QWaveNet: Quantum-Enhanced Wavelet Network for Time Series Forecasting

Fan Zhang
Shandong Technology and Business University
Yantai 264005, China
zhangfan@sdtbu.edu.cn

Shijun Chen
Shandong Technology and Business University
Yantai 264005, China
2024410055@sdtbu.edu.cn

Meijia Wang
Shandong Technology and Business University
Yantai 264005, China
2024410048@sdtbu.edu.cn

Shimin Fan
Shandong Technology and Business University
Yantai 264005, China
2024410061@sdtbu.edu.cn

Zexuan Ma
Shandong Technology and Business University
Yantai 264005, China
2024410053@sdtbu.edu.cn

Hua Wang*
Ludong University
Yantai 264025, China
hwa229@163.com

Abstract

Series decomposition aims to separate long-term trajectories and periodic patterns in time series, thereby improving the specificity and interpretability of forecasting models. However, mainstream approaches usually equate this functional separation with an operational split between a smoothed baseline and fluctuating residuals. One line of methods approximates the trend via low-pass smoothing, which assumes long-term trajectories to be purely smooth and may misassign nonlinear evolution, stage-wise transitions, and structural fluctuations to the residual. The other relies on heuristic nonlinear decompositions to capture complex dynamics, but is sensitive to noise and hyperparameters, leading to unstable and mixed component boundaries. To address this issue, we propose a forecasting framework, QWaveNet. It employs a quantum convolutional neural network to learn Dynamic Evolution, a long-term structural representation beyond low-pass smoothing priors, which preserves nonlinear evolution, stage-wise transitions, and structural fluctuations more coherently, while its complementary Evolutionary Residual captures localized variations and periodic dynamics. Guided by these complementary components, QWaveNet further performs adaptive multi-scale decomposition and reconstruction to support forecasting. Experiments on multiple benchmark datasets demonstrate the effectiveness of QWaveNet.

1 Introduction

Across machine learning, learning informative representations and modeling complex dependencies have become important research directions in diverse applications Yao et al. [2026], Wang et al.

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

[2024a, 2026], Xiao et al. [2026a], Zhou et al. [2026]. In current time series modeling, series decomposition is of great importance [Cheng et al., 2025, Das and Barman, 2025]. On the one hand, it can break down a complex time series process into components with clear functional roles. The long-term trend describes the overall evolution of the series over a longer time scale, while periodic patterns reveal recurring structures. On the other hand, a well-designed decomposition can improve the efficiency of information extraction. It allows the model to learn global structure and local variations in a more targeted manner, and reduces the risk that different types of information are obscured within the same representation. Therefore, designing an effective decomposition mechanism to better exploit the value of different information components is a key issue for improving time series forecasting performance.

However, many trend–seasonality decomposition methods operate in practice more like splitting a series into a low-pass smoothed baseline and a fluctuation residual. In terms of mechanism, one class of methods relies on linear filtering assumptions, such as the moving average (MA) Zhou et al. [2022], Zeng et al. [2023] and the Hodrick-Prescott (HP) filter Qin et al. [2022]. These methods usually treat the low-pass smoothed term as the trend and regard the remaining part as the seasonal residual. It should be noted that an operational split based on "smooth vs. non-smooth" is not equivalent to the true division of labor between long-term trend and periodic patterns. A long-term trend is not necessarily strictly smooth. It may involve nonlinear evolution and stage-wise structural changes, which appear in the time domain as structured undulations. Such undulations are not recurring periodic patterns. If "trend" is constrained a priori to be a purely smooth term, these structural undulations are often misassigned to the residual channel. This breaks long-term information into fragments and mixes it with non-periodic disturbances. Meanwhile, the residual may still retain slowly drifting long-term information. As a result, the decomposition boundary becomes ambiguous and fails to produce a clear functional separation. This, in turn, limits the subsequent forecasting model in targeted modeling and effective integration of long-term structure and periodic patterns. Another class of methods attempts to capture nonlinear structure through more flexible strategies. For example, STL Cleveland et al. [1990] uses LOESS smoothing to estimate the long-term trend and periodic patterns. LMD Nasir et al. [2023] extracts product-function components by iteratively removing the local mean. SSA Wang et al. [2024b] obtains the dominant structural and periodic components via singular value decomposition. Although these methods improve the ability to represent nonlinear changes to some extent, their nonlinearity often depends on heuristic rules, predefined basis functions, or adaptive sifting procedures. As a result, the decomposition outcomes are sensitive to parameter choices and noise levels. They may suffer from component mixing and instability. Therefore, it remains difficult for these methods to achieve a stable and clear separation between the long-term trend and periodic patterns.

As illustrated in Figure 1, decompositions based on linear filtering often fail to extract long-term trend information sufficiently. Some structural changes remain in the original series, making the residual channel highly similar in shape to the original series. This reflects the leakage and misallocation of long-term information in the decomposition results. In contrast, more flexible nonlinear decompositions can absorb part of the nonlinear variation. However, the extracted trend term still shows noticeable high-frequency oscillations, suggesting that some periodic patterns may remain. This leads to incomplete characteristics in the extracted periodic component, indicating that the decomposition is not yet thorough. Overall, both classes of methods suffer, to varying degrees, from incomplete decomposition and boundary mixing. They struggle to simultaneously highlight stable periodic patterns and interpretable long-term structural changes. This reduces the efficiency with which forecasting models can exploit structured information. Relying solely on existing decomposition paradigms is often insufficient to reveal the differences and synergy of multi-scale information in real-world series. Therefore, a more flexible and more expressive mechanism is needed to achieve more reliable functional separation and integration of information.

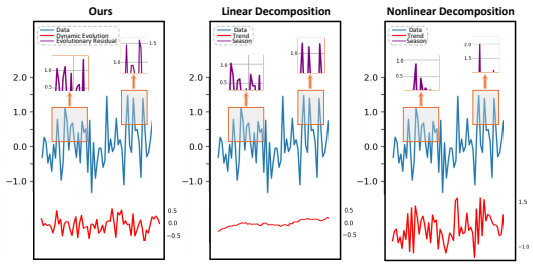


Figure 1: Comparison of decomposition methods in real-world dataset ETTh1. From left to right: our proposed quantum-guided decomposition, linear decomposition using moving average, and nonlinear decomposition using LMD.

To address the above challenges, we argue that it is necessary to establish clearer and more generalizable component boundaries. Specifically, the separation of the long-term trend should not be constrained a priori to a single low-pass smoothed baseline. Instead, it should be allowed to accommodate both linear and nonlinear evolution, phase-wise turning points, and structural undulations. A stable representation should be driven by the data's own dynamic characteristics Xiao et al. [2026b], Zhang et al. [2026], Xiao and Hou [2026]. On this basis, once the long-term trend is more completely extracted and effectively decoupled from other variations, the boundary for representing periodic patterns can be "purified". The periodic component can then focus on capturing recurring oscillatory patterns, without being contaminated by residual slow drift or structural changes. Motivated by this insight, this paper further explores a unified framework that supports both deep representation learning and feature decomposition, with the goal of improving decomposition stability and enhancing the efficiency of structured information utilization in downstream forecasting.

Inspired by the above criteria, we introduce Quantum Machine Learning (QML) as the representation learning module. With the development of quantum computing software stacks and simulators Bergholm et al. [2018], a Parameterized Quantum Circuit can be embedded into forecasting models to provide a structurally constrained nonlinear mapping for time series. The unitary transformations and entanglement structures induced by quantum gate sequences can create different representation geometries under a compact parameterization. This offers a structured way to learn the differences between long-term and short-term information.

Based on this idea, we propose a new long-term representation extracted from time series, termed Dynamic Evolution (DE). Functionally, it corresponds to the characterization of long-term trend-like information in traditional decomposition. However, we do not constrain it a priori to be a purely smoothed baseline. Specifically, we use the measurement outputs produced by a quantum convolutional network as DE. Leveraging the high-dimensional feature mapping of quantum states and the global interactions enabled by gate structures, this representation can accommodate long-term changes that combine linear and nonlinear dynamics, as well as phase-wise structural undulations. We then define the remaining part after removing DE from the original series as the Evolutionary Residual (ER). It is designed to carry more local variations and disturbance information, and it complements the long-term representation.

After obtaining these complementary representations, we further introduce a wavelet transform with an adaptive guidance mechanism to integrate them into forecasting and to organize multi-scale features in a stable manner. As a classical signal processing tool, the wavelet transform relies on a multi-scale analysis framework and localized basis functions. It enables time–frequency localized representations, and thus captures local patterns and frequency variations at different time scales. In this work, we treat the wavelet domain as the information carrier. We use the DE and the ER signals to guide the wavelet decomposition process. In this way, long-term and short-term information can be unfolded more orderly in multi-scale coefficients, and feature generalization for forecasting is achieved through reconstruction and fusion.

In summary, we propose QWaveNet, which combines quantum machine learning with classical signal processing. Specifically, it adopts a quantum convolutional neural network (QCNN) and a wavelet transform with adaptive guidance for time series modeling. This design addresses the bottlenecks of existing models in information decomposition and utilization. It provides an innovative solution to the limitations of traditional approaches and opens a new perspective for applying quantum machine learning.

Our main contributions can be summarized as follows:

- We propose a component partitioning perspective for time series modeling and introduce a QCNN as a structured representation learning module. This enables a more thorough decomposition and complementary representations of long-term and short-term information.
- We design a wavelet transform module with an adaptive guidance mechanism. It preserves multi-scale feature extraction while adaptively adjusting the decomposition process based on upstream representations, thereby organizing and utilizing decomposed information more effectively.
- We integrate quantum-inspired feature learning and classical wavelet-based multi-scale analysis into a unified forecasting framework. Extensive comparisons and ablation studies

on multiple datasets validate its effectiveness, providing a new practical solution for time series modeling that combines quantum-inspired models with signal processing.

2 Related Work

QML originated at the intersection of quantum computing and classical machine learning, aiming to accelerate the training and inference processes of machine learning algorithms through the potential of quantum computing. Quantum computing possesses unique properties such as superposition and entanglement, which enable the parallel processing of large-scale data for specific types of problems. The applications of QML are extensive and span multiple domains, including chemistry Motta and Rice [2022], computer vision Larasati et al. [2022], and natural language processing Widdows et al. [2024]. In the field of time series, the application of QML is still in its early stages, but some studies have demonstrated its significant potential Khan et al. [2024], Choudhary et al. [2025]. For instance, a quantum model using amplitude embedding has been proposed to solve time series classification problems Cuéllar et al. [2023]. Hybrid quantum architectures have also been introduced for time series forecasting, including Quantum Neural Networks and Hybrid Quantum Neural Networks Rivera-Ruiz et al. [2022]. Furthermore, the L-QLSTM, which combines quantum Long Short-Term Memory (LSTM) with Linear layers, has been shown to improve performance in carbon price forecasting Cao et al. [2023]. A detailed analysis can be found in Appendix B.

Signal processing techniques are various methods used to analyze, transform, enhance, or restore signals. They are widely applied across multiple fields. In the field of time series, the application of signal processing techniques has also been increasing. For example, FilterNet proposes applying two learnable filters to time series forecasting, which efficiently enhances high-frequency signals, thereby improving predictive performance Yi et al. [2024]. Additionally, Empirical Mode Decomposition has been used for time series decomposition and compared with other existing decomposition methods Van Jaarsveldt et al. [2023]. Fractional Fourier Transform has also been introduced to time series prediction by RNNs, with experimental results showing it outperforms ordinary Fourier Transform Koç and Koç [2022]. Furthermore, a combination of ridge regression, high-order FCM, and empirical wavelet transform has been used together to complete time series forecasting Mohammadi et al. [2023]. Spatiotemporal forecasting also explores evolving graph patterns and adaptive graph learning Wang et al. [2023a], Ma et al. [2026].

Currently, there is limited research that combines signal processing techniques with QML. However, integrating the advantages of signal processing with the potential of QML offers a new approach to time series analysis and forecasting. Our approach achieves this by incorporating signal processing techniques, particularly in the feature decomposition and utilization processes of time series, while harnessing the powerful capabilities of quantum computing to enhance the accuracy of the model. This combination leverages the strengths of both fields, enabling more efficient and precise analysis of time series data.

3 Method

To enable effective decomposition and full utilization of different information components in time series, we propose QWaveNet, as shown in Figure 2. It consists of three core modules: Quantum-Enhanced Feature Extraction (QFE), Guided Adaptive Wavelet Transform (GAWT), and Optimize-Driven Prediction (ODP). Before feature modeling, we normalize each variable to a unified scale and embed the sequence into a fixed-dimensional representation space, providing stable and consistent inputs for subsequent modules.

3.1 Quantum-Enhanced Feature Extraction

In quantum machine learning, although a variety of architectures are available, we choose QCNN as the core feature extraction module mainly because it can capture local dependencies in time series and enhance the representation of complex patterns through a structured mapping formed by parameterized quantum gates. Compared with fully connected networks, the convolutional structure provides local perception and parameter sharing, which is better suited for modeling temporal and spatial dependencies. Building on this, QCNN further introduces coupling and a nonlinear

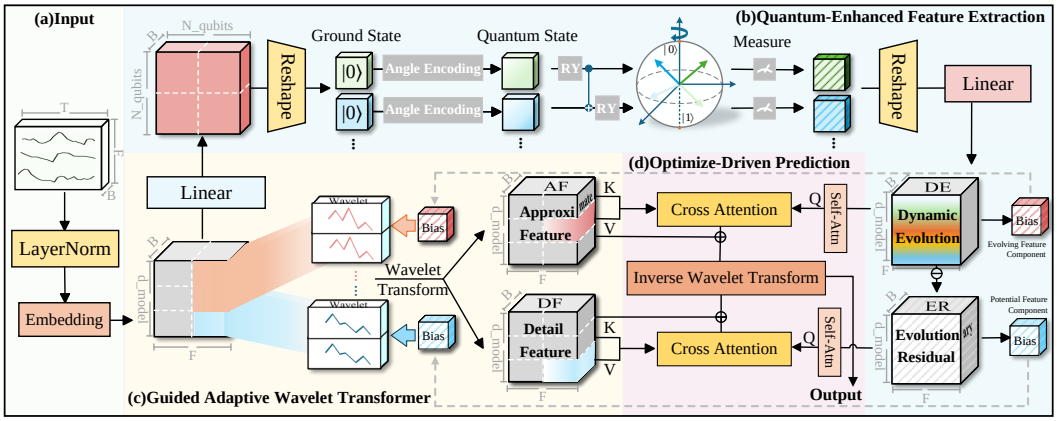


Figure 2: Overall framework of QWaveNet.

representation space induced by quantum gate sequences. This enables richer feature interactions and allows the model to better characterize complex nonlinear variations.

To reduce the simulation cost of quantum circuits and mitigate the risk of overfitting caused by high-dimensional mappings, we first apply an overall compact transformation to the input data $X = [x_1, x_2, \ldots, x_v]$ before feeding it into the quantum convolution module. This compact transformation operates across multiple aspects and dimensions, but it follows a unified linear transformation form:$\hat{X} = X \cdot \prod_{i=0}^{v} W_i$, where v represents the dimensionality of the input data, and W_i represents the linear transformation matrix. The compact mapping is designed to preserve the dominant information structure while suppressing redundant noise, thereby providing more stable inputs for subsequent representation learning.

In QCNN, qubits play a central role, analogous to neurons in classical convolutional neural networks. The raw data are mapped onto qubits through quantum state encoding. Each qubit represents a feature of the input data or a combination of features. Classical data must be converted into quantum states before they can be processed by quantum computation. Initially, all qubits are in the ground state:

$$|\psi(0)\rangle = |0\rangle^{\otimes n}. \tag{1}$$

We use angle encoding to map the compacted data features into quantum states, with different features corresponding one-to-one to qubits. The specific mapping approach we use is the RY rotation gate, which is a basic rotation gate in quantum gates that rotates the qubit's state around the Y-axis by an angle θ. Specifically, the RY gate rotates the state of the qubit, transforming it from the ground state to a new quantum state. Its standard matrix is given by:

$$RY(\theta) = \begin{bmatrix} \cos(\frac{\theta}{2}) & -\sin(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{bmatrix}. \tag{2}$$

After applying the rotation gate, the state of the qubit undergoes the following transformation:

$$|\psi_i\rangle = RY(\theta_i)|0\rangle = \cos\left(\frac{\theta_i}{2}\right)|0\rangle + \sin\left(\frac{\theta_i}{2}\right)|1\rangle, \tag{3}$$

where the rotation angle θ_i is derived from the feature x_i through a linear transformation:

$$\theta_i = \alpha x_i + \beta, \quad |\psi_{\text{enc}}(X)\rangle = \bigotimes_{i=1}^{n} RY(\theta_i)\,|0\rangle^{\otimes n}, \tag{4}$$

here, α is a scaling factor that controls the relationship between x_i and the qubit rotation angle, and β is an offset that shifts the range of rotation angles to ensure that θ_i lies within an appropriate interval. In implementation, we use the angle-encoding interface provided by PennyLane to directly map the normalized features to rotation angles. Therefore, α and β can be regarded as scale and bias parameters that are implicitly determined by the data preprocessing and encoding steps.

In traditional CNNs, the convolution operation performs a weighted summation on the input data using a sliding window to extract local features. In QCNN, we replace the traditional convolution operation with parameterized quantum gates. Specifically, we first use single-qubit RY rotation gates to rotate the state of each input qubit around a specified axis. We then introduce CNOT gates to establish entanglement between qubits and extract multi-dimensional features from the series. The formulation is given as follows:

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$
$$U_{\text{conv}}(\omega) = \prod_{i=1}^{n-1} \left(CNOT_{i,i+1} \cdot RY(\omega_i)_{\,i+1}\right), \quad |\psi(X)\rangle = U_{\text{conv}}(\omega)\,|\psi_{\text{enc}}(X)\rangle. \tag{5}$$

Qubit measurement is a crucial step in obtaining the results of quantum computations. To analyze the state of a qubit, we often use Pauli measurements. The Pauli-Z measurement is one of the most common measurements, as it can describe whether the qubit is in the $|0\rangle$ or $|1\rangle$ state, the specific formulas are as follows:

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \langle Z\rangle = \langle\psi|Z|\psi\rangle. \tag{6}$$

The result of a Pauli-Z measurement is either +1 or -1, corresponding to the qubit being in the $|0\rangle$ or $|1\rangle$ state, respectively. Through measurement, we can obtain the expectation value of the qubit, which represents the distribution between the $|0\rangle$ and $|1\rangle$ states. Finally, we perform Pauli-Z expectation measurements on the qubits to obtain the measurement outputs, which we define as DE. This quantity is a long-term–related latent representation learned by the parameterized quantum circuit under the guidance of the forecasting objective. It is used to characterize more stable evolution information with greater global consistency in the series. We provide a theoretical proof in Appendix A to illustrate how R_y encoding and CNOT entanglement enable the quantum circuit to produce cross-variable coupled responses, thereby supporting richer long-term structural representations in DE.

$$Z_k = I^{\otimes(k-1)} \otimes Z \otimes I^{\otimes(n-k)}, \quad k = 1, 2, \ldots, n,$$
$$z_k(X) = \langle\psi(X)|Z_k|\psi(X)\rangle \cdot W', \quad k = 1, 2, \ldots, n, \quad \text{DE}(X) = [z_1(X), z_2(X), \ldots, z_n(X)], \tag{7}$$

where W' represents the linear transformation matrix. Furthermore, we construct a complementary representation, the ER, in a residual manner to carry relatively local variations and disturbances: $\text{ER}(X) = X - \text{DE}(X)$. These two representations are jointly used as inputs to the subsequent wavelet-based multi-scale organization module, supporting collaborative modeling of information at different scales.

3.2 Guided Adaptive Wavelet Transform

After obtaining these complementary representations, a key question is how to incorporate them effectively into forecasting, and how to achieve stable organization and controllable reconstruction of information at different time scales within a unified framework. To this end, we further introduce a wavelet transform with an adaptive guidance mechanism Michau et al. [2022]. This design enables a structured expansion of multi-scale components in the wavelet domain and provides a consistent multi-scale interface for subsequent fusion modeling.

Traditional discrete wavelet transform typically uses fixed filters (e.g., Haar and Daubechies), whose coefficients are determined by a predefined mother wavelet and are difficult to adapt to the data structure. As a result, fixed filters may fail to capture useful information sufficiently in time series with mixed multi-scale characteristics. On the other hand, fully adaptive schemes that rely solely on learnable filters enlarge the optimization space. This makes early training more vulnerable to noise and local variations, leading to slow or unstable convergence.

To address this issue, we use the two complementary representations obtained upstream, DE and ER, as structural guidance. They provide priors for learning the wavelet filters, making the decomposition

more aligned with long-term and local information, and offering a consistent representation interface for subsequent fusion modeling.

Concretely, we abstract features from DE and the ER. We apply linear projections to compress the dimensionality while preserving key information, yielding two guidance vectors, denoted as the Evolving Feature Component and the Potential Feature Component. We then treat these guidance vectors as bias terms for the filter parameters and add them to the baseline weights of classical wavelet filters, thereby obtaining guided initial filter weights. The corresponding formulation is as follows:

$$\begin{aligned} &Bias_D = Component_D \cdot W'_1, \quad Bias_E = Component_E \cdot W'_2, \\ &\hat{W}_{low} = W_{low} + Bias_D, \quad \hat{W}_{high} = W_{high} + Bias_E, \end{aligned} \tag{8}$$

where $Component_D$ and $Component_E$ are shorthand for the Dynamic Evolution and Evolutionary Residual, respectively. Additionally, W'_1 and W'_2 represent the linear transformation weights, while W_{high} and W_{low} represent the original weights. The terms $\hat{W}_{high}$ and $\hat{W}_{low}$ represent the structurally guided weights.

The core idea of this design is that the guided weight bias adjusts the initial state of the wavelet filters as a prior. It provides a more reasonable starting point for training, making optimization more stable and convergence smoother. It also encourages filter updates to focus on structural directions consistent with the upstream representations. It should be emphasized that the guidance term does not replace the adaptive learning of the filters. As training proceeds, the filter coefficients are further refined in a data-driven manner to better characterize the global variations and local fluctuations of the series.

Regarding the specific wavelet operation, we apply a single-level Discrete Wavelet Transform (DWT) to decompose the original data features X using a pair of low-pass filter h_0 and high-pass filter h_1. This process results in the Approximate Feature (AF) and Detail Feature (DF). The specific process is as follows:$F_a[i] = \sum_{m=0}^{v-1} x_m \cdot h_0[i-m], \quad F_d[i] = \sum_{m=0}^{v-1} x_m \cdot h_1[i-m]$, where, F_a represents AF and F_d represents DF.

3.3 Optimize-Driven Prediction

To fully integrate the complementary strengths of QCNN and wavelet transforms, we explicitly inject the quantum-extracted dynamic representations into the wavelet decomposition and reconstruction process. To align and fuse the two types of representations effectively, we employ cross-attention to establish interactions between the quantum representations and the wavelet coefficients. It adaptively selects key information via similarity-based weighting, rather than relying on fixed rules, thereby improving cross-module information exchange and utilization. In addition, we introduce a causal mask to satisfy the autoregressive constraint, ensuring that information flows only from the past to the future and enhancing the rationality of the forecasting process.

The effectiveness of cross-attention depends heavily on how Query (Q), Key (K), and Value (V) are constructed. A well-designed QKV scheme can focus on critical information and directly shape the output representation. In this work, we use the quantum components enhanced by one layer of self-attention as Q, and take the wavelet-decomposed features as K and V. Similarity-based weighting is then applied to achieve quantum-guided refinement of the wavelet features. Specifically, we let Dynamic Evolution interact with the approximation features, and let the Evolutionary Residual interact with the detail features. We compute similarities and perform fusion for the two pairs separately. In this way, the model improves generalization while strengthening feature expressiveness, enabling it to capture both global patterns and local fluctuations.

Finally, we apply a residual connection between the fused representations obtained by cross-attention and the original wavelet coefficients. The result is used as the final multi-scale generalized coefficients, which are then fed into the inverse discrete wavelet transform (IDWT) to produce the reconstructed output $Y = [y_1, y_2, \ldots, y_v]$. This procedure ensures a smooth transition of multi-scale information while enabling collaborative forecasting of the long-term trend and short-term fluctuations. The formulation is given as follows: $y_i = \sum_{m=0}^{v-1} (F_a[m] \cdot h_0[i-m] + F_d[m] \cdot h_1[i-m])$.

Table 1: Averaged forecasting results. **Bold** indicates the best results, and underlining indicates the second-best results. The complete results are in the Appendix E.

Datasets	Ours		SimpleTM		iTransformer		PatchTST		FEDformer		MICN		TimesNet		ETSformer	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.406**	**0.428**	0.418	0.434	0.454	0.448	0.413	0.434	0.428	0.454	0.440	0.462	0.458	0.455	0.542	0.510
ETTh2	**0.330**	**0.381**	0.338	0.387	0.385	0.417	0.331	**0.381**	0.388	0.434	0.403	0.437	0.414	0.427	0.439	0.452
ETTm1	**0.352**	**0.380**	0.356	0.387	0.407	0.410	0.353	0.382	0.382	0.422	0.387	0.411	0.400	0.406	0.429	0.425
ETTm2	0.258	0.318	0.267	0.324	0.288	0.332	**0.256**	**0.317**	0.292	0.343	0.284	0.340	0.291	0.333	0.293	0.342
ECL	**0.157**	**0.252**	0.160	0.256	0.161	0.257	0.159	0.253	0.207	0.321	0.187	0.295	0.193	0.295	0.208	0.323
Weather	**0.223**	**0.261**	0.230	0.265	0.271	0.297	0.226	0.264	0.310	0.357	0.243	0.299	0.259	0.287	0.271	0.334
Solar	**0.186**	**0.244**	0.189	0.250	0.197	0.262	0.256	0.298	0.243	0.350	0.213	0.266	0.244	0.334	0.603	0.615
Exchange	0.378	0.410	0.387	0.419	0.458	0.469	0.405	0.426	0.477	0.477	**0.315**	**0.404**	0.416	0.443	0.410	0.427

Table 2: Averaged ablation experiments. **Bold** indicates the best results. The complete results are in the Appendix E.

Datasets	Ours		w/o QFE		w/o GAWT		CNN(1 Layer)		CNN(2 Layer)		CNN(3 Layer)		Attn→Add	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	**0.352**	**0.380**	0.359	**0.380**	0.363	0.382	0.369	0.393	0.371	0.397	0.379	0.401	0.359	0.379
ECL	**0.157**	**0.252**	0.161	0.255	0.166	0.260	0.188	0.293	0.173	0.274	0.173	0.271	0.161	0.255
Solar	**0.186**	**0.244**	0.253	0.320	0.252	0.315	0.207	0.281	0.201	0.286	0.217	0.289	0.255	0.324

4 Experiment

4.1 Forecasting Performance

Experiment Setup In the overall experiment, we used 8 classic time series datasets, including ETTh1, ETTh2, ETTm1, ETTm2, Electricity (ECL), Solar_Energy (Solar), Weather, and Exchange. The detailed descriptions of each dataset are provided in Appendix C. At the same time, we compared our model with 7 widely recognized benchmarks, including SimpleTM Chen et al. [2025], iTransformer Liu et al. [2023], PatchTST Nie et al. [2022], FEDformer Zhou et al. [2022], MICN Wang et al. [2023b], TimesNet Wu et al. [2022], ETSformer Woo et al. [2022]. We employed Mean Absolute Error (MAE) and Mean Squared Error (MSE) as evaluation metrics, attempting to comprehensively assess the practical application value of the model from multiple perspectives.

Main Results The forecasting results are reported in Table 1, with the corresponding radar chart provided in Appendix G. Lower MAE and MSE indicate better performance. Overall, QWaveNet achieves superior forecasting performance across multiple benchmark datasets. See the Appendix H for an analysis of when QWaveNet is more effective and where its gains are limited.

Compared with SimpleTM, QWaveNet reduces MSE by up to 8.36% and MAE by up to 5.2% on ECL across all forecasting horizons. It also consistently improves performance on ETTh2, ETTm2, and Solar, with the maximum MSE reduction reaching around 5% on each dataset. Compared with strong baselines such as iTransformer, PatchTST, and FEDformer, QWaveNet also demonstrates better feature generalization and information utilization. On Solar, it reduces MSE and MAE by up to 6.65% and 7.18% compared with iTransformer. Relative to PatchTST and FEDformer, it achieves average MSE reductions of 27.11% and 23.36%, and average MAE reductions of 17.95% and 30.26%, respectively. On ETTh2, QWaveNet further reduces the average MSE by 14.47% over iTransformer and 15.02% over FEDformer. These results demonstrate its strong forecasting capability and practical value for real-world time series forecasting.

Ablation Study In this study, we conduct a systematic ablation analysis of the key designs in QWaveNet, focusing on the impacts of QFE, GAWT, and ODP on overall performance, and the results are shown in Table 2. Specifically, we construct multiple comparison settings:

(1) replacing the QFE with the traditional MA decomposition method (w/o QFE); (2) removing GAWT (w/o GAWT) (Regarding the additive bias in GAWT's internal Equation 8, we explored alternative methods; our approach yielded the best results. Detailed results can be found in the Table 9); (3) replacing the quantum convolution structure in QFE with classical CNNs of different depths (1/2/3 layers), denoted as CNN (1 Layer), CNN (2 Layer), and CNN (3 Layer), to examine whether the structural benefits brought by the quantum module can be matched by an equivalent classical convolution; and (4) replacing cross-attention fusion with simple summation (Attn→Add) to evaluate

the role of attention in multi-source feature alignment and fusion. For the three CNN variants, we further conduct parameter searches and report the best results after careful tuning in the Table 8. The results show that removing either QFE or GAWT increases the forecasting error. When the quantum convolution is replaced by CNNs, performance degrades further and exhibits a consistent decline as the network depth increases. In addition, replacing attention with summation also leads to a clear performance drop. Taken together, these findings verify the critical contributions of QFE, GAWT, and the attention-driven fusion strategy to improving forecasting accuracy and feature utilization efficiency.

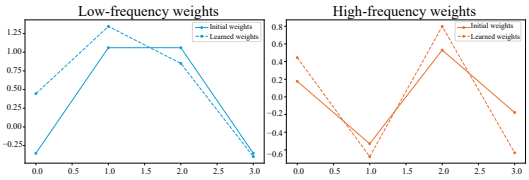


Figure 3: Variations in high-frequency and low-frequency weights of the GAWT before and after training.

4.2 Miscellaneous Analysis

Wavelet Weight Variation In our GAWT, the adaptivity and guidance of wavelet decomposition play a critical role in the model's overall performance. To visualize the detailed changes in wavelet decomposition before and after training, we present the wavelet weights in Figure 3. The wavelet type used in the experiments is Bior3.3. By observing the variations in low-frequency and high-frequency weights, it is evident that there are substantial differences between the pre-training and post-training states. This demonstrates that during training, the model effectively captures and adjusts features through the wavelet transform and, under the optimization of quantum computation, further enhances the representations of both low-frequency and high-frequency components.

Table 3: Low-frequency ratio of DE/Trend and high-frequency ratio of ER/Season on the Solar dataset.

Model	DE/Trend low-frequency ratio	ER/Season high-frequency ratio
Quantum	72.53%	64.62%
Wavelet	25.34%	74.41%
Moving-average	31.96%	71.53%
STL	33.11%	70.68%

Table 4: Performance comparison of QWaveNet across different numbers of qubits.

Dataset	ETTh1		ECL		Weather	
Metric	MSE	MAE	MSE	MAE	MSE	MAE
2 qubits	0.412	0.431	0.158	0.252	0.224	0.262
4 qubits	0.408	0.429	0.157	0.252	0.224	0.263
6 qubits	0.418	0.437	0.157	0.252	0.224	0.263

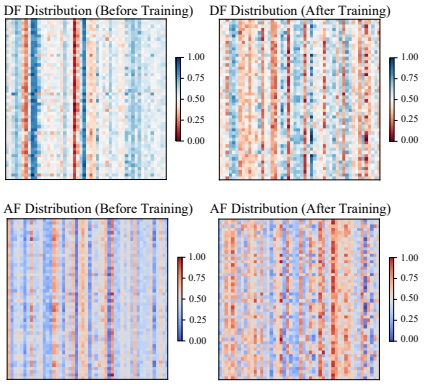


Figure 4: Normalized AF/DF distributions before and after training with a shared color scale.

Feature Decomposition Using the Solar dataset as an example, we visualize the coefficient distributions of AF and DF before and after training in Figure 4. It can be observed that the distribution patterns of both change noticeably after training. DF shifts from a relatively concentrated high-response structure before training to a more dispersed and balanced distribution. Meanwhile, the structural salience and contrast of AF increase, exhibiting clearer patterns. Furthermore, from a frequency-domain perspective, we quantify the above long/local-term functional separation by computing the low-frequency ratio of DE/Trend and the high-frequency ratio of ER/Season under different decomposition methods (see Table 3). The results show that the DE obtained by the quantum module has a higher low-frequency ratio, while ER also maintains a high high-frequency ratio. This indicates that slowly varying structures and rapid disturbances are more likely to be allocated to complementary components. Taken together, the coefficient visualizations and frequency-domain statistics suggest that the model not only reshapes the multi-scale energy layout in the wavelet domain, but also yields a more stable carrier for long-term information and a more concentrated characterization of local variations at the representation level.

Parameter Sensitivity From the averaged results in Table 4, the model maintains relatively stable performance across different numbers of qubits, with slight variations. On ETTh1, which has a

relatively small number of channels, increasing the number of qubits from 2 to 4 reduces both MSE and MAE, suggesting that moderate quantum embedding capacity may help capture cross-variable dependencies. However, further increasing the number to 6 leads to a slight performance degradation, potentially due to redundant representations. In contrast, ECL and Weather have higher channel dimensionality and more complex multivariate dependencies. Their performance remains consistently stable across different qubit configurations, suggesting that effective feature representations can be maintained with a limited number of qubits, potentially benefiting from the compact feature mapping. Overall, these results highlight the importance of balancing representation capacity and model complexity. A detailed analysis can be found in Appendix J.

5 Conclusion

This study proposes QWaveNet, a quantum-enhanced framework for time series forecasting. It aims to improve feature utilization under multi-scale information mixing. QWaveNet first uses QCNN to learn two complementary representations, DE and ER. It then performs adaptive wavelet-based decomposition and reconstruction, followed by cross-feature fusion for collaborative multi-scale modeling. Experiments show that QWaveNet improves forecasting accuracy and generalization, offering a modeling approach that combines quantum learning with signal processing.

Acknowledgements

This work was supported in part by the following: the National Natural Science Foundation of China under Grant Nos. U24A20219, 62272281, U24A20328, the Yantai Natural Science Foundation under Grant No. 2024JCYJ034, and the Youth Innovation Technology Project of Higher School in Shandong Province under Grant No. 2023KJ212.

Competing Interests. The authors declare no competing interests.

References

Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum science and technology*, 4(4):043001, 2019.

Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M Sohaib Alam, Guillermo Alonso-Linaje, B AkashNarayanan, Ali Asadi, et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018.

Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1):015004, 2022.

Yuji Cao, Xiyuan Zhou, Xiang Fei, Huan Zhao, Wenxuan Liu, and Junhua Zhao. Linear-layer-enhanced quantum long short-term memory for carbon price forecasting. *Quantum Machine Intelligence*, 5(2):26, 2023.

Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

Hui Chen, Viet Luong, Lopamudra Mukherjee, and Vikas Singh. Simpletm: A simple baseline for multivariate time series forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025.

Mingyue Cheng, Zhiding Liu, Xiaoyu Tao, Qi Liu, Jintao Zhang, Tingyue Pan, Shilong Zhang, Panjing He, Xiaohan Zhang, Daoyu Wang, et al. A comprehensive survey of time series forecasting: Concepts, challenges, and future directions. *Authorea Preprints*, 2025.

Prashant Kumar Choudhary, Nouhaila Innan, Muhammad Shafique, and Rajeev Singh. Hqnn-fsp: A hybrid classical-quantum neural network for regression-based financial stock market prediction. *arXiv preprint arXiv:2503.15403*, 2025.

Robert B Cleveland, William S Cleveland, Jean E McRae, Irma Terpenning, et al. Stl: A seasonal-trend decomposition. *J. off. Stat*, 6(1):3–73, 1990.

Iris Cong, Soonwon Choi, and Mikhail D Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019.

Manuel P Cuéllar, C Cano, Luis G Baca Ruíz, and Lorenzo Servadei. Time series quantum classifiers with amplitude embedding. *Quantum Machine Intelligence*, 5(2):45, 2023.

Pierre-Luc Dallaire-Demers and Nathan Killoran. Quantum generative adversarial networks. *Physical Review A*, 98(1):012324, 2018.

Pankaj Das and Samir Barman. Perspective chapter: An overview of time series decomposition and its applications. *Applied and Theoretical Econometrics and Financial Crises*, page 59, 2025.

Saad Zafar Khan, Nazeefa Muzammil, Salman Ghafoor, Haibat Khan, Syed Mohammad Hasan Zaidi, Abdulah Jeza Aljohani, and Imran Aziz. Quantum long short-term memory (qlstm) vs. classical lstm in time series forecasting: a comparative study in solar power forecasting. *Frontiers in Physics*, 12:1439180, 2024.

Emirhan Koç and Aykut Koç. Fractional fourier transform in time series prediction. *IEEE Signal Processing Letters*, 29:2542–2546, 2022.

Harashta Tatimma Larasati, Howon Kim, et al. Trends of quantum computing applications to computer vision. In *2022 International Conference on Platform Technology and Service (PlatCon)*, pages 7–12. IEEE, 2022.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in neural information processing systems*, 35:5816–5828, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature physics*, 10(9):631–633, 2014.

Jiaming Ma, Binwu Wang, Guanjun Wang, Kuo Yang, Zhengyang Zhou, Pengkun Wang, Xu Wang, and Yang Wang. Less but more: Linear adaptive graph learning empowering spatiotemporal forecasting. *Advances in Neural Information Processing Systems*, 38:58382–58418, 2026.

Gabriel Michau, Gaetan Frusque, and Olga Fink. Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proceedings of the National Academy of Sciences*, 119(8):e2106598119, 2022.

Hossein Abbasian Mohammadi, Sedigheh Ghofrani, and Ali Nikseresht. Using empirical wavelet transform and high-order fuzzy cognitive maps for time series forecasting. *Applied Soft Computing*, 135:109990, 2023.

Mario Motta and Julia E Rice. Emerging quantum computing algorithms for quantum chemistry. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(3):e1580, 2022.

Jawaria Nasir, Muhammad Aamir, Zahoor Ul Haq, Shehzad Khan, Muhammad Yusuf Amin, and Muhammad Naeem. A new approach for forecasting crude oil prices based on stochastic and deterministic influences of lmd using arima and lstm models. *IEEE Access*, 11:14322–14339, 2023.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Jae-Eun Park, Brian Quanz, Steve Wood, Heather Higgins, and Ray Harishankar. Practical application improvement to quantum svm: theory to practice. *arXiv preprint arXiv:2012.07725*, 2020.

Wenbo Qiao, Jiaming Zhao, and Peng Zhang. Quantum time-index models with reservoir for time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 1173–1184, 2025.

Quande Qin, Zhaorong Huang, Zhihao Zhou, Yu Chen, and Weigang Zhao. Hodrick–prescott filter-based hybrid arima–slfns model with residual decomposition scheme for carbon price forecasting. *Applied Soft Computing*, 119:108560, 2022.

Mayra Alejandra Rivera-Ruiz, Andres Mendez-Vazquez, and José Mauricio López-Romero. Time series forecasting with quantum machine learning architectures. In *Mexican international conference on artificial intelligence*, pages 66–82. Springer, 2022.

Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. The quest for a quantum neural network. *Quantum Information Processing*, 13(11):2567–2586, 2014.

Alessandro Sebastianelli, Daniela Alessandra Zaidenberg, Dario Spiller, Bertrand Le Saux, and Silvia Liberata Ullo. On circuit-based hybrid quantum neural networks for remote sensing imagery classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:565–580, 2021.

Cole Van Jaarsveldt, Gareth W Peters, Matthew Ames, and Mike Chantler. Tutorial on empirical mode decomposition: Basis decomposition and frequency adaptive graduation in non-stationary time series. *IEEE Access*, 11:94442–94478, 2023.

Binwu Wang, Yudong Zhang, Xu Wang, Pengkun Wang, Zhengyang Zhou, Lei Bai, and Yang Wang. Pattern expansion and consolidation on evolving graphs for continual traffic prediction. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2223–2232, 2023a.

Binwu Wang, Jiaming Ma, Pengkun Wang, Xu Wang, Yudong Zhang, Zhengyang Zhou, and Yang Wang. Stone: A spatio-temporal ood learning framework kills both spatial and temporal shifts. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2948–2959, 2024a.

Binwu Wang, Gaoyun Lin, Jiaming Ma, Qihe Huang, Zhengyang Zhou, Xu Wang, Pengkun Wang, and Yang Wang. Crisp: A spectral-based interaction strategy for multivariate time series forecasting. In *Forty-third International Conference on Machine Learning*, 2026. URL `https://openreview.net/forum?id=avYLuILgcr`.

Huiqiang Wang, Jian Peng, Feihu Huang, Jince Wang, Junhui Chen, and Yifei Xiao. Micn: Multi-scale local and global context modeling for long-term series forecasting. In *The eleventh international conference on learning representations*, 2023b.

Jing Wang, Xuhong Peng, Jindong Wu, Youde Ding, Barkat Ali, Yizhou Luo, Yiting Hu, and Keyao Zhang. Singular spectrum analysis (ssa) based hybrid models for emergency ambulance demand (ead) time series forecasting. *IMA Journal of Management Mathematics*, 35(1):45–64, 2024b.

Dominic Widdows, Aaranya Alexander, Daiwei Zhu, Chase Zimmerman, and Arunava Majumder. Near-term advances in quantum natural language processing. *Annals of Mathematics and Artificial Intelligence*, 92(5):1249–1272, 2024.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Etsformer: Exponential smoothing transformers for time-series forecasting. *arXiv preprint arXiv:2202.01381*, 2022.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

Canran Xiao and Liwei Hou. Prototype-aligned federated soft-prompts for continual web personalization. In *Proceedings of the ACM Web Conference 2026*, pages 6743–6754, 2026.

Canran Xiao, Jiabao Dou, Zhiming Lin, Zong Ke, and Liwei Hou. From points to coalitions: Hierarchical contrastive shapley values for prioritizing data samples. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 15995–16003, 2026a.

Canran Xiao, Tianxiang Xu, Siyuan Ma, Yiyang Jiang, Haoyu Gao, and Yuhan Wu. Reversible primitive–composition alignment for continual vision–language learning. In *International Conference on Learning Representations*, volume 2026, pages 88700–88722, 2026b.

Zhen Yao, Bo Lang, Zhihao Zheng, Haoyan Xu, Zhengtao Yao, Zhikang Dong, and Mooi Choo Chuah. From semantic to reasoing: A survey of language-guided segmentation in the multimodal era. 2026.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: Harnessing frequency filters for time series forecasting. *Advances in Neural Information Processing Systems*, 37:55115–55140, 2024.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

Jiayu Zhang, Chuangxin Zhao, Canran Xiao, Ruibo Duan, Wenyi Mo, Haoyu Gao, and Wenshuo Wang. Pi-cca: Prompt-invariant cca certificates for replay-free continual multimodal learning. In *The Fourteenth International Conference on Learning Representations*, volume 3, 2026.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

Heng Zhou, Jing Tang, Yanshu Li, Canran Xiao, Liwei Hou, Zong Ke, Jiawei Yao, et al. Comem: Compositional concept-graph memory for vision–language adaptation. In *International Conference on Learning Representations*, volume 2026, pages 20009–20032, 2026.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pages 27268–27286. PMLR, 2022.

A Theoretical Proof of Quantum Coupling

Regarding how quantum circuits exploit their structured nonlinear mapping capability to capture sequential features, our description is as follows.

Taking the simplest 2-qubit case as an example, two variables s_1 and s_2 start from the initial state $|00\rangle$. By applying R_y encoding, the inputs are mapped into a trigonometric nonlinear space. The states of the two qubits are then given by:

$$R_y(s_1)\,|0\rangle = \cos\left(\frac{s_1}{2}\right)|0\rangle + \sin\left(\frac{s_1}{2}\right)|1\rangle \tag{9}$$

$$R_y(s_2)\,|0\rangle = \cos\left(\frac{s_2}{2}\right)|0\rangle + \sin\left(\frac{s_2}{2}\right)|1\rangle \tag{10}$$

This shows that after R_y encoding, each variable is mapped into a trigonometric nonlinear representation composed of $\cos(\cdot)$ and $\sin(\cdot)$. Their joint state is given by:

$$|\psi_0\rangle = (R_y(s_1)\,|0\rangle) \otimes (R_y(s_2)\,|0\rangle) \tag{11}$$

After expansion, we obtain:

$$\begin{aligned} |\psi_0\rangle = {} & \cos\frac{s_1}{2}\cos\frac{s_2}{2}\,|00\rangle + \cos\frac{s_1}{2}\sin\frac{s_2}{2}\,|01\rangle \\ & + \sin\frac{s_1}{2}\cos\frac{s_2}{2}\,|10\rangle + \sin\frac{s_1}{2}\sin\frac{s_2}{2}\,|11\rangle\,. \end{aligned} \tag{12}$$

Now, we add a CNOT gate between the two qubits, where the first qubit is used as the control qubit and the second qubit as the target qubit. Its action on the computational basis states is given by

$|00\rangle \to |00\rangle$, $|01\rangle \to |01\rangle$, $|10\rangle \to |11\rangle$, and $|11\rangle \to |10\rangle$. Therefore, after applying the CNOT gate, we obtain:

$$|\psi_1\rangle = \cos\frac{s_1}{2}\cos\frac{s_2}{2}|00\rangle + \cos\frac{s_1}{2}\sin\frac{s_2}{2}|01\rangle + \sin\frac{s_1}{2}\sin\frac{s_2}{2}|10\rangle + \sin\frac{s_1}{2}\cos\frac{s_2}{2}|11\rangle. \quad (13)$$

Compared with the previous state, the coefficients of $|10\rangle$ and $|11\rangle$ are swapped. This means that the state of the second qubit is no longer determined only by s_2, but is also affected by s_1.

We further examine the Z measurement on the second target qubit:

$$P(\text{second qubit} = 0) = \cos^2\frac{s_1}{2}\cos^2\frac{s_2}{2} + \sin^2\frac{s_1}{2}\sin^2\frac{s_2}{2}, \quad (14)$$

$$P(\text{second qubit} = 1) = \cos^2\frac{s_1}{2}\sin^2\frac{s_2}{2} + \sin^2\frac{s_1}{2}\cos^2\frac{s_2}{2}. \quad (15)$$

The difference between the two is

$$\langle Z_2\rangle = \cos^2\frac{s_1}{2}\left(\cos^2\frac{s_2}{2} - \sin^2\frac{s_2}{2}\right) + \sin^2\frac{s_1}{2}\left(\sin^2\frac{s_2}{2} - \cos^2\frac{s_2}{2}\right). \quad (16)$$

Using the identity

$$\cos^2\frac{x}{2} - \sin^2\frac{x}{2} = \cos x, \quad (17)$$

we can simplify it as

$$\langle Z_2\rangle = \left(\cos^2\frac{s_1}{2} - \sin^2\frac{s_1}{2}\right)\cos(s_2) = \cos(s_1)\cos(s_2). \quad (18)$$

Therefore, we finally obtain

$$\langle Z_2\rangle = \cos(s_1)\cos(s_2). \quad (19)$$

This shows that after introducing the CNOT gate, the measurement output changes from a single-variable response to a cross-variable coupled response. The output after quantum transformation is finally used as DE.

This supports our claim that the DE representation is not intended to capture a single, purely low-pass smooth trend in the traditional sense, but rather a more realistic long-term structure. Such a structure may include slow linear evolution, nonlinear variations, stage-wise turning points, and organized structural fluctuations. Therefore, the value of the quantum circuit in our method lies in providing a coupled representation mechanism, allowing these long-term factors, which are otherwise easily fragmented or misclassified by low-pass smoothing priors, to be more completely encoded into DE, thereby forming clearer component boundaries.

B More Discussion on Quantum Machine Learning

As an interdisciplinary field that integrates quantum computing with artificial intelligence, Quantum Machine Learning (QML) has rapidly evolved into a prominent frontier in computational science. Its core concept lies in leveraging the fundamental principles of quantum superposition and entanglement to represent and process information within a high-dimensional Hilbert space far beyond the capacity of classical systems.

By encoding classical data into quantum states, QML models can exploit quantum parallelism to simultaneously explore multiple representations of the data, thereby enabling more expressive and compact feature mappings. Furthermore, the intrinsic entanglement among qubits facilitates the modeling of highly correlated and nonlinear dependencies that are often intractable for classical machine learning methods. These properties allow QML to provide a new computational paradigm for tasks requiring high-dimensional feature interactions, such as pattern recognition and temporal sequence modeling, offering significant potential for achieving exponential advantages in representational efficiency and computational scalability.

From a developmental perspective, the evolution of QML can be broadly divided into three stages. The first stage, the theoretical foundation phase, primarily focused on the quantization of classical learning algorithms, such as the Quantum Support Vector Machine (Quantum SVM) Park et al. [2020] and Quantum Principal Component Analysis (Quantum PCA) Lloyd et al. [2014]. Research during this period emphasized demonstrating the potential for exponential acceleration of quantum computing

in specific mathematical operations, including matrix inversion and eigenvalue decomposition. The second stage, characterized by the exploration of hybrid architectures, emerged with the advent of Variational Quantum Algorithms (VQAs) Cerezo et al. [2021]. Researchers began integrating quantum circuits with classical neural networks, forming Parameterized Quantum Circuits (PQCs) Benedetti et al. [2019] that enable quantum–classical collaborative learning through gradient-based optimization. These methods have been widely applied to classification, regression, and generative modeling tasks, establishing the foundation for hybrid quantum learning systems. The third stage, the application expansion phase, has been driven by the advancement of Noisy Intermediate-Scale Quantum (NISQ) Bharti et al. [2022] devices, which have enabled QML to be applied to real-world problems such as molecular energy level prediction, image recognition, text understanding, and time-series modeling, providing empirical support for the practical feasibility of quantum algorithms.

In terms of model architecture, researchers have proposed various Quantum Neural Network (QNN) Schuld et al. [2014] variants to simulate different types of learning processes. For example, the Quantum Convolutional Neural Network (QCNN) Cong et al. [2019] performs multi-layer feature extraction through quantum gate operations, while the Quantum Generative Adversarial Network (Quantum GAN) Dallaire-Demers and Killoran [2018] demonstrates strong distribution-fitting capabilities in generative modeling. Additionally, Hybrid Quantum Neural Networks (HQNNs) Sebastianelli et al. [2021] introduce learnable mappings between classical and quantum layers, enabling interaction between high-dimensional quantum spaces and low-dimensional classical feature spaces.

Against this research background, the model proposed in this study exhibits clear relevance and innovation within the evolutionary trajectory of QML. Existing QML research has primarily focused on algorithmic design and the verification of quantum acceleration feasibility. Although these studies have shown potential advantages across multiple domains, their applications in time-series forecasting remain relatively limited. In current forecasting research, information decomposition and utilization continue to face bottlenecks, while quantum computing, with its high-dimensional feature representation and intrinsic parallelism, provides a novel perspective and solution to overcome these limitations. Fundamentally, our proposed QWaveNet incorporates quantum computational principles into the entire process of decomposition, fusion, and reconstruction, achieving an organic integration of quantum information processing and time-series analysis. This fusion paradigm transcends the limitations of traditional feature decomposition methods, enabling the model to fully exploit dynamic dependencies among time–frequency features and thereby enhance predictive performance.

C Dataset Descriptions

To comprehensively evaluate the effectiveness and generalization ability of the proposed model, we conducted experiments on eight widely used benchmark datasets in the field of time series forecasting, covering diverse real-world domains such as energy, meteorology, and economics. These datasets [Zhou et al., 2021, Liu et al., 2022] include ETT (ETTh1, ETTh2, ETTm1, ETTm2), Electricity, Weather, Solar_Energy and Exchange. Each dataset exhibits distinct sampling frequencies, feature dimensions, and dynamic characteristics, providing a thorough validation environment for long-term and multivariate forecasting tasks. The specific introduction is as follows.

ETT (Electricity Transformer Temperature). The dataset is one of the most commonly used benchmarks for long-term time-series forecasting. It records transformer operating conditions from two regions in China between July 2016 and July 2018. ETTh1 and ETTh2 are sampled hourly, while ETTm1 and ETTm2 are sampled every 15 minutes, providing different temporal resolutions. Each sample contains a target variable, Oil Temperature (OT), along with six load-related features such as High-Useful Load (HUFL) and High-Unuseful Load (HULL). This dataset captures typical industrial time-series characteristics with evident seasonal trends, periodic fluctuations, and non-stationary dynamics, making it suitable for evaluating models' abilities to handle long-term dependencies and nonlinear variations.

Electricity. The Electricity dataset (ECL) contains hourly electricity consumption records from 321 clients, commonly used for multivariate electricity load forecasting. The dataset exhibits clear daily and weekly periodic patterns, as well as heterogeneous consumption behaviors among different clients. These characteristics make it suitable for evaluating models' abilities to capture complex temporal patterns and cross-variable dependencies in high-dimensional time series.

Weather. The Weather dataset consists of meteorological observations collected throughout 2020, sampled every 10 minutes. It contains 21 meteorological variables, including temperature, humidity, air pressure, and wind-related measurements. This dataset features strong multivariate coupling and non-stationary behaviors, making it suitable for evaluating forecasting performance under complex meteorological dynamics.

Solar_Energy. The Solar_Energy dataset contains solar power production records from 137 photovoltaic plants, sampled at 10-minute intervals. It exhibits strong daily periodicity and temporal variability, reflecting fluctuations in solar power generation. The primary forecasting task aims to predict future solar power output based on historical generation patterns, providing a benchmark for modeling strong periodicity and nonlinear temporal dynamics.

Exchange. The Exchange dataset records daily exchange rates of eight foreign currencies against the U.S. dollar, including the currencies of Australia, the United Kingdom, Canada, Switzerland, China, Japan, New Zealand, and Singapore. It features strong volatility and non-stationarity, making it suitable for evaluating forecasting models in financial and economic time series scenarios. This dataset is commonly used to assess nonlinear dependency modeling and multivariate correlation learning.

D Code for Reproducibility

Code is available at https://github.com/ShijunChen01/QWaveNet.

E Experimental Setting and Results

To ensure a fair and consistent comparison, we adopt the same training, validation, and testing splits for our model and all baseline methods. For the proposed model, the number of qubits is set within the range of 2 to 4 across different experimental configurations. The learning rate is tuned within the range of 0.0008 to 0.01. Unless otherwise specified, the remaining training and evaluation settings are kept consistent across methods. Our experiments are conducted on NVIDIA GeForce RTX 3090 GPUs with CUDA 12.4. The system runs CentOS Linux 7 with an Intel Core i7-7700K CPU.

This section supplements the experimental results that were not fully presented in the main text. It reports the complete MSE/MAE comparisons for all datasets under different forecasting horizons, and includes all baseline methods (see Table 5 and Table 6).

To further investigate the contribution of the DE/ER guidance mechanism, we introduce an additional ablation setting, termed Zeroed Quantum Guidance, in which the DE/ER guidance signals are removed while preserving the wavelet structure and the remaining components of QWaveNet. This controlled setting allows us to examine whether the performance improvements arise solely from the wavelet-based decomposition or also benefit from the quantum-derived guidance information.

As shown in Table 7, removing the DE/ER guidance leads to higher average forecasting errors across all three datasets. The performance degradation is particularly evident on ETTh1, where the guidance mechanism provides more noticeable improvements, especially for longer forecasting horizons. Similar trends are observed on ETTh2 and ETTm2, although the differences are relatively smaller.

These results indicate that the performance of QWaveNet does not solely depend on the wavelet decomposition structure. Instead, the DE/ER guidance provides additional information that contributes to adaptive multi-scale feature modeling and improves forecasting accuracy. This further supports the effectiveness of integrating quantum-derived representations with wavelet-based decomposition.

Table 8 presents additional comparisons between QWaveNet and CNN-based replacement baselines with different depths. In these variants, the original quantum module is replaced by 1-layer, 2-layer, and 3-layer CNNs, and each CNN baseline is carefully tuned by searching hidden sizes over $\{32, 64, 128\}$ and kernel sizes over $\{3, 5, 7\}$. Despite this careful tuning, QWaveNet consistently outperforms all CNN-based alternatives in terms of average MSE and MAE across the evaluated datasets. This indicates that the superiority of QWaveNet cannot be attributed merely to network capacity or hyperparameter selection.

Table 5: Forecasting results for prediction lengths $L \in \{96, 192, 336, 720\}$. For a fair comparison for all baselines, the lookback window T is searched among $\{96, 192, 336, 512, 720\}$.

Dataset	Method	Ours		SimpleTM		iTransformer		PatchTST		FEDformer		MICN		TimesNet		ETSformer	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.367**	0.397	**0.367**	**0.394**	0.386	0.405	0.370	0.400	0.376	0.415	0.398	0.427	0.384	0.402	0.494	0.479
	192	**0.412**	**0.423**	0.423	**0.423**	0.441	0.436	0.413	0.429	0.423	0.446	0.430	0.453	0.436	0.429	0.538	0.504
	336	**0.400**	**0.424**	0.416	0.441	0.487	0.458	0.422	0.440	0.444	0.462	0.440	0.460	0.491	0.469	0.574	0.521
	720	**0.447**	**0.467**	0.466	0.477	0.503	0.491	**0.447**	0.468	0.469	0.492	0.491	0.509	0.521	0.521	0.562	0.535
	Avg	**0.406**	**0.428**	0.418	0.434	0.454	0.448	0.413	0.434	0.428	0.454	0.440	0.462	0.458	0.455	0.542	0.510
ETTh2	96	**0.273**	**0.336**	0.281	0.346	0.311	0.363	0.274	0.337	0.332	0.374	0.299	0.364	0.340	0.374	0.340	0.391
	192	**0.324**	**0.376**	0.330	0.381	0.391	0.413	0.341	0.382	0.407	0.446	0.422	0.441	0.402	0.414	0.430	0.439
	336	**0.324**	**0.383**	0.340	0.388	0.415	0.437	0.329	0.384	0.400	0.447	0.447	0.474	0.452	0.452	0.485	0.479
	720	0.398	0.430	0.401	0.431	0.424	0.455	**0.379**	**0.422**	0.412	0.469	0.442	0.467	0.462	0.468	0.500	0.497
	Avg	**0.330**	**0.381**	0.338	0.387	0.385	0.417	0.331	**0.381**	0.388	0.434	0.403	0.437	0.414	0.427	0.439	0.452
ETTm1	96	**0.291**	**0.344**	0.292	0.346	0.334	0.368	0.293	0.346	0.326	0.390	0.316	0.364	0.338	0.375	0.375	0.398
	192	**0.329**	**0.366**	0.334	0.374	0.377	0.391	0.333	0.370	0.365	0.415	0.363	0.390	0.374	0.387	0.408	0.410
	336	**0.364**	**0.389**	0.365	0.395	0.426	0.420	0.369	0.392	0.391	0.425	0.408	0.426	0.410	0.411	0.435	0.428
	720	0.426	0.421	0.434	0.434	0.491	0.459	**0.416**	**0.420**	0.446	0.458	0.459	0.464	0.478	0.450	0.499	0.462
	Avg	**0.352**	**0.380**	0.356	0.387	0.407	0.410	0.353	0.382	0.382	0.422	0.387	0.411	0.400	0.406	0.429	0.425
ETTm2	96	0.169	0.258	0.171	0.259	0.180	0.264	**0.166**	**0.256**	0.180	0.271	0.179	0.275	0.187	0.267	0.189	0.280
	192	**0.223**	0.297	0.229	0.305	0.250	0.309	**0.223**	**0.296**	0.252	0.318	0.262	0.326	0.249	0.309	0.253	0.319
	336	0.276	0.331	0.284	0.339	0.311	0.348	**0.274**	**0.329**	0.324	0.364	0.305	0.353	0.321	0.351	0.314	0.357
	720	0.364	**0.384**	0.382	0.393	0.412	0.407	**0.362**	0.385	0.410	0.420	0.389	0.407	0.408	0.403	0.414	0.413
	Avg	0.258	0.318	0.267	0.324	0.288	0.332	**0.256**	**0.317**	0.292	0.343	0.284	0.340	0.291	0.333	0.293	0.342
ECL	96	**0.128**	**0.222**	0.140	0.234	0.131	0.227	0.129	**0.222**	0.186	0.302	0.164	0.269	0.168	0.272	0.187	0.304
	192	**0.140**	**0.237**	**0.140**	0.240	0.152	0.248	0.147	0.240	0.197	0.311	0.177	0.285	0.184	0.289	0.199	0.315
	336	0.164	**0.258**	**0.160**	**0.258**	0.170	0.267	0.163	0.259	0.213	0.328	0.193	0.304	0.198	0.300	0.212	0.329
	720	0.196	0.291	0.201	0.294	**0.190**	**0.284**	0.197	0.290	0.233	0.344	0.212	0.321	0.220	0.320	0.233	0.345
	Avg	**0.157**	**0.252**	0.160	0.256	0.161	0.257	0.159	0.253	0.207	0.321	0.187	0.295	0.193	0.295	0.208	0.323
Weather	96	**0.148**	**0.198**	0.150	0.205	0.191	0.239	0.149	**0.198**	0.238	0.314	0.161	0.229	0.172	0.220	0.197	0.281
	192	**0.190**	**0.239**	0.193	0.241	0.219	0.263	0.194	0.241	0.275	0.329	0.220	0.281	0.219	0.261	0.237	0.312
	336	**0.244**	**0.281**	0.250	**0.281**	0.284	0.311	0.245	0.282	0.339	0.377	0.278	0.331	0.280	0.306	0.298	0.353
	720	**0.311**	**0.326**	0.326	0.333	0.389	0.375	0.314	0.334	0.389	0.409	0.311	0.356	0.365	0.359	0.352	0.388
	Avg	**0.223**	**0.261**	0.230	0.265	0.271	0.297	0.226	0.264	0.310	0.357	0.243	0.299	0.259	0.287	0.271	0.334
Solar	96	**0.170**	**0.230**	0.179	0.237	0.179	0.248	0.224	0.278	0.201	0.304	0.188	0.252	0.219	0.314	0.258	0.371
	192	**0.184**	**0.247**	**0.184**	0.250	0.197	0.266	0.253	0.298	0.237	0.337	0.215	0.280	0.231	0.322	0.608	0.606
	336	**0.192**	**0.249**	0.193	0.257	0.202	0.263	0.273	0.306	0.254	0.362	0.222	0.267	0.246	0.337	0.758	0.705
	720	**0.199**	**0.250**	**0.199**	0.255	0.209	0.271	0.272	0.308	0.280	0.397	0.226	0.264	0.280	0.363	0.789	0.779
	Avg	**0.186**	**0.244**	0.189	0.250	0.197	0.262	0.256	0.298	0.243	0.350	0.213	0.266	0.244	0.334	0.603	0.615
Exchange	96	**0.084**	**0.203**	0.090	0.212	0.108	0.239	0.086	0.208	0.136	0.276	0.102	0.235	0.107	0.234	0.085	0.204
	192	0.177	**0.299**	0.181	0.302	0.253	0.376	0.195	0.316	0.256	0.369	**0.172**	0.316	0.226	0.344	0.182	0.303
	336	0.355	0.430	0.363	0.440	0.390	0.471	0.342	0.425	0.426	0.464	**0.272**	**0.407**	0.367	0.448	0.348	0.428
	720	0.894	0.710	0.914	0.721	1.080	0.789	0.998	0.756	1.090	0.800	**0.714**	**0.658**	0.964	0.746	1.025	0.774
	Avg	0.378	0.410	0.387	0.419	0.458	0.469	0.405	0.426	0.477	0.477	**0.315**	**0.404**	0.416	0.443	0.410	0.427

Table 6: Full results of ablation experiments.

Datasets	Methods	Ours		w/o QFE		w/o GAWT		CNN(1 Layer)		CNN(2 Layer)		CNN(3 Layer)		Attn→Add	
	Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	0.291	0.344	0.300	0.348	0.310	0.352	0.301	0.352	0.309	0.359	0.314	0.367	0.292	0.339
	192	0.329	0.366	0.335	0.366	0.337	0.365	0.344	0.380	0.353	0.387	0.348	0.382	0.335	0.365
	336	0.364	0.389	0.375	0.389	0.376	0.390	0.383	0.403	0.382	0.404	0.382	0.402	0.383	0.396
	720	0.426	0.421	0.427	0.416	0.430	0.419	0.446	0.438	0.440	0.436	0.471	0.454	0.427	0.415
	Avg	0.352	0.380	0.359	0.380	0.363	0.382	0.369	0.393	0.371	0.397	0.379	0.401	0.359	0.379
ECL	96	0.128	0.222	0.135	0.229	0.141	0.236	0.149	0.255	0.150	0.255	0.147	0.247	0.132	0.227
	192	0.140	0.237	0.141	0.239	0.146	0.244	0.173	0.282	0.168	0.275	0.166	0.268	0.141	0.239
	336	0.164	0.258	0.165	0.259	0.170	0.265	0.204	0.309	0.167	0.262	0.178	0.270	0.165	0.259
	720	0.196	0.291	0.202	0.292	0.207	0.296	0.225	0.325	0.205	0.304	0.201	0.300	0.204	0.297
	Avg	0.157	0.252	0.161	0.255	0.166	0.260	0.188	0.293	0.173	0.274	0.173	0.271	0.161	0.255
Solar	96	0.170	0.230	0.221	0.294	0.220	0.289	0.209	0.263	0.185	0.273	0.218	0.266	0.220	0.296
	192	0.184	0.247	0.251	0.320	0.251	0.319	0.194	0.283	0.198	0.290	0.227	0.310	0.252	0.324
	336	0.192	0.249	0.267	0.326	0.267	0.324	0.217	0.299	0.218	0.302	0.217	0.302	0.274	0.340
	720	0.199	0.250	0.274	0.338	0.270	0.327	0.207	0.277	0.203	0.278	0.208	0.280	0.274	0.335
	Avg	0.186	0.244	0.253	0.320	0.252	0.315	0.207	0.281	0.201	0.286	0.217	0.289	0.255	0.324

We added experiments to evaluate the original additive bias design (Equation 8) against two strategies: multiplicative gating using the quantum outputs as gates and direct replacement of the wavelet weights with the quantum outputs. Our approach (see Table 9) delivers the most stable overall performance, indicating it is a more suitable fusion strategy and not an arbitrary heuristic.

F Prediction Visualization

In this experiment, we conducted a visualization analysis of the model's prediction results in Figure 5a to intuitively demonstrate its performance in time series forecasting tasks. By comparing the predicted values with the actual observations, it can be clearly observed that the predicted curves closely align with the real curves in both overall trends and local fluctuations, indicating the model's ability to

Table 7: Ablation study on the effectiveness of DE/ER guidance. Zeroed Quantum Guidance removes the DE/ER guidance signals while preserving the remaining model architecture. The best results are highlighted in bold.

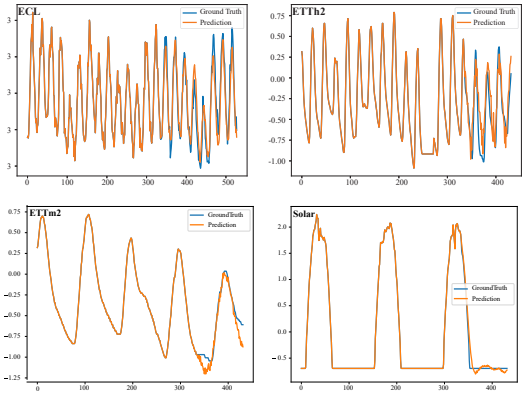
Datasets	**Methods** / **Metrics**	**Ours**		**Zeroed Quantum Guidance**	
		MSE	**MAE**	**MSE**	**MAE**
ETTh1	96	0.367	0.397	0.376	0.405
	192	0.412	0.423	0.423	0.434
	336	0.400	0.424	0.405	0.429
	720	0.447	0.467	0.469	0.477
	Avg	0.406	0.428	0.418	0.436
ETTh2	96	0.273	0.336	0.274	0.338
	192	0.324	0.376	0.329	0.378
	336	0.324	0.383	0.327	0.385
	720	0.398	0.430	0.403	0.436
	Avg	0.330	0.381	0.333	0.384
ETTm2	96	0.169	0.258	0.169	0.260
	192	0.223	0.297	0.228	0.302
	336	0.276	0.331	0.283	0.335
	720	0.364	0.384	0.368	0.387
	Avg	0.258	0.318	0.262	0.321

Table 8: Comparison with CNN-based replacement baselines of different depths. The original quantum module is replaced by 1-layer, 2-layer, and 3-layer CNNs. In particular, the 3-layer CNN results are obtained after careful tuning, with hidden size searched over {32,64,128} and kernel size over {3,5,7}. Results show that our method consistently achieves better overall forecasting performance.

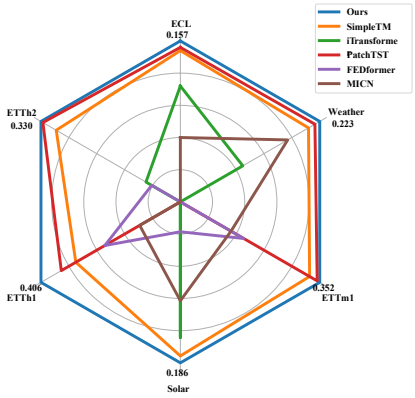
Datasets	Metrics	Ours		CNN(3 Layer)		CNN(2 Layer)		CNN(1 Layer)	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.367**	**0.397**	0.394	0.421	0.392	0.421	0.382	0.412
	192	**0.412**	**0.423**	0.418	0.433	0.433	0.443	0.434	0.442
	336	**0.400**	**0.424**	0.410	0.435	0.410	0.436	0.409	0.432
	720	**0.447**	**0.467**	0.501	0.497	0.505	0.501	0.477	0.483
	avg	**0.406**	**0.428**	0.431	0.447	0.435	0.450	0.425	0.442
ETTh2	96	**0.273**	**0.336**	0.289	0.350	0.300	0.368	0.296	0.363
	192	**0.324**	**0.376**	0.352	0.396	0.341	0.398	0.329	0.379
	336	**0.324**	**0.383**	0.335	0.393	0.330	0.395	0.327	0.387
	720	0.398	**0.430**	0.407	0.438	0.406	0.440	**0.397**	0.432
	avg	**0.330**	**0.381**	0.346	0.394	0.344	0.400	0.337	0.390
ETTm1	96	**0.291**	**0.344**	0.302	0.357	0.305	0.360	0.303	0.354
	192	**0.329**	**0.366**	0.347	0.381	0.352	0.391	0.343	0.379
	336	**0.364**	**0.389**	0.384	0.406	0.387	0.407	0.388	0.408
	720	**0.426**	**0.421**	0.439	0.434	0.442	0.438	0.445	0.440
	avg	**0.352**	**0.380**	0.368	0.394	0.371	0.399	0.370	0.395
ETTm2	96	**0.169**	**0.258**	0.179	0.271	0.179	0.270	0.178	0.266
	192	**0.223**	**0.297**	0.239	0.308	0.230	0.305	0.230	0.301
	336	**0.276**	**0.331**	0.290	0.342	0.289	0.342	0.299	0.348
	720	**0.364**	**0.384**	0.375	0.392	0.385	0.395	0.366	0.388
	avg	**0.258**	**0.318**	0.271	0.328	0.271	0.328	0.268	0.326
ECL	96	**0.128**	**0.222**	0.140	0.242	0.145	0.249	0.149	0.254
	192	**0.140**	**0.237**	0.153	0.257	0.166	0.274	0.180	0.287
	336	**0.164**	**0.258**	0.182	0.284	0.174	0.273	0.201	0.304
	720	**0.196**	**0.291**	0.203	0.304	0.206	0.306	0.217	0.319
	avg	**0.157**	**0.252**	0.170	0.272	0.173	0.275	0.187	0.291
Weather	96	**0.148**	**0.198**	0.152	0.205	0.151	0.205	0.153	0.206
	192	**0.190**	**0.239**	0.198	0.250	0.197	0.251	0.198	0.248
	336	**0.244**	**0.281**	0.250	0.292	0.249	0.290	0.248	0.289
	720	**0.311**	**0.326**	0.329	0.349	0.326	0.344	0.335	0.345
	avg	**0.223**	**0.261**	0.232	0.274	0.231	0.272	0.233	0.272

Table 9: Comparison of different integration schemes for incorporating the quantum module into the wavelet decomposition in the GAWT module (Equation 8), including our additive-bias design, a multiplicative-gating variant that uses the quantum outputs as gates to multiplicatively modulate the original wavelet weights, and a direct replacement variant that uses quantum outputs to replace the original wavelet weights.

Datasets	Metrics	Ours(additive-bias)		Multiplicative Gating		Direct Replacement	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.367**	**0.397**	0.369	0.399	0.381	0.410
	192	**0.412**	**0.423**	0.414	0.424	0.443	0.449
	336	**0.400**	**0.424**	0.409	0.432	0.411	0.435
	720	**0.447**	**0.467**	0.466	0.474	0.472	0.477
	avg	**0.406**	**0.428**	0.415	0.432	0.427	0.443
ETTh2	96	**0.273**	**0.336**	0.274	0.337	0.278	0.340
	192	**0.324**	**0.376**	0.328	0.378	0.362	0.397
	336	**0.324**	**0.383**	0.328	0.385	0.331	0.389
	720	0.398	0.430	**0.393**	**0.428**	0.405	0.437
	avg	**0.330**	**0.381**	0.331	0.382	0.344	0.391
ETTm1	96	**0.291**	0.344	0.299	0.350	0.291	**0.343**
	192	0.329	**0.366**	0.328	0.367	**0.327**	0.367
	336	**0.364**	**0.389**	0.366	0.390	0.368	0.393
	720	**0.426**	**0.421**	0.430	0.422	0.427	0.422
	avg	**0.352**	**0.380**	0.356	0.382	0.354	0.381
ETTm2	96	**0.169**	**0.258**	0.170	0.260	0.173	0.263
	192	**0.223**	**0.297**	0.226	0.299	0.226	0.299
	336	**0.276**	**0.331**	0.278	0.334	0.281	0.338
	720	**0.364**	**0.384**	0.381	0.394	0.377	0.395
	avg	**0.258**	**0.318**	0.264	0.322	0.264	0.324
ECL	96	**0.128**	**0.222**	0.129	0.223	**0.128**	**0.222**
	192	**0.140**	**0.237**	**0.140**	**0.237**	**0.140**	**0.237**
	336	**0.164**	**0.258**	0.166	**0.258**	0.166	0.259
	720	0.196	0.291	0.198	0.291	**0.195**	**0.290**
	avg	**0.157**	**0.252**	0.158	0.252	0.157	0.252
Weather	96	**0.148**	**0.198**	0.149	**0.198**	**0.148**	**0.198**
	192	**0.190**	**0.239**	**0.190**	**0.239**	**0.190**	**0.239**
	336	0.244	0.281	0.244	**0.280**	**0.241**	**0.280**
	720	**0.311**	**0.326**	0.315	0.332	0.315	0.331
	avg	**0.223**	**0.261**	0.224	0.262	0.224	0.262



(a) Visualization of prediction results.



(b) Radar chart of MSE comparison across models.

Figure 5: Prediction visualizations and MSE radar comparison.

accurately capture the dynamic characteristics of the data. Notably, even at abrupt change points and within non-stationary intervals, the predicted curves remain closely aligned with the true values, demonstrating the model's strong generalization and adaptability in handling complex nonlinear and rapidly changing temporal patterns. This visualization not only provides direct evidence of the effectiveness of our method in modeling complex time series data but also highlights the synergistic advantages of the model in extracting and integrating different informational components.

G Performance Visualization

To provide a more intuitive comparison of model performance, a radar chart is used to visualize the MSE values across different datasets. As shown in Figure 5b, QWaveNet achieves the most balanced and superior performance overall, with larger coverage across the radar axes. This indicates that our model consistently outperforms baseline methods in terms of forecasting accuracy, stability, and generalization capability, demonstrating its robustness and effectiveness in capturing complex temporal dynamics.

H Analysis of the Relationship Between Data Properties and Model Gains

To further clarify under which data properties QWaveNet can achieve more pronounced performance gains, we conduct supplementary analyses on both real-world and synthetic datasets. QWaveNet is designed to alleviate the misallocation and entanglement between long-term structures and local variations in traditional decomposition methods, and to establish a clearer division of information through the DE/ER modules and the wavelet-guided mechanism. Therefore, we further analyze the relationship between model gains and different data regimes from the perspective of data properties, thereby providing stronger empirical support for the experimental conclusions.

For the real-world datasets used in the main text, we characterize their data properties from four dimensions. First, long-term structure strength measures the dominance of slowly varying components in a sequence using the energy ratio of the low-frequency band. Second, periodic structure strength measures the prominence of repetitive patterns using the energy ratio of the dominant spectral peak. Third, cross-scale mixing strength measures whether sequence energy is distributed across multiple temporal scales using normalized wavelet energy entropy. Fourth, channel heterogeneity characterizes the dispersion of the above metrics across different channels. Among these properties, we focus on cross-scale mixing strength. This is because when components at different temporal scales are simultaneously prominent, the decomposition boundaries in the sequence are usually more ambiguous, and long-term structures and local variations are more likely to become entangled, which is directly related to the motivation of QWaveNet. Long-term structure strength, periodic structure strength, and channel heterogeneity are used as auxiliary properties to further explain the sources of performance differences across datasets. The calculation methods of these properties are shown in the Figure 6.

The analysis results on real-world datasets in the Table 10 show that the performance gains of QWaveNet generally exhibit a relationship with cross-scale mixing strength that is consistent with the motivation of the method. When cross-scale mixing is stronger in the data, the advantage of QWaveNet is usually more pronounced; when cross-scale mixing is weaker, the model gains are relatively limited. A typical example is the Solar dataset, which has high cross-scale mixing strength and low channel heterogeneity, corresponding to the largest average improvement over all baselines in the main experiments, with average improvements of 23.21% in MSE and 21.80% in MAE. In contrast, the Exchange dataset has extremely weak cross-scale mixing strength, and its corresponding average improvement is also the smallest, with average improvements of 6.38% in MSE and 6.01% in MAE. This phenomenon suggests that cross-scale mixing strength can explain, at an overall trend level, the differences in QWaveNet's gains across different datasets. It should be noted that model gains are not independently determined by a single data property. Cross-scale mixing strength provides the most direct overall trend, while long-term structure strength, periodic structure strength, and channel heterogeneity further affect the strength of this trend. For example, when the data simultaneously contains pronounced long-term structures and local fluctuations, QWaveNet can more easily obtain a clear division of information through wavelet-guided decomposition. When the statistical properties vary greatly across channels, different channels may have inconsistent requirements for the decomposition mechanism, thereby affecting the overall gain. Therefore, the

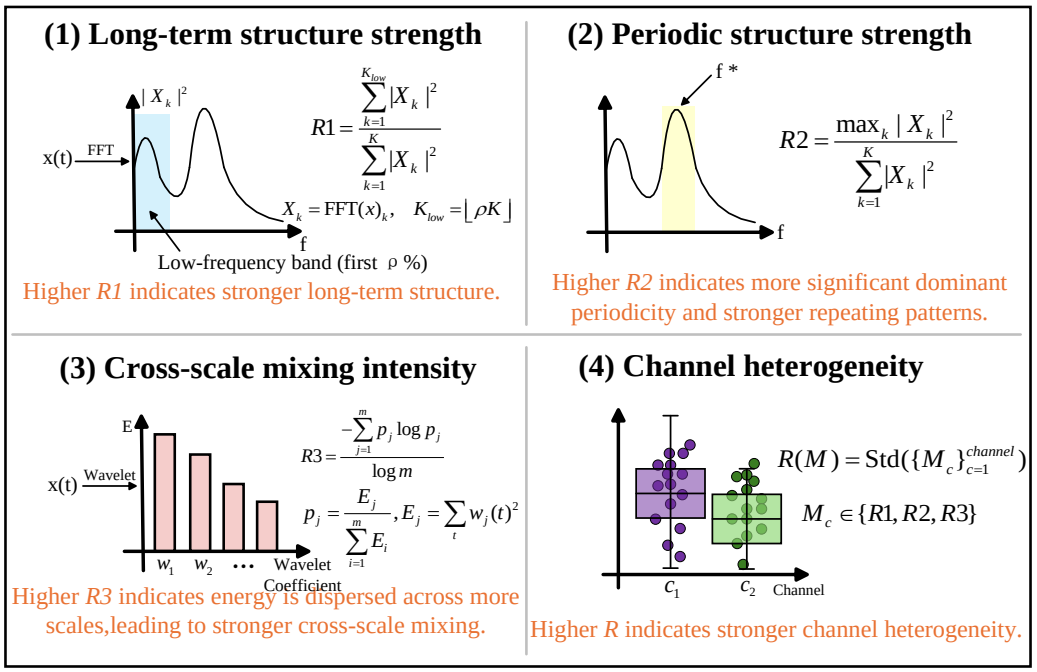


Figure 6: The figure illustrates how the four metrics are computed: long-term structure strength, periodicity structure strength, cross-scale mixing, and channel heterogeneity, measured by the low-frequency energy ratio, main spectral peak ratio, wavelet transform entropy, and inter-channel standard deviation, respectively.

Table 10: Long-term Structure Strength, Periodic Structure Strength, Cross-scale Mixing Intensity, and their corresponding Channel Heterogeneity on real-world datasets. The last two columns report the average MSE and MAE improvements over all baselines.

Datasets	Long-term Structure Strength	Channel Heterogeneity	Periodic Structure Strength	Channel Heterogeneity	Cross-scale Mixing Intensity	Channel Heterogeneity	Avg MSE Imp. over All Baselines	Avg MAE Imp. over All Baselines
ETTh2	0.922	0.045	0.377	0.246	0.268	0.131	13.68%	8.75%
ETTh1	0.626	0.293	0.270	0.150	0.570	0.236	9.12%	6.12%
ETTm1	0.936	0.035	0.270	0.150	0.424	0.189	8.70%	6.26%
ETTm2	0.974	0.020	0.376	0.246	0.177	0.087	8.11%	4.53%
ECL	0.252	0.206	0.514	0.206	0.697	0.117	13.47%	10.93%
Weather	0.848	0.256	0.342	0.246	0.388	0.328	12.65%	12.08%
Solar	0.959	0.012	0.617	0.016	**0.538**	0.014	**23.21%**	**21.80%**
Exchange	0.996	0.002	0.558	0.200	**0.034**	0.018	**6.38%**	**6.01%**

results on real-world datasets are better interpreted as a comprehensive phenomenon analysis, rather than attributing performance differences entirely to a single metric.

To more directly verify the relationship between cross-scale mixing strength and the gains of QWaveNet, we further construct synthetic datasets that control only the cross-scale mixing strength. Specifically, we generate three groups of synthetic datasets, denoted as LowMix, MediumMix, and HighMix. The detailed synthesis process is shown in the Figure 7 and the properties results of the dataset are presented in Table 11. The three groups of data share the same long-term evolution term and noise level, and the strength of multi-scale coexistence is controlled only by gradually increasing the energy ratios of components at different temporal scales, so that the cross-scale mixing strength increases sequentially. To ensure a fair comparison, since the synthetic datasets are relatively close to the ETTh2 dataset in terms of the number of channels and sequence length, QWaveNet and SimpleTM directly use their respective parameter settings on ETTh2, without additional tuning on the synthetic data.

The results on the synthetic datasets in the Table 12 further verify the above conclusion. As the cross-scale mixing strength gradually increases from LowMix to MediumMix and HighMix, the average gain of QWaveNet over SimpleTM continuously increases, reaching 0.52%, 6.54%, and 10.95%,

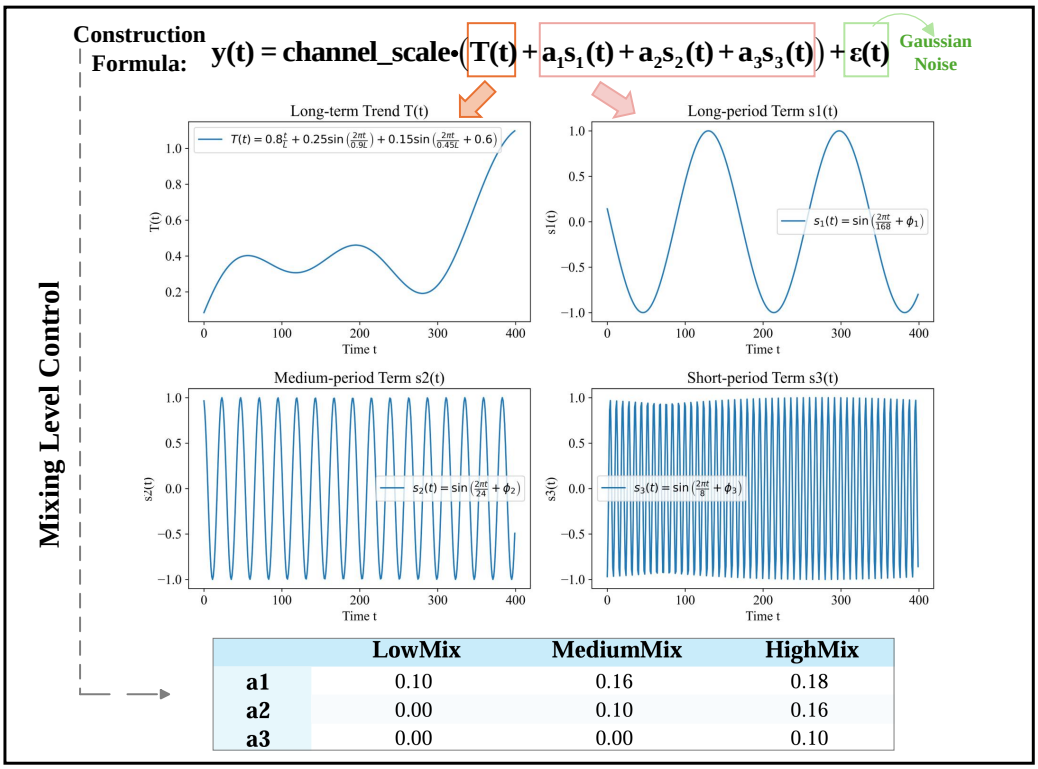


Figure 7: The figure illustrates the construction of the three synthetic datasets. Each consists of a long-term trend, periodic components, and noise. Here, ϕ represents a random phase, ensuring that all channels maintain a certain degree of heterogeneity. By adjusting cross-scale mixing intensity through the amplitudes of long-period, medium-period, and short-period components, we generate datasets with low, medium, and high mixing levels.

Table 11: Long-term Structure Strength, Periodic Structure Strength, Cross-scale Mixing Intensity, and their corresponding Channel Heterogeneity on synthetic datasets.

Datasets	Long-term Structure Strength	Channel Heterogeneity	Periodic Structure Strength	Channel Heterogeneity	Cross-scale Mixing Intensity	Channel Heterogeneity
LowMix	0.938	0.007	0.305	0.003	0.186	0.015
MediumMix	0.876	0.006	0.248	0.002	0.323	0.015
HighMix	0.750	0.005	0.201	0.002	0.497	0.015

respectively. Meanwhile, from the perspective of QWaveNet's own performance across different synthetic datasets, the average gains of MediumMix and HighMix over LowMix also continue to increase, reaching 47.27% and 66.85%, respectively. This indicates that when components at different temporal scales coexist more prominently and cross-scale mixing becomes stronger, the decomposition and wavelet-guided mechanisms of QWaveNet can play a more substantial role, leading to more significant improvements in forecasting performance. In summary, the supplementary analyses on both real-world and synthetic datasets jointly show that the advantage of QWaveNet is more pronounced under data regimes with stronger cross-scale mixing, while its gains are relatively limited when cross-scale mixing is weak. This conclusion is consistent with the design motivation of our method and avoids overclaiming its applicability.

I Computational Efficiency

In terms of computational efficiency, we report the model size (Total Params) and theoretical computational cost (GFLOPs) under a unified setting with an input length of 336 (see Table 13). Overall,

Table 12: Performance comparison between Ours and SimpleTM on synthetic datasets.

Datasets	Method	Ours		SimpleTM		Ours vs SimpleTM	
	Metrics	MSE	MAE	MSE	MAE	Avg MSE Imp.	Avg MAE Imp.
LowMix	96	0.270	0.413	0.264	0.409		
	192	0.266	0.410	0.267	0.410		
	336	0.275	0.417	0.279	0.420		
	720	0.329	0.456	0.336	0.460		
	Avg	0.285	0.424	0.286	0.425	0.52%	0.17%
MediumMix	96	0.144	0.301	0.144	0.301		
	192	0.140	0.298	0.151	0.309		
	336	0.145	0.303	0.172	0.331		
	720	0.172	0.329	0.177	0.334		
	Avg	0.150	0.308	0.161	0.319	6.54%	3.48%
MediumMix Dataset over the LowMix Dataset	Avg Imp.	47.3%	27.5%	-	-		
HighMix	96	0.090	0.241	0.091	0.241		
	192	0.088	0.236	0.099	0.251		
	336	0.092	0.241	0.106	0.259		
	720	0.108	0.261	0.128	0.285		
	Avg	0.094	0.245	0.106	0.259	10.95%	5.38%
HighMix Dataset over the LowMix Dataset	Avg Imp.	66.8%	42.3%	-	-		

Table 13: Computational efficiency comparison.

Datasets	Methods	Total_Params	GFLOPs
Weather	**Ours**	37210	0.459
	SimpleTM	42560	1.203
	iTransformer	903008	15.853
Solar	**Ours**	126790	14.593
	SimpleTM	197024	53.942
	iTransformer	3378784	59.222
ETTh1	**Ours**	31124	0.153
	SimpleTM	20536	0.234
	iTransformer	903008	1.270

QWaveNet maintains a compact parameter size and low computational complexity across the three datasets.

On Weather, for example, QWaveNet requires only 37,210 parameters and 0.459 GFLOPs, which is substantially lower than iTransformer (903,008 parameters and 15.853 GFLOPs). It is also lower than SimpleTM (42,560 parameters and 1.203 GFLOPs). On Solar, QWaveNet achieves 14.593 GFLOPs, which is again markedly lower than SimpleTM (53.942) and iTransformer (59.222), and it also uses fewer parameters. On ETTh1, QWaveNet uses 0.153 GFLOPs, lower than SimpleTM (0.234) and far below iTransformer (1.27).

These results indicate that QWaveNet can perform multi-scale modeling with more controllable computational and parameter costs while maintaining competitive forecasting performance.

I.1 Scalability Analysis

To further investigate the computational scalability of QWaveNet, we evaluate the effects of two factors: the number of qubits and the prediction horizon. We report the total number of model parameters and the theoretical computational cost in GFLOPs, following the evaluation setting used in the preceding computational efficiency comparison.

Scalability with respect to the number of qubits. We first increase the number of qubits from 4 to 8 on ETTh1, Weather, and Solar, while keeping the remaining experimental configurations unchanged. The results are reported in Table 14.

Table 14: Scalability analysis of QWaveNet with different numbers of qubits.

Datasets	Qubits	Total Params	GFLOPs
ETTh1	4	31,124	0.153
	6	34,388	0.155
	8	36,896	0.158
Weather	4	37,210	0.459
	6	44,700	0.473
	8	49,616	0.479
Solar	4	126,790	14.593
	6	524,212	14.920
	8	639,688	14.969

As shown in Table 14, increasing the number of qubits leads to additional model parameters, while the reported GFLOPs exhibit relatively modest growth. On ETTh1 and Weather, increasing the qubit number from 4 to 8 raises the parameter count by approximately 18.5% and 33.3%, respectively, whereas the GFLOPs increase by only 3.3% and 4.4%. On Solar, the increase in parameter count is more pronounced, while the GFLOPs increase by approximately 2.6%.

These results indicate that, under the evaluated configurations, increasing the number of qubits has a relatively limited impact on the reported theoretical computational cost. However, the parameter growth varies across datasets, highlighting the need to balance quantum representation capacity and model complexity.

Scalability with respect to the prediction horizon. We further investigate the influence of prediction horizon on computational cost. Specifically, we fix the number of qubits at 4 and evaluate QWaveNet on Solar under different prediction horizons, while keeping the remaining configurations unchanged. The results are presented in Table 15.

Table 15: Scalability analysis of QWaveNet with different prediction horizons on Solar.

Datasets	Horizon	Total Params	GFLOPs
Solar	96	126,790	14.593
	192	414,976	15.737
	336	424,336	17.030
	720	449,296	20.478

As shown in Table 15, both the parameter count and GFLOPs increase with the prediction horizon. When the horizon increases from 96 to 720, the GFLOPs rise from 14.593 to 20.478, corresponding to an increase of approximately 40.3%, while the parameter count increases from 126,790 to 449,296. The growth in GFLOPs is substantially smaller than the proportional increase in the prediction horizon, indicating that the reported theoretical computational cost grows relatively moderately over the evaluated range.

Overall, these scalability experiments complement the computational efficiency comparison by illustrating how model size and theoretical computational cost change under different quantum embedding dimensions and forecasting horizons. The results suggest that QWaveNet can maintain relatively moderate growth in reported GFLOPs under the examined configurations, although increasing the number of qubits or extending the prediction horizon introduces additional model complexity. The reported GFLOPs do not, by themselves, establish equivalent scalability in wall-clock execution time or on physical quantum hardware.

Table 16: The full results of QWaveNet at different numbers of qubits.

	Dataset	ETTh1		ECL		Weather	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE
2 qubits	**96**	0.371	0.400	0.128	0.222	0.150	0.201
	192	0.412	0.423	0.140	0.237	0.190	0.239
	336	0.399	0.423	0.164	0.258	0.244	0.281
	720	0.467	0.477	0.200	0.293	0.311	0.326
	Avg	0.412	0.431	0.158	0.252	0.224	0.262
4 qubits	**96**	0.367	0.397	0.128	0.222	0.148	0.198
	192	0.418	0.429	0.140	0.237	0.190	0.239
	336	0.400	0.424	0.164	0.258	0.244	0.281
	720	0.447	0.467	0.196	0.291	0.314	0.332
	Avg	0.408	0.429	0.157	0.252	0.224	0.263
6 qubits	**96**	0.389	0.417	0.128	0.222	0.150	0.199
	192	0.412	0.423	0.140	0.237	0.191	0.240
	336	0.408	0.432	0.165	0.259	0.243	0.280
	720	0.463	0.475	0.195	0.291	0.313	0.331
	Avg	0.418	0.437	0.157	0.252	0.224	0.263

J Parameter Sensitivity

Table 16 presents the complete experimental results on the ETTh1, ECL, and Weather datasets using 2, 4, and 6 qubits under four forecasting horizons (96, 192, 336, and 720). Overall, the model maintains relatively stable forecasting performance across different qubit configurations, with slight dataset-dependent variations. This suggests that the proposed quantum embedding mechanism can provide effective feature representations under different levels of quantum capacity.

For ETTh1, which has a relatively small number of channels, increasing the number of qubits from 2 to 4 improves the average forecasting performance. In particular, the improvement is more evident at the longest forecasting horizon, suggesting that a moderate increase in quantum embedding capacity may help capture cross-variable dependencies and long-term temporal patterns. However, further increasing the number of qubits to 6 leads to a slight performance degradation. This may be attributed to redundant representations introduced by additional quantum capacity, which does not necessarily contribute useful information to the forecasting task.

In contrast, ECL and Weather have higher channel dimensionality and more complex multivariate dependencies. Despite their higher-dimensional inputs, the model maintains consistently stable performance across different qubit configurations, with only marginal variations across forecasting horizons. This suggests that effective feature representations can be maintained with a limited number of qubits, potentially benefiting from the compact feature mapping applied before quantum encoding. By compressing the original inputs into a more compact representation, this mechanism may reduce the need for a proportional increase in quantum embedding dimensions while preserving information relevant to forecasting.

In summary, the results indicate that the influence of qubit number varies with dataset characteristics and is not directly determined by the original channel dimensionality. A moderate quantum embedding capacity can improve forecasting performance in certain settings, while a compact configuration may already provide sufficient representation capability in others. These findings highlight the importance of balancing quantum representation capacity and model complexity to achieve stable and effective time series forecasting.

K Robustness

To evaluate the stability of the decomposed representations under noise perturbations and their interpretable functional separation, we inject Gaussian noise into the input series at different ratios (5%, 10%, 30%, and 50%). We then compute the Pearson correlation coefficient and the relative L_2 error between the perturbed and clean Dynamic Evolution (DE) and Evolutionary Residual (ER) (see Table 17). The results show that DE maintains an extremely high correlation with the clean series

Table 17: Pearson correlation and relative L_2 error for DE and ER at different noise levels.

Noise	DE_correlation	ER_correlation	DE_L2	ER_L2
5%	1.0000	0.9991	0.0026	0.0428
10%	1.0000	0.9963	0.0049	0.0861
30%	0.9999	0.9682	0.0134	0.2578
50%	0.9997	0.9183	0.0245	0.4303

(close to 1) across all noise levels, and its relative L_2 error increases only slowly as the noise intensity grows. This indicates that DE tends to preserve stable low-frequency structures and is insensitive to random perturbations.

In contrast, the correlation of ER gradually decreases as noise increases, and its relative L_2 error rises more noticeably, exhibiting a smooth degradation trend. This observation is consistent with the role of ER in mainly carrying high-frequency variations and local disturbances. Since noise affects high-frequency components first, ER is more sensitive to noise, yet its degradation remains controllable. Overall, this experiment verifies that DE and ER exhibit clear and consistent scale-specific roles under noise perturbations, and it also demonstrates the robustness and interpretability of the learned representations.

L Details of Frequency-Domain Analysis for Feature Decomposition

We conduct a frequency-separation analysis to quantify the scale-specific roles of the two learned representations. Specifically, under the same data preprocessing, input length, forecasting horizons, and evaluation protocol as the main experiments, we extract DE and ER from the model outputs. We then compute the real fast Fourier transform (FFT) of these sequences and convert it to power spectra. Using a fixed cutoff ratio (25%), we split the frequency axis into low- and high-frequency bands, and report the low-frequency energy ratio of DE and the high-frequency energy ratio of ER. These ratios are aggregated to obtain the final statistics reported in Table 3, and we repeat the same procedure for other decomposition baselines to ensure a fair comparison.

M Mean and Standard Deviation

Table 18 reports the mean and standard deviation of MSE and MAE on real-world datasets over multiple runs. Overall, the results show that the proposed method achieves stable performance across different datasets and prediction horizons. The standard deviations are consistently small compared with the corresponding mean values, indicating that the model is insensitive to random initialization and exhibits good robustness. In particular, on ECL, Weather, Solar, ETTm1, ETTm2, and ETTh2, the standard deviations are mostly close to 0.000–0.003, demonstrating highly stable forecasting behavior. Although ETTh1 and Exchange present slightly larger deviations at some long prediction horizons, the variance remains limited relative to their mean errors, suggesting that the performance fluctuations are still controlled.

N Comparison with Existing Quantum Time-Series Forecasting Models

To further evaluate the effectiveness of the proposed quantum feature extraction mechanism, we conduct additional comparisons with existing hybrid quantum-classical models for time series forecasting. Specifically, we consider two representative approaches: Linear-Layer-Enhanced Quantum Long Short-Term Memory (L-QLSTM) and QuantumTime. The former is evaluated through a controlled module replacement within QWaveNet, whereas the latter provides a comparison with an existing quantum time-series forecasting framework.

Table 18: Mean and standard deviation of MSE and MAE on real-world datasets.

Datasets	Method	Mean		SD	
	Metrics	MSE	MAE	MSE	MAE
ETTm1	96	0.291	0.344	0.001	0.001
	192	0.329	0.367	0.001	0.001
	336	0.364	0.388	0.000	0.001
	720	0.427	0.421	0.002	0.001
	Avg	0.353	0.380	0.001	0.001
ETTm2	96	0.169	0.258	0.001	0.001
	192	0.224	0.299	0.002	0.002
	336	0.280	0.333	0.003	0.002
	720	0.365	0.385	0.001	0.001
	Avg	0.260	0.319	0.001	0.001
ETTh1	96	0.369	0.399	0.002	0.002
	192	0.422	0.432	0.011	0.009
	336	0.400	0.424	0.004	0.004
	720	0.462	0.474	0.013	0.006
	Avg	0.413	0.432	0.006	0.004
ETTh2	96	0.272	0.336	0.001	0.000
	192	0.327	0.377	0.003	0.001
	336	0.326	0.385	0.003	0.003
	720	0.399	0.431	0.002	0.001
	Avg	0.331	0.382	0.002	0.001
ECL	96	0.128	0.222	0.000	0.000
	192	0.140	0.237	0.000	0.000
	336	0.164	0.258	0.000	0.000
	720	0.196	0.291	0.001	0.000
	Avg	0.157	0.252	0.000	0.000
Weather	96	0.149	0.199	0.001	0.002
	192	0.189	0.239	0.001	0.000
	336	0.242	0.280	0.002	0.001
	720	0.314	0.330	0.003	0.003
	Avg	0.224	0.262	0.001	0.001
Solar	96	0.172	0.233	0.002	0.004
	192	0.184	0.247	0.000	0.001
	336	0.192	0.251	0.001	0.002
	720	0.198	0.250	0.001	0.001
	Avg	0.187	0.245	0.000	0.002
Exchange	96	0.086	0.205	0.002	0.001
	192	0.181	0.301	0.004	0.002
	336	0.347	0.426	0.012	0.006
	720	0.903	0.714	0.009	0.005
	Avg	0.379	0.411	0.004	0.003

N.1 Comparison with L-QLSTM

We first compare the proposed QCNN-based feature extraction module with L-QLSTM, a hybrid quantum-classical recurrent architecture. The implementation of L-QLSTM is adapted from the authors' official Quantum-LSTM repository.

For a controlled comparison, we replace only the QCNN-based feature extractor in QWaveNet with the L-QLSTM module, while retaining the remaining components of the forecasting pipeline. Specifically, L-QLSTM processes the embedded feature sequence recurrently. At each time step, the current input and the previous hidden state are concatenated and projected into a four-qubit representation through a shared linear embedding layer. Four independent variational quantum circuits are then used to model the forget, input, candidate, and output gates, followed by separate linear output projections. An additional output adapter maps the resulting hidden representations to the dimensionality of the original QCNN outputs, allowing the subsequent modules to remain unchanged.

Table 19: Forecasting performance comparison between QWaveNet and the L-QLSTM replacement on ETTh1 and ETTh2. Lower MSE and MAE indicate better performance.

Datasets	Horizon	QWaveNet		L-QLSTM	
		MSE	MAE	MSE	MAE
ETTh1	96	0.367	0.397	0.380	0.411
	192	0.412	0.423	0.419	0.430
	336	0.400	0.424	0.412	0.435
	720	0.447	0.467	0.476	0.482
	Avg	0.406	0.428	0.422	0.439
ETTh2	96	0.273	0.336	0.276	0.338
	192	0.324	0.376	0.332	0.380
	336	0.324	0.383	0.331	0.386
	720	0.398	0.430	0.407	0.436
	Avg	0.330	0.381	0.336	0.385

Table 20: Forecasting performance comparison between QWaveNet and QuantumTime on ECL, Weather, and ETTm2. Lower MSE and MAE indicate better performance.

Datasets	Horizon	QWaveNet		QuantumTime	
		MSE	MAE	MSE	MAE
ECL	96	0.128	0.222	0.134	0.231
	192	0.140	0.237	0.148	0.243
	336	0.164	0.258	0.166	0.262
	720	0.196	0.291	0.198	0.297
	Avg	0.157	0.252	0.162	0.258
Weather	96	0.148	0.198	0.166	0.223
	192	0.190	0.239	0.209	0.263
	336	0.244	0.281	0.252	0.300
	720	0.311	0.326	0.299	0.338
	Avg	0.223	0.261	0.232	0.281
ETTm2	96	0.169	0.258	0.165	0.256
	192	0.223	0.297	0.223	0.301
	336	0.276	0.331	0.275	0.338
	720	0.364	0.384	0.425	0.401
	Avg	0.258	0.318	0.272	0.324

As shown in Table 19, QWaveNet consistently outperforms the L-QLSTM Cao et al. [2023] replacement in terms of both MSE and MAE across all evaluated forecasting horizons on ETTh1 and ETTh2. The performance differences are particularly evident for longer forecasting horizons, suggesting that the proposed QCNN-based feature extractor is more effective in capturing temporal representations within the current forecasting framework.

Furthermore, L-QLSTM requires four variational quantum circuits at each recurrent time step, resulting in substantially higher computational overhead. In our implementation, its training time on ETTh2 with a prediction horizon of 720 is approximately 50 times that of QWaveNet. Due to this computational cost, we conduct the L-QLSTM comparison on ETTh1 and ETTh2.

These results support the effectiveness of the proposed QCNN-based feature extraction design, which achieves better forecasting performance with lower training overhead than the recurrent quantum alternative under the evaluated settings.

N.2 Comparison with QuantumTime

We further compare QWaveNet with QuantumTime Qiao et al. [2025], a representative hybrid quantum-classical approach for time series forecasting. This comparison complements the experiments against classical forecasting baselines and provides additional evidence of the competitiveness of the proposed framework among quantum-based methods.

As shown in Table 20, QWaveNet achieves lower average MSE and MAE than QuantumTime on all three evaluated datasets. The improvements are particularly consistent on ECL, where QWaveNet achieves lower forecasting errors across all prediction horizons. On Weather, QWaveNet also provides better overall performance, although QuantumTime obtains a lower MSE at the longest prediction horizon. On ETTm2, the two methods exhibit comparable performance at shorter horizons, while QWaveNet achieves a more noticeable advantage for long-horizon forecasting.

It should be noted that the two methods adopt different historical input configurations. QuantumTime adjusts the historical input length according to the prediction horizon, potentially using longer historical contexts for long-term forecasting. In contrast, QWaveNet selects its input length from a fixed set of candidate values, rather than systematically increasing the historical window with the prediction horizon. Therefore, this comparison reflects the respective forecasting settings and should not be interpreted as a strictly controlled comparison of individual architectural components.

Nevertheless, the results demonstrate that QWaveNet achieves competitive forecasting performance against an existing quantum-based forecasting approach. Together with the L-QLSTM replacement experiments, these findings further support the effectiveness of the proposed QCNN-based representation learning mechanism and its integration with wavelet-based multi-scale modeling.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims in the abstract and introduction are consistent with the scope of the paper. The proposed contributions are supported by the experimental results and analyses reported in the Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses the limitations and applicability of the proposed method through an analysis of the relationship between data properties and model gains on both real-world and synthetic datasets. This analysis helps clarify under what data conditions the proposed method is expected to be more or less effective.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper provides a theoretical proof in Appendix A. The proof illustrates how R_y encoding and CNOT entanglement enable the quantum circuit to produce cross-variable coupled responses, thereby supporting richer long-term structural representations in DE.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper provides the experimental details required to reproduce the main results, with additional information included in the Appendix E. The code URL is also provided to facilitate reproduction.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The experiments are based on publicly available datasets, and the implementation code is made accessible through the URL provided in the paper. The Appendix E further describes the experimental setup needed to reproduce the reported results.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper reports the relevant experimental settings in the Appendix E, including the configurations used for training and evaluation. These details allow readers to understand how the reported results were obtained.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper reports standard deviations in the Appendix M to characterize the variability of the experimental results. These statistics are provided for the relevant experiments supporting the main empirical conclusions.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper reports the computational resources used for the experiments, including the hardware environment in the Appendix E and computational efficiency in the Appendix I. These details provide sufficient guidance for reproducing the main experimental results and understanding the associated computational cost.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This work strictly conforms, in every respect, with the NeurIPS Code of Ethics. It uses standard public datasets and does not involve unethical data collection, human-subject risks, or unsafe deployment.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This is a technical study on time series forecasting and does not directly target sensitive applications or deployment scenarios. Therefore, a detailed broader societal impact discussion is not applicable.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The datasets and models used in this work are standard for time series forecasting and do not pose known high-risk safety or misuse concerns. Therefore, special safeguards for responsible release are not required.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper uses publicly available datasets and open-source code, and the corresponding sources are properly cited in the main text. The terms of use and licenses of these assets are respected.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The paper releases implementation code to support reproducibility of the proposed method. The code is accompanied by documentation and experimental settings that help users reproduce the main results.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This work does not involve crowdsourcing or research with human subjects. As a result, participant instructions, screenshots, and compensation details are not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing or human-subject research. Therefore, there are no participant risks to disclose and no IRB or equivalent approval is required.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not used as an important, original, or non-standard component of the core methodology in this research. Therefore, declaration of LLM usage for the proposed method is not applicable.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.