REVISITING NOISE DESIGN IN TIME SERIES DIFFUSION MODELS: ADAPTIVE SCHEDULE WITH TEMPORAL PRIOR

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models have demonstrated remarkable generative capabilities and been increasingly adopted in time series analysis. Recent research primarily focus on developing effective architectures and training protocols to improve performance on downstream tasks. Despite the critical role of noise in diffusion models, it has received comparatively limited attention, hindering time series diffusion models from fully realizing their potential. In this work, we propose a principled approach to design Adaptive Noise for Time Series diffusion models (ANTS) based on the inherent and unique properties of data. Our core insight is that an efficient noise schedule should distribute corruption process to maintain uniform information degradation across diffusion steps, eliminating redundant noise levels that contribute little learnable signal. It improves the precision of noise estimation and facilitates better traceability of diffusion. Furthermore, we incorporate informative statistic priors to capture temporal dependencies and enhance sample fidelity during reverse process. Our method incurs only marginal additional offline computational cost and can be seamlessly integrated into frameworks based on standard diffusion process. Experiments on generation and forecasting tasks across various datasets demonstrate that ANTS improves generative quality and predictive performance over baselines, validating its effectiveness for time series diffusion models. Our code is available at https://anonymous.4open.science/r/ANTS/.

1 INTRODUCTION

Time series data are ubiquitous across a wide range of real-world domains, including energy consumption, traffic flow and environmental monitoring (Yuan & Qiao, 2024; Wang et al., 2025; Yao et al., 2026b). In recent years, diffusion models that learn to reverse a progressive noising process have demonstrated remarkable generative capabilities (Ho et al., 2020). Their effectiveness in probabilistic modeling has been extensively validated in fields like computer vision (Esteves & Makadia, 2026; Okada et al., 2024), attracting increasing attention from time series community and facilitating various applications (Li et al., 2025; Alcaraz & Strodthoff, 2023).

While early works directly borrow diffusion frameworks from other domains, recent studies have developed specialized architectures and training protocols that enhance time series denoising and achieve promising results (Park et al., 2026; Wang et al., 2025). Despite the progress, most existing methods still rely on the conventional noise design during diffusion, where noise schedule governing the magnitude of isotropic noise injected at each step is prescribed by a linear or cosine-like curve (Nichol & Dhariwal, 2021). Although some researches explore alternative schedules (Lin et al., 2024; Esteves & Makadia, 2026), they are not well suited to the property of time series. Unlike image whose spectrum can be characterized by relatively smooth spatial-frequency statistics (Torralba & Oliva, 2003), time series exhibits substantial dataset-specific variation in spectral energy distributions, with informative components concentrated in low-frequency regions and pronounced peaks at particular frequencies. Consequently, a fixed noise schedule may fail to align corruption process with diverse temporal structures (Yao et al., 2026a), ultimately leading to suboptimal performance.

As illustrated in Figure 1, under the fixed cosine schedule at same step, the Flow time series undergoes substantially more severe corruption, whereas the Energy time series experiences relatively

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.

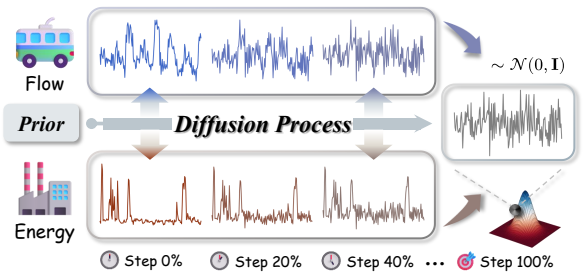

Figure 1: Fixed schedule struggles to adapt to heterogeneous time series across domains, hindering appropriate noise for uniform information degradation.

mild degradation and retains more informative temporal structure. This discrepancy highlights the unique diffusion dynamics driven by dataset. Moreover, the Flow time series rapidly collapse into a noise-like state in early diffusion stage, suggesting that the remaining training steps become meaningless and redundant, as excessive noise has destroyed informative content required for effective learning. In contrast, insufficient noise injection may preserve overwhelming original content, making early steps ineffective due to inadequate corruption. Intuitively, an efficient diffusion process should allocate corruption across steps such that the information of noised time series decreases at a controlled and approximately uniform rate. This design can make adjacent diffusion states more distinguishable and ensure that each step contributes meaningful corruption, thereby facilitating the efficiency of generative process and high-quality samples (Okada et al., 2024). Although previous work (Lee et al., 2024) attempt to alleviate the issue by searching over parametric candidate schedule pool, such approach is constrained by the predefined space and may not adapt efficiently and flexibly to various datasets.

In this work, we propose **A**daptive **N**oise for **T**ime **S**eries (ANTS) that adjusts to dataset-specific non-stationarity distribution transition to optimize diffusion process. Specifically, we leverage time series statistic to characterize the degradation of information toward noise with low non-stationarity as noise level increases. We then resample schedule to enforce a constant rate of change in distribution to explicitly construct structure-aware diffusion trajectory. To better match noise with adaptive schedule and improve stability of generative process, we further incorporate temporal properties into prior. The resulting paradigm is equivalent to whitened diffusion process and can be seamlessly integrated into existing DDPM-based models with marginal computational overhead.

Our main contributions are summarized as follows:

- We develop a tight noise schedule that adaptively follows non-stationarity dynamics of time series. By enforcing uniform transition of distribution across diffusion steps, the schedule improves traceability of diffusion trajectory and facilitates higher-quality samples.
- We design specialized prior with structured distribution that encodes intrinsic scale information and temporal dependencies of time series datasets. It is combined with the proposed schedule to establish a model-agnostic paradigm that only requires offline precomputation once and enables seamless integration with existing diffusion architectures.
- We integrate ANTS with advanced generative models for time series generation and forecasting. Experiments across various datasets demonstrate its effectiveness in improving model performance. Further analyses provide insight into the underlying mechanism.

2 Related Work

Time Series Diffusion Models. Denoising diffusion probabilistic models (DDPMs) (Ho et al., 2020) are generative models that learn to reverse a noising process. Their powerful capabilities have been widely demonstrated (Esteves & Makadia, 2026; Okada et al., 2024). DiffusionTS (Yuan & Qiao, 2024) employs disentangled temporal representations to synthesize samples. DiM-TS (Yao et al., 2026b) further incorporates selective state space models into diffusion process. In probabilistic forecasting, diffusion models are adopted to capture future sequence distribution and enable uncertainty quantification (Alcaraz & Strodthoff, 2023; Wang et al., 2025). TimeDiff (Shen & Kwok, 2023) employs future mixup strategy during training to mitigate boundary disharmony, while CW-Diff (Yang et al., 2026) improves training protocol by introducing conditional whitening. Despite these advances, existing methods primarily focus on developing architectures, leaving the critical component of diffusion—noise—relatively underexplored. While the objective is to recover time series, a more efficient and traceable corruption process may facilitate the process (Okada et al., 2024). Our design is therefore complementary to these methods and can unlock their potential.

Noise Schedule. It determines noise levels used to corrupt data at different steps. Early works made noise magnitude increase linearly with adjacent diffusion step (Ho et al., 2020). However, such strategy may cause data dominated by low-frequency components to be prematurely corrupted into a noise-like state. To achieve a smoother transition, cosine schedule (Nichol & Dhariwal, 2021) applies slower changes at the beginning and end of diffusion process and has been widely adopted in existing works. CRS (Okada et al., 2024) then controls noise evolution by introducing discrepancy-based measure. Spectrally-guided schedule (Esteves & Makadia, 2026) proposes a per-instance noise strategy that adapts to spectral diversity of image. Considering the fundamental structural differences between time series and other modalities, ANT (Lee et al., 2024) selects noise schedule hyperparameters from candidates based on statistics. MA-TSD (Wang et al., 2025) utilizes moving averages to preserve low-frequency components and designs a compensated schedule according to variance reduction. However, these approaches typically require manual tuning or specialized framework, limiting adaptability and flexible integration with existing diffusion models. In contrast, our method establishes a direct connection between noise and dataset characteristics, requiring minimal intervention while improving overall sample quality.

3 PRELIMINARIES

Diffusion models are a class of generative models that consist of forward process and backward denoising process. The forward process is a Markov process that gradually perturbs data distribution by progressively injecting noise into input $X_0 \sim q(X)$, transforming it into a Gaussian distribution $X_1 \sim \mathcal{N}(0, I)$. It can be described by a stochastic differential equation (SDE):

$$dX_\tau = -\frac{1}{2}\beta_\tau X_\tau d\tau + \sqrt{\beta_\tau} dW_\tau, \quad \tau \in [0, 1], \tag{1}$$

where $\beta_\tau > 0$ controls the noise intensity and W_τ denotes a Brownian motion. For clarity, τ represents the diffusion time, whereas t denotes time series index throughout the manuscript. Following the Ornstein–Uhlenbeck process, the marginal distribution of X_τ can be derived in closed form:

$$X_\tau \overset{d}{=} \alpha_\tau X_0 + \sigma_\tau \epsilon, \quad \epsilon \sim N(0, I) \tag{2}$$

where $\alpha_\tau := \exp\left\{-\int_0^\tau \beta_u du/2\right\}$ and $\sigma_\tau^2 := 1 - \alpha_\tau^2$ are determined by noise schedule. For example, a cosine schedule defines $\alpha_\tau = \cos(\pi\tau/2)$. The backward process aims to learn the inverse transition dynamics and recover original data by tracing the reverse trajectory:

$$d\overleftarrow{X}_\tau = \left[-\frac{1}{2}\beta_\tau \overleftarrow{X}_\tau - \beta_\tau \nabla_x \log p_\tau(\overleftarrow{X}_\tau)\right] d\tau + \sqrt{\beta_\tau}\, d\overleftarrow{W}_\tau. \tag{3}$$

Since the score function is intractable (Ho et al., 2020), it is approximated by a neural network s_θ parameterized by θ. Hence, the denoising network is optimized by minimizing the objective:

$$\mathbb{E}_{X_0,\tau,\epsilon} \| s_\theta\left(\alpha_\tau X_0 + \sigma_\tau \epsilon, \tau\right) + \epsilon/\sigma_\tau \|^2. \tag{4}$$

After training, unconditional samples can be generated by substituting the intractable score function $\nabla_x \log p_\tau(\overleftarrow{X}\tau)$ with learned estimator $s_\theta(X_\tau, \tau)$ and iteratively denoising X_τ according to Eq. 3.

4 METHODOLOGY

In this section, we present the proposed adaptive noise design ANTS as depicted in Figure 2. We first discuss motivation and theoretical derivation of noise schedule with uniform evolution. Then, we introduce discrepancy measure to quantify distribution change of time series and describe procedure for dataset-specific adaptive schedule. Finally, we enhance noise by injecting scale information and prior structural temporal dependencies. Collectively, these components form the effective framework that can be integrated into diffusion-based architectures to improve time series modeling. Detailed workflow is summarized in Algorithm 1 in Appendix B.

4.1 UNIFORM EVOLUTION SCHEDULING

Our main contribution is a data driven adaptive noise schedule that adapts to the evolution of informative content. By smoothly allocating noise across diffusion stages, it regulates the corruption process to achieve a balanced information transition while avoiding abrupt and unpredictable changes, ultimately facilitating a more accurate approximation of the forward process.

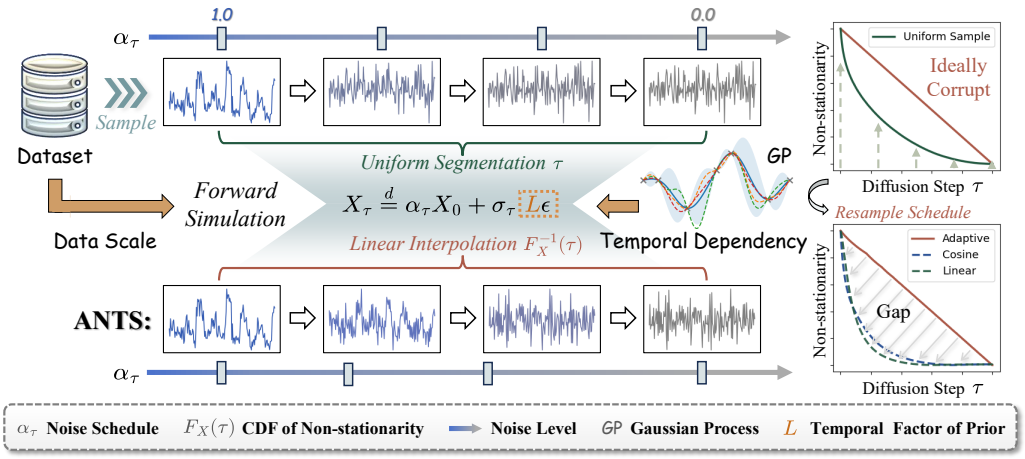

Figure 2: Workflow of ANTS. We simulate forward process with prior incorporating temporal scale and dependency. α_τ is first initialized by uniformly sampled τ to obtain non-stationarity evolution, then we use CDF F_X to resample the adaptive schedule of ANTS for appropriate corruption over τ.

Let $\mathcal{D}(\tau, \tau + \Delta\tau)$ denote discrepancy between diffused time series distributions $q(X_\tau)$ and $q(X_{\tau+\Delta\tau})$. An ideal diffusion trajectory should maintain a consistent distribution transition rate, where each step introduces a comparable level of degradation. Therefore, it is expected to follow:

$$\mathcal{D}(\tau, \tau + \Delta\tau) = \frac{\mathcal{D}(\tau, \tau + \Delta\tau) - \mathcal{D}(\tau, \tau)}{\Delta\tau}\Delta\tau \simeq \frac{\partial \mathcal{D}(\tau, \tau')}{\partial \tau'}\bigg|_{\tau'=\tau} \Delta\tau = c, \tag{5}$$

where c is the constant rate depending on discrepancy measurement $\mathcal{D}$, and $\mathcal{D}(\tau, \tau) = 0$.

In practice, time series distribution evolution does not naturally follow such transition. Directly sampling τ uniformly may lead to uneven degradation of structures. To this end, noise schedule α is introduced to control the corruption and make diffusion process toward desired direction. To explicitly reflect the dependence of measurement on schedule, the objective is reformulated as:

$$\mathcal{D}(\alpha_\tau, \alpha_{\tau+\Delta\tau}) \simeq \frac{\partial \mathcal{D}(\alpha_\tau, \alpha_{\tau'})}{\partial \tau'}\bigg|_{\tau'=\tau} \Delta\tau = v(\alpha)\frac{d\alpha_\tau}{d\tau}\Delta\tau = c, \tag{6}$$

where $v(\alpha)$ denotes probability distribution change rate of diffused time series induced by the noise schedule. Consequently, the optimal noise schedule should satisfy:

$$\frac{d\alpha_\tau}{d\tau} = c \cdot v(\alpha)^{-1}, \quad c = -\int_{\alpha_{\min}}^{\alpha_{\max}} v(\alpha) d\alpha, \tag{7}$$

where $\alpha_{\min} = \alpha(1)$ and $\alpha_{\max} = \alpha(0)$ denote the minimum and maximum values of α, respectively.

This formulation is closely related to quantum annealing (Morita & Nishimori, 2008; Okada et al., 2024), where the adiabatic theorem provides traceability of the optimal solution along interpolation between initial state with trivial optimum and target cost function. Let α_τ denote the weight of target function and $1 - \alpha_\tau$ that of initial function. The inverse traceability is formulated by $v(\alpha)\frac{d\alpha_\tau}{d\tau}$ as defined in Eq.6. Prior studies (Roland & Cerf, 2002) show that optimal α_τ should minimize the maximum of traceability, which is equivalent to maintaining consistent throughout interpolation. Under our context, this principle implies that an optimal noise schedule should maintain an approximately uniform evolution of discrepancy measure $\mathcal{D}$ across diffusion steps.

4.2 Integrated Absolute Autocorrelation Time

While α_t is derived by Eq.7, the measure $\mathcal{D}$ that quantifies the discrepancy of diffused time series plays a critical role in determining $v(\alpha)$ and explicitly shaping noise schedule. Due to the intrinsic structure differences between time series and other modalities like images and natural language, directly utilizing discrepancy measures from them may fail to capture unique characteristics of time series. To accurately quantify the degradation of time series under different noise levels, we introduce the Integrated Absolute Autocorrelation Time (IAAT).

IAAT is a variation of the Integrated Autocorrelation Time (IAT) (Madras & Sokal, 1988; Lee et al., 2024) which aggregates dependency across multiple temporal scales and characterizes the effective memory of time series. Since negative autocorrelation can also encode meaningful information, we adopt the absolute autocorrelation function to avoid cancellation between positive and negative dependencies. Let $\mathcal{D}(\alpha_\tau, \alpha_{\tau+\Delta\tau}) = \psi_{\text{IAAT}}(X_\tau) - \psi_{\text{IAAT}}(X_{\tau+\Delta\tau})$, the ψ_{IAAT} can be defined as:

$$\psi_{\text{IAAT}}(X_\tau) = 1 + 2\sum_{k=1}^{\infty} |\rho_k|, \quad \rho_k = \text{Cov}(X_\tau^t, X_\tau^{t+k})/\text{Var}(X_\tau^t). \tag{8}$$

Natural time series typically exhibit nontrivial temporal dependencies reflected by non-stationary such as trend, periodic pattern, and long-range correlation. They lead to nonzero autocorrelation across a range of lags and ψ_{IAAT} is generally larger than temporally independent white noise with low non-stationarity. During diffusion, the non-stationarity gradually vanishes and ψ_{IAAT} monotonically decreases toward theoretical boundary (see proof in Section C.1). Therefore, the evolution trajectory of ψ_{IAAT} serves as an effective proxy for measuring distribution transition of diffused time series.

4.3 Adaptive Noise Schedule

Once the forward process is specified, $v(\alpha)$ can be estimated by simulating on dataset. The procedure is independent of model parameters and can be precomputed offline before training. Given a predefined discretization steps T_d, we first construct α on uniformly spaced grid over interval $[0, 1]$ as $\alpha_\tau = 1 - \tau, \tau \in \{0, 1/T_d, \ldots, 1\}$. For each instance, we then compute corresponding ψ_{IAAT} at different noise level of τ and average values across entire training set. This yields a dataset-specific IAAT trajectory that implicitly reflects the evolution of distribution throughout diffusion process.

As illustrated in Figure 2, the evolution of ψ_{IAAT} tends to be large in early region ($\tau \to 0$) where non-stationary undergoes substantial degradation. In contrast, it becomes smaller in later region ($\tau \to 1$) where diffused data have already approached noise-like state and further corruption produces limited non-stationary changes. Consequently, the adaptive schedule would allocate more steps to region with rapidly changing dynamics, yielding a denser discretization of early stage.

Different time series exhibit diverse trajectories due to various domain and granularity. It is difficult to describe them using a unified formulation as image (Esteves & Makadia, 2026). Although identifying a specific function could facilitate the derivation of analytical solution for $v(\alpha)$, empirical curves are not well approximated by simple function (e.g. x^n). Conversely, adopting highly complex form may introduce fitting costs and limit efficient generalization across datasets. To this end, we approximate the IAAT curve via linear interpolation. Let $Z = \max(\psi_{\text{IAAT}}) - 1$ be the normalization constant and v as a probability distribution function (PDF). We define the cumulative distribution function (CDF) F_X and the resampled adaptive noise schedule α_τ as follows:

$$F_X(\tau) = \frac{1}{Z}(\psi_{\text{IAAT}}(1-\tau) - 1) = \int_0^\tau \frac{v(u)}{Z} du, \quad \alpha_\tau = F_X^{-1}(\tau), \quad \tau \in \{0, 1/T_s, \ldots, 1\}, \tag{9}$$

where T_s is the required denoising step. It achieves optimal traceability between consecutive steps as stated in theorem in Appendix C.2. The resulting schedule will cover low noise levels more often, emphasizing the preservation of coarse temporal structures over fine-grained details within high-frequency. However, generating high-quality samples requires balancing attention at different scales. Following prior works (Okada et al., 2024; Esteves & Makadia, 2026), we employ a mixed strategy that combines the adaptive schedule α_τ with other complementary schedule (e.g. cosine).

4.4 Time Series Prior

As standard diffusion forces Gaussian noise as terminal distribution, the reverse process is not explicitly guided by inherent properties of time series (Han et al., 2022). Moreover, the IAAT may exhibit abrupt change between sampling steps for some datasets, making the discretization error more significant and degrading model performance. From optimal transport perspective (Park et al., 2026), reducing the Kullback-Leibler divergence (KLD) between prior and target distribution can alleviate the difficulty of learning transformation and facilitate better match. Motivated by this insight, we design a temporal-aware prior that replace the mean and covariance of $N(0, I)$.

Employing neural networks to predict these parameters (Yang et al., 2026; Wang et al., 2025) could introduce complexity and computational overhead. To enable efficient generalization, we instead estimate them directly from dataset statistic. Since most models perform input normalization with zero mean, we focus on estimating covariance Σ_X. The covariance at each index t is represented as Σ_X^t, and the distribution of data-dependent prior can be obtained as:

$$X_1 \sim N(\mu, \Sigma_X), \quad \mu = 0, \quad \Sigma_X := [\Sigma_X^1, \dots, \Sigma_X^T] \tag{10}$$

where $\Sigma_X^t \in \mathbb{R}^{d\times d}$ is diagonal matrix along the channel dimension, T denotes the length of window.

To comprehensively extract temporal dynamics and correlations, we further employ Gaussian Process (GP) into prior. GP is a classical non-parametric approach that measures joint variability among arbitrary temporal points over functions (Biloš et al., 2023). For given timestamps $\Omega = \{1, \dots, T\}$ and radial basis function (RBF) kernel $\mathcal{K}(i, j) = \eta \exp(-|i - j|^2)$, sample noise from GP prior as:

$$X_1 \sim N(\mu, \Sigma(\Omega)), \quad \Omega = \{1, \dots, T\}, \tag{11}$$

where $\Sigma(\Omega)_{i,j} = \Sigma_X^{i,j} + \mathcal{K}(i, j)$. Let L denote the lower-triangular Cholesky factor of $\Sigma(\Omega)$ under channel independence, satisfying $\Sigma(\Omega) = LL^\top$. The forward process is reformulated as:

$$d(\mathbf{X}_\tau - \mu) = -\frac{1}{2}\beta_\tau(\mathbf{X}_\tau - \mu)d\tau + \sqrt{\beta_\tau} \cdot L \cdot d\mathbf{W}_\tau, \quad \tau \in [0, 1]. \tag{12}$$

The terminal distribution X_1 becomes closer to target, while the enhanced prior captures scale statistics and cross-timestep correlations, addressing heteroscedasticity and preserving temporal dependencies. Therefore, we simulate the evolution of diffused time series under this formulation to construct adaptive noise schedule, which can then be combined and efficiently integrated into existing frameworks. Specially, the process can be equivalently conducted in the whitened space, which remains consistent with the standard diffusion formulation in Eq.2 (see proof in Appendix C.4).

5 Experiments

In this section, we conduct experiments to evaluate the effectiveness of ANTS. We focus on two representative tasks: ❶ **time series generation**, which aligns with the objective of generative modeling and can be viewed as a special case of imputation, and ❷ **time series forecasting**, one of the most widely studied applications in time series analysis. We reproduce baselines using official implementations and make adjustments if necessary. All experiments are conducted on a machine equipped with NVIDIA V100 GPU and 32GB memory.

Datasets: We adopt widely used benchmarks spanning diverse domains and temporal resolutions. For time series generation, we evaluate on four real-world datasets: ETTh1 (Zhou et al., 2021), Energy (Candanedo et al., 2017), Wind, and KDDCup (Godahewa et al., 2021). The generation models are expected to learn the underlying data distribution from the entire dataset. For time series forecasting, we employ public datasets: ETTh1, ETTm2, ECL, and Traffic (Zhou et al., 2021; Wu et al., 2021). The training, validation, and test sets are split with a ratio of $3:1:1$ for ETT, while the remaining datasets follow a ratio of $7:1:2$. Detailed descriptions are provided in Appendix D.

Baselines: Since we simulate on diffusion process of conventional DDPMs, we select five advanced diffusion models with corresponding frameworks for evaluation. Notably, ANTS is not limited to these methods and can generalize to other models. We evaluate the original models (denoted as Raw) and integrate them with our proposed method (denoted as ANTS) to assess effectiveness. For generation task, we consider DiffusionTS (Yuan & Qiao, 2024), PaD-TS (Li et al., 2025), and DiM-TS (Yao et al., 2026b) as baselines. As for forecasting task, we adopt DiffusionTS, SSSD (Alcaraz & Strodthoff, 2023), and TimeDiff (Shen & Kwok, 2023).

Metrics: We evaluate generative performance with popular quantitative metrics from perspectives of distribution diversity and dependencies fidelity: Context-Fréchet Inception Distance (Context-FID) (Paul et al., 2021), Correlational Score (Liao et al., 2020), and Discriminative Score (Yoon et al., 2019). For forecasting evaluation, we adopt Continuous Ranked Probability Score (CRPS) (Matheson & Winkler, 1976) and Quantile Interval Coverage Error (QICE) (Han et al., 2022) to measure probabilistic forecasting performance, while Mean Squared Error (MSE) and Mean Absolute Error (MAE) are employed to evaluate deterministic point forecasting accuracy. Reported results are averaged over 5 runs. Formal definitions of these metrics are included in Appendix D.2.

Table 1: Generation experiment results. The better results are underlined.

Model	Dataset	Context-FID Score (↓)		Correlational Score (↓)		Discriminative Score (↓)	
		Raw	ANTS	Raw	ANTS	Raw	ANTS
DiffusionTS	ETTh1	$1.370_{\pm 0.085}$	$\underline{1.247}_{\pm 0.095}$	$0.100_{\pm 0.002}$	$\underline{0.092}_{\pm 0.006}$	$0.177_{\pm 0.008}$	$\underline{0.163}_{\pm 0.008}$
	Energy	$0.123_{\pm 0.007}$	$\underline{0.093}_{\pm 0.005}$	$\underline{0.420}_{\pm 0.060}$	$0.603_{\pm 0.114}$	$0.288_{\pm 0.017}$	$\underline{0.283}_{\pm 0.029}$
	Wind	$0.766_{\pm 0.021}$	$\underline{0.559}_{\pm 0.028}$	$0.074_{\pm 0.004}$	$\underline{0.072}_{\pm 0.005}$	$0.135_{\pm 0.012}$	$\underline{0.125}_{\pm 0.012}$
	KddCup	$3.365_{\pm 0.205}$	$\underline{2.454}_{\pm 0.141}$	$11.099_{\pm 0.900}$	$\underline{9.733}_{\pm 1.197}$	$0.477_{\pm 0.014}$	$\underline{0.445}_{\pm 0.031}$
PaD-TS	ETTh1	$1.636_{\pm 0.095}$	$\underline{1.520}_{\pm 0.083}$	$0.100_{\pm 0.007}$	$\underline{0.095}_{\pm 0.010}$	$0.173_{\pm 0.003}$	$\underline{0.170}_{\pm 0.011}$
	Energy	$0.269_{\pm 0.033}$	$\underline{0.166}_{\pm 0.013}$	$0.561_{\pm 0.128}$	$\underline{0.515}_{\pm 0.061}$	$0.252_{\pm 0.118}$	$\underline{0.202}_{\pm 0.109}$
	Wind	$1.467_{\pm 0.121}$	$\underline{0.390}_{\pm 0.037}$	$0.095_{\pm 0.006}$	$\underline{0.043}_{\pm 0.003}$	$0.219_{\pm 0.086}$	$\underline{0.052}_{\pm 0.027}$
	KddCup	$1.087_{\pm 0.110}$	$\underline{1.071}_{\pm 0.070}$	$7.063_{\pm 2.343}$	$\underline{6.570}_{\pm 1.340}$	$0.489_{\pm 0.009}$	$\underline{0.429}_{\pm 0.006}$
DiM-TS	ETTh1	$0.083_{\pm 0.004}$	$\underline{0.073}_{\pm 0.002}$	$0.024_{\pm 0.005}$	$\underline{0.020}_{\pm 0.003}$	$0.010_{\pm 0.006}$	$\underline{0.009}_{\pm 0.005}$
	Energy	$0.142_{\pm 0.005}$	$\underline{0.095}_{\pm 0.007}$	$0.507_{\pm 0.153}$	$\underline{0.488}_{\pm 0.147}$	$0.314_{\pm 0.134}$	$\underline{0.278}_{\pm 0.047}$
	Wind	$0.247_{\pm 0.008}$	$\underline{0.221}_{\pm 0.016}$	$0.043_{\pm 0.011}$	$\underline{0.040}_{\pm 0.007}$	$0.048_{\pm 0.015}$	$\underline{0.041}_{\pm 0.014}$
	KddCup	$0.171_{\pm 0.009}$	$\underline{0.112}_{\pm 0.010}$	$8.612_{\pm 1.328}$	$\underline{8.368}_{\pm 2.405}$	$0.061_{\pm 0.012}$	$\underline{0.056}_{\pm 0.010}$

Table 2: Forecasting experiment results. The better results are underlined.

Model	Dataset	CRPS (↓)		QICE (↓)		MSE (↓)		MAE (↓)	
		Raw	ANTS	Raw	ANTS	Raw	ANTS	Raw	ANTS
DiffusionTS	ETTh1	0.608	<u>0.597</u>	6.505	<u>6.267</u>	1.103	<u>1.078</u>	0.781	<u>0.772</u>
	ETTm2	1.098	<u>0.947</u>	10.456	<u>8.900</u>	2.523	<u>2.100</u>	1.299	<u>1.133</u>
	ECL	<u>0.615</u>	0.617	<u>4.279</u>	4.361	1.058	<u>1.053</u>	0.839	<u>0.836</u>
	Traffic	0.656	<u>0.625</u>	6.960	<u>6.228</u>	1.450	<u>1.437</u>	0.805	<u>0.800</u>
SSSD	ETTh1	0.548	<u>0.537</u>	5.219	<u>4.664</u>	0.828	<u>0.814</u>	0.697	<u>0.688</u>
	ETTm2	0.795	<u>0.626</u>	8.850	<u>3.787</u>	1.801	<u>1.376</u>	1.053	<u>0.839</u>
	ECL	0.374	<u>0.364</u>	5.497	<u>5.317</u>	0.469	<u>0.454</u>	0.481	<u>0.477</u>
	Traffic	0.610	<u>0.409</u>	13.446	<u>6.148</u>	1.289	<u>0.984</u>	0.650	<u>0.511</u>
TimeDiff	ETTh1	0.467	<u>0.461</u>	14.906	<u>14.450</u>	0.532	<u>0.529</u>	0.482	<u>0.480</u>
	ETTm2	<u>0.309</u>	0.310	<u>11.436</u>	12.779	0.269	<u>0.256</u>	0.332	<u>0.325</u>
	ECL	0.750	<u>0.739</u>	15.501	<u>15.096</u>	0.869	<u>0.867</u>	0.764	<u>0.763</u>
	Traffic	0.772	<u>0.760</u>	15.423	<u>14.929</u>	1.354	<u>1.350</u>	<u>0.786</u>	<u>0.786</u>

5.1 Time Series Generation

We perform unconditional generation with length 168, the results are presented in Table 1. Among time series diffusion models concerned, ANTS generally improves the quality and fidelity of generated instances in most cases. Specifically, it achieves improvements of over 20% in Context-FID and an average reduction of approximately 15% in Discriminative Score. Real-world time series exhibit periodic patterns reflected in high non-stationary are not explicitly considered by fixed noise schedules in raw baselines. Therefore, the diffusion process may contain redundant steps that operate on nearly pure noise, making it challenging for reverse process to recover original time series. Notably, ANTS empowers models to achieve superior performance over other baselines in some scenarios. It highlights that architectural improvements and noise design represent two complementary directions of diffusion models for time series modeling. Figure 3 provides an intuitive case on the Wind dataset.

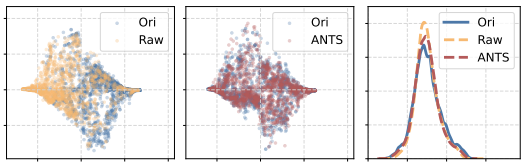

Figure 3: T-SNE plots and KDE distributions on original and synthetic time series by PaD-TS.

5.2 Time Series Forecasting

Table 2 reports the forecasting results with prediction horizon 192 and historical window 168. As observed, ANTS improves forecasting performance across most datasets and baselines, demonstrating its broad applicability. Quantitatively, ANTS substantially reduces probabilistic forecasting errors, achieving average improvements of over 6.9% and 12.4% on CRPS and QICE, respectively. It indicates that ANTS facilitates more accurate modeling of predictive distributions and better captures temporal dependencies. For MSE and MAE, ANTS also achieves statistically significant improve-

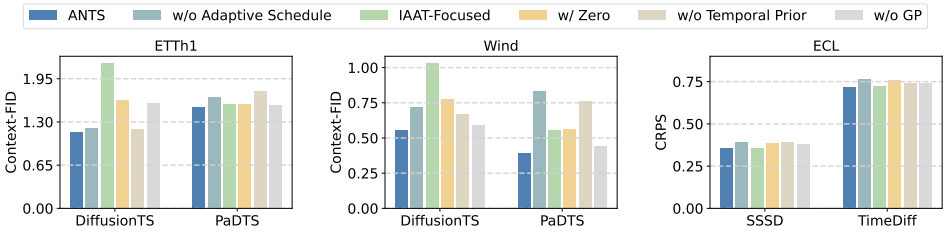

Figure 4: Ablation study of ANTS on ETTh1, Wind (generation) and ECL (forecasting).

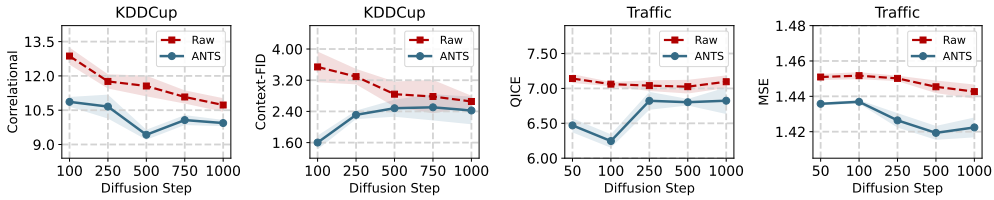

Figure 5: Influence of denoising diffusion steps on KDDCup (generation) and Traffic (forecasting).

ments, underscoring its effectiveness across diverse scenarios. It is expected that ANTS provides marginal improvements in minor cases, as the adaptive objective focuses on specific statistic property and not fully capture multi-scale structures. In addition, the original schedule may already be close to the target, leaving limited room for further enhancement (as stated in Appendix E).

5.3 Ablation Study

To validate the effectiveness of each component in ANTS, we design a series of ablation variants based on DiffusionTS for comprehensive analysis: ❶ **w/o Adaptive Schedule**: This variant adopts default schedule in the baselines instead of the proposed adaptive schedule. ❷ **IAAT-Focused**: The schedule is obtained by enforcing a linear decay of IAAT without incorporating complementary schedule. ❸ **w/ Zero**: We replace schedule with the zero terminal signal-to-noise ratio (SNR) schedule widely adopted in previous studies (Lee et al., 2024; Lin et al., 2024). ❹ **w/o Temporal Prior**: This variant removes the data-dependent prior and adopts Gaussian noise as the terminal distribution. ❺ **w/o GP**: The Gaussian Process is removed from the prior construction. The resulting isotropic prior only preserves data-dependent scale information.

As evidenced by Figure 4, both **w/o Adaptive Schedule** and **IAAT-Focused** variants consistently lead to substantial performance degradation, indicating that dataset-specific statistic evolution patterns and the preservation of general diffusion characteristics are indispensable. Our mixed schedule which combines two components provides complementary benefits. Although **w/ Zero** achieves some improvement, it still lacks consideration of unique property of time series and consequently yields suboptimal performance. Further performance gains stem from injecting scale information into prior. While incorporating GP provides moderate improvement, it enhances isotropic prior with temporal dependencies and still plays a crucial role.

5.4 Analysis

Sensitivity of Statistic Metric. Existing analysis provides various statistic metrics that quantify non-stationarity. However, there is no universally optimal metric that measure corruption degree of time series. To validate the rationality of adopting IAAT and examine the sensitivity of ANTS to different measurements, we replace IAAT with widely used Lag1AC and VarAC (Witt et al., 1998; Lee et al., 2024). Table 3 reports the averaged results in forecasting. We observe that different statistics lead to comparable improvements over original models, demonstrating that the effectiveness mainly originates from the principle of aligning diffusion trajectory with dataset-specific information evolution. Overall, IAAT that covers all time lags can capture more fine-grained variations and result in the most effective performance. Please refer to Appendix D.3 for concrete results.

Table 3: Statistic metrics comparison.

Metric	Method	Model		
		DiffTS	SSSD	TimeDiff
CRPS	Raw	0.744	0.582	0.575
	Lag1AC	0.703	0.504	0.572
	VarAC	0.699	0.532	0.580
	IAAT	0.697	0.484	0.568
MAE	Raw	0.931	0.720	0.594
	Lag1AC	0.891	0.661	0.592
	VarAC	0.884	0.677	0.599
	IAAT	0.886	0.629	0.589

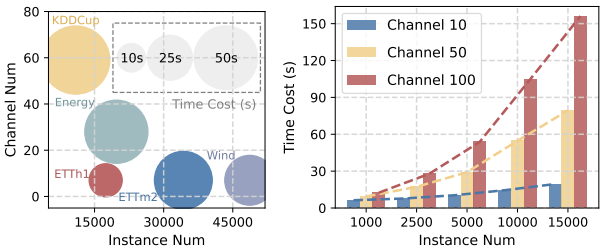

Figure 6: Computational overhead of ANTS.

Impact of Denoising Step T_s. As shown in Figure 5, under the challenging large-scale datasets, ANTS consistently outperforms raw models across different T_s settings. The performance gap widens under smaller step budgets. The advantage stems from the tight noise schedule which enforces a more uniform and informative distribution transition, enabling reasonable denoising even with limited denoising steps. It allows ANTS to reach optimal performance with fewer T_s than raw models, eliminating unnecessary steps and thereby reducing training costs while improving inference efficiency and overall performance. When T_s exceeds the optimal steps, ANTS exhibits performance degradation that may be attributed to amplifying discretization errors or overfitting. In contrast, the original schedule may benefit from additional steps, as larger T_s allows the change among steps to become less abrupt and more gradual.

Efficiency Analysis. The computational overhead introduced by ANTS arises from two aspects. ❶ ANTS requires computing dataset-specific evolution of non-stationarity under different noise levels, where the offline procedure cost depends on the number of samples and variables. ❷ ANTS introduces temporal prior that requires construction of structured noise injection. Since the latter does not involve prediction by model and incurs negligible cost, we focus analysis on the former. Figure 6 reports the time of computing IAAT evolution across datasets and scales. The results show that ANTS introduces less than 1.5% additional computational overhead relative to the total training time of original models, which is negligible and practically acceptable given the performance gains. And the computation only needs to be done once before training. Furthermore, our analysis reveals that ANTS can achieve superior performance with fewer steps. It indicates that ANTS can improve efficiency in practice by reducing unnecessary steps during training and inference.

5.5 Discussion

Despite the promising results, several limitations remain and provide opportunities for further exploration. For example, although ANTS enforces uniform distribution change across steps, the denoising step T_s still influences model performance. Determining the optimal T_s in advance remains a problem. Moreover, while our contribution provides theoretically motivated and empirically effective paradigm for incorporating temporal property into noise design, it does not establish a formal guarantee that the adaptive schedule is optimal for sample generation quality. We hope this work encourages the community to investigate noise design in time series diffusion models and pursue solutions that explicitly connect diffusion trajectories with sample quality across domains.

6 Conclusion

In this work, we propose a dataset-specific adaptive noise design for time series diffusion models by leveraging the inherent statistic of time series. We develop a lightweight procedure to derive adaptive noise schedule, which enforce a uniform distribution transition throughout diffusion process. It improves the controllability and traceability of injected noise. Furthermore, we introduce prior with scale information and Gaussian Process, allowing the diffusion process to better preserve valuable temporal dependencies. Extensive experimental results demonstrate that ANTS can be seamlessly integrated with advanced time series diffusion models and generally enhance performance across diverse settings with negligible computational overhead. It provides a complementary perspective to existing architectures, revealing the potential of noise design for future investigation.

7 AI USE STATEMENT

In this work, we used generative AI tools solely for the purpose of improving the grammar, clarity, and fluency of the manuscript. The content, structure, technical contributions, experiments, analysis, and all scientific writing were entirely conceived, drafted, and validated by the human authors. The AI tools were not involved in research ideation, experimental design, data analysis, or any aspect of the scientific content creation. All outputs generated by the AI tools were reviewed and edited by the authors to ensure accuracy and correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Juan Miguel Lopez Alcaraz and Nils Strodthoff. Diffusion-based time series imputation and forecasting with structured state space models. *Transactions on Machine Learning Research*, 2023.

Marin Biloš, Kashif Rasul, Anderson Schneider, Yuriy Nevmyvaka, and Stephan Günnemann. Modeling temporal data as continuous functions with stochastic process diffusion. In *international conference on machine learning*, pp. 2452–2470. PMLR, 2023.

Luis M Candanedo, Véronique Feldheim, and Dominique Deramaix. Data driven prediction models of energy use of appliances in a low-energy house. *Energy and buildings*, 140:81–97, 2017.

Carlos Esteves and Ameesh Makadia. Spectrally-guided diffusion noise schedules. *arXiv preprint arXiv:2603.19222*, 2026.

Rakshitha Godahewa, Christoph Bergmeir, Geoffrey I Webb, Rob J Hyndman, and Pablo Montero-Manso. Monash time series forecasting archive. *arXiv preprint arXiv:2105.06643*, 2021.

Xizewen Han, Huangjie Zheng, and Mingyuan Zhou. Card: Classification and regression diffusion models. *Advances in Neural Information Processing Systems*, 35:18100–18115, 2022.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Seunghan Lee, Kibok Lee, and Taeyoung Park. Ant: Adaptive noise schedule for time series diffusion models. *Advances in Neural Information Processing Systems*, 37:121099–121130, 2024.

Yang Li, Han Meng, Zhenyu Bi, Ingolv T Urnes, and Haipeng Chen. Population aware diffusion for time series generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18520–18529, 2025.

Yuxin Li, Wenchao Chen, Xinyue Hu, Bo Chen, Mingyuan Zhou, et al. Transformer-modulated diffusion models for probabilistic multivariate time series forecasting. In *International Conference on Learning Representations*, volume 2024, pp. 18604–18622, 2024.

Shujian Liao, Hao Ni, Lukasz Szpruch, Magnus Wiese, Marc Sabate-Vidales, and Baoren Xiao. Conditional sig-wasserstein gans for time series generation. *arXiv preprint arXiv:2006.05421*, 2020.

Shanchuan Lin, Bingchen Liu, Jiashi Li, and Xiao Yang. Common diffusion noise schedules and sample steps are flawed. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5392–5399. IEEE, 2024.

Neal Madras and Alan D Sokal. The pivot algorithm: A highly efficient monte carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50(1):109–186, 1988.

James E Matheson and Robert L Winkler. Scoring rules for continuous probability distributions. *Management science*, 22(10):1087–1096, 1976.

Satoshi Morita and Hidetoshi Nishimori. Mathematical foundation of quantum annealing. *Journal of Mathematical Physics*, 49(12), 2008.

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PmLR, 2021.

Shuntaro Okada, Ryota Yoshihashi, Hirokatsu Kataoka, Tomohiro Tanaka, et al. Constant rate scheduling: A general framework for optimizing diffusion noise schedule via distributional change. *arXiv preprint arXiv:2411.12188*, 2024.

Jinseong Park, Seungyun Lee, Woojin Jeong, Yujin Choi, and Jaewook Lee. Timebridge: Better diffusion prior design with bridge models for time series generation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1135–1146, 2026.

Jeha Paul, Bohlke-Schneider Michael, Mercado Pedro, Kapoor Shubham, Singh Nirwan Rajbir, Flunkert Valentin, Gasthaus Jan, and Januschowski Tim. Psa-gan: Progressive self attention gans for synthetic time series. *arXiv preprint arXiv:2108.00981*, 2021.

Jérémie Roland and Nicolas J Cerf. Quantum search by local adiabatic evolution. *Physical Review A*, 65(4):042308, 2002.

Lifeng Shen and James Kwok. Non-autoregressive conditional diffusion models for time series prediction. In *International Conference on Machine Learning*, pp. 31016–31029. PMLR, 2023.

Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391, 2003.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Chenxi Wang, Linxiao Yang, Zhixian Wang, Liang Sun, and Yi Wang. A non-isotropic time series diffusion model with moving average transitions. In *Forty-second International Conference on Machine Learning*, 2025.

Stanisław Weglarczyk. Kernel density estimation and its application. In *ITM web of conferences*, volume 23, pp. 00037. EDP Sciences, 2018.

Annette Witt, Jürgen Kurths, and A Pikovsky. Testing stationarity in time series. *physical Review E*, 58(2):1800, 1998.

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.

Yanfeng Yang, Siwei Chen, Pingping Hu, Zhaotong Shen, Yingjie Zhang Zhang, Zhuoran Sun, Shuai Li, Ziqi Chen, and Kenji Fukumizu. Conditionally whitened generative models for probabilistic time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 53263–53298, 2026.

Zihao Yao, Qi Zheng, Jiankai Zuo, and Yaying Zhang. Towards a unified generative model for scarce time series with domain experts. *arXiv preprint arXiv:2606.15172*, 2026a.

Zihao Yao, Jiankai Zuo, and Yaying Zhang. Dim-ts: Bridge the gap between selective state space models and time series for generative modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 16137–16144, 2026b.

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. In *International Conference on Learning Representations*, volume 2024, pp. 41582–41610, 2024.

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 NOTATIONS

For clarity and ease of understanding, the main notations throughout our manuscript are summarized and presented in Table 4.

Table 4: The notations commonly used in the manuscript.

Notation	Definition
X_0	Input time Series of diffusion models
X_1	Prior noise of diffusion process
τ	Time step of diffusion process
α_τ	Noise schedule that control noise level
$\mathcal{D}$	Distribution discrepancy measurement
$v(\alpha)$	Distribution evolution rate of α on $\mathcal{D}$
ψ_{IAAT}	Integrated Absolute Autocorrelation Time
T_d	Steps for simulating forward process
T_s	Denoising diffusion steps
F_X	Cumulative distribution function of v
T	Time series window length
Σ_X^t	Covariance matrix at temporal index t
$\Sigma(\Omega)$	Improved covariance matrix of prior
L	Lower-triangular Cholesky factor of $\Sigma(\Omega)$

B ALGORITHM OF ANTS

Algorithm 1 presents the pseudocode for obtaining the adaptive noise schedule for a given dataset. The procedure is independent of model parameters and only needs to be performed once prior to training the diffusion model. Notably, once the procedure is established, it can also be flexibly adapted to other measurements that quantify differences in data distribution, enabling appropriate biased degradation throughout diffusion process.

C DERIVATION

In this section, we provide a theoretical foundation for the proposed adaptive diffusion framework. We establish propositions and theorems that formalize the relationship between diffusion degradation, noise scheduling and temporal prior. **Proposition 1** shows that autocorrelation and ψ_{IAAT} gradually decrease with coefficient α_τ, providing a theoretical basis for using IAAT as a proxy for the degradation of temporal information. Building on it, **Theorem 1** presents that the inverse CDF achieves a minimax degradation between consecutive steps, offering the justification for adaptive schedule that achieves optimal traceability. **Proposition 2** formally motivates the use of data-dependent covariance statistics to construct the temporal prior. **Theorem 2** further shows that temporal-aware diffusion is exactly equivalent to standard diffusion process in the whitened space, allowing the prior to incorporate dataset statistics without changing the fundamental formulation. Finally, **Proposition 3** establishes invariance of KL divergence under invertible whitening transformation, ensuring that the change of representation does not alter underlying objective.

C.1 MONOTONIC DEGRADATION OF IAAT

Proposition 1. *Consider the forward diffusion process $X_\tau = \alpha_\tau X_0 + \sqrt{1-\alpha_\tau^2}\epsilon$, where X_0 is a second-order stationary time series with unit variance, $\epsilon \sim N(0, I)$ is Gaussian noise. Assume that the autocorrelation function of X_0 is absolutely summable. Then, for any nonzero temporal lag k, the autocorrelation coefficient of the diffused time series satisfies:*

$$\rho_k(\tau) = \alpha_\tau^2 \rho(0) \tag{13}$$

Consequently, the IAAT of the diffused time series is given by:

$$\psi_{\text{IAAT}}(X_\tau) = 1 + \alpha_\tau^2[\psi_{\text{IAAT}}(X_0) - 1]. \tag{14}$$

Algorithm 1 Adaptive optimization of ANTS on basis of IAAT

Require: Dataset $\mathcal{M} = \{X\}_{i=1}^{|\mathcal{M}|}$, RBF kernel of Gaussian Process $\mathcal{K}(i,j)$, number of steps T_d for simulating forward process, number of denoising steps T_s for diffusion model, complementary schedule α^m.

Output: Lower-triangular factor L for prior, and noise schedule α_τ

▷ *Obtain prior factor L with scale information and temporal dynamics* ◁
Initialize: $S_C \leftarrow \{\}$
for each i in $\{1, 2, \ldots, |\mathcal{M}|\}$ **do**
  Calculate X covariance: $\Sigma_X = [\Sigma_X^1, \ldots, \Sigma_X^T]$
  Store sample covariance: $S_C \leftarrow \Sigma_X$
Calculate dataset averaged covariance: $\Sigma_X = \text{mean}(S_C)$
Inject Gaussian Process: $\Sigma_X^{i,j} = \Sigma_X^{i,j} + \mathcal{K}(i,j)$
Obtain prior factor: $L = \text{Cholesky Decomposition}(\Sigma_X)$

▷ *Obtain dataset-specific noise schedule α_τ* ◁
▷ *Simulate forward distribution evolution* ◁
Initialize: $S_D \leftarrow \{\}$
Construct α_τ with uniformly spaced 1d-grid: $\alpha_\tau = 1 - \tau, \tau \in \{0, 1/T_d, \ldots, 1\}$
for each i in $\{1, 2, \ldots, |\mathcal{M}|\}$ **do**
  Initialize: $S_X \leftarrow \{\}$
  Draw noise ϵ from Gaussian distribution $N(0, I)$
  for each τ in $\{0, 1/T_d, \ldots, 1\}$ **do**
    Obtain diffused time series: $X_\tau = \alpha_\tau X_0 + \sigma_\tau L\epsilon$
    Calculate IAAT of data: $\psi_{\text{IAAT}}(X_\tau) = 1 + 2\sum_{k=1}^{\infty} |\rho_k|$
    Store sample at different stage: $S_X \leftarrow \psi_{\text{IAAT}}(X_\tau)$
  Store diffused sample: $S_D \leftarrow S_{X^i}$
Obtain dataset diffusion evolution trajectory with τ: $\psi_{\text{IAAT}}(\tau) = \text{mean}(S_D)$

▷ *Resample schedule with linear degradation* ◁
Calculate normalization constant: $Z = \max(\psi_{\text{IAAT}}) - 1$
Define cumulative distribution function (CDF): $F_X(\tau) = \frac{1}{Z}(\psi_{\text{IAAT}}(1-\tau) - 1)$
Interpolate CDF F_X with uniformly spaced 1d-grid: $\{0, 1/T_s, \ldots, 1\}$
for each τ in $\{0, 1/T_s, \ldots, 1\}$ **do**
  Obtain adaptive noise schedule: $\alpha_\tau = F_X^{-1}(\tau)$
  Obtain mixed schedule: $\alpha_\tau = \frac{1}{2}(\alpha_\tau + \alpha_\tau^m)$

Therefore, if α_τ monotonically decreases as the diffusion process proceeds, the IAAT monotonically decreases toward its theoretical minimum of 1.

Proof. For any nonzero lag k the covariance between two temporal positions of the diffused time series can be written as:

$$\text{Cov}\left(X_\tau^t, X_\tau^{t+k}\right) = \text{Cov}\left(\alpha_\tau X_0^t + \sqrt{1-\alpha_\tau^2}\epsilon_t, \alpha_\tau X_0^{t+k} + \sqrt{1-\alpha_\tau^2}\epsilon_{t+k}\right). \tag{15}$$

Since the Gaussian noise is temporally independent and independent of the original time series, the cross-covariance terms involving ϵ vanish for $k > 0$. Thus, we have:

$$\text{Cov}(X_\tau^t, X_\tau^{t+k}) = \alpha_\tau^2 \,\text{Cov}(X_0^t, X_0^{t+k}). \tag{16}$$

As both X_0 and injected noise have unit variance, satisfying:

$$\text{Var}\left(X_\tau^t\right) = \alpha_\tau^2 \,\text{Var}\left(X_0^t\right) + \left(1-\alpha_\tau^2\right), \quad \text{Var}\left(\epsilon_t\right) = \alpha_\tau^2 + \left(1-\alpha_\tau^2\right) = 1. \tag{17}$$

Therefore, the autocorrelation coefficient at lag k is:

$$\rho_k(\tau) = \frac{\text{Cov}\left(X_\tau^t, X_\tau^{t+k}\right)}{\text{Var}\left(X_\tau^t\right)} = \alpha_\tau \frac{\text{Cov}\left(X_0^t, X_0^{t+k}\right)}{\text{Var}\left(X_0^t\right)} = \alpha_\tau^2 \rho_k(0). \tag{18}$$

Since $\alpha_\tau \geq 0$, substituting this relationship into IAAT yields:

$$\psi_{\text{IAAT}}(X_\tau) = 1 + 2\sum_{k=1}^{\infty} |\rho_k(\tau)| = 1 + 2\alpha_\tau^2 \sum_{k=1}^{\infty} |\rho_k(0)|. \tag{19}$$

As $\psi_{\mathrm{IAAT}}(0) = 2\sum_{k=1}^{\infty} |\rho_k(0)| + 1$, hence:

$$\psi_{\mathrm{IAAT}}(X_\tau) = 1 + \alpha_\tau^2[\psi_{\mathrm{IAAT}}(0) - 1]. \tag{20}$$

Since $\psi_{\mathrm{IAAT}}(0) \geq 1$, when α_τ decreases monotonically during diffusion, $\psi_{\mathrm{IAAT}}(\tau)$ also decreases monotonically. In the terminal limit $\alpha_\tau \to 0$, we obtain $\psi_{\mathrm{IAAT}}(\tau) \to 1$, which corresponds to temporally independent white noise.

□

C.2 Optimal Traceability

Theorem 1. *Let $g(\tau) = \psi_{IAAT}(0) - \psi_{IAAT}(\tau)$ denote the cumulative temporal-dependency degradation induced by the diffusion process, with $g(0) = 0$ and $g(1) = Z > 0$. For a diffusion process discretized into T_s steps, define the normalized cumulative degradation function (CDF) as $F_X(\tau) = g(\tau)/Z$ and construct the diffusion time points according to the inverse CDF as:*

$$\tau_i = F_X^{-1}(i/T_s), \quad i = \{0, 1, \ldots, T_s\}. \tag{21}$$

Then, every adjacent diffusion interval produces an identical amount of IAAT degradation:

$$\Delta_i = g(\tau_i) - g(\tau_{i-1}) = \frac{Z}{T_s}. \tag{22}$$

Moreover, among all possible T_s-step discretizations of the diffusion interval, this schedule minimizes the maximum IAAT degradation between any two consecutive diffusion steps. Specifically,

$$\min_{\{\tau_i\}} \max_i |\Delta_i| = \frac{Z}{T_s}. \tag{23}$$

Proof. By constructing $F_X(\tau_i) = \frac{i}{T_s}$, it follows that:

$$g(\tau_i) = \frac{iZ}{T_s}. \tag{24}$$

Taking difference between $g(\tau_i)$ and $g(\tau_{i-1})$ gives:

$$\Delta_i = g(\tau_i) - g(\tau_{i-1}) = \frac{iZ}{T_s} - \frac{(i-1)Z}{T_s} = \frac{Z}{T_s}. \tag{25}$$

Thus, the proposed inverse-CDF discretization assigns exactly the same amount of IAAT degradation to every diffusion step.

We next prove its minimax optimality. Consider an arbitrary discretization satisfying:

$$0 = \tilde{\tau}_0 < \tilde{\tau}_1 < \cdots < \tilde{\tau}_{T_s} = 1. \tag{26}$$

Because the intervals form a complete partition of the diffusion trajectory, their degradation values satisfy:

$$\sum_{i=1}^{T_s} \tilde{\Delta}_i = \sum_{i=1}^{T_s} g(\tilde{\tau}_i) - g(\tilde{\tau}_{i-1}) = g(1) - g(0) = Z. \tag{27}$$

Suppose that:

$$\max_i \tilde{\Delta}_i < \frac{Z}{T_s}. \tag{28}$$

Then the sum of all T_s degradation intervals would satisfy:

$$\sum_{i=1}^{T_s} \tilde{\Delta}_i < T_s \frac{Z}{T_s} = Z, \tag{29}$$

which contradicts the fact that their sum is exactly Z. Therefore, every possible T_s-step discretization must satisfy:

$$\max_i \tilde{\Delta}_i \geq \frac{Z}{T_s}. \tag{30}$$

The proposed inverse-CDF schedule achieves $\Delta_i = \frac{Z}{T_s}$ for every i. Hence, its maximum interval degradation is exactly the theoretical lower bound:

$$\min_{\{\tau_i\}} \max_i |\tilde{\Delta}_i| = \frac{Z}{T_s}. \tag{31}$$

Thus, the proposed schedule is minimax-optimal with respect to the maximum IAAT degradation per diffusion step. □

C.3 Moment-Matched Gaussian Prior

Proposition 2. *Let P_X be an arbitrary data distribution with finite mean $\mu_X = \mathbb{E}_{P_X}[X]$ and positive-definite covariance matrix $\Sigma_X = \mathrm{Cov}_{P_X}(X)$. Consider the family of Gaussian distributions $Q_{\mu,\Sigma} = \mathcal{N}(\mu, \Sigma)$. Then the Gaussian distribution that minimizes the Kullback–Leibler (KL) divergence from the data distribution is the moment-matched Gaussian:*

$$Q^* = \mathcal{N}(\mu_X, \Sigma_X). \tag{32}$$

That is:

$$(\mu^*, \Sigma^*) = \arg\min_{\mu,\Sigma} D_{\mathrm{KL}}\left(P_X \| \mathcal{N}(\mu, \Sigma)\right) = (\mu_X, \Sigma_X). \tag{33}$$

Proof. By definition, we have:

$$D_{\mathrm{KL}}(P_X \| Q_{\mu,\Sigma}) = \mathbb{E}_{P_X}\left[\log P_X(X) - \log Q_{\mu,\Sigma}(X)\right]. \tag{34}$$

The first term $\mathbb{E}_{P_X}[\log P_X(X)]$ does not depend on μ or Σ. Therefore, minimizing the KL divergence is equivalent to minimizing the negative expected log-likelihood under the Gaussian distribution. The log-density of $Q_{\mu,\Sigma}$ is:

$$\log Q_{\mu,\Sigma}(X) = -\frac{1}{2}\left[(X-\mu)^\top \Sigma^{-1}(X-\mu) + \log|\Sigma| + d\log(2\pi)\right]. \tag{35}$$

Consequently, minimizing the KL divergence is equivalent to minimizing:

$$J(\mu, \Sigma) = \frac{1}{2}\mathbb{E}_{P_X}\left[(X-\mu)^\top \Sigma^{-1}(X-\mu)\right] + \frac{1}{2}\log|\Sigma|. \tag{36}$$

For a fixed covariance matrix Σ, the first-order condition with respect to μ is:

$$\frac{\partial J}{\partial \mu} = -\Sigma^{-1}\left(\mathbb{E}[X] - \mu\right). \tag{37}$$

Setting this derivative to zero gives:

$$\mu^* = \mathbb{E}[X] = \mu_X. \tag{38}$$

Substituting $\mu = \mu_X$, similarly, we have:

$$\Sigma^* = \Sigma_X. \tag{39}$$

Hence, the KL-optimal Gaussian approximation to an arbitrary data distribution is the Gaussian distribution whose mean and covariance exactly match those of the data. □

C.4 Diffusion Equivalence under Conditional Whitening

Theorem 2. *Let the temporal-aware forward diffusion process be defined as:*

$$d(X_\tau - \mu) = -\frac{1}{2}\beta_\tau (X_\tau - \mu)d\tau + \sqrt{\beta_\tau} L dW_\tau, \tag{40}$$

where L is a full-rank lower-triangular matrix satisfying $\Sigma = LL^\top$. Define the whitened variable $X_\tau^{\mathrm{CW}} = L^{-1}(X_\tau - \mu)$. Then X_τ^{CW} follows the standard diffusion process:

$$dX_\tau^{\mathrm{CW}} = -\frac{1}{2}\beta_\tau X_\tau^{\mathrm{CW}} d\tau + \sqrt{\beta_\tau} dW_\tau. \tag{41}$$

Therefore, the temporal-aware diffusion process in the original data space is exactly equivalent to standard diffusion in the whitened space, and the original data can be recovered through the inverse transformation:

$$X_\tau = L X_\tau^{\mathrm{CW}} + \mu. \tag{42}$$

Proof. Define the affine transformation:

$$X_\tau - \mu = L X_\tau^{\mathrm{CW}}. \tag{43}$$

Since L and μ are independent of the diffusion time τ, differentiating both sides gives:

$$d(X_\tau - \mu) = L dX_\tau^{\mathrm{CW}}. \tag{44}$$

Substituting the forward diffusion process yields:

$$L dX_\tau^{\mathrm{CW}} = -\frac{1}{2}\beta_\tau (X_\tau - \mu) d\tau + \sqrt{\beta_\tau} L dW_\tau. \tag{45}$$

Since $X_\tau - \mu = L X_\tau^{\mathrm{CW}}$, we obtain:

$$L dX_\tau^{\mathrm{CW}} = -\frac{1}{2}\beta_\tau L X_\tau^{\mathrm{CW}} d\tau + \sqrt{\beta_\tau} L dW_\tau. \tag{46}$$

Since L is full rank, L^{-1} exists. Multiplying both sides by L^{-1} gives:

$$dX_\tau^{\mathrm{CW}} = -\frac{1}{2}\beta_\tau X_\tau^{\mathrm{CW}} d\tau + \sqrt{\beta_\tau} dW_\tau. \tag{47}$$

This is exactly the standard diffusion SDE in the whitened space. Therefore, the temporal covariance encoded by L is completely absorbed into the linear transformation, while the diffusion dynamics in the transformed space retain the standard isotropic form.

Then, the inverse transformation follows:

$$X_\tau = L X_\tau^{\mathrm{CW}} + \mu. \tag{48}$$

Hence, solving the standard diffusion process in the whitened space and transforming the generated samples back to the original space is mathematically equivalent to performing diffusion directly under the temporal-aware covariance structure. □

C.5 KL Invariance under Whitening

Proposition 3. *Let P_X and Q_X be two distributions on X, and let $Y = L^{-1}(X - \mu)$ be an invertible affine transformation with nonsingular L. Denote the corresponding transformed distributions by P_Y and Q_Y. Then:*

$$D_{\mathrm{KL}}(P_X \| Q_X) = D_{\mathrm{KL}}(P_Y \| Q_Y). \tag{49}$$

Proof. The Jacobian determinant of the inverse transformation is:

$$\left| \frac{\partial X}{\partial Y} \right| = |\det(L)|. \tag{50}$$

Therefore, the transformed densities satisfy:

$$p_Y(y) = p_X(Ly + \mu)|L|, \quad q_Y(y) = q_X(Ly + \mu)|L|, \tag{51}$$

The KL divergence in the transformed space is:

$$D_{\mathrm{KL}}(P_Y \| Q_Y) = \int p_Y(y) \log \frac{p_Y(y)}{q_Y(y)} dy. \tag{52}$$

Substituting the transformed densities gives:

$$D_{\mathrm{KL}}(P_Y \| Q_Y) = \int p_X(Ly + \mu)|L| \log \frac{p_X(Ly + \mu)|L|}{q_X(Ly + \mu)|L|} dy. \tag{53}$$

Changing variables from y back to $x = Ly + \mu$, with $dx = |L| dy$, gives:

$$D_{\mathrm{KL}}(P_Y \| Q_Y) = \int p_X(x) \log \frac{p_X(x)}{q_X(x)} dx. \tag{54}$$

Therefore, the KL divergence between original data and whitened data is equivalent. □

D EXPERIMENT DETAILS

D.1 DATASET DESCRIPTION

We conduct generation and forecasting experiments to simulate diverse scenarios encountered in real-world downstream applications. To ensure consistency with previous works, we adopt widely used public datasets that are available online for evaluation. These datasets span a broad range of application domains and temporal resolutions, providing diverse temporal characteristics for a comprehensive evaluation of the proposed method. The dataset statistics are summarized in Table 5, and detailed descriptions are provided below:

- **ETT**: It monitors the oil temperature and six power load characteristics from 2016 to 2018. The ETTh1 and ETTm2 datasets are collected at hourly and 15-minute intervals, respectively.
- **Energy**: A UCI appliances energy prediction dataset with 28 features related to household energy consumption.
- **KDDCup**: This dataset consists of air quality indices (AQIs) for Beijing and London, used in the KDDCup 2018.
- **Wind:** It comprises wind farm generation sampled at 15-minute intervals across seasons spanning over 2020 to 2021, provided in PaddleSpatial.
- **ECL**: It collects the hourly electricity consumption of 321 residents from 2012 to 2014.
- **Traffic:** It collects the hourly occupancy rate of different lanes on San Francisco freeway from 2015 to 2016.

Table 5: The statistics of benchmark datasets.

Dataset	Samples	Variables	Granularity	Domain	Source
ETTh1	17420	7	1 hour	Energy	(Zhou et al., 2021)
Energy	19735	28	10 min	Energy	(Candanedo et al., 2017)
Wind	48673	7	15 min	Sensor	(PaddleSpatial)
KDDCup	10920	59	1 hour	Nature	(Godahewa et al., 2021)
ETTm2	34273	7	15 min	Energy	(Zhou et al., 2021)
ECL	26304	321	1 hour	Electricity	(Wu et al., 2021)
Traffic	17544	862	1 hour	Transport	(Wu et al., 2021)

D.2 EVALUATION METRICS

The evaluation metrics are selected according to the characteristics of each task, with complementary metrics adopted to assess different aspects of model performance and mitigate potential evaluation bias. For the time series generation task, we employ quantitative metrics from two perspectives: (1) the distributional similarity between synthetic and real time series, and (2) the fidelity of temporal and channel-wise dependencies (Yuan & Qiao, 2024; Weglarczyk, 2018; Van der Maaten & Hinton, 2008). Detailed interpretations and corresponding formulations are presented below.

1) Context-FID score. It replaces the original image-based FID metric with a time series representations learned by TS2Vec (Yuan & Qiao, 2024). Specifically, both synthetic and real time series are first encoded using a pre-trained TS2Vec model, after which the FID score is computed in the resulting representation space. A lower score indicates a closer alignment between the global and contextual distributions of synthetic and real time series, and thus correspond to higher-quality generation.

2) Discriminative score. It is trained on a post-hoc LSTM-based time series classifier (clf) in a supervised manner on the training sets constructed from synthetic and real data, where synthetic samples are labeled as 0 and real samples as 1. The metric is then computed on the corresponding test sets:

$$\left| \frac{\sum_{n=1}^{S}(0 = \text{clf}(x_n^f)) + \sum_{n=1}^{S}(1 = \text{clf}(x_n^r))}{2S} - 0.5 \right|, \tag{55}$$

where S is set length, x_n^r is real data, x_n^f is synthetic data.

3) Correlational score. We first estimate the covariance of ith and j^{th} channel of time series:

$$\text{cov}_{i,j} = \frac{1}{L}\sum_{t=1}^{L} x_i^t x_j^t - \left(\frac{1}{L}\sum_{t=1}^{L} x_i^t\right)\left(\frac{1}{L}\sum_{t=1}^{L} x_j^t\right). \tag{56}$$

Then the difference of covariance between the synthetic data and original data is computed by:

$$\frac{1}{10}\sum_{i,j}^{d}\left|\frac{\text{cov}_{i,j}^r}{\sqrt{\text{cov}_{i,i}^r}\,\text{cov}_{j,j}^r} - \frac{\text{cov}_{i,j}^f}{\sqrt{\text{cov}_{i,i}^f}\,\text{cov}_{j,j}^f}\right|. \tag{57}$$

While early forecasting methods primarily focus on predicting the conditional mean of future sequences given historical observations, they contribute little to uncertainty quantification. To provide a more comprehensive evaluation, we additionally employ probabilistic forecasting metrics, which assess the quality of predicted distributions and the model's ability to quantify uncertainty.

1) Quantile Interval Coverage Error (QICE) It quantifies the deviation between the proportion of true data contained within each quantile intervals (QIs) and the optimal proportion, which is $1/M$ for all intervals (Li et al., 2024). To compute QICE, we divide the generated y-samples into M quantile intervals with approximately equal sizes. Subsequently, quantile values corresponding to each QI boundary are determined. It can be expressed as:

$$\text{QICE} := \frac{1}{M}\sum_{m=1}^{M}\left|r_m - \frac{1}{M}\right|, \quad \text{where } r_m = \frac{1}{N}\sum_{n=1}^{N}\mathbb{1}_{y_n \geq \hat{y}_n^{\text{low}_m}} \cdot \mathbb{1}_{y_n \leq \hat{y}_n^{\text{high}_m}}. \tag{58}$$

$\mathbb{1}_{\text{condition}}$ is an indicator function, which evaluates to 1 if the condition holds, and 0 otherwise. The terms $\hat{y}_n^{\text{low}_m}$ and $\hat{y}_n^{\text{high}_m}$ represent the lower and upper boundaries of the mth quantile interval, respectively.

2) Continuous ranked probability score (CRPS) It measures the compatibility of a cumulative distribution function (CDF) F with an observation x as (Matheson & Winkler, 1976):

$$\text{CRPS}(F, x) = \int_{\mathbb{R}} \left(F(z) - \mathbb{1}\{x \leq z\}\right)^2 \mathrm{d}z, \tag{59}$$

where $\mathbb{1}_{z<q}$ is an indicator function. Employing the empirical CDF of F , i.e. $\hat{F}(z) = \frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\left\{x^{0,s} \leq z\right\}$ with S samples $x^{0,s} \sim F$ as a natural approximation of the predictive CDF, CRPS can be directly computed by samples from DDPM.

3) Mean Squared Error (MSE) It is a classic metric for evaluating time series forecasting performance, which measures the average squared error between the predicted and ground-truth values. A lower MSE indicates more accurate point forecasts and smaller prediction errors. It is defined as:

$$\text{MSE} = \frac{1}{T}\frac{1}{N}\sum_{i=1}^{T}\sum_{j=1}^{N}(y_j^i - \hat{y}_j^i)^2 \tag{60}$$

where y_j^i is the observed ground-truth value of jth variable at the ith time step, while $\hat{y}_j^i$ represents the corresponding model's predicted value.

4) Mean Absolute Error (MAE) Instead of squaring the prediction errors, it computes the absolute values, making it less sensitive to outliers than MSE. Similar to MSE, a lower MAE indicates more accurate forecasts. It is defined as:

$$\text{MAE} = \frac{1}{T}\frac{1}{N}\sum_{i=1}^{T}\sum_{j=1}^{N}|y_j^i - \hat{y}_j^i| \tag{61}$$

D.3 Statistics of Non-stationarity

Existing works provide different statistic metrics to quantify non-stationarity of time series. To support the rationality of adopting IAAT in our design, we replace it with Lag1AC that quantifies the

correlation between adjacent observations and VarAC that measures the variance of the autocorrelation over different lags (Witt et al., 1998; Lee et al., 2024). They are defined as follows:

$$\psi_{\text{Lag1AC}} = \rho_1 = \frac{\sum_{t=1}^{T-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}, \tag{62}$$

where x_t and x_{t+1} are consecutive timestamps of time series x.

$$\psi_{\text{VarAC}} = \frac{1}{m}\sum_{k=1}^{m}(\rho_k - \bar{\rho})^2, \tag{63}$$

where $\bar{\rho}$ is the mean of the autocorrelation over considered m lags. The corresponding forecasting results on probabilistic and deterministic metrics are presented in Table 6 and 7, respectively.

We observe that our method is robust to different statistical measurements. It can be mainly attributed to our carefully designed noise schedule, which induces a linear degradation along a specific distribution trajectory. It enables diffusion models to explicitly preserve specific temporal attributes during the corruption process, facilitating more effective recovery from degraded time series during the reverse process. While IAAT achieves the best performance, we hypothesize that the advantage stems from comprehensive consideration of autocorrelation across all lags, whereas Lag1AC considers only a single lag and VarAC focuses on variance change rather than exact values.

Table 6: Forecasting performance comparison of non-stationarity statistics on CRPS and QICE.

Model	Dataset	CRPS (↓)				QICE (↓)			
		Raw	Lag1AC	VarAC	IAAT	Raw	Lag1AC	VarAC	IAAT
DiffusionTS	ETTh1	0.608	0.604	0.604	0.597	6.505	6.337	6.209	6.267
	ETTm2	1.098	0.954	0.936	0.947	10.456	9.019	8.878	8.900
	ECL	0.615	0.628	0.627	0.617	4.279	4.891	4.898	4.361
	Traffic	0.656	0.624	0.628	0.625	6.960	6.239	6.440	6.228
SSSD	ETTh1	0.548	0.539	0.528	0.537	5.219	5.310	3.950	4.664
	ETTm2	0.795	0.658	0.808	0.626	8.850	2.638	5.793	3.787
	ECL	0.374	0.362	0.356	0.364	5.497	4.997	4.464	5.317
	Traffic	0.610	0.457	0.435	0.409	13.446	7.618	7.847	6.148
TimeDiff	ETTh1	0.467	0.483	0.489	0.461	14.906	14.326	14.392	14.450
	ETTm2	0.309	0.315	0.322	0.310	11.436	13.759	13.942	12.779
	ECL	0.750	0.738	0.745	0.739	15.501	15.146	15.303	15.096
	Traffic	0.772	0.756	0.765	0.760	15.423	14.902	15.105	14.929

Table 7: Forecasting performance comparison of non-stationarity statistics on MSE and MAE.

Model	Dataset	MSE (↓)				MAE (↓)			
		Raw	Lag1AC	VarAC	IAAT	Raw	Lag1AC	VarAC	IAAT
DiffusionTS	ETTh1	1.103	1.092	1.089	1.078	0.781	0.780	0.780	0.772
	ETTm2	2.523	2.088	2.014	2.100	1.299	1.139	1.113	1.133
	ECL	1.058	1.079	1.073	1.053	0.839	0.846	0.844	0.836
	Traffic	1.450	1.438	1.438	1.437	0.805	0.800	0.800	0.800
SSSD	ETTh1	0.828	0.827	0.796	0.814	0.697	0.686	0.680	0.688
	ETTm2	1.801	1.507	2.092	1.376	1.053	0.909	1.045	0.839
	ECL	0.469	0.452	0.443	0.454	0.481	0.477	0.466	0.477
	Traffic	1.289	1.066	1.052	0.984	0.650	0.570	0.517	0.511
TimeDiff	ETTh1	0.532	0.592	0.596	0.529	0.482	0.506	0.510	0.480
	ETTm2	0.269	0.272	0.276	0.256	0.332	0.332	0.335	0.325
	ECL	0.869	0.870	0.872	0.867	0.764	0.762	0.765	0.763
	Traffic	1.354	1.351	1.355	1.353	0.786	0.782	0.787	0.786

E VISUALIZATION RESULTS

To provide an intuitive understanding of the proposed method, we visualize the non-stationarity curves and forward diffusion processes across different datasets. As shown in Figure 7, there is generally substantial gap between the conventional cosine/linear schedules and the adaptive schedule. The original time series may prematurely degrade into noise with low non-stationarity even at relatively low noise levels or diffusion steps, indicating that many steps are ineffective and may impair model capability. In contrast, ANTS adaptively allocates appropriate noise levels according to the denoising steps, improving model while achieving optimal performance with fewer steps. It potentially reduces both training and inference time cost. Notably, the cosine schedule is already close to the ideal linear degradation in a few cases (e.g. ECL). It explains the marginal improvement of ANTS in minor time series forecasting experiments.

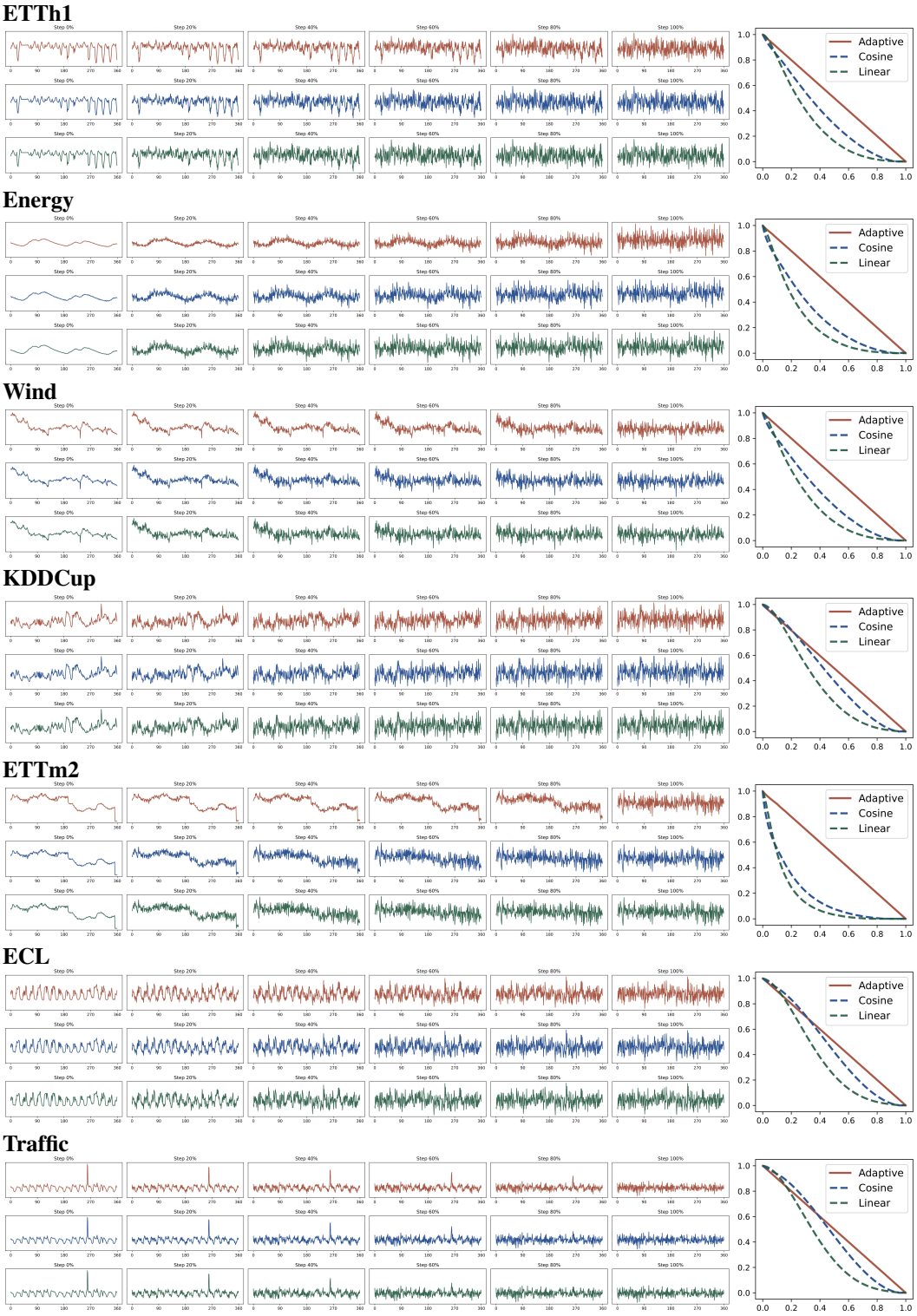

Figure 7: Overall visualization comparison of different schedules across datasets.