VMD algorithm [53] is an adaptive time series decomposition method that effectively resolves the modal aliasing problem and produces precise signal separation. In order to efficiently split the time series into low-frequency and high-frequency modes, VMD employs an iterative process to find the variational model's optimal solution. This allows it to ascertain the center frequency and bandwidth of each IMF. In particular, the objective function of the VMD algorithm is to minimize the sum of the estimated bandwidths of the

K IMFs and to satisfy the constraints to ensure that the sum of the K IMFs is equal to the input signal f. The mathematical expression for this process is given below:

$$\min_{u_k, w_k} = \sum_{k=1}^{K} \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-jw_k t} \right\|_2^2 \tag{1}$$

$$s.t. \sum_{k=1}^{K} u_k = f$$

where ∂_t represents an impulse function, u_k represents the k-th IMF obtained from the signal decomposition, w_k denotes the center frequency of the k-th mode, and j refers to the imaginary unit. The objective of this formulation is to minimize the bandwidth of each u_k.

To cope with this problem, the quadratic penalty parameter α and the Lagrange multiplier operator $\lambda(t)$ are introduced to subtly transform the original constrained variational problem into an unconstrained variational problem. Among them, the expression of the augmented Lagrangian function is given as follows:

$$L = \alpha \sum_{k=1}^{K} \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-jw_k t} \right\|_2^2$$
$$+ \left\| f(t) - \sum_{k=1}^{K} u_k \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k \right\rangle \tag{2}$$

where α represents a weight parameter, $\langle \cdot \rangle$ is inner product operation, ∂_t denotes the differentiation with respect to time, λ represents the Lagrange multiplier, and $\| \cdot \|_2$ represents the L2 regularization to avoid over-fitting of the results.

For the solution of this complex problem, the IMF decomposition was conducted by transferring it from the time domain to the frequency domain, resulting in the following findings:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(t)}{2}}{1 + 2\alpha(w - w_k)^2} \tag{3}$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k(w)|^2}{\int_0^\infty |\hat{u}_k(w)|^2} \tag{4}$$

When the number of decomposed IMFs is K, the sequence of operations of the VMD algorithm is as follows:

(1) Firstly, initialization is carried out by setting $u_k^1, w_k^1, \hat{\lambda}^1$, and the number of iterations n.
(2) By applying Equations (3) and (4), update u_k and w_k until k reaches K.
(3) Update the operator λ using $\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left(\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) \right)$.
(4) Keep iterating until the convergence condition $\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 / \| \hat{u}_k^n \|_2^2 < e$ is reached, or until the maximum number of iterations n is reached, or else end the iteration and then return to step (2).

As previously demonstrated, a finite set of intrinsic modal functions u_k with specific sparse properties can be obtained in a non-recursive manner. It is worth noting that the VMD algorithm exhibits increased robustness in suppressing noise due to its integration of a Wiener filter to dynamically update these modes.

2.2. Transformer Model

The Google team proposed Transformer [40] as a significant classical model for natural language processing in 2017. To manage dependencies between inputs and outputs, the model uses a self-attention mechanism without relying on the sequential structure of Recurrent Neural Networks [36]. This architectural choice allows the model to be trained in parallel and to capture global information. In the areas of visual image processing and natural language processing, such as machine translation, Transformer has excelled and produced excellent results. The overall structure of the Transformer model is illustrated in Figure 1. The structure of Transformer is mainly composed of multiple Encoder and

Decoder layers stacked on top of each other. Each layer is constructed in a uniform manner. The information is generated by a single Encoder and subsequently transmitted to all other Encoders, after which it is relayed to a single Decoder. The Encoder module incorporates both multi-head attention and a feed-forward neural network. The Decoder module incorporates an additional masked multi-head attention mechanism. Each Encoder and Decoder layer is followed by a residual join and a layer normalization section, which is designed to address the challenges encountered when training multi-layer neural networks. In this manner, information from the preceding layer is transmitted seamlessly to the subsequent layer, thereby enhancing the model's performance.
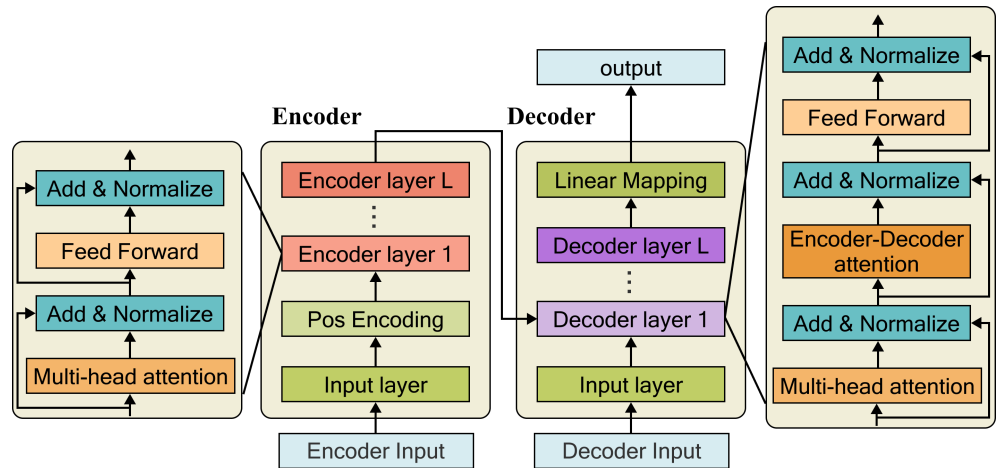
Figure 1. Transformer Model structure.

(1) Pos Encoding

Before inputting the model, the input features must be positionally encoded. This is because Transformer models do not use sequential information like RNNs, but rely on global information. This means that the model cannot automatically understand the sequential information in the input features, which is often crucial for tasks such as time series.

To solve this problem, Transformer introduces sine and cosine functions for position encoding. The exact mathematical formulas are given below:

$$PE_{(pos,2i)} = \sin\left(pos/10,000^{2i/d_{model}}\right) \tag{5}$$

$$PE_{(pos,2i+1)} = \cos\left(pos/10,000^{2i/d_{model}}\right) \tag{6}$$

where pos represents the position of the current word in the sequence, i denotes the index of each element in the vector, and d_{model} denotes the dimension of the word vector. Note that the sine function is used for encoding and at odd positions the cosine function is used for encoding. This mechanism of positional encoding helps the model to understand the positional information of the input features.

(2) Attention Mechanism

The core idea of the attention mechanism is to learn the importance of different elements and weigh them, given a set of <key, value> pairs and a query vector. By calculating the similarity between the query vector and each key, the weighting coefficients for each key can be obtained; then, these weighting coefficients are applied to the corresponding values, which ultimately gives the output of the attention. The specific mathematical formula is as follows:

$$\text{Attention (Query, Source)} = \sum_{i=1}^{L_X} \text{Similarity(Query , Key}_i) * \text{Value}_i \tag{7}$$

where L_X denotes the length of the data.

The purpose of the self-attention mechanism is to capture the correlation between vectors, whereas the multi-attention mechanism consists of multiple self-attention mechanisms that help to capture richer feature information.

(3)   Feed-Forward Network

Each layer of the Encoder and Decoder is embedded in a feed-forward network, the structure of which consists of two linearly connected layers processed with a ReLU activation function between them, as mathematically expressed below:

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{8}$$

3. Methods

This section will provide a detailed explanation of the following: (1) the overall structure of the MoCoformer model, as illustrated in Figure 2; (2) the Multi-Mode Decomp Block; (3) the Mode Independence Attention, as illustrated in Figure 3.
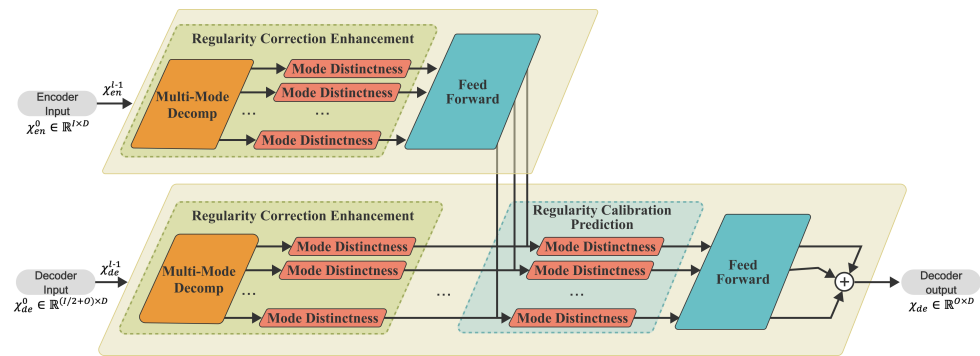

Figure 2. MoCoformer structure. MoCoformer consists of N Encoders and M Decoders. The Regularity Correction Enhancement module (including Multi-Mode Decomp Block and Mode Independence Attention) and the Regularity Calibration Prediction module (including Mode Independence Attention) are used to extract irregular features.
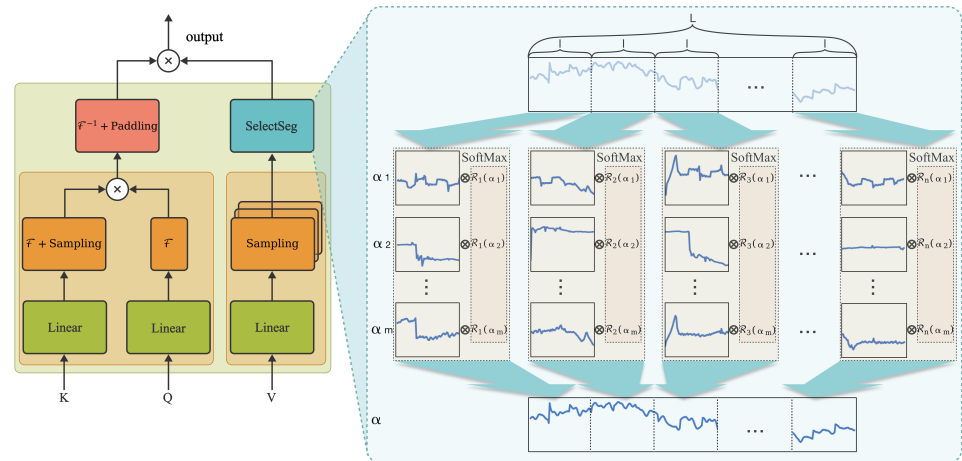

Figure 3. Mode Independence Attention Structure (**left**) and SelectSeg Structure (**right**). We perform Mode Independence Attention on each decomposed subsequence to capture the correlation between data points in the time series.

As illustrated in Figure 2, the Encoder in the MoCoformer model employs a multi-layer structure: $\mathcal{X}_{en}^l = \text{Encoder}(\mathcal{X}_{en}^{l-1})$, where $l \in \{1, 2, \ldots, N\}$ represents the output of the second-layer Encoder. Each Encoder layer consists of a Regularity Correction Enhance-

ment and Feed-Forward. $\mathcal{X}_{en}^0 \in \mathbb{R}^{I \times D}$ denotes the embedded historical sequence. The specification of the Encoder can be expressed as follows:

$$\mathcal{D}_{en}^{l,1,1}, \mathcal{D}_{en}^{l,1,2}, \ldots \mathcal{D}_{en}^{l,1,n} = \text{Multi-ModeDecomp}\left(\mathcal{X}_{en}^{l-1}\right) \tag{9}$$

$$\mathcal{D}_{en}^{l,2,1}, \mathcal{D}_{en}^{l,2,2}, \ldots \mathcal{D}_{en}^{l,2,n} = \text{ModeIndependence}\left(\mathcal{D}_{en}^{l,1,1}, \mathcal{D}_{en}^{l,1,2}, \ldots \mathcal{D}_{en}^{l,1,n}\right) \tag{10}$$

$$\mathcal{D}_{en}^{l,3,1}, \mathcal{D}_{en}^{l,3,2}, \ldots \mathcal{D}_{en}^{l,3,n} = \text{FeedForward}\left(\mathcal{D}_{en}^{l,1,1}, \mathcal{D}_{en}^{l,1,2}, \ldots \mathcal{D}_{en}^{l,1,n}\right) \tag{11}$$

where $\mathcal{D}_{en}^{l,i,m}, i \in \{1,2\}, m \in \{1,2,\ldots,n\}$ represents the n-th decomposed sequence after the i-th module in the first layer.

The Decoder also employs a multi-layer structure, for example, $\mathcal{X}_{de}^l = \text{Decoder}(\mathcal{X}_{de}^{l-1})$, where $l \in \{1,2,\ldots,M\}$, indicating the output of the l-th Decoder layer. Each Decoder layer includes Regularity Correction Enhancement, Regularity Calibration Prediction, and Feed-Forward. The specifications of the Decoder can be represented as follows:

$$\mathcal{D}_{de}^{l,1,1}, \mathcal{D}_{de}^{l,1,2}, \ldots \mathcal{D}_{de}^{l,1,n} = \text{Multi-ModeDecomp}\left(\mathcal{X}_{de}^{l-1}\right) \tag{12}$$

$$\mathcal{D}_{de}^{l,2,1}, \mathcal{D}_{de}^{l,2,2}, \ldots \mathcal{D}_{de}^{l,2,n} = \text{ModeIndependence}\left(\mathcal{D}_{de}^{l,1,1}, \mathcal{D}_{de}^{l,1,2}, \ldots \mathcal{D}_{de}^{l,1,n}\right) \tag{13}$$

$$\mathcal{D}_{de}^{l,3,1}, \ldots \mathcal{D}_{de}^{l,3,n} = \text{ModeIndependence}\left(\left(\mathcal{D}_{de}^{l,2,1}, \mathcal{D}_{en}^{l,3,1}\right), \ldots \left(\mathcal{D}_{de}^{l,2,n}, \mathcal{D}_{en}^{l,3,n}\right)\right) \tag{14}$$

$$\mathcal{D}_{de}^l = \sum_{i=1}^n \left(\text{FeedForward}\left(W_{l,i} * \mathcal{D}_{de}^{l,3,i}\right)\right) \tag{15}$$

$$\mathcal{X}_{en}^l = \mathcal{D}_{de}^{l-1} \tag{16}$$

where $\mathcal{D}_{de}^{l,i,m}, i \in \{1,2\}, m \in \{1,2,\ldots,n\}$ represents the m-th decomposed sequence after the i-th module in the first layer. $W_{l,i}, i \in \{1,2,3\}$ denotes the weight for the i-th decomposed component of $\mathcal{T}_{de}^{l,i}$.

The final prediction result is obtained by summing all the decomposed components. Specifically, it is obtained by multiplying each decomposed component by its corresponding weight and then summing them.

3.1. Multi-Mode Decomp Block

To explore the correlation information in solar wind time series, we introduce a time series decomposition module based on variational mode decomposition (VMD). This module is derived from the Whale Optimization Algorithm (WOA) proposed by Mirjalili et al. [54], which simulates the foraging behavior of whale pods, decomposing historical time series into multiple component sequences, with each component sequence representing a unique time oscillation pattern. This Multi-Mode decomposition technique is an effective tool for revealing the complex internal structure of data, thereby enhancing the accuracy of feature interpretation and analysis.

Assume that the initial optimal position is the current prey's location. Other predators initially move towards this target position and continuously update their positions:

$$\boldsymbol{D} = |\boldsymbol{C} \cdot \boldsymbol{X}^*(t) - \boldsymbol{X}(t)| \tag{17}$$

$$\boldsymbol{X}(t+1) = \boldsymbol{X}^*(t) - \boldsymbol{A} \cdot \boldsymbol{D} \tag{18}$$

where t represents the current iteration number, $\boldsymbol{X}^*(t) = \left(X_1^*, X_2^*, \cdots, X_{dim}^*\right)$ represents the current position of the whale, and $\boldsymbol{A} \cdot \boldsymbol{D}$ represents the movement step length when enclosing prey during each iteration:

$$A = 2a \cdot \text{rand} - a \tag{19}$$

$$C = 2 \cdot \text{rand} \tag{20}$$

where rand represents a random number generated between [0, 1], and as the loop iteration increases, a acting as the contraction factor will linearly decrease from 2 to 0. Its expression is as follows:

$$a = (2 - 2t/T_{\max}) \tag{21}$$

where $T_{\max}$ represents the maximum number of iterations set for the run.

Whales achieve continuous shrinking of the surrounding area during the prey search process through Equations (19) and (20). At the same time, each whale updates its distance from the target position through a spiral pattern, and the mathematical model simulating the spiral pattern update can be expressed as follows:

$$X(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \tag{22}$$

where $D' = |X^*(t) - X(t)|$ represents the distance between the i-th whale and the current optimal position, b is a constant coefficient, and l is a random number in the range of [1, −1].

To synchronize the contraction of the surrounding and the spiral update, the selection update of the two methods is achieved through probability. The mathematical expression for this is as follows:

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D \\ D' \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \end{cases} \tag{23}$$

When $|A| \geqslant 1$, randomly selecting whales to move away from the current optimal whale enhances the algorithm's global search capability. The mathematical expression for this is as follows:

$$D = |C \cdot X_{\text{rand}} - X(t)| \tag{24}$$

$$X(t+1) = X_{\text{rand}} - A \cdot D \tag{25}$$

where X_{rand} represents the vector of the randomly selected whale's position.

Multi-Mode Decomposition Steps The combination optimization of VMD parameters k and α is carried out through the following steps:

1. Initialize the parameters k and α as k = 8 and α = 2000, respectively, to avoid incomplete signal decomposition.
2. Perform VMD decomposition on the oscillation signal.
3. Calculate the fitness value for each combination of parameters and update the best fitness value when it exceeds the current value.
4. Determine whether to terminate the iteration. If t is less than X_{rand}, increment t by 1 and update the position of the whales. Otherwise, terminate the iteration and save the best results.

In summary, the Multi-Mode Decomp Block is improved using the WOA method to extract IMF components with minimum envelope entropy, enhancing its ability to capture deep-level intrinsic correlations in sequences. This enhancement boosts global search capabilities and robustness, aiding in better time series analysis and decomposition.

3.2. Mode Independence Attention

Figure 3 illustrates the self-attention mechanism, which employs parallel connections to systematically reduce the impact of feature irregularities on predictions. This attention mechanism operates based on individual modes, calculating attention distribution independently. It assesses correlations among distinct sequences and assigns varying weights

according to the correlation strength. Finally, we combine these weighted sequences to generate the final predicted sequence.

This model adopts the representation scheme of the classical Transformer model. The inputs consist of queries, keys, and values, denoted as $q \in \mathbb{R}^{L \times D}, k \in \mathbb{R}^{L \times D}$, and $v \in \mathbb{R}^{L \times D}$, respectively. In the cross-attention, the Decoder obtains queries through $q = x_{\text{en}} \cdot w_{\text{q}}$, where $w_{\text{q}} \in \mathbb{R}^{D \times D}$. The Encoder obtains keys and values through $k = x_{\text{en}} \cdot w_{\text{k}}$ and $v = x_{\text{en}} \cdot w_{\text{v}}$, respectively, where $w_{\text{k}}, w_{\text{v}} \in \mathbb{R}^{D \times D}$. The classical attention can be formalized as a dot product and softmax operation based on queries and keys. It calculates the weights between queries and keys, and then applies these weights to the corresponding values for weighted summation, resulting in the final context embedding.

$$\text{Atten}(q, k, v) = \text{Softmax}\left(\frac{qk^{\text{T}}}{\sqrt{\text{d}_q}}\right)v \tag{26}$$

Query–Key Interaction In the computation of query and key, a linear projection matrix $E_i \in \mathbb{R}^{D \times K}$ is employed. Initially, the original key $k \in \mathbb{R}^{L \times D}$ undergoes projection to a key of dimension (K × D) through linear mapping, facilitating dimension transformation. Next, perform FFT transformations on k and q to bring them into the frequency domain for further computation. The ultimate linear self-attention operation, illustrated in Figure 3, incorporates a sequence of linear projections and scaled dot-product attention calculations.

$$\tilde{\mathcal{Q}} = \mathcal{F}(q) \tag{27}$$

$$\tilde{\mathcal{K}}\& = \mathcal{F}(E_i \cdot k) \tag{28}$$

SelectSeg Initially, m linear projection matrices $F_{i,j} \in \mathbb{R}^{D \times K}$ (where j = 1, 2, ..., m) are introduced for projecting the original value $v \in \mathbb{R}^{L \times D}$ to a projection value layer of dimension (K × D) through linear mapping. This process achieves dimension transformation, generating m-value matrices that have undergone diverse filtering.

$$\tilde{\mathcal{V}}_j = \mathcal{F}_{i,j} \cdot (v) \tag{29}$$

where i represents the layer number, and j = 1, 2, ..., m represents the different projection matrices formed.

The m value matrices are input into the SelectSeg layer, where each matrix is partitioned into n subsequences of length l. The SelectSeg layer is a matrix of length l. For each subsequence, correlations with the other (m − 1) sets of values are computed using a correlation formula.

$$\mathcal{R}_{\alpha_i \alpha_j} = \frac{1}{l} \sum_{t=1}^{l} (\chi_{\text{i}} - \chi_{\text{j}}) \tag{30}$$

$$\mathcal{R}_{\text{p}}(\alpha_q) = \frac{1}{\sum_{t=1}^{\text{m}} \mathcal{R}_{\alpha_t \alpha_q}} \tag{31}$$

$$\mathcal{R}_{\text{p}}(\alpha_1), \ldots, \mathcal{R}_{\text{p}}(\alpha_m) = \text{SoftMax}\left(\mathcal{R}_{\text{p}}(\alpha_1), \ldots, \mathcal{R}_{\text{p}}(\alpha_m)\right) \tag{32}$$

where $\mathcal{R}_{\alpha_i \alpha_j}$—where i, j = 1, 2, ..., m—the calculation is based on the similarity between two values, with a larger value indicating lower similarity. Similarly, for $\mathcal{R}_{\text{p}}(\alpha_q)$, where q = 1, 2, ..., m and p = 1, 2, ..., n, the total similarity between each value matrix α_q in the n-th subsequence and others is computed. As the numerical value increases, it indicates a decrease in irregularity and a higher weightage. Weighted sums are calculated for all value matrices within the same subsequence, resulting in the final subsequence matrix. Finally, the concatenation of all prediction subsequences forms the ultimate value matrix $\tilde{v}$.

Formulating the Formula Following the outlined procedure, we derive the computational outcomes for query–key interactions and a value matrix with reduced irregularity. Subsequently, we employ an effective self-attention mechanism through attentive calcu-

lations on the sequence data, facilitating the capture of crucial contextual information.

$$\begin{aligned}
\overline{\text{head}_i} &= \text{Atten}(\tilde{\mathcal{Q}}, \tilde{\mathcal{K}}, \tilde{\mathcal{V}}) \\
&= \mathcal{F}^{-1}\left(\sigma\left(\text{Padding}\left(\tilde{\mathcal{Q}} \cdot \tilde{\mathcal{K}^{-1}}\right)\right) \cdot \tilde{\mathcal{V}}\right)
\end{aligned} \tag{33}$$

where σ represents the activation function.

In summary, the Mode Independence Attention employs independent distribution computation, random subsequence extraction, correlation comparison, and enhancement to capture relationships between sequences. This design aims to reduce the negative impact of irregular temporal patterns on prediction accuracy and enhance the accuracy of solar wind time series prediction.

3.3. Complexity Analysis

The time and space complexity of VMD and WOA in the Multi-Mode Decomp Block in MoCoformer is O(L). In the Mode Independence Attention, although the FFT transformation complexity is O(L log(L)), our model achieves fast execution by using pre-selected Fourier basis sets, thus reducing the complexity of query–key interactions to O(L). For the SelectSeg computation, the complexity grows linearly with the length of the prediction due to the fixed number and length of randomly selected patterns, so the overall complexity is O(L).

4. Results

4.1. Datasets

The dataset utilized in this study was provided by the Space Environment Center of the National Aeronautics and Space Administration (NASA) and comprise a multitude of parameters pertaining to the solar wind and the Earth's magnetic field. Data on solar wind from the omni dataset were chosen, and correlation analyses were carried out between the IMF and Field Magnitude data for the time frame spanning from 1 January 2006 to 31 December 2015. The omni-Field-Magnitude dataset is more concerned with changes in the Earth's magnetic field, which is important for geophysical research and navigation. In contrast, the omni-IMF dataset is more closely related to solar activity and can be used to predict the effects of solar activity on the Earth, such as geomagnetic storms.

The elements included in the omni-Field-Magnitude dataset are DST Index, Bulk flow longitude, Flow Pressure, MAC, Proton density, IMF, and Proton temperature. A visual presentation of the dataset is illustrated in Figure 4. By analyzing the variations in mean magnetic field strength, scientists can gain insight into the behavior of the solar wind's magnetic field, including the occurrence of magnetic storms, the presence of magnetic structures, and the effect of solar activity on Earth's magnetosphere. The features exhibit significant fluctuations and extreme instability, including significant temporal irregularities.
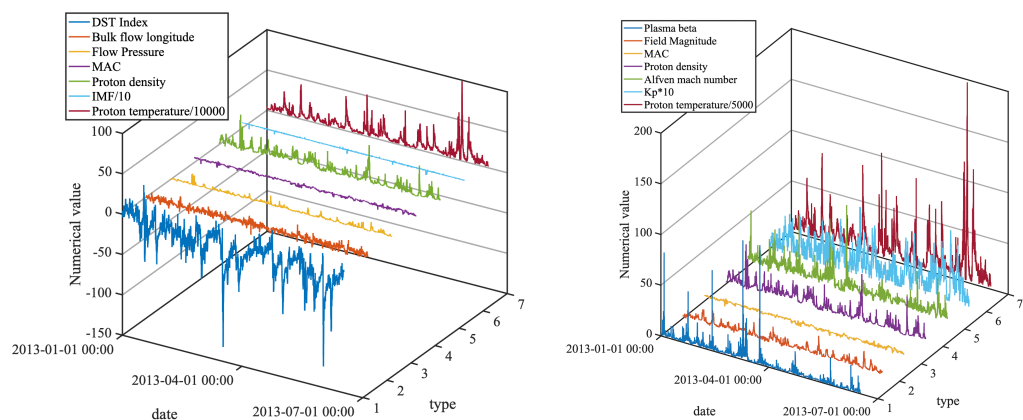
Figure 4. Visualization view of omni-Field-Magnitude dataset (**left**) and omni-IMF dataset (**right**). To make the images more harmonious, the Proton temperature is scaled down.

The elements included in the omni-IMF dataset are Plasma beta, Field Magnitude, MAC, Proton density, Alfven mach number, Kp*10, and Proton temperature. A visual presentation of the dataset is illustrated in Figure 4. The omni-IMF dataset represents a valuable resource for the study of the behavior and properties of the solar wind, with a particular focus on the interplanetary magnetic field. The dataset provides an opportunity to explore the complex relationship between the solar wind and Earth's magnetic environment, contributing to the advancement of understanding of space weather and its impact on Earth. It can be observed that these features exhibit significant volatility, and their time series show clear instability and strong temporal irregularities.

To more effectively illustrate the irregularity of the solar wind sequence, it is compared to common time series, as illustrated in Figure 5. The Electricity dataset contains hourly electricity consumption data from 321 customers from 2012 to 2014. The Exchange_rate dataset collects daily exchange rate data from 1990 to 2016 for eight countries. The ETT dataset contains load characteristic data for seven types of oil and power transformers from July 2016 to July 2018. A comparison of these datasets reveals that the solar wind sequence exhibits a higher temporal irregularity than other common sequences.
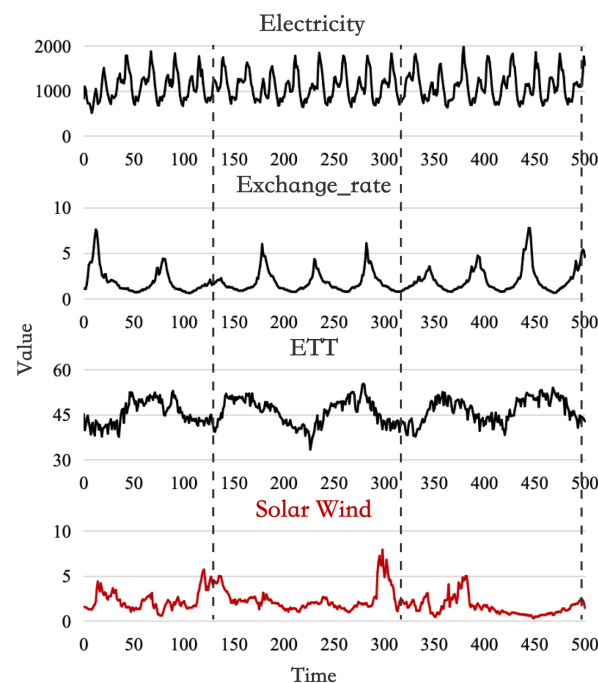
Figure 5. Visualization comparison of regularity between commonly used time series data and solar wind data.

4.2. Evaluation Metrics

Three key evaluation metrics for regression problems—namely, MAE, MAPE, and RMSE—are used to provide a comprehensive assessment of the model's fit performance. The mathematical definitions of these metrics are provided below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{34}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{35}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{36}$$

where $\hat{y}_i$ represents the observed values and y_i represents the predicted values, RMSE and MAE are used to measure the magnitude of the error between the predicted series and the actually observed series, and MAPE is used as a measure of the percentage of error relative to the observed values. RMSE focuses on the overall size of the prediction error and is more sensitive to large errors, while MAE is concerned with the average level of prediction error, is insensitive to outliers, and places more emphasis on the overall average performance. MAPE emphasizes the average level of relative error, expresses error as a percentage, and is suitable for comparing data on different scales. For a given model, the smaller the values of these metrics, the higher the predictive accuracy of the model and the better it fits the real data.

4.3. Multivariate Results

Table 1 compiles the evaluation outcomes for multivariate prediction on the two solar wind time series datasets. The five main methods included in the evaluation are the empirical solar wind method WSA [10]; the common deep learning algorithms LSTM [37]; GRU [38]; and the Transformer-based long-term series prediction methods Transformer [40], Autoformer [50], and FEDformer [51], along with the method MoCoformer proposed in this study. The experimental configuration maintains a consistent input length of 96, with fixed prediction lengths of 96, 192, 336, and 720 for both training and evaluation, enhancing comparability across the methods.

Table 1. The forecasting results for multivariate long-term series, with an input length of I = 96 and prediction lengths of O $\in$ {96, 192, 336, 720}. Lower values of MAE, RMSE, and MAPE indicate better performance, and the optimal results are highlighted in bold.

Omni-Field-Magnitude Dataset											
Predict Length	**96**			**192**			**336**			**720**		
Models	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
MoCoformer	**0.669**	**1.047**	**2.741**	**0.712**	**1.087**	**2.588**	**0.721**	**1.058**	**2.754**	**0.682**	**1.016**	**2.598**
FEDformer	0.688	1.062	2.752	0.712	1.090	2.792	0.743	1.127	2.896	0.719	1.080	2.950
Autoformer	0.724	1.111	3.207	0.752	1.147	3.223	0.765	1.152	2.889	0.733	1.085	2.916
Transformer	0.681	1.084	3.569	0.716	1.094	3.151	0.728	1.102	2.862	0.690	1.022	2.687
GRU	0.693	1.072	2.978	0.723	1.096	3.171	0.747	1.231	2.798	0.708	1.069	2.757
LSTM	0.699	1.085	3.011	0.730	1.094	3.055	0.754	1.214	2.856	0.721	1.053	2.780
WSA	0.678	1.069	3.723	0.714	1.099	3.138	0.730	1.095	2.861	0.687	1.021	2.655
omni-IMF dataset											
Predict Length	96			192			336			720		
Models	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
MoCoformer	**0.596**	**0.850**	**2.715**	**0.628**	**0.889**	**2.593**	**0.639**	**0.894**	**2.474**	**0.628**	**0.888**	**2.415**
FEDformer	0.622	0.897	2.721	0.641	0.922	2.791	0.659	0.950	2.797	0.644	0.935	2.661
Autoformer	0.658	0.932	2.938	0.650	0.938	2.814	0.661	0.948	2.768	0.641	0.930	2.597
Transformer	0.610	0.888	3.298	0.643	0.907	2.761	0.643	0.907	2.552	0.642	0.905	2.444
GRU	0.616	0.917	2.897	0.635	0.918	2.744	0.648	0.926	2.667	0.636	0.902	2.486
LSTM	0.623	0.930	2.913	0.644	0.929	2.781	0.655	0.930	2.690	0.645	0.910	2.483
WSA	0.601	0.859	2.954	0.640	0.896	2.687	0.644	0.903	2.492	0.632	0.896	2.437

In the context of varying prediction durations, it is crucial to acknowledge that the training and test datasets will differ. For instance, the randomized time window of Case96 (eight days) and of Case192 (twelve days) could serve as reasonable baselines for training various quarterly cases. It is crucial to address the issue of continuity, as time series are inherently distinctive. The sliding window method is employed, whereby the final 30% of a continuous data segment is designated as the test set while the initial 70% is utilized as the training set. This method ensures temporal continuity between the training and test sets.

The investigation presents an analysis of MAE, RMSE, and MAPE metrics in different models across varying prediction intervals. Across all metrics, MoCoformer demonstrated

the highest level of performance. Remarkably, in the omni-Field-Magnitude dataset, MoCo-former surpassed the second-ranked model with an average improvement of roughly 0.76% in MAE, 1.51% in RMSE, and 3.59% in MAPE. Likewise, within the omni-IMF dataset, MoCoformer displayed an average enhancement of about 1.40% in MAE, 1.19% in RMSE, and 1.37% in MAPE compared to its closest competitor. These findings underscore the exceptional predictive power of MoCoformer in long-term sequence forecasting, showcasing its superior adaptability to temporal irregularities in solar wind sequences and resulting in more precise predictions compared to alternative models.

4.4. Univariate Results

Table 2 compiles univariate prediction evaluation results for various methods on the two solar wind time series datasets. Experimental settings maintain a fixed input length of 96, with prediction lengths at 96, 192, 336, and 720 for training and evaluation.

Table 2. The forecasting results for univariate long-term series, with an input length of I = 96 and prediction lengths of O ∈ {96, 192, 336, 720}. Lower values of MAE, RMSE, and MAPE indicate better performance, and the optimal results are highlighted in bold.

Omni-Field-Magnitude Dataset											
Predict Length	**96**			**192**			**336**			**720**		
Models	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
MoCoformer	**0.699**	**0.926**	**2.408**	**0.711**	**0.968**	**2.156**	**0.731**	**0.988**	**1.983**	**0.733**	**0.968**	**1.850**
FEDformer	0.726	1.020	2.815	0.742	1.045	2.589	0.776	1.070	2.678	0.792	1.057	2.614
Autoformer	0.782	1.047	3.119	0.776	1.065	2.727	0.774	1.057	2.742	0.792	1.057	2.614
Transformer	0.714	0.954	2.612	0.725	0.986	2.258	0.758	1.009	1.997	0.739	0.987	1.897
GRU	0.736	0.978	2.758	0.749	0.994	2.534	0.762	1.037	2.179	0.788	0.993	1.985
LSTM	0.743	0.983	2.745	0.748	1.032	2.629	0.767	1.044	2.295	0.794	1.006	1.996
WSA	0.753	0.970	2.440	0.760	1.003	2.402	0.784	1.023	2.036	0.758	0.975	1.905
omni-IMF dataset											
Predict Length	96			192			336			720		
Models	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
MoCoformer	**0.583**	**0.812**	**2.128**	**0.588**	**0.860**	**2.464**	**0.594**	**0.865**	**2.316**	**0.607**	**0.833**	**2.471**
FEDformer	0.650	0.851	2.528	0.669	0.968	2.643	0.657	0.967	2.965	0.660	0.985	2.771
Autoformer	0.632	0.935	2.574	0.649	0.956	2.633	0.639	0.955	2.722	0.667	0.891	2.795
Transformer	0.628	0.931	2.450	0.660	0.963	2.570	0.652	0.969	2.535	0.628	0.950	2.999
GRU	0.624	0.896	2.385	0.637	0.902	2.586	0.651	0.913	2.602	0.668	0.922	2.684
LSTM	0.633	0.917	2.527	0.641	0.934	2.599	0.657	0.954	2.671	0.674	0.957	2.767
WSA	0.642	0.930	2.287	0.658	0.943	2.610	0.641	0.939	2.900	0.658	0.965	2.964

The research presents the evaluation results of MAE, RMSE, and MAPE metrics across different models and prediction horizons. MoCoformer demonstrated superior performance across a range of prediction horizons. Specifically, on the omni-Field-Magnitude dataset, MoCoformer exhibited a performance boost of around 2.16% in MAE, 1.92% in RMSE, and 2.30% in MAPE compared to the second-ranked model. Likewise, on the omni-IMF dataset, MoCoformer showcased a significant improvement of about 13.54% in MAE, 4.27% in RMSE, and 10.36% in MAPE over the second-ranked model. These results highlight the exceptional predictive capabilities of MoCoformer in long-term sequence forecasting.

To validate the effectiveness of the proposed model in long-term solar wind sequence prediction applications, a comprehensive case study is conducted. Figure 6 illustrates the visual results of univariate prediction experiments conducted on the omni-Field-Magnitude dataset, showcasing comparisons between the outputs of multiple models and ground truth curves. This experiment aims to evaluate the predictive capability and accuracy of the proposed model in real-world applications.
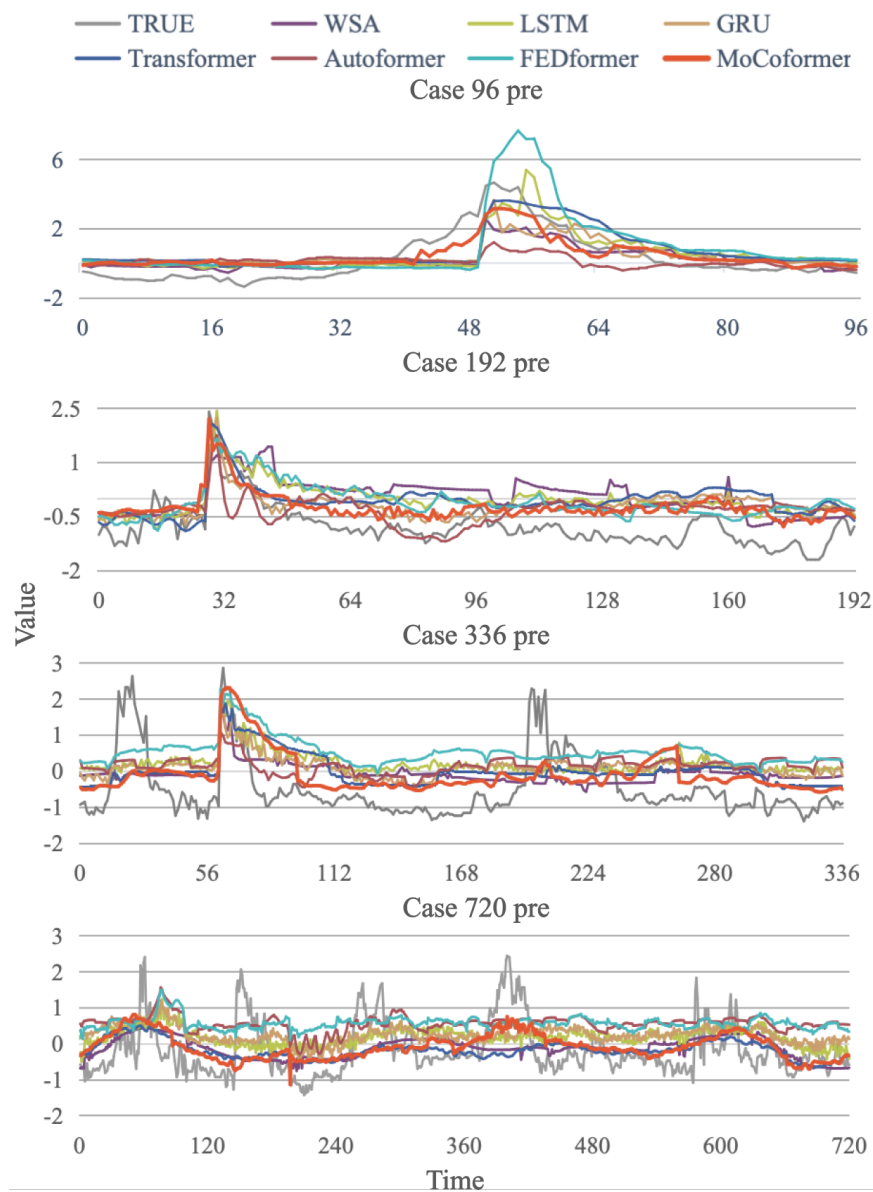
Figure 6. Visualization of the predicted results of various models with the ground truth for different prediction lengths on the omni-Field-Magnitude dataset experiment.

Figure 6 illustrates that MoCoformer accurately reflects the trends and fluctuations in the true values, exhibiting superior performance in capturing the peaks and troughs of the time series compared to other models. In particular, in Cases 96 and 196, MoCoformer demonstrates enhanced sensitivity in capturing changes in the true values compared to other models, enabling accurate predictions of increases and decreases in the corresponding periods. However, with regard to numerical values, it appears that MoCoformer tends to be relatively conservative in its predictions. For example, in Case 720, although MoCoformer is able to identify changes in true values, it still encounters difficulties in modeling large fluctuations in them. This observation is consistent with the results of the three evaluation metrics, which indicate that while there have been notable improvements in MAE and MAPE, RMSE values remain substantial.

The results of a univariate prediction experiment conducted on the omni-IMF dataset are presented in Figure 7. The figure depicts comparative curves between the ground truth and the outputs of various models.
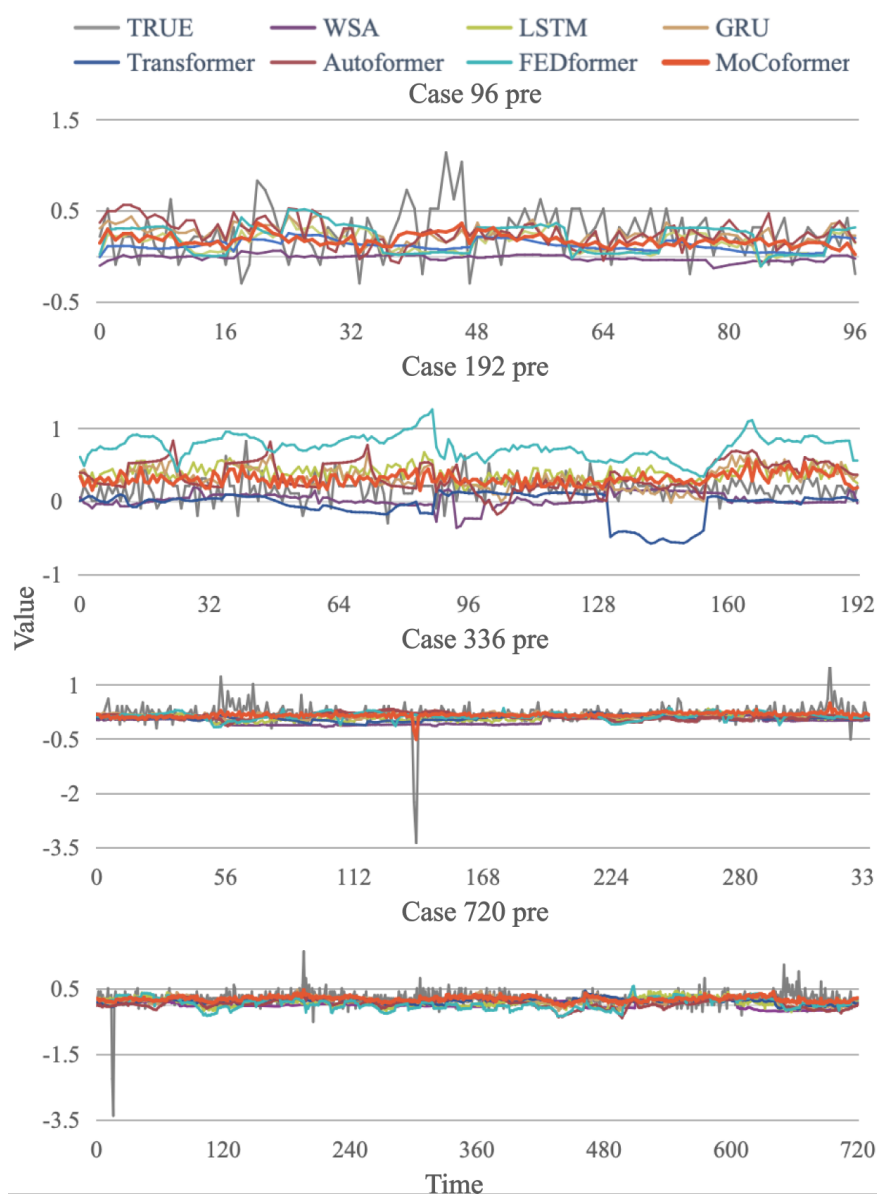
Figure 7. Visualization of the predicted results of various models with the ground truth for different prediction lengths on the omni-IMF dataset experiment.

Figure 7 illustrates that the MoCoformer model outperforms the other models in capturing trends and changes in the true values, which is particularly evident in Cases 192 and 720. It is also noteworthy that although the MoCoformer model excels in trend capture, its numerical predictions tend towards conservatism, implying a tendency for more conservative numerical outputs. This becomes evident when modeling prominent exceptional values, as illustrated in Case 336. Moreover, the limited size of the omni-IMF dataset contrasts with the relatively higher prominence of exceptional values. No model demonstrates superior performance in predicting these exceptional values. This observation is consistent with the trend observed in the three assessment metrics. The MAE and MAPE demonstrate a more pronounced improvement, while the RMSE shows a comparatively slower improvement. Overall, there is considerable scope for further enhancement and improvement of the models' performance on the omni-IMF dataset, particularly in the prediction of exceptional values.

In comparison to other models, MoCoformer demonstrates superior performance in identifying trends and fluctuations in actual values. It is able to recognize trends with exceptional precision. Nevertheless, its numerical predictions frequently exhibit a tendency

towards conservatism, as evidenced by the table, which shows that there are notable improvements in MAE and MAPE, yet RMSE values remain relatively high. In conclusion, it can be stated that MoCoformer makes a significant contribution to long-term sequence prediction in the solar wind domain. However, the complexity of solar wind data presents numerous challenges for long-term solar wind sequence prediction research.

4.5. Ablation Studies

For module validation, univariate prediction experiments were conducted on the omni-Field-Magnitude dataset. The proposed models and their variants are summarized below for comparative analysis.

- **MoCo-rM**: removed the Multi-Mode Decomp Block from the model.
- **MoCo-rMrSTL**: replaced the Multi-Mode Decomp Block with the STL decomposition module.
- **MoCo-rCrT**: substituted the Mode Independence Attention proposed in this chapter with the auto-correlation module from the Transformer model.
- **MoCoformer**: The original model proposed in this chapter.

The experimental results of each module combined with the backbone network structure are shown in Table 3.

Table 3. Comparison of ablation studies results for univariate prediction based on the omni-IMF dataset.

Method	96			192			336			720		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
MoCo-rM	0.747	1.016	2.407	0.746	1.025	2.554	0.747	1.061	2.382	0.771	1.026	2.411
MoCo-rMrSTL	0.769	1.038	3.027	0.776	1.043	2.724	0.768	1.048	2.686	0.788	1.045	2.593
MoCo-rCrT	0.715	0.938	2.230	0.720	0.988	2.258	0.735	**0.986**	1.997	0.739	**0.966**	**1.844**
MoCoformer	**0.699**	**0.926**	**2.408**	**0.711**	**0.968**	**2.156**	**0.731**	0.988	**1.983**	**0.733**	0.968	1.850

The table illustrates the significance of the Multi-Mode Decomp Block in enhancing model performance across various prediction lengths, thereby underscoring its positive influence on long-term solar wind forecasting. Higher numerical values were observed upon replacement of the STL module with this block, indicating the challenge of capturing temporal irregularities through conventional trend–seasonal decomposition for comprehensive pattern comprehension. Furthermore, the substitution of Mode Independence Attention with the Transformer's auto-correlation module resulted in inferior outcomes in shorter prediction lengths, thereby underscoring the beneficial impact of the proposed Mode Independence Attention in solar wind prediction tasks.

However, as the length of the predicted time series increases, the predictions of the two models converge. This indicates that as the complexity of the solar wind increases in longer time series predictions, the predictions of both modules gradually become more similar.
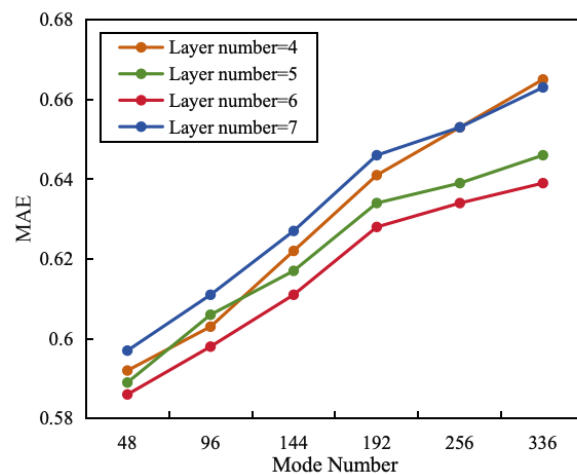
4.6. Multi-Mode Decomp Block Decomposition Experiment

The number of layers in the Multi-Mode Decomp Block is of significant importance in influencing the prediction process and the resulting outcomes. An increase in the number of layers allows for a more precise capture of diverse frequency components and modal features, thereby enhancing the modal information extracted from the original signal. The selection of the optimal layer count necessitates a delicate balance in order to ensure optimal model performance. A deficiency in the number of layers may result in the loss of information, whereas an excess could introduce noise or irrelevant details. The results of the experiments conducted on the layer count in the Multi-Mode Decomp Block are presented in Figure 8, which illustrates the impact of this variable on the performance of the model.

The experimental results, corroborated by Figure 9, highlight the significance of choosing an optimal number of decomposition layers. The figure illustrates that when

the number of IMF layers is set to six, optimal outcomes are achieved, with the essential information captured while avoiding noise and unnecessary details. The acquisition of information may be inadequate with five or fewer layers, resulting in less accurate predictions. Conversely, the introduction of seven or more layers may result in the inclusion of excessive details or noise, which could have a detrimental impact on the performance of the model. This underscores the pivotal importance of determining an optimal number of layers for specific conditions.

Figure 8. The visualization of MAE results of different decomposition layers in the Multi-Mode Decomp Block based on the omni-IMF dataset experiment.
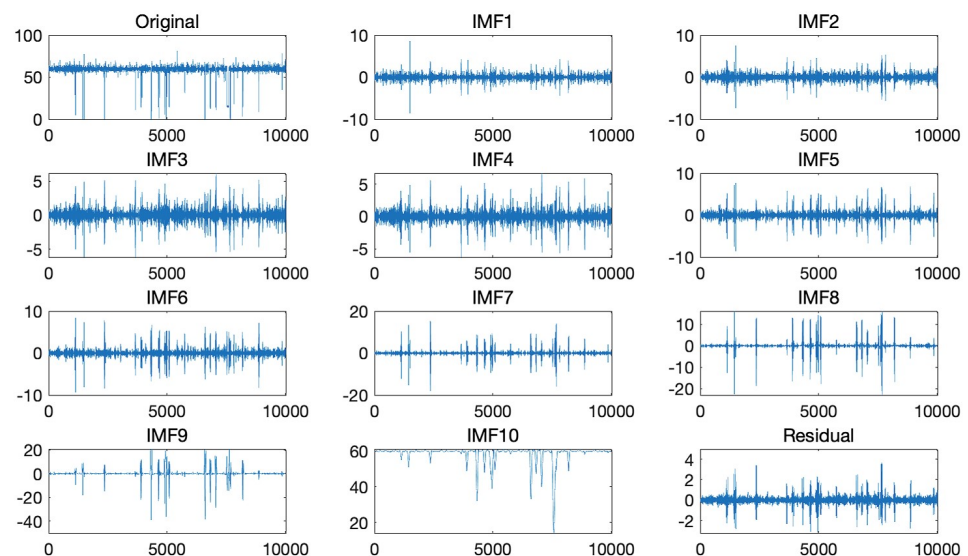
Figure 9. The visualization of different decomposition layers in the Multi-Mode Decomp Block based on the omni-IMF dataset experiment.

4.7. Subsequence Length Experiment

The accuracy of the predicted time series in the SelectSeg module of Mode Independence Attention is intricately linked to the subdivision size for the prediction length. This size determines the proportion of irregular segments in the final forecast and is crucial for overall accuracy. An inadequate subdivision length may be insufficient to effectively reduce the proportion of highly uncertain segments, while an excessively large length could inaccurately diminish the weight of effective segments. Experiments were conducted on the impact of the subdivision length in the SelectSeg module, and the findings are presented in Figure 10.
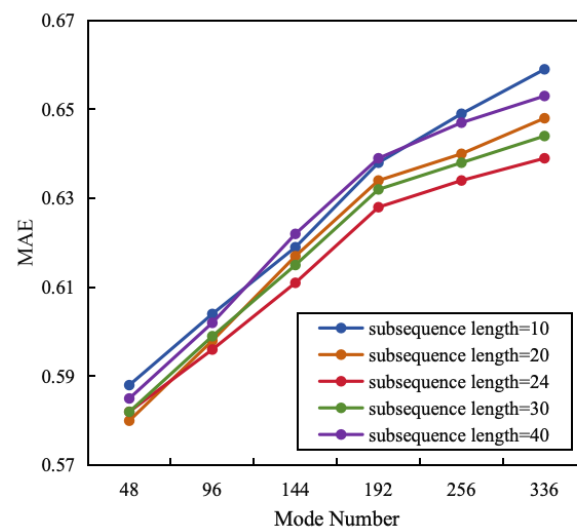
Figure 10. Visualization of the MAE results for different subsequence length selections in the SelectSeg module of the Mode Independence Attention experiment based on the omni-IMF dataset experiment.

The results demonstrate a significant improvement in model performance resulting from strategic parameter selection. Although the impact of subsequence length selection in SelectSeg is relatively minor in short-term sequence prediction, it becomes more pronounced as the prediction sequence length increases. This underscores the importance of meticulously selecting the subsequence length in a manner that is conducive to enhanced prediction efficacy.

4.8. Limitations

In this study, we have conducted preliminary explorations and research into the prediction method of long-term solar wind series, considering the irregularity and uncertainty of time. However, due to the limitations of the authors' abilities, the optimal approach has not yet been identified. Further exploration and research are required to address the remaining issues.

(1) The datasets included in this study encompass seven feature variables. It should be noted, however, that the solar wind is affected by a multitude of factors that extend beyond the scope of this study. Moreover, the datasets employed in this study have an hourly temporal resolution. Nevertheless, there is a plethora of more granular information on solar wind features currently available. Consequently, future studies may wish to consider incorporating this additional detail on solar wind forecasts as an input variable that affects the forecast results.

(2) During the experimental process, it was observed that the long-term series prediction exhibited reduced sensitivity to outliers. This indicates that the dataset contains a certain number of outliers, and the proposed model is unable to predict these outliers effectively. This is an issue that requires further in-depth research and exploration. Consequently, future research should focus on differentiating between erroneous values and outliers in solar wind data and investigate how to enhance the detection and processing capabilities of the model in the presence of a considerable number of outliers in the dataset.

(3) Solar wind features encompass a diverse array of information types, including data, images, and frequencies. Future research will explore the potential of multimodal studies to fully leverage the diverse feature resources available and facilitate information sharing across different modalities. Integrating multiple information sources can lead to more accurate prediction and analysis of the solar wind sequence, thereby enhancing the performance and reliability of prediction models.

5. Conclusions

In this study, a long-term solar wind correlation coefficient prediction method based on Multi-Mode Decomp Block and Mode Independence Attention is employed to construct a deep learning model for long-term series prediction. The Multi-Mode Decomp Block exhibits outstanding performance in extracting information from multiple time modes, providing a more comprehensive feature analysis that aids in better understanding the complexity of the features. The Mode Independence Attention calculations on a per-mode basis enhance the feature analysis and contribute to a deeper understanding of feature complexity.

The experimental section predicts solar wind correlation coefficients for two datasets, encompassing multivariate as well as univariate sequences with different prediction lengths of 96, 192, 336, and 720 steps, respectively. The experimental comparison of the proposed model with existing models demonstrates that the proposed model, MoCoformer, outperforms the others in all cases and metrics. This validates the effectiveness of the model in predicting long-duration solar wind sequences. Furthermore, ablation experiments were conducted on the model components, demonstrating the efficacy of the two novel modules proposed in this chapter for solar wind dataset prediction. Moreover, adjustments and experimental comparisons of the model's crucial parameters were conducted to optimize their fit with solar wind predictions. The experimental results presented in this series provide additional evidence of the feasibility and superior performance of the methods proposed in this chapter.

It is regrettable that, despite the capacity of Mocoformer to diminish temporal irregularities in forecasts compared to state-of-the-art models, there persists a degree of uncertainty in solar wind long-time series forecasts due to the complexity of solar wind data, among other challenges. This is true regardless of the accuracy of the measurements and models used for forecasting. This has necessitated a more comprehensive investigation of solar wind.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in NASA GODDARD SPACE FLIGHT CENTER at https://omniweb.gsfc.nasa.gov (accessed on 29 May 2024).

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

References

1. Zhou, Y.; Feng, X. Numerical study of the propagation characteristics of coronal mass ejections in a structured ambient solar wind. *J. Geophys. Res. Space Phys.* **2017**, *122*, 1451–1462. [CrossRef]
2. Shen, F.; Yang, Z.; Zhang, J.; Wei, W.; Feng, X. Three-dimensional MHD simulation of solar wind using a new boundary treatment: comparison with in situ data at Earth. *Astrophys. J.* **2018**, *866*, 18. [CrossRef]
3. Guo, X.C.; Zhou, Y.C.; Wang, C.; Liu, Y.D. Propagation of Large-Scale Solar Wind Events in the Outer Heliosphere From a Numerical MHD Simulation. *Earth Planet. Phys.* **2021**, *5*, 223–231. [CrossRef]
4. Odstrcil, D. Modeling 3-D solar wind structure. *Adv. Space Res.* **2003**, *32*, 497–506. [CrossRef]
5. Tóth, G.; Sokolov, I.V.; Gombosi, T.I.; Chesney, D.R.; Clauer, C.R.; De Zeeuw, D.L.; Hansen, K.C.; Kane, K.J.; Manchester, W.B.; Oehmke, R.C.; et al. Space Weather Modeling Framework: A new tool for the space science community. *J. Geophys. Res. Space Phys.* **2005**, *110*. [CrossRef]
6. Wu, C.C.; Fry, C.D.; Wu, S.T.; Dryer, M.; Liou, K. Three-dimensional global simulation of interplanetary coronal mass ejection propagation from the Sun to the heliosphere: Solar event of 12 May 1997. *J. Geophys. Res. Space Phys.* **2007**, *112*. [CrossRef]
7. Detman, T.; Smith, Z.; Dryer, M.; Fry, C.D.; Arge, C.N.; Pizzo, V. A hybrid heliospheric modeling system: Background solar wind. *J. Geophys. Res. Space Phys.* **2006**, *111*. [CrossRef]

8. Feng, X.; Zhou, Y.; Wu, S.T. A novel numerical implementation for solar wind modeling by the modified conservation element/solution element method. *Astrophys. J.* **2007**, *655*, 1110. [CrossRef]

9. Bussy-Virat, C.D.; Ridley, A.J. Predictions of the solar wind speed by the probability distribution function model. *Space Weather* **2014**, *12*, 337–353. [CrossRef]

10. Wang, Y.M.; Sheeley, N.R., Jr. Solar wind speed and coronal flux-tube expansion. *Astrophys. J.* **1990**, *355*, 726–732. [CrossRef]

11. Arge, C.N.; Pizzo, V.J. Improvement in the prediction of solar wind conditions using near-real time solar magnetic field updates. *J. Geophys. Res. Space Phys.* **2000**, *105*, 10465–10479. [CrossRef]

12. Riley, P.; Linker, J.A.; Mikić, Z. An empirically-driven global MHD model of the solar corona and inner heliosphere. *J. Geophys. Res. Space Phys.* **2001**, *106*, 15889–15901. [CrossRef]

13. Altschuler, M.D.; Newkirk, G.Magnetic fields and the structure of the solar corona: I: Methods of calculating coronal fields. *Sol. Phys.* **1969**, *9*, 131–149. [CrossRef]

14. Owens, M.J.; Challen, R.; Methven, J.; Henley, E.; Jackson, D.R. A 27 day persistence model of near-Earth solar wind conditions: A long lead-time forecast and a benchmark for dynamical models. *Space Weather* **2013**, *11*, 225–236. [CrossRef]

15. Innocenti, M.E.; Lapenta, G.; Vršnak, B.; Crespon, F.; Skandrani, C.; Temmer, M.; Veronig, A.; Bettarini, L.; Markidis, S.; Skender, M. Improved forecasts of solar wind parameters using the Kalman filter. *Space Weather* **2011**, *9*. [CrossRef]

16. Liu, D.D.; Huang, C.; Lu, J.Y.; Wang, J.S. The hourly average solar wind velocity prediction based on support vector regression method. *Mon. Not. R. Astron. Soc.* **2011**, *413*, 2877–2882. [CrossRef]

17. Garrett, H.B.; Dessler, A.J.; Hill, T.W. Influence of solar wind variability on geomagnetic activity. *J. Geophys. Res.* **1974**, *79*, 4603–4610. [CrossRef]

18. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]

19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

20. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298. [CrossRef]

21. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [CrossRef]

22. Carbonneau, R.; Laframboise, K.; Vahidov, R. Application of machine learning techniques for supply chain demand forecasting. *Eur. J. Oper. Res.* **2008**, *184*, 1140–1154. [CrossRef]

23. Huang, X.; Mehrkanoon, S.; Suykens, J.A.K. Support vector machines with piecewise linear feature mapping. *Neurocomputing* **2013**, *117*, 118–127. [CrossRef]

24. Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]

25. Lahouar, A.; Slama, J.B.H. Hour-ahead wind power forecast based on random forests. *Renew. Energy* **2017**, *109*, 529–541. [CrossRef]

26. Shi, Z.; Ma, H.; Ren, M.; Wu, T.; Andrew, J.Y. A learning-based two-stage optimization method for customer order scheduling. *Comput. Oper. Res.* **2021**, *136*, 105488. [CrossRef]

27. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An introduction to decision tree modeling. *J. Chemom. Soc.* **2004**, *18*, 275–285. [CrossRef]

28. 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]

29. Gregor, K.; Danihelka, I.; Mnih, A.; Blundell, C.; Wierstra, D. Deep autoregressive networks. In Proceedings of the International Conference on Machine Learning, Beijing, China, 22–24 October 2014; pp. 1242–1250.

30. Chiarella, C.; He, X.Z.; Hommes, C. A dynamic analysis of moving average rules. *J. Econ. Dyn. Control* **2006**, *30*, 1729–1753. [CrossRef]

31. Benjamin, M.A.; Rigby, R.A.; Stasinopoulos, D.M. Generalized autoregressive moving average models. *J. Am. Stat. Assoc.* **2003**, *98*, 214–223. [CrossRef]

32. Box, G.E.P.; Pierce, D.A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Am. Stat. Assoc.* **1970**, *65*, 1509–1526. [CrossRef]

33. Poggi, P.; Marchetti, C.; Scelsi, R. Automatic morphometric analysis of skeletal muscle fibers in the aging man. *Anat. Rec.* **1987**, *217*, 30–34. [CrossRef] [PubMed]

34. Liu, H.; Shi, J. Applying ARMA—GARCH approaches to forecasting short-term electricity prices. *Energy Econ.* **2013**, *37*, 152–166. [CrossRef]

35. Magadum, R.B.; Bilagi, S.; Bhandarkar, S.; Patil, A.; Joshi, A. Short-term wind power forecast using time series analysis: Autoregressive moving-average model (ARMA). In *Recent Developments in Electrical and Electronics Engineering: Select Proceedings of ICRDEEE 2022*; Springer Nature: Singapore, 2023; pp. 319–341.

36. Grossberg, S. Recurrent neural networks. *Scholarpedia* **2013**, *8*, 1888. [CrossRef]

37. Graves, A.; Graves, A. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.

38. Chung, J.; Gulcehre, C.; Cho, K.H.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.

39. Backhus, J.; Rao, A.R.; Venkatraman, C.; Padmanabhan, A.; Kumar, A.V.; Gupta, C. Equipment Health Assessment: Time Series Analysis for Wind Turbine Performance. *Appl. Sci.* **2024**, *14*, 3270. [CrossRef]

40. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]

41. Child, R.; Gray, S.; Radford, A.; Sutskever, I. Generating long sequences with sparse transformers. *arXiv* **2019**, arXiv:1904.10509.

42. Fedus, W.; Zoph, B.; Shazeer, N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *J. Mach. Learn. Res.* **2022**, *23*, 1–39.

43. Rae, J.W.; Potapenko, A.; Jayakumar, S.M.; Lillicrap, T.P. Compressive transformers for long-range sequence modelling. *arXiv* **2019**, arXiv:1911.05507.

44. Pratt, W.K.; Kane, J.; Andrews, H.C. Hadamard transform image coding. *Proc. IEEE* **1969**, *57*, 58–68. [CrossRef]

45. Wang, S.; Li, B.Z.; Khabsa, M.; Fang, H.; Ma, H. Linformer: Self-attention with linear complexity. *arXiv* **2020**, arXiv:2006.04768.

46. Shaw, P.; Uszkoreit, J.; Vaswani, A. Self-attention with relative position representations. *arXiv* **2018**, arXiv:1803.02155.

47. Li, S.; Jin, X.; Xuan, Y.; Zhou, X.; Chen, W.; Wang, Y.X.; Yan, X. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. [CrossRef]

48. Liu, S.; Yu, H.; Liao, C.; Li, J.; Lin, W.; Liu, A.X.; Dustdar, S. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In Proceedings of the International Conference on Learning Representations, Virtual, 3 May 2021.

49. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 11106–11115.

50. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.

51. Kreifeldt, J.G. An analysis of surface-detected EMG as an amplitude-modulated noise. In Proceedings of the Internacional Conference Medicine and Biological Engineering, Chicago, IL, USA, 26–30 August 1989.

52. Cao, H.; Huang, Z.; Yao, T.; Wang, J.; He, H.; Wang, Y. InParformer: Evolutionary decomposition transformers with interactive parallel attention for long-term time series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37; pp. 6906–6915.

53. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [CrossRef]

54. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]