

000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027 028 029 030 031 WAVELETDIFF: MULTILEVEL WAVELET DIFFUSION FOR TIME SERIES GENERATION

005 **Anonymous authors**

006 Paper under double-blind review

009 ABSTRACT

011 Time series are ubiquitous in many applications that involve forecasting, classi-
012 fication and causal inference tasks, such as healthcare, finance, audio signal pro-
013 cessing and climate sciences. Still, large, high-quality time series datasets remain
014 scarce. Synthetic generation can address this limitation; however, current models
015 confined either to the time or frequency domains struggle to reproduce the inher-
016 ently multi-scaled structure of real-world time series. We introduce WaveletDiff,
017 a novel framework that trains diffusion models *directly on wavelet coefficients*
018 to exploit the inherent multi-resolution structure of time series data. The model
019 combines dedicated transformers for each decomposition level with cross-level at-
020 tention mechanisms that enable selective information exchange between temporal
021 and frequency scales through adaptive gating. It also incorporates energy preser-
022 vation constraints for individual levels based on Parseval’s theorem to preserve
023 spectral fidelity throughout the diffusion process. Comprehensive tests across
024 six real-world datasets from energy, finance, and neuroscience domains demon-
025 strate that WaveletDiff consistently outperforms state-of-the-art time-domain and
026 frequency-domain generative methods on both short and long time series across
027 five diverse performance metrics. For example, WaveletDiff achieves discrim-
028 inative scores and Context-FID scores that are 3 \times smaller on average than the
029 second-best baseline across all datasets.

030 1 INTRODUCTION

031 Time series data arises in diverse practical settings, including healthcare Lee et al. (2019a); van der
032 Schaar Lab (2019), finance Sezer et al. (2020); Özbayoglu et al. (2020), climate sciences Dinku
033 (2019); Climate Change AI (2021), audio processing Mitra & Zualkernan (2025) and engineer-
034 ing Susto et al. (2020); Lei et al. (2020); Carvalho et al. (2022). Due to various constraints, acquiring
035 sufficiently high-quality labeled time-series datasets remains a challenge Wang et al. (2021); Desai
036 et al. (2025). The problem may be mitigated through synthetic time series generation, which also
037 offers promising solutions for data augmentation Wen et al. (2021); Le Guennec et al. (2018); Ryu
038 et al. (2023), privacy preservation Wang et al. (2020); Jordon et al. (2024); Nosowsky & Giordano
039 (2006), forecasting Taga et al. (2025) and simulations Nikolenko (2021); El Emam et al. (2022).

040 Current time series generation methods predominantly operate either directly in the time domain or
041 frequency domain, and come with different advantages and limitations. Time-domain approaches,
042 including those based on GANs Yoon et al. (2019); Pei et al. (2021); EskandariNasab et al. (2024),
043 autoregressive Salinas et al. (2020) and diffusion models Lim et al. (2023); Narasimhan et al. (2024);
044 Sikder et al. (2024) are well-suited for modeling local temporal patterns, but struggle with long-
045 term dependencies and preservation of important spectral characteristics. To address time-domain
046 induced limitations, recent approaches have increasingly leveraged frequency-domain analysis, of-
047 ten along with temporal modeling Tian et al. (2020); Chi et al. (2024); Crabbé et al. (2024); Huang
048 et al. (2024). These methods are of relevance since many real-world time series tend to exhibit
049 higher localization in the frequency rather than the time domain. Representative methods include
050 FourierFlow Alaa et al. (2021), which applies normalizing flows to Fourier representations, Diffu-
051 sionTS Yuan & Qiao (2024), which combines Fourier decompositions with diffusion models, and
052 various frequency-enhanced transformers Zhou et al. (2022a;b); Xu et al. (2024); Yi et al. (2023)
053 that use both spectral and temporal analyses. However, these approaches typically process time and
frequency domain information either in a separate manner or impose trade-offs between temporal

resolution and spectral coherence. They are also not able to simultaneously capture both local and global time and spectral patterns, which is crucial for synthesizing realistic time series.

Wavelet transforms represent a natural approach to address the above issue by creating a multi-resolution representation that simultaneously captures both temporal and spectral information Mallat (1989); Cohen (2001). Unlike the Fourier transform, which capture global frequency properties, wavelets maintain temporal localization while also providing useful decompositions into multiple frequency bands Rioul & Flandrin (1992); Daubechies (1988). This results in a highly versatile time-frequency hierarchical representation Mallat (1989). As a result, wavelet-based analyses have been used with success for various signal processing applications including speech recognition, financial trends analysis, image processing, and biomedical signal analysis Daubechies (1992); Vetterli & Kovačević (1995); Burrus et al. (1998). Despite these results, a handful of known wavelet-based approaches for time series generation have failed to provide improvements over Fourier-based methods Takahashi & Mizuno (2024); Kazemi & Meidani (2022); Zhao et al. (2018). This may be attributed to the fact that, almost exclusively, the methods treat wavelet coefficients as image structures and then follow-up by applying standard image generation techniques such as convolutional neural networks or image-based diffusion models. While potentially useful for data-poor applications and highly specialized time series, indirect time series → wavelet → image conversion methods in general suffer from pattern distortions caused by noninvertible image features.

A more adequate approach based on wavelet decompositions is to run diffusion models directly in the wavelet domain, which is a new direction proposed in this work. For diffusion, methods such as denoising diffusion probabilistic models (DDPMs) Ho et al. (2020) which have demonstrated remarkable success in image Dhariwal & Nichol (2021), audio Kong et al. (2021), and text generation Austin et al. (2021), are considered state-of-the-art for time-series generation. However, these diffusion models are tailor-made for highly specific time-series formats (e.g., audio or financial data), and may not be suitable for other modalities. This motivates us to implement a new wavelet-space diffusion model, termed WaveletDiff, which is universally applicable as it inherently respects different multi-level structures. Unlike frequency-domain approaches, WaveletDiff also captures temporal patterns at different scales simultaneously. Our key innovations lie in running forward diffusion processes for each wavelet level individually and in parallel, following fine-tuned exponential noise mechanisms and using dedicated level-transformer denoising networks combined with a cross-level attention mechanism that enables information exchange between different decomposition levels. This design preserves the hierarchical nature of wavelet representations while allowing the model to learn complex inter-scale dependencies crucial for realistic time series generation.

Our technical contributions include:

1. A diffusion framework that operates directly in the wavelet domain and tunes the noise addition process to the approximation and detail levels, identifies the most suitable choice of mother wavelet for different time series and uses level-specific loss functions and transformers.
2. A cross-scale attention mechanisms that enables information flow between different temporal scales while preserving their individual properties.
3. A wavelet-aware loss weighting mechanism that prevents some levels from dominating the training objective through level-specific balancing strategies.
4. A new evaluation metric based on the Dynamic Time-Warping distance and extensive comparative analysis of both short and long time series datasets, including ETTh1, ETTh2, Stocks, Exchange Rate, fMRI, EEG Zhou et al. (2021); Lai et al. (2018); Roesler (2013). The results show significant performance gains of WaveletDiff compared to Fourier and time-based methods, which are roughly three-fold on average for discriminative and Context-FID scores.
5. The first empirical evaluation of reproducibility of diffusion models for time series, akin to recent efforts reported for images Zhang et al. (2024b); Li et al. (2024); Kadkhodaie et al. (2024).

2 RELATED WORK

Time Series Generation with Diffusion Models. Early generative AI methods for time series focused on conditional generation tasks such as forecasting and imputation Rasul et al. (2021); Tashiro et al. (2021); Li et al. (2022); Yang et al. (2024), while recent approaches target unconditional time series generation Shen & Kwok (2023); Barancikova et al. (2025). The above methods employ various architectural choices including RNNs, transformers, and specialized denoising networks to handle the sequential nature of temporal data Kong et al. (2021), and almost exclusively operate in the time domain. This limits their ability to capture global and local spectral properties.

108 **Frequency Domain Approaches for Time Series.** Recent works have demonstrated that real-world
 109 time series are more localized in the frequency domain, making spectral diffusion more effective
 110 than time diffusion Crabbé et al. (2024). Various frequency-based approaches include lightweight
 111 models using complex-valued operations Xu et al. (2024), frequency-enhanced transformers com-
 112 bining discrete Fourier transform (DFT) with attention mechanisms Zhou et al. (2022b;a), and spe-
 113 cialized MLP architectures for frequency learning Yi et al. (2023). Additional methods incorpo-
 114 rate spectral filtering Zhang et al. (2024a), multi-resolution frequency analysis Wang et al. (2024),
 115 and normalizing flows in Fourier domain Alaa et al. (2021). Additional unconditional genera-
 116 tion approaches include interpretable diffusion models that combine *trend and seasonality* com-
 117 ponents Yuan & Qiao (2024) and latent diffusion models that operate in compressed latent spaces
 118 for more efficient generation Qian et al. (2024). These methods often require separate processing
 119 pipelines for temporal and frequency components, limiting their ability to simultaneously capture
 120 multi-scale temporal-spectral relationships.
 121

122 **Wavelet-Based Time Series Modeling.** Wavelet transforms provide multi-resolution time-
 123 frequency representation capabilities Mallat (1989); Daubechies (1992); Addison (2017) and have
 124 been extensively used in time series analysis Percival & Walden (2000); Sang (2013); Patrik et al.
 125 (2015). In generative modeling, wavelets have shown promise across various domains through di-
 126 rect coefficient processing Phung et al. (2022); Hu et al. (2023); Guth et al. (2022). For time series
 127 forecasting, wavelets have been employed to enhance traditional forecasting models through their
 128 multi-resolution time-frequency analysis capabilities Zhou et al. (2025); Sasal et al. (2022); Arabi
 129 et al. (2024); Schlüter & Deuschle (2010). For generation tasks, existing methods predominantly
 130 convert wavelet coefficients to image representations for processing with standard computer vision
 131 techniques Takahashi & Mizuno (2024); Kazemi & Meidani (2022). However, this indirect approach
 132 may not fully exploit the hierarchical multi-scale structure of wavelet decompositions, where each
 133 level captures distinct temporal and spectral characteristics.
 134

3 METHODOLOGY

3.1 WAVELET REPRESENTATIONS OF TIME SERIES

136 A multivariate time series dataset $\mathbf{X} \in \mathbb{R}^{N \times T \times D}$ with N samples, T timesteps and D features (e.g.,
 137 opening price, closing price, high/low, volume for financial data) comprises time series of the form
 138 $\mathbf{x}^{(i)} = [\mathbf{x}_0^{(i)}, \mathbf{x}_1^{(i)}, \dots, \mathbf{x}_{T-1}^{(i)}] \in \mathbb{R}^{T \times D}$, where $i \in [1, N]$. The Discrete Wavelet Transform (DWT)
 139 decomposes each time series through a cascade of high-pass and low-pass filtering operations fol-
 140 lowed by downsampling. The decomposition utilizes a scaling function $\phi(t)$ and its associated
 141 *mother wavelet* $\psi(t)$. The mother wavelet is characterized by its order p , which ensures that the
 142 wavelet is orthogonal to all polynomials of degree less than p (hence, p determines the number of
 143 vanishing moments). Higher-order wavelets provide better frequency localization but require longer
 144 filters. The filter length F represents the number of nonzero coefficients in the discrete filters, which
 145 depends on the wavelet family and order p (e.g., $F = 2p$ for Daubechies wavelets). More details are
 146 available in Appendix A. These functions satisfy the two-scale relations:
 147

$$\psi(t) = \sqrt{2} \sum_{k=0}^{F-1} g_k \phi(2t - k), \text{ where } \phi(t) \text{ satisfies } \phi(t) = \sqrt{2} \sum_{k=0}^{F-1} h_k \phi(2t - k), \quad (1)$$

148 and where $\{g_k\}_{k=0}^{F-1}$ and $\{h_k\}_{k=0}^{F-1}$ are the high-pass and low-pass filter coefficients, respectively,
 149 with the relationship $g_k = (-1)^k h_{F-1-k}$ ensuring orthogonality. The DWT performs recursive
 150 decomposition over L levels. Starting with the approximation coefficients $\mathbf{A}^{(0)} = \mathbf{X}$, at each level
 151 $l \in [1, L]$, we apply high-pass and low-pass filters followed by temporal-dimension downsampling:
 152

$$\mathbf{C}_{:,m,:}^{(l)} = \sum_k g_k \mathbf{A}_{:,2m-k,:}^{(l-1)} \text{ (detail coeff.)}, \quad \mathbf{A}_{:,m,:}^{(l)} = \sum_k h_k \mathbf{A}_{:,2m-k,:}^{(l-1)} \text{ (approximate coeff.)}, \quad (2)$$

153 where m indexes the downsampled time dimension and the operation is applied independently across
 154 all N samples and D features. Boundary effects are handled using symmetric extension, where the
 155 signal is mirrored at the tails to ensure sufficient coefficients for filtering operations. This decompo-
 156 sition yields the wavelet coefficient representation:
 157

$$\text{DWT}(\mathbf{X}) = \{\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(L)}, \mathbf{A}^{(L)}\}, \quad (3)$$

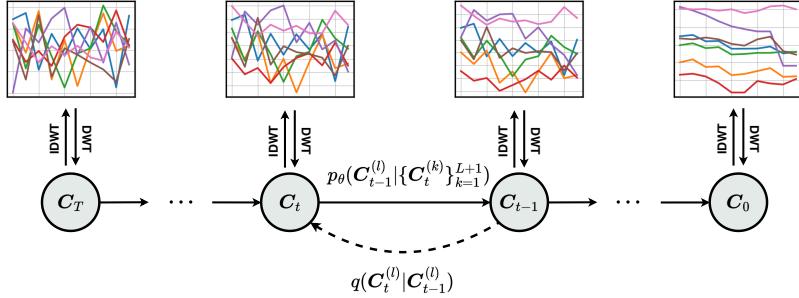


Figure 1: Direct wavelet coefficient diffusion, where the forward process proceeds independently at each decomposition level, while the reverse process integrates information across all levels to perform collective denoising.

where $\mathbf{C}^{(l)} \in \mathbb{R}^{N \times d_l \times D}$, for $l \in [1, L]$, are the detail coefficients and $\mathbf{A}^{(L)} \in \mathbb{R}^{N \times d_L \times D}$ are the approximation coefficients. For consistency with diffusion notation, we write $\mathbf{C}^{(L+1)} = \mathbf{A}^{(L)}$.

The wavelet order p is chosen based on the sequence length to ensure sufficient coefficients at each level, with longer sequences accommodating higher-order wavelets for better frequency localization. The coefficient dimension at each level l is calculated recursively as:

$$d_l = \lfloor \frac{d_{l-1} + F - 1}{2} \rfloor, \quad l = 1, \dots, L, \quad (4)$$

where $d_0 = T$ is the original sequence length. This formula accounts for the filter overlap (requiring $F - 1$ additional boundary coefficients) and dyadic downsampling (division by 2) inherent to the wavelet decomposition process. The number of decomposition levels L is determined based on the sequence length T to ensure sufficiently coefficients are available at each level while maintaining meaningful frequency separation, and is set to $L = \max \left(3, \min \left(7, \left\lfloor \log_2 \left(\frac{T}{F-1} \right) \right\rfloor \right) \right)$ in practice.

To reconstruct time series from diffusion-generated wavelet coefficients $\{\hat{\mathbf{C}}^{(1)}, \dots, \hat{\mathbf{C}}^{(L)}, \hat{\mathbf{C}}^{(L+1)}\}$, we apply the Inverse Discrete Wavelet Transform (IDWT). The reconstruction proceeds from the coarsest level to the finest level. Starting with $\hat{\mathbf{A}}^{(L)} = \hat{\mathbf{C}}^{(L+1)}$, for each $l = L, \dots, 1$, we compute:

$$\hat{\mathbf{A}}_{:,m,:}^{(l-1)} = \sum_k \tilde{h}_{m-2k} \hat{\mathbf{A}}_{:,k,:}^{(l)} + \sum_k \tilde{g}_{m-2k} \hat{\mathbf{C}}_{:,k,:}^{(l)}, \quad (5)$$

where \tilde{h} and \tilde{g} are the synthesis filters used for reconstruction. For orthogonal wavelets, these take the form $\tilde{h}_k = h_{-k}$ and $\tilde{g}_k = g_{-k}$, while for biorthogonal wavelets, they are independently designed dual filters that ensure perfect reconstruction. The reconstruction combines the current approximation $\hat{\mathbf{A}}^{(l)}$ with the detail coefficients $\hat{\mathbf{C}}^{(l)}$ through upsampling and filtering. The inverse transform can hence be written as:

$$\hat{\mathbf{X}} = \text{IDWT}(\{\hat{\mathbf{C}}^{(1)}, \dots, \hat{\mathbf{C}}^{(L)}, \hat{\mathbf{A}}^{(L)}\}), \quad (6)$$

where $\hat{\mathbf{X}} = \hat{\mathbf{A}}^{(0)} \in \mathbb{R}^{N \times T \times D}$ is the reconstructed time series.

3.2 WAVELET-SPACE DIFFUSION FRAMEWORK

We propose to run the diffusion process on wavelet coefficients using Denoising Diffusion Probabilistic Models (DDPM), as shown in Figure 1. The forward diffusion process in the wavelet domain gradually adds Gaussian noise to coefficients at all levels independently:

$$q(\mathbf{C}_t^{(l)} | \mathbf{C}_0^{(l)}) = \mathcal{N}(\mathbf{C}_t^{(l)}; \sqrt{\alpha_t} \mathbf{C}_0^{(l)}, (1 - \bar{\alpha}_t) \mathbf{I}), \quad l = 1, \dots, L + 1, \quad (7)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ follow standard DDPM schedules. Specifically, we adopt an exponential noise schedule $\beta_t = \beta_{\text{start}} + (\beta_{\text{end}} - \beta_{\text{start}}) \cdot (1 - e^{-\gamma \cdot t})$, where γ is the exponential decay rate, $t \in [0, 1]$ is the normalized timestep, and β_{start} and β_{end} are tuneable hyperparameters. Exponential schedules are better suited for wavelet-based time series generation than cosine schedules. This may be because cosine schedules start slowly, peak mid-epoch, and then decrease, with

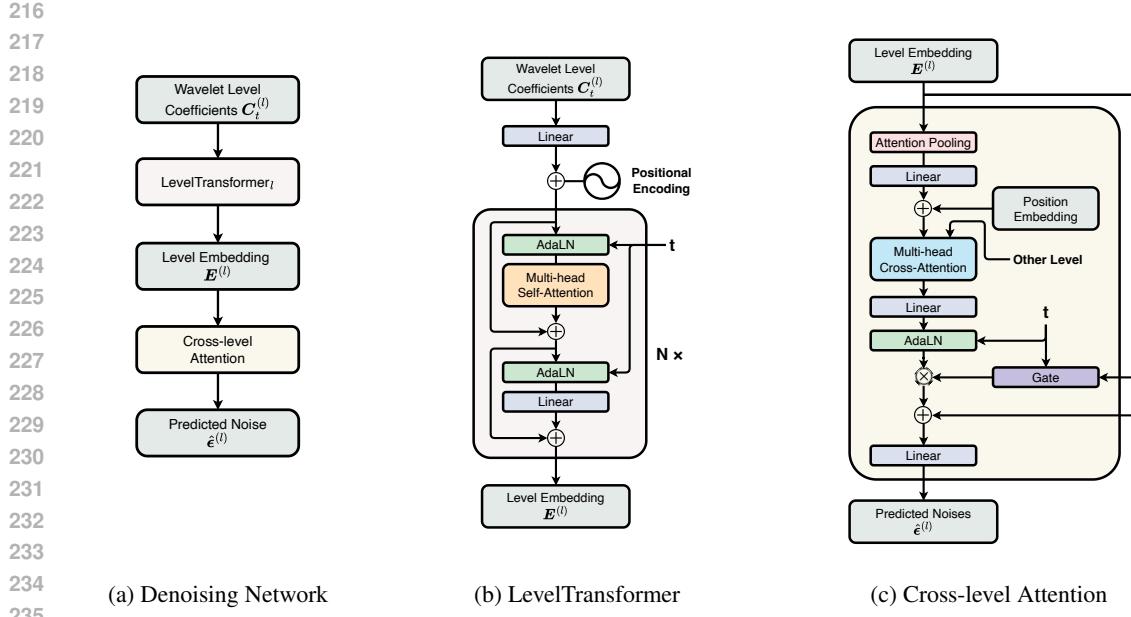


Figure 2: The wavelet coefficients are independently processed by LevelTransformers to obtain level-specific embeddings. These embeddings are obtained through interaction across levels via a cross-level attention module based on adaptive gating mechanisms.

this smooth behavior stabilizing high-dimensional data like images. In contrast, time series and their wavelet decompositions have significantly lower dimensions, benefiting from more aggressive noise injection when coupled with transformer-based denoising models used at early backwards steps.

We parameterize the reverse process using a cross-level transformer network that employs cross-attention to enable communication across levels:

$$p_\theta(C_{t-1}^{(l)} | \{C_t^{(k)}\}_{k=1}^{L+1}) = \mathcal{N}(C_{t-1}^{(l)}; \mu_\theta(\{C_t^{(k)}\}_{k=1}^{L+1}, t), \Sigma_\theta), \quad l = 1, \dots, L+1, \quad (8)$$

where $\mu_\theta(\{C_t^{(k)}\}_{k=1}^{L+1}, t)$ represents the predicted mean of the reverse diffusion process, parameterized by the neural network θ and conditioned on all wavelet levels and the diffusion timestep t , and Σ_θ is the predicted covariance matrix. Following the DDPM framework, we fix $\Sigma_\theta = \beta_t \mathbf{I}$ and train the denoising network f_θ to predict the added noise ϵ using the mean square error (MSE) as the loss:

$$\hat{\epsilon}^{(l)} = f_\theta(\{C_t^{(k)}\}_{k=1}^{L+1}, t), \quad l = 1, \dots, L+1 \quad (9)$$

$$\mathcal{L}_{\text{recon}} = \mathbb{E}_{C, t, \epsilon} \left[\sum_{l=1}^{L+1} w_l \cdot \|\epsilon^{(l)} - \hat{\epsilon}^{(l)}\|^2 \right] \quad (10)$$

where $\epsilon^{(l)}$ and $\hat{\epsilon}^{(l)}$ are the true and predicted noise at level l , and w_l are level-specific weights ensuring balanced contribution across scales.

To preserve data spectra for long sequence generation, we optionally introduce an energy conservation penalty based on Parseval’s theorem:

$$\mathcal{L}_{\text{energy}} = \mathbb{E}_{C, t, \epsilon} \left[\sum_{l=1}^{L+1} |\mathcal{E}^{(l)} - \hat{\mathcal{E}}^{(l)}| \right], \quad (11)$$

where $\mathcal{E}^{(l)} = \sum_{n=1}^N \sum_{j=1}^{d_l} \sum_{k=1}^D (C_{n,j,k}^{(l)})^2$ represents the true energy at wavelet level l , and $\hat{\mathcal{E}}^{(l)} = \sum_{n=1}^N \sum_{j=1}^{d_l} \sum_{k=1}^D (\hat{C}_{n,j,k}^{(l)})^2$ represents the predicted energy at wavelet level l . By enforcing energy preservation at each decomposition level individually, the constraints stabilize training and preserve the natural energy distribution across frequency scales. The overall training objective combines both the reconstruction and energy terms $\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{energy}} \mathcal{L}_{\text{energy}}$, where λ_{energy} denotes the weight of the energy loss term. For short sequences, the base reconstruction loss is typically sufficient since

270 the spectral energy drift is minimal over limited temporal horizons. The energy preservation term
 271 mostly benefits datasets with strong low-frequency trends and smooth spectral characteristics (e.g.,
 272 ETT1, Exchange Rate), while high-volatility datasets with abrupt changes (e.g., Stocks) are better
 273 reproduced through the reconstruction loss alone.

274 The denoising network f_θ uses dedicated transformers for each wavelet level. We adopt Adaptive
 275 Layer Normalization (AdaLN) Peebles & Xie (2022) as the normalization layer. For level l , the
 276 coefficients $C_t^{(l)}$ are processed through a specialized transformer,
 277

$$278 \quad \mathbf{E}^{(l)} = \text{LevelTransformer}_l(C_t^{(l)}, \mathbf{t}), \quad l = 1, \dots, L + 1, \quad (12)$$

280 where \mathbf{t} denotes the diffusion time embedding and $\mathbf{E}^{(l)} \in \mathbb{R}^{N \times h_l \times D}$ represents the output level
 281 embeddings for level l , with h_l denoting the embedding dimension at that level. The embeddings of
 282 each level are aggregated through attention-based pooling. Cross-level attention operates on these
 283 aggregated representations, allowing each level to adaptively incorporate contextual information
 284 from other scales through learned gating mechanisms (Figure 2).

285 4 EXPERIMENTS

286 4.1 EXPERIMENTAL SETTINGS

289 **Benchmarks** We compare WaveletDiff to several state-of-the-art time series generation methods,
 290 including FourierDiffusion Crabbé et al. (2024), Diffusion-TS Yuan & Qiao (2024), TimeGAN Yoon
 291 et al. (2019), and SigDiffusion Barancikova et al. (2025).

293 **Datasets** We use six real-world datasets to evaluate our method, covering energy, finance, and
 294 neuroscience domains. **ETTh1** and **ETTh2** Zhou et al. (2021) are electricity transformer datasets
 295 containing oil temperature and six power load features recorded hourly from 2016 to 2018. **Stocks** is
 296 a multivariate financial time series dataset containing historical Google stock market data with price
 297 and volume features from 2004 to 2019. **Exchange Rate** Lai et al. (2018) contains daily exchange
 298 rates of eight countries from 1990 to 2016. **fMRI** is the NetSim dataset containing simulated BOLD
 299 time series data for evaluating network modeling methods in functional magnetic resonance imaging.
 300 **EEG** Roesler (2013) contains multichannel electroencephalogram recordings that measure brain
 301 electrical activity over time, offering information about neural dynamics and cognitive processes.
 302 For more details, refer to Appendix B.1.

303 **Metrics** We evaluate generation quality using five complementary metrics. The **discriminative**
 304 **score** measures similarity between real and generated samples by training a binary classifier to
 305 distinguish them Yoon et al. (2019). The **predictive score** assesses the utility of synthetic data
 306 for forecasting real sequences using mean absolute error. **Context-Fréchet inception distance**
 307 (**Context-FID**) Paul et al. (2022) quantifies distributional distance using TS2Vec Yue et al. (2022)
 308 embeddings following Yuan & Qiao (2024). The **correlational score** evaluates temporal depen-
 309 dencies by comparing cross-correlation matrices. Additionally, we propose a novel metric, termed
 310 the **Dynamic Time Warping Jensen-Shannon Distance (DTW-JS distance)**, which combines the
 311 temporal alignment features of DTW with the JS divergence for distributional comparison. DTW
 312 aims to capture optimal temporal alignments between two time series sequences x and y by mini-
 313 mizing

$$314 \quad \text{DTW}(x, y) = \min_{\pi} \sum_{(i, j) \in \pi} |x_i - y_j|, \quad (13)$$

316 where π is a warping path allowing flexible temporal matching. We first create a reference set \mathcal{M}
 317 by randomly sampling sequences from the union of the real \mathcal{R} and generated \mathcal{G} datasets, which are
 318 matched in size. For each sequence s in the real dataset, we compute its mean DTW distance to all
 319 sequences in the reference set: $d_{\mathcal{R}}(s) = \frac{1}{|\mathcal{M}|} \sum_{r \in \mathcal{M}} \text{DTW}(s, r)$. We perform the same calculation
 320 for each generated sequence to obtain $d_{\mathcal{G}}(s)$. This creates two collections of mean distances across
 321 all choices of s , which we convert into empirical distributions $D_{\mathcal{R}}$ and $D_{\mathcal{G}}$. We then apply Jensen-
 322 Shannon divergence to compare these distance distributions:

$$323 \quad \text{DTW-JS}(D_{\mathcal{R}}, D_{\mathcal{G}}) = \frac{1}{2} [\text{KL}(D_{\mathcal{R}} || D_{\mathcal{M}}) + \text{KL}(D_{\mathcal{G}} || D_{\mathcal{M}})] \quad (14)$$

324 where $D_M = \frac{1}{2}(D_{\mathcal{R}} + D_{\mathcal{G}})$ is the mixture of the two distance distributions. Small DTW-JS values
 325 indicate that the real and generated samples are “distributionally” similar in terms of their temporal
 326 patterns. More details can be found in Appendix B.2.
 327

328 4.2 SHORT SEQUENCE TIME SERIES GENERATION

330 We follow the evaluation setup of TimeGAN Yoon et al. (2019) and Diffusion-TS Yuan & Qiao
 331 (2024) to assess generation quality against baseline models. All datasets are segmented into se-
 332 quences of length 24 using a sliding window with stride 1. For evaluation, we generate samples
 333 matching the size of the original training data for each dataset to ensure fair evaluation. Training
 334 configurations and times, as well as model complexity are discussed in Appendices B.4 and B.5.
 335

336 Table 1: Time series generation performance comparison on short sequences (length 24).

Metric	Methods	ETTh1	ETTh2	Stocks	Exchange Rate	fMRI	EEG
Discriminative Score (Lower the Better)	WaveletDiff	0.005±.005	0.008±.007	0.005±.004	0.004±.001	0.087±.077	0.006±.008
	FourierDiffusion	0.019±.007	0.016±.006	0.024±.003	0.015±.009	0.196±.013	0.016±.007
	Diffusion-TS	0.071±.002	0.038±.008	0.087±.008	0.032±.002	0.188±.018	0.304±.177
	TimeGAN	0.127±.047	0.106±.035	0.091±.047	0.257±.070	0.499±.001	0.161±.063
	SigDiffusions	0.353±.023	0.381±.048	0.371±.027	0.324±.055	0.482±.018	0.500±.000
Predictive Score (Lower the Better)	WaveletDiff	0.119±.002	0.106±.004	0.037±.000	0.037±.002	0.100±.000	0.000±.000
	FourierDiffusion	0.120±.005	0.111±.003	0.037±.000	0.040±.001	0.100±.000	0.000±.000
	Diffusion-TS	0.120±.004	0.107±.003	0.037±.000	0.037±.002	0.100±.000	0.001±.000
	TimeGAN	0.152±.015	0.128±.005	0.038±.000	0.064±.005	0.124±.002	0.000±.000
	SigDiffusions	0.131±.002	0.125±.003	0.040±.001	0.089±.006	0.105±.000	0.000±.000
Context-FID Score (Lower the Better)	WaveletDiff	0.020±.001	0.023±.002	0.018±.002	0.006±.000	0.104±.006	0.006±.000
	FourierDiffusion	0.031±.002	0.024±.003	0.093±.010	0.054±.013	0.169±.005	0.012±.001
	Diffusion-TS	0.151±.007	0.054±.002	0.187±.016	0.056±.007	0.106±.003	0.017±.001
	TimeGAN	0.661±.041	0.157±.011	0.110±.012	0.660±.042	1.404±.114	0.018±.001
	SigDiffusions	2.413±.179	1.053±.099	3.494±.383	1.691±.157	6.576±.210	0.022±.001
Correlational Score (Lower the Better)	WaveletDiff	0.043±.008	0.083±.016	0.005±.003	0.060±.020	1.177±.031	1.811±.963
	FourierDiffusion	0.046±.009	0.095±.016	0.013±.003	0.072±.019	1.184±.023	3.544±.626
	Diffusion-TS	0.051±.007	0.089±.022	0.009±.007	0.115±.016	1.382±.036	4.764±.107
	TimeGAN	0.202±.010	0.185±.015	0.053±.003	0.416±.018	29.562±.067	8.820±.121
	SigDiffusions	0.210±.010	0.430±.025	0.070±.005	0.943±.024	15.389±.064	4.389±.257
DTW-JS distance (Lower the Better)	WaveletDiff	0.101±.016	0.064±.014	0.106±.027	0.121±.029	0.191±.043	0.055±.011
	FourierDiffusion	0.105±.022	0.073±.014	0.138±.024	0.130±.021	0.286±.042	0.067±.017
	Diffusion-TS	0.111±.020	0.087±.012	0.153±.019	0.139±.031	0.237±.042	0.220±.013
	TimeGAN	0.155±.030	0.097±.042	0.142±.028	0.231±.025	0.215±.037	0.632±.049
	SigDiffusions	0.259±.024	0.273±.034	0.377±.068	0.376±.036	0.693±.000	0.293±.125

356 As shown in Table 1, our method consistently outperforms all baseline methods across all datasets
 357 and metrics. While FourierDiffusion achieves competitive performance on certain datasets, Wavelet-
 358 Diff demonstrates superior and more consistent results across all evaluation scenarios. Notably, it
 359 achieves 3× lower discriminative and Context-FID scores on average than the second-best baseline
 360 across all datasets. In our evaluations, we also tested different wavelet families and selected Sym-
 361 let wavelets for Stocks, Coiflets wavelets for fMRI, and Daubechies wavelets for other datasets.
 362 In general, even when universally adopting Daubechies wavelets, WaveletDiff outperforms other
 363 diffusion paradigms. More details regarding the influence of the wavelet basis function on genera-
 364 tive performance are available in Appendix B.3. Furthermore, unlike methods such as DiffusionTS
 365 that require nonuniversal trend and seasonality decompositions, our wavelet approach automatically
 366 identifies time-frequency patterns. The cross-level attention mechanism also allows for reconstruc-
 367 tion of fine-grained temporal details while maintaining global spectral coherence. This is particularly
 368 evident in terms of consistent improvements of Context-FID scores, which measure distributional
 369 similarity using learned temporal representations.

370 To highlight our model’s ability to capture real data distributions, we present t-SNE embeddings
 371 and probability density plots for ETTh1 and Stocks in Figures 3 and 4, respectively (see also Ap-
 372 pendix C.1). The density plots reveal near-perfect alignment between real and generated distribu-
 373 tions, outperforming FourierDiffusion and Diffusion-TS.

374 4.3 LONG TIME SERIES GENERATION

375 We assess the performance of WaveletDiff for long time series generation by segmenting datasets
 376 into sequences of length 32, 64, and 128, again using a sliding window with stride 1. The same
 377 evaluation protocol is applied, where we generate samples matching the size of the original training

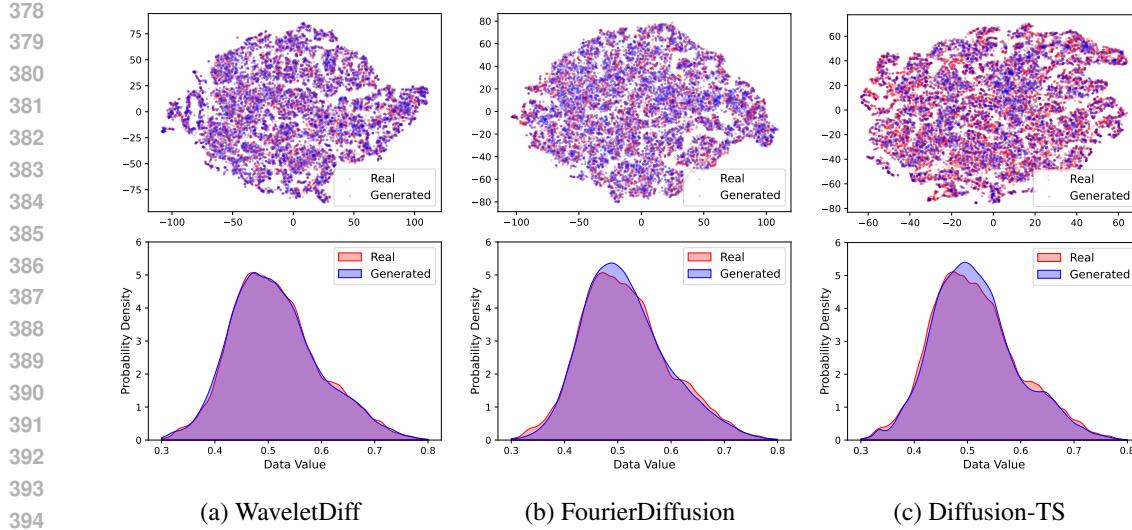


Figure 3: t-SNE visualization and probability distributions of training/synthetic data for ETTh1.

395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431

data for each dataset. As shown in Table 2, WaveletDiff once again offers consistent performance improvements across all settings. Here we used the spectral energy preservation term based on Parseval’s theorem with a loss weight $\lambda_{\text{energy}} = 0.3$ on ETTh1 and Exchange Rate, but not on Stocks due to its high volatility.

4.4 ABLATION STUDY

To validate the effectiveness of different architectural components of WaveletDiff, we conduct an ablation study comparing our full model against four variants: (1) *Predicting coefficients rather than noise*: Instead of predicting noise as in standard DDPM, this variant directly predicts the wavelet coefficients themselves, following the approach in Diffusion-TS Yuan & Qiao (2024) which suggests this method outperforms noise prediction. (2) *Removing cross-attention*: This variant disables information exchange between different wavelet decomposition levels, but maintains level-specific transformer models. (3) *Using Cosine instead of Exponential schedules*: This variant replaces exponential with cosine noise scheduling during diffusion. (4) *Using DDIM sampling*: This variant uses deterministic DDIM rather than DDPM sampling during inference. The results in Table 7 in the Appendix reveal that *cross-level attention* is the universally most critical architectural component.

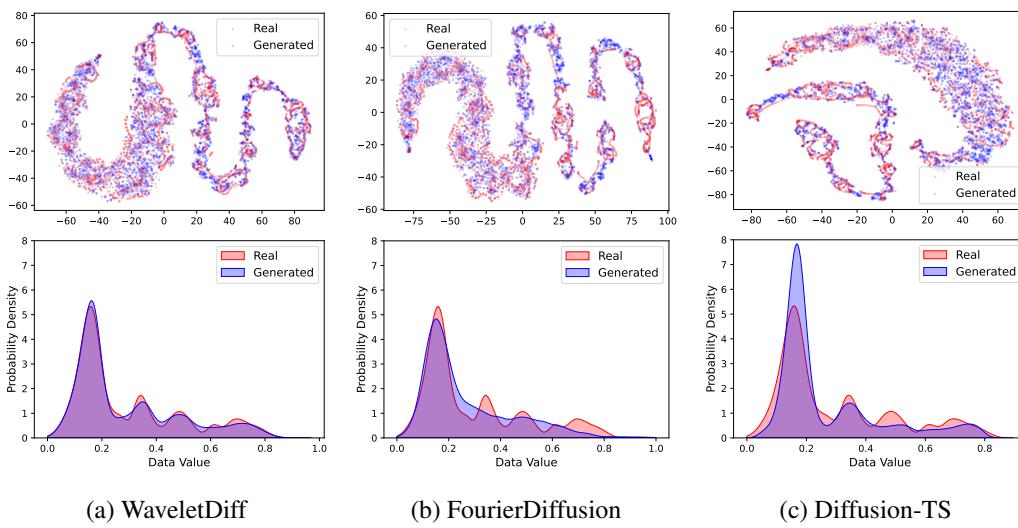


Figure 4: t-SNE visualization and probability distributions of training/synthetic data for Stocks.

Table 2: Time series generation performance comparison on long sequences.

Dataset	Metric	Length	WaveletDiff	FourierDiffusion	Diffusion-TS	TimeGAN	SigDiffusions
ETTh1	Discriminative Score (Lower the Better)	32	0.016±.001	0.030±.004	0.078±.003	0.128±.036	0.346±.033
		64	0.028±.009	0.048±.004	0.079±.010	0.116±.088	0.294±.156
		128	0.034±.037	0.113±.006	0.159±.006	0.299±.148	0.462±.035
	Predictive Score (Lower the Better)	32	0.119±.001	0.119±.005	0.119±.003	0.126±.009	0.129±.000
		64	0.114±.007	0.114±.004	0.120±.004	0.125±.004	0.129±.002
		128	0.113±.005	0.112±.007	0.116±.005	0.177±.015	0.129±.002
	Context-FID Score (Lower the Better)	32	0.038±.005	0.048±.003	0.204±.011	0.599±.044	2.875±.027
		64	0.088±.005	0.135±.010	0.265±.012	0.978±.114	6.622±.354
		128	0.256±.014	0.356±.021	0.805±.094	11.813±.851	11.596±.800
	Correlational Score (Lower the Better)	32	0.050±.004	0.056±.019	0.064±.014	0.118±.013	0.180±.013
Stocks		64	0.054±.009	0.052±.006	0.059±.010	0.307±.015	0.200±.023
		128	0.059±.021	0.072±.010	0.083±.004	1.098±.005	0.235±.015
	DTW-JS Distance (Lower the Better)	32	0.095±.022	0.099±.035	0.113±.022	0.226±.019	0.235±.017
		64	0.095±.028	0.105±.031	0.123±.034	0.208±.017	0.199±.031
		128	0.105±.035	0.134±.022	0.122±.017	0.262±.051	0.122±.021
	Discriminative Score (Lower the Better)	32	0.006±.004	0.022±.012	0.099±.012	0.197±.025	0.357±.027
		64	0.007±.003	0.032±.018	0.099±.008	0.152±.020	0.324±.044
		128	0.015±.008	0.086±.036	0.141±.011	0.270±.124	0.339±.007
	Predictive Score (Lower the Better)	32	0.037±.000	0.037±.000	0.038±.000	0.037±.000	0.040±.001
		64	0.036±.000	0.036±.000	0.037±.000	0.038±.000	0.039±.000
Exchange Rate		128	0.036±.000	0.038±.000	0.037±.000	0.070±.007	0.040±.000
	Context-FID Score (Lower the Better)	32	0.026±.006	0.087±.007	0.256±.029	0.449±.042	3.403±.373
		64	0.047±.005	0.151±.026	0.369±.065	0.336±.046	4.229±.495
		128	0.080±.012	0.379±.025	0.417±.077	3.231±.325	5.472±.004
	Correlational Score (Lower the Better)	32	0.002±.002	0.011±.001	0.017±.007	0.094±.006	0.075±.004
		64	0.003±.001	0.013±.005	0.020±.002	0.098±.003	0.052±.004
		128	0.004±.002	0.162±.011	0.021±.006	0.621±.006	0.091±.004
	DTW-JS Distance (Lower the Better)	32	0.112±.025	0.118±.021	0.137±.026	0.182±.026	0.301±.060
		64	0.136±.021	0.139±.008	0.136±.018	0.155±.031	0.261±.013
		128	0.112±.013	0.127±.020	0.116±.004	0.420±.015	0.281±.058
fMRI	Discriminative Score (Lower the Better)	32	0.011±.005	0.018±.013	0.031±.006	0.254±.064	0.314±.024
		64	0.020±.005	0.038±.015	0.028±.005	0.277±.046	0.300±.007
		128	0.026±.008	0.092±.032	0.046±.007	0.106±.064	0.276±.015
	Predictive Score (Lower the Better)	32	0.035±.002	0.040±.002	0.036±.002	0.069±.006	0.085±.007
		64	0.035±.001	0.041±.001	0.035±.002	0.056±.005	0.078±.008
		128	0.034±.003	0.044±.002	0.034±.002	0.048±.003	0.074±.005
	Context-FID Score (Lower the Better)	32	0.013±.001	4.057±.648	0.037±.003	1.038±.144	1.853±.164
		64	0.022±.003	0.129±.065	0.056±.005	1.136±.114	1.834±.235
		128	0.052±.003	0.264±.008	0.063±.004	0.849±.087	2.079±.168
	Correlational Score (Lower the Better)	32	0.064±.010	0.109±.037	0.091±.044	0.456±.016	1.065±.033
EEG		64	0.066±.024	0.096±.026	0.097±.012	0.421±.038	1.042±.027
		128	0.065±.026	0.173±.011	0.101±.019	0.237±.035	1.001±.046
	DTW-JS Distance (Lower the Better)	32	0.108±.037	0.116±.028	0.129±.023	0.182±.023	0.375±.045
		64	0.132±.015	0.142±.028	0.136±.010	0.195±.036	0.305±.024
		128	0.124±.018	0.146±.029	0.145±.028	0.224±.020	0.306±.022

With its removal causing discriminative scores and Context-FID scores to degrade on average by approximately 4 \times and 3.5 \times , respectively. While cosine noise scheduling and DDIM sampling show competitive performance on certain datasets, they exhibit instability in neuroscience domain datasets fMRI and EEG. Additionally, our analysis confirms that coefficient prediction consistently underperforms the standard DDPM noise prediction paradigm across all datasets and metrics, in contrast to the findings of Diffusion-TS which is applied directly to the time domain.

Reproducibility Analysis. Inspired by recent diffusion model reproducibility studies Zhang et al. (2024b); Li et al. (2024); Kadkhodaie et al. (2024), we examine whether this phenomenon extends to time series generation. To this end, we train model pairs with architectural variations, generate samples from identical Gaussian noise using deterministic DDIM sampling, and conclude that time series diffusion models exhibit reproducibility across all representation domains (see Appendix D).

5 CONCLUSION

We introduced WaveletDiff, a wavelet-space diffusion model that operates directly on wavelet coefficients with dedicated transformers for each decomposition level and cross-level attention mechanisms. Our approach captures multi-scale temporal patterns and preserves spectral characteristics through energy preservation training objectives. Extensive experiments across six diverse datasets demonstrate that WaveletDiff consistently outperforms state-of-the-art baselines across all evaluation metrics and sequence lengths, achieving discriminative scores and Context-FID scores that are 3 \times smaller on average than the second-best baseline.

486 THE USE OF LARGE LANGUAGE MODELS (LLMs)
487488 LLMs assisted with drafting portions of the text, correcting spelling and grammatical errors, and
489 improving clarity and style. In addition, LLMs were used in a limited capacity to assist with de-
490 bugging during code development. All technical contributions, experimental design, mathematical
491 formulations, and core insights are the original work of the authors. The authors take full responsi-
492 bility for the final content, including any LLM-assisted text or code that was subsequently reviewed
493 and validated.494
495 ETHICS STATEMENT
496497 To the best of our knowledge, our work does not raise any ethical concerns. WaveletDiff is a time
498 series generation method that operates on synthetic data generation for applications such as data
499 augmentation, privacy preservation, and forecasting research. The datasets used in our experiments
500 (ETTh1, ETTh2, Stocks, Exchange Rate, fMRI, and EEG) are publicly available research datasets
501 that do not contain sensitive personal information. Our method does not involve human subjects,
502 and the synthetic time series generation capability could potentially benefit privacy preservation
503 by enabling the creation of synthetic datasets that maintain statistical properties while protecting
504 individual privacy.505
506 REPRODUCIBILITY STATEMENT
507508 We provide comprehensive implementation details to ensure reproducibility of our results. Model
509 architecture specifications, hyperparameters, and training configurations are detailed in Section 3.2
510 and Appendix B.4. Dataset descriptions and preprocessing steps are provided in Section 4.1 and
511 Appendix B.1. Evaluation metrics and experimental protocols are specified in Section 4.1 and
512 Appendix B.2. Wavelet selection criteria and mother wavelet analysis are documented in Ap-
513 pendix B.3. Computational requirements and training procedures are outlined in Appendix B.5.
514 All experimental settings, including cross-level attention mechanisms, energy preservation
515 constraints, and ablation study configurations, are thoroughly described in the main paper and corre-
516 sponding appendix sections to enable complete reproduction of our findings. Our code is available
517 at <https://anonymous.4open.science/r/WaveletDiff-27E9/>.518 REFERENCES
519520 Paul S Addison. *The illustrated wavelet transform handbook: introductory theory and applications*
521 *in science, engineering, medicine and finance*. CRC press, 2017.
522 Ahmed M Alaa, Alex James Chan, and Mihaela van der Schaar. Generative time-series modeling
523 with fourier flows. In *International Conference on Learning Representations*, 2021.
524 Dona Arabi, Jafar Bakhshaliyev, Ayse Coskuner, Kiran Madhusudhanan, and Kami Serdar
525 Uckardes. Wave-mask/mix: Exploring wavelet-based augmentations for time series forecasting,
526 2024. URL <https://arxiv.org/abs/2408.10951>.
527 Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured
528 denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing*
529 *Systems*, 34:17981–17993, 2021.
530 Barbora Barancikova, Zhuoyue Huang, and Christopher Salvi. Sigdiffusions: Score-based diffusion
531 models for time series via log-signature embeddings. In *The Thirteenth International Confer-
532 ence on Learning Representations*, 2025. URL <https://openreview.net/forum?id=Y8KK9kjgIK>.
533 C Sidney Burrus, Ramesh A Gopinath, and Haitao Guo. *Introduction to wavelets and wavelet*
534 *transforms: a primer*. Prentice Hall, 1998.
535 Thyago P Carvalho, Fabrizzio AAM N Soares, Roberto Vita, Ronaldo da P Francisco, João P Basto,
536 and Secundino GS Alcalá. Challenges in predictive maintenance—a review. *CIRP Journal of*
537 *Manufacturing Science and Technology*, 40:53–67, 2022.

540 Guoxuan Chi, Zheng Yang, Chenshu Wu, Jingao Xu, Yuchong Gao, Yunhao Liu, and Tony Xiao
 541 Han. Rf-diffusion: Radio signal generation via time-frequency diffusion, 2024. URL <https://arxiv.org/abs/2404.09140>.
 542

543 Climate Change AI. Data gaps that matter for AI in climate action. *Climate Change AI*, 2021.
 544 Available at: <https://www.climatechange.ai/dev/datagaps>.
 545

546 Leon Cohen. The uncertainty principle for the short-time Fourier transform and wavelet transform.
 547 pp. 217–245, 2001.
 548

549 Jonathan Crabbé, Serge Assaad, Jean-Philippe Thiran, and Jens Sjölund. Time series diffusion in
 550 the frequency domain. In *Proceedings of the 41st International Conference on Machine Learning*,
 551 pp. 9491–9510, 2024.

552 Ingrid Daubechies. Time-frequency localization operators: A geometric phase space approach.
 553 volume 34, pp. 605–612, 1988.
 554

555 Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.
 556

557 Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. Time series generation under data
 558 scarcity: A unified generative modeling approach. *arXiv preprint arXiv:2505.12850*, 2025.
 559

560 Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances
 561 in neural information processing systems*, 34:8780–8794, 2021.

562 Tufa Dinku. Challenges with availability and quality of climate data in africa. *Climate Services*, 14:
 6–15, 2019.
 563

564 Khaled El Emam, Lucy Mosquera, and Richard Hopetroff. Synthetic data generation for realtime
 565 data pipelines. *Online Scientific Research*, 2022.
 566

567 MohammadReza EskandariNasab, Shah Muhammad Hamdi, and Soukaina Filali Boubrahimi. Se-
 568 riesgan: Time series generation via adversarial and autoregressive learning, 2024. URL <https://arxiv.org/abs/2410.21203>.
 569

570 Florentin Guth, Simon Coste, Valentin De Bortoli, and Stephane Mallat. Wavelet score-based gen-
 571 erative modeling. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.),
 572 *Advances in Neural Information Processing Systems*, volume 35, pp. 478–491. Curran Associates,
 573 Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/03474669b759f6d38cdca6fb4eb905f4-Paper-Conference.pdf.
 574

575 Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances
 576 in neural information processing systems*, volume 33, pp. 6840–6851, 2020.
 577

578 Jingyu Hu, Ka-Hei Hui, Zhengzhe Liu, Ruihui Li, and Chi-Wing Fu. Neural wavelet-domain diffu-
 579 sion for 3d shape generation, inversion, and manipulation, 2023. URL <https://arxiv.org/abs/2302.00190>.
 580

581 Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with
 582 irregular and scale-invariant patterns. In *The Twelfth International Conference on Learning Rep-
 583 resentations*, 2024. URL <https://openreview.net/forum?id=CdjnzWsQax>.
 584

585 James Jordon, Lukasz Szpruch, Florimond Houssiau, Mirko Bottarelli, Giovanni Cherubin, Carsten
 586 Maple, Samuel N Cohen, and Adrian Weller. Synthetic data for privacy-preserving clinical risk
 587 prediction. *Scientific Reports*, 14(1):24119, 2024.
 588

589 Zahra Kadkhodaie, Florentin Guth, Eero P Simoncelli, and Stéphane Mallat. Generalization in
 590 diffusion models arises from geometry-adaptive harmonic representations. In *The Twelfth Inter-
 591 national Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=ANvmVS2Yr0>.
 592

593 Amir Kazemi and Hadi Meidani. Time series synthesis via multi-scale patch-based generation of
 594 wavelet scalogram, 2022. URL <https://arxiv.org/abs/2211.02620>.
 595

594 Zhifeng Kong, Wei Ping, Jiahui Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile
 595 diffusion model for audio synthesis. In *International Conference on Learning Representations*,
 596 2021.

597

598 Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-
 599 term temporal patterns with deep neural networks, 2018. URL <https://arxiv.org/abs/1703.07015>.

600

601 Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. Data augmentation using synthetic
 602 data for time series classification with deep residual networks. *arXiv preprint arXiv:1808.02455*,
 603 2018.

604

605 Changhee Lee, Jinsung Yoon, and Mihaela van der Schaar. Dynamic-DeepHit: A deep learning
 606 approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE
 607 Transactions on Biomedical Engineering*, 67(1):122–133, 2019a.

608

609 Gregory R. Lee, Ralf Gommers, Filip Wasilewski, Kai Wohlfahrt, and Aaron O’Leary. Pywavelets:
 610 A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019b.
 doi: 10.21105/joss.01237.

611

612 Yaguo Lei, Bo Yang, Xinwei Jiang, Feng Jia, Naipeng Li, and Asoke K Nandi. A survey of predictive
 613 maintenance: Systems, purposes and approaches. *IEEE Access*, 8:74305–74334, 2020.

614

615 Xiang Li, Yixiang Dai, and Qing Qu. Understanding generalizability of diffusion models requires
 616 rethinking the hidden gaussian structure. In *The Thirty-eighth Annual Conference on Neural
 617 Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=Sk2duBGvrK>.

618

619 Yusuke Li, Xinyuan Lu, Yiping Wang, and Dejing Dou. Conditional score-based diffusion models
 620 for probabilistic time series imputation. *Advances in Neural Information Processing Systems*, 35:
 621 24804–24816, 2022.

622

623 Haksoo Lim, Minjung Kim, Sewon Park, and Noseong Park. Regular time-series generation using
 624 sgm, 2023. URL <https://arxiv.org/abs/2301.08518>.

625

626 Stephane G Mallat. *A theory for multiresolution signal decomposition: the wavelet representation*,
 627 volume 11. IEEE, 1989.

628

629 Rohan Mitra and Imran Zualkernan. Music generation using deep learning and generative ai: a
 systematic review. *IEEE Access*, 2025.

630

631 Sai Shankar Narasimhan, Shubhankar Agarwal, Oguzhan Akcin, Sujay Sanghavi, and Sandeep
 632 Chinchali. Time weaver: a conditional time series generation model. In *Proceedings of the
 633 41st International Conference on Machine Learning*, ICML’24. JMLR.org, 2024.

634

635 Sergey I Nikolenko. Synthetic data generation for machine learning. *arXiv preprint
 arXiv:1909.11512*, 2021.

636

637 Rachel Nosowsky and Thomas J Giordano. The Health Insurance Portability and Accountability
 638 Act of 1996 (HIPAA) privacy rule: implications for clinical research. *Annual review of medicine*,
 57:575–590, 2006.

639

640 Ahmet Murat Özbayoğlu, Mehmet Ugur Gudelek, and Ömer Berat Sezer. Deep learning for financial
 641 applications: A survey. *Applied Soft Computing*, 93:106384, 2020.

642

643 Sleziačk Patrik, Hlavčová Kamila, and Jan Szolgay. Advantages of a time series analysis using
 644 wavelet transform as compared with a fourier analysis. *Slovak Journal of Civil Engineering*, 23,
 06 2015. doi: 10.1515/sjce-2015-0010.

645

646 Jeha Paul, Bohlke-Schneider Michael, Mercado Pedro, Kapoor Shubham, Singh Nirwan Rajbir,
 647 Flunkert Valentin, Gasthaus Jan, and Januschowski Tim. Psa-gan: Progressive self attention gans
 for synthetic time series, 2022. URL <https://arxiv.org/abs/2108.00981>.

648 William Peebles and Saining Xie. Scalable diffusion models with transformers. *arXiv preprint*
 649 [arXiv:2212.09748](https://arxiv.org/abs/2212.09748), 2022.

650 Hengzhi Pei, Kan Ren, Yuqing Yang, Chang Liu, Tao Qin, and Dongsheng Li. Towards generating
 651 real-world time series data, 2021. URL <https://arxiv.org/abs/2111.08386>.

653 Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4
 654 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press,
 655 Cambridge, UK, 2000. ISBN 9780521685085.

656 Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image gener-
 657 ators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,
 658 pp. 10199–10208, 2022.

660 Jian Qian, Bingyu Xie, Biao Wan, Minhao Li, Miao Sun, and Patrick Yin Chiang. Timeldm: Latent
 661 diffusion model for unconditional time series generation, 2024. URL <https://arxiv.org/abs/2407.04211>.

663 Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising
 664 diffusion models for multivariate probabilistic time series forecasting. *International Conference*
 665 *on Machine Learning*, pp. 8857–8868, 2021.

667 Olivier Rioul and Patrick Flandrin. Time-scale energy distributions: a general class extending
 668 wavelet transforms. *IEEE Transactions on Signal Processing*, 40(7):1746–1757, 1992.

669 Oliver Roesler. EEG Eye State. UCI Machine Learning Repository, 2013. DOI:
 670 <https://doi.org/10.24432/C57G7J>.

672 Hyun Ryu et al. SimPSI: A simple strategy to preserve spectral information in time series data
 673 augmentation. *arXiv preprint arXiv:2312.05790*, 2023.

674 David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic fore-
 675 casting with autoregressive recurrent neural networks. In *International Journal of Forecasting*,
 676 volume 36, pp. 1181–1191, 2020.

678 Yan-Fang Sang. A review on the applications of wavelet transform in hydrology time series analysis.
 679 *Atmospheric Research*, 122:8–15, 2013. doi: 10.1016/j.atmosres.2012.11.003.

680 Lena Sasal, Tanujit Chakraborty, and Abdenour Hadid. W-transformers: A wavelet-based trans-
 681 former framework for univariate time series forecasting. In *2022 21st IEEE International Confer-
 682 ence on Machine Learning and Applications (ICMLA)*, pp. 671–676. IEEE, December 2022. doi:
 683 10.1109/icmla55696.2022.00111. URL <http://dx.doi.org/10.1109/ICMLA55696.2022.00111>.

685 Stephan Schlüter and Carola Deuschle. Using wavelets for time series forecasting: Does it pay off?
 686 2010. URL <https://api.semanticscholar.org/CorpusID:122828772>.

688 Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial time series fore-
 689 casting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing*,
 690 90:106181, 2020.

692 Lifeng Shen and James T Kwok. Non-autoregressive conditional diffusion models for time series
 693 prediction. In *International Conference on Machine Learning*, pp. 30739–30762, 2023.

694 Md Fahim Sikder, Resmi Ramachandranpillai, and Fredrik Heintz. Transfusion: Generating
 695 long, high fidelity time series using diffusion models with transformers, 2024. URL <https://arxiv.org/abs/2307.12667>.

697 Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using
 698 large learning rates, 2018. URL <https://arxiv.org/abs/1708.07120>.

700 Gian Antonio Susto, Andrea Schirru, Simone Pampuri, Seán McLoone, and Alessandro Beghi. Ma-
 701 chine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0.
Sustainability, 12(19):8211, 2020.

702 Ege Onur Taga, Muhammed Emrullah Ildiz, and Samet Oymak. TimePFN: Effective multivariate
 703 time series forecasting with synthetic data. In *Proceedings of the AAAI Conference on Artificial*
 704 *Intelligence*, volume 39, pp. 20761–20769, 2025.

705 Tomonori Takahashi and Takayuki Mizuno. Generation of synthetic financial time series by diffusion
 706 models. *arXiv preprint arXiv:2401.03692*, 2024.

708 Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. CSdi: Conditional score-based
 709 diffusion models for probabilistic time series imputation. *Advances in Neural Information Pro-*
 710 *cessing Systems*, 34:24804–24816, 2021.

712 Qiao Tian, Yi Chen, Zewang Zhang, Heng Lu, Linghui Chen, Lei Xie, and Shan Liu. TFGAN: Time
 713 and frequency domain based generative adversarial network for high-fidelity speech synthesis,
 714 2020. URL <https://arxiv.org/abs/2011.12206>.

715 van der Schaar Lab. Time series in healthcare: challenges and solutions. <https://www.vanderschaar-lab.com/time-series-in-healthcare/>, 2019. Accessed: April
 716 12, 2023.

719 Martin Vetterli and Jelena Kovačević. *Wavelets and subband coding*. Prentice Hall, 1995.

720 Jinliang Wang, Weifeng Zhang, Cheng Li, and Yanfeng Liu. Exploring time series analysis in
 721 frequency domain with complex-valued spectral attention and bidirectional variable mamba. pp.
 722 1–25. Springer, 2024.

724 Qiyao Wang, Ahmed Farahat, Chetan Gupta, and Sheng Zheng. Deep time series models for scarce
 725 data. *Neurocomputing*, 456:504–518, 2021.

726 Shuo Wang, Carsten Rudolph, Surya Nepal, Marthie Grobler, and Shangqi Chen. Part-GAN:
 727 Privacy-preserving time-series sharing. pp. 578–593, 2020.

729 Qingsong Wen, Liang Sun, Fan Yang, Xiuqiang Song, Jingkun Gao, Xue Wang, and Huan Xu.
 730 Time series data augmentation for deep learning: A survey. *Proceedings of the International*
 731 *Joint Conference on Artificial Intelligence*, pp. 4653–4660, 2021.

732 Zhijian Xu, Atul Sharma, Yicheng Zhang, Sercan O. Arik, and Tomas Pfister. Fits: Modeling time
 733 series with 10k parameters. In *International Conference on Learning Representations*, 2024.

735 Xinyu Yang, Yu Sun, Xiaojie Yuan, and Xinyang Chen. Frequency-aware generative mod-
 736 els for multivariate time series imputation. In A. Globerson, L. Mackey, D. Bel-
 737 grave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural In-*
 738 *formation Processing Systems*, volume 37, pp. 52595–52623. Curran Associates, Inc.,
 739 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/5e364212327fa3a59ae3595b025c469f-Paper-Conference.pdf.

741 Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing
 742 Cao, and Zhendong Niu. Frequency-domain MLPs are more effective learners in time series fore-
 743 casting. In *Advances in Neural Information Processing Systems*, volume 36, pp. 76656–76671,
 744 2023.

746 Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversar-
 747 ial networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran
 748 Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.

750 Xinyu Yuan and Yan Qiao. Diffusion-TS: Interpretable diffusion for general time series generation.
 751 In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=4h1apFj099>.

754 Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and
 755 Bixiong Xu. Ts2vec: Towards universal representation of time series, 2022. URL <https://arxiv.org/abs/2106.10466>.

756 Hengyi Zhang, Zheng Wang, and Weiwei Su. Atfnet: Adaptive time-frequency ensembled network
 757 for long-term time series forecasting. *arXiv preprint arXiv:2404.05192*, 2024a.
 758

759 Huijie Zhang, Jinfan Zhou, Yifu Lu, Minzhe Guo, Peng Wang, Liyue Shen, and Qing Qu. The
 760 emergence of reproducibility and consistency in diffusion models. In Ruslan Salakhutdinov, Zico
 761 Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp
 762 (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of
 763 *Proceedings of Machine Learning Research*, pp. 60558–60590. PMLR, 21–27 Jul 2024b. URL
 764 <https://proceedings.mlr.press/v235/zhang24cn.html>.
 765

766 Yi Zhao, Yanyan Shen, Yanmin Zhu, and Junjie Yao. Forecasting wavelet transformed time series
 767 with attentive neural networks. In *2018 IEEE International Conference on Data Mining (ICDM)*,
 768 pp. 1452–1457, 2018. doi: 10.1109/ICDM.2018.00201.
 769

770 Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang.
 771 Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-
 772 Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp.
 773 11106–11115. AAAI Press, 2021.
 774

775 Tian Zhou, Ziqing Ma, Xue Wang, Qingsong Wen, Liang Sun, and Rong Jin. Film: Frequency
 776 improved legendre memory model for long-term time series forecasting. *Advances in Neural
 777 Information Processing Systems*, 35:12677–12690, 2022a.
 778

779 Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency
 780 enhanced decomposed transformer for long-term series forecasting. *International Conference on
 781 Machine Learning*, pp. 27268–27286, 2022b.
 782

783 Ziyu Zhou, Jiaxi Hu, Qingsong Wen, James T. Kwok, and Yuxuan Liang. Multi-order wavelet
 784 derivative transform for deep time series forecasting, 2025. URL <https://arxiv.org/abs/2505.11781>.
 785

810 A MOTHER WAVELET FAMILIES
811

812 Different wavelet families provide distinct characteristics of multi-scale decompositions through
813 their specific filter coefficients $\{h_k\}$ and $\{g_k\}$ in the two-scale relations (Equation 1). In our ex-
814 periments, we used five representative wavelet families, each satisfying different relationships be-
815 tween their order p and filter length F . The vanishing moment properties and definitions presented here
816 follow the PyWavelets framework implementation Lee et al. (2019b).

817
818 A.1 DAUBECHIES WAVELETS (DB)
819

820 Daubechies wavelets of order p (e.g., db- p) provide orthogonality, compact support, and exactly p
821 vanishing moments for the wavelet function $\psi(t)$, with filter length $F = 2p$. The scaling func-
822 tion $\phi(t)$ has zero vanishing moments for orthogonal wavelets. The filter coefficients $\{h_k\}_{k=0}^{F-1}$ are
823 derived from polynomial factorization to maximize regularity and maintain compact support.

824 For db2 (Daubechies-2 with $p = 2, F = 4$), the low-pass filter coefficients equal:
825

$$826 \quad h_0 = \frac{1 + \sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, \quad (15)$$

$$827 \quad h_2 = \frac{3 - \sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1 - \sqrt{3}}{4\sqrt{2}}. \quad (16)$$

831 The high-pass coefficients satisfy $g_k = (-1)^k h_{F-1-k}$. The orthogonality and vanishing moment
832 conditions ensure that
833

$$834 \quad \sum_{k=0}^{F-1} h_k = \sqrt{2}, \quad (17)$$

$$835 \quad \sum_{k=0}^{F-1} h_k h_{k+2m} = \delta_{m,0}, \quad (18)$$

$$836 \quad \sum_{k=0}^{F-1} k^j h_k = 0 \quad \text{for } j = 1, \dots, p-1, \quad (19)$$

843 where $\delta_{m,0}$ is the Kronecker delta function, defined as:
844

$$845 \quad \delta_{m,0} = \begin{cases} 1 & \text{if } m = 0 \\ 0 & \text{if } m \neq 0. \end{cases} \quad (20)$$

846
847 A.2 SYMLETS (SYM)

848 Symlets are modified Daubechies wavelets designed to improve symmetry while maintaining or-
849 thogonality and compact support. With a filter length $F = 2p$, Symlets have the same vanishing
850 moment properties as Daubechies wavelets, i.e., p vanishing moments for the wavelet function $\psi(t)$
851 and zero vanishing moments for the scaling function $\phi(t)$. They minimize an asymmetry measure
852 A that quantifies the deviation from perfect symmetry, namely
853

$$854 \quad A = \sum_{k=0}^{F-1} k \cdot |h_k|^2 - \frac{F-1}{2} \sum_{k=0}^{F-1} |h_k|^2, \quad (21)$$

855 where the first term represents the weighted center of mass of the filter coefficients, while the second
856 term represents the theoretical center for a perfectly symmetric filter. A lower value of A indicates
857 better symmetry.
858

859 For sym2 ($p = 2, F = 4$), the coefficients are optimized versions of db2 coefficients, satisfying the
860 same orthogonality conditions but with improved phase linearity and near-symmetric properties for
861 better temporal localization.
862

864 A.3 COIFLETS (COIF)
865

866 Coiflets of order p are designed with balanced vanishing moments for both the scaling $\phi(t)$ and
867 wavelet function $\psi(t)$, with filter length $F = 6p$. The wavelet function $\psi(t)$ has $2p$ vanishing
868 moments while the scaling function $\phi(t)$ has $2p - 1$ vanishing moments, providing more balance for
869 the moments when compared to Daubechies wavelets (in which case the scaling function has zero
870 vanishing moments). The filter coefficients satisfy extended moment conditions of the form

$$871 \quad \sum_{k=0}^{F-1} h_k = \sqrt{2}, \quad (22)$$

$$874 \quad \sum_{k=0}^{F-1} k^j h_k = 0 \quad \text{for } j = 1, \dots, 2p-1, \quad (23)$$

$$877 \quad \sum_{k=0}^{F-1} g_k = 0, \quad (24)$$

$$880 \quad \sum_{k=0}^{F-1} k^j g_k = 0 \quad \text{for } j = 1, \dots, 2p. \quad (25)$$

883 For coif1 ($p = 1, F = 6$), the wavelet function has two vanishing moments and the scaling function
884 has one vanishing moment. The six filter coefficients provide enhanced moment balancing between
885 analysis and synthesis operations, with the scaling function having non-zero vanishing moments
886 unlike orthogonal Daubechies wavelets.

887 A.4 BIORTHOGONAL WAVELETS (BIOR)
888

889 Biorthogonal wavelets use different filters for decomposition and reconstruction, denoted as
890 bior_{p_r, p_d} where p_r and p_d are the orders that determine the vanishing moment properties. For a
891 general bior_{p_r, p_d} wavelet, we have the following:

- 893 • *Wavelet function* $\psi(t)$: p_r vanishing moments;
- 894 • *Scaling function* $\phi(t)$: p_d vanishing moments;
- 895 • *Filter lengths*: These depend on the specific bior_{p_r, p_d} configuration and are not given by
896 simple formulas like those of other wavelet families.

897 For decomposition one uses a low-pass filter $\{h_k\}$ and a high-pass filter $\{g_k\}$, while for recon-
898 struction one uses a low-pass filter $\{\tilde{h}_k\}$ (denoted with tilde) and a high-pass filter $\{\tilde{g}_k\}$. The tilde
899 notation $\tilde{\cdot}$ indicates the dual (reconstruction) filters that are different from the primal (decomposition)
900 filters.

902 For bior2.2 ($p_r = p_d = 2$), both decomposition and reconstruction wavelet functions have two
903 vanishing moments, and both scaling functions have two vanishing moments, which result in perfect
904 symmetry. The filters also satisfy the perfect reconstruction condition,

$$905 \quad \sum_k h_k \tilde{h}_{k+2m} + g_k \tilde{g}_{k+2m} = \delta_{m,0} \quad (26)$$

908 In the z-domain, where $H(z)$, $G(z)$, $\tilde{H}(z)$, and $\tilde{G}(z)$ are the z-transforms of the respective filter
909 sequences, the perfect reconstruction condition is succinctly summarized as

$$910 \quad H(z)\tilde{H}(z^{-1}) + H(-z)\tilde{H}(-z^{-1}) = 2. \quad (27)$$

912 A.5 REVERSE BIORTHOGONAL WAVELETS (RBIO)
913

914 Reverse biorthogonal wavelets (rbior_{p_r, p_d}) interchange the decomposition and reconstruction filter
915 roles compared to standard biorthogonal wavelets, according to:

$$916 \quad h_k^{\text{rbio}} = \tilde{h}_k^{\text{bior}} \quad (28)$$

$$917 \quad g_k^{\text{rbio}} = \tilde{g}_k^{\text{bior}} \quad (29)$$

918 For reverse biorthogonal wavelets, the vanishing moment assignment follows the same pattern as
 919 biorthogonal wavelets:
 920

- 921 • *Wavelet function* $\psi(t)$ has p_r vanishing moments.
- 922 • *Scaling function* $\phi(t)$ has p_d vanishing moments.

924 For rbio2.2 ($p_r = p_d = 2$), both the wavelet function $\psi(t)$ and scaling function $\phi(t)$ have two
 925 vanishing moments each, maintaining symmetric properties.

926 The choice of wavelet family affects the sparsity and localization properties of the decomposition,
 927 with symmetric wavelets (bior, rbio) providing better phase preservation, while orthogonal wavelets
 928 (db, sym) ensure energy conservation through orthogonality.

930 B EXPERIMENTAL DETAILS

931 B.1 DATASETS

934 Table 3 lists detailed properties of the datasets used in our experiments, and their repository links.

936 Dataset	# of Samples	Dim	Source
937 ETTh1	17420	7	https://github.com/zhouhaoyi/ETDataset
938 ETTh2	17420	7	https://github.com/zhouhaoyi/ETDataset
939 Stocks	3685	6	https://finance.yahoo.com/quote/GOOG
940 Exchange Rate	7588	8	https://github.com/laiguokun/multivariate-time-series-data
941 fMRI	10000	50	https://www.fmrib.ox.ac.uk/datasets/netsim
942 EEG	14980	14	https://archive.ics.uci.edu/dataset/264/eeg+eye+state

943 Table 3: Summary of the dataset types and statistics.
 944

945 B.2 METRICS

948 **Discriminative Score.** The discriminative score captures how difficult it is for a classifier to dis-
 949 tinguish between real and generated samples. The score is measured by $|\text{acc} - 0.5|$, where acc is
 950 the classification accuracy. A score close to 0 indicates that real and generated samples are indistin-
 951 guishable to the classifier, while a score close to 0.5 indicates they are very different. We follow the
 952 setup of TimeGAN Yoon et al. (2019) using a 2-layer GRU-based neural network as the classifier,
 953 trained with binary cross-entropy loss to distinguish between real (label=1) and synthetic (label=0)
 954 sequences.

955 **Predictive Score.** The predictive score captures how useful generated samples are for the forecasting
 956 task on real data. The score is measured by the mean absolute error (MAE) between predicted values
 957 and ground-truth values on test data. We follow TimeGAN Yoon et al. (2019) using a 2-layer GRU-
 958 based sequence predictor trained on synthetic data to predict the next time step features, evaluated
 959 on real sequences. Lower MAE values indicate better predictive utility of the generated samples.

960 **Context-FID.** Paul et al. (2022) The Fréchet Inception Distance (FID) measures the distance be-
 961 tween two multivariate Gaussian distributions, i.e.,

$$962 \text{FID}(X, Y) = \|\mu_X - \mu_Y\|^2 + \text{Tr}(\Sigma_X + \Sigma_Y - 2(\Sigma_X \Sigma_Y)^{1/2}), \quad (30)$$

964 where μ_X, μ_Y are the means and Σ_X, Σ_Y are the covariance matrices of the two distributions.
 965 Context-FID adapts this to time series by replacing the Inception-v3 features with time series fea-
 966 tures. We follow Diffusion-TS Yuan & Qiao (2024) using TS2Vec Yue et al. (2022) representations
 967 as the features. We extract embeddings from both real and generated sequences using a trained
 968 TS2Vec encoder, then compute FID in the embedding space. Lower Context-FID values indicate
 969 better distributional similarity.

970 **Correlational Score.** This metric assesses temporal dependencies by comparing cross-correlation
 971 matrices between real and generated data. For sequences with D features, we compute the sam-
 972 ple covariance matrix for each dataset, convert them to correlation matrices, and then measure the

972 average absolute difference across all feature pairs according to
 973

$$974 \quad \text{Correlational Score} = \frac{1}{10} \sum_{i=1}^D \sum_{j=1}^D |\rho_{i,j}^{real} - \rho_{i,j}^{generated}|, \quad (31)$$

$$975$$

$$976$$

977 where $\rho_{i,j}^{real}$ and $\rho_{i,j}^{generated}$ are the correlation coefficients between features i and j for real and
 978 generated data, respectively. Note that we follow the Diffusion-TS Yuan & Qiao (2024) setup using
 979 the factor $\frac{1}{10}$, although $\frac{1}{D^2}$ could provide better normalization across different feature dimensions.
 980 The former choice of normalization ensures direct comparability with prior work.

981 **DTW-JS Distance.** We propose a Dynamic Time Warping Jensen-Shannon Distance (DTW-JS
 982 distance) metric, which combines DTW’s temporal alignment capabilities with Jensen-Shannon di-
 983 vergence for distributional comparison. DTW computes the optimal alignment distance between
 984 two time series sequences x and y by minimizing

$$985 \quad \text{DTW}(x, y) = \min_{\pi} \sum_{(i,j) \in \pi} |x_i - y_j| \quad (32)$$

$$986$$

$$987$$

988 where x and y are two time series sequences, π represents a warping path consisting of index pairs
 989 (i, j) that map elements from sequence x to sequence y , and the path must satisfy DTW constraints:
 990 monotonicity (indices only increase), continuity (no skipping), and boundary conditions (path starts
 991 at $(1, 1)$ and ends at $(|x|, |y|)$). The warping allows sequences to be stretched or compressed along
 992 the time axis to find the best alignment, enabling DTW to handle sequences of different lengths and
 993 account for temporal shifts or speed variations between similar patterns.

994 For our metric, we create a reference set \mathcal{M} by randomly sampling from both real samples \mathcal{R} and
 995 generated samples \mathcal{G} (i.e., by taking the union of samples of these two sets, and ensuring that both
 996 sets have the same number of elements). For each sample s in the real set \mathcal{R} and generated set \mathcal{G} ,
 997 we compute its mean DTW distance to all samples in the reference set:

$$998 \quad d(s) = \frac{1}{|\mathcal{M}|} \sum_{r \in \mathcal{M}} \text{DTW}(s, r) \quad (33)$$

$$999$$

$$1000$$

1001 This creates two collections of mean DTW distances, which we histogram across different samples
 1002 s to form distance distributions $D_{\mathcal{R}}$ and $D_{\mathcal{G}}$ for real and generated samples, respectively. We then
 1003 apply Jensen-Shannon divergence to compute the distance between the two distance distributions,
 1004 i.e.,

$$1005 \quad \text{DTW-JS}(D_{\mathcal{R}}, D_{\mathcal{G}}) = \frac{1}{2} [\text{KL}(D_{\mathcal{R}} || D_M) + \text{KL}(D_{\mathcal{G}} || D_M)] \quad (34)$$

$$1006$$

1007 where $D_M = \frac{1}{2}(D_{\mathcal{R}} + D_{\mathcal{G}})$ is the mixture of the two distance distributions. This approach mea-
 1008 sures distributional similarity between real and generated samples while accounting for temporal
 1009 alignment flexibility, providing a robust evaluation metric that captures both temporal structure and
 1010 statistical properties.

1011 B.3 WAVELET BASIS FUNCTION ANALYSIS

1012 The choice of mother wavelet significantly influences the multi-scale decomposition characteristics
 1013 and subsequent series generation quality. Different wavelet families exhibit distinct properties in
 1014 terms of orthogonality, compact support, and smoothness, which directly affect the sparsity and lo-
 1015 calization of coefficients in the wavelet domain. This choice becomes especially critical when deal-
 1016 ing with diverse dataset characteristics, as different signal types require wavelets that can optimally
 1017 capture their specific temporal-spectral patterns. As shown in Table 4, we systematically evaluated
 1018 five representative wavelet families: Daubechies (db), Symlets (sym), Coiflets (coif), Biorthogonal
 1019 (bior), and reverse Biorthogonal wavelets (rbio).

1020 The results reveal dataset-specific preferred wavelets: Symlets works best for the Stocks dataset,
 1021 likely due to their enhanced symmetry properties that better capture the near-symmetric fluctuations
 1022 characteristic of financial time series. Coiflets demonstrate the best performance on the Exchange
 1023 Rate dataset, benefiting from balanced vanishing moments for both scaling and wavelet functions,
 1024 which effectively capture the smooth yet complex dynamics of currency fluctuations. For the re-
 1025 maining datasets (ETTh1, ETTh2, fMRI, EEG), Daubechies wavelets consistently provide the best

overall performance. This ability to adapt the wavelet basis to match dataset characteristics represents an important advantage of WaveletDiff over frequency-domain approaches, which are constrained to use fixed Fourier basis functions regardless of the underlying signal properties, limiting their capacity to optimally represent diverse temporal patterns across different domains.

Metrics	Wavelet	ETTh1	ETTh2	Stocks	Exchange Rate	fMRI	EEG
Discriminative Score (Lower the Better)	db	0.005±.005	0.008±.007	0.013±.007	0.004±.001	0.175±.071	0.006±.008
	sym	0.023±.005	0.023±.005	0.005±.004	0.011±.009	0.196±.066	0.007±.003
	coif	0.025±.010	0.031±.004	0.017±.009	0.073±.010	0.087±.077	0.014±.014
	bior	0.022±.008	0.033±.002	0.012±.004	0.051±.012	0.273±.007	0.008±.004
	rbio	0.057±.009	0.074±.008	0.010±.008	0.094±.010	0.129±.127	0.017±.009
Predictive Score (Lower the Better)	db	0.119±.002	0.106±.004	0.037±.000	0.037±.002	0.100±.000	0.000±.000
	sym	0.117±.004	0.107±.003	0.037±.000	0.035±.003	0.100±.000	0.000±.000
	coif	0.115±.005	0.106±.004	0.037±.000	0.036±.003	0.100±.000	0.000±.000
	bior	0.122±.002	0.109±.004	0.037±.000	0.037±.001	0.100±.000	0.000±.000
	rbio	0.121±.004	0.104±.001	0.037±.000	0.036±.001	0.100±.000	0.000±.000
Context-FID Score (Lower the Better)	db	0.020±.001	0.023±.002	0.024±.004	0.006±.000	0.104±.003	0.006±.000
	sym	0.052±.004	0.051±.006	0.018±.002	0.009±.001	0.122±.007	0.011±.001
	coif	0.079±.008	0.069±.009	0.018±.003	0.108±.008	0.104±.006	0.006±.000
	bior	0.049±.003	0.156±.016	0.016±.001	0.088±.013	0.119±.004	0.006±.001
	rbio	0.161±.005	0.225±.041	0.016±.002	0.175±.027	0.176±.008	0.007±.001
Correlational Score (Lower the Better)	db	0.043±.008	0.083±.016	0.006±.003	0.060±.020	1.073±.005	1.811±.963
	sym	0.055±.008	0.073±.025	0.005±.003	0.066±.012	1.172±.048	2.164±.533
	coif	0.036±.006	0.064±.011	0.007±.004	0.167±.032	1.177±.031	1.971±.969
	bior	0.048±.011	0.094±.009	0.005±.003	0.137±.020	1.147±.033	3.034±.759
	rbio	0.051±.008	0.099±.026	0.003±.004	0.161±.018	1.402±.034	1.959±.707
DTW-JS Distance (Lower the Better)	db	0.101±.016	0.064±.014	0.121±.013	0.121±.029	0.199±.043	0.055±.011
	sym	0.123±.009	0.067±.023	0.106±.027	0.132±.017	0.283±.035	0.049±.015
	coif	0.104±.015	0.086±.015	0.115±.016	0.157±.022	0.191±.011	0.062±.017
	bior	0.117±.013	0.095±.019	0.109±.009	0.129±.033	0.280±.011	0.050±.009
	rbio	0.110±.021	0.086±.033	0.119±.024	0.144±.037	0.464±.021	0.068±.022

Table 4: Mother Wavelet Selection

B.4 TRAINING CONFIGURATION DETAILS

We provide training configuration information needed for reproducibility. The WaveletDiff model uses an embedding dimension of 256 with 8 attention heads across 8 transformer layers, a time embedding dimension of 128, and dropout rate of 0.1. The approximation level transformer uses twice as many embedding dimensions (512) and twice the number of layers (16) to capture the critically important low-frequency information.

The diffusion process employs 1000 timesteps with an exponential noise schedule. The exponential noise schedule is of the form

$$\beta_t = \beta_{start} + (\beta_{end} - \beta_{start}) \cdot (1 - e^{-\gamma \cdot t}) \quad (35)$$

where $\beta_{start} = 0.0001$, $\beta_{end} = 0.02$, $\gamma = 2.0$ is the exponential decay rate, $t \in [0, 1]$ is the normalized timestep, and $T = 1000$ is the total number of timesteps. The coefficient-weighted loss strategy assigns an approximation coefficient weight of 2.0 to emphasize low-frequency components. We train for 5000 epochs with batch sizes 512.

For optimization, we use the AdamW optimizer with initial learning rate 2×10^{-4} and weight decay 1×10^{-5} . We employ a one-cycle learning rate schedule Smith & Topin (2018) with cosine annealing strategy. The learning rate follows a two-phase schedule: Linear warm-up for the first 30% of training, then cosine annealing for the remaining 70%:

$$lr(e) = \begin{cases} lr_{base} + (lr_{max} - lr_{base}) \times \frac{e}{p \times E}, & \text{if } e \leq p \times E; \\ lr_{final} + (lr_{max} - lr_{final}) \times \left(\frac{1 + \cos(\pi \times \frac{e - p \times E}{(1 - p) \times E})}{2} \right), & \text{if } e > p \times E. \end{cases} \quad (36)$$

where e is the current epoch, $lr_{base} = 4 \times 10^{-5}$, $lr_{max} = 1 \times 10^{-3}$, $lr_{final} = 4 \times 10^{-9}$, $p = 0.3$, and $E = 5000$ total epochs.

1080 B.5 COMPUTATIONAL SETUP AND TRAINING TIME ANALYSIS
1081

1082 WaveletDiff experiments were conducted on a single NVIDIA H100 GPU with 80GB memory using
1083 PyTorch 2.7.1 with CUDA 11.8, with training performed for 5000 epochs. Our WaveletDiff model
1084 contains approximately 63M trainable parameters. FourierDiffusion baseline results were obtained
1085 on a Tesla T4 GPU due to hardware constraints, following their original configuration and training
1086 settings. We note that the difference in GPU hardware was unavoidable due to High-Performance
1087 Computing (HPC) resource availability and the PyTorch Lightning deployment used for WaveletD-
1088 iff, and is not selected to give any unfair advantage to our model. Tables 5 and 6 present the training
1089 times for WaveletDiff and FourierDiffusion across different datasets and sequence lengths, respec-
1090 tively. Despite the large model and long training, WaveletDiff completes training in a relatively short
1091 time using only a single GPU, indicating that compute time is clearly not a bottleneck.

1092
1093 Table 5: Training times (hours:minutes:seconds) for WaveletDiff across different datasets and se-
1094 quence lengths on single NVIDIA H100 GPU.
1095

sequence length	ETTh1	ETTh2	Stocks	Exchange Rate	fMRI	EEG
24	3:45:54	3:36:34	1:07:05	1:36:59	2:40:13	3:22:06
32	3:45:33	3:38:23	1:02:49	1:46:32	2:42:41	3:24:07
64	3:24:46	4:19:04	1:12:34	2:03:44	3:15:34	4:13:58
128	5:06:03	6:27:03	1:33:53	2:59:33	4:32:52	6:01:14

1104 Table 6: Training times (hours:minutes:seconds) for FourierDiffusion across different datasets and
1105 sequence lengths on single NVIDIA Tesla T4 GPU.
1106

sequence length	ETTh1	ETTh2	Stocks	Exchange Rate	fMRI	EEG
24	1:09:51	1:19:01	1:57:23	2:21:39	2:21:38	2:39:03
32	1:38:56	1:39:13	2:10:41	2:30:54	2:27:28	3:47:09
64	4:49:37	4:49:39	3:14:13	2:31:45	3:58:26	5:21:55
128	10:14:05	10:14:04	6:35:18	3:46:43	6:29:26	7:16:49

Metrics	Methods	ETTh1	ETTh2	Stocks	Exchange Rate	fMRI	EEG
Discriminative Score (Lower the Better)	WaveletDiff	0.005±.005	0.008±.007	0.005±.004	0.004±.001	0.087±.077	0.006±.008
	coefficient prediction	0.017±.013	0.040±.030	0.027±.014	0.059±.010	0.277±.012	0.020±.013
	w/o cross attention	0.055±.054	0.028±.017	0.016±.012	0.052±.011	0.179±.016	0.006±.003
	cosine noise scheduler	0.024±.014	0.021±.019	0.094±.013	0.012±.005	0.112±.037	0.500±.000
Predictive Score (Lower the Better)	DDIM sampling	0.135±.094	0.017±.005	0.010±.003	0.006±.004	0.494±.006	0.024±.017
	WaveletDiff	0.119±.002	0.106±.004	0.037±.000	0.037±.002	0.100±.000	0.000±.000
	coefficient prediction	0.120±.003	0.111±.003	0.037±.000	0.038±.003	0.100±.000	0.000±.000
	w/o cross attention	0.119±.002	0.106±.002	0.037±.000	0.037±.001	0.100±.000	0.000±.000
Context-FID Score (Lower the Better)	cosine noise scheduler	0.119±.003	0.107±.003	0.037±.000	0.037±.002	0.103±.000	0.172±.239
	DDIM sampling	0.118±.005	0.106±.004	0.037±.000	0.036±.002	0.101±.000	0.000±.000
	WaveletDiff	0.020±.001	0.023±.002	0.018±.002	0.006±.000	0.104±.006	0.006±.000
	coefficient prediction	0.027±.003	0.059±.004	0.053±.007	0.085±.008	0.131±.010	0.009±.001
Correlational Score (Lower the Better)	w/o cross attention	0.118±.006	0.044±.004	0.039±.006	0.042±.004	0.170±.008	0.008±.001
	cosine noise scheduler	0.125±.005	0.050±.009	0.116±.021	0.011±.002	0.329±.022	44246±3637
	DDIM sampling	0.135±.006	0.049±.004	0.008±.002	0.006±.000	2.394±.043	0.009±.001
	WaveletDiff	0.043±.008	0.083±.016	0.005±.003	0.060±.020	1.177±.031	1.811±.963
DTW-JS Distance (Lower the Better)	coefficient prediction	0.056±.010	0.097±.029	0.006±.003	0.182±.019	2.005±.051	2.650±.314
	w/o cross attention	0.046±.016	0.078±.016	0.006±.004	0.092±.035	1.755±.062	1.993±.738
	cosine noise scheduler	0.050±.014	0.105±.026	0.029±.010	0.070±.009	2.175±.071	1.487±.579
	DDIM sampling	0.057±.011	0.094±.018	0.005±.003	0.080±.031	2.113±.044	0.844±.150

1133 Table 7: Ablation study results for key WaveletDiff architectural components.

1134
1135

C VISUALIZATION

1136

C.1 T-SNE AND DATA DISTRIBUTION ON SHORT SEQUENCE GENERATION

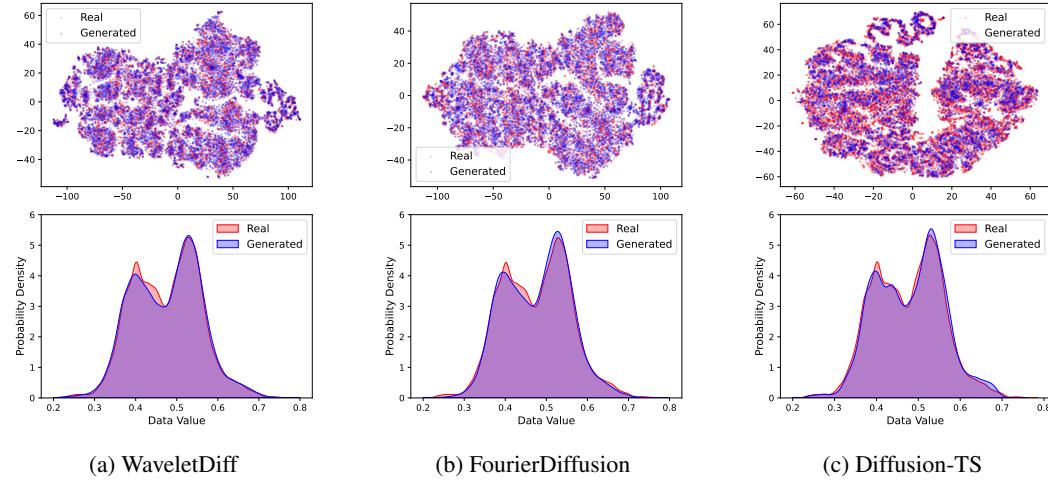
1137
1138 We present the t-SNE and data distribution visualization of short sequence generation on ETTh2,
1139 Exchange Rate, fMRI, and EEG dataset in Figure 5, 6, 7, and 8. The results demonstrate that
1140 WaveletDiff consistently outperforms baseline methods in capturing the underlying data distribu-
1141 tions across all datasets.1157
1158

Figure 5: t-SNE visualization and probability distribution of data values on ETTh2 dataset.

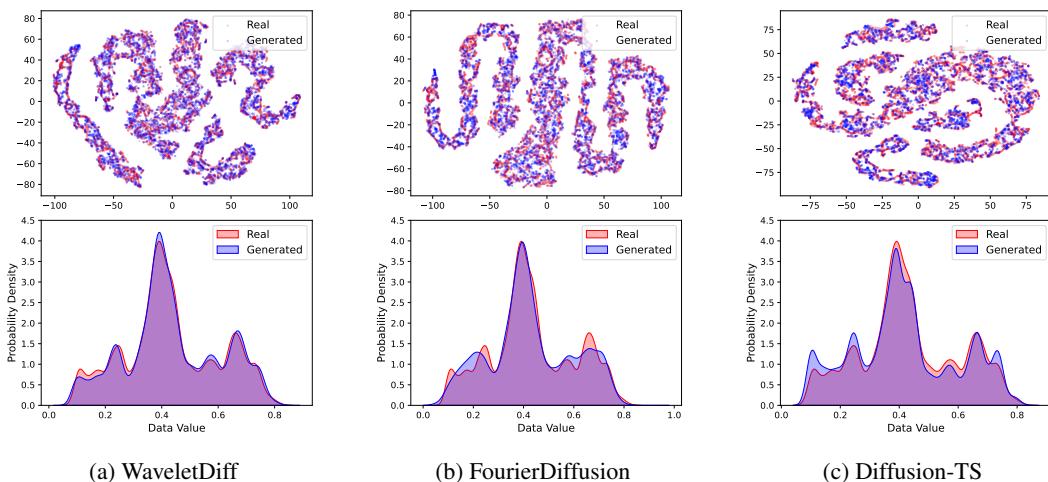
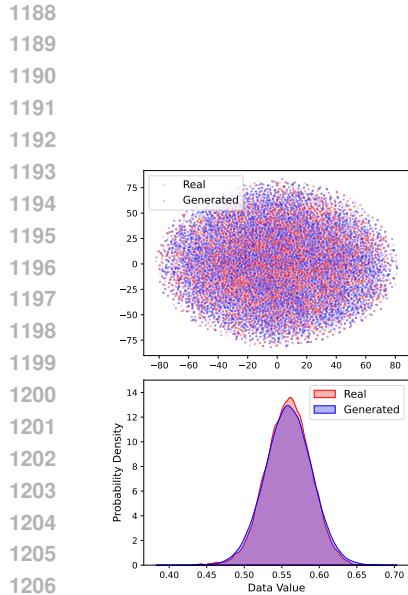
1159
1160

Figure 6: t-SNE visualization and probability distribution of data values on Exchange Rate dataset.

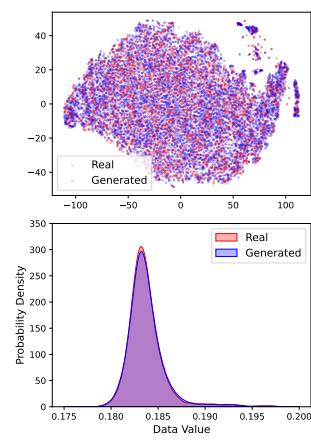
1181

D DIFFUSION MODEL REPRODUCIBILITY ANALYSIS

1182
1183 Inspired by recent work on reproducibility in diffusion models for images Zhang et al. (2024b); Li
1184 et al. (2024); Kadkhodaie et al. (2024), we examine whether this phenomenon extends to time series
1185 generation. Specifically, we train pairs of models with slightly different architectures or configura-
1186 tions, then generate samples from the same fixed Gaussian noise input using deterministic DDIM
1187 sampling. For each pair of generated sequences (x_1, x_2) , we compute their similarity using dy-
1188 namic time warping (DTW) distance. To quantify reproducibility, we follow Zhang et al. (2024b)



(a) WaveletDiff (b) FourierDiffusion (c) Diffusion-TS



(a) WaveletDiff (b) FourierDiffusion (c) Diffusion-TS

Figure 7: t-SNE visualization and probability distribution of data values on fMRI dataset.

1242 and define the RP score as
 1243

$$1244 \text{RP score} := \mathbb{P}(\text{DTW}(x_1, x_2) < \overline{\text{DTW}}_{\text{rand}}), \quad (37)$$

1245 where $\overline{\text{DTW}}_{\text{rand}}$ denotes the average DTW distance between randomly chosen sequence pairs generated
 1246 by the two models. Thus, the RP score measures the probability that two models produce more
 1247 similar samples from the same noise than would be expected by chance. An RP score greater than
 1248 0.5 indicates reproducibility.
 1249

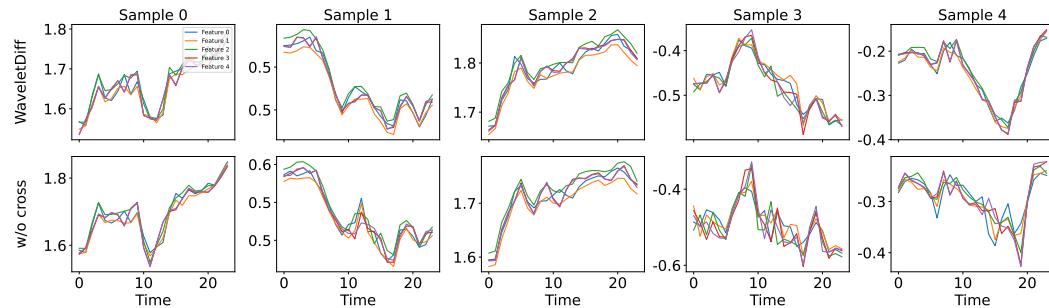
1250 Unlike image generation, which commonly uses U-Net architectures for comparison, time series
 1251 generation employs diverse architectures. We examine the RP score across different model variations
 1252 for both WaveletDiff and FourierDiffusion. Table 8 demonstrates that reproducibility exists
 1253 for time series data regardless of the representation domain (wavelet, time, or Fourier). To the best
 1254 of our knowledge, we are the first to examine the reproducibility phenomenon specifically for time
 1255 series generation.
 1256

Datasets	Model 1	Model 2	RP score
Stocks	WaveletDiff	WaveletDiff w/o cross-attention	1.0
		WaveletDiff + cosine noise scheduler	0.66
	FourierDiffusion	FourierDiffusion on time domain	0.665
		FourierDiffusion using LSTM score model	0.805
Exchange Rate	WaveletDiff	WaveletDiff w/o cross-attention	0.999
		WaveletDiff + cosine noise scheduler	0.619
	FourierDiffusion	FourierDiffusion on time domain	0.610
		FourierDiffusion using LSTM score model	0.945

1265 Table 8: Reproducibility scores for different model variations demonstrate that time series diffusion
 1266 models exhibit reproducibility across architectural changes and representation domains.
 1267

1268 D.1 IMPACT OF ARCHITECTURAL VARIATIONS

1269 We evaluate how architectural modifications affect reproducibility by removing the cross-attention
 1270 module from WaveletDiff. Figures 9 and 10 show that the model maintains strong reproducibility
 1271 despite this significant architectural change, with generated sequences from identical noise exhibiting
 1272 nearly identical patterns.
 1273



1285 Figure 9: Reproducibility comparison on Stocks dataset using identical initial noise (volume feature
 1286 excluded for clarity).
 1287

1288 D.2 IMPACT OF MOTHER WAVELET SELECTION

1289 We further investigate how different mother wavelet choices affect reproducibility. As shown in Figures 11 and 12, Daubechies and Symlets wavelets demonstrate high reproducibility, while Coiflets,
 1290 Biorthogonal, and Reverse Biorthogonal wavelets also exhibit good consistency. This indicates that
 1291 wavelet choice significantly affects the learned distribution, with similar wavelet families (e.g., or-
 1292 thogonal wavelets like Daubechies and Symlets) producing more comparable results than dissimilar
 1293 families.
 1294

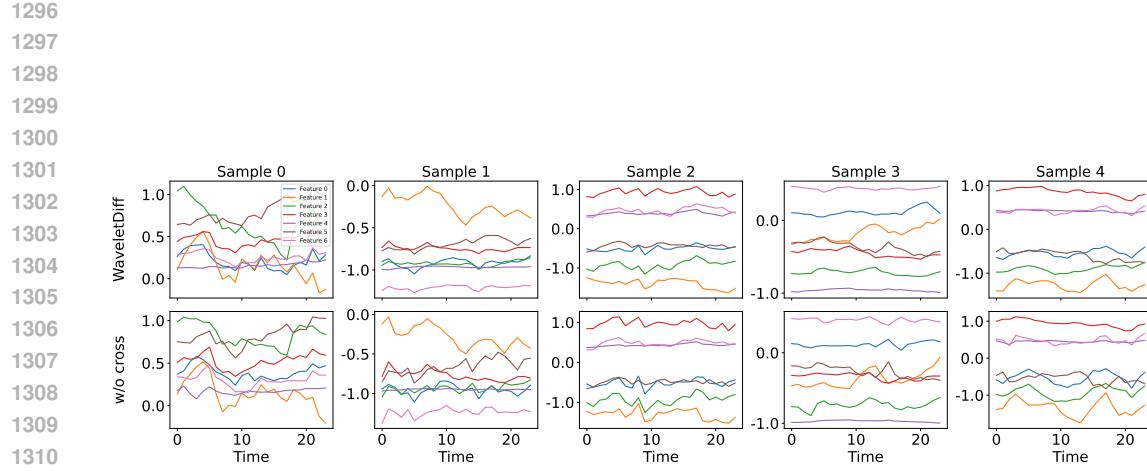


Figure 10: Reproducibility comparison on Exchange Rate dataset using identical initial noise.

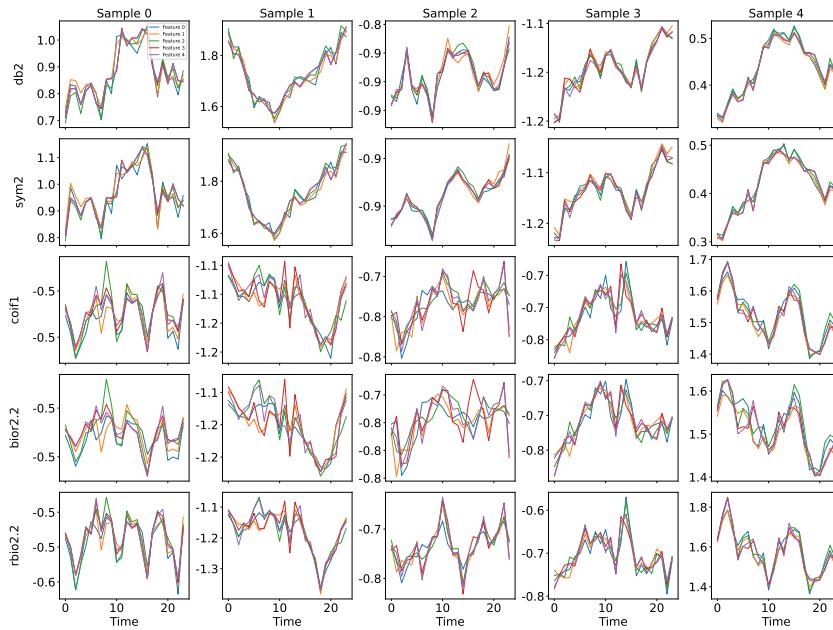


Figure 11: Wavelet family comparison on Stocks dataset demonstrating varying reproducibility across different mother wavelets.

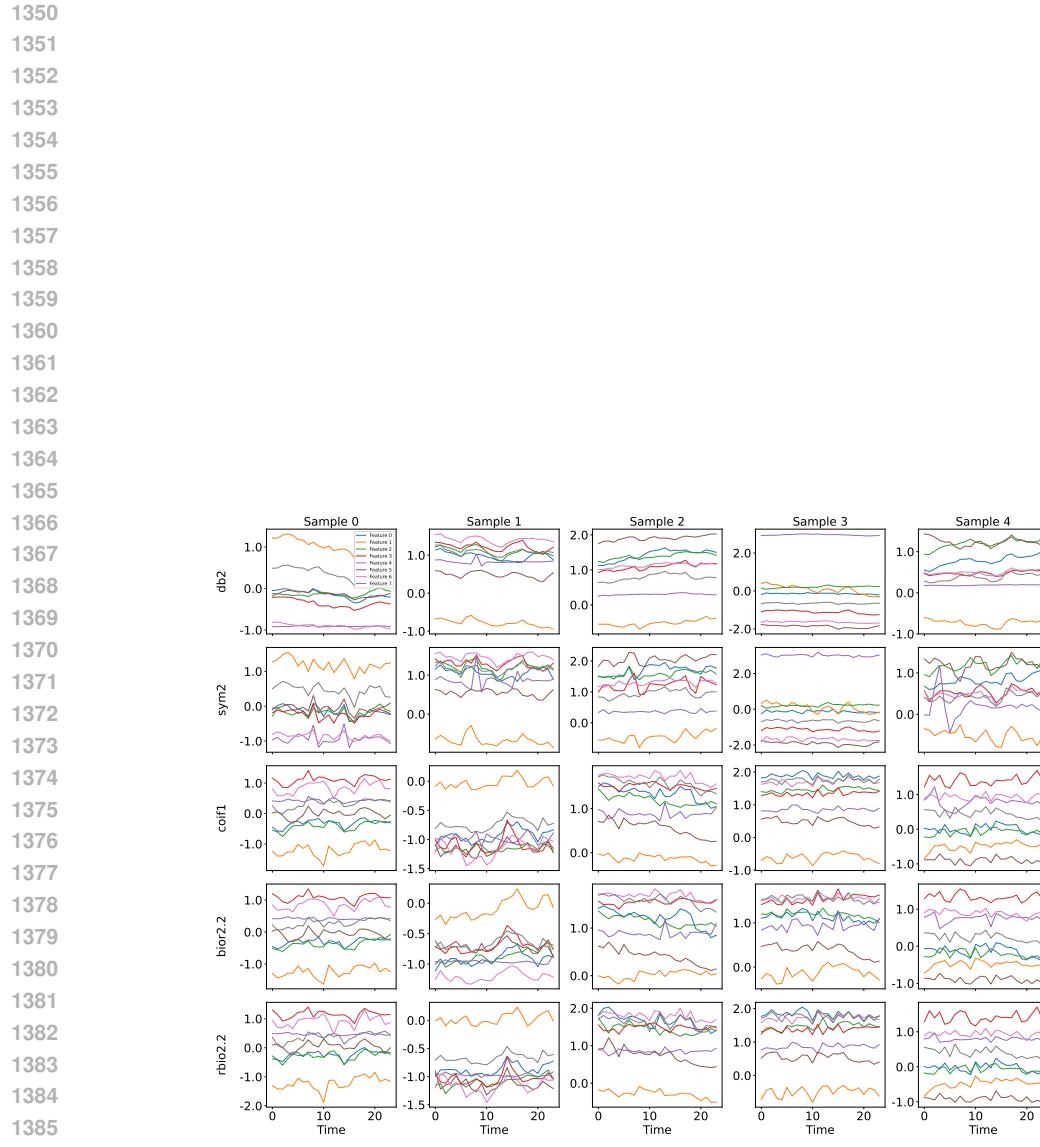


Figure 12: Wavelet family comparison on Exchange Rate dataset demonstrating varying reproducibility across mother wavelets.