Received: 27 September 2025 | Revised: 18 March 2026 | Accepted: 27 March 2026 | Published online: 5 August 2026

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2026, Vol. 00(00) 1–8
DOI: 10.47852/bonviewJDSIS62027770

BON VIEW PUBLISHING

Time Series Prediction Network for Modeling Channel Correlation Based on Wavelet Transform

Hanyu Jiang[1,2] and Zhichao Zheng[3,*]

[1]*State Key Laboratory of Robotics and Intelligent Systems, Chinese Academy of Sciences, China*

[2]*University of Chinese Academy of Sciences, China*

[3]*School of Electrical Engineering and Automation, Wuhan University, China*

Abstract: With the advancement of technology, multi-channel time series prediction tasks have become a focal point of research. However, due to the complex and intertwined correlations between channels, it is challenging to model their interactions accurately and flexibly. Furthermore, most existing methods typically model data from either the time domain or frequency domain alone, leading to inefficient utilization of information. To overcome these challenges, we introduce MCCWT-Net, a multi-channel time series forecasting framework built upon the wavelet transform. Unlike transformer-based models, MCCWT-Net does not rely on self-attention mechanisms, resulting in a more efficient training process. The wavelet transform, when applied in the frequency domain, can be used to extract and reflect features with predictability (such as periodicity) and act as a filter to reduce noise, instead of relying solely on raw historical data. Our model incorporates a specific number of channels and, while modeling long-term features, effectively captures inter-channel correlation information in the frequency domain through wavelet transform, combining it with information from the time domain. Experimental results demonstrate that our model achieves advanced performance across multiple datasets.

Keywords: time series prediction, wavelet transform, channel relationships

1. Introduction

Time series forecasting [1–10] provides estimates of future trends or events [11, 12], allowing practitioners to make informed decisions in advance. Owing to this capability, it has become an essential technique in numerous scientific and engineering applications. Typical applications include predicting the remaining useful life of industrial machinery [13, 14], estimating future disease trends [15, 16], weather forecasting [17–19], energy consumption management [20–24], traffic flow prediction, and financial investment. In recent years, transformers have demonstrated outstanding performance across numerous domains, especially in time series forecasting, where accurately capturing long-term dependencies and complex temporal patterns is essential. This success can largely be attributed to their encoder–decoder architecture and powerful attention mechanisms, which allow the model to selectively focus on the most relevant parts of an input sequence rather than relying solely on sequential processing. By dynamically assigning importance weights to different time steps, transformers can effectively model both short-term fluctuations and long-range temporal correlations among data points. Such capabilities make them particularly well suited for forecasting tasks involving multivariate data, irregular patterns, and nonlinear dynamics. Furthermore, the parallel computation enabled by the attention mechanism significantly improves training efficiency compared with traditional recurrent structures, allowing transformers to scale effectively to large datasets and high-dimensional inputs. As a result, they have achieved remarkable success in a wide range of real-world applications, including energy consumption prediction, traffic flow analysis, financial trend modeling, and weather forecasting. These advantages have led to their widespread adoption among researchers and practitioners who continue to explore improved architectures, hybrid designs, and domain-specific adaptations to further enhance forecasting accuracy and robustness.

Time series prediction tasks [5, 6], such as weather forecasting, power demand prediction, and financial forecasting, can assist individuals in making better decisions by providing insights into future trends. Nonetheless, time series forecasting poses significant challenges because real-world sequences usually lack stationarity and are prone to noise interference [1]. In addition, time series prediction, especially multivariate time series forecasting, often involves interdependencies between variables. Modeling the correlations between these variables presents a formidable challenge.

The PatchTST [25] model draws inspiration from Vision Transformer [26], partitioning temporal data into patch-like subsequences and employing these segments as token inputs for transformer-based frameworks. Following the channel-independent setup in References [27, 28], it achieves excellent performance. Recently, simple channel-independent linear models, such as Dliner [27], have been designed, achieving comparable performance to transformer-based solutions. Furthermore, transformer-driven approaches for temporal forecasting, such as Informer [29], Autoformer [30], Reformer [31], and Pyraformer [32], do not explicitly model the channels in time series data. This neglect of inter-channel relationships can lead to information loss because channels are evidently highly correlated in certain datasets [10]. This paper models the relationships among variables from a frequency-domain perspective because frequency-domain methods have inherent

*Corresponding author: Zhichao Zheng, School of Electrical Engineering and Automation, Wuhan University, China. Email: zzc0101@whu.edu.cn

advantages over time-domain approaches. First, channel modeling in the frequency domain naturally separates high-frequency noise [1], thereby improving the quality of feature representations. Second, the ability of the frequency domain to capture periodic patterns is typically much stronger than that of the time domain [10, 33–35]. In the time domain, periodicity must be expressed through waveform repetition, requiring the model to remember such repetitions over long temporal spans. In the frequency domain, however, periodic signals correspond to distinct peaks in the spectrum, making periodic structures much easier to identify and model.

To model the complex and highly coupled correlations among different variables and to fully leverage information from both the frequency and time domains, thereby improving overall information utilization, we introduce a non-transformer architecture, called MCCWT-Net, a multi-channel time series forecasting framework built upon the wavelet transform. In this paper, we decompose time series prediction into a trend component and a seasonal component through sequence decomposition. The trend component is relatively easy to predict, so we use a multi-layer perceptron (MLP) for this task. For the seasonal component, due to the complexity among variables, we adopt wavelet transform to model the relationships between various variables in the time series prediction task in the frequency domain and subsequently integrate it with time-domain information to forecast seasonal patterns, thereby improving prediction performance. Experiments show that despite its low computational complexity and small number of parameters, our method achieves excellent performance, outperforming seven time series prediction methods on five datasets. Ablation experiments also verify the effectiveness of our method.

2. Method

2.1. Problem statement

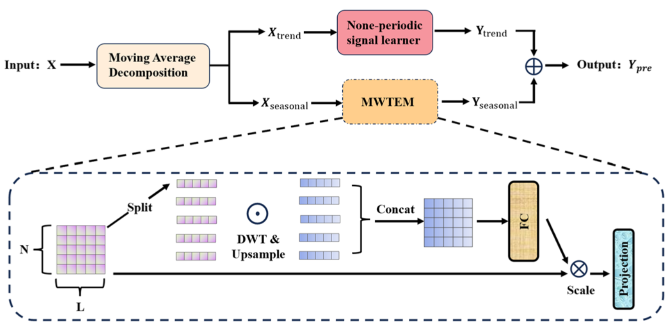
Time series forecasting [2] may be formalized by considering a look-back interval of length L, where the input corresponding to time ttt is written as $X_{t-L+1:t} = \{x_{t-L+1}, \ldots, x_t\}$, while the model aims to predict the future sequence $X_{t-L+1:t+O}$, where O denotes the prediction horizon. In long-term forecasting scenarios, O is generally larger than $L (O > L)$.

2.2. Model architecture

As depicted in Figure 1, MCCWT-Net employs a sequence decomposition strategy to achieve efficient time series forecasting. At the beginning, a moving average operation is employed to suppress short-term variations while capturing the long-term trend. For an input time series $X \in \mathbb{R}^{I \times C}$, the procedure is described as follows:

$$X_{\text{trend}} = AvgPool(\text{Padding}\ (X))_{\text{kernel}} \tag{1}$$

Figure 1
MCCWT-Net architecture diagram



$$X_{\text{seasonal}} = X - X_{\text{trend}}. \tag{2}$$

In this formulation, X_{trend} and $X_{\text{seasonal}} \in \mathbb{R}^{I \times C}$ correspond to the trend and seasonal components of the time series, respectively. The AvgPool operation is applied with proper padding to maintain the same output length as that of the input sequence. The decomposed trend component is passed through the non-periodic signal learner, which transforms it into the trend prediction Y_{trend} for the output sequence. Following the smoothing operation, the trend part can be regarded as a stable and slowly varying signal, which is generally easier to model than the seasonal part characterized by strong periodic fluctuations. Hence, a single MLP layer is employed to estimate the trend component, formulated as:

$$Y_{\text{trend}} = W \cdot X_{\text{trend}} + b. \tag{3}$$

In this context, X_{trend} and Y_{trend} denote the input and target sequences used for forecasting the trend component, respectively. W denotes the weights assigned to the MLP, while b is the bias in the MLP layer. Conversely, the seasonal component exhibits rapid variations and short-term randomness, and the model requires higher resolution and more complex modeling techniques; otherwise, it cannot effectively simulate the time series.

Recent studies suggest that the variables in the seasonal component of time series often show significant inter-variable correlations. Nevertheless, most channel-independent models employ identical parameters for all channels, neglecting the diversity of inter-channel relationships. While such parameter sharing simplifies model design, it overlooks the distinct behavior of each channel and consequently limits performance, particularly when channel dependencies are weak or vary significantly across the dataset.

To overcome the above limitations, we introduce a multi-channel wavelet transform enhancement module (MWTEM) designed for long-range time series prediction. The proposed module integrates features from the time and frequency domains to model the inter-channel correlations of the seasonal component in the time series. By feeding the seasonal component into the MWTEM, the resulting output X_{seasonal} is produced. Finally, X_{seasonal} is added to X_{trend}, resulting in the final output of our model for time series prediction.

2.3. Discrete wavelet transform (DWT)

Due to its compact support, binary formulation, and orthogonal properties, the Haar wavelet transform [36, 37] has been widely adopted in applications such as image compression, edge detection, and digital logic design. The first-level one-dimensional Haar wavelet and its corresponding scaling function can be expressed as follows:

$$\begin{cases} \phi_1(x) = \frac{1}{\sqrt{2}}\phi_{1,0}(x) + \frac{1}{\sqrt{2}}\phi_{1,1}(x) \\ \psi_1(x) = \frac{1}{\sqrt{2}}\phi_{1,0}(x) - \frac{1}{\sqrt{2}}\phi_{1,1}(x) \end{cases}. \tag{4}$$

Here, $\phi_{j,k}(x)$ can be defined as:

$$\phi_{j,k}(x) = \sqrt{2^j}\phi\left(2^j x - k\right), k = 0, 1, \cdots, 2^j - 1. \tag{5}$$

Here, j refers to the Haar basis level, and k specifies the index of the corresponding basis function. In addition, the definition of $\phi_{0,0}(x)$ [37] is:

$$\phi_{0,0}(x) = \phi_0(x) = \begin{cases} 0, & x < 0 \\ 1, & 0 \leq x < 1k = 2^j - 1. \\ 0, & x \geq 1 \end{cases} \tag{6}$$

Thus, the level-1 Haar transform can be represented in terms of the level-0 Haar basis function:

$$\begin{cases}\phi_1(x) = \phi_0(2x) + \phi_0(2x-1) \\ \psi_1(x) = \phi_0(2x) - \phi_0(2x-1)\end{cases}. \tag{7}$$

This indicates that the input signal is divided into two parts, each of which is downsampled to half of its original size, but the overall length remains unchanged. This suggests that the structure of the wavelet transform is lossless. The two components can be viewed as corresponding to low-pass and high-pass filtering operations. The low-pass filter isolates the low-frequency part of the signal, which describes the global trend or smooth regions, while the high-pass filter emphasizes the high-frequency part, highlighting abrupt variations and detailed structures.

Here, we provide a mathematical explanation for why wavelet transform is effective in modeling inter-channel relationships in time series forecasting.

$$x(t) = \sum_s \sum_\tau W(s,\tau)\psi_{s,\tau}(t), \tag{8}$$

where $x(t)$ is the original time-domain signal, s denotes the scale, τ represents the translation in time, determining the position of the wavelet along the time axis, $\psi_{s,\tau}(t)$ is the wavelet basis function, and $W(s,\tau)$ is the wavelet coefficient. The wavelet transform decomposes a signal into multiple scale components, allowing the model to capture both high-frequency fluctuations and low-frequency trends. Modeling channel correlations on these decomposed components enables the network to learn cross-variable dependencies at different frequency scales.

2.4. Multi-channel wavelet transform enhancement module (MWTEM)

In Figure 1, the detailed internal structure of our MWTEM module can be seen. The workflow of MWTEM can be described as follows:

$$v_0, v_1, \ldots, v_{n-1} = \mathrm{Split}(X_{t-L+1:t}) \tag{9}$$

$$\mathrm{Freq}^i = \mathrm{DWT}_j\left(v^i\right) \tag{10}$$

$$\mathrm{Freq} = \mathrm{stack}\left(\mathrm{Freq}^0, \mathrm{Freq}^1, \ldots, \mathrm{Freq}^{n-1}\right) \tag{11}$$

$$F_{\mathrm{cr}} = \sigma(W_2\delta(W_1\mathrm{Freq})) \tag{12}$$

$$g(X_{t-L+1:t}) = F_{\mathrm{cr}} \tag{13}$$

$$X_{t+1:t+o} = f(X_{t-L+1:t} \cdot g(X_{t-L+1:t})). \tag{14}$$

The process starts by decomposing the input matrix $X \in R^{N\times L}$ into n channel-specific vectors $\{v_0, v_1, \ldots, v_{n-1}\} \in R^{1\times L}$ through a split along the channel dimension, with n representing the total channel count. The underlying motivation for conducting channel decomposition is that when using the wavelet transform to extract frequency information from the entire sequence, the information within each channel might not be extracted independently, potentially leading to distortion due to interference from other channels.

Next, a wavelet transform [36, 37] is applied to each channel component. The application of the wavelet transform yields both high-frequency and low-frequency outputs, resulting in a twofold increase in channels and a corresponding reduction of the sequence length by 50%, so that the dimension becomes $R^{2\times L/2}$. However, the total amount of information remains unchanged. Therefore, we perform upsampling on the channel components after the wavelet transform to restore the original sequence length, resulting in a dimension of $R^{2\times L/2}$. As presented in Equation (11), the frequency channel vectors are concatenated along the channel axis to form the tensor $\mathtt{Fre} \in \mathbb{R}^{N\times L}$. A fully connected module is then employed to learn the cross-channel relationships in the frequency domain, resulting in $F\,cr \in \mathbb{R}^{N\times L}$. By performing this step, the model uncovers the frequency interactions present across channels. It then outputs the corresponding frequency-adaptive weight vector, denoted as $F\,cr$ or $g(x)$, conditioned on the input data.

The process concludes by computing the element-wise product of the input sequence and the frequency correlation tensor, producing a weighted representation, so that the dimension is still $R^{N\times L}$. Consequently, the network can adapt to the frequency-domain features of the input without requiring the extra computation associated with an inverse transform. To control the output length for prediction, this weighted representation is further projected through a fully connected layer $f(x)$, resulting in the seasonal forecast $Y_{\mathrm{seasonal}} \in R^{N\times O}$, where O denotes the prediction horizon. In this way, MWTEM can capture fine-grained frequency-domain information for each channel independently, minimizing interference from other channels. Subsequently, an MLP is applied to model the frequency characteristics of each channel, facilitating interactions among channels and integrating latent frequency features. An adaptive weight vector, derived from the frequency-domain information of the input, is thus computed. The vector is then multiplied element-wise with the input and processed through a linear projection to generate the ultimate output of the network.

Consequently, the proposed approach strengthens the model's capacity to capture and represent frequency-domain features while combining them with temporal information, resulting in improved prediction accuracy.

2.5. Loss function

MSE intuitively measures the discrepancy between predictions and ground truth, and its continuous differentiability ensures smooth and stable gradients that facilitate optimization. Therefore, we adopt MSE as the loss function in this work. The loss function is defined as follows:

$$L\left(Y, \widehat{Y}\right) = \frac{1}{|Y|}\sum_{i=1}^{|Y|}\left(y_{(i)} - \hat{y}_{(i)}\right)^2, \tag{15}$$

where $\widehat{Y}$ denotes the model's predictions and Y represents the corresponding ground-truth values.

3. Experiments

3.1. Experimental settings

We evaluated our approach on five real-world multivariate time series datasets, summarized in Table 1. We compared our method,

Table 1
Statistics of datasets

Dataset	Total variables	Sample rate	Total records
Electricity	321	1 h	26304
ETTm1	7	15 min	69680
Exchange	8	1 Day	7588
Traffic	862	1 h	17544
Weather	21	10 min	52695

MCCWT-Net, with several other approaches, including LSTM [38], Informer, FEDformer [33–35], Autoformer, LightTS [39], DSformer, and iTransformer, to evaluate its performance. Below is a description [40] of the datasets used in our experiments:

Electricity: Hourly electricity consumption data for 321 customers, covering the period between 2012 and 2014, are provided.

ETTm1: It includes 15-min interval measurements of transformer power load and oil temperature collected between 2016 and 2018.

Exchange: It contains daily currency exchange data for eight nations collected between 1990 and 2016.

Traffic: It includes hourly records of highway traffic occupancy collected by 862 sensors in the San Francisco Bay Area between 2015 and 2016.

Weather: It provides time-series data of 21 weather-related variables collected every 10 min at the Max Planck Institute for Biogeochemistry.

The input sequence length was set to $T = 96$, and prediction horizons were chosen from $S \in \{96,192,336,720\}$. Normalization was applied independently to individual time slots to avoid information leakage across the dataset. In addition, to ensure training reproducibility, we enabled deterministic training during the training process. This setting ensures that on our device, the training results remain exactly the same as long as no other settings are changed. Evaluation was conducted using MSE [41] and MAE [42] metrics. Experiments were executed on a single NVIDIA RTX 4090 GPU (24 GB) with PyTorch [43], ensuring reproducibility. This consistent setup allows for direct comparison of results, minimizing variability and ensuring that performance differences are attributable to the methods being tested, rather than experimental conditions.

3.2. Comparative study

Multivariate long-term forecasting experiments are more challenging than univariate ones and provide a better test of a model's predictive performance in complex environments. Our multivariate long-term forecasting experimental results are displayed in Table 2. As shown, our proposed model, MCCWT-Net, achieves advanced results across nearly all benchmarks and prediction horizons, demonstrating both its effectiveness and stability. Specifically, MCCWT-Net achieved the best performance on the Traffic and Electricity datasets, while on the Weather dataset, it performed best for prediction horizons of 336 and 720. On the ETTm1 dataset, it excelled at prediction horizons of 96, 192, and 336. For other prediction lengths, although we did not achieve the absolute best performance, the results indicate that we attained the second-best or still competitive performance. Moreover, as the prediction horizon increased, our model consistently exhibited strong stability. For example, compared to MCCWT-Net, LSTM showed a sharp increase in prediction error on the Weather dataset as the prediction length grew. Specifically, the MSE for a prediction length of 720 increased by 77.94% compared to a prediction length of 96, while MCCWT-Net's MSE only grew by 13.44%. The findings indicate that our model is highly capable of accurately forecasting multivariate time series over extended horizons. Table 3 also presents an evaluation of its computational efficiency. Our model demonstrates a balance between computational efficiency and model complexity. Compared with iTransformer [44] and DSformer [45], our model achieves higher computational efficiency and better performance. On the Electricity dataset, our model surpasses iTransformer in overall performance while using only 4.8% of its FLOPs and 25.1% of its parameters.

Table 2

Results for multivariate long-term forecasting are shown, with the top-performing value highlighted in bold black, the second-best in red, and a dash (–) indicating that the model does not apply to the dataset

Models		Ours		LSTM		Informer		FEDformer		Autoformer		LightTS		DSformer		iTransformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	**0.475**	**0.327**	0.843	0.453	0.719	0.391	0.587	0.366	0.613	0.388	0.615	0.391	0.532	0.344	0.554	0.358
	192	**0.485**	**0.335**	0.847	0.453	0.696	0.379	0.604	0.373	0.616	0.382	0.601	0.382	0.557	0.355	0.541	0.344
	336	**0.501**	**0.343**	0.853	0.455	0.777	0.420	0.621	0.383	0.622	0.387	0.613	0.386	0.580	0.383	0.551	0.358
	720	**0.539**	**0.364**	1.500	0.805	0.864	0.472	0.626	0.382	0.660	0.408	0.658	0.407	0.624	0.380	0.591	0.379
Weather	96	0.175	0.228	0.369	0.406	0.847	0.752	0.217	0.296	0.266	0.336	0.182	0.242	**0.143**	**0.197**	0.174	0.214
	192	**0.217**	0.271	0.416	0.435	1.204	0.895	0.276	0.336	0.307	0.367	0.227	0.287	0.276	0.299	0.219	**0.254**
	336	**0.267**	**0.315**	0.455	0.454	1.672	1.036	0.339	0.380	0.359	0.395	0.282	0.334	0.329	0.331	0.278	0.296
	720	**0.334**	**0.368**	0.535	0.520	2.478	1.310	0.403	0.428	0.419	0.428	0.352	0.386	-	-	0.349	0.398
Electricity	96	**0.159**	**0.261**	0.375	0.437	0.274	0.368	0.193	0.308	0.201	0.317	0.207	0.307	0.195	0.274	0.198	0.277
	192	**0.174**	**0.275**	0.442	0.473	0.296	0.386	0.201	0.315	0.222	0.334	0.213	0.316	0.197	0.279	0.192	0.277
	336	**0.191**	**0.294**	0.439	0.473	0.300	0.394	0.214	0.329	0.231	0.338	0.230	0.333	0.215	0.292	0.218	0.299
	720	**0.222**	**0.320**	0.980	0.814	0.373	0.439	0.246	0.355	0.254	0.361	0.265	0.360	0.265	0.349	0.262	0.337
Exchange	96	0.087	0.214	1.453	1.049	0.847	0.752	0.148	0.278	0.197	0.323	0.116	0.262	0.170	0.320	**0.084**	**0.202**
	192	**0.161**	**0.296**	1.846	1.179	1.204	0.895	0.271	0.380	0.300	0.369	0.215	0.359	0.335	0.450	0.175	0.298
	336	**0.339**	**0.433**	2.136	1.231	1.672	1.036	0.460	0.500	0.509	0.524	0.377	0.466	0.652	0.641	0.361	0.435
	720	**0.763**	**0.661**	2.984	1.427	2.478	1.310	1.195	0.841	1.447	0.941	0.831	0.699	-	-	0.846	0.693
ETTm1	96	**0.327**	**0.369**	0.863	0.664	0.672	0.571	0.379	0.419	0.505	0.475	0.374	0.400	0.336	0.389	0.346	0.380
	192	**0.327**	**0.369**	1.113	0.776	0.795	0.669	0.426	0.441	0.553	0.496	0.400	0.407	0.392	0.399	0.381	0.395
	336	**0.405**	**0.420**	1.267	0.823	1.212	0.871	0.438	0.438	0.445	0.459	0.621	0.537	0.456	0.440	0.419	0.425
	720	0.489	0.480	1.324	0.858	1.166	0.823	0.543	0.490	0.671	0.561	0.527	0.502	0.487	0.494	**0.486**	**0.456**

Table 3
Comparison of model efficiency on the Weather and Electricity datasets for a prediction horizon of 336. FLOPs indicate the total floating-point operations, while Params refer to the number of model parameters

	Weather		Electricity	
Methods	**FLOPs (M)**	**Params (M)**	**FLOPs (M)**	**Params (M)**
Ours	4.7	0.2	201.0	1.6
iTrans-former	274.0	6.5	4188.0	6.5
DSformer	46.0	241.0	960.0	858.0

3.3. Ablation study

Ablation experiments were performed to assess the contributions of the DWT and the channel relationship modeling (CR) within the MCCWT-Net framework. Figures 2 and 3 summarize the results of our ablation experiments for different prediction lengths on the Weather and Electricity datasets.

Figure 2
The ablation study results on the Electricity dataset, where "w/o CR+DWT" refers to the model without channel relationships and discrete wavelet transform, and "w/o DWT" refers to the model with only the absence of the discrete wavelet transform

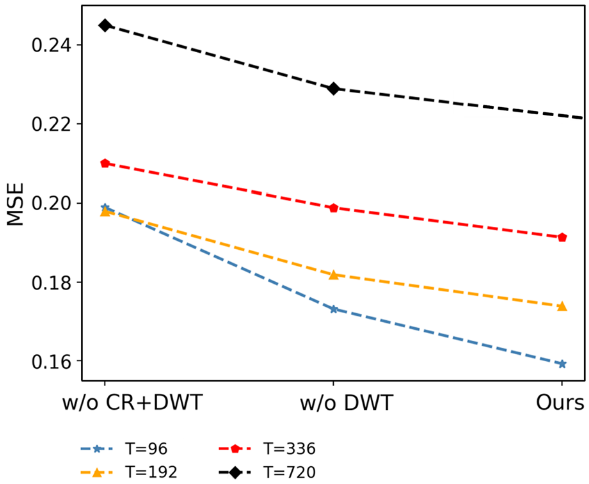
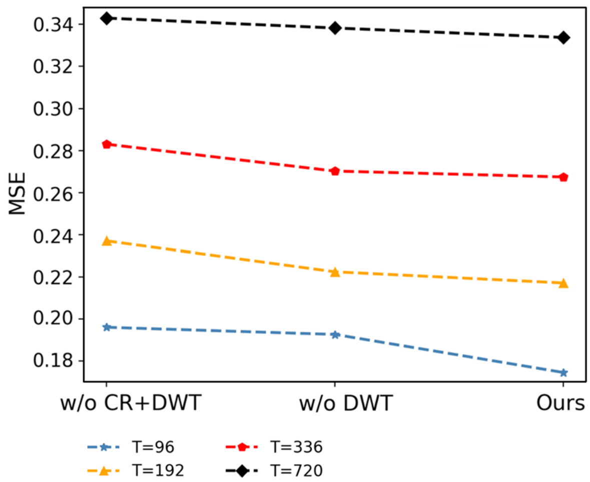


Figure 3
The ablation study results on the Weather dataset, where "w/o CR+DWT" refers to the model without channel relationships and discrete wavelet transform, and "w/o DWT" refers to the model with only the absence of the discrete wavelet transform



The experimental results show that removing both DWT and CR leads to the highest prediction errors. Retaining CR while removing DWT further improves prediction accuracy, demonstrating that modeling the relationships between variables in multivariate time-series data is a key technique for enhancing prediction accuracy. When both CR and DWT are retained, the lowest prediction errors and the highest prediction accuracy are achieved across all prediction lengths, confirming the synergistic effect of modeling inter-variable relationships and using wavelet transforms to capture frequency dependencies. It is worth noting that the Weather dataset contains 21 variables, while the Electricity dataset includes as many as 321 variables, resulting in substantial differences in data complexity and inter-variable interactions between the two benchmarks. The relatively smaller number of variables in the Weather dataset limits the diversity of cross-channel dependencies that can be exploited during feature learning. In contrast, the Electricity dataset exhibits significantly higher dimensionality, where each variable corresponds to a distinct electricity consumption pattern, often influenced by shared temporal trends such as daily cycles, seasonal variations, and collective demand fluctuations. This high-dimensional setting naturally introduces richer and more intricate relationships among variables, including both strong correlations and subtle complementary patterns. Consequently, the synergy brought by modeling inter-channel dependencies becomes more pronounced in the Electricity dataset. In particular, the application of wavelet transforms enables the model to analyze signals across multiple frequency bands, effectively capturing both global trends and localized variations across channels. Such multi-scale representations enhance the model's ability to exploit cross-variable correlations and suppress noise, thereby improving forecasting performance. Therefore, the richer relational structure inherent in the Electricity dataset makes the use of wavelet-based channel relationship modeling more effective and beneficial compared with datasets containing fewer variables.

3.4. Forecasting results visualization

In the experiments on the Electricity dataset, we fixed the input sequence length and varied the prediction horizons, visualizing the long-term forecasting outcomes of MCCWT-Net in Figures 4–7. The results indicate that MCCWT-Net consistently outperforms

Figure 4
Forecasting results on the Electricity dataset (input: 96, prediction: 96)

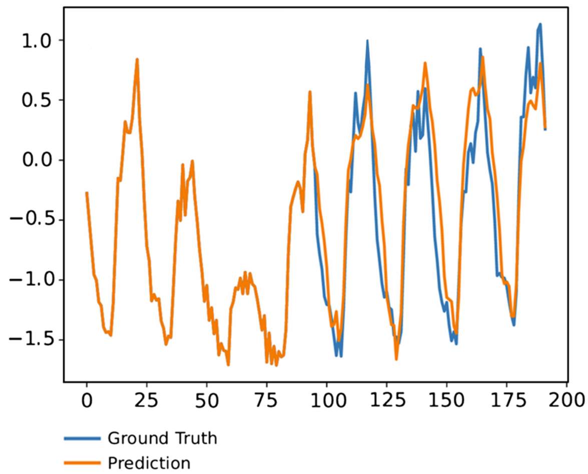


Figure 5
Forecasting results on the Electricity dataset (input: 96, prediction: 192)

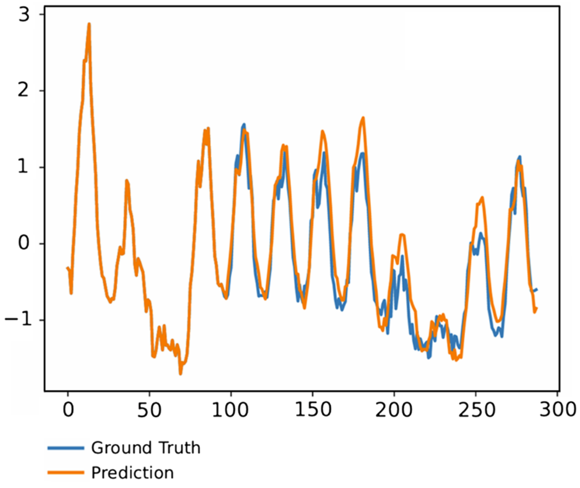


Figure 6
Forecasting results on the Electricity dataset (input: 96, prediction: 336)

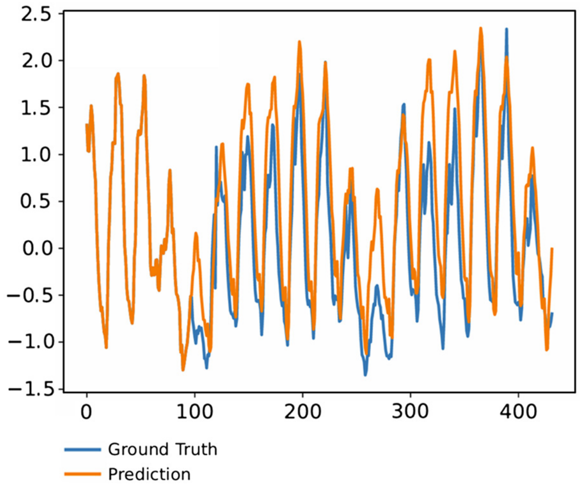
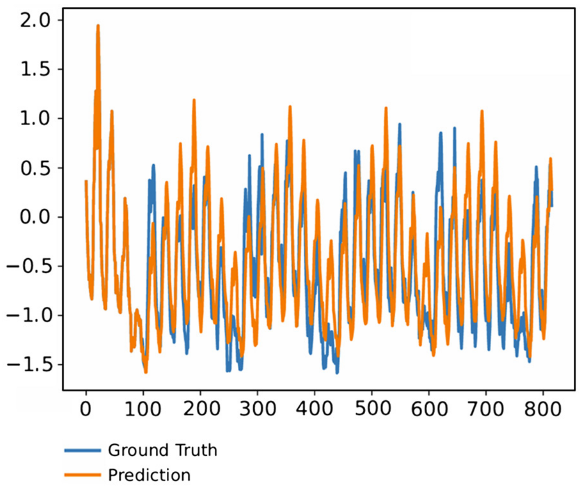


Figure 7
Forecasting results on the Electricity dataset (input: 96, prediction: 720)



baseline methods across all prediction lengths, effectively capturing complex periodic patterns and long-range dependencies. The superior performance mainly stems from the model's capability to incorporate frequency-domain information, which helps in capturing heterogeneous frequency patterns among variables. Furthermore, by learning inter-channel relationships, MCCWT-Net can more accurately represent the effects of different variables on the target variable, resulting in a more informative representation of the multivariate time series.

4. Conclusion

This paper proposes a non-transformer architecture, MCCWT-Net, which is a multi-channel time series forecasting framework that integrates wavelet transform with a channel modeling mechanism. The model innovatively captures inter-variable dependencies in the frequency domain and fuses frequency-domain representations with temporal information, thereby enhancing multivariate modeling capability. The effectiveness of MCCWT-Net primarily stems from two aspects. First, the wavelet transform extracts periodic and predictable components in the frequency domain while suppressing high-frequency noise, enabling the model to better capture complex temporal patterns. Second, the channel modeling mechanism explicitly learns inter-variable dependencies within the frequency domain, avoiding the information loss commonly observed in channel-independent approaches. This frequency–time fusion strategy allows the model to exhibit stronger robustness and generalization ability when dealing with highly dynamic and non-stationary sequences.

Compared with recent mainstream channel-independent approaches (e.g., DLinear), MCCWT-Net explicitly models cross-variable correlations in the frequency domain, leading to superior long-term forecasting performance on datasets characterized by strong coupling and high variability. Relative to frequency-enhanced models such as FEDformer, the proposed framework maintains a simpler architectural design, resulting in fewer parameters and higher computational efficiency, while achieving stable performance across multiple benchmarks.

Despite its strong empirical results, MCCWT-Net still has certain limitations. Although our model outperforms seven baseline methods across five datasets, there remains a performance gap compared with the latest state-of-the-art models. In addition, the current architecture is relatively simple and does not incorporate transformer-based modules, which—despite their higher computational cost—are known to substantially improve modeling capacity. Therefore, future work will focus on integrating transformer-based mechanisms to model inter-variable relationships in the frequency domain and further narrow the performance gap with recent advanced approaches.

Acknowledgement

The authors sincerely thank the editor and anonymous reviewers for their valuable suggestions, which helped in improving the quality of this paper. I am grateful to Professor Zhao Wentao for providing research funding; even though we did not work closely together, your guidance has been truly inspiring. Many thanks to Dr. Ma Fuyu for providing equipment and technical support, and for accompanying me throughout my academic journey. Special appreciation also goes to my pet hamster, Little Lee-Yan—your cute, round presence brought endless joy, even if your occasional nibbles sometimes disrupted my work. I am deeply sorry for my negligence that allowed you to escape from the cage and fall into the toilet where you drowned. Whenever I think of this incident, I am overwhelmed with grief.

Funding Support

This work is supported by the National Natural Science Foundation of China (U23A20645).

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The ETT dataset was acquired at https://github.com/zhouhaoyi/ETDataset. The Exchange dataset was acquired at https://github.com/thuml/Autoformer. The Weather dataset was acquired at https://www.bgc-jena.mpg.de/wetter/. The Traffic dataset was acquired at http://pems.dot.ca.gov/. The Electricity dataset was acquired at https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014.

Author Contribution Statement

Hanyu Jiang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Zhichao Zheng:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Supervision.

References

[1] Stefenon, S. F., Seman, L. O., Aquino, L. S., & Coelho, L. D. S. (2023). Wavelet-Seq2Seq-LSTM with attention for time series forecasting of level of dams in hydroelectric power plants. *Energy*, *274*, 127350. https://doi.org/10.1016/j.energy.2023.127350

[2] Jiang, M., Zeng, P., Wang, K., Liu, H., Chen, W., & Liu, H. (2023). FECAM: Frequency enhanced channel attention mechanism for time series forecasting. *Advanced Engineering Informatics*, *58*, 102158. https://doi.org/10.1016/j.aei.2023.102158

[3] Nagy, A. M., & Simon, V. (2021). Improving traffic prediction using congestion propagation patterns in smart cities. *Advanced Engineering Informatics*, *50*, 101343. https://doi.org/10.1016/j.aei.2021.101343

[4] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

[5] Sun, S., Wang, S., & Wei, Y. (2020). A new ensemble deep learning approach for exchange rates forecasting and trading. *Advanced Engineering Informatics*, *46*, 101160. https://doi.org/10.1016/j.aei.2020.101160

[6] Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied Intelligence*, *6*(3), 205–213. https://doi.org/10.1007/BF00126626

[7] Zhu, J., Liu, D., Chen, H., Liu, J., & Tao, Z. (2025). DTSFormer: Decoupled temporal-spatial diffusion transformer for enhanced long-term time series forecasting. *Knowledge-Based Systems*, *309*, 112828. https://doi.org/10.1016/j.knosys.2024.112828

[8] Qiu, X., Wu, X., Lin, Y., Guo, C., Hu, J., & Yang, B. (2025). DUET: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, 1185–1196. https://doi.org/10.1145/3690624.3709325

[9] Tian, X., Huang, S., Xiao, J., Wang, H., & Liu, Y. (2025). SDVS-Net: A spatial dilated convolution and variable self-attention network for multivariate long-term time series forecasting. *Neurocomputing*, *619*, 129148. https://doi.org/10.1016/j.neucom.2024.129148

[10] Luo, Y., Zhang, S., Lyu, Z., & Hu, Y. (2025). TFDnet: Time–frequency enhanced decomposed network for long-term time series forecasting. *Pattern Recognition*, *162*, 111412. https://doi.org/10.1016/j.patcog.2025.111412

[11] Zhang, F., Wang, M., Zhang, W., & Wang, H. (2025). THATSN: Temporal hierarchical aggregation tree structure network for long-term time-series forecasting. *Information Sciences*, *692*, 121659. https://doi.org/10.1016/j.ins.2024.121659

[12] Li, L., Zhou, X., Hu, G., Li, S., & Jia, D. (2024). A recurrent spatio-temporal graph neural network based on latent time graph for multi-channel time series forecasting. *IEEE Signal Processing Letters*, *31*, 2875–2879. https://doi.org/10.1109/LSP.2024.3479917

[13] Jing, T., Zheng, P., Xia, L., & Liu, T. (2022). Transformer-based hierarchical latent space VAE for interpretable remaining useful life prediction. *Advanced Engineering Informatics*, *54*, 101781. https://doi.org/10.1016/j.aei.2022.101781

[14] Zhou, J., Qin, Y., Chen, D., Liu, F., & Qian, Q. (2022). Remaining useful life prediction of bearings by a new reinforced memory GRU network. *Advanced Engineering Informatics*, *53*, 101682. https://doi.org/10.1016/j.aei.2022.101682

[15] Wang, Y., Lv, Z., Sheng, Z., Sun, H., & Zhao, A. (2022). A deep spatio-temporal meta-learning model for urban traffic revitalization index prediction in the COVID-19 pandemic. *Advanced Engineering Informatics*, *53*, 101678. https://doi.org/10.1016/j.aei.2022.101678

[16] Ghassemi, M., Pimentel, M., Naumann, T., Brennan, T., Clifton, D., Szolovits, P., & Feng, M. (2015). A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *29*(1), 446–453. https://doi.org/10.1609/aaai.v29i1.9209

[17] Doycheva, K., Horn, G., Koch, C., Schumann, A., & König, M. (2017). Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning. *Advanced Engineering Informatics*, *33*, 427–439. https://doi.org/10.1016/j.aei.2016.11.001

[18] Huang, M., Lin, R., Huang, S., & Xing, T. (2017). A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. *Advanced Engineering Informatics*, *33*, 89–95. https://doi.org/10.1016/j.aei.2017.05.003

[19] Zeng, P., Hu, G., Zhou, X., Li, S., Liu, P., & Liu, S. (2022). Muformer: A long sequence time-series forecasting model based on modified multi-head attention. *Knowledge-Based Systems*, *254*, 109584. https://doi.org/10.1016/j.knosys.2022.109584

[20] Yang, F., Yan, K., Jin, N., & Du, Y. (2023). Multiple households energy consumption forecasting using consistent modeling with privacy preservation. *Advanced Engineering Informatics*, *55*, 101846. https://doi.org/10.1016/j.aei.2022.101846

[21] Runge, J., & Zmeureanu, R. (2022). Deep learning forecasting for electric demand applications of cooling systems in buildings.

Advanced Engineering Informatics, *53*, 101674. https://doi.org/10.1016/j.aei.2022.101674

[22] Abdallah, M., Talib, M. A., Hosny, M., Waraga, O. A., Nasir, Q., & Arshad, M. A. (2022). Forecasting highly fluctuating electricity load using machine learning models based on multimillion observations. *Advanced Engineering Informatics*, *53*, 101707. https://doi.org/10.1016/j.aei.2022.101707

[23] Alizadeh, R., Allen, J. K., & Mistree, F. (2022). Surrogate models and time series for flow prediction on the red river dam network. In *Proceedings of the ASME 2022 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Volume 3B: 48th Design Automation Conference*. https://doi.org/10.1115/DETC2022-88163

[24] Bentsen, L. Ø., Warakagoda, N. D., Stenbro, R., & Engelstad, P. (2023). Spatio-temporal wind speed forecasting using graph networks and novel transformer architectures. *Applied Energy*, *333*, 120565. https://doi.org/10.1016/j.apenergy.2022.120565

[25] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 1–24.

[26] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., …, & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 1–21.

[27] Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(9), 11121–11128. https://doi.org/10.1609/aaai.v37i9.26317

[28] Han, L., Ye, H.-J., & Zhan, D.-C. (2024). The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, *36*(11), 7129–7142. https://doi.org/10.1109/TKDE.2024.3400008

[29] Zhou, H., Li, J., Zhang, S., Zhang, S., Yan, M., & Xiong, H. (2023). Expanding the prediction capacity in long sequence time-series forecasting. *Artificial Intelligence*, *318*, 103886. https://doi.org/10.1016/j.artint.2023.103886

[30] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, 1717.

[31] Kitaev, N., Kaiser, L., & Levskaya, A. (2020). Reformer: The efficient transformer. In *International Conference on Learning Representations*, 1–12.

[32] Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A. X., & Dustdar, S. (2022). Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International Conference on Learning Representations*, 1–20.

[33] Yi, K., Zhang, Q., Fan, W., Wang, S., Wang, P., He, H., …, & Niu, Z. (2023). Frequency-domain MLPs are more effective learners in time series forecasting. In *37th International Conference on Neural Information Processing Systems*, 3349.

[34] Liu, P., Wu, B., Li, N., Dai, T., Lei, F., Bao, J., …, & Xia, S.-T. (2024). WFTNet: Exploiting global and local periodicity in long-term time series forecasting. In *2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5960–5964. https://doi.org/10.1109/ICASSP48485.2024.10446883

[35] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, *162*, 27268–27286.

[36] Zhang, Z., Guo, D., Zhou, S., Zhang, J., & Lin, Y. (2023). Flight trajectory prediction enabled by time-frequency wavelet transform. *Nature Communications*, *14*(1), 5258. https://doi.org/10.1038/s41467-023-40903-9

[37] Xu, G., Liao, W., Zhang, X., Li, C., He, X., & Wu, X. (2023). Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation. *Pattern Recognition*, *143*, 109819. https://doi.org/10.1016/j.patcog.2023.109819

[38] Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., & Borodulin, A. (2024). Applications of long short-term memory (LSTM) networks in polymeric sciences: A review. *Polymers*, *16*(18), 2607. https://doi.org/10.3390/polym16182607

[39] Campos, D., Zhang, M., Yang, B., Kieu, T., Guo, C., & Jensen, C. S. (2023). LightTS: Lightweight time series classification with adaptive ensemble distillation. *Proceedings of the ACM on Management of Data*, *1*(2), 171. https://doi.org/10.1145/3589316

[40] Jiang, M., Wang, K., Sun, Y., Chen, W., Xia, B., & Li, R. (2023). MLGN: Multi-scale local-global feature learning network for long-term series forecasting. *Machine Learning: Science and Technology*, *4*(4), 045059. https://doi.org/10.1088/2632-2153/ad1436

[41] Zhang, W., Huang, J., Wang, R., Wei, C., Huang, W., & Qiao, Y. (2024). Integration of Mamba and transformer—MAT for long-short range time series forecasting with application to weather dynamics. In *2024 International Conference on Electrical, Communication and Computer Engineering*, 1–6.

[42] Wang, Z., Kong, F., Feng, S., Wang, M., Yang, X., Zhao, H., …, & Zhang, Y. (2025). Is Mamba effective for time series forecasting? *Neurocomputing*, *619*, 129178. https://doi.org/10.1016/j.neucom.2024.129178

[43] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., …, & Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 721.

[44] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 1–25.

[45] Yu, C., Wang, F., Shao, Z., Sun, T., Wu, L., & Xu, Y. (2023). DSformer: A double sampling transformer for multivariate time series long-term prediction. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 3062–3072. https://doi.org/10.1145/3583780.3614851

How to Cite: Jiang, H., & Zheng, Z. (2026). Time Series Prediction Network for Modeling Channel Correlation Based on Wavelet Transform. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS62027770