Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)

EVENTTSF: Event-Aware Non-Stationary Time Series Forecasting

Yunfeng Ge[1,2], **Ming Jin**[2*], **Yiji Zhao**[3], **Hongyan Li**[1], **Bo Du**[2], **Chang Xu**[4] and **Shirui Pan**[2*]

[1]Xidian University, China
[2]Griffith University, Australia
[3]Yunnan University, China
[4]Microsoft Research Asia, China

Abstract

Time series forecasting is vital in diverse sectors such as energy and transportation, where non-stationary dynamics are deeply intertwined with external events in other modalities such as texts. However, incorporating natural language-based external events to improve non-stationary forecasting remains largely unexplored, as most approaches still rely on a single modality, resulting in limited contextual knowledge and model underperformance. Enabling fine-grained multimodal interactions between temporal and textual data is challenged by two fundamental issues: (1) the gap in modeling interactions among discrete external events and continuous time series in a unified framework; (2) classical uniform diffusion timestep ignores event-induced non-stationary variability, leading to imbalanced denoising difficulty across diffusion stages. In this work, we propose event-aware non-stationary time series forecasting (**EVENTTSF**), an autoregressive diffusion framework that integrates historical time series and textual events via step-wise diffusion. To mitigate the imbalanced denoising difficulty of uniform timestep sampling, EVENTTSF uses an event-aware flow-matching timestep conditioned on event semantics. Extensive experiments on 7 synthetic and real-world datasets show that EVENTTSF outperforms 12 non-stationary time series forecasting baselines, achieving average gains of $\mathbf{41.3}\%$ in probabilistic forecasting and $\mathbf{27.5}\%$ in deterministic forecasting across all evaluation metrics.

1 Introduction

Time series forecasting is critical in sectors such as energy, transportation, and meteorology [Liang *et al.*, 2024b], where accurate forecasting enables effective decision making and resource management. However, real-world forecasting faces longstanding challenges due to non-stationarity and distribution shifts [Box and Jenkins, 1976], where the underlying time series distribution changes over time. While recent advances, including normalization techniques (e.g., RevIN [Kim *et al.*, 2021]), model decomposition (e.g., Koopa [Liu *et al.*, 2023b]), and online learning (e.g., OneNet [Wen *et al.*, 2023]), have partially addressed these issues, they typically operate on a single modality, **overlooking that the change is entangled with external events from other modalities**, particularly texts. This leaves valuable contextual knowledge largely unexploited, resulting in limited forecasting performance as evidenced in Figure 1 (a). The distinct segment-wise temporal distributions triggered by event type in Figure 1 (b) highlight the need for event-aware forecasting. To capture interactions between discrete textual events and continuous time series, we propose EVENTTSF, an autoregressive generative architecture (Figure 1 (c)).

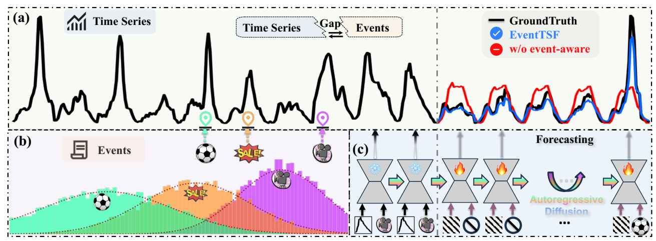

Figure 1: **Motivation.** Events induce distribution shifts. **(a)** Event-unaware forecasting degrades. **(b)** Events induce distinct temporal distributions. **(c)** EVENTTSF jointly models events and time series.

In many real-world systems, upcoming events such as holidays or planned promotions are known in advance of the forecast horizon, providing actionable exogenous priors to anticipate distribution shifts that cannot be inferred from historical observations alone. Recent efforts have explored multimodal integration for time series forecasting using large language models via exogenous information and agentic interactions [Qiu *et al.*, 2026; Jiang *et al.*, 2025]. However, these methods struggle to model non-stationarity driven by evolving temporal dynamics and complex event interactions. **First**, existing methods utilize general and coarse-grained textual context for forecasting [Zhang *et al.*, 2025b], but struggle to capture fine-grained event-induced distribution shifts. **Second**, existing approaches rely on a deterministic event representation and produce point forecasts, limiting their ability to model event-induced uncertainty under non-stationary set-

*Correspondence to Ming Jin (mingjinedu@gmail.com) and Shirui Pan (s.pan@griffith.edu.au).

tings. Recent advances in flow matching and autoregressive generation offer some promising tools for event-aware forecasting [Lipman *et al.*, 2022; Liu *et al.*, 2022a].

However, two key challenges remain unresolved for multimodal event-aware forecasting: ❶ **Fine-grained temporal-event interaction**. Existing methods [Narasimhan *et al.*, 2024] rely on static global "meta context", which fails to synchronize and align fine-grained discrete events with continuous time series at the segment level. In contrast, modeling fine-grained temporal and event interactions helps capture cascading and lagged dependencies between time series and event dynamics. Specifically, aligning chronologically discrete ordered event segments with the fixed-length time series windows enables step-wise event–series interaction, rather than conditioning on a single global context. ❷ **Event-agnostic timestep sampling**. Current diffusion-based time series forecasting models [Ye *et al.*, 2025; Yuan and Qiao, 2024] typically sample diffusion timesteps from a fixed uniform distribution. However, uniform sampling can skew denoising difficulty across diffusion stages, undermining stability. Denoising difficulty is closely tied to the signal-to-noise ratio (SNR) of the denoiser's learning target [Choi *et al.*, 2022]. External events can induce non-stationary shifts in magnitude and variance, yielding heterogeneous SNR and thus varied event-dependent denoising difficulty. Consequently, event-agnostic uniform timestep sampling cannot adapt to event-dependent dynamics, leading to imbalanced training difficulty and unstable denoiser optimization.

To address these challenges, we propose EVENTTSF, a diffusion-based autoregressive architecture for event-aware non-stationary time series forecasting. EVENTTSF jointly models historical time series and textual events to generate probabilistic forecasts under non-stationary dynamics, maintaining a learnable multimodal hidden state at each step and incrementally incorporating event information. To mitigate event-agnostic uniform diffusion timestep sampling, we introduce an event-aware mechanism that learns an event-conditioned reparameterization, mapping a base uniform sample to an event-dependent timestep via learnable distribution parameterized by event-conditioned networks. This adaptive sampling strategy balances denoising difficulty across diffusion stages. The implementation code is available[1]. Our key contributions are summarized as follows:

- **Paradigm Reformulation**. We formulate event-aware non-stationary time series forecasting as a paradigm that incorporates multimodal event information to better address distribution shifts in real-world forecasting.
- **Methodological Innovation**. We develop EVENTTSF, an autoregressive diffusion model that enables fine-grained event–series alignment and mitigates denoising imbalance via event-aware timestep sampling.
- **Superior Performance**. EVENTTSF consistently improves forecasting performance across diverse scenarios, improving probabilistic forecasting by 41.3% and deterministic forecasting by 27.5% on average overall.

[1] https://github.com/WinfredGe/EventTSF

2 Definition

Given a univariate time series $\mathbf{X}_{1:L} = \{x_l\}_{l=1}^{L}$ of length L, we define a set of N chronologically ordered fine-grained textual event embeddings $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_s, \dots, \mathbf{c}_N\}$. Each event embedding $\mathbf{c}_s$ corresponds to an aligned non-overlapping consistent time segment $\mathbf{x}_s = \{x_l\}_{l=i_s}^{j_s} \subset \mathbf{X}_{1:L}$, where i_s and j_s denote the start and end timestamps of event segment s. The multimodal time series datasets is defined as $\mathcal{D} = \{(\mathbf{x}_s, \mathbf{c}_s)\}_{s=1}^{N}$, satisfying the constraints: (1) $\bigcup_{s=1}^{N} \mathbf{x}_s \subseteq \mathbf{X}_{1:L}$(complete coverage), and (2) $\mathbf{x}_s \cap \mathbf{x}_{s'} = \emptyset$ for all $s \neq s'$(non-overlapping segments).

We address event-aware non-stationary time series forecasting, where the underlying time series distribution exhibits event-conditioned shifts across different event categories,

$$P(\mathcal{X} \mid \mathbf{c}_i) \neq P(\mathcal{X} \mid \mathbf{c}_j), \quad \text{if} \quad \mathbf{c}_i \neq \mathbf{c}_j.$$

Given p historical time series and event pairs $\mathcal{H}_{\mathbf{x}} = \{\mathbf{x}_l\}_{l=s-p+1}^{s}$ and $\mathcal{H}_{\mathbf{c}} = \{\mathbf{c}_l\}_{l=s-p+1}^{s}$, and q future event descriptions $\mathcal{F}_{\mathbf{c}} = \{\mathbf{c}_l\}_{l=s+1}^{s+q}$, the objective is to forecast the following q future time series predictions $\hat{\mathcal{F}}_{\mathbf{x}} = \mathcal{G}_\theta(\mathcal{H}_{\mathbf{x}}, \mathcal{H}_{\mathbf{c}}, \mathcal{F}_{\mathbf{c}})$, where $\hat{\mathcal{F}}_{\mathbf{x}}$ is the estimated $\mathcal{F}_{\mathbf{x}} = \{\mathbf{x}_l\}_{s+1}^{s+q}$; $\mathcal{G}_\theta$ denotes an event-aware forecasting model that captures non-stationary dynamics conditioned on external events.

3 Methodology

As shown in Figure 2, we introduce EVENTTSF, an autoregressive diffusion architecture designed for event-aware non-stationary time series forecasting. EVENTTSF addresses events and time series synchronization and joint modeling. EVENTTSF integrates event-aware diffusion timestep, allowing event information to modulate the denoising process.

3.1 Multimodal Autoregressive Diffusion

Event Synchronization. Discrete events often exert influence on continuous time series in real-world forecasting. For example, traffic demand often surges during public celebrations; network bandwidth usage can spike during highly anticipated live broadcasts; and product sales rise during planned promotional campaigns. However, due to the lack of an effective joint representation, continuous time series and stochastic discrete events have been studied as two separate research topics: time series and temporal point process. To bridge this modality gap, we propose event synchronization, which constructs a unified event-series representation under a shared temporal granularity. Specifically, as illustrated in the left part of Figure 2, we partition the time series into non-overlapping segment $\mathbf{x}_s$. according to selected window size, and align each segment with an event embedding $\mathbf{c}_s$ that occurs within the same interval. If no event occurs in a segment window, we assign a NULL event. Synchronized event–time series pairs $\mathcal{D} = \{(\mathbf{x}_s, \mathbf{c}_s)\}_{s=1}^{N}$ are then used as inputs to our autoregressive diffusion model. Specifically, this model is trained on the synchronized multimodal segments produced by the forward flow matching process in Section 3.2.

Autoregressive Diffusion. The integration of autoregressive modeling and diffusion processes has proven effective for temporal data modeling [Chen *et al.*, 2024]. We design an

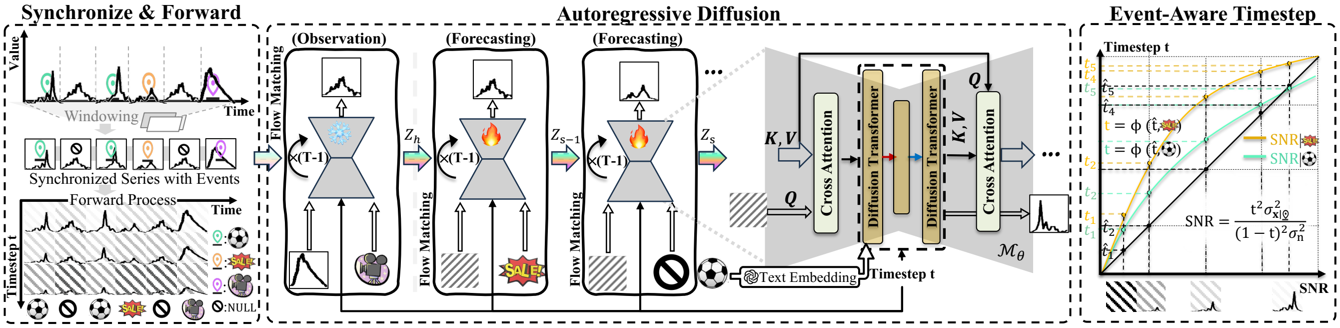

Figure 2: Overview of EVENTTSF. **(Left)** Continuous time series is synchronized with discrete events via fixed-length windowing. The forward diffusion process injects noise with event-aware timestep. **(Middle)** An autoregressive diffusion architecture jointly models events and time series. At each step, it conditions on observed time series history and aligned event embedding, and progressively denoising future segments in an autoregressive manner. **(Right)** Since SNR characterizes denoising difficulty and varies across events, we replace uniform timestep sampling by reparameterizing the uniformly sampled base timestep $\hat{t}$ into an event-aware timestep $t = \phi(\hat{t}, \mathbf{c})$.

autoregressive diffusion model EVENTTSF for event-aware time series forecasting tasks, as illustrated in the middle of Figure 2. This model regards the probabilistic time series forecasting as an autoregressive fine-grained segments generation. Specifically, the model generates the segements $\mathbf{x}_s$ at forecasting step s via a conditional diffusion denoising process. This denoising is initialized by a latent state $\mathbf{Z}_h$ that summarizes historical time series and event observations. The hidden state $\mathbf{Z}_s$ is updated at each step in an autoregressive manner through interactions with temporal segments and event semantics across the forecasting horizon.

During the training stage, we autoregressively optimize a flow-matching denoiser $\mathcal{M}_\theta$ conditioned on latent state $\mathbf{Z}_{s-1}$, noised time series segments $\mathbf{x}_s^t$, event-aware diffusion timestep t, and textual event embedding $\mathbf{c}_s$ encoded by a frozen OpenAI *text-embedding-3* model. Specifically, at each autoregressive step s and diffusion timestep t, the flow matching denoiser learns the velocity field $\mathbf{v}_s^t$ and updated latent state $\mathbf{Z}_s$ from $\mathbf{v}_s^t, \mathbf{Z}_s = \mathcal{M}_\theta(\mathbf{Z}_{s-1}, \mathbf{x}_s^t, \mathbf{c}_s, t)$ conditioned on the latent state $\mathbf{Z}_{s-1}$ and the event representation $\mathbf{c}_s$. The denoiser $\mathcal{M}_\theta$ is shared across all autoregressive steps, while step specific dynamics are captured through the evolving latent state $\mathbf{Z}_s$ and event representation $\mathbf{c}_s$. This design enables fine-grained relationship modeling between synchronized temporal patterns and event semantics. The detailed training procedure is summarized in Algorithm 1.

During the inference stage, we first compute the historical state $\mathbf{Z}_h$ from the clean historical time series $\mathcal{H}_x$ and the corresponding historical events $\mathcal{H}_c$, which is denoted in Section 2. During the forecasting stage, the historical state $\mathbf{Z}_h$ serves as the initial state $\mathbf{Z}_0$. At each autoregressive step s, the model takes a noise input $\mathbf{n}_s$ and an exogenous event $\mathbf{c}_s$ to generate the denoised segment $\mathbf{x}_s$ via the learned flow matching denoiser $\mathcal{M}_\theta(\mathbf{Z}_{s-1}, \mathbf{n}_s, \mathbf{c}_s, t)$, coupled with an vanilla ODE solver. The architecture of the denoiser is diffusion transformer, and the sampling procedure is detailed in Algorithm 2. Overall, EVENTTSF autoregressively incorporates dynamic event descriptions, enabling flexible and event-aware forecasting in evolving environments.

3.2 Flow Matching with Event-Aware Timestep

Flow Matching. We use the flow matching [Liu *et al.*, 2022a; Ge *et al.*, 2025] to jointly train textual events and time series. As illustrated in the left part of Figure 2, the forward diffusion process injects noise according to an event-aware diffusion timestep t. At each autoregressive step s and diffusion timestep $t \in [0, 1]$, flow matching constructs data trajectories that transform samples from a simple prior distribution $p(\mathbf{x}_s^0) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ into the target data distribution $p(\mathbf{x}_s^1)$. An optimal transport path, realized through Rectified Flow [Liu *et al.*, 2022a], guides this transformation process:

$$\mathbf{x}_s^t = (1-t)\mathbf{x}_s^0 + t\mathbf{x}_s^1. \tag{1}$$

The instantaneous velocity vector field along this path $\mathbf{v}_s^t$ with respect to diffusion timestep t is given by:

$$\mathbf{v}_s^t = \frac{d\mathbf{x}_s^t}{dt} = \mathbf{x}_s^1 - \mathbf{x}_s^0. \tag{2}$$

A neural network $\mathbf{v}_\theta(\mathbf{x}_s^t, \mathbf{Z}_{s-1}, \mathbf{c}_s, t)$ approximates the target vector field $\mathbf{v}_s^t$. To enhance generation quality, $\mathbf{v}_\theta$ is designed to integrate the conditional information [Lipman *et al.*, 2022]. In our settings, this conditional information includes the historical state $\mathbf{Z}_{s-1}$ of the time series and the textual condition $\mathbf{c}_s$. During sampling, the model generates $\hat{\mathbf{x}}_s^1$ by first sampling noise $\mathbf{x}_s^0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and then solving the learned velocity field $\mathbf{v}_\theta$ via a numerical ODE solver.

Event-Aware Timestep. Real-world time series exhibit event-dependent dynamics, and different events can induce distinct signal magnitudes and variances, leading to heterogeneous denoising difficulty. Denoising difficulty is closely tied to the SNR of the denoiser's learning target [Choi *et al.*, 2022]; accordingly, we quantify denoising difficulty via SNR. Under the standard noise injection in Equation 1, the SNR at diffusion timestep t is given by:

$$\mathrm{SNR}(t; \mathbf{c}) = \frac{t^2\sigma^2_{\mathbf{x}|\mathbf{c}}}{(1-t)^2\sigma^2_{\mathbf{n}}}, \tag{3}$$

where $\sigma^2_{\mathbf{x}|\mathbf{c}}$ and $\sigma^2_{\mathbf{n}}$ denote the variances of the clean signal and noise, respectively. Notably, $\sigma^2_{\mathbf{x}|\mathbf{c}}$ can vary across event

Algorithm 1 Training

Require: Training samples $(\mathbf{x}_s, \mathbf{c}_s)$; number of time series and event pairs per sample S; learnable initial state $\mathbf{Z}_0$; flow matching model $\mathcal{M}_\theta$; the learnable event-aware reparameterization function ϕ; learning rate η; optimal transport path OT; Uniform distribution $\mathcal{U}(0,1)$.

Initialize predicted velocity set $\hat{\mathcal{V}} \leftarrow \{\}$.
Initialize groundtruth velocity set $\mathcal{V} \leftarrow \{\}$.
for $s = 1, \ldots, S$ **do**
  Sample $\hat{t} \sim \mathcal{U}(0,1)$
  $t \leftarrow \phi(\hat{t}, \mathbf{c}_s)$
  Define $\mathbf{x}_s^t = OT(\mathbf{x_s}, t)$ and $\mathbf{v}_s^t = \frac{d\mathbf{x}_s^t}{d\mathbf{t}}$
  Union Update $(\mathbf{Z}_s, \hat{\mathbf{v}}_s^t) = \mathcal{M}_\theta(\mathbf{Z}_{s-1}, \mathbf{x}_s^t, \mathbf{c}_s, t)$
  $\mathcal{V} \leftarrow \mathcal{V} \cup \{\mathbf{v}_s^t\}$
  $\hat{\mathcal{V}} \leftarrow \hat{\mathcal{V}} \cup \{\hat{\mathbf{v}}_s^t\}$
end for
$\mathcal{L} = \mathrm{MSE}(\mathcal{V}, \hat{\mathcal{V}})$
$\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$

conditions, implying that the effective denoising difficulty at the same t may differ under different events.

However, the diffusion timestep t is typically sampled from a condition-agnostic distribution, such as uniform distribution $t \sim \mathcal{U}(0,1)$. Such uniform sampling fails to allocate denoising difficulty appropriately; it ignores event-induced heterogeneity, resulting in uneven difficulty allocation for the denoiser learning. As illustrated in the right panel of Figure 2, SNR varies nonlinearly with the diffusion timestep t increase, indicating that denoising difficulty varies nonlinearly across different diffusion stages. Moreover, the timestep–SNR curve is event-dependent, resulting in heterogeneous denoising difficulty under different event conditions.

To address these issues, we reparameterize uniformly sampled base timestep via an event-conditioned mapping, such that the effective denoising difficulty is more evenly allocated in an event-modulated manner. Specifically, we map a base timestep $\hat{t} \in [0,1]$ drawn from a Uniform distribution $\hat{t} \sim \mathcal{U}(0,1)$ into an event-adaptive timestep $t \in [0,1]$, whose distribution follows a Beta law $t|\mathbf{c} \sim \mathrm{Beta}(\alpha(\mathbf{c}), \beta(\mathbf{c}))$:

$$t = \phi(\hat{t}, \mathbf{c}) = F^{-1}_{\mathrm{Beta}(\alpha(\mathbf{c}),\beta(\mathbf{c}))}(\hat{t}), \tag{4}$$

where ϕ is the event-aware reparameterization function parameterized by Θ, and $F^{-1}_{\alpha,\beta}(\cdot)$ is the inverse cumulative distribution function of a Beta distribution with parameters α and β. This design ensures: (i) $t \in [0,1]$, (ii) ϕ is strictly monotone increasing in $\hat{t}$, preserving the intrinsic ordering of diffusion steps, and (iii) the mapping shape can be flexibly adjusted through (α, β) to realize condition-dependent adaptive schedules.

The Beta distribution parameters are predicted from the event representation $\mathbf{c}$ via two lightweight linear networks, f_{Θ_α} and g_{Θ_β}, followed by $\mathrm{softplus}(\cdot)$ activation:

$$\begin{aligned} \alpha(\mathbf{c}) &= \mathrm{softplus}\big(f_{\Theta_\alpha}(\mathbf{c})\big) + \epsilon, \\ \beta(\mathbf{c}) &= \mathrm{softplus}\big(g_{\Theta_\beta}(\mathbf{c})\big) + \epsilon, \end{aligned} \tag{5}$$

Algorithm 2 Sampling (Forecasting)

Require: Initial state $\mathbf{Z}_0$; flow matching model $\mathcal{M}_\theta$;reparameterization function ϕ; historical p time series and event pairs $(\mathcal{H}_x, \mathcal{H}_c)$; evolving event sequence $\{\mathbf{c}_s\}_{s=p+1}^q = \mathcal{F}_\mathbf{c}$; Gaussian distribution $\mathcal{N}(0, \mathbf{I})$; total number of denoising steps T and its incremental $\Delta_{\hat{t}} = \frac{1}{T}$.

Initialize forecasting set $\hat{\mathcal{X}} \leftarrow \{\}$.
for $s = 1, \ldots, p$ **do**
  $(\mathbf{Z}_s, \_) = \mathcal{M}_\theta(\mathbf{Z}_{s-1}, \mathbf{x}_s, \mathbf{c}_s, t = 1)$
end for
for $s = p+1, \ldots, q$ **do**
  $\mathbf{n}_s^0 \sim \mathcal{N}(0, \mathbf{I})$
  for $t = \Delta_{\hat{t}}, 2\Delta_{\hat{t}}, \ldots, 1$ **do**
    $\Delta_t \leftarrow \phi(t - \Delta_{\hat{t}}, \mathbf{c}_s)$
    $(\mathbf{Z}_s, \hat{\mathbf{v}}_s^{\Delta_t}) = \mathcal{M}_\theta(\mathbf{Z}_{s-1}, \mathbf{n}_s^{\Delta_t}, \mathbf{c}_s, t)$
    $\mathbf{n}_s^t \leftarrow \mathbf{n}_s^{\Delta_t} + \mathrm{ODE}(\hat{\mathbf{v}}_s^{\Delta_t}, \Delta_t)$ ▷ Refinement
  end for
  $\hat{\mathbf{x}}_s \leftarrow \mathbf{n}_s^1$
  $\hat{\mathcal{X}} \leftarrow \hat{\mathcal{X}} \cup \{\hat{\mathbf{x}}_s\}$
end for
return $\hat{\mathcal{X}}$

where $f_{\Theta_\alpha}(\cdot)$ and $g_{\Theta_\beta}(\cdot)$ denote two separate lightweight learnable linear networks parameterized by Θ_α and Θ_β, respectively; $\mathrm{softplus}(\cdot)$ enforces $\alpha(\mathbf{c}), \beta(\mathbf{c}) > 0$; and $\epsilon > 0$ is a small constant added for numerical stability. By conditioning $\alpha(\mathbf{c}), \beta(\mathbf{c})$ on the event representation $\mathbf{c}$, the model adapts timestep conditioned on heterogeneous event semantics, thereby modulating the denoising process and enabling event-specific restoration dynamics for the denoising targets.

4 Experiments

We evaluate the performance of EVENTTSF and address core research questions. Experiments are conducted on 7 datasets spanning 12 models. Configurations follow NsDiff [Ye *et al.*, 2025] and standard TSF protocols [Wu *et al.*, 2023]. Results are averaged over three independent runs.

- **RQ1**: How effectively does EVENTTSF forecast non-stationary multimodal data in both deterministic and probabilistic settings?
- **RQ2**: When baselines can access event data, can they achieve performance comparable to EVENTTSF?
- **RQ3**: How effective are the components, textual event conditioning and event-aware timestep in enhancing the method's performance?
- **RQ4**: How does EVENTTSF's training efficiency compare to baselines and across forecasting horizons?
- **RQ5**: What insights can be gained from visualizations with vs. without events?

4.1 Experimental Settings

Datasets. The experiments are evaluated on one synthetic dataset and six real-world datasets spanning traffic, weather and atmospheric physics sector. These include Synthetic

Method	Metric	Synthetic	Atmosphere Physics	Traffic Public	Traffic News	Temperature–Rainfall Houston	Temperature–Rainfall San Fran.	Temperature–Rainfall New York	Avg.
CSDI	CRPS	0.5769	1.3840	0.4359	0.4471	0.5742	0.5986	0.8201	0.6910
	WQL	0.3174	0.7452	0.2341	0.2445	0.3149	0.3274	0.4509	0.3763
	MAE	0.8901	2.1477	0.7623	0.6585	0.8165	0.8905	1.1509	1.0452
TimeDiff	CRPS	0.6718	1.3876	0.6179	0.4977	0.8092	0.7194	1.0869	0.8272
	WQL	0.3380	0.7475	0.3111	0.2509	0.4067	0.3618	0.5456	0.4231
	MAE	0.6858	2.1343	0.6320	0.5114	0.8233	0.7335	1.1012	0.9459
TMDM	CRPS	0.1842	1.6422	0.2796	0.3059	0.5500	0.5532	0.6857	0.6001
	WQL	0.0983	0.8233	0.1393	0.2012	0.2978	0.2987	0.3723	0.3187
	MAE	0.2899	1.6568	0.3598	0.4480	0.7297	0.7305	0.9258	0.7344
NsDiff	CRPS	0.2101	1.2933	0.2825	0.4516	0.4924	0.6588	0.4585	0.5496
	WQL	0.1131	0.6794	0.1406	0.3680	0.2769	0.2785	0.3612	0.3168
	MAE	0.2657	1.5355	0.3508	0.6199	0.7015	0.6836	0.9451	0.7289
DiffusionTS	CRPS	0.6211	0.7151	0.5401	0.4437	0.6065	0.5453	0.8363	0.6154
	WQL	0.3361	1.1631	0.2377	0.2388	0.3273	0.2938	0.4470	0.4348
	MAE	0.8818	3.8218	0.5930	0.5988	0.8280	0.7314	1.1070	1.2231
EVENTTSF	CRPS	**0.0776**	**0.2857**	**0.1908**	**0.2943**	**0.2670**	**0.5183**	**0.2895**	**0.2747**
	WQL	**0.0395**	**0.1464**	**0.1006**	**0.1561**	**0.1405**	**0.2702**	**0.1532**	**0.1438**
	MAE	**0.0812**	**0.3544**	**0.2314**	**0.3561**	**0.3178**	**0.5869**	**0.3500**	**0.3254**

Table 1: Probabilistic forecasting results on seven event-aware datasets. We report CRPS↓, WQL↓, and MAE↓ (lower is better) for each dataset, and the final column summarizes the average performance across datasets. Best results are highlighted in **bold**.

Dataset, Atmospheric Physics-Weather Events Dataset, Traffic–Public Events Dataset [Liang *et al.*, 2024a], Temperature–Rainfall Events Dataset [Lee *et al.*, 2025] and Traffic–News Events Dataset [Wang *et al.*, 2024].

Evaluation Metrics. We use comprehensive metrics for different forecasting tasks. For *deterministic forecasting*, we use Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for evaluation. For *probabilistic forecasting*, we evaluate performance using Continuous Ranked Probability Score (CRPS) and Weighted Quantile Loss (WQL), and report MAE for point summaries.

Baselines. To evaluate EVENTTSF's effectiveness, we benchmark it against probabilistic diffusion forecasting models and deterministic deep time series forecasting models. For the probabilistic diffusion forecasting models, CSDI [Tashiro *et al.*, 2021], TimeDiff [Shen and Kwok, 2023], TMDM [Li *et al.*, 2024], DiffusionTS [Yuan and Qiao, 2024], and NsDiff [Ye *et al.*, 2025] are included. For the deterministic deep time series forecasting, we include representative models, including iTransformer [Liu *et al.*, 2023a], Koopa [Liu *et al.*, 2023b], NSTransformer [Liu *et al.*, 2022b], PatchTST [Nie *et al.*, 2022], TimesNet [Wu *et al.*, 2023], and other used baselines. NsDiff [Ye *et al.*, 2025], Koopa [Liu *et al.*, 2023b], and NSTransformer [Liu *et al.*, 2022b] are designed for the non-stationary time series forecasting.

4.2 Performance Comparison on Probabilistic and Deterministic Forecasting (RQ1)

EVENTTSF achieves superior performance across both deterministic and probabilistic forecasting tasks. Table 1 compares probabilistic forecasting performance across seven datasets using CRPS, WQL, and average MAE, where lower values indicate better performance. The results highlight EVENTTSF's superior performance across diverse event-aware forecasting settings, with average improvements of 41.3% and 27.5% across all datasets and metrics for probabilistic and deterministic forecasting, respectively. Specifically, EVENTTSF delivers notable gains, lowering the CRPS on the Synthetic dataset from TMDM's 0.1842 to 0.0776, a 58.4% improvement, and reducing the CRPS on the Atmosphere Physics dataset from DiffusionTS's 0.7151 to 0.2857, a 65.1% improvement. Table 2 reports deterministic forecasting results across datasets. On the Synthetic dataset, EVENTTSF achieves an MAE of 0.0812, an 74.1% reduction from Autoformer's 0.3141; on the averaged dataset results, EVENTTSF achieves an RMSE of 0.5846, denoting a 35.8% reduction from DLinear's 0.9088. Deterministic forecasting models perform competitively on the Traffic–News dataset, where the limited contextual relevance of long-form, low-information textual event representations significantly weakens the effectiveness of event-aware forecasting.

4.3 Event-Enhanced Baseline Performance (RQ2)

Most existing non-stationary time series forecasting baselines cannot natively incorporate external textual events; therefore, we augment both probabilistic and deterministic non-stationary time series forecasting baselines with the aligned external textual event embeddings for a fairer evaluation. For the probabilistic forecasting baselines, we replace the denoiser's transformer block with the crossformer block to enable event-conditioned cross-attention. For the deterministic forecasting baselines, we linearly project event embeddings

Method	Metric	Synthetic	Atmosphere Physics	Traffic Public	Traffic News	Temperature–Rainfall Houston	Temperature–Rainfall San Fran.	Temperature–Rainfall New York	Avg.
Autoformer	MAE	0.3141	2.0739	0.4296	0.3536	0.7127	0.7011	0.9930	0.7969
	MSE	0.2612	6.5844	0.5258	0.2144	0.8194	0.8168	1.5016	1.5319
	RMSE	0.5161	2.5656	0.7294	0.4673	0.9062	0.9043	1.2215	1.0443
DLinear	MAE	0.3614	1.4269	0.4262	0.3165	0.7277	0.6016	0.9305	0.6844
	MSE	0.3022	3.5741	0.5786	0.1641	0.7980	0.5992	1.1781	1.0278
	RMSE	0.5493	1.8910	0.7617	0.4054	0.8943	0.7748	1.0852	0.9088
iTransformer	MAE	0.3821	1.5398	0.4419	0.3136	0.7239	0.6275	0.9420	0.7101
	MSE	0.3244	4.2576	0.6249	0.1657	0.8524	0.6970	1.3635	1.1836
	RMSE	0.5705	2.0634	0.7909	0.4067	0.9231	0.8358	1.1689	0.9656
Koopa	MAE	0.3233	1.6498	0.4161	**0.3057**	0.7226	0.6189	0.8582	0.6992
	MSE	0.2680	5.2355	0.5285	**0.1640**	0.8509	0.6462	1.1296	1.2604
	RMSE	0.5188	2.2884	0.7271	**0.4054**	0.9227	0.7852	1.0635	0.9587
NSTransformer	MAE	0.3261	1.7127	0.4569	0.3626	0.7698	0.6878	0.9752	0.7559
	MSE	0.2762	5.7245	0.6027	0.2198	0.9331	0.7514	1.3522	1.4086
	RMSE	0.5256	2.3902	0.7763	0.4672	0.9695	0.8677	1.1627	1.0227
PatchTST	MAE	0.3228	1.8450	0.4029	0.3371	0.7743	0.6456	0.9150	0.7490
	MSE	0.2614	5.7711	0.4693	0.1985	0.9516	0.7223	1.2445	1.3741
	RMSE	0.5118	2.4012	0.6890	0.4457	0.9750	0.8492	1.1135	0.9979
TimesNet	MAE	0.3181	1.5708	0.4283	0.3263	0.7159	0.6367	0.9269	0.7033
	MSE	0.2633	4.4105	0.5914	0.1861	0.8446	0.7087	1.3148	1.1885
	RMSE	0.5132	2.1030	0.7699	0.4321	0.9187	0.8462	1.1496	0.9618
EVENTTSF	MAE	**0.0812**	**0.3544**	**0.2314**	0.3561	**0.3500**	**0.3178**	**0.5869**	**0.3254**
	MSE	**0.0235**	**0.3668**	**0.2243**	0.3936	**0.3512**	**0.3202**	**1.1526**	**0.4046**
	RMSE	**0.1535**	**0.6056**	**0.4736**	0.6273	**0.5926**	**0.5659**	**1.0736**	**0.5846**

Table 2: Deterministic forecasting results on seven event-aware datasets. We report MAE↓, MSE↓, and RMSE↓ (lower is better) for each dataset, and the last column summarizes the average performance across datasets. Best results are highlighted in **bold**.

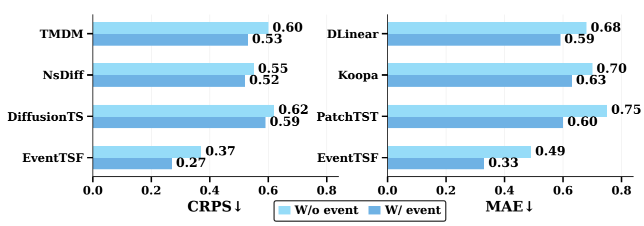

Figure 3: Baselines with Event Insert. (**Left**) Probabilistic forecasting(CRPS↓); (**Right**)Deterministic forecasting (MAE↓)

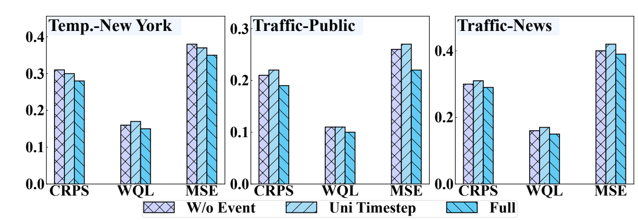

Figure 4: Ablation. Evaluating textual events and event-aware diffusion timestep against the full model.

into the input embedding space and add them to the input tokens as a residual conditioning signal.

As reported in Fig. 3, incorporating events consistently improves the performance of both probabilistic and deterministic baselines. Even with event access, all baselines remain substantially inferior to EVENTTSF. For probabilistic forecasting, incorporating event information yields a CRPS reduction of 0.07 for TMDM, corresponding to an 11.7% relative improvement. In comparison, EVENTTSF achieves a larger CRPS reduction of 0.10, translating to a 27.1% improvement. For deterministic forecasting, under the MAE metric, EVENTTSF with textual event conditioning consistently outperforms the other event-augmented baselines, delivering an average gain of 0.28 corresponding to an 45.9% relative improvement. Overall, even with the textual event access, none of the event-enhanced baselines achieves performance comparable to EVENTTSF. This indicates that injecting event embeddings simply is insufficient to exploiting event information and achieve event-aware modeling.

4.4 Ablation Study (RQ3)

Figure 4 reports ablation results across three datasets under three experiment configurations and evaluated with three metrics. Removing textual events causes performance degradation: WQL increases 10.71% on the Temperature–New York

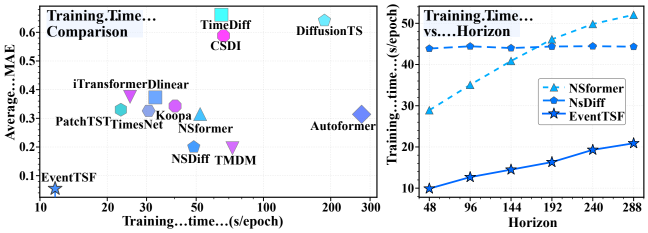

Figure 5: Training Efficiency. **(Left)** Training time vs. MAE. **(Right)** Training time scalability with horizon length.

events dataset. Moreover, replacing classical uniform diffusion timestep sampling with event-aware diffusion timestep yields consistent gains, reducing CRPS by 9.21%, WQL by 12.50%, and MSE by 10.42%. Additional detailed experimental on the effectiveness of event-aware diffusion timestep sampling are provided in Figure 2. Overall, the ablation results consistently confirm the necessity of each component, demonstrating that the complete architecture is essential for event-aware non-stationary time series forecasting.

4.5 Efficiency Analysis (RQ4)

Figure 5 analyzes training efficiency across forecasting models and scalability with horizon lengths. EVENTTSF achieves the shortest training time and lowest forecasting error across all datasets. It shows $1.13\times$ faster training and $4.85\times$ lower forecasting error versus the second-best baseline on the Synthetic dataset. NSTransformer shows linear training time growth with horizon length, while NsDiff maintains consistently high training time across all horizons. In contrast, EVENTTSF achieves both lower training time and sublinear growth with increasing horizons. Because it generates forecasts segment-wise rather than point-wise. Results confirm method's superior efficiency in event-aware non-stationary forecasting, benefiting long-horizon scenarios.

4.6 Event-Aware Forecasting Visualization (RQ5)

Figure 6 illustrates event-aware forecasting visualization on the Synthetic and Traffic-Public datasets. Textual event incorporation significantly improves accuracy. On the Synthetic dataset, the event-aware model accurately captures abrupt transitions while the event-unaware model produces smoothed responses. On the Traffic-Public dataset, event information improves the alignment with ground truth under noisy conditions, although improvements are more pronounced on the synthetic data. Results demonstrate that event information provides essential semantic cues for precise forecasting across both synthetic dataset and real-world dataset.

5 Related Work

5.1 Non-stationary Time Series Forecasting

Non-stationarity challenges time series forecasting; deep learning tackles it through normalization, model design, and learning theory. **(I) Normalization plugins** are designed from time domains [Kim *et al.*, 2021; Fan *et al.*, 2023] and frequency domains [Ye *et al.*, 2024; Liu *et al.*, 2025b]. **(II) Models** include tailored architectures [Liu *et al.*, 2022b; Liu *et al.*, 2024b], such as frequency and statistical decompositions [Yi *et al.*, 2023; Luo *et al.*, 2025; Ge *et al.*, 2024], and dynamic system modeling [Liu *et al.*, 2023b; Wang *et al.*, 2022]. **(III) Learning-theoretic approaches** handle distribution shifts via domain generation [Liu *et al.*, 2024a], adaptation [Kim *et al.*, 2025; Li *et al.*, 2022], and online learning [Wen *et al.*, 2023; Zhan *et al.*, 2025]. However, incorporating external modality for non-stationary time series forecasting remains largely unexplored.

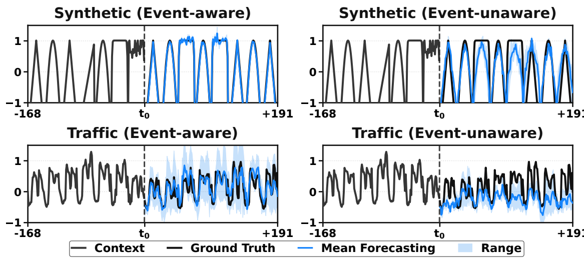

Figure 6: Visualization of event-aware forecasting on the Synthetic and Traffic–Public datasets. External events improve accuracy.

5.2 Multimodal for Time series Forecasting

Incorporating textual information into time series forecasting has gained attention across finance, traffic, and power systems [Xu *et al.*, 2021; Liang *et al.*, 2024a; Lian *et al.*, 2026; Bai *et al.*, 2024], showing promise in early LLM-based forecasting [Xue and Salim, 2023; Jin *et al.*, 2024]. Subsequently,verious fusion methods have emerged exploring model architecture [Xu *et al.*, 2024], learning strategies [Liu *et al.*, 2025a], and benchmarks [Williams *et al.*, 2024]. Recently, retrieval-augmented [Jiang *et al.*, 2025; Zhang *et al.*, 2025a], reasoning [Guan *et al.*, 2026], and agent-based systems [Lee *et al.*, 2025; Wang *et al.*, 2024] leverage LLMs' contextual analysis capabilities. However, these approaches rely heavily on LLMs for modality fusion, potentially limiting their interactive generative expressiveness. Autoregressive diffusion architectures offer probabilistic forecasting capabilities while supporting textual conditioning. Combining autoregressive mechanisms with diffusion modeling remains unexplored for event-aware time series forecasting.

6 Conclusion

This work pioneers the integration of external multimodal knowledge to tackle non-stationarity in time series forecasting. EVENTTSF captures fine-grained time series and event interactions via an autoregressive diffusion architecture with event-aware timestep. This novel architecture addresses the gap in synchronizing and modeling discrete event and continuous time series. Event-aware timestep reallocate timesteps conditioned on the event so that the denoiser sees a more balanced difficulty distribution across diffusion stages. Success in modeling cross-modal non-stationary dynamics opens up a direction for non-stationary time series analysis.

References

[Bai *et al.*, 2024] Yun Bai, Simon Camal, and Andrea Michiorri. News and load: A quantitative exploration of natural language processing applications for forecasting day-ahead electricity system demand. *IEEE Transactions on Power Systems*, 39:6222–6234, 2024.

[Box and Jenkins, 1976] George EP Box and Gwilym M Jenkins. Time series analysis. forecasting and control. *Holden-Day Series in Time Series Analysis*, 1976.

[Chen *et al.*, 2024] Boyuan Chen, Diego Martí Monsó, Yilun Du, Max Simchowitz, Russ Tedrake, and Vincent Sitzmann. Diffusion forcing: Next-token prediction meets full-sequence diffusion. *Advances in Neural Information Processing Systems*, 37:24081–24125, 2024.

[Choi *et al.*, 2022] Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In *CVPR*, pages 11472–11481, 2022.

[Fan *et al.*, 2023] Wei Fan, Pengyang Wang, Dongkun Wang, Dongjie Wang, Yuanchun Zhou, and Yanjie Fu. Dish-ts: a general paradigm for alleviating distribution shift in time series forecasting. In *AAAI*, volume 37, pages 7522–7529, 2023.

[Ge *et al.*, 2024] Yunfeng Ge, Yingxin Zhang, Keyi Shi, and Hongyan Li. A moment cross predictor for non-stationary mobile traffic forecasting. In *ICCC*, pages 2059–2064. IEEE, 2024.

[Ge *et al.*, 2025] Yunfeng Ge, Jiawei Li, Yiji Zhao, Haomin Wen, Zhao Li, Meikang Qiu, Hongyan Li, Ming Jin, and Shirui Pan. T2s: high-resolution time series generation with text-to-series diffusion models. In *IJCAI*, pages 5208–5216, 2025.

[Guan *et al.*, 2026] Tong Guan, Zijie Meng, Dianqi Li, Shiyu Wang, Chao-Han Huck Yang, Qingsong Wen, Zuozhu Liu, Sabato Marco Siniscalchi, Ming Jin, and Shirui Pan. Timeomni-1: Incentivizing complex reasoning with time series in large language models. In *ICLR*, 2026.

[Jiang *et al.*, 2025] Yushan Jiang, Wenchao Yu, Geon Lee, Dongjin Song, Kijung Shin, Wei Cheng, Yanchi Liu, and Haifeng Chen. Explainable multi-modal time series prediction with llm-in-the-loop. *arXiv preprint arXiv:2503.01013*, 2025.

[Jin *et al.*, 2024] Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang, Bin Yang, Jindong Wang, Shirui Pan, and Qingsong Wen. Position: What can large language models tell us about time series analysis. In *ICML*. MLResearchPress, 2024.

[Kim *et al.*, 2021] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *ICLR*, 2021.

[Kim *et al.*, 2025] HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *AAAI*, volume 39, pages 17868–17876, 2025.

[Lee *et al.*, 2025] Geon Lee, Wenchao Yu, Kijung Shin, Wei Cheng, and Haifeng Chen. Timecap: Learning to contextualize, augment, and predict time series events with large language model agents. In *AAAI*, volume 39, pages 18082–18090, 2025.

[Li *et al.*, 2022] Wendi Li, Xiao Yang, Weiqing Liu, Yingce Xia, and Jiang Bian. Ddg-da: Data distribution generation for predictable concept drift adaptation. In *AAAI*, volume 36, pages 4092–4100, 2022.

[Li *et al.*, 2024] Yuxin Li, Wenchao Chen, Xinyue Hu, Bo Chen, Mingyuan Zhou, et al. Transformer-modulated diffusion models for probabilistic multivariate time series forecasting. In *ICLR*, 2024.

[Lian *et al.*, 2026] Xinglin Lian, Yu Zheng, Yan Liu, Fan Zhou, Chunlei Peng, and Xinbo Gao. Contextual masking distillation for network traffic anomaly detection. *IEEE Transactions on Information Forensics and Security*, 2026.

[Liang *et al.*, 2024a] Yuebing Liang, Yichao Liu, Xiaohan Wang, and Zhan Zhao. Exploring large language models for human mobility prediction under public events. *Computers, Environment and Urban Systems*, 112:102153, 2024.

[Liang *et al.*, 2024b] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *KDD*, pages 6555–6565, 2024.

[Lipman *et al.*, 2022] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[Liu *et al.*, 2022a] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

[Liu *et al.*, 2022b] 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.

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

[Liu *et al.*, 2023b] Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with koopman predictors. *Advances in neural information processing systems*, 36:12271–12290, 2023.

[Liu *et al.*, 2024a] Haoxin Liu, Harshavardhan Kamarthi, Lingkai Kong, Zhiyuan Zhao, Chao Zhang, and B Aditya Prakash. Time-series forecasting for out-of-distribution generalization using invariant learning. *arXiv preprint arXiv:2406.09130*, 2024.

[Liu *et al.*, 2024b] Peiyuan Liu, Beiliang Wu, Yifan Hu, Naiqi Li, Tao Dai, Jigang Bao, and Shu-tao Xia. Timebridge: Non-stationarity matters for long-term time series forecasting. *arXiv preprint arXiv:2410.04442*, 2024.

[Liu *et al.*, 2025a] Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In *AAAI*, volume 39, pages 18915–18923, 2025.

[Liu *et al.*, 2025b] Qinglong Liu, Cong Xu, Wenhao Jiang, Kaixuan Wang, Lin Ma, and Haifeng Li. Timestacker: A novel framework with multilevel observation for capturing nonstationary patterns in time series forecasting. In *ICML*, 2025.

[Luo *et al.*, 2025] Yuxiao Luo, Songming Zhang, Ziyu Lyu, and Yuhan Hu. Tfdnet: Time–frequency enhanced decomposed network for long-term time series forecasting. *Pattern Recognition*, 162:111412, 2025.

[Narasimhan *et al.*, 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*, 2024.

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

[Qiu *et al.*, 2026] Xiangfei Qiu, Yuhan Zhu, Zhengyu Li, Xingjian Wu, Bin Yang, and Jilin Hu. Dag: A dual correlation network for time series forecasting with exogenous variables. In *ICML*, 2026.

[Shen and Kwok, 2023] Lifeng Shen and James Kwok. Non-autoregressive conditional diffusion models for time series prediction. In *ICML*, pages 31016–31029. PMLR, 2023.

[Tashiro *et al.*, 2021] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in neural information processing systems*, 34:24804–24816, 2021.

[Wang *et al.*, 2022] Rui Wang, Yihe Dong, Sercan Ö Arik, and Rose Yu. Koopman neural forecaster for time series with temporal distribution shifts. *arXiv preprint arXiv:2210.03675*, 2022.

[Wang *et al.*, 2024] Xinlei Wang, Maike Feng, Jing Qiu, Jinjin Gu, and Junhua Zhao. From news to forecast: Integrating event analysis in llm-based time series forecasting with reflection. *Advances in Neural Information Processing Systems*, 37:58118–58153, 2024.

[Wen *et al.*, 2023] Qingsong Wen, Weiqi Chen, Liang Sun, Zhang Zhang, Liang Wang, Rong Jin, Tieniu Tan, et al. Onenet: Enhancing time series forecasting models under concept drift by online ensembling. *Advances in Neural Information Processing Systems*, 36:69949–69980, 2023.

[Williams *et al.*, 2024] Andrew Robert Williams, Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Jithendaraa Subramanian, Roland Riachi, James Requeima, Alexandre Lacoste, Irina Rish, Nicolas Chapados, et al. Context is key: A benchmark for forecasting with essential textual information. *arXiv preprint arXiv:2410.18959*, 2024.

[Wu *et al.*, 2023] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *ICLR*, 2023.

[Xu *et al.*, 2021] Wentao Xu, Weiqing Liu, Chang Xu, Jiang Bian, Jian Yin, and Tie-Yan Liu. Rest: Relational event-driven stock trend forecasting. In *WWW*, pages 1–10, 2021.

[Xu *et al.*, 2024] Zhijian Xu, Yuxuan Bian, Jianyuan Zhong, Xiangyu Wen, and Qiang Xu. Beyond trend and periodicity: Guiding time series forecasting with textual cues. *arXiv e-prints*, pages arXiv–2405, 2024.

[Xue and Salim, 2023] Hao Xue and Flora D Salim. Promptcast: A new prompt-based learning paradigm for time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36:6851–6864, 2023.

[Ye *et al.*, 2024] Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. *Advances in Neural Information Processing Systems*, 37:31350–31379, 2024.

[Ye *et al.*, 2025] Weiwei Ye, Zhuopeng Xu, and Ning Gui. Non-stationary diffusion for probabilistic time series forecasting. *arXiv preprint arXiv:2505.04278*, 2025.

[Yi *et al.*, 2023] Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.

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

[Zhan *et al.*, 2025] Tianxiang Zhan, Ming Jin, Yuanpeng He, Yuxuan Liang, Yong Deng, and Shirui Pan. Continuous evolution pool: Taming recurring concept drift in online time series forecasting. *arXiv preprint arXiv:2506.14790*, 2025.

[Zhang *et al.*, 2025a] Huanyu Zhang, Chang Xu, Yi-Fan Zhang, Zhang Zhang, Liang Wang, and Jiang Bian. Timeraf: Retrieval-augmented foundation model for zero-shot time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Zhang *et al.*, 2025b] Xiyuan Zhang, Boran Han, Haoyang Fang, Abdul Fatir Ansari, Shuai Zhang, Danielle C Maddix, Cuixiong Hu, Andrew Gordon Wilson, Michael W Mahoney, Hao Wang, et al. Does multimodality lead to better time series forecasting? *arXiv preprint arXiv:2506.21611*, 2025.