DOWNSAMPLING TIME SERIES VIA AUTO-REGRESSION

Anonymous authors
Paper under double-blind review

ABSTRACT

High-resolution time series are essential for granular analysis but often prohibitively expensive to process. While low-resolution representations offer a pragmatic alternative, conventional direct downsampling acts as a destructive projection that irreversibly disrupts latent dynamical structures, leading to significant performance collapse in downstream tasks. We challenge this long-standing hegemony by proposing that the transition to lower resolutions should be an active, distribution-aware synthesis rather than a passive, deterministic pruning. To realize this paradigm shift, we introduce `VARes`, a framework that reformulates downsampling as a hierarchical auto-regressive process. Empirical results demonstrate that data obtained with `VARes` consistently outperforms traditionally downsampled counterparts by amplifying task-critical features and effectively denoising signals. Our work establishes a new downsampling standard, proving that our method yields data that is better suited for machine learning models on downstream time series tasks than conventional downsampling approaches.

1 INTRODUCTION

The explosion of high-resolution (HR) time series data has created a paradoxical challenge in modern AI: while HR sequences offer granular insights, their immense computational footprint and storage requirements often render them impractical for large-scale deployment or long-term historical analysis. Consequently, low-resolution (LR) representations remain the pragmatic backbone of real-world systems. Traditionally, the transition from HR to LR has been treated as a trivial preprocessing step, almost exclusively handled via direct downsampling (e.g., decimation or pooling). In this paper, we challenge the pervasive reliance on passive downsampling as the de facto standard, arguing that it acts as a destructive projection that irreversibly compromises the structural integrity of temporal data for downstream tasks.

The failure of conventional downsampling is most evident in downstream task performance. As shown in Figure 1, when state-of-the-art models perform time series imputation on data downsampled by a factor of 2, they suffer a consistent and significant performance collapse compared to the original HR baseline. This degradation suggests that naive downsampling is not merely summarizing the data; it is actively disrupting the underlying dynamical features and high-frequency dependencies required for accurate reconstruction. We posit that this is a fundamental limitation of deterministic, local operators which ignore the global data distribution.

We propose a radical alternative: active, distribution-aware synthesis. Instead of treating LR data as a pruned version of the HR original, we view it as a sample from a resolution-aware generative manifold. In principle, a generative model that captures the true distribution of time series can synthesize LR sequences that satisfy the statistical invariants of the HR source, thereby protecting task-critical information. However, current generative paradigms Ang et al. (2023)—such as GANsYoon et al. (2019); Xu et al. (2020); Li et al. (2022); Seyfi et al. (2022), VAEsLee et al. (2023); Naiman et al. (2023); Chen et al. (2024); Gu et al., and Diffusion modelsLim et al. (2023); Yuan & Qiao (2024); Naiman et al. (2024); Galib et al. (2024); Narasimhan et al. (2024a); Gonen et al. (2025); Narasimhan et al. (2024b); Li et al. (2025a)—are predominantly resolution-static. They are designed for fixed-scale modeling and cannot natively navigate the continuum between coarse and fine temporal resolutions without relying on the very downsampling techniques we seek to replace.

To bridge this gap, we introduce `VARes`, a framework that reformulates downsampling as a hierarchical auto-regressive process. While inspired by the Visual AutoRegressive (VAR) paradigm Tian et al.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

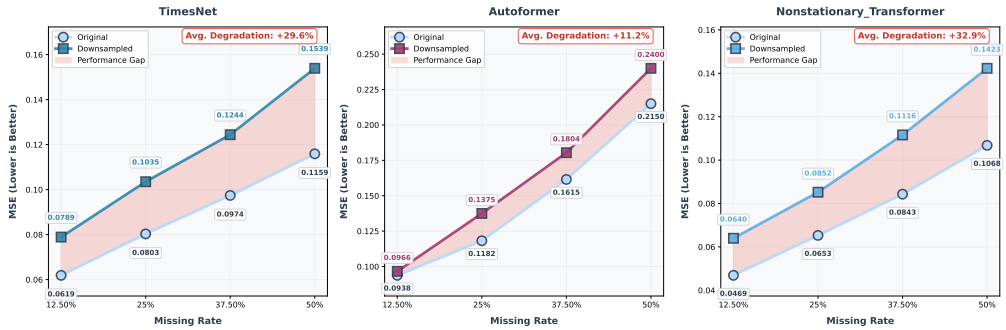

Figure 1: Imputation performance comparison between original-resolution data and $2\times$ downsampled data across three representative time series imputation models under different temporal resolutions.

(2024), VARes addresses the unique structural gap of 1D temporal data through a temporal-spatial co-embedding. By treating resolution as a latent hierarchy, VARes allows for the fluid downsampling of sequences across scales while preserving underlying dynamical invariants. Our technical contributions include: (i) Delay-embedded manifold mapping: A novel mechanism that aligns 1D temporal trajectories with 2D spatial lattices, ensuring that the local topology of the time series is preserved during coarse-to-fine transitions. (ii) Channel-independent projections: A strategy to mitigate inter-variable interference in multivariate settings, preventing feature smearing during downsampling.

Experimental results validate our challenge to the downsampling status quo. First, we demonstrate that LR data obtained with VARes consistently outperforms conventional downsampled data across four major downstream tasks, revealing that task-relevant temporal features are better preserved within our framework. Second, qualitative analysis shows that our model does not just preserve but can amplify critical high- and low-frequency components, effectively denoising the sequence while maintaining its predictive power.

2 Related Work

Time Series Generation. Existing methods primarily use four generative frameworks: GANs with adversarial learning Yoon et al. (2019); Xu et al. (2020); Li et al. (2022); Seyfi et al. (2022); VAEs with variational inference Desai et al. (2021); Lee et al. (2023); Naiman et al. (2023); Chen et al. (2024); Gu et al.; flow-based models with reversible transformations Alaa et al. (2021); Deng et al. (2020); and diffusion models with iterative denoising Lim et al. (2023); Huang et al. (2024); Yuan & Qiao (2024); Naiman et al. (2024); Galib et al. (2024); Nikitin et al. (2024); Narasimhan et al. (2024a); Gonen et al. (2025); Fadlon et al. (2025); Narasimhan et al. (2024b); Li et al. (2025a). Although effective at modeling statistical and local dynamics, these approaches are generally resolution-static and do not natively model transitions across temporal scales. They therefore rely on pre-sampled targets or deterministic downsampling, which can discard high-frequency dependencies and latent dynamical structure. This motivates a hierarchical, distribution-aware framework for resolution transitions.

Multi-scale Time Series. Conventional multi-resolution methods rely on deterministic reduction, including interval sampling along the sample dimension (US-S) Shao et al. (2022) or temporal dimension (US-T) Wu et al. (2021). Frequency-aware alternatives include Weighted Moving Average (WMA) for local trend preservation Hyndman & Athanasopoulos (2018), Discrete Wavelet Transform (DWT) for low-frequency decomposition Daubechies (1992); Zhou et al. (2021), and anti-aliasing Butterworth filtering (BW-FB) with scale-dependent cut-off α Butterworth et al. (1930). These methods are efficient but passive: they reduce or smooth the signal rather than synthesize a distribution-consistent representation, potentially removing high-frequency information and dynamical invariants. VARes instead treats resolution reduction as an active generative process. We further discuss time-series representation learning and feature compression in Appendix L.

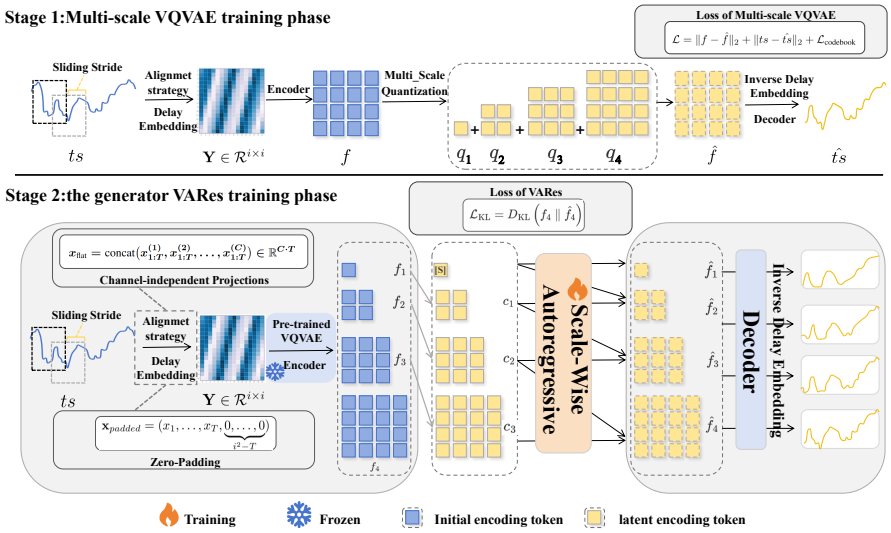

Figure 2: Overview of the proposed VARes.

3 METHODOLOGY

Motivation: Why VAR for Downsampling? Inspired by the multi-scale residual connection design of the image generation model Visual AutoRegressive (VAR) models Tian et al. (2024), we leverage this architecture to accomplish our downsampling task. Since this model achieves feature learning across scales by segmenting the overall feature map of the original scale through residual connections and the superposition of features learned at each scale, it enables, first, reducing the difficulty of feature learning at each scale through a task-segmentation approach, thereby simplifying the feature learning task for each segment. Second, since the feature map learned at a given scale is the superposition of feature maps learned at that scale and all preceding scales, it represents not only the feature learning task assigned to that scale but also the superposition of features learned at the original scale by all preceding scales. Therefore, it is feasible to directly use this as a coarser-grained reconstruction feature map to achieve high-quality downsampling tasks.

3.1 DOWNSAMPLING MODEL

Delay-embedded Manifold Mapping To harness the multi-scale generative power of VAR models Tian et al. (2024) for time series, two fundamental challenges must be addressed: structural alignment and temporal preservation. Standard VAR models are optimized for 2D spatial lattices, whereas time series data resides in a 1D temporal domain. Bridging this gap requires an encoding mechanism that can map 1D sequences into 2D manifolds without sacrificing inherent temporal dependencies.

A straightforward reshaping of 1D data into 2D matrices often disrupts local temporal continuities. To provide a stronger inductive bias for the image-based generator, we employ Delay Embedding Takens (2006), a method rooted in dynamical systems theory for reconstructing phase spaces. A brief theoretical discussion regarding the operational conditions for delay embedding is provided in Appendix F.

Mathematically, given a univariate time series $\mathbf{x} = \{x_1, x_1, \ldots, x_T\} \in \mathcal{R}^T$, we transform it into a 2D representation $\mathbf{Y} \in \mathcal{R}^{i \times i}$ that mimics the topology of an image. This process is parameterized by the embedding dimension i and the delay stride j. To ensure the sequence length matches the target matrix dimensions, we apply zero-padding to the original series, denoted as $\mathbf{x}_{padded}$. The transformation is formulated as follows:

$$\mathbf{x}_{padded} = (x_1, \ldots, x_T, \underbrace{0, \ldots, 0}_{i^2-T})$$

$$\mathbf{Y} = \begin{bmatrix} x_1 & x_{1+j} & \cdots & x_{1+(i-1)j} \\ x_2 & x_{2+j} & \cdots & x_{2+(i-1)j} \\ \vdots & \vdots & \ddots & \vdots \\ x_i & x_{i+j} & \cdots & x_{i+(i-1)j} \end{bmatrix} \tag{1}$$

The rationale for using Delay Embedding extends beyond simple reshaping. In the context of modern temporal modeling Nie (2022); Wu et al. (2022), this transformation is conceptually equivalent to hierarchical patch extraction. Each column of $\mathbf{Y}$ can be viewed as a "temporal patch" of size i, while the row-wise transitions capture long-range dependencies across a stride j.

By mapping 1D data into this 2D structured format, we unlock the following advantages for the VAR framework:

- Spatial-Temporal Locality: The 2D convolutional kernels in VAR can simultaneously capture intra-patch correlations (local fluctuations within a window) and inter-patch dependencies (global trends across time steps) through shared-weight filters.
- Resolution Scalability: The multi-scale residual connection in VAR can now operate on the 2D matrix at varying granularities, which naturally corresponds to learning the time series at different sampling frequencies (resolutions).
- Feature Fusion: This representation facilitates the fusion of latent temporal features within a single sample with the cross-temporal patterns across the entire sequence, effectively aligning the 1D temporal task with the 2D hierarchical learning paradigm of VAR.

We provide a more comprehensive evaluation of the *delayed embedding manifold mapping* under realistic non-stationary and distribution shift conditions in Appendix I. The results substantiate that the proposed mapping maintains stable and consistent predictive performance even under complex non-stationary dynamics.

Channel-independent Projections A fundamental discrepancy exists between multivariate time series and image-centric models: the former typically feature a domain-specific number of channels C, whereas the latter are optimized for a fixed (e.g., RGB) input dimenionality. To bridge this gap, we evaluate and refine two primary alignment strategies.

Limitations of Projection-based Alignment. One intuitive approach is to employ dynamic convolutional layers to project variable channels into a unified latent space, similar to ImagenFew Gonen et al. (2025). However, in a two-stage framework, this paradigm introduces significant bottlenecks: (i) Optimization Instability: The auxiliary 2D convolutions coupled with a multi-scale VQVAE van den Oord et al. (2017) increase training volatility due to mutual weight interference. (ii) Information Dissipation: The non-invertible nature of convolutional projections and subsequent shape restoration inevitably leads to the degradation of fine-grained temporal features. (iii) Gradient Decoupling: The independent training of the VQVAE and the downsampler prevents end-to-end optimization of the projection layers, further hindering reconstruction fidelity.

Challenges in Spatial Stacking. Alternatively, stacking channels into spatial dimensions (width/height) via interpolation or padding introduces undesirable complexity. Given the diversity of channel counts across domains, such transformations often lack common factors with target image resolutions, resulting in inconsistent feature scales and aliasing artifacts that impede model convergence.

Channel-Independent Flattening. Drawing inspiration from the channel-independent modeling success in PatchTST Nie (2022), we propose a robust alignment via Single-Channel Flattening. Specifically, we concatenate all C channels into a singular elongated sequence before applying delay embedding:

$$x_{\text{flat}} = \text{concat}\big(x_{1:T}^{(1)}, x_{1:T}^{(2)}, \ldots, x_{1:T}^{(C)}\big) \in \mathbb{R}^{C \cdot T}, \tag{2}$$

where T denotes the temporal length per channel. This approach offers three distinct advantages: 1) Domain Agnosticism: It natively handles heterogeneous channel counts without structural modification. 2) Feature Preservation: By treating each channel as an independent temporal segment

Table 1: Comparison of different training strategies on the Stocks dataset. Here, *load_VAR_d16* refers to fine-tuning the second-stage `VARes` model using ImageNet-pretrained weights, while *train_from_scratch* denotes direct training on time series data without pretrained initialization.

Dataset	Exp_Setting	DS ↓	PS ↓	C-FID ↓
Stocks	load_VAR_d16	0.500	0.103	28.340
	train_from_scratch	**0.497**	**0.081**	**21.989**

within the 2D manifold, we circumvent the information loss inherent in convolutional projection. 3) Experimental Fidelity: Consistent with downstream forecasting requirements, this method maintains high-quality reconstruction at native resolutions, ensuring that cross-channel correlations are preserved in the latent space through the subsequent generative process.

Training Multi-Scale VQVAE from Scratch. Both the multiscale VQVAE van den Oord et al. (2017) in the first stage of VAR Tian et al. (2024) and the main VAR image generator in the second stage provide model weights pre-trained on the ImageNet dataset. It is worth investigating whether data in the time series domain shares overlapping features with image domain data. Specifically, could using small-scale time series data for two-stage fine-tuning on the multiscale VQVAE and VAR weights—trained on the large-scale ImageNet dataset—enhance the downsampling quality of our time-series model?

Initially inspired by the success of VisionTS Chen et al. (2025), we attempted to directly fine-tune the pre-trained weights of the multi-scale VQVAE and VAR Tian et al. (2024) models with small-scale time series datasets. We designed single-input variables and channel rendering to align with the image domain, aiming to transfer feature learning expertise from the image domain to achieve better downsampling quality. However, experimental results revealed that this approach yielded poor performance, as demonstrated in Table 1.

As shown in Table 1, freezing the first-stage VQVAE and attempting to either fine-tune on ImageNet pre-trained weights or train from scratch in the second stage resulted in poor generation quality. Therefore, we directly trained the multi-scale VQVAE from scratch using time-series data and modified the original loss function to address the misalignment in feature learning directions between image and time-series domains:

$$\mathcal{L} = \|f - \hat{f}\|_2 + \|ts - \hat{ts}\|_2 + \mathcal{L}_{\text{codebook}}, \tag{3}$$

where f is the feature map obtained from the encoder at the original scale, $\hat{f}$ is the feature map after reconstructed. ts is the ground-truth time series, $\hat{ts}$ is the reconstructed time series obtained by restoring data after decoding through the multi-scale VQVAE, and $\mathcal{L}_{\text{codebook}}$ is the codebook quantization loss.

The redesigned loss function comprises two major components: reconstruction loss and quantization loss. The core concept involves placing a multi-scale VQVAE before and after the main VARes downsampler, which achieves two functions—constructing the latent feature space during training and mapping outputs to the numerical space. Therefore, the loss function is constructed by addressing both reconstruction performance (reconstruction loss) and feature mapping learning effectiveness (quantization loss). Given the inherent differences between reconstructing time series and feature maps, the reconstruction loss designed for this section comprises two components, namely temporal reconstruction loss and feature map reconstruction loss.

To bridge the gap between general-purpose and specialized temporal modeling, we refine the architecture through domain-specific adaptations in normalization, codebook-based loss objectives, and data range processing; specific implementations and technical formulations are deferred to Appendix E.

4 Experiments

4.1 Impact of Direct Downsampling on Downstream Tasks

To evaluate the effectiveness of $2\times$ downsampled data, we conduct experiments across four downstream tasks: long-term forecasting, missing value imputation, anomaly detection, and classification.

Table 2: Summary of performance impact: Original vs. 2× Downsampled. $\Delta\%$ represents the relative performance change (positive indicates degradation for MSE, negative for Accuracy).

Task	**Metric**	**Original (Avg.)**	**Downsampled (Avg.)**	$\Delta\%$
Forecasting	MSE ↓	0.382	0.449	+17.5%
Imputation	MSE ↓	0.098	0.124	+26.5%
Anomaly	F_1 ↑	0.842	0.773	-8.2%
Classification	Acc ↑	0.756	0.752	-0.5%

Table 3: Imputation Accuracy of different strategies on multiple datasets. $\Delta\%$ represents the average error increase of the best-performing baseline relative to Ours (PG). Detailed per-task results are in Appendix Table 13.

MSE ↓	**Strategy**	**TimesNet (Avg.)**	**Autoformer (Avg.)**	**NST (Avg.)**	$\Delta\%$
	PG (Ours)	**5.8e-06**	**7.3e-03**	**9.9e-05**	
	US-S	2.6e-04	2.4e-02	5.2e-04	237%
	US-T	3.2e-04	2.2e-02	5.9e-04	212%
Stocks	WMA	2.3e-04	2.1e-02	5.0e-04	198%
	DWT-A	4.2e-04	5.4e-02	9.2e-04	644%
	BW-FB 1.5	3.9e-04	2.1e-02	7.2e-04	203%
	BW-FB 2.0	4.6e-04	2.1e-02	8.8e-04	207%
	PG (Ours)	**1.2e-05**	**1.4e-02**	**1.6e-04**	
	US-S	1.5e-04	1.6e-01	7.1e-03	1094%
	US-T	1.4e-04	6.2e-02	6.0e-03	389%
Sine	WMA	1.7e-04	6.2e-02	6.0e-03	387%
	DWT-A	2.8e-04	1.1e-01	1.1e-02	788%
	BW-FB 1.5	3.9e-04	2.1e-02	7.2e-04	61%
	BW-FB 2.0	1.6e-04	6.5e-02	5.8e-03	409%
	PG (Ours)	**3.2e-05**	**3.2e-03**	**1.5e-04**	
	US-S	5.5e-03	7.1e-02	6.6e-03	2378%
	US-T	5.0e-03	5.0e-02	7.1e-03	1752%
Energy	WMA	2.9e-03	4.0e-02	4.2e-03	1315%
	DWT-A	5.5e-03	1.0e-01	7.8e-03	3348%
	BW-FB 1.5	6.7e-03	6.6e-02	1.0e-02	2368%
	BW-FB 2.0	9.6e-03	7.6e-02	1.5e-02	2864%
	PG (Ours)	**3.4e-06**	**4.9e-04**	**2.1e-05**	
	US-S	2.6e-05	3.1e-02	1.4e-04	5982%
	US-T	4.1e-05	3.8e-02	1.3e-04	7373%
Mujoco	WMA	1.3e-04	3.9e-02	2.1e-04	7465%
	DWT-A	1.0e-04	8.2e-02	2.9e-04	15943%
	BW-FB 1.5	4.5e-05	3.7e-02	1.3e-04	7153%
	BW-FB 2.0	4.9e-05	3.7e-02	1.3e-04	7117%

Table 4: Prediction Accuracy of different data construction strategies on multiple datasets. $\Delta\%$ represents the average error increase of the best-performing baseline relative to Ours (PG). Detailed per-task results are in Appendix Table 14.

MSE ↓	**Strategy**	**TimesNet (Avg.)**	**Autoformer (Avg.)**	**NST (Avg.)**	$\Delta\%$
	PG (Ours)	**9.6e-06**	**2.8e-03**	**2.4e-05**	
	US-S	5.0e-04	7.6e-03	6.5e-04	212%
	US-T	4.9e-04	5.4e-03	4.9e-04	129%
Stocks	WMA	3.9e-04	8.2e-03	3.5e-04	221%
	DWT-A	7.9e-04	1.1e-02	6.9e-04	359%
	BW-FB 1.5	6.5e-04	5.7e-03	6.4e-04	151%
	BW-FB 2.0	8.6e-04	6.1e-03	8.5e-04	179%
	PG (Ours)	**9.0e-06**	**3.3e-02**	**2.9e-05**	
	US-S	7.0e-05	4.1e-01	1.6e-03	1138%
	US-T	3.0e-05	1.2e-01	6.8e-04	267%
Sine	WMA	4.3e-05	1.0e-01	5.2e-04	208%
	DWT-A	7.2e-05	2.4e-01	1.5e-03	618%
	BW-FB 1.5	5.1e-05	1.1e-01	7.2e-04	232%
	BW-FB 2.0	4.6e-05	1.0e-01	7.2e-04	211%
	PG (Ours)	**1.5e-06**	**3.0e-03**	**6.7e-07**	
	US-S	1.7e-05	2.0e-02	5.2e-05	568%
	US-T	7.7e-06	1.8e-02	2.4e-05	491%
Energy	WMA	6.3e-06	1.6e-02	1.6e-05	440%
	DWT-A	1.2e-05	2.8e-02	3.6e-05	841%
	BW-FB 1.5	1.0e-05	2.1e-02	3.4e-05	593%
	BW-FB 2.0	1.4e-05	2.4e-02	5.2e-05	712%
	PG (Ours)	**1.7e-07**	**2.1e-05**	**6.9e-07**	
	US-S	1.5e-06	1.9e-03	2.4e-06	8770%
	US-T	1.3e-06	2.6e-03	2.6e-06	11806%
Mujoco	WMA	3.7e-06	4.3e-03	2.6e-06	19617%
	DWT-A	1.0e-05	5.3e-03	6.0e-06	24664%
	BW-FB 1.5	4.3e-06	2.7e-03	3.0e-06	12285%
	BW-FB 2.0	4.4e-06	2.9e-03	3.1e-06	13223%

We employ representative models (e.g., TimesNetWu et al. (2022), TimeMixerWang et al. (2024)) and multiple benchmark datasets (e.g., ETT, PSM, and UEA datasets). Detailed dataset descriptions, full experimental tables and parameters of the downstream task models are provided in Appendix A.

Results & Analysis. As summarized in Table 2, direct downsampling yields distinct impacts across tasks:

- *Fine-grained Tasks (Forecasting, Imputation, Anomaly Detection)*: A consistent and significant performance degradation is observed. For instance, forecasting MSE increases by over 10% on average. This indicates that these tasks rely heavily on high-frequency variations and local dynamic patterns, which are inevitably destroyed by simple subsampling.
- *Global-feature Tasks (Classification)*: Surprisingly, downsampling maintains or even slightly improves accuracy on several datasets (e.g., FaceDetection). This suggests that classification models prioritize global discriminative profiles over precise time-step alignment, and downsampling may act as a noise-reduction regularizer.

These results validate that direct downsampling is inadequate for most resolution-sensitive time series tasks.

4.2 QUANTITATIVE VALIDITY OF VARES DOWNSAMPLING

We evaluate the utility of LR data obtained with VARes against traditional downsampling strategies across four domains: Stocks (financial), Energy (physical), MuJoCo (robotics), and Sine (synthetic). We use three state-of-the-art backbones—TimesNetWu et al. (2022), AutoformerWu et al. (2021), and Non-stationary Transformer (NST)Liu et al. (2022)—to perform downstream forecasting and imputation. Full dataset descriptions, metric definitions, parameters of our models, and baseline configurations are detailed in Appendix B. Considering that downsampling via VARes introduces a certain degree of stochasticity compared to traditional deterministic downsampling, we incorporate multi-dimensional statistical metrics in Appendix H to discuss the impact of this stochasticity on the stability and confidence intervals of downstream predictive variables.

Superiority in Preserving Task-Critical Dynamics. The empirical evaluation across four benchmark datasets for imputation (Table 3) and forecasting (Table 4) reveals that LR data obtained with VARes consistently mitigates the "performance collapse" inherent in passive downsampling. Unlike deterministic pruning, our active synthesis ensures that latent dynamical structures remain intact at lower resolutions.

Order-of-Magnitude Performance Gains: Across all backbones, VARes achieves a consistent reduction in MSE. In the imputation task, VARes demonstrates robust superiority; for instance, on the Energy dataset (TimesNet, 50% mask), it achieves an MSE of 4.64×10^{-5}, outperforming the best passive baseline WMA (3.63×10^{-3}) by nearly two orders of magnitude. This gap becomes even more transformative in forecasting, especially on the complex MuJoCo system. On the Autoformer backbone (Horizon 96), VARes reaches an MSE of 8.8×10^{-6}, a $160\times$ improvement over the best-performing passive baseline (1.4×10^{-3}).

Systemic Advantage in Complex Manifolds: The lead maintained by VARes is not localized but systemic. In the Energy dataset, characterized by high-dimensional non-linear coupling, VARes reduces average forecasting error by 607.5% across all models. Even in the relatively periodic Sine dataset, our method maintains a $2\times$ to $10\times$ performance lead, proving that active synthesis is beneficial regardless of the signal's intrinsic predictability.

These results suggest that passive methods treat data reduction as a simple signal-processing problem (filtering/sampling), which inadvertently destroys the "dynamical signature" of the series. By contrast, VARes treats the transition to low resolution as a distribution-preserving synthesis task, thereby providing downstream models with a distilled yet structurally complete surrogate of the high-resolution truth.

Resilience to Information Decay and Sparsity. A crucial hallmark of a robust representation is its ability to maintain structural integrity as task difficulty increases—either through longer temporal horizons or higher data sparsity.

Scaling Performance Margins in Long-Term Forecasting: The performance lead of VARes scales significantly with task complexity. In the MuJoCo dataset (Non-stationary Transformer), the relative error reduction of VARes compared to the strongest passive baseline widens from 18.1% ($T = 96$) to 81.9% ($T = 336$). This widening "performance moat" indicates that VARes captures long-term dynamical invariants that are often discarded as noise by passive filters.

Decoupling Error Growth from Sparsity: This stability is even more evident in the imputation task. In the Energy dataset, as the mask ratio increases from 12.5% to 50%, the MSE of the strongest baseline (downsample_sample) increases by 181%, whereas VARes exhibits a significantly more controlled degradation. This disparity proves that VARes does not rely on local point-wise correlations, which vanish under high sparsity, but instead leverages a learned global manifold to reconstruct missing segments.

The consistent ability of VARes to suppress error growth under both temporal extension and spatial sparsity confirms that our framework successfully bridges the structural gap between resolutions. We have transformed the low-resolution representation from a "degraded sample" into a "robust summary" that remains predictive and reconstructible even when the available information is severely constrained. Our method can also generate data at the same resolution as the original data while maintaining

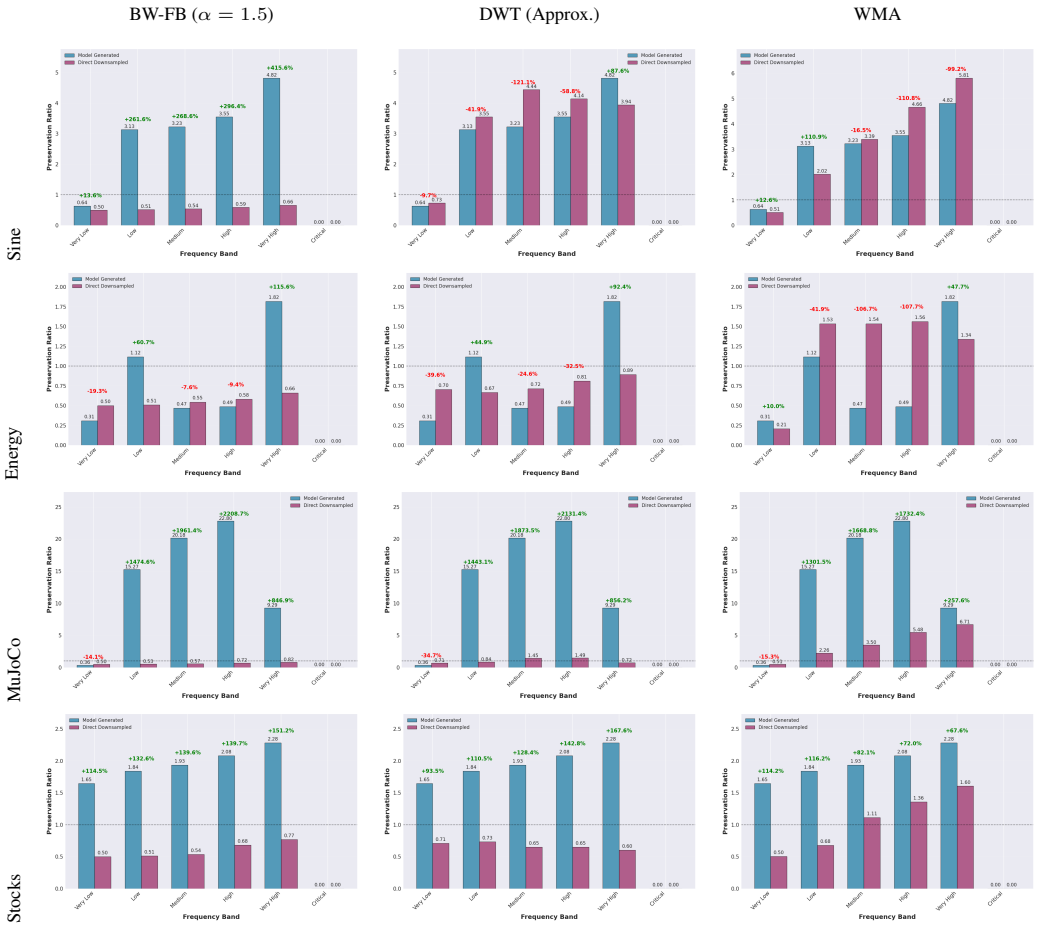

Figure 3: Comparison of three frequency-aware smoothing strategies (columns) across four datasets (rows).

competitive quality; see Appendix D for details. Furthermore, we quantify the computational overhead of VARes and discuss the trade-off between performance gains and efficiency (see Appendix J).

4.3 Qualitative Analysis: Spectral Integrity and Feature Amplification

To explain the substantial performance gains, we perform spectral fingerprinting to analyze the frequency preservation of obtained series. We compare VARes against frequency-aware baselines (e.g., anti-aliasing, wavelet decomposition, and smoothing) to prove that effective low-resolution modeling requires an active rather than a passive subtractive approach (see Figure 3).

From Passive Structural Decay to Active Feature Amplification Traditional filters (e.g., BW-FB, DWT) treat data reduction as a deterministic pruning process, prioritizing noise suppression at the expense of structural complexity. This results in an irreversible loss of high-frequency innovations—critical signals for state transitions—leading to performance collapse. In contrast, VARes treats the transition as a distribution-aware synthesis, consistently achieving frequency preservation ratios exceeding 1.0. Notably, in the MuJoCo dataset, VARes achieves a $+2208.7\%$ enhancement in high-frequency bands, actively reinforcing the latent dynamical signature for downstream models. This empirical trait correlates with performance gains; crucially, it was not an explicit design goal but discovered while analyzing spectral properties to explain such advantages. A detailed elaboration is provided in Appendix G.

Preserving the Geometric Essence of Dynamics To move beyond surface-level metrics, we conduct a rigorous audit of the underlying dynamical invariants preserved within the synthesized

Table 5: Quantitative evaluation of structural fidelity for low-resolution series ($r = 2, 4$) obtained with VARes.

Dataset	r	Pearson Correlation	DTW	PSD Cosine
Stocks	2	9.8e-01	3.5e-03	9.3e-01
	4	9.2e-01	6.3e-03	9.7e-01
Mujoco	2	7.4e-01	1.2e-02	9.6e-01
	4	5.3e-01	1.8e-02	9.0e-01
Energy	2	9.2e-01	7.5e-03	9.9e-01
	4	4.4e-01	1.5e-02	7.3e-01
Sine	2	8.6e-01	1.7e-03	1.0e+00
	4	6.2e-01	8.6e-03	9.9e-01

surrogates. Quantitative analysis (Table 5) across multiple resolutions reveals that VARes consistently maintains high Pearson correlation (up to 0.98), minimal DTW distances ($0.001 \sim 0.018$), and robust PSD cosine similarity (exceeding 0.93). These results demonstrate that VARes successfully preserves a suite of practically relevant invariants—including temporal alignment, spectral structure, and global signal dependency—which collectively encapsulate the essential geometry and evolution patterns of the original signals under resolution changes. By treating the low-resolution transition as a distribution-aware synthesis rather than a pruning process, our method ensures that the core oscillatory signatures remain invariant. A more granular analysis of these underlying dynamical indices is provided in Appendix F.

Feature Preservation as a Proxy for Performance Our analysis reveals a direct correlation between spectral preservation and predictive accuracy. Among baselines, Weighted Moving Average (WMA) is the strongest competitor because it is the only method capable of partially retaining or amplifying frequency features. However, while WMA's preservation is a passive byproduct of its weighting, VARes' enhancement is a precise, learned synthesis. This validates our core thesis: maintaining spectral richness is the primary driver of downstream success, and VARes establishes a new standard by pushing this principle to its the limit.

5 Conclusions

This paper introduces VARes, a hierarchical auto-regressive framework that redefines downsampling by shifting from destructive passive downsampling to active generative synthesis. By leveraging delay-embedded manifold mapping, VARes preserves temporal alignment, spectral structure, and global signal dependency through the downsampling process, effectively overcoming the structural degradation and information loss inherent in traditional pruning methods.

Empirical evaluations across forecasting and imputation tasks confirm that VARes consistently achieves state-of-the-art performance over traditional downsampling strategies, demonstrating superior robustness in maintaining both temporal structure and high-frequency integrity. Future work will mechanistically decouple the advantages of active synthesis to formalize the preservation of scale-invariant dynamical features, establishing a theoretical foundation for next-generation, scale-aware time series foundation models.

AI use statement

In this work, we used generative AI tools for language translation and for drafting, editing, and polishing the manuscript. We have not used generative AI tools for proposing or refining research hypotheses, designing or providing feedback on research methodology or experiments, or interpreting experimental results, and the remaining required disclosure tasks are not applicable to this work.

Additionally, we used generative AI tools for debugging and partially modifying experimental code. We have reviewed and verified all AI-assisted work, and we take responsibility for the final content of this work, including all text, claims, code, and experimental results.

REFERENCES

Ahmed Alaa, Alex James Chan, and Mihaela van der Schaar. Generative time-series modeling with fourier flows. In *International Conference on Learning Representations*, 2021.

Yihao Ang, Qiang Huang, Yifan Bao, Anthony K. H. Tung, and Zhiyong Huang. Tsgbench: Time series generation benchmark. *Proc. VLDB Endow.*, 17(3):305–318, November 2023. ISSN 2150-8097.

Disharee Bhowmick, Ranjith Ramanathan, and Sathyanarayanan N Aakur. Stats: Structure-aware temporal sequence summarization via statistical window merging. *Pattern Recognition Letters*, 2026.

Stephen Butterworth et al. On the theory of filter amplifiers. *Wireless Engineer*, 7(6):536–541, 1930.

Mouxiang Chen, Lefei Shen, Zhuo Li, Xiaoyun Joy Wang, Jianling Sun, and Chenghao Liu. VisionTS: Visual masked autoencoders are free-lunch zero-shot time series forecasters. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=5DSj3MfWrB.

Zhicheng Chen, FENG SHIBO, Zhong Zhang, Xi Xiao, Xingyu Gao, and Peilin Zhao. Sdformer: Similarity-driven discrete transformer for time series generation. *Advances in Neural Information Processing Systems*, 37:132179–132207, 2024.

Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

Ruizhi Deng, Bo Chang, Marcus A Brubaker, Greg Mori, and Andreas Lehrmann. Modeling continuous stochastic processes with dynamic normalizing flows. *Advances in Neural Information Processing Systems*, 33:7805–7815, 2020.

Abhyuday Desai, Cynthia Freeman, Zuhui Wang, and Ian Beaver. Timevae: A variational auto-encoder for multivariate time series generation. *arXiv preprint arXiv:2111.08095*, 2021.

Gal Fadlon, Idan Arbiv, Nimrod Berman, and Omri Azencot. A diffusion model for regular time series generation from irregular data with completion and masking. *arXiv preprint arXiv:2510.06699*, 2025.

II STOCHASTIC FORCING. Delay embeddings for forced systems.

Asadullah Hill Galib, Pang-Ning Tan, and Lifeng Luo. Fide: Frequency-inflated conditional diffusion model for extreme-aware time series generation. *Advances in Neural Information Processing Systems*, 37:114434–114457, 2024.

Tal Gonen, Itai Pemper, Ilan Naiman, Nimrod Berman, and Omri Azencot. Time series generation under data scarcity: A unified generative modeling approach. *arXiv preprint arXiv:2505.20446*, 2025.

Shuqi Gu, Chuyue Li, Baoyu Jing, and Kan Ren. Verbalts: Generating time series from texts. In *Forty-second International Conference on Machine Learning*.

Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with irregular and scale-invariant patterns. In *The Twelfth International Conference on Learning Representations*, 2024.

Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

Zhichen Lai, Dalin Zhang, Huan Li, Christian S Jensen, Hua Lu, and Yan Zhao. Lightcts*: Lightweight correlated time series forecasting enhanced with model distillation. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):8695–8710, 2024.

Daesoo Lee, Sara Malacarne, and Erlend Aune. Vector quantized time series generation with a bidirectional prior model. *arXiv preprint arXiv:2303.04743*, 2023.

Hao Li, Yuhao Huang, Chang Xu, Viktor Schlegel, Renhe Jiang, Riza Batista-Navarro, Goran Nenadic, and Jiang Bian. Bridge: Bootstrapping text to control time-series generation via multi-agent iterative optimization and diffusion modelling. *arXiv preprint arXiv:2503.02445*, 2025a.

Xiaomin Li, Vangelis Metsis, Huangyingrui Wang, and Anne Hee Hiong Ngu. Tts-gan: A transformer-based time-series generative adversarial network. In *International conference on artificial intelligence in medicine*, pp. 133–143. Springer, 2022.

Yuqi Li, Chuanguang Yang, Hansheng Zeng, Zeyu Dong, Zhulin An, Yongjun Xu, Yingli Tian, and Hao Wu. Frequency-aligned knowledge distillation for lightweight spatiotemporal forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7262–7272, 2025b.

Haksoo Lim, Minjung Kim, Sewon Park, and Noseong Park. Regular time-series generation using sgm. *arXiv preprint arXiv:2301.08518*, 2023.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in neural information processing systems*, 35: 9881–9893, 2022.

Ilan Naiman, N Benjamin Erichson, Pu Ren, Michael W Mahoney, and Omri Azencot. Generative modeling of regular and irregular time series data via koopman vaes. *arXiv preprint arXiv:2310.02619*, 2023.

Ilan Naiman, Nimrod Berman, Itai Pemper, Idan Arbiv, Gal Fadlon, and Omri Azencot. Utilizing image transforms and diffusion models for generative modeling of short and long time series. *Advances in Neural Information Processing Systems*, 37:121699–121730, 2024.

Sai Shankar Narasimhan, Shubhankar Agarwal, Oguzhan Akcin, Sujay Sanghavi, and Sandeep Chinchali. Time weaver: A conditional time series generation model. *arXiv preprint arXiv:2403.02682*, 2024a.

Sai Shankar Narasimhan, Shubhankar Agarwal, Litu Rout, Sanjay Shakkottai, and Sandeep P Chinchali. Constrained posterior sampling: Time series generation with hard constraints. *arXiv preprint arXiv:2410.12652*, 2024b.

Y Nie. A time series is worth 64words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Alexander Nikitin, Letizia Iannucci, and Samuel Kaski. Tsgm: a flexible framework for generative modeling of synthetic time series. *Advances in Neural Information Processing Systems*, 37: 129042–129061, 2024.

Tim Sauer, James A Yorke, and Martin Casdagli. Embedology. *Journal of statistical Physics*, 65(3): 579–616, 1991.

Ali Seyfi, Jean-Francois Rajotte, and Raymond Ng. Generating multivariate time series with common source coordinated gan (cosci-gan). *Advances in neural information processing systems*, 35: 32777–32788, 2022.

Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 1567–1577, 2022.

Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pp. 366–381. Springer, 2006.

Keyu Tian, Yi Jiang, Zehuan Yuan, Bingyue Peng, and Liwei Wang. Visual autoregressive modeling: Scalable image generation via next-scale prediction. *Advances in neural information processing systems*, 37:84839–84865, 2024.

Patara Trirat, Yooju Shin, Junhyeok Kang, Youngeun Nam, Jihye Na, Minyoung Bae, Joeun Kim, Byunghyun Kim, and Jae-Gil Lee. Universal time-series representation learning: A survey. *arXiv preprint arXiv:2401.03717*, 2024.

Aaron van den Oord, Oriol Vinyals, and koray kavukcuoglu. Neural discrete representation learning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. *arXiv preprint arXiv:2405.14616*, 2024.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. *arXiv preprint arXiv:2202.01575*, 2022.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

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.

Tianlin Xu, Li Kevin Wenliang, Michael Munn, and Beatrice Acciaio. Cot-gan: Generating sequential data via causal optimal transport. *Advances in neural information processing systems*, 33:8798–8809, 2020.

Jinsung Yoon, Daniel Jarrett, and Mihaela Van der Schaar. Time-series generative adversarial networks. *Advances in neural information processing systems*, 32, 2019.

Xinyu Yuan and Yan Qiao. Diffusion-ts: Interpretable diffusion for general time series generation. *arXiv preprint arXiv:2403.01742*, 2024.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8980–8987, 2022.

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, pp. 11106–11115, 2021.

A Experimental Details and Full Results for Section 4.1

For various downstream time series tasks, we selected multiple corresponding benchmark datasets to evaluate the effectiveness of $2\times$ downsampled time series obtained through direct downsampling of real data in long-term forecasting, missing value imputation, anomaly detection, and classification.

A.1 Datasets and Metrics

As discussed in Section 4.1, we utilize a diverse set of benchmarks to validate the impact of resolution.

- **Forecasting & Imputation**: We use ETTh1, ETTh2, ETTm1, and ETTm2. ETTh datasets are hourly sampled, while ETTm are 15-minute sampled, providing a mix of long-term and short-term dynamics.
- **Classification**: Nine datasets from the UEA archive (including EthanolConcentration, Heartbeat, JapaneseVowels, PEMS-SF, SelfRegulationSCP1, SelfRegulationSCP2, SpokenArabicDigits, UWaveGestureLibrary, and FaceDetection covering physiological signals and gesture recognition are used.
- **Anomaly Detection**: Experiments are conducted on the PSM dataset.

Metrics include Mean Squared Error (MSE) for regression-based tasks, and Accuracy, Precision, Recall, and F_1-score for classification and detection.

A.2 Reference Models

To ensure generalizability, we select state-of-the-art models for each task:

- **Forecasting**: TimesNet, TimeMixer, TimeXer.
- **Imputation**: TimesNet, Autoformer, Nonstationary Transformer.
- **Classification**: TimesNet, Autoformer, Informer.
- **Anomaly Detection**: TimesNet, Autoformer.

A.3 Comprehensive Experimental Results

The following tables provide the full per-dataset and per-model results summarized in the main text.

Table 6: Prediction performance (MSE) comparison between original-resolution data and $2\times$ downsampled data across different prediction lengths. Best results between the two resolutions are highlighted in bold.

MSE ↓	Resolution	TimesNet				TimeMixer				TimeXer			
		96	192	336	720	96	192	336	720	96	192	336	720
ETTh1	Not downsampled	**0.389**	**0.439**	**0.494**	**0.518**	**0.376**	**0.442**	0.497	**0.477**	**0.386**	**0.430**	0.472	**0.504**
	Downsampled	0.458	0.477	0.496	0.645	0.410	0.478	**0.460**	0.620	0.423	0.441	**0.440**	0.789
ETTh2	Not downsampled	**0.338**	**0.404**	0.452	**0.435**	**0.289**	**0.368**	**0.419**	**0.440**	**0.285**	**0.363**	0.417	**0.448**
	Downsampled	0.408	0.426	**0.445**	0.578	0.359	0.379	0.429	0.544	0.356	0.395	**0.409**	0.544
ETTm1	Not downsampled	**0.333**	**0.407**	**0.415**	**0.490**	**0.318**	**0.363**	**0.390**	**0.451**	**0.320**	**0.362**	**0.397**	**0.452**
	Downsampled	0.371	0.424	0.430	0.521	0.323	0.376	0.416	0.510	0.341	0.375	0.418	0.492
ETTm2	Not downsampled	**0.188**	**0.251**	**0.322**	**0.418**	**0.176**	**0.237**	**0.297**	**0.398**	**0.171**	**0.237**	**0.297**	**0.399**
	Downsampled	0.239	0.343	0.385	0.478	0.226	0.293	0.364	0.446	0.225	0.300	0.360	0.427

A.4 Parameters of the downstream task models

In our experiments, all models share a common set of hyperparameters, where e_layers and d_layers denote the numbers of encoder and decoder layers, $factor$ controls the sparsity level in attention, and d_model and d_ff represent the model and feed-forward dimensions; additional parameters such as top_k (for TimesNet) and p_hidden_dims (for Nonstationary Transformer) correspond to model-specific components. Across different models, we largely follow their standard configurations

Table 7: Comparison of directly downsampled and non-downsampled low-resolution data under different missing ratios. Imputation performance (MSE) is reported, and lower values indicate better performance.

MSE ↓	Resolution	TimesNet				Autoformer				Nonstationary Transformer			
		12.5%	25%	37.5%	50%	12.5%	25%	37.5%	50%	12.5%	25%	37.5%	50%
ETTh1	Not downsampled	**0.062**	**0.080**	**0.097**	**0.116**	**0.094**	**0.118**	**0.162**	**0.215**	**0.047**	**0.065**	**0.084**	**0.107**
	Downsampled	0.079	0.104	0.124	0.154	0.097	0.138	0.180	0.240	0.064	0.085	0.112	0.142
ETTh2	Not downsampled	**0.040**	**0.047**	**0.054**	**0.062**	**0.189**	**0.221**	**0.263**	**0.367**	**0.040**	**0.046**	**0.052**	**0.058**
	Downsampled	0.043	0.053	0.063	0.073	0.257	0.301	0.391	0.500	0.040	0.049	0.057	0.067
ETTm1	Not downsampled	**0.019**	**0.023**	**0.029**	**0.036**	**0.120**	**0.111**	0.124	**0.136**	**0.019**	**0.023**	**0.028**	**0.035**
	Downsampled	0.020	0.030	0.038	0.047	0.156	0.115	**0.123**	0.173	0.022	0.030	0.040	0.051
ETTm2	Not downsampled	**0.019**	**0.021**	**0.023**	**0.026**	**0.253**	0.298	**0.325**	**0.351**	**0.018**	**0.020**	**0.022**	**0.025**
	Downsampled	0.021	0.024	0.026	0.030	0.272	**0.297**	0.371	0.354	0.021	0.024	0.028	0.031

Table 8: Classification accuracy comparison between not downsampled and downsampled settings across different datasets. Higher values indicate better performance.

Accuracy ↑	Resolution	TimesNet	Autoformer	Informer
Ethanol Concentration	Original	0.297	0.293	0.274
	Downsampled	**0.333**	**0.318**	**0.333**
Heartbeat	Original	**0.800**	**0.722**	**0.781**
	Downsampled	0.748	0.718	0.767
Japanese Vowels	Original	**0.978**	**0.949**	**0.970**
	Downsampled	0.962	0.924	0.968
PEMS-SF	Original	**0.867**	**0.780**	**0.856**
	Downsampled	0.759	0.701	0.805
SelfRegulation SCP1	Original	**0.911**	0.536	**0.898**
	Downsampled	0.857	**0.551**	0.884
SelfRegulation SCP2	Original	**0.556**	0.528	**0.550**
	Downsampled	0.544	**0.633**	0.533
SpokenArabic Digits	Original	0.987	**0.977**	0.985
	Downsampled	**0.991**	0.975	**0.986**
UWaveGesture Library	Original	**0.881**	**0.531**	**0.859**
	Downsampled	0.794	0.450	0.788
FaceDetection	Original	0.675	0.578	0.667
	Downsampled	**0.682**	**0.579**	**0.683**

with minor dataset-specific adjustments: TimesNet adopts relatively smaller model sizes to match dataset scale, while Nonstationary Transformer and Autoformer use larger and consistent configurations to ensure stable performance. Overall, hyperparameters are kept aligned with their original implementations with minimal tuning to ensure fair comparisons. We will include these details in the revised manuscript for clarity.

We apply minor dataset-specific adjustments: TimesNet uses smaller sizes to match data scale, while Nonstationary Transformer and Autoformer use larger, consistent configurations. Overall, hyperparameters remain aligned with standard settings to ensure fair comparisons.

B Experimental Details and Full Results for Section B.1

B.1 Quantitative Analysis of Generated Low-Resolution Data Validity

Datasets. We evaluated the utility of our generated variable-resolution time series on four widely used time series generation datasets, including Stocks, Energy, Mujoco, and Sine. The first dataset,

Table 9: Anomaly detection performance on the PSM dataset under different resolution settings. Higher values indicate better performance.

Dataset	Metric ↑	Resolution	TimsNet	Autoformer
PSM	Accuracy	Not downsampled	**0.986**	**0.942**
		Downsampled_odd	0.970	0.904
		Downsampled_even	0.962	0.937
	Precision	Not downsampled	**0.985**	**0.999**
		Downsampled_odd	0.982	0.998
		Downsampled_even	0.984	0.999
	Recall	Not downsampled	**0.963**	**0.790**
		Downsampled_odd	0.909	0.654
		Downsampled_even	0.879	0.773
	F1-score	Not downsampled	**0.974**	**0.882**
		Downsampled_odd	0.944	0.791
		Downsampled_even	0.928	0.872

Table 10: Hyperparameter configurations for TimesNet across different datasets.

Dataset	**e_layers**	**d_layers**	**factor**	**d_model**	**d_ff**	**top_k**
Energy	2	1	3	32	64	5
Mujoco	2	1	3	16	32	5
Sine	2	1	3	16	32	5
Stocks	2	1	3	16	32	5

Stocks, comprises Google stock price data from 2004 to 2019, with six channels representing the highest price, lowest price, opening price, closing price, adjusted closing price, and trading volume. The second dataset, Energy, is the Appliances energy prediction dataset from the UCI Machine Learning Repository. It features 28 channels and presents a learning challenge due to its real-world complexity and explicit physical significance. The third dataset, Mujoco, is a multivariate time series simulation dataset generated by simulating the dynamics of articulated robotic systems using a physics engine. The fourth dataset, Sine, is a synthetic dataset composed of sine-wave time series. Its five channels each possess independent frequency and phase parameters.

Metrics. To validate the utility of the generated low-resolution time series in downstream tasks, we evaluate them on two typical tasks: time series forecasting and missing value imputation. We employ the widely adopted mean squared error (MSE) metric across these two categories to quantify the deviation of forecasted and imputed values relative to the real sequence.

Reference Models. We evaluate the utility of the generated low-resolution time series on downstream tasks by adopting widely used forecasting and imputation models, including TimesNet, Autoformer, and the Nonstationary Transformer, as baselines.

Experimental Setup. We compare the performance of our model's proactively generated low-resolution time series with directly downsampled time series data in long-term forecasting and missing value imputation tasks. For the original-resolution time series, after applying identical data preprocessing and sliding window sampling, we implement two distinct downsampling approaches. The first subsampling approach reduces the original resolution to half by subsampling along the sample count dimension. This preserves the internal resolution within each sample. The second subsampling approach reduces the sequence length dimension of each sample, effectively halving the resolution of the real data samples. Both subsampling methods employ interval sampling, achieving $2\times$ downsampling by retaining only odd-indexed rows. This dual-subsampling design avoids bias from dimensional or shape inconsistencies in the results.

Table 11: Hyperparameter configurations for Nonstationary Transformer across different datasets.

Dataset	e_layers	d_layers	factor	d_model	d_ff	p_hidden_dims
Energy	2	1	3	128	2048	256 256
Mujoco	2	1	3	128	2048	256 256
Sine	2	1	3	128	2048	256 256
Stocks	2	1	3	128	2048	256 256

Table 12: Hyperparameter configurations for Autoformer across different datasets.

Dataset	e_layers	d_layers	factor	d_model	d_ff
Energy	2	1	3	512	2048
Mujoco	2	1	3	512	2048
Sine	2	1	3	512	2048
Stocks	2	1	3	512	2048

For the validity verification experiments on the generated data at half the original resolution, the prediction lengths for the long-term prediction task were set to 96, 192, and 336, while the missing rate settings for the imputation task were set to 12.5%, 25%, 37.5%, and 50%. For the validity verification experiments on the generated quarter-resolution data, the prediction lengths for the long-term prediction task were set to 96 and 192, while the missing rate settings for the imputation task remained consistent with those in the half-resolution experiments.

Half-Resolution Data Used For Downstream Task Experiments. This set of experiments does not aim to compare the merits of different time series models. Instead, it treats models as a unified evaluation tool to verify whether proactively generated low-resolution time series provide a superior input representation for time series imputation and forecasting tasks compared to directly downsampling the original high-resolution sequences.

As demonstrated by the results in Tables 13 and 14, the proactively generated low-resolution time series consistently outperformed sequences obtained through direct downsampling in both time series imputation and long-term forecasting tasks across all datasets, missing rate settings, prediction time step configurations, and different backbone models. This conclusion remains stable across models with significantly different architectures, indicating that the performance improvement does not stem from specific model designs but rather from the fact that proactively generated low-resolution data preserves more critical features required for downstream tasks. Furthermore, this advantage is observed in both synthetic and real-world data, demonstrating the robust performance of proactively generated low-resolution data across varying data complexities and temporal dynamics.

The results of the imputation experiments shown in Table 13 reveal that the performance gains achieved by proactively generated low-resolution data compared to direct downsampling methods vary across different models and datasets. This variation is reasonable given differences in data characteristics, learning difficulty, and model architectural design. On the sine and Energy datasets, downstream tasks generally perform better with temporal downsampling than with spatial downsampling in most scenarios. Given the strong trendiness of the Energy dataset, and the strong periodicity of the sine dataset, reducing temporal resolution by downsampling the sample temporality for both datasets does not significantly disrupt their primary trends and periodic structures. Consequently, temporality downsampling has a relatively minor adverse impact on downstream imputation performance. Notably, on the Energy dataset, the TimesNet model achieved an average improvement of two orders of magnitude in the imputation performance of its proactively generated low-resolution data. This is because the Energy dataset itself is a real-world dataset with high feature learning difficulty, and the TimesNet model inherently possesses a frequency-sensitive design. Consequently, this model exhibits high sensitivity to the preservation levels of different frequency features in the training data.

Table 13: Mean Squared Error (MSE) comparison of different data construction strategies across datasets under varying missing rates. Missing entries correspond to experiments that have not yet finished.

		TimesNet				Autoformer				Nonstationary Transformer			
Dataset	Strategy	12.5%	25%	37.5%	50%	12.5%	25%	37.5%	50%	12.5%	25%	37.5%	50%
Stocks	**PG (Ours)**	**4.33e-06**	**4.96e-06**	**6.49e-06**	**7.43e-06**	**1.11e-02**	**8.50e-03**	**5.70e-03**	**3.80e-03**	**1.02e-04**	**9.31e-05**	**9.94e-05**	**1.00e-04**
	US-S	2.14e-04	2.37e-04	2.79e-04	3.00e-04	2.16e-02	1.66e-02	2.64e-02	3.19e-02	3.81e-04	4.66e-04	5.55e-04	6.67e-04
	US-T	2.72e-04	3.01e-04	3.46e-04	3.77e-04	1.91e-02	2.82e-02	1.78e-02	2.34e-02	4.99e-04	5.49e-04	6.16e-04	6.77e-04
	WMA	1.91e-04	1.89e-04	2.60e-04	2.68e-04	2.22e-02	2.03e-02	1.89e-02	2.38e-02	4.05e-04	4.49e-04	5.36e-04	5.92e-04
	DWT-A	3.22e-04	3.66e-04	4.53e-04	5.42e-04	3.91e-02	3.09e-02	3.61e-02	1.08e-01	7.04e-04	8.46e-04	1.00e-03	1.13e-03
	BW-FB (α=1.5)	3.73e-04	3.44e-04	3.89e-04	4.51e-04	2.21e-02	1.98e-02	1.89e-02	2.43e-02	6.58e-04	6.72e-04	7.49e-04	8.08e-04
	BW-FB (α=2.0)	3.97e-04	4.23e-04	4.76e-04	5.61e-04	2.21e-02	1.97e-02	1.89e-02	2.44e-02	7.92e-04	8.37e-04	9.18e-04	9.84e-04
Sine	**PG (Ours)**	**8.47e-06**	**1.03e-05**	**1.25e-05**	**1.58e-05**	**8.68e-03**	**7.40e-03**	**1.26e-02**	**2.63e-02**	**1.11e-04**	**1.48e-04**	**1.88e-04**	**2.01e-04**
	US-S	1.16e-04	1.22e-04	1.47e-04	2.00e-04	2.37e-01	9.61e-02	1.65e-01	1.38e-01	2.99e-03	8.11e-03	1.01e-02	7.15e-03
	US-T	1.50e-04	1.10e-04	1.14e-04	1.84e-04	3.26e-02	3.74e-02	7.83e-02	9.99e-02	3.07e-03	6.33e-03	6.95e-03	7.61e-03
	WMA	1.43e-04	1.42e-04	1.37e-04	2.42e-04	3.19e-02	4.16e-02	7.07e-02	1.03e-01	2.87e-03	6.57e-03	7.68e-03	6.88e-03
	DWT-A	2.81e-04	2.31e-04	2.70e-04	3.35e-04	4.51e-02	9.15e-02	1.42e-01	1.71e-01	5.61e-03	1.27e-02	1.43e-02	1.17e-02
	BW-FB (α=1.5)	3.19e-04	3.53e-04	4.17e-04	4.79e-04	2.21e-02	1.98e-02	1.89e-02	2.43e-02	6.58e-04	6.73e-04	7.50e-04	8.08e-04
	BW-FB (α=2.0)	1.42e-04	1.32e-04	1.81e-04	1.85e-04	3.46e-02	4.44e-02	8.41e-02	9.66e-02	2.78e-03	6.49e-03	6.63e-03	7.21e-03
Energy	**PG (Ours)**	**1.96e-05**	**3.05e-05**	**3.15e-05**	**4.64e-05**	**4.65e-03**	**4.19e-03**	**2.17e-03**	**1.70e-03**	**1.08e-04**	**1.37e-04**	**1.80e-04**	**1.90e-04**
	US-S	4.41e-03	5.36e-03	6.01e-03	6.35e-03	4.16e-02	7.40e-02	9.56e-02	7.35e-02	6.38e-03	6.56e-03	6.66e-03	6.94e-03
	US-T	3.49e-03	4.80e-03	4.85e-03	6.66e-03	2.91e-02	4.12e-02	6.95e-02	6.14e-02	6.75e-03	7.03e-03	7.19e-03	7.32e-03
	WMA	1.88e-03	3.26e-03	2.97e-03	3.64e-03	1.45e-02	4.15e-02	5.64e-02	4.92e-02	3.98e-03	4.15e-03	4.31e-03	4.51e-03
	DWT-A	3.81e-03	5.04e-03	6.27e-03	7.02e-03	4.09e-02	7.57e-02	1.48e-01	1.46e-01	7.23e-03	7.78e-03	7.90e-03	8.37e-03
	BW-FB (α=1.5)	4.05e-03	5.90e-03	7.70e-03	9.20e-03	2.99e-02	6.70e-02	1.01e-01	6.76e-02	9.81e-03	9.81e-03	1.02e-02	1.04e-02
	BW-FB (α=2.0)	5.87e-03	8.39e-03	1.10e-02	1.31e-02	3.42e-02	6.99e-02	1.07e-01	9.09e-02	1.45e-02	1.45e-02	1.46e-02	1.49e-02
Mujoco	**PG (Ours)**	**2.15e-06**	**2.48e-06**	**3.55e-06**	**5.55e-06**	**3.65e-04**	**4.54e-04**	**5.23e-04**	**6.21e-04**	**1.86e-05**	**1.81e-05**	**2.35e-05**	**2.57e-05**
	US-S	1.64e-05	2.67e-05	2.37e-05	3.80e-05	2.42e-02	3.31e-02	3.01e-02	3.73e-02	1.10e-04	1.28e-04	1.46e-04	1.69e-04
	US-T	3.07e-05	3.36e-05	4.34e-05	5.73e-05	4.46e-02	4.05e-02	3.81e-02	3.03e-02	9.88e-05	1.30e-04	1.26e-04	1.55e-04
	WMA	1.22e-04	1.27e-04	1.37e-04	1.54e-04	3.94e-02	4.57e-02	4.13e-02	2.83e-02	1.99e-04	2.21e-04	2.18e-04	1.92e-04
	DWT-A	6.86e-05	9.61e-05	1.02e-04	1.36e-04	7.42e-02	1.01e-01	9.61e-02	5.78e-02	2.49e-04	2.72e-04	2.92e-04	3.32e-04
	BW-FB (α=1.5)	3.04e-05	3.67e-05	5.17e-05	6.15e-05	3.51e-02	4.45e-02	4.00e-02	2.93e-02	1.04e-04	1.28e-04	1.28e-04	1.52e-04
	BW-FB (α=2.0)	3.65e-05	4.37e-05	5.51e-05	5.93e-05	3.42e-02	4.45e-02	4.00e-02	2.94e-02	1.04e-04	1.26e-04	1.36e-04	1.55e-04

Observing the results of the long-term prediction experiments shown in Table 14, the average improvement factor across different prediction lengths ranges from 5 to 20 times across nearly all datasets and models. Specifically, the TimesNet model achieved approximately 50-fold improvement on the Stocks dataset, while the Autoformer model reached roughly 100-fold improvement on the Mujoco dataset. Analyzing this phenomenon from the perspective of dataset frequency characteristics, it likely stems from the pronounced high-frequency features in both the Stocks and Mujoco datasets. However, both downsampling methods cause significant degradation to the high-frequency components of these datasets.

Quarter-Resolution Data Used For Downstream Task Experiments. As shown in Tables 15 and 16, our actively generated low-resolution data maintains performance advantages on both downstream tasks across the vast majority of experimental settings. This demonstrates that the feature retention capability of our actively generated low-resolution data persists even at quarter-resolution generated data.

As shown in the filling experiment results presented in Table 15, the average improvement in filling accuracy across multiple missing rates for most datasets and models ranges between 5 and 36 times. Notably, the TimesNet model achieved improvement factors ranging from 10 to 125 times. This suggests that, due to its specialized design and explicit frequency-aware processing, TimesNet is more sensitive than other models to whether frequency-specific features are preserved in the training data.

In the long-term forecasting experiments presented in Table 16, the average improvement factor across different prediction lengths ranged from 2 to 40 times across nearly all datasets and models. Similar to the half-resolution scenario, the TimesNet model continues to demonstrate significantly higher performance gains on the Stocks dataset compared to other models and datasets, achieving 150–300 times improvement. This further validates the analysis from the half-resolution experiments, highlighting the pronounced high-frequency features in the Stocks dataset and the heightened sensitivity of TimesNet to whether features from different frequency bands are preserved in the training data.

Table 14: MSE comparison of different data construction strategies across four datasets.

Dataset	Strategy	TimesNet			Autoformer			Nonstationary Transformer		
		96	192	336	96	192	336	96	192	336
Stocks	**PG (Ours)**	**9.50e-06**	**8.10e-06**	**1.12e-05**	**2.10e-03**	**4.30e-03**	**1.90e-03**	**8.30e-06**	**2.32e-05**	**4.19e-05**
	US-S (Sample)	5.37e-04	5.06e-04	4.53e-04	1.01e-02	6.60e-03	6.10e-03	7.36e-04	7.44e-04	4.78e-04
	US-T (Temporal)	5.18e-04	5.31e-04	4.11e-04	6.20e-03	5.50e-03	4.60e-03	5.75e-04	5.13e-04	3.95e-04
	WMA	4.76e-04	3.76e-04	3.17e-04	1.52e-02	5.20e-03	4.30e-03	4.06e-04	3.63e-04	2.93e-04
	DWT (Approx.)	9.14e-04	8.13e-04	6.46e-04	1.41e-02	1.09e-02	9.10e-03	8.35e-04	6.93e-04	5.50e-04
	BW-FB+HF (α=1.5)	7.60e-04	6.61e-04	5.32e-04	7.70e-03	4.90e-03	4.60e-03	7.34e-04	6.52e-04	5.36e-04
	BW-FB+HF (α=2.0)	9.45e-04	9.36e-04	6.89e-04	8.50e-03	5.10e-03	4.70e-03	9.65e-04	8.79e-04	6.94e-04
Sine	**PG (Ours)**	**9.06e-06**	**9.02e-06**	**9.01e-06**	**5.61e-02**	**3.81e-02**	**5.09e-03**	**1.79e-05**	**3.01e-05**	**4.01e-05**
	US-S (Sample)	9.31e-05	8.13e-05	3.56e-05	9.72e-01	1.21e-01	1.32e-01	1.13e-03	1.74e-03	2.05e-03
	US-T (Temporal)	2.12e-05	4.42e-05	2.61e-05	1.50e-01	1.31e-01	8.21e-02	4.16e-04	5.79e-04	1.04e-03
	WMA	3.01e-05	4.25e-05	5.65e-05	1.30e-01	1.03e-01	7.13e-02	5.00e-04	5.26e-04	5.26e-04
	DWT (Approx.)	4.29e-05	1.14e-04	6.01e-05	2.92e-01	2.24e-01	1.92e-01	9.80e-04	1.30e-03	2.15e-03
	BW-FB+HF (α=1.5)	2.68e-05	4.45e-05	8.26e-05	1.24e-01	1.20e-01	8.31e-02	4.70e-04	5.36e-04	1.16e-03
	BW-FB+HF (α=2.0)	2.82e-05	3.78e-05	7.34e-05	1.19e-01	1.16e-01	7.15e-02	4.70e-04	5.34e-04	1.16e-03
Energy	**PG (Ours)**	**1.25e-06**	**1.39e-06**	**1.79e-06**	**1.88e-03**	**1.99e-03**	**5.09e-03**	**6.50e-07**	**6.10e-07**	**7.50e-07**
	US-S (Sample)	2.03e-05	1.38e-05	1.62e-05	2.09e-02	2.09e-02	1.79e-02	6.53e-05	4.18e-05	4.88e-05
	US-T (Temporal)	7.87e-06	7.38e-06	7.87e-06	1.89e-02	1.73e-02	1.66e-02	2.23e-05	3.41e-05	1.48e-05
	WMA	5.84e-06	6.44e-06	6.71e-06	1.87e-02	1.56e-02	1.41e-02	1.96e-05	1.47e-05	1.33e-05
	DWT (Approx.)	1.37e-05	1.07e-05	1.26e-05	2.81e-02	2.90e-02	2.71e-02	3.26e-05	4.36e-05	3.20e-05
	BW-FB+HF (α=1.5)	1.02e-05	1.07e-05	1.01e-05	2.15e-02	2.03e-02	2.02e-02	2.92e-05	3.53e-05	3.81e-05
	BW-FB+HF (α=2.0)	1.41e-05	1.38e-05	1.39e-05	2.60e-02	2.57e-02	2.10e-02	5.47e-05	4.82e-05	5.24e-05
Mujoco	**PG (Ours)**	**1.30e-07**	**1.70e-07**	**2.10e-07**	**8.80e-06**	**3.41e-05**	**1.93e-05**	**5.90e-07**	**5.70e-07**	**9.20e-07**
	US-S (Sample)	9.20e-07	1.21e-06	2.31e-06	1.42e-03	1.69e-03	2.62e-03	2.57e-06	2.21e-06	2.41e-06
	US-T (Temporal)	1.21e-06	1.62e-06	1.18e-06	3.31e-03	1.02e-03	3.38e-03	1.08e-06	1.88e-06	4.81e-06
	WMA	1.09e-06	2.26e-06	7.81e-06	1.82e-03	7.15e-03	3.78e-03	7.20e-07	1.97e-06	5.09e-06
	DWT (Approx.)	1.91e-06	4.26e-06	2.45e-05	4.54e-03	3.81e-03	7.65e-03	2.59e-06	4.41e-06	1.11e-05
	BW-FB+HF (α=1.5)	1.16e-06	2.29e-06	9.57e-06	1.71e-03	3.21e-03	3.07e-03	6.30e-07	2.35e-06	6.16e-06
	BW-FB+HF (α=2.0)	1.52e-06	2.14e-06	9.57e-06	1.60e-03	3.22e-03	3.78e-03	6.10e-07	2.36e-06	6.41e-06

Table 15: Performance comparison of different Quarter-resolution data construction strategies across datasets, missing rates, and backbone models.

MSE ↓	Data Construction Strategies	TimesNet				Autoformer				Nonstationary Transformer			
		12.50%	25%	37.50%	50%	12.50%	25%	37.50%	50%	12.50%	25%	37.50%	50%
Stocks	proactively_generated	**2.80e-06**	**3.50e-06**	**4.80e-06**	**5.30e-06**	**3.64e-03**	**2.62e-03**	**3.22e-03**	**3.32e-03**	**2.74e-05**	**4.09e-05**	**6.01e-05**	**8.39e-05**
	downsample_sample	2.64e-04	3.31e-04	3.70e-04	4.07e-04	2.25e-02	2.56e-02	1.72e-02	3.11e-02	4.70e-04	6.22e-04	7.25e-04	8.62e-04
	downsample_resolution	3.79e-04	3.36e-04	3.96e-04	4.64e-04	2.08e-02	2.11e-02	2.24e-02	2.70e-02	5.94e-04	6.13e-04	7.05e-04	7.67e-04
Sine	proactively_generated	**1.38e-05**	**1.92e-05**	**3.16e-05**	**3.68e-05**	**2.22e-02**	**1.17e-02**	**2.21e-02**	**5.31e-02**	**2.62e-04**	**2.94e-04**	**3.29e-04**	**4.12e-04**
	downsample_sample	1.56e-04	2.13e-04	2.95e-04	4.06e-04	2.05e-01	1.62e-01	1.20e-01	8.75e-02	3.15e-03	8.77e-03	1.62e-02	1.66e-02
	downsample_resolution	7.65e-05	1.42e-04	1.94e-04	3.52e-04	1.72e-02	2.52e-02	3.55e-02	3.39e-02	4.16e-03	1.01e-02	1.39e-02	1.26e-02
Energy	proactively_generated	**2.69e-05**	**3.52e-05**	**5.81e-05**	**8.97e-05**	**7.55e-03**	**8.24e-03**	**5.65e-03**	**5.54e-03**	**2.99e-04**	**3.92e-04**	**4.61e-04**	**5.27e-04**
	downsample_sample	5.35e-03	6.16e-03	6.42e-03	6.60e-03	6.13e-02	9.43e-02	1.29e-01	8.15e-02	6.66e-03	7.01e-03	7.04e-03	7.23e-03
	downsample_resolution	5.44e-03	6.55e-03	6.91e-03	7.53e-03	2.59e-02	3.26e-02	4.89e-02	4.67e-02	7.38e-03	7.64e-03	7.83e-03	8.04e-03
Mujoco	proactively_generated	**1.60e-06**	**2.40e-06**	**2.60e-06**	**4.20e-06**	**4.03e-04**	**5.66e-04**	**8.62e-04**	**1.12e-03**	**3.15e-05**	**3.55e-05**	**3.52e-05**	**3.82e-05**
	downsample_sample	2.18e-05	2.67e-05	3.25e-05	3.75e-05	2.10e-02	2.01e-02	3.27e-02	3.47e-02	1.32e-04	1.89e-04	1.96e-04	2.18e-04
	downsample_resolution	7.66e-05	8.22e-05	1.13e-04	1.58e-04	2.97e-02	3.50e-02	3.72e-02	2.99e-02	1.84e-04	2.35e-04	2.57e-04	3.09e-04

B.2 HYPERPARAMETERS OF THE VARES MODEL

We present tables(Table 17,Table 18) showing the hyperparameter selections for two-stage training across four datasets.

Table 16: Performance comparison of different Quarter-resolution data construction strategies under varying prediction lengths.

MSE ↓	Data Construction Strategies	TimesNet		Autoformer		Nonstationary Transformer	
		96	192	96	192	96	192
Stocks	proactively_generated	**2.38e-06**	**3.48e-06**	**2.17e-03**	**1.38e-03**	**6.42e-06**	**1.65e-05**
	downsample_sample	1.06e-03	8.76e-04	1.20e-02	5.68e-03	9.28e-04	9.99e-04
	downsample_resolution	5.62e-04	4.08e-04	5.43e-03	5.13e-03	4.18e-04	3.39e-04
Sine	proactively_generated	**6.99e-06**	**7.38e-06**	**2.08e-02**	**7.77e-02**	**1.32e-04**	**1.52e-04**
	downsample_sample	2.67e-04	2.64e-04	1.33e-01	1.24e-01	1.78e-03	3.99e-03
	downsample_resolution	4.58e-05	6.21e-05	1.09e-01	1.09e-01	1.64e-03	9.49e-04
Energy	proactively_generated	**1.81e-06**	**2.25e-06**	**3.79e-04**	**5.86e-04**	**6.47e-06**	**2.03e-06**
	downsample_sample	3.74e-05	5.64e-05	2.00e-02	2.03e-02	1.38e-04	2.06e-04
	downsample_resolution	5.51e-06	5.87e-06	1.57e-02	2.20e-02	3.56e-05	2.58e-05
Mujoco	proactively_generated	**9.00e-08**	**2.60e-07**	**3.83e-05**	**2.52e-04**	**9.20e-07**	5.78e-06
	downsample_sample	1.72e-06	2.21e-06	3.16e-03	2.32e-03	2.69e-06	**3.57e-06**
	downsample_resolution	3.82e-06	1.33e-05	2.35e-03	4.44e-03	1.62e-06	1.20e-05

Table 17: Hyperparameter selections for Multi-scale VQVAE training across four datasets.

Dataset	**vocab_size**	**z_channels**	**ch**	**ch_mult**	**batch_size**	**patch_nums**
Energy	4096	32	16	(1, 1, 2, 2, 4)	36	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)
Mujoco	4096	32	16	(1, 1, 2, 2, 4)	128	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)
Sine	4096	32	16	(1, 1, 2, 2, 4)	128	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)
Stocks	4096	32	16	(1, 1, 2, 2, 4)	256	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)

Table 18: Hyperparameter configurations for the generative model across four datasets.

Dataset	**embed_dim**	**attention heads**	**batch_size**	**Model Size (M)**	**patch_nums**
Energy	512	8	36	41.32	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)
Mujoco	512	8	128	41.18	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)
Sine	512	8	128	41.32	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)
Stocks	256	4	256	6.17	(1, 2, 3, 4, 5, 6, 8, 10, 13, 16)

C EXPERIMENTAL DETAILS AND FULL RESULTS FOR SECTION 4.3

To avoid overly weak baseline methods, we considered multiple frequency-aware downsampling strategies, including anti-aliasing filtering, local polynomial smoothing, and wavelet-based decomposition. After enhancing the frequency intensity of the original resolution data using the aforementioned methods, we proceeded with downsampling. All approaches generated signals with a temporal resolution equal to half that of the original sequence.

To provide a rigorous explanation for the order-of-magnitude gains observed in Section 5.1, we analyze the frequency preservation ratio of synthesized series across diverse datasets. This spectral "fingerprinting" reveals that the transition to lower resolutions must be a generative rather than a subtractive process to maintain dynamical integrity.

From Passive Structural Decay to Active Feature Amplification A fundamental divergence between `VARes` and traditional methods lies in their treatment of spectral entropy. With the exception of WMA, most conventional downsampling and anti-aliasing filters (e.g., BW-FB with $\alpha = 1.5$, DWT) function as deterministic operators that prioritize noise suppression at the cost of structural complexity. In the MuJoCo and Energy datasets, these passive filters exhibit a systemic attenuation of high-frequency components, with preservation ratios frequently dropping significantly below the original resolution baseline. This irreversible loss of high-frequency "innovations" explains the performance collapse in forecasting, as these bands often carry the critical signals for state transitions.

Conversely, `VARes` consistently exhibits preservation ratios exceeding 1.0 across nearly all bands—notably achieving a +2208.7% enhancement in the High band for MuJoCo. This demonstrates that our method does not merely "protect" the signal but actively synthesizes a reinforced representation, ensuring that the latent dynamical signature remains salient for downstream backbones even under reduced granularity.

Bridging the Structural Gap via Temporal-Spatial Co-embedding The ability to amplify frequency components without introducing artifacts is rooted in our temporal-spatial co-embedding mechanism. While passive methods treat data reduction as a 1D signal-processing problem, `VARes` re-encodes temporal oscillations into a structurally consistent 2D latent lattice.

By aligning high-frequency temporal fluctuations with spatial lattices, `VARes` transforms volatile oscillations into stable structural invariants. In complex mechanical systems like MuJoCo, joint interactions are no longer "pruned" as noise; instead, they are distilled into the low-resolution surrogate as task-critical features. This re-encoding allows the synthesized series to act as a "robust summary" that maintains the energy and phase information necessary for long-term consistency.

Feature Preservation as a Proxy for Performance A pivotal finding in our qualitative study is the unique behavior of WMA, which provides an empirical bridge between spectral characteristics and predictive accuracy. Unlike deterministic filters, WMA demonstrates a relative capacity to retain or even amplify certain frequency features, which directly translates to its status as the strongest baseline in long-horizon forecasting and high-sparsity imputation.

Crucially, while our method shares this characteristic of feature amplification, it does so with significantly higher magnitude and precision. While WMA's preservation is a passive byproduct of its weighting scheme, `VARes`' enhancement is a distribution-aware synthesis. The fact that WMA—the only baseline to maintain high-frequency energy—also happens to be the only baseline that competes at all, validates our core thesis: maintaining spectral richness is the primary driver of downstream success. `VARes` pushes this principle to its logical limit, achieving a spectral density that no passive averaging can replicate, thus establishing a new generative standard for variable-resolution modeling.

Table 19: Quantitative comparison across different metrics and datasets. The best and second-best results are highlighted in red and blue, respectively.

Metric	Method	Stocks	Energy	Sine	Mujoco
DS↓	TimeGAN	0.106	0.086	0.315	0.014
	TimeVAE	0.039	0.057	0.149	0.015
	Diffusion-TS	0.044	0.059	0.088	0.022
	ImagenTime	0.029	0.069	0.015	0.009
	ImagenFew_seq24	0.027	0.377	0.015	0.014
	Our	0.005	0.056	0.028	0.007
PS↓	TimeGAN	0.015	0.037	0.154	0.009
	TimeVAE	0.013	0.035	0.014	0.012
	Diffusion-TS	0.010	0.030	0.012	0.005
	ImagenTime	0.009	0.031	0.009	0.005
	ImagenFew_seq24	0.037	0.251	0.096	0.036
	Our	0.009	0.034	0.008	0.003
C-FID↓	TimeGAN	3.285	7.400	24.934	2.324
	TimeVAE	0.390	1.314	7.824	0.125
	Diffusion-TS	0.449	2.102	3.858	0.407
	ImagenTime	0.239	0.181	0.221	0.039
	ImagenFew_seq24	0.069	0.517	0.084	0.072
	Our	0.186	0.414	0.219	0.015

D Experimental Details and Full Results for of Our Original-Resolution TS

We benchmarked our method against state-of-the-art GANYoon et al. (2019), VAEDesai et al. (2021), and diffusion-based modelsYuan & Qiao (2024); Naiman et al. (2024); Gonen et al. (2025) across four datasets (Table 19). Further descriptions of the model selected and experimental details are provided in Appendix D. Results show our model is highly competitive in fixed-resolution generation, achieving the best Discriminative Scores (DS) on three datasets—most notably on Stocks (0.005 vs. 0.027). It also leads in Predictive Scores (PS) on Mujoco and Sine, indicating superior preservation of temporal dynamics. While specialized models like ImagenTime lead in specific metrics (e.g., C-FID on Energy), our approach consistently maintains a 'Top-2' presence, proving its reliability in capturing complex temporal patterns across resolutions.

To evaluate the original-resolution generation quality of our proposed method, we conduct a comprehensive quantitative comparison against several state-of-the-art (SOTA) generative models, including GAN-based (*TimeGAN*), VAE-based (*TimeVAE*), and recent diffusion-based benchmarks (*Diffusion-TS*, *ImagenTime*, and *ImagenFew*). Table 5 summarizes the results across four diverse real-world datasets using a variety of statistical and distributional metrics.

As shown in Table 19, Our model maintains highly competitive, and in several instances superior, performance on fixed-resolution time series generation compared to recent state-of-the-art baselines. Notably, in terms of the *Discriminative Score (DS)*, our model achieves the best performance on three of the four datasets. Specifically, on the Stocks dataset, our model reaches a DS of 0.005, representing a substantial improvement over the next-best baseline, *ImagenFew* (0.027). This suggests a particular proficiency in capturing the complex, non-linear dependencies inherent in financial manifolds. Furthermore, our model yields leading *Predictive Scores (PS)* on Mujoco and Sine, indicating that the generated sequences preserve the underlying temporal dynamics essential for downstream predictive tasks. While specialized models like *ImagenTime* exhibit strengths in specific metrics such as C-FID on the Energy dataset, our approach maintains a more consistent "Top-2" presence across the majority of distributional evaluations, confirming its efficacy as a reliable generator capable of capturing complex temporal dynamics across resolutions.

E REMARK

To adapt the VAR from a general-purpose model to a specialized one, we first removed the adaptive normalization parameter adjustment module within the model that injects conditional embeddings for different labels. Next, to model the codebook index of tokens at various scales within the latent space constructed by the multi-scale VQVAE, we computed the KL divergence between the probability density distribution of codebook indexes for all tokens output by VAR and the true codebook of all tokens as the generator's loss function. Finally, to address the numerical range discrepancy between image and temporal data during input training and post-generation (typically $[0, 1]$ for temporal domains versus $[0, 256]$ for images), we modified the data range processing of the decoder's output heads. These adjustments for domain-specific nuances enable features learned in earlier stages to be effectively transferred for improved downstream generation quality.

F THEORETICAL CLARIFICATION AND QUANTITATIVE ANALYSIS OF DYNAMICAL INVARIANTS

We first clarify that the delay embedding used in our work is conceptually consistent with classical delay embedding in dynamical systems, but serves a different purpose. In the theoretical setting (e.g., Takens-style embedding), delay embedding aims to guarantee a topologically faithful reconstruction of the underlying phase space under strict assumptions such as smooth dynamics and sufficiently dense observations. In contrast, our objective is not to establish topological conjugacy, but to construct a representation that transforms 1D time series into a 2D form suitable for learning, while preserving practically relevant temporal and structural characteristics. In this sense, delay embedding in our framework should be understood as a structure-preserving transformation rather than a strict phase-space reconstruction tool.

We acknowledge that classical embedding results rely on conditions such as smooth dynamics and sufficient sampling density. At the same time, extensions in the literature (e.g., Sauer et al. (1991); FORCING) suggest that these conditions can be relaxed in practical settings, and that delay-coordinate mappings can still preserve meaningful structural properties under mild assumptions. In our case, the time series considered exhibit sufficient temporal continuity at the given sampling rates, along with clear temporal dependencies (e.g., autocorrelation and spectral structure), which makes delay-based representations empirically stable under moderate resolution changes.

Importantly, our method does not aim to establish strict topological reconstruction guarantees. Instead, delay embedding is used as a structured transformation that preserves salient temporal organization—such as ordering, dependency, and frequency patterns—while enabling effective learning in a 2D representation space. We will revise the manuscript to clarify this distinction and better position our claims in line with both theoretical insights and practical usage.

To make this notion concrete, we explicitly characterize the "dynamical invariants" preserved by our method through quantitative analysis across resolutions. First, global signal similarity is largely maintained, as evidenced by high Pearson correlation (e.g., 0.98/0.92 on Stocks), indicating that overall amplitude structure is well preserved at moderate downsampling ratios. Second, temporal alignment is consistently retained, with all DTW distances remaining extremely low (0.001–0.018), demonstrating that the generated low-resolution sequences preserve the temporal ordering and shape of the original signals even under elastic distortions. Third, and most importantly, the spectral structure is strongly preserved: PSD cosine similarity remains high across datasets (e.g., 0.93–0.99 for Stocks/Energy/Sine), indicating that dominant frequencies and energy distributions are largely invariant under resolution changes. These results suggest that the core oscillatory patterns and frequency-domain characteristics—key aspects of dynamical behavior—are maintained.

Rather than asserting strict preservation of dynamical invariants in the formal sense of dynamical systems theory, our method preserves a set of practically relevant invariants, including temporal alignment, spectral structure, and global signal dependency, which together capture the essential geometry and evolution patterns of the time series under resolution changes. We will incorporate this clarification and an expanded discussion of the underlying assumptions and limitations in the revised version to better align our claims with both theory and empirical evidence.

Table 20: Distributional statistics (std, IQR, P90, P95, max, median) of MSE for different data construction strategies on the Energy dataset.

Dataset	Horizon	Strategy	std	IQR	P90	P95	max	median
Energy	96	PG (Ours)	4.0E-06	7.5E-07	3.0E-06	5.3E-06	3.8E-04	3.2E-07
		US-S (Sample)	6.6E-05	1.2E-05	6.3E-05	8.7E-05	1.8E-03	3.5E-06
		US-T (Temporal)	1.9E-05	5.7E-06	2.4E-05	3.5E-05	1.2E-03	2.3E-06
		WMA	1.4E-05	4.7E-06	1.6E-05	2.3E-05	4.3E-04	2.0E-06
		DWT (Approx.)	3.0E-05	1.2E-05	4.0E-05	5.6E-05	1.2E-03	5.0E-06
		BW-FB+HF ($\alpha = 1.5$)	2.9E-05	6.7E-06	3.1E-05	4.6E-05	2.0E-03	2.6E-06
		BW-FB+HF ($\alpha = 2.0$)	4.2E-05	8.7E-06	4.4E-05	7.1E-05	2.8E-03	3.4E-06
	192	PG (Ours)	3.3E-06	8.9E-07	4.0E-06	6.7E-06	2.4E-04	3.4E-07
		US-S (Sample)	3.4E-05	9.4E-06	4.2E-05	5.5E-05	7.4E-04	4.6E-06
		US-T (Temporal)	1.4E-05	5.9E-06	2.5E-05	3.5E-05	5.4E-04	2.6E-06
		WMA	1.0E-05	5.7E-06	2.0E-05	2.7E-05	2.3E-04	2.7E-06
		DWT (Approx.)	1.8E-05	9.4E-06	3.2E-05	4.5E-05	4.8E-04	4.5E-06
		BW-FB+HF ($\alpha = 1.5$)	1.9E-05	8.1E-06	3.4E-05	5.2E-05	8.2E-04	3.8E-06
		BW-FB+HF ($\alpha = 2.0$)	2.6E-05	9.9E-06	4.1E-05	7.4E-05	1.2E-03	4.6E-06
	336	PG (Ours)	3.1E-06	1.0E-06	5.5E-06	8.6E-06	1.5E-04	3.9E-07
		US-S (Sample)	3.3E-05	1.0E-05	6.0E-05	7.5E-05	4.5E-04	4.5E-06
		US-T (Temporal)	1.4E-05	6.0E-06	2.5E-05	4.1E-05	3.0E-04	2.7E-06
		WMA	8.4E-06	6.0E-06	1.9E-05	2.5E-05	1.4E-04	3.4E-06
		DWT (Approx.)	1.6E-05	1.2E-05	3.8E-05	5.2E-05	2.8E-04	6.0E-06
		BW-FB+HF ($\alpha = 1.5$)	1.6E-05	8.0E-06	3.2E-05	4.8E-05	4.2E-04	3.9E-06
		BW-FB+HF ($\alpha = 2.0$)	2.4E-05	1.0E-05	4.1E-05	7.6E-05	6.4E-04	4.6E-06

G Analysis and Interpretation of High-Frequency Amplification

We clarify that the reported high-frequency amplification (e.g., 2208.7% on MuJoCo) should not be interpreted as arbitrary or uncontrolled signal distortion; instead, it is a byproduct of the generative process. Importantly, this phenomenon was not an explicit design objective of our model. Rather, it was observed during post-hoc analysis when examining the frequency characteristics of the generated variable-resolution time series to better understand their empirical advantages.

In this sense, the amplification reflects how the model redistributes representational emphasis across frequency components, potentially highlighting patterns that are suppressed by deterministic downsampling, rather than introducing physically implausible signals.

It is also important to note that our work is exploratory in nature. The downstream tasks serve to evaluate the utility of the generated variable-resolution time series. From this perspective, the observed behavior can be understood as an empirical characteristic of the generated data that is associated with improved downstream performance. Nevertheless, we fully agree that in applications where data fidelity is critical, additional constraints would be necessary. Incorporating frequency regularization or domain-specific priors to control such amplification represents an important direction for future work.

H Evaluation Scope: Stochasticity and Stability

We extend the evaluation from mean-based metrics (e.g., MSE) to distributional statistics, including standard deviation, IQR, tail quantiles ($P90/P95$), and worst-case error(Table 20,Table 21,Table 22,Table 23).

Across datasets with diverse characteristics, we observe a consistent pattern that stochastic generation improves the shape of the error distribution rather than only its mean. On stationary data (Sines),

Table 21: Distributional statistics (std, IQR, P90, P95, max, median) of MSE for different data construction strategies on the Mujoco dataset.

Dataset	Horizon	Strategy	std	IQR	P90	P95	max	median
Mujoco	96	PG (Ours)	1.5E-07	1.1E-07	2.7E-07	3.5E-07	4.5E-06	9.2E-08
		US-S (Sample)	1.2E-05	1.2E-07	3.7E-07	8.8E-07	7.8E-04	4.4E-08
		US-T (Temporal)	1.2E-05	1.1E-07	4.0E-07	1.8E-06	4.4E-04	3.9E-08
		WMA	1.3E-05	8.2E-08	3.9E-07	1.9E-06	5.3E-04	2.9E-08
		DWT (Approx.)	2.0E-05	1.1E-07	5.3E-07	2.9E-06	8.7E-04	4.1E-08
		BW-FB+HF ($\alpha = 1.5$)	1.1E-05	1.1E-07	4.6E-07	1.8E-06	4.8E-04	3.9E-08
		BW-FB+HF ($\alpha = 2.0$)	1.3E-05	1.5E-07	5.8E-07	2.3E-06	4.9E-04	5.0E-08
	192	PG (Ours)	1.6E-07	1.4E-07	3.2E-07	4.1E-07	4.0E-06	1.3E-07
		US-S (Sample)	1.5E-05	2.6E-07	8.9E-07	2.5E-06	9.4E-04	1.1E-07
		US-T (Temporal)	1.3E-05	3.1E-07	2.6E-06	6.2E-06	8.9E-04	4.9E-08
		WMA	1.2E-05	3.8E-07	3.8E-06	1.0E-05	4.9E-04	5.9E-08
		DWT (Approx.)	2.2E-05	9.0E-07	6.9E-06	1.8E-05	9.5E-04	1.6E-07
		BW-FB+HF ($\alpha = 1.5$)	1.1E-05	4.0E-07	3.7E-06	9.5E-06	4.7E-04	5.7E-08
		BW-FB+HF ($\alpha = 2.0$)	1.1E-05	3.8E-07	3.6E-06	9.0E-06	4.6E-04	5.8E-08
	336	PG (Ours)	2.8E-07	1.8E-07	4.1E-07	5.1E-07	1.2E-05	1.6E-07
		US-S (Sample)	1.1E-05	4.1E-07	2.8E-06	8.4E-06	2.9E-04	1.2E-07
		US-T (Temporal)	2.6E-05	8.8E-06	2.9E-05	5.2E-05	4.9E-04	2.6E-06
		WMA	1.7E-05	6.6E-06	2.1E-05	3.6E-05	2.9E-04	1.9E-06
		DWT (Approx.)	5.4E-05	2.0E-05	6.7E-05	1.2E-04	8.8E-04	5.9E-06
		BW-FB+HF ($\alpha = 1.5$)	1.9E-05	7.3E-06	2.4E-05	4.2E-05	3.3E-04	2.1E-06
		BW-FB+HF ($\alpha = 2.0$)	2.0E-05	7.4E-06	2.4E-05	4.3E-05	3.5E-04	2.0E-06

PG achieves comparable performance to deterministic baselines, indicating that stochasticity does not harm stability. On weakly non-stationary data (Mujoco), PG maintains lower variance. On more challenging non-stationary settings (Energy), PG significantly reduces dispersion and improves robustness under distribution shifts. On noisy signals (ECG), PG effectively suppresses tail errors. Finally, under strong distribution shifts (Stocks), PG yields the most pronounced gains, dramatically reducing variance and extreme errors, transforming heavy-tailed error distributions into more concentrated ones.

These results suggest a unified conclusion: stochastic generation consistently contracts the error distribution and improves robustness as data complexity increases, while remaining competitive in simpler regimes.

We also note that in a few cases on simpler or highly regular datasets, PG may exhibit slightly higher variance than deterministic baselines. This is expected, as deterministic methods can exploit strong regularity through fixed mappings, whereas stochastic generation introduces controlled variability. However, this trade-off is generally modest, and in most non-stationary and noisy scenarios, PG tends to provide more stable error distributions, while deterministic methods are more prone to larger variance and heavier tails.

Overall, our results demonstrate that stochasticity does not introduce instability; instead, it provides an implicit ensemble effect that reduces variance, suppresses extreme errors, and leads to more reliable predictions under distribution shifts.

I Extended Empirical Validation on Non-Stationarity and Distribution Shift

In our original experimental setup, Stocks and Energy datasets were selected as they are representative of real-world time series exhibiting non-stationarity and distribution shift. Specifically, Stocks capture

Table 22: Distributional statistics (std, IQR, P90, P95, max, median) of MSE for different data construction strategies on the Sines dataset.

Dataset	Horizon	Strategy	std	IQR	P90	P95	max	median
Sines	96	PG (Ours)	3.0E-05	4.4E-06	1.4E-05	2.5E-05	1.1E-03	3.8E-06
		US-S (Sample)	9.8E-05	5.3E-05	1.7E-04	2.3E-04	9.5E-04	6.2E-05
		US-T (Temporal)	2.0E-05	9.6E-06	4.0E-05	5.9E-05	2.7E-04	1.4E-05
		WMA	2.7E-05	1.6E-05	6.0E-05	8.9E-05	3.1E-04	2.1E-05
		DWT (Approx.)	5.3E-05	1.6E-05	6.2E-05	1.1E-04	5.8E-04	3.0E-05
		BW-FB+HF ($\alpha = 1.5$)	2.8E-05	1.0E-05	4.0E-05	8.0E-05	2.8E-04	1.9E-05
		BW-FB+HF ($\alpha = 2.0$)	4.3E-05	1.5E-05	5.5E-05	1.0E-04	9.9E-04	2.3E-05
	192	PG (Ours)	2.6E-05	3.9E-06	1.4E-05	2.6E-05	1.1E-03	3.5E-06
		US-S (Sample)	7.6E-05	2.4E-05	9.8E-05	1.6E-04	8.7E-04	4.3E-05
		US-T (Temporal)	5.0E-05	1.8E-05	7.2E-05	1.2E-04	9.5E-04	2.9E-05
		WMA	1.3E-04	3.1E-05	9.4E-05	1.4E-04	3.4E-03	3.3E-05
		DWT (Approx.)	2.3E-04	3.3E-05	1.5E-04	2.4E-04	1.5E-02	7.6E-05
		BW-FB+HF ($\alpha = 1.5$)	5.4E-05	1.7E-05	6.5E-05	8.7E-05	1.5E-03	2.4E-05
		BW-FB+HF ($\alpha = 2.0$)	3.5E-05	1.8E-05	6.4E-05	8.6E-05	4.5E-04	2.5E-05
	336	PG (Ours)	8.5E-06	4.2E-06	1.4E-05	2.4E-05	5.3E-04	3.9E-06
		US-S (Sample)	1.5E-04	6.7E-05	2.9E-04	4.8E-04	2.0E-03	1.5E-04
		US-T (Temporal)	2.0E-05	1.4E-05	4.9E-05	6.0E-05	4.0E-04	2.3E-05
		WMA	4.1E-05	1.9E-05	8.1E-05	9.9E-05	8.2E-04	5.6E-05
		DWT (Approx.)	1.3E-04	3.5E-05	1.1E-04	1.5E-04	6.4E-03	5.1E-05
		BW-FB+HF ($\alpha = 1.5$)	1.1E-04	3.4E-05	1.0E-04	1.7E-04	2.9E-03	6.2E-05
		BW-FB+HF ($\alpha = 2.0$)	6.2E-05	3.9E-05	9.8E-05	1.4E-04	1.7E-03	4.8E-05

strong distributional changes such as regime shifts and volatility dynamics, while Energy reflects structured non-stationarity driven by seasonal and behavioral patterns. To further strengthen our empirical validation, we additionally incorporate the ECG dataset. ECG signals also exhibit typical real-world non-stationary characteristics, including temporal variability and inter-subject distribution differences. Therefore, the combined use of these datasets provides a more comprehensive evaluation under realistic non-stationary and distribution-shift conditions.

We have expanded our experiments to include an electrocardiogram (ECG) dataset and present the results. As shown by the experimental results(Table 24,Table 25), the variable-resolution data generated by `VARes` continues to demonstrate a significant performance advantage in downstream tasks on this newly added ECG dataset from the healthcare domain.

To empirically validate the robustness of the delay-embedded manifold mapping under non-stationary conditions, we extend our evaluation beyond mean-based metrics (e.g., MSE) to distributional statistics, including standard deviation, interquartile range (IQR), tail quantiles ($P90/P95$), and worst-case error(Table 20,Table 21,Table 22,Table 23).

On strongly non-stationary datasets such as Energy, Stocks, and ECG, PG consistently demonstrates significantly reduced variance (up to $100\times$ lower std), compressed interquartile ranges, and substantially lighter tails compared to baselines. Notably, on the Stocks dataset, which exhibits severe distribution shifts and heavy-tailed behavior, baseline methods show large discrepancies between median and $P95$, whereas PG maintains a tightly concentrated error distribution. Furthermore, as the forecasting horizon increases (thus amplifying distribution shift), PG maintains stable variance and bounded tail errors, while competing methods degrade substantially.

These results suggest that the proposed delay-embedded manifold mapping preserves local topological structures even under non-stationary dynamics, enabling stable predictions. Combined with channel-independent projections and hierarchical autoregressive modeling, the method effectively mitigates

Table 23: Distributional statistics (std, IQR, P90, P95, max, median) of MSE for different data construction strategies on the Stocks dataset.

Dataset	Horizon	Strategy	std	IQR	P90	P95	max	median
Stocks	96	PG (Ours)	1.2E-05	7.0E-06	1.9E-05	2.8E-05	2.5E-04	6.4E-06
		US-S (Sample)	2.3E-03	7.2E-05	7.1E-04	1.6E-03	2.4E-02	6.5E-05
		US-T (Temporal)	2.0E-03	8.0E-05	7.9E-04	1.9E-03	1.6E-02	6.1E-05
		WMA	1.7E-03	9.3E-05	6.4E-04	1.9E-03	1.5E-02	8.3E-05
		DWT (Approx.)	3.4E-03	1.7E-04	1.4E-03	4.1E-03	3.2E-02	1.4E-04
		BW-FB+HF ($\alpha = 1.5$)	3.0E-03	9.9E-05	1.1E-03	2.7E-03	2.6E-02	8.4E-05
		BW-FB+HF ($\alpha = 2.0$)	3.8E-03	1.2E-04	1.3E-03	3.4E-03	3.2E-02	9.1E-05
	192	PG (Ours)	9.0E-06	6.0E-06	1.5E-05	2.2E-05	1.1E-04	5.7E-06
		US-S (Sample)	2.0E-03	7.6E-05	7.8E-04	1.5E-03	1.8E-02	8.5E-05
		US-T (Temporal)	2.0E-03	1.3E-04	8.0E-04	2.6E-03	1.7E-02	7.9E-05
		WMA	1.3E-03	9.4E-05	6.2E-04	2.0E-03	1.2E-02	6.2E-05
		DWT (Approx.)	2.8E-03	2.3E-04	1.5E-03	4.2E-03	2.5E-02	1.4E-04
		BW-FB+HF ($\alpha = 1.5$)	2.6E-03	1.1E-04	8.9E-04	3.2E-03	2.1E-02	7.2E-05
		BW-FB+HF ($\alpha = 2.0$)	3.7E-03	1.7E-04	1.2E-03	4.4E-03	3.1E-02	9.4E-05
	336	PG (Ours)	9.4E-06	7.9E-06	2.3E-05	2.8E-05	1.2E-04	8.3E-06
		US-S (Sample)	1.6E-03	1.2E-04	8.0E-04	1.5E-03	1.7E-02	1.0E-04
		US-T (Temporal)	1.4E-03	1.1E-04	7.7E-04	1.8E-03	1.1E-02	6.4E-05
		WMA	9.6E-04	1.0E-04	5.5E-04	1.4E-03	8.0E-03	6.6E-05
		DWT (Approx.)	1.9E-03	2.2E-04	1.3E-03	2.8E-03	1.6E-02	1.3E-04
		BW-FB+HF ($\alpha = 1.5$)	1.8E-03	1.2E-04	9.1E-04	2.3E-03	1.5E-02	7.7E-05
		BW-FB+HF ($\alpha = 2.0$)	2.4E-03	1.3E-04	1.2E-03	3.0E-03	2.0E-02	8.5E-05

Table 24: Mean squared error (MSE) comparison of different data downsampling strategies on the ECG dataset across three forecasting horizons (96, 192, 336).

Dataset	Strategy	96	192	336
ECG	PG (Ours)	8.99E-04	1.26E-03	1.16E-03
	US-S (Sample)	6.11E-03	6.78E-03	7.91E-03
	US-T (Temporal)	2.86E-03	3.29E-03	3.65E-03
	WMA	3.25E-03	3.26E-03	4.08E-03
	DWT (Approx.)	6.17E-03	5.20E-03	7.58E-03
	BW-FB+HF ($\alpha = 1.5$)	3.40E-03	2.55E-03	3.89E-03
	BW-FB+HF ($\alpha = 2.0$)	3.49E-03	2.61E-03	3.44E-03

inter-variable interference and accumulative distribution shift, leading to improved robustness in real-world non-stationary scenarios.

J Computational Complexity and Performance-Cost Trade-off

Although `VARes` introduces additional computational overhead due to its two-stage generative synthesis pipeline(Table 26), it achieves substantial performance gains that justify the cost. Specifically, while the training process incurs a one-time offline cost (e.g., up to $\sim$182 hours on the Energy dataset), the inference overhead remains moderate and can be amortized across large-scale usage. All experiments are conducted on a single NVIDIA GeForce RTX 3090 GPU (24GB VRAM) to ensure a consistent benchmark for execution time.

Table 25: MSE comparison of different data downsampling strategies on the ECG dataset across different downsampling ratios (12.50%, 25%, 37.50%, 50%).

Dataset	Strategy	12.50%	25%	37.50%	50%
ECG	PG (Ours)	5.90E-04	6.14E-04	6.05E-04	5.76E-04
	US-S (Sample)	8.89E-04	1.23E-03	1.95E-03	2.76E-03
	US-T (Temporal)	1.97E-03	3.07E-03	4.88E-03	8.43E-03
	WMA	9.98E-04	1.36E-03	1.69E-03	2.57E-03
	DWT (Approx.)	1.01E-03	1.13E-03	1.55E-03	2.40E-03
	BW-FB+HF ($\alpha = 1.5$)	7.62E-04	1.14E-03	1.81E-03	2.53E-03
	BW-FB+HF ($\alpha = 2.0$)	8.03E-04	1.27E-03	1.75E-03	2.50E-03

Table 26: Training and inference time breakdown of VARes across different datasets.

Dataset	VQVAE Training	VAR Training	Inference Phase
Sine	9h 16min 19s	23h 06min 36s	0h 04min 06s
Stocks	1h 07min 43s (80 epochs)	3h 44min 32s (160 epochs)	0h 05min 43s
Energy	66h 17min 11s (180 epochs)	116h 05min 04s (180 epochs)	1h 04min 51s
Mujoco	10h 08min 04s (100 epochs)	59h 27min 00s (180 epochs)	0h 15min 51s

More importantly, `VARes` consistently delivers orders-of-magnitude improvements in downstream tasks. As shown in our results, the relative error reduction ranges from $2\times$ to over $200\times$, and even reaches 6000%–25000% improvement on challenging datasets such as Mujoco. This indicates a favorable performance-per-compute trade-off to some extent. Compared to traditional downsampling methods, which are computationally cheap but suffer from severe information loss, `VARes` significantly enhances signal quality and robustness under non-stationary and distribution-shift conditions.

Furthermore, the additional computational cost is primarily incurred during offline training, making it practical for real-world deployment. Under resource-constrained scenarios, flexible configurations can be adopted to balance efficiency and accuracy.

In summary, this work is primarily exploratory in nature, aiming to demonstrate that Active Variable-Resolution Time Series Generation can provide a more effective data representation for downstream tasks, thereby offering a new perspective beyond conventional fixed-resolution processing. We acknowledge that this benefit comes with increased computational cost. Nevertheless, we consider this trade-off acceptable at the current stage, as the primary goal is to validate the effectiveness of the paradigm. In future work, we plan to incorporate model optimization techniques such as quantization and pruning to reduce the computational overhead and improve efficiency.

K Robustness Validation and Extended Evaluation on Downstream Models

First, the core of our work is to demonstrate that the variable-resolution data obtained via `VARes` is more effective in downstream tasks than data obtained from traditional deterministic downsampling methods. Therefore, our experimental design is not intended to compare the relative merits of different downstream time-series models, but rather to treat these models as a unified evaluation tool. Consequently, it is reasonable to select classic, robust, and high-performing models to validate the central argument of our work.

Furthermore, to ensure that the performance improvement does not stem from a specific model architecture, we selected three models for the downstream task validation experiments that ranked among the top performers in both the forecasting and missing value imputation tasks within the *Time-Series-Library* repository.

Finally, to further substantiate our findings, we incorporated *Time-MoE* to verify whether VARes retains its advantages under state-of-the-art configurations. The experimental results demonstrate that when using the *Time-MoE* model for prediction, our variable-resolution data still achieves a significant performance improvement compared to data obtained from the original deterministic downsampling method, further highlighting the robustness and advantages of our proposed approach.

L Time Series Representation Learning and Feature Compression

Time series representation learning has seen substantial progress in recent years. TS2Vec Yue et al. (2022) learns robust per-timestamp representations via hierarchical contrastive learning on augmented context views, while CoST Woo et al. (2022) introduces disentangled seasonal-trend representations. Tang et al. Trirat et al. (2024) provide a comprehensive survey of the field.

However, these methods typically operate on raw sequences at full resolution, leading to computational inefficiency on long or noisy data. To address this, feature compression and sequence summarization methods have emerged. STaTS Bhowmick et al. (2026) achieves multi-resolution adaptive compression via statistical change point detection, reaching 30$\times$ compression while retaining 85–90% of original performance. LightCTS Lai et al. (2024) adopts knowledge distillation for lightweight forecasting in resource-constrained environments. Frequency-aligned distillation Li et al. (2025b) further extracts multi-scale spectral features to guide compact student models.

The compression methods above focus primarily on computational efficiency rather than structural preservation along the resolution dimension. More closely related are time series super-resolution and upsampling methods, which reconstruct high-resolution signals from low-resolution inputs. Our work addresses a complementary question: how to reduce time series resolution while preserving dynamical structure for downstream tasks. We formalize this as structure-preserving downsampling — unlike naive uniform sampling or low-pass filtering, our method VARes actively learns to retain latent dynamical invariants. This reframes resolution reduction from "information loss" to "information distillation," complementing feature compression in objective while departing from it in methodology.