MERATSF: EVENT-AWARE RESIDUAL MODELING FOR SPARSE EXOGENOUS TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incorporating exogenous variables into time series forecasting remains challenging, as real-world exogenous information is often sparse and temporally heterogeneous in its effects. Unlike conventional covariates, sparse exogenous events occur infrequently and abruptly, inducing transient disturbances, horizon-dependent responses, and long-term distribution shifts. These effects can be viewed as scale-dependent deviations from the underlying normal temporal dynamics, which are difficult to capture through direct correlation-based modeling. To address this challenge, we propose **MeraTSF**, a multi-scale residual refinement framework that models exogenous effects as structured residual deviations from normal temporal patterns. MeraTSF decomposes temporal dynamics across multiple scales and uses exogenous information to refine the corresponding residual representations adaptively, enabling the model to capture heterogeneous temporal effects induced by sparse exogenous events. Extensive experiments on multiple real-world datasets show that MeraTSF consistently outperforms state-of-the-art methods across diverse datasets and forecasting settings. Moreover, MeraTSF integrates seamlessly into existing forecasting architectures, highlighting its strong plug-and-play capability and broad applicability.

1 INTRODUCTION

Time series forecasting (Zhang et al., 2026; Cao et al., 2026; Wu et al., 2026a) is a fundamental task in numerous real-world applications, including economic analysis (Newbold, 2014), energy management (Fiot & Dinuzzo, 2016), and traffic flow prediction (Yu et al., 2018; Zheng et al., 2020). It aims to infer future temporal dynamics from historical observations and plays a crucial role in data-driven decision-making. However, real-world time series are often affected by various exogenous factors, introducing additional complexity into forecasting. For example, ride-hailing demand (Geng et al., 2019) is not only determined by historical mobility patterns but also influenced by external factors such as weather (e.g., rainy and snowy), holidays, and large-scale events (e.g., major sporting events and concerts). Therefore, effectively incorporating exogenous variables is essential for capturing contextual information and improving forecasting accuracy.

Exogenous-aware forecasting methods (Zhou et al., 2025; Chen et al., 2026; Lim et al., 2021) have recently achieved significant progress by developing advanced architectures (Vaswani et al., 2017; He et al., 2016) to model interactions between exogenous and endogenous variables. For instance, TimeXer (Wang et al., 2024c) represents exogenous variables as variable tokens and utilizes cross-attention to capture interactions with endogenous variables. GCGNet (Li et al., 2026) constructs relational graphs to model temporal and channel-wise dependencies between predictions and ground truth, while DAG (Qiu et al., 2026) captures cross-time and cross-channel dependencies between exogenous and endogenous variables.

However, exogenous effects in real-world scenarios are far more complex and heterogeneous than simple correlations. In particular, the impacts of exogenous variables cannot always be captured through direct associations with endogenous dynamics, especially for sparse and discrete factors such as holidays and large-scale events. Such scenarios motivate the problem of sparse exogenous time series forecasting, where infrequent external events may induce complex and heterogeneous effects on the target series.

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.

Moreover, different exogenous variables influence temporal dynamics at distinct temporal scales, leading to substantially different response patterns. As illustrated in Figure 1, weather may introduce localized and transient disturbances, large-scale events may trigger irregular responses across multiple forecasting horizons, while holidays may cause persistent distribution shifts lasting for several days. These exogenous effects can be viewed as scale-dependent deviations from normal temporal patterns, which are difficult to capture using strategies of modeling correlation. Motivated by this insight, we first decouple normal temporal dynamics from exogenous event effects: a base forecasting component learns the underlying normal temporal patterns, while exogenous variables serve as residual refinement signals that capture event-induced deviations.

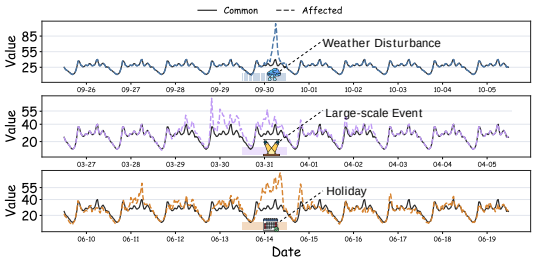


Figure 1: The multi-scale exogenous effects.

Based on this formulation, we introduce **MeraTSF**, a multi-scale residual refinement framework that models exogenous effects from a residual-centric perspective. MeraTSF first decomposes the temporal dynamics into trends and residuals at multiple temporal scales. The trend branches capture normal temporal patterns at different scales, while the residual branches characterize exogenous influences through the evolution of forecasting residuals. By progressively refining residual signals across multiple temporal scales, MeraTSF captures diverse exogenous-induced effects, including transient perturbations, irregular responses, and persistent distribution shifts. This disentangled formulation offers a new perspective on sparse exogenous forecasting, where exogenous effects are modeled as residual deviations from normal temporal dynamics rather than direct predictors of future values. Extensive experiments on 13 real-world datasets demonstrate that MeraTSF outperforms state-of-the-art forecasting methods across diverse forecasting scenarios, while also exhibiting strong plug-and-play capability across existing forecasting architectures.

In summary, our contributions are as follows:

- We propose a new prediction paradigm that is specifically designed to effectively exploit sparse and heterogeneous exogenous information for time series forecasting.
- We introduce a simple exogenous residual refinement mechanism that leverages forecasting residuals to uncover and exploit information from sparse exogenous variables.
- Building upon this mechanism, we propose MeraTSF, a multi-scale residual refinement framework that effectively captures heterogeneous exogenous effects.
- We conduct extensive experiments on 13 real-world datasets with exogenous variables, demonstrating that MeraTSF achieves SOTA performance across various scenarios.

2 Related Work

Time Series Forecasting. Early forecasting methods primarily relied on statistical techniques, including ARIMA (Box & Pierce, 1970), exponential smoothing (Hyndman et al., 2008), Kalman filtering (Harvey & Harvey, 1990), and vector autoregression (Toda & Phillips, 1994), which characterize temporal dependencies through explicit statistical assumptions. With the rapid development of deep learning, recent forecasting methods can be broadly categorized into channel-independent and multivariate approaches (Wang et al., 2024b; Shao et al., 2024; Zhang et al., 2024). Channel-independent methods (Niu et al., 2026; Huang et al., 2025; Lin et al., 2025b; Li et al., 2023; Shao et al., 2022) model each variable separately while sharing parameters across channels, achieving competitive forecasting performance with improved efficiency. For example, DLinear (Zeng et al., 2023) decomposes time series into trend and residual components and models them separately, providing a simple yet effective forecasting paradigm. In contrast, multivariate approaches (Luo & Wang, 2024; Yu et al., 2024; Han et al., 2024; Huang et al., 2023; Hu et al., 2025) explicitly capture interactions among variables to exploit cross-variable information. For instance, Crossformer (Zhang & Yan, 2023) jointly models dependencies across temporal and variable dimensions. More recently, large-scale time series foundation models (Das et al., 2023b; Liu et al., 2025a; 2026;

Auer et al., 2026; Wu et al., 2026a) and LLM-based forecasting approaches (Jin et al., 2024; Siru et al., 2025; Liu et al., 2025b) have emerged as promising directions toward general-purpose forecasting systems. Despite these advances, existing forecasting models predominantly focus on learning endogenous temporal dynamics from historical observations, while the effective utilization of exogenous information remains largely underexplored.

Forecasting with Exogenous Variables. Exogenous variables provide valuable contextual signals for real-world time series forecasting (Hao et al., 2026; Das et al., 2023a; Lim et al., 2021). Exogenous variables provide valuable contextual information for real-world time series forecasting (Hao et al., 2026; Das et al., 2023a; Lim et al., 2021). Existing approaches can be broadly categorized into two groups. The first group incorporates exogenous information by directly concatenating or fusing covariates with the endogenous series (Das et al., 2023a; Lim et al., 2021; Ansari et al., 2025). While straightforward, such approaches offer limited interpretability and may fail to fully exploit the heterogeneous information contained in exogenous variables. The second group introduces architecture-specific interaction modules to explicitly model the interactions between endogenous and exogenous variables (Wang et al., 2024c; Li et al., 2026; Qiu et al., 2026; Chen et al., 2026). These methods generally rely on modeling correlations between the two types of variables, which provides a degree of interpretability but can be challenging in real-world settings. In particular, exogenous variables are often sparse and highly heterogeneous, and their effects on the endogenous series can vary substantially across temporal scales, making stable and meaningful correlation structures difficult to establish.

In contrast, MeraTSF does not rely on explicit correlation modeling between endogenous and exogenous variables. Instead, it provides an effective mechanism for exploiting sparse and heterogeneous exogenous information, making it better suited to the complex characteristics of real-world forecasting scenarios.

3 METHODOLOGY

Definition of forecasting task. In exogenous-aware time series forecasting, we denote the historical endogenous series by $X^{\text{endo}} \in \mathbb{R}^{D\times T}$, the historical exogenous variables by $X^{\text{exo}} \in \mathbb{R}^{K\times T}$, and the known future exogenous variables by $Y^{\text{exo}} \in \mathbb{R}^{K\times H}$. And D denotes the number of endogenous variables, K denotes the number of exogenous variables, T is the lookback length, and H is the forecasting horizon. Given X^{endo}, X^{exo}, and Y^{exo}, the objective is to predict the future endogenous sequence $Y^{\text{endo}} \in \mathbb{R}^{D\times H}$.

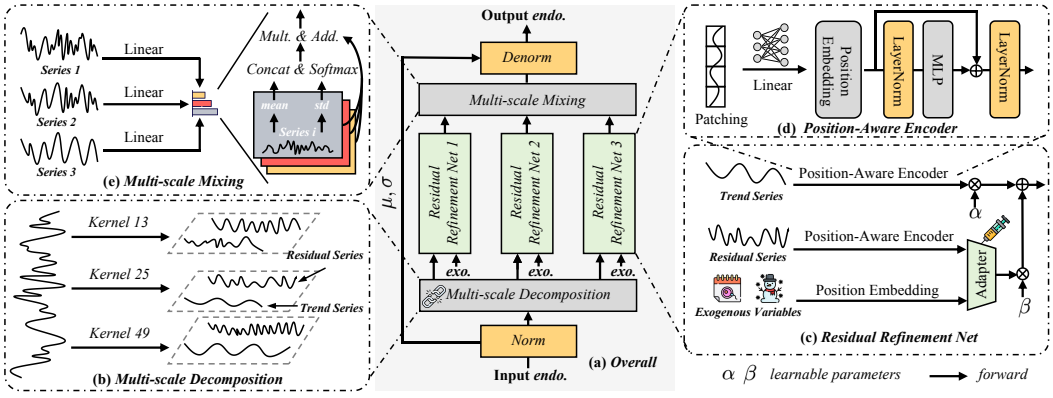


Figure 2: Model architecture of the proposed MeraTSF.

Overview of MeraTSF. Exogenous variables can affect endogenous dynamics over heterogeneous temporal scales. Some induce localized and transient perturbations, while others lead to persistent distribution shifts. MeraTSF addresses this heterogeneity by refining endogenous residual representations across multiple temporal scales. As illustrated in Figure 2, MeraTSF first normalizes the historical endogenous series and decomposes it using multiple moving-average kernels. At each scale, the input is decomposed into a trend branch and a residual branch, which are separately represented as patch tokens. The Multi-Scale Residual Refinement module then leverages historical

and known-future exogenous information as contextual cues to refine the residual representations at each scale. The refined residual representation is subsequently fused with its corresponding trend representation to produce a scale-specific temporal feature. Finally, the Multi-Scale Mixing module assigns instance-dependent weights to the scale-specific forecasts based on their feature statistics and aggregates them into the final prediction.

The key idea is to inject exogenous information primarily into the residual component, which captures deviations from the intrinsic endogenous dynamics, while preserving the trend branch to retain the dominant temporal structure. This design allows MeraTSF to capture exogenous effects across different temporal scales without disrupting the underlying endogenous dynamics.

3.1 Multi-scale Residual Refinement

Normalization. Real-world time series often exhibit distribution shifts across different forecasting windows. To mitigate this issue, we apply RevIN (Kim et al., 2021) to each endogenous variable. For channel d, the normalized input is denoted by $\bar{X}_d^{\text{endo}}$. The forecasting backbone produces prediction $\bar{Y}_d^{\text{endo}}$, which is transformed to the original scale through inverse normalization to obtain the final prediction $\hat{Y}_d^{\text{endo}}$. Formally,

$$\bar{X}_d^{\text{endo}} = \text{Norm}\left(X_d^{\text{endo}}\right), \qquad \hat{Y}_d^{\text{endo}} = \text{DeNorm}\left(\bar{Y}_d^{\text{endo}}\right). \tag{1}$$

Multi-scale decomposition. Decomposing time series into trend and residual components (Lin et al., 2024a) facilitates the modeling of complex temporal dynamics. Let $\mathcal{K} = \{k_1, \ldots, k_S\}$ denote a set of moving-average kernel sizes corresponding to S temporal scales. A smaller kernel captures fine-grained local variations by preserving high-frequency temporal patterns, whereas a larger kernel aggregates information over a broader temporal context and thus yields a smoother trend. Consequently, the residual component at a larger scale reflects deviations from longer-term temporal patterns. For each scale s, the normalized time series are decomposed into a trend component and a residual component:

$$T_s = \mathcal{M}_{k_s}\left(\bar{X}^{\text{endo}}\right), \qquad R_s = \bar{X}^{\text{endo}} - T_s, \tag{2}$$

where $\mathcal{M}_{k_s}$ denotes a centered moving average with boundary replication. Both components are partitioned into temporal patches of length p with stride q using the patching operator $\mathcal{P}$. The trend patches are first projected into the token space and augmented with learnable positional embeddings Π^{endo}. Formally,

$$\bar{T}_s = \mathcal{P}(T_s)W_\psi + \Pi^{\text{endo}} \in \mathbb{R}^{D\times N\times d}, \tag{3}$$

$$\hat{T}_s = \text{LayerNorm}\big(\bar{T}_s + \text{MLP}\big(\text{LayerNorm}(\bar{T}_s)\big)\big) \in \mathbb{R}^{D\times N\times d}. \tag{4}$$

The residual is encoded analogously, yielding the corresponding residual tokens $\hat{R}_s$. Separate encoders allow the rapidly varying residual dynamics and slowly varying trend dynamics to be represented independently, preventing their distinct temporal characteristics from being forced into a shared representation.

Exogenous patch encoding. We concatenate the historical and future-known exogenous variables along the temporal dimension, and partition them into temporal patches, where the variable and temporal dimensions are flattened and projected into the token space. To establish correspondence with the endogenous residual representation, the exogenous tokens are subsequently projected into the residual token space and aligned with the N residual patch positions. Specifically,

$$Z = \text{Flatten}[\mathcal{P}(\text{Concat}(X^{\text{exo}}, Y^{\text{exo}}))]\, W_\varphi + \Pi^{\text{exo}} \in \mathbb{R}^{N_z\times d}, \tag{5}$$

$$\bar{Z} = \Big[\text{LayerNorm}(Z)^\top W_\phi\Big]^\top \in \mathbb{R}^{N\times d}, \tag{6}$$

where Π^{exo} also denotes the learnable positional embedding. The aligned exogenous representation $\bar{Z}$ is subsequently broadcast across the D endogenous variables.

Residual adaptation. At each temporal scale, the aligned exogenous representation provides contextual information for characterizing residual dynamics. Since exogenous variables may exhibit heterogeneous predictive relevance across variables and temporal contexts, directly incorporating them into the residual representation can obscure the intrinsic structure of endogenous dynamics.

We therefore employ exogenous variables as a conditioning signal to adapt the residual representation, rather than directly replacing it.

Specifically, MeraTSF constructs an exogenous-conditioned correction by jointly modeling the residual state, the aligned exogenous context, and their multiplicative interaction. The interaction term, i.e., $\hat{R}_s \odot \bar{Z}$, captures conditional dependencies between endogenous residual states and exogenous contexts, while a gate G_s regulates the magnitude of the resulting correction. Formally, we define

$$G_s = \text{sigmoid}\Big(\text{Linear}\Big(\text{Concat}\Big(\hat{R}_s, \bar{Z}\Big)\Big)\Big) \in \mathbb{R}^{D\times N\times d}, \tag{7}$$

$$\widetilde{R}_s = \hat{R}_s + G_s \odot \text{MLP}\Big(\text{Concat}\Big(\hat{R}_s, \bar{Z}, \hat{R}_s \odot \bar{Z}\Big)\Big) \in \mathbb{R}^{D\times N\times d}. \tag{8}$$

The additive skip connection retains the residual representation and constrains the contribution of exogenous information to an adaptive correction. Consequently, the conditioning strength can vary across endogenous variables, patch positions, feature dimensions, and temporal scales, allowing the model to account for the heterogeneous relevance of exogenous variables.

For each temporal scale, the adapted residual tokens and the corresponding trend tokens are separately decoded into temporal features and then combined using learnable scalar coefficients:

$$F_s^r = \text{Dec}_s^r(\widetilde{R}_s), \qquad F_s^t = \text{Dec}_s^t(\hat{T}_s), \qquad F_s = \alpha F_s^t + \beta F_s^r \in \mathbb{R}^{D\times L}, \tag{9}$$

where L denotes the output dimension of the decoder, and $\alpha, \beta \in \mathbb{R}$ are learnable scalars.

3.2 Multi-scale Mixing

At each temporal scale, the corresponding feature representation is first projected onto the forecasting horizon via a scale-specific temporal projection, yielding $\bar{Y}_s \in \mathbb{R}^{D\times H}$. The relative predictive importance of different temporal scales may vary across instances. To capture such instance-dependent scale preferences, we construct an instance-level descriptor from the global statistics of the feature representation at each scale:

$$h = \text{Concat}(\text{Mean}(F_1), \text{Std}(F_1), \ldots, \text{Mean}(F_S), \text{Std}(F_S)) \in \mathbb{R}^{2S}. \tag{10}$$

Compared with other statistical measures, the mean and standard deviation effectively characterize the effects induced by multi-scale temporal variations.

As illustrated in Figure 1, short-term fluctuations primarily manifest as changes in amplitude and can exhibit substantially larger magnitudes, whereas long-term variations are characterized more by shifts in the underlying data distribution. The mean and standard deviation therefore provide complementary statistics for capturing these scale-dependent characteristics. Accordingly, we obtain instance-specific weights for different temporal scales, formally defined as follows:

$$\boldsymbol{\omega} = [\omega_1, \ldots, \omega_S] = \text{softmax}(\text{MLP}(h)) \in \mathbb{R}^S, \qquad \bar{Y}^{\text{endo}} = \sum_{s=1}^{S} \omega_s \bar{Y}_s \in \mathbb{R}^{D\times H}. \tag{11}$$

The final normalized forecast is obtained by adaptively mixing the scale-specific predictions. The normalized scale weights determine the relative contribution of each temporal scale, ensuring that the final prediction remains a convex combination of the scale-specific forecasts.

Loss Function. Following DAG (Qiu et al., 2026) and GCGNet (Li et al., 2026), MeraTSF adopts the mean absolute error (i.e., Eq.(12)) as the forecasting objective to ensure a fair comparison.

$$\mathcal{L}_{\text{total}} = \left\| Y^{\text{endo}} - \hat{Y}^{\text{endo}} \right\|_1 \tag{12}$$

4 Experiment

Dataset. We conduct experimental evaluations on one large-scale dataset and 12 small-scale datasets (Li et al., 2026; Qiu et al., 2026). The large-scale Ride-Hailing[1] dataset spans a continuous four-year period with a 30-minute sampling interval, resulting in 14,025,600 time-stamped

[1] https://anonymous.4open.science/r/ridebench-338A/datasets

observations across 200 areas and encompassing three representative sparse exogenous variables: *weather disturbance*, *holiday effect*, and *large-scale event impact*. This provides a comprehensive benchmark for evaluating the impact of sparse and heterogeneous exogenous factors. More details of these datasets are provided in Appendix A.1.

Baselines. We conduct a comprehensive comparison with 11 baseline models, including methods that natively support exogenous variables (e.g., DAG (Qiu et al., 2026), GCGNet (Li et al., 2026), XLinear (Chen et al., 2026), CrossLinear (Zhou et al., 2025), TimeXer (Wang et al., 2024c), CATS (Lu et al., 2024), and TiDE (Das et al., 2023a), TFT (Lim et al., 2021)) and methods without explicit exogenous variable modeling (e.g., TimeBase (Huang et al., 2025), TimeFilter (Hu et al., 2025), DUET (Qiu et al., 2025), TimeMixer (Wang et al., 2024a), and PatchTST (Nie et al., 2023)). To ensure a fair comparison, all exogenous-aware methods are evaluated using the same input, including historical endogenous variables, historical exogenous variables, and future exogenous variables. For methods that do not natively support future exogenous variables, we incorporate future exogenous information following the architectures in TFB (Qiu et al., 2024) and RideBench. More details are provided in Appendix A.1.

Setup. All experiments are implemented in Python 3.11 with PyTorch and conducted on a cluster of eight NVIDIA RTX A6000 GPUs. For the Ride-Hailing dataset, we focus on evaluating performance under different exogenous scenarios, where models use 1,344 historical steps to predict the next 336 steps. For the small-scale datasets, due to the absence of such scenario-specific evaluations, we follow the evaluation protocol of DAG and conduct both short-term and long-term forecasting experiments. Specifically, the short-term setting uses 168 historical steps for predicting the next 24 steps, while the long-term setting uses 720 historical steps for predicting the next 360 steps. To comprehensively evaluate ExoRAF's performance, the Weighted Mean Absolute Percentage Error (WMAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) metrics are employed. For additional details, please refer to the Appendix A.1.

4.1 Main Results

Table 1: Results on the real-world Ride-Hailing dataset. The best and second best results are highlighted in **bold** and underlined. The full results of other methods are provided in Table 13.

Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
PatchTST	0.1079	4.132	13.38	0.1391	5.821	20.84	0.1507	6.249	20.37	0.1130	4.983	16.56
TimeMixer	0.1113	4.261	13.47	0.1411	5.902	20.90	0.1548	6.420	20.42	0.1159	5.113	16.73
TimeBase	0.1222	4.674	13.68	0.1437	6.008	20.38	0.1553	6.437	19.67	0.1209	5.342	16.19
TimeFilter	0.1099	4.206	13.30	0.1390	5.812	20.67	0.1494	6.187	20.00	0.1137	5.012	16.42
DUET	0.1147	4.397	13.57	0.1437	6.011	20.97	0.1539	6.379	20.23	0.1193	5.267	16.86
TiDE	0.1043	3.996	12.43	0.1293	5.397	18.96	0.1282	5.305	17.29	0.1104	4.866	15.81
CrossLinear	0.1051	4.032	12.65	0.1269	5.296	18.74	0.1328	5.505	18.43	0.1100	4.854	15.87
XLinear	0.1075	4.124	12.86	0.1300	5.431	19.42	0.1326	5.493	18.52	0.1144	5.055	16.30
CATS	0.1103	4.225	12.95	0.1321	5.510	19.53	0.1379	5.703	18.76	0.1159	5.106	16.13
DAG	0.1090	4.174	13.30	0.1388	5.803	20.64	0.1476	6.115	19.91	0.1141	5.033	16.55
TimeXer	0.1093	4.207	12.93	0.1388	5.804	20.14	0.1371	5.687	18.42	0.1172	5.177	16.40
MeraTSF	**0.1027**	**3.943**	**12.01**	**0.1217**	**5.074**	**17.77**	**0.1261**	**5.216**	**16.65**	**0.1069**	**4.712**	**15.14**
Improve	-1.53%	-1.33%	-3.38%	-4.10%	-4.19%	-5.18%	-1.64%	-1.68%	-3.70%	-2.82%	-2.93%	-4.24%

Results in Sparse Exogenous Settings. Table 1 shows the results of MeraTSF on the large-scale ride-hailing dataset with sparse exogenous variables. MeraTSF achieves the best performance across all three forecasting scenarios, which differ in both the degree of sparsity and the magnitude of exogenous effects. Notably, MeraTSF exhibits a particularly pronounced advantage in scenarios involving sparse exogenous variables and irregular responses to large-scale events, highlighting its effectiveness and robustness in challenging forecasting settings. Furthermore, the forecasting visualizations (see Appendix A.7) consistently show that MeraTSF produces more accurate forecasts than existing methods.

Results in Non-Sparse Exogenous Settings. Table 2 summarizes the performance of MeraTSF on 12 small-scale datasets with non-sparse exogenous variables. MeraTSF consistently surpasses existing exogenous-aware time series forecasting methods, highlighting its robustness across non-sparse exogenous settings. This result demonstrates that MeraTSF is not restricted to sparse exogenous forecasting and can also effectively generalize to non-sparse exogenous forecasting scenarios. we also conduct an efficiency analysis of the MeraTSF, with the results reported in Appendix A.2.

Table 2: Average results on 12 real-world datasets with exogenous variables. The best results are **bold**, and the second best results are underline. Full results are available in Table 14.

Method	MeraTSF		DAG		GCGNet		TimeXer		TFT		TiDE		DUET		CrossLinear		PatchTST	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	**0.322**	**0.322**	0.362	0.344	0.370	0.348	0.418	0.371	0.379	0.375	0.443	0.400	0.411	0.408	0.371	0.387	0.390	0.396
PJM	**0.080**	**0.180**	0.093	**0.180**	0.095	0.187	0.108	0.198	0.114	0.207	0.142	0.246	0.102	0.197	0.112	0.223	0.133	0.259
BE	**0.415**	**0.277**	0.423	0.279	0.431	0.294	0.452	0.290	0.454	0.291	0.498	0.325	0.515	0.354	0.479	0.337	0.577	0.432
FR	**0.399**	**0.219**	0.414	**0.219**	0.415	0.234	0.427	0.241	0.504	0.257	0.484	0.281	0.496	0.327	0.483	0.298	0.588	0.410
DE	**0.341**	**0.363**	0.370	0.370	0.401	0.389	0.475	0.418	0.489	0.446	0.499	0.447	0.482	0.430	0.485	0.452	0.501	0.455
Energy	**0.092**	**0.232**	0.124	0.267	0.131	0.277	0.163	0.315	0.130	0.283	0.153	0.302	0.203	0.367	0.239	0.402	0.226	0.377
Sdwpfm1	**0.397**	**0.449**	0.423	0.461	0.424	0.457	0.701	0.609	0.482	0.474	0.483	0.507	0.599	0.570	0.426	0.502	0.435	0.502
Sdwpfm2	**0.441**	**0.479**	0.477	0.485	0.475	0.486	0.803	0.653	0.476	0.488	0.486	0.516	0.514	0.490	0.533	0.573	0.510	0.547
Sdwpfh1	**0.432**	**0.484**	0.448	0.486	0.450	0.500	0.746	0.643	0.479	0.491	0.453	0.508	0.539	0.516	0.557	0.593	0.468	0.527
Sdwpfh2	**0.484**	**0.516**	0.523	0.530	0.520	0.536	0.891	0.719	0.566	0.521	0.599	0.583	0.647	0.566	0.538	0.574	0.549	0.580
Colbun	0.101	0.169	**0.098**	**0.154**	0.107	0.175	0.145	0.235	0.238	0.297	0.164	0.227	0.198	0.266	0.126	0.195	0.239	0.309
Rapel	0.247	**0.302**	**0.230**	0.305	0.306	0.307	0.344	0.362	0.305	0.333	0.320	0.351	0.269	0.326	0.252	0.313	0.269	0.332
1^{st} Count	**10**	**11**	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0

4.2 Ablation Study and Analysis

Table 3: Ablation study of different kernels in MeraTSF. The best results are highlighted in **bold**.

Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
w/o *ke. 13*	0.1036 +0.87%	3.974 +0.78%	12.12 +0.91%	0.1240 +1.85%	5.169 +1.84%	18.01 +1.33%	0.1266 +0.39%	5.234 +0.34%	17.12 +2.75%	**0.1068** -0.09%	**4.706** -0.13%	**14.96** -1.20%
w/o *ke. 25*	0.1033 +0.58%	3.959 +0.40%	12.19 +1.48%	0.1234 +1.38%	5.147 +1.42%	18.17 +2.20%	0.1295 +2.63%	5.356 +2.61%	17.25 +3.48%	0.1070 +0.09%	4.717 +0.11%	14.97 -1.14%
w/o *ke. 49*	0.1036 +0.87%	3.976 +0.83%	12.30 +2.36%	0.1237 +1.62%	5.158 +1.63%	18.25 +2.63%	0.1291 +2.32%	5.342 +2.36%	17.76 +6.25%	0.1079 +0.93%	4.756 +0.93%	15.26 +0.79%
Full	**0.1027**	**3.943**	**12.01**	**0.1217**	**5.074**	**17.77**	**0.1261**	**5.216**	**16.65**	0.1069	4.712	15.14

Ablation Study of Different Kernels. As shown in Table 3, removing any moving average kernel degrades the performance, indicating that residual information at different scales are complementary. Specifically, removing the short-scale kernel, i.e., *kernel 13*, has a larger impact on weather disturbance scenarios, suggesting that weather disturbance mainly introduces short-term fluctuations, while removing the long-scale kernel, i.e., *kernel 49*, leads to the largest degradation in holiday effect scenarios, reflecting the long-lasting effects of holidays. The contribution of *kernel 13* to holiday scenarios further indicates that short-term fluctuations may still exist within long-term holiday patterns.

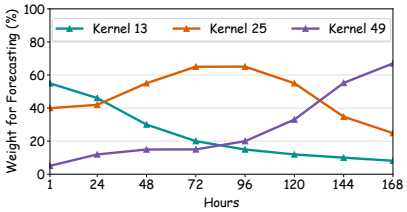


Figure 3: Weight of three kernels.

As shown in Figure 3, the forecasting kernel weights reveal clear scale-dependent temporal patterns. *kernel 13* dominates in short-term horizons, where rapid fluctuations and transient variations are more prominent, whereas *kernel 49* becomes increasingly influential as the horizon grows, capturing long-term dependencies. These results demonstrate that different exogenous scenarios indeed operate at distinct temporal scales, validating the effectiveness of MeraTSF.

Ablation Study of Different Components. As shown in Table 4, removing either the residual refinement net (i.e., w/o *RRN*) or exogenous variables (i.e., w/o *exo.*) leads to a similar degradation in performance. This suggests that the performance gains mainly come from adapting the residual component with exogenous variables, rather than from the decoupling architecture of RRN itself. Furthermore, the results of w/o *future exo.* indicate that the forecasting improvements primarily benefit from future exogenous information. To further investigate the role of residual refinement, we apply the adapter to the trend branch. However, this variant performs worse than residual refinement, suggesting that exogenous variables do not directly modify the global trend dynamics but instead introduce conditional deviations from regular temporal patterns. Together, these four ablation stud-

ies demonstrate that the effectiveness of our method stems from the residual refinement mechanism, rather than from the specific architectural designs of individual modules.

Moreover, removing the multi-scale mixing component (i.e., w/o *MM*) degrades performance, indicating that different samples may rely on different temporal dependency scales and thus require adaptive multi-scale pattern modeling. This finding is also consistent with the results in Figure 3.

Table 4: Ablation study of different components in MeraTSF.

Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
w/o *RRN*	0.1107	4.240	13.35	0.1397	5.843	20.73	0.1475	6.115	19.82	0.1150	5.075	16.58
w/o *exo.*	0.1102	4.219	13.43	0.1400	5.859	20.80	0.1525	6.324	20.30	0.1152	5.082	16.65
w/o *future exo.*	0.1122	4.305	13.16	0.1398	5.843	20.47	0.1499	6.218	19.23	0.1161	5.125	16.40
on Trend	0.1051	4.033	12.38	0.1262	5.265	18.63	0.1305	5.399	17.36	0.1097	4.840	15.50
w/o *MM*	0.1038	3.980	12.17	0.1231	5.137	18.10	0.1274	5.271	16.95	0.1085	4.788	15.19
Full	**0.1027**	**3.943**	**12.01**	**0.1217**	**5.074**	**17.77**	**0.1261**	**5.216**	**16.65**	**0.1069**	**4.712**	**15.14**

Explanation of MeraTSF. To further explain the effectiveness of residual refinement in MeraTSF, we analyze it from an information gain perspective. As shown in Figure 4(a), across all temporal scales, exogenous variables consistently provide larger predictive gains for the residual component than for the trend component. This observation suggests that, under a single-branch conditional refinement paradigm, prioritizing residual refinement more effectively recovers missing predictive information and reduces predictive uncertainty. Meanwhile, as shown in Figure 4(b), refining both the trend and residual branches simultaneously yields no additional performance improvement. Together, these findings indicate that the primary contribution of exogenous information lies in explaining residual deviations from intrinsic endogenous dynamics, while the trend branch already preserves the dominant temporal structure, making residual-only refinement a more effective design.

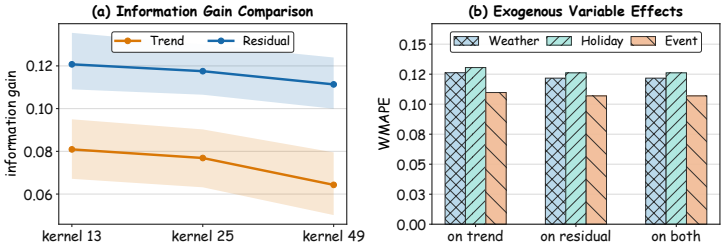


Figure 4: Understanding the effectiveness of MeraTSF from an information gain perspective. (a) Information gains from incorporating exogenous variables into the trend and residual branches across different temporal scales. (b) Performance comparison of trend-only, residual-only, and joint refinement of the trend and residual branches.

Table 5: Ablation study of MeraTSF with different backbones on the Ride-Hailing dataset.

Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
DLinear	0.1171	4.483	13.48	0.1441	6.029	20.64	0.1545	6.403	19.78	0.1186	5.234	16.29
+*MeraTSF*	0.1030 **-12.04%**	3.946 **-11.97%**	12.36 **-8.37%**	0.1240 **-13.95%**	5.172 **-14.22%**	18.20 **-11.84%**	0.1288 **-16.63%**	5.330 **-16.75%**	18.13 **-8.35%**	0.1083 **-8.68%**	4.773 **-8.82%**	15.52 **-4.70%**
PatchTST	0.1079	4.132	13.38	0.1391	5.821	20.84	0.1507	6.249	20.37	0.1130	4.983	16.56
+*MeraTSF*	0.1015 **-5.93%**	3.889 **-5.89%**	12.42 **-7.22%**	0.1236 **-11.14%**	5.155 **-11.44%**	18.46 **-11.39%**	0.1288 **-14.53%**	5.330 **-14.70%**	18.44 **-9.48%**	0.1085 **-3.98%**	4.785 **-3.97%**	15.46 **-6.64%**
iTransformer	0.1117	4.276	13.47	0.1410	5.899	20.88	0.1518	6.290	20.16	0.1169	5.159	16.79
+*MeraTSF*	0.1033 **-7.52%**	3.959 **-7.40%**	12.46 **-7.51%**	0.1245 **-11.70%**	5.195 **-11.93%**	18.43 **-11.71%**	0.1325 **-12.71%**	5.487 **-12.77%**	18.52 **-8.17%**	0.1091 **-6.67%**	4.811 **-6.75%**	15.63 **-6.95%**
SparseTSF	0.1134	4.342	13.60	0.1420	5.941	20.99	0.1553	6.442	20.53	0.1176	5.193	16.75
+*MeraTSF*	0.1024 **-9.70%**	3.926 **-9.60%**	12.36 **-9.15%**	0.1240 **-12.68%**	5.176 **-12.88%**	18.36 **-12.53%**	0.1284 **-17.32%**	5.316 **-17.49%**	18.10 **-11.84%**	0.1088 **-7.48%**	4.797 **-7.63%**	15.48 **-7.56%**
Avg. Improve	-8.80%	-8.71%	-8.06%	-12.37%	-12.62%	-11.87%	-15.30%	-15.43%	-9.46%	-6.70%	-6.79%	-6.46%

Plug-and-Play Integration of MeraTSF. To further validate the effectiveness of MeraTSF, we integrate it with different backbone models and evaluate their performance on the Ride-Hailing dataset. As shown in Table 5, MeraTSF improves all backbones, demonstrating its general applicability across different backbones. In particular, integrating MeraTSF with iTransformer (Liu et al., 2024)

achieves the most significant improvement. This is because iTransformer primarily focuses on modeling dependencies among endogenous variables, while MeraTSF does not explicitly exploit such inter-variable correlations. Therefore, integrating MeraTSF with iTransformer further improves the overall performance, which highlights the limitation (see Appendix A.4) discussed in this paper.

Moreover, combining MeraTSF with SparseTSF (Lin et al., 2024b) brings notable improvement, especially in large-scale event scenarios. SparseTSF learns periodic dependencies through cross-period sparse sampling; however, large-scale events are often irregular and do not align with fixed periodic patterns. MeraTSF complements this limitation by capturing low-frequency, exogenous-driven deviations through residual refinement, leading to more robust predictions under complex scenarios. To further evaluate the generalizability of MeraTSF in real-world scenarios, we conduct extensive experiments across multiple practical settings, with results provided in Appendix A.3.

Efficiency Analysis. Forecasting in ride-hailing platform typically involves large-scale areas, multiple prediction targets, and repeated rolling-window inference. Therefore, even models achieving competitive forecasting accuracy may face practical deployment challenges if they incur excessive inference overhead. To investigate the trade-off between forecasting performance and computational cost, Figure 5 compares representative models in terms of prediction accuracy (i.e., the average WMAPE values for the overall and the three scenarios) and inference time on the Ride-Hailing dataset. MeraTSF achieves superior forecasting performance while maintaining comparable inference costs to existing methods, demonstrating its effectiveness and deployment potential in real-world scenarios. As shown in Figure 6, MeraTSF demonstrates superior forecasting performance over existing exogenous-aware methods when exogenous variables are sparse. It also maintains strong performance in the presence of multiple heterogeneous and sparse exogenous variables. Corresponding visualizations are provided in Appendix A.7.

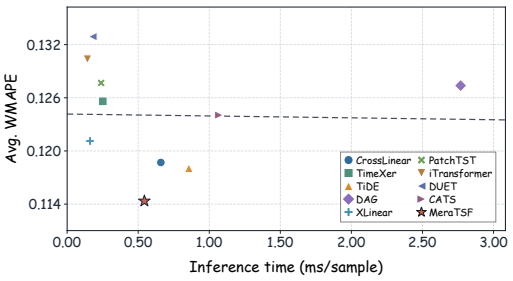


Figure 5: Efficiency of MeraTSF.

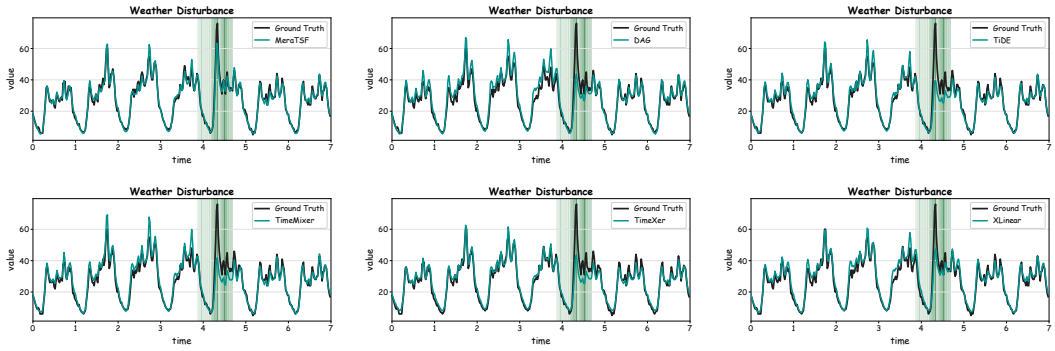


Figure 6: Forecasting visualization on the Ride-Hailing dataset under weather disturbance.

5 Conclusion

In this paper, we investigate the heterogeneous temporal effects of exogenous variables in time series forecasting and propose MeraTSF, a multi-scale residual refinement framework. MeraTSF decomposes endogenous dynamics across multiple temporal scales, performs exogenous-conditioned residual refinement, and adaptively aggregates scale-specific forecasts to model both transient perturbations and persistent distributional shifts. Extensive experiments on 13 public datasets demonstrate the effectiveness of MeraTSF across diverse domains and exogenous-aware settings. As future work, we will investigate spatial heterogeneity in ride-hailing forecasting, focusing on latent endogenous factors associated with different spatial contexts. This direction may extend MeraTSF toward jointly modeling heterogeneous temporal and spatial effects.

AI USE STATEMENT

In this work, we used a generative AI tool (i.e., ChatGPT) solely for assisting with manuscript writing refinement, including checking spelling errors, correcting grammar, and improving language clarity. We did not use generative AI tools for research ideation, methodology development, data analysis, experimental design, implementation, or generation of scientific results. All AI-assisted revisions were carefully reviewed and verified by the authors. We take responsibility for the final content of this work, including all text, claims, and artifacts.

REPRODUCIBILITY STATEMENT

We promise that all experimental results can be reproduced. We have released our model code in an anonymous repository: https://anonymous.4open.science/r/MeraTSF-main

REFERENCES

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, et al. Chronos-2: From univariate to universal forecasting. *arXiv preprint arXiv:2510.15821*, 2025.

Andreas Auer, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. Tirex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. *Advances in Neural Information Processing Systems*, 38:57529–57580, 2026.

George EP Box and David A Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332):1509–1526, 1970.

Fanpu Cao, Lu Dai, Jindong Han, and Hui Xiong. Enhancing multivariate time series forecasting with global temporal retrieval. *arXiv preprint arXiv:2602.10847*, 2026.

Xinyang Chen, Huidong Jin, Yu Huang, and Zaiwen Feng. Xlinear: A lightweight and accurate mlp-based model for long-term time series forecasting with exogenous inputs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026.

Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tide: Time-series dense encoder. *arXiv preprint arXiv:2304.08424*, 2023a.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688*, 2023b.

Jean-Baptiste Fiot and Francesco Dinuzzo. Electricity demand forecasting by multi-task learning. *IEEE Transactions on Smart Grid*, 9(2):544–551, 2016.

Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3656–3663, 2019.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. Softs: Efficient multivariate time series forecasting with series-core fusion. *Advances in Neural Information Processing Systems*, 37: 64145–64175, 2024.

Xixuan Hao, Guicheng Li, Daiqiang Wu, Xusen Guo, Yumeng Zhu, Zhichao Zou, Peng Zhen, Yao Yao, and Yuxuan Liang. Enhancing ride-hailing forecasting at didi with multi-view geospatial representation learning from the web. In *Proceedings of the ACM Web Conference 2026*, pp. 8200–8211, 2026.

Andrew C Harvey and Andrew Harvey. Forecasting, structural time series models and the kalman filter. 1990.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Ao Hu, Liangjian Wen, Jiang Duan, Yong Dai, YAN HE, Dongkai Wang, Jun Wang, Yukun Zhang, Ruoxi Jiang, and Zenglin Xu. Pmdformer: Patch-mean decoupling information transformer for long-term forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. Timefilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *International Conference on Machine Learning*, pp. 24893–24911. PMLR, 2025.

Qihe Huang, Lei Shen, Ruixin Zhang, Shouhong Ding, Binwu Wang, Zhengyang Zhou, and Yang Wang. Crossgnn: Confronting noisy multivariate time series via cross interaction refinement. *Advances in Neural Information Processing Systems*, 36:46885–46902, 2023.

Qihe Huang, Zhengyang Zhou, Kuo Yang, Zhongchao Yi, Xu Wang, and Yang Wang. TimeBase: The power of minimalism in efficient long-term time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 26227–26246, 2025.

Rob Hyndman, Anne Koehler, Keith Ord, and Ralph Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer, 2008.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. In *International conference on learning representations*, volume 2024, pp. 23857–23880, 2024.

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 *International conference on learning representations*, 2021.

Zhe Li, Shiyi Qi, Yiduo Li, and Zenglin Xu. Revisiting long-term time series forecasting: An investigation on linear mapping. *arXiv preprint arXiv:2305.10721*, 2023.

Zhengyu Li, Xiangfei Qiu, Yuhan Zhu, Xingjian Wu, Jilin Hu, Chenjuan Guo, and Bin Yang. Gcgnet: Graph-consistent generative network for time series forecasting with exogenous variables. *arXiv preprint arXiv:2603.08032*, 2026.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4): 1748–1764, 2021.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. *Advances in Neural Information Processing Systems*, 37:106315–106345, 2024a.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: modeling long-term time series forecasting with 1k parameters. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 30211–30226, 2024b.

Shengsheng Lin, Haojun Chen, Haijie Wu, Chunyun Qiu, and Weiwei Lin. Temporal query network for efficient multivariate time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025a.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Feiyu Zhao, Ruichao Mo, and Haotong Zhang. Segrnn: Segment recurrent neural network for long-term time series forecasting. *IEEE Internet of Things Journal*, 2025b.

Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. Moirai 2.0: When less is more for time series forecasting. *arXiv preprint arXiv:2511.11698*, 2025a.

Chenxi Liu, Qianxiong Xu, Hao Miao, Sun Yang, Lingzheng Zhang, Cheng Long, Ziyue Li, and Rui Zhao. Timecma: Towards llm-empowered multivariate time series forecasting via cross-modality alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18780–18788, 2025b.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 11116–11140, 2024.

Yong Liu, Xingjian Su, Shiyu Wang, Haoran Zhang, Haixuan Liu, Yuxuan Wang, Zhou Ye, Yang Xiang, Jianmin Wang, and Mingsheng Long. Timer-s1: A billion-scale time series foundation model with serial scaling. *arXiv preprint arXiv:2603.04791*, 2026.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Jiecheng Lu, Xu Han, Yan Sun, and Shihao Yang. Cats: Enhancing multivariate time series forecasting by constructing auxiliary time series as exogenous variables. In *Forty-first International Conference on Machine Learning*, 2024.

Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *International Conference on Machine Learning*, 2024.

Paul Newbold. *Forecasting economic time series*. Academic press, 2014.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

Yiming Niu, Jinliang Deng, and Yongxin Tong. Phaseformer: From patches to phases for efficient and effective time series forecasting. In *International Conference on Learning Representations*, 2026.

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S Jensen, Zhenli Sheng, et al. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *Proceedings of the VLDB Endowment*, 17(9): 2363–2377, 2024.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1185–1196, 2025.

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 *International Conference on Machine Learning*, 2026.

Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pp. 4454–4458, 2022.

Zezhi Shao, Fei Wang, Yongjun Xu, Wei Wei, Chengqing Yu, Zhao Zhang, Di Yao, Tao Sun, Guangyin Jin, Xin Cao, et al. Exploring progress in multivariate time series forecasting: Comprehensive benchmarking and heterogeneity analysis. *IEEE Transactions on Knowledge and Data Engineering*, 37(1):291–305, 2024.

Zhong Siru, Ruan Weilin, Ming Jin, Li Huan, Wen Qingsong, and Liang Yuxuan. Time-vlm: Exploring multimodal vision-language models for augmented time series forecasting. In *Forty-Second International Conference on Machine Learning (ICML 2025)*. Proceedings of Machine Learning Research, 2025.

Hiro Y Toda and Peter CB Phillips. Vector autoregression and causality: a theoretical overview and simulation study. *Econometric reviews*, 13(2):259–285, 1994.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 38626–38652, 2024a.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. *arXiv preprint arXiv:2407.13278*, 2024b.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. *Advances in Neural Information Processing Systems*, 37:469–498, 2024c.

Xingjian Wu, Jianxin Jin, Wanghui Qiu, Peng Chen, Yang Shu, Bin Yang, and Guo Guo. Aurora: Towards universal generative multimodal time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 103239–103277, 2026a.

Zhaowang Wu, Kaixin Deng, and Hua Yan. Timeseed: Effective time series forecasting with sparse endogenous variables. 2026b.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 3634–3640, 2018.

Guoqi Yu, Jing Zou, Xiaowei Hu, Angelica I Aviles-Rivero, Jing Qin, and Shujun Wang. Revitalizing multivariate time series forecasting: Learnable decomposition with inter-series dependencies and intra-series variations modeling. In *International Conference on Machine Learning*, pp. 57818–57841. PMLR, 2024.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

Jiawen Zhang, Xumeng Wen, Zhenwei Zhang, Shun Zheng, Jia Li, and Jiang Bian. Probts: Benchmarking point and distributional forecasting across diverse prediction horizons. *Advances in Neural Information Processing Systems*, 37:48045–48082, 2024.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.

Yunhao Zhang, Wenyao Hu, Jiale Zheng, Lujia Pan, and Junchi Yan. Mmpd: Diverse time series forecasting via multi-mode patch diffusion loss. In *International Conference on Learning Representations (ICLR)*, 2026.

Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 1234–1241, 2020.

Pengfei Zhou, Yunlong Liu, Junli Liang, Qi Song, and Xiangyang Li. Crosslinear: Plug-and-play cross-correlation embedding for time series forecasting with exogenous variables. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 4120–4131, 2025.

A APPENDIX

This appendix provides additional details, analyses, future work, forecasting visualizations, robustness evaluations, and full experimental results for MeraTSF. It is organized as follows:

- Appendix A.1 presents descriptions of the datasets, baselines, and evaluation metrics.
- Appendix A.2 provides more analyses on efficiency and generalization of MeraTSF.
- Appendix A.3 shows the applicability of MeraTSF in real-world scenarios.
- Appendix A.4 provides a discussion of the limitations of MeraTSF and future work.
- Appendix A.5 evaluates the robustness of MeraTSF.
- Appendix A.7 presents forecasting visualizations on the Ride-Hailing dataset.
- Appendix A.6 reports the full forecasting results on the 12 small-scale datasets and the large-scale Ride-Hailing dataset.

A.1 EXPERIMENTAL DETAILS

Table 6: Summary of datasets. Exo. and Endo. denote exogenous and endogenous variables.

Dataset	Time-stamps	Frequency	#Num	Endo. Descriptions	Exo. Descriptions
NP	52,416	1 hour	2	Nord Pool Electricity Price	Grid Load, Wind Power
PJM	52,416	1 hour	2	Pennsylvania-New Jersey-MarylandElectricity Price	System Load, SyZonal COMED load
BE	52,416	1 hour	2	Belgium's Electricity Price	Generation, System Load
FR	52,416	1 hour	2	France's Electricity Price	Generation, System Load
DE	52,416	1 hour	2	German's Electricity Price	Wind power, Ampirion zonal load
Energy	13,064	1 day	5	Thermoelectric	Battery, Geothermal, Hydroelectric, Solar, Wind
Sdwpfh1/2	14,641	1 hour	6	Active Power	Climate Feature
Sdwpfm1/2	29,282	30 min	6	Active Power	Climate Feature
Colbun	2,958	1 day	2	Water Level	Precipitation, Tributary Inflow
Rapel	3,366	1 day	2	Water Level	Precipitation, Tributary Inflow
Ride-Hailing	**14,025,600**	30 min	9	Online Platform Data	Temporal Indicators, Area ID, Scenario Factors

Datasets. As shown in Table 6, we conduct experiments on 13 real-world datasets with exogenous variables, where future exogenous information is either directly available or can be accurately estimated. These datasets span diverse application domains, including traffic forecasting, electricity markets, power generation, hydropower, and wind energy. Specifically, we adopt five small-scale electricity price forecasting datasets (NP, PJM, BE, FR, and DE) from TimeXer[2] (Wang et al., 2024c), and seven small-scale datasets collected by DAG[3] (Qiu et al., 2026), including Energy, Colbun, Rapel, Sdwpfh1, Sdwpfh2, Sdwpfm1, and Sdwpfm2. For small-scale datasets, we follow the same data splitting protocol as DAG, partitioning the data into training, validation, and test sets with a ratio of 7:1:2.

We further evaluate MeraTSF on the Ride-Hailing dataset, which contains ride-hailing time series from 200 representative areas. Compared with existing benchmarks, Ride-Hailing offers diverse endogenous and exogenous information and better captures the complexity of real-world forecasting scenarios. The exogenous variables include temporal indicators (e.g., half-hour of the day and day of the week), area IDs, weather conditions, and holiday-related factors, including public holidays, traditional festivals, Western festivals, and two types of large-scale events. Table 7 summarizes the occurrence rates of the three exogenous scenarios within the Ride-Hailing dataset. For the Ride-Hailing dataset, we follow the same data splitting protocol as RideBench[4], partitioning the data into training, validation, and test sets with a ratio of 7:1.5:1.5.

Table 7: Occurrence rates of exogenous events in the Ride-Hailing dataset.

Exogenous Event	Rate
Weather Disturbance	9.16%
Public Holiday	9.99%
Traditional Holiday	2.46%
Western Festival	3.49%
Large-scale Event 1	0.30%
Large-scale Event 2	0.23%

[2]https://github.com/thuml/TimeXer

[3]https://github.com/decisionintelligence/DAG

[4]https://anonymous.4open.science/r/ridebench-338A

Baselines. On the large-scale Ride-Hailing dataset, we further compare MeraTSF with a broader range of baselines, including DLinear (Zeng et al., 2023), TimeBase (Huang et al., 2025), SegRNN (Lin et al., 2025b), SparseTSF (Lin et al., 2024b), PMDformer (Hu et al., 2026), TimeFilter (Hu et al., 2025), Crossformer (Zhang & Yan, 2023), TQNet (Lin et al., 2025a), ModernTCN (Luo & Wang, 2024), and iTransformer (Liu et al., 2024). The detailed results are reported in Table 13. Models are optimized with AdamW (Loshchilov & Hutter, 2019) under a unified hyperparameter search protocol. For each method, we search the learning rate from $\{3 \times 10^{-5}, 1 \times 10^{-4}, 3 \times 10^{-4}, 1 \times 10^{-3}, 3 \times 10^{-3}\}$ and report the results under the best configuration. We set the batch size to 128 for the Ride-Hailing dataset. For other datasets, we use dataset-specific batch sizes and further perform a sensitivity analysis (see Figure 11) to investigate the impact of batch size variations on model performance. Early stopping is applied if the validation performance does not improve for three consecutive epochs.

Evaluation Metrics. We follow RideBench for the Ride-Hailing dataset and evaluate point forecasting performance using three standard metrics: Weighted Mean Absolute Percentage Error (WMAPE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). WMAPE measures scale-normalized absolute error, MAE measures average absolute error, and RMSE emphasizes large deviations. Formally,

$$\text{WMAPE} = \frac{\sum_{i=1}^{n} |y_i - \tilde{y}_i|}{\sum_{i=1}^{n} |y_i|}, \qquad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \tilde{y}_i|, \qquad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2}, \tag{13}$$

where y_i and $\tilde{y}_i$ denote the ground-truth and predicted values of the i-th evaluated point, and n is the total number of evaluated points. For the 12 small-scale datasets, we use the same evaluation metrics as DAG (Qiu et al., 2026) and GCGNet (Li et al., 2026), namely Mean Absolute Error (MAE) and Mean Squared Error (MSE), to ensure a fair comparison. The definition of MSE is given as follows:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2, \tag{14}$$

where y_i and $\tilde{y}_i$ denote the ground-truth and predicted values of the i-th evaluated point, and n is the total number of evaluated points.

A.2 More Analyses

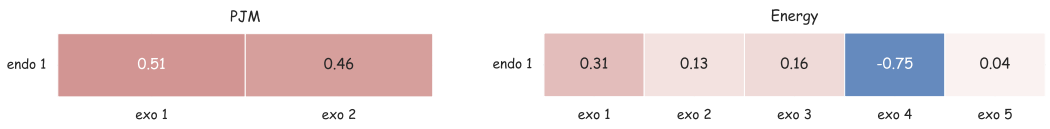


Figure 7: Correlation between endogenous and exogenous variables on the PJM and Energy datasets.

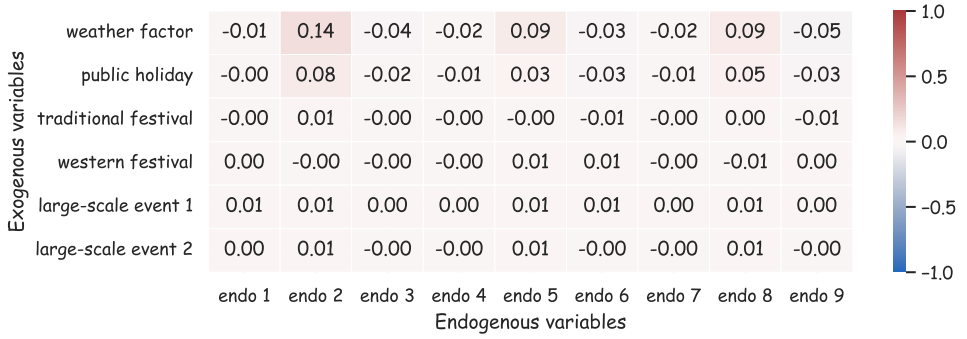


Figure 8: Correlation between endogenous and exogenous variables on the Ride-Hailing dataset.

Existing Exogenous-aware Methods vs. MeraTSF. Existing exogenous-aware methods (Li et al., 2026; Qiu et al., 2026; Wu et al., 2026b) have achieved strong performance on several benchmarks (see Table 2). However, they exhibit limited effectiveness in exploiting exogenous variables on large-scale real-world datasets such as Ride-Hailing (see Table 1). As shown in Figure 7, public datasets often exhibit stronger correlations between exogenous variables and endogenous variables, which can be effectively captured by dependency modeling. In contrast, such correlations become weaker in the Ride-Hailing dataset (see Figure 8), making it challenging for existing methods. On the one hand, discrete exogenous variables often fail to establish stable and generalizable correlations with the target variables. On the other hand, even continuous exogenous variables (e.g., weather factors) may exhibit time-varying correlations with endogenous variables, making correlation-based modeling insufficient.

In contrast, MeraTSF does not rely on explicit correlations between exogenous and endogenous variables. Instead, it leverages a residual refinement mechanism to capture the latent influence of exogenous factors on endogenous variations, thereby avoiding the interference caused by discrete variables and dynamic correlation shifts. Consequently, MeraTSF achieves superior performance on the Ride-Hailing dataset.

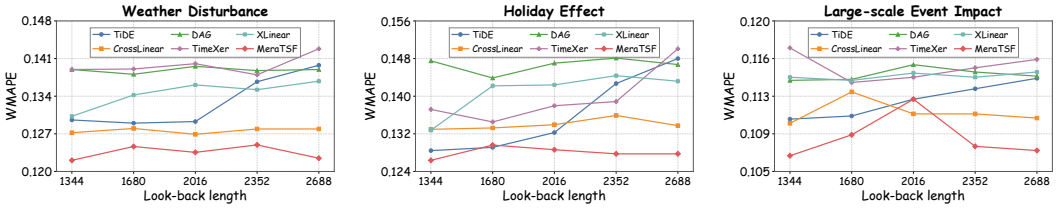


Figure 9: Result of MeraTSF and other models on the Ride-Hailing with different look-back lengths.

Performance with various Look-back Length. Since different forecasting tasks require different historical contexts, we evaluate models with varying look-back lengths. While longer look-back length provide more information for capturing long-term dependencies, they may also introduce noise. Figure 9 presents the performance of different models with varying look-back lengths on the Ride-Hailing dataset, where the output length is fixed to 336 steps.

We observe that: ***i)*** MeraTSF outperforms existing methods across all settings, demonstrating robust forecasting capability; ***ii)*** existing methods rely on inter-variable dependency modeling, where long look-back length may increase the difficulty of extracting useful cross-variable patterns. In contrast, MeraTSF leverages channel-independent residual refinement to reduce reliance on cross-variable interactions and achieve more effective historical information utilization.

A.3 Application Analysis

Table 8: Zero-shot area forecasting results of MeraTSF on Ride-Hailing dataset. The best and second best results are highlighted in **bold** and underlined.

Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
TimeMixer	0.1106	4.323	12.52	0.1422	6.108	19.42	0.1556	6.584	20.18	0.1136	5.122	17.23
CrossLinear	0.1054	4.124	11.71	0.1270	5.437	17.24	0.1338	5.654	18.05	0.1127	5.087	16.90
PatchTST	0.1076	4.205	12.51	0.1408	6.043	19.42	0.1524	6.446	20.32	0.1124	5.068	17.36
DAG	0.1080	4.220	12.32	0.1391	5.966	19.18	0.1462	6.175	19.53	0.1138	5.133	17.26
XLinear	0.1083	4.232	12.14	0.1339	5.737	18.49	0.1429	6.042	18.87	0.1134	5.111	16.96
TimeXer	0.1089	4.275	12.01	0.1387	5.948	18.65	0.1388	5.869	18.49	0.1144	5.161	17.05
TiDE	0.1095	4.273	12.42	0.1413	6.064	19.31	0.1511	6.382	19.86	0.1137	5.123	17.15
MeraTSF	**0.1034**	**4.046**	**11.47**	**0.1234**	**5.279**	**16.96**	**0.1313**	**5.538**	**17.57**	**0.1100**	**4.956**	**16.63**
Improve	-1.90%	-1.89%	-2.05%	-2.83%	-2.91%	-1.62%	-1.87%	-2.05%	-2.66%	-2.14%	-2.21%	-1.60%

Zero-Shot Area Forecasting. Ride-hailing platforms typically exhibit strong spatial heterogeneity (Hao et al., 2026), where demand patterns vary across areas due to differences in commuting structures, weather sensitivity, and local supply-demand dynamics. Therefore, the ability to generalize to unseen areas is crucial for practical deployment. We conduct zero-shot area forecasting

experiments on the Ride-Hailing dataset. Specifically, the 200 areas are randomly divided into two disjoint subsets, with 150 areas used for training and validation and the 50 areas used for testing. As shown in Table 8, MeraTSF outperforms different backbones (Zhou et al., 2025; Nie et al., 2023). These results suggest that MeraTSF transcends mere function approximation, demonstrating its ability to uncover and model the complex interaction patterns between endogenous and exogenous variables.

Forecasting with Increasingly Sparse Exogenous Variables. To further evaluate the robustness of MeraTSF under increasingly sparse observed exogenous events, we randomly remove 5%, 15%, and 25% of event instances from each exogenous event variable in the Ride-Hailing dataset. This procedure reduces the effective event density available to the model while preserving the original event semantics. As shown in Table 9, MeraTSF consistently outperforms existing methods across all sparsity levels, highlighting its robustness to severely sparse exogenous variables. This advantage is particularly pronounced for sparse large-scale events, which constitute only 0.53% of the data. Even in this challenging setting, MeraTSF reduces MAE by 2.78% compared with the strongest baseline, demonstrating its ability to extract reliable predictive signals from highly sparse exogenous information.

Table 9: Forecasting full results of MeraTSF under further sparsified exogenous variables on the Ride-Hailing dataset. The best and second best results are highlighted in **bold** and underlined.

	Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
	Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
TiDE	0%	0.1043	3.996	12.43	0.1293	5.397	18.96	0.1282	5.305	17.29	0.1104	4.866	15.81
	5%	0.1302	5.022	13.84	0.1560	6.534	20.95	0.1652	6.864	20.14	0.1306	5.784	16.64
	15%	0.1094	4.186	13.27	0.1400	5.854	20.72	0.1452	6.015	19.61	0.1144	5.045	16.50
	25%	0.1292	4.980	13.82	0.1550	6.491	20.95	0.1648	6.844	20.15	0.1302	5.766	16.65
	Avg.	0.1183	4.546	13.34	0.1451	6.069	20.40	0.1509	6.257	19.30	0.1214	5.365	16.40
DAG	0%	0.1090	4.174	13.30	0.1388	5.803	20.64	0.1476	6.115	19.91	0.1141	5.033	16.55
	5%	0.1092	4.182	13.26	0.1380	5.772	20.56	0.1471	6.098	19.82	0.1149	5.072	16.56
	15%	0.1090	4.174	13.26	0.1380	5.769	20.59	0.1462	6.056	19.76	0.1134	5.003	16.47
	25%	0.1089	4.173	13.29	0.1384	5.789	20.63	0.1473	6.104	19.89	0.1146	5.055	16.56
	Avg.	0.1090	4.176	13.28	0.1383	5.783	20.61	0.1471	6.093	19.85	0.1143	5.041	16.54
TimeXer	0%	0.1093	4.207	12.93	0.1388	5.804	20.14	0.1371	5.687	18.42	0.1172	5.177	16.40
	5%	0.1105	4.252	13.06	0.1399	5.848	20.31	0.1390	5.765	18.78	0.1155	5.099	16.49
	15%	0.1105	4.250	13.06	0.1402	5.864	20.35	0.1383	5.738	18.76	0.1154	5.097	16.42
	25%	0.1096	4.219	12.74	0.1371	5.730	19.75	0.1346	5.584	18.04	0.1139	5.032	16.00
	Avg.	0.1100	4.232	12.95	0.1390	5.812	20.14	0.1373	5.694	18.50	0.1155	5.101	16.33
XLinear	0%	0.1075	4.124	12.86	0.1300	5.431	19.42	0.1326	5.493	18.52	0.1144	5.055	16.30
	5%	0.1124	4.320	13.16	0.1368	5.720	19.95	0.1475	6.118	19.32	0.1169	5.165	16.44
	15%	0.1125	4.318	13.18	0.1370	5.726	19.99	0.1477	6.127	19.37	0.1171	5.175	16.49
	25%	0.1126	4.322	13.21	0.1373	5.743	20.07	0.1478	6.130	19.43	0.1173	5.181	16.53
	Avg.	0.1113	4.271	13.10	0.1353	5.655	19.86	0.1439	5.967	19.16	0.1164	5.144	16.44
CrossLinear	0%	0.1051	4.032	12.65	0.1269	5.296	18.74	0.1328	5.505	18.43	0.1100	4.854	15.87
	5%	0.1057	4.054	12.68	0.1277	5.330	18.91	0.1314	5.442	18.14	0.1111	4.898	15.93
	15%	0.1068	4.101	12.80	0.1284	5.363	18.96	0.1345	5.581	18.62	0.1134	5.010	16.11
	25%	0.1063	4.075	12.76	0.1286	5.369	19.05	0.1330	5.511	18.40	0.1115	4.922	16.02
	Avg.	0.1060	4.066	12.72	0.1279	5.340	18.92	0.1329	5.510	18.40	0.1115	4.921	15.98
MeraTSF	0%	0.1027	3.943	12.01	0.1217	5.074	17.77	0.1261	5.216	16.65	0.1069	4.712	15.14
	5%	0.1039	3.984	12.46	0.1266	5.282	18.50	0.1309	5.419	18.30	0.1078	4.748	15.51
	15%	0.1049	4.022	12.58	0.1280	5.346	18.74	0.1336	5.535	18.43	0.1094	4.822	15.74
	25%	0.1053	4.035	12.66	0.1288	5.380	18.88	0.1345	5.572	18.62	0.1101	4.855	15.85
	Avg.	**0.1042**	**3.996**	**12.43**	**0.1263**	**5.271**	**18.47**	**0.1313**	**5.436**	**18.00**	**0.1086**	**4.784**	**15.56**
	Improve	-1.67%	-1.71%	-2.32%	-1.27%	-1.29%	-2.34%	-1.24%	-1.35%	-2.16%	-2.65%	-2.78%	-2.64%

Forecasting with Sparse Endogenous Variables. In real-world applications, time series data are often incomplete due to limitations in data collection or transmission (Wu et al., 2026b; Lin et al., 2024b). To evaluate the robustness of MeraTSF to sparse endogenous variables, we conduct experiments by randomly masking endogenous variables at different missing ratios (i.e., 5%, 15%, and 25%). For the missing endogenous variables, we employ zero imputation. And the full results are shown in Table 10.

MeraTSF consistently outperforms existing methods across different missing ratios, demonstrating its robustness to increasingly sparse endogenous variables. We attribute this robustness to its multi-scale architecture (Wang et al., 2024a; Wu et al., 2026b): the short-scale branch is designed to capture local and transient variations, while the long-scale branch focuses on smooth and persistent

trends. Consequently, when the representations at one scale are adversely affected by missing endogenous variables, the complementary information captured at other scales can still contribute to the final prediction.

Table 10: Results under randomly missing values of MeraTSF on Ride-Hailing dataset. The best and second best results are highlighted in **bold** and underlined.

	Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
	Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
SparseTSF	0%	0.1134	4.342	13.60	0.1420	5.941	20.99	0.1553	6.442	20.53	0.1176	5.193	16.75
	5%	0.1164	4.460	13.64	0.1436	6.009	20.97	0.1563	6.482	20.40	0.1213	5.363	16.85
	15%	0.1226	4.706	13.87	0.1483	6.198	21.10	0.1606	6.655	20.51	0.1289	5.702	17.21
	25%	0.1321	5.070	14.16	0.1554	6.482	21.27	0.1665	6.886	20.58	0.1386	6.121	17.60
	Avg.	0.1211	4.645	13.82	0.1473	6.158	21.08	0.1597	6.616	20.51	0.1266	5.595	17.10
TimeXer	0%	0.1093	4.207	12.93	0.1388	5.804	20.14	0.1371	5.687	18.42	0.1172	5.177	16.40
	5%	0.1095	4.217	12.87	0.1344	5.614	19.46	0.1381	5.728	18.86	0.1149	5.075	16.19
	15%	0.1122	4.327	12.94	0.1360	5.679	19.41	0.1389	5.765	18.68	0.1178	5.209	16.40
	25%	0.1138	4.384	13.17	0.1383	5.776	19.75	0.1398	5.798	19.01	0.1212	5.358	16.68
	Avg.	0.1112	4.284	12.98	0.1369	5.718	19.69	0.1385	5.745	18.74	0.1178	5.205	16.42
XLinear	0%	0.1075	4.124	12.86	0.1300	5.431	19.42	0.1326	5.493	18.52	0.1144	5.055	16.30
	5%	0.1111	4.263	13.10	0.1359	5.681	19.98	0.1429	5.924	18.93	0.1159	5.115	16.31
	15%	0.1127	4.329	13.16	0.1373	5.736	20.04	0.1445	5.990	19.02	0.1176	5.190	16.40
	25%	0.1139	4.374	13.19	0.1379	5.763	20.06	0.1452	6.019	19.02	0.1185	5.230	16.43
	Avg.	0.1113	4.273	13.08	0.1353	5.653	19.88	0.1413	5.857	18.87	0.1166	5.148	16.36
CrossLinear	0%	0.1051	4.032	12.65	0.1269	5.296	18.74	0.1328	5.505	18.43	0.1100	4.854	15.87
	5%	0.1066	4.090	12.72	0.1279	5.339	18.73	0.1337	5.540	18.42	0.1124	4.963	15.97
	15%	0.1082	4.152	12.82	0.1291	5.391	18.86	0.1358	5.634	18.48	0.1136	5.021	16.04
	25%	0.1090	4.187	12.84	0.1299	5.423	18.89	0.1370	5.685	18.73	0.1155	5.101	16.22
	Avg.	0.1072	4.115	12.76	0.1285	5.362	18.81	0.1348	5.591	18.52	0.1129	4.985	16.03
MeraTSF	0%	0.1027	3.943	12.01	0.1217	5.074	17.77	0.1261	5.216	16.65	0.1069	4.712	15.14
	5%	0.1050	4.028	12.48	0.1257	5.245	18.28	0.1304	5.399	18.39	0.1102	4.858	15.58
	15%	0.1072	4.113	12.56	0.1273	5.307	18.32	0.1327	5.495	18.52	0.1125	4.961	15.66
	25%	0.1081	4.145	12.66	0.1294	5.393	18.62	0.1351	5.591	18.72	0.1127	4.964	15.72
	Avg.	**0.1058**	**4.057**	**12.43**	**0.1260**	**5.255**	**18.25**	**0.1311**	**5.425**	**18.07**	**0.1106**	**4.874**	**15.53**
	Improve	-1.38%	-1.41%	-2.59%	-1.89%	-2.00%	-2.96%	-2.78%	-2.96%	-2.40%	-2.04%	-2.23%	-3.12%

Forecasting without Future Exogenous Variables. In real-world applications, future exogenous variables are often unavailable or difficult to obtain in a timely manner. To evaluate the robustness of MeraTSF under this practical setting, we conduct experiments on 12 small-scale datasets using only historical endogenous and exogenous variables.

As shown in Table 11, MeraTSF does not outperform DAG (Qiu et al., 2026) under this setting. This is mainly because MeraTSF relies on future exogenous information to guide residual refinement, and the absence of such information limits its ability to accurately capture exogenous-driven residual patterns. Therefore, its advantage is naturally limited when future exogenous variables are unavailable. Nevertheless, these results also highlight that MeraTSF can effectively leverage future exogenous information and achieve more accurate forecasting when it is accessible.

A.4 Limitation and Future Work

Limitation. Although MeraTSF demonstrates its efficacy across diverse scenarios, it still has several practical limitations worth noting. ***i)*** **Limited Information Utilization.** As the look-back length increases, MeraTSF does not consistently improve, indicating limited capability in exploiting additional historical information. Moreover, due to its channel-independent modeling paradigm, MeraTSF does not explicitly capture correlations among endogenous variables. ***ii)*** **Sensitivity to Future Exogenous Information.** As shown in Table 11, when future exogenous variables are unavailable, MeraTSF underperforms DAG (Qiu et al., 2026). This is mainly because the absence of future exogenous information weakens the contextual guidance provided by exogenous variables, making it more challenging to capture residual patterns driven by exogenous variables.

Future Work. While MeraTSF effectively captures temporal heterogeneity induced by exogenous factors, the spatial heterogeneity arising from intrinsic differences across urban areas remains unexplored. Figure 10 reveals substantial variations in mobility patterns across different areas under normal operating conditions: some areas exhibit pronounced morning and evening commuting peaks, whereas others demonstrate stronger nighttime activity. These spatial variations are likely driven by

diverse urban characteristics, including functional zoning, land-use patterns, commuting behaviors, commercial activities, and population density. Such complex spatial variations introduce fundamental challenges in learning reliable inter-area dependencies and developing forecasting models with strong generalization capabilities across heterogeneous urban environments. Future work will investigate spatially adaptive representations to characterize latent area-specific heterogeneity, enabling more effective modeling of spatial dependencies and improving cross-area prediction generalization.

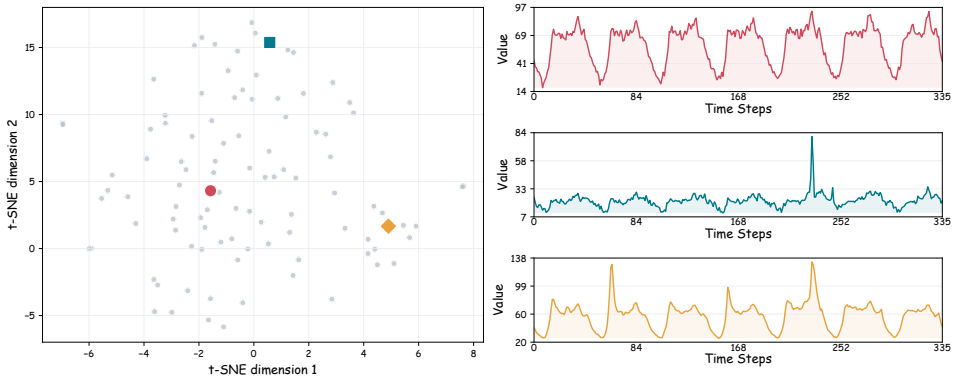


Figure 10: Heterogeneous patterns across different areas on the Ride-Hailing dataset.

Table 11: Results on 12 real-world datasets under the setting where future exogenous variables are not available. The best results are **bold**, and the second best results are underline.

Method		MeraTSF		DAG		GCGNet		TimeXer		TFT		TiDE		DUET		CrossLinear		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.214**	**0.252**	0.237	0.259	0.240	0.260	0.270	0.292	0.330	0.321	0.297	0.309	0.262	0.271	0.246	0.282	0.283	0.296
	360	0.636	0.494	**0.601**	0.500	0.620	0.502	0.609	**0.474**	0.964	0.655	0.638	0.523	0.626	0.494	0.655	0.505	0.630	0.505
	Avg	0.425	**0.373**	**0.419**	0.380	0.430	0.381	0.440	0.383	0.647	0.488	0.467	0.416	0.444	0.383	0.451	0.394	0.457	0.401
PJM	24	**0.068**	0.163	0.072	**0.162**	0.072	0.164	0.096	0.191	0.129	0.231	0.105	0.213	0.082	0.178	0.096	0.198	0.106	0.212
	360	0.189	0.263	**0.180**	**0.259**	0.191	0.269	0.186	0.267	0.270	0.310	0.210	0.292	0.198	0.273	0.197	0.283	0.189	0.273
	Avg	0.128	0.213	**0.126**	**0.210**	0.131	0.216	0.141	0.229	0.200	0.270	0.158	0.253	0.140	0.226	0.147	0.241	0.148	0.243
BE	24	0.372	0.248	0.376	**0.244**	**0.365**	0.259	0.392	0.253	0.400	0.267	0.492	0.322	0.376	0.253	0.389	0.253	0.405	0.264
	360	0.554	0.351	**0.547**	0.350	0.553	0.354	0.563	0.349	0.727	0.442	0.602	0.375	0.571	0.357	0.564	**0.346**	0.564	0.354
	Avg	0.463	0.299	0.462	**0.297**	**0.459**	0.306	0.477	0.301	0.563	0.354	0.547	0.348	0.473	0.305	0.477	0.300	0.485	0.309
FR	24	0.364	0.193	**0.345**	**0.190**	0.352	0.199	0.366	0.195	0.445	0.231	0.415	0.254	0.359	0.208	0.397	0.204	0.397	0.223
	360	0.548	0.310	0.526	0.307	**0.519**	**0.293**	0.542	0.300	0.624	0.345	0.574	0.325	0.577	0.317	0.555	0.310	0.542	0.306
	Avg	0.456	0.252	**0.435**	0.249	**0.435**	**0.246**	0.454	0.247	0.535	0.288	0.494	0.290	0.468	0.262	0.476	0.257	0.470	0.264
DE	24	**0.410**	**0.401**	0.422	0.411	0.413	0.402	0.501	0.445	0.576	0.452	0.465	0.433	0.424	0.403	0.438	0.423	0.503	0.450
	360	0.787	0.570	**0.784**	**0.563**	0.816	0.600	0.818	0.568	0.792	0.578	0.823	0.604	0.895	0.623	0.832	0.592	0.889	0.604
	Avg	**0.599**	**0.486**	0.603	0.487	0.614	0.501	0.659	0.507	0.684	0.515	0.644	0.519	0.660	0.513	0.635	0.508	0.696	0.527
Energy	24	0.102	**0.245**	0.112	0.257	0.112	0.259	0.138	0.293	0.352	0.463	0.117	0.265	0.108	0.254	**0.102**	0.246	0.108	0.254
	360	0.229	0.378	**0.189**	**0.342**	0.195	0.351	0.206	0.360	0.399	0.497	0.207	0.358	0.211	0.362	0.299	0.425	0.298	0.428
	Avg	0.166	0.312	**0.150**	**0.300**	0.154	0.305	0.172	0.326	0.376	0.480	0.162	0.311	0.160	0.308	0.201	0.336	0.203	0.341
Sdwpfm1	24	0.567	0.495	**0.532**	0.512	0.537	0.509	0.558	0.533	0.596	0.531	0.561	0.540	0.550	**0.495**	0.549	0.553	0.557	0.557
	360	0.918	**0.663**	**0.845**	0.695	0.885	0.682	**0.845**	0.684	1.372	0.925	0.864	0.725	0.898	0.672	1.070	0.834	0.910	0.745
	Avg	0.743	**0.579**	**0.689**	0.604	0.711	0.595	0.701	0.609	0.984	0.728	0.713	0.633	0.724	0.583	0.809	0.694	0.733	0.651
Sdwpfm2	24	0.658	0.542	**0.609**	0.548	0.630	0.547	0.627	0.570	0.749	0.657	0.636	0.582	0.612	**0.537**	0.635	0.586	0.648	0.610
	360	1.018	**0.712**	**0.962**	0.766	0.983	0.739	0.978	0.736	1.340	0.877	0.995	0.783	1.028	0.745	0.965	0.788	1.020	0.819
	Avg	0.838	**0.627**	**0.786**	0.657	0.807	0.643	0.803	0.653	1.044	0.767	0.816	0.682	0.820	0.641	0.800	0.687	0.834	0.714
Sdwpfh1	24	0.713	0.576	**0.647**	0.592	0.651	0.578	0.651	0.587	0.754	0.670	0.719	0.641	0.673	**0.575**	0.687	0.652	0.669	0.629
	360	0.913	0.691	**0.820**	0.695	0.830	**0.677**	0.841	0.700	0.831	0.727	0.897	0.757	0.886	0.712	0.849	0.752	0.939	0.804
	Avg	0.813	0.633	**0.733**	0.643	0.741	**0.628**	0.746	0.643	0.793	0.698	0.808	0.699	0.779	0.644	0.768	0.702	0.804	0.717
Sdwpfh2	24	0.835	**0.633**	**0.739**	0.635	0.798	0.645	0.820	0.677	0.811	0.714	0.783	0.674	0.975	0.694	0.810	0.718	0.769	0.680
	360	1.056	0.749	1.000	0.755	0.974	0.757	**0.962**	0.761	1.042	0.779	1.056	0.829	1.039	**0.735**	1.102	0.830	1.280	0.925
	Avg	0.945	**0.691**	**0.870**	0.695	0.886	0.701	0.891	0.719	0.926	0.746	0.919	0.751	1.007	0.715	0.956	0.774	1.025	0.802
Colbun	10	0.068	**0.091**	0.072	0.103	0.070	0.094	0.084	0.146	0.449	0.279	0.095	0.140	0.073	0.106	**0.068**	0.101	0.083	0.125
	30	0.237	0.291	**0.161**	**0.208**	0.168	0.236	0.181	0.293	0.663	0.494	0.282	0.341	0.221	0.290	0.190	0.289	0.217	0.316
	Avg	0.153	0.191	**0.117**	**0.155**	0.119	0.165	0.132	0.219	0.556	0.386	0.188	0.240	0.147	0.198	0.129	0.195	0.150	0.221
Rapel	10	0.206	0.228	**0.195**	**0.215**	0.217	0.221	0.301	0.308	0.238	0.274	0.238	0.270	0.221	0.225	0.201	0.224	0.215	0.232
	30	0.381	0.380	0.327	0.347	0.313	**0.346**	0.387	0.416	0.377	0.394	0.409	0.444	0.387	0.386	**0.278**	0.374	0.436	0.436
	Avg	0.294	0.304	0.261	**0.281**	0.265	0.284	0.344	0.362	0.308	0.334	0.323	0.357	0.304	0.306	**0.240**	0.299	0.325	0.334

A.5 Parameter Sensitivity

We conduct sensitivity analyses to evaluate the robustness of MeraTSF. Table 12 presents results across 12 small-scale datasets under different random seeds (2021, 2022, and 2023). For rigorous evaluation and reproducibility, we report the averaged results across two Python environments (Python 3.10 and Python 3.11) for each seed. MeraTSF exhibits stable performance with low variance, indicating its robustness to random initialization. Figure 11 illustrates the performance of MeraTSF with different patch sizes and batch sizes. The results show that MeraTSF achieves consistent performance across diverse hyperparameter configurations, indicating its robustness and insensitivity to variations in these parameters.

Table 12: Performance of MeraTSF under different random seeds. The Avg row averages the two forecasting horizons for each seed. Mean and Std in the Avg row are computed across the per-seed averages, with Std denoting the sample standard deviation. The best result is highlighted in **bold**.

Seed		2021		2022		2023		Mean		Std	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	0.161	0.211	0.161	**0.207**	**0.158**	0.208	0.160	0.209	0.002	0.002
	360	**0.482**	**0.430**	0.504	0.442	0.504	0.434	0.497	0.435	0.013	0.006
	Avg	**0.322**	**0.321**	0.333	0.325	0.331	0.321	0.328	0.322	0.006	0.002
PJM	24	0.055	**0.144**	**0.054**	0.145	0.056	0.146	0.055	0.145	0.001	0.001
	360	0.105	0.214	**0.103**	0.216	**0.103**	**0.213**	0.104	0.214	0.001	0.002
	Avg	0.080	**0.179**	**0.079**	0.181	0.080	0.180	0.079	0.180	0.001	0.001
BE	24	0.348	0.228	0.351	0.229	**0.346**	**0.227**	0.348	0.228	0.003	0.001
	360	0.482	0.312	0.484	0.317	**0.473**	**0.311**	0.480	0.313	0.006	0.004
	Avg	0.415	0.270	0.418	0.273	**0.410**	**0.269**	0.414	0.271	0.004	0.002
FR	24	**0.323**	**0.172**	0.331	0.174	0.324	0.173	0.326	0.173	0.004	0.001
	360	**0.475**	0.265	0.478	0.266	**0.475**	**0.264**	0.476	0.265	0.002	0.001
	Avg	**0.399**	**0.219**	0.405	0.220	0.400	**0.219**	0.401	0.219	0.003	0.001
DE	24	0.251	0.314	**0.244**	**0.310**	0.258	0.317	0.251	0.314	0.007	0.003
	360	**0.431**	**0.411**	0.444	0.417	0.432	0.412	0.436	0.413	0.007	0.003
	Avg	**0.341**	**0.363**	0.344	0.364	0.345	0.365	0.343	0.364	0.002	0.001
Energy	24	0.051	**0.174**	**0.050**	0.175	0.052	0.177	0.051	0.175	0.001	0.001
	360	**0.132**	**0.289**	0.138	0.291	0.143	0.298	0.138	0.293	0.005	0.005
	Avg	**0.092**	**0.232**	0.094	0.233	0.098	0.238	0.094	0.234	0.003	0.003
Sdwpfm1	24	0.350	**0.401**	0.357	0.404	**0.345**	**0.401**	0.350	0.402	0.006	0.002
	360	**0.445**	**0.497**	0.450	0.501	0.448	0.501	0.448	0.500	0.003	0.002
	Avg	0.398	**0.449**	0.404	0.453	**0.397**	0.451	0.399	0.451	0.004	0.002
Sdwpfm2	24	0.405	0.439	0.400	0.442	**0.390**	**0.431**	0.398	0.437	0.008	0.005
	360	0.512	0.533	0.513	0.535	**0.492**	**0.526**	0.506	0.532	0.012	0.005
	Avg	0.459	0.486	0.457	0.489	**0.441**	**0.479**	0.452	0.484	0.010	0.005
Sdwpfh1	24	0.385	0.441	**0.380**	**0.436**	**0.380**	0.440	0.382	0.439	0.003	0.002
	360	0.499	0.537	0.494	0.534	**0.483**	**0.527**	0.492	0.533	0.008	0.005
	Avg	0.442	0.489	0.437	0.485	**0.432**	**0.484**	0.437	0.486	0.005	0.003
Sdwpfh2	24	**0.433**	**0.468**	0.436	0.481	0.437	0.477	0.435	0.475	0.002	0.006
	360	**0.536**	**0.556**	0.540	0.561	0.556	0.566	0.544	0.561	0.010	0.005
	Avg	**0.485**	**0.512**	0.488	0.521	0.497	0.522	0.490	0.518	0.006	0.005
Colbun	10	**0.055**	0.086	0.058	**0.083**	0.058	0.087	0.057	0.086	0.001	0.002
	30	**0.146**	0.251	0.159	0.258	0.149	**0.247**	0.151	0.252	0.007	0.005
	Avg	**0.101**	0.169	0.109	0.171	0.104	**0.167**	0.104	0.169	0.004	0.002
Rapel	10	0.177	0.219	**0.174**	0.213	0.182	**0.209**	0.178	0.214	0.004	0.005
	30	**0.317**	0.385	**0.317**	0.383	**0.317**	**0.378**	0.317	0.382	0.001	0.004
	Avg	0.247	0.302	**0.246**	0.298	0.250	**0.294**	0.247	0.298	0.002	0.004

A.6 Full Results

Table 13 further presents the performance comparison results on the large-scale Ride-Hailing dataset. The full results across different forecasting horizons on 12 small-scale datasets are provided in Table 14.

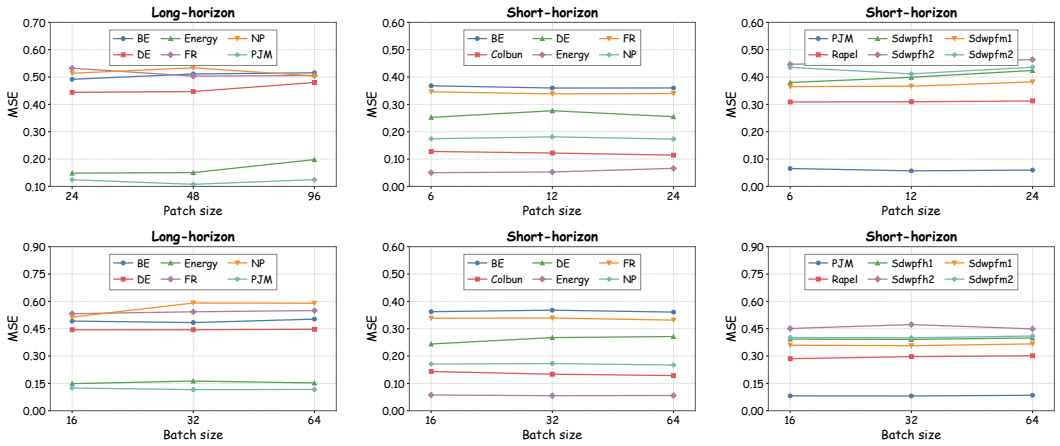


Figure 11: Parameter sensitivity studies of main hyper-parameters in MeraTSF.

Table 13: Full results on the Ride-Hailing dataset. The best and second best results are highlighted in **bold** and underlined, respectively.

	Scenarios	Overall			Weather Disturbance			Holiday Effect			Large-scale Event Impact		
	Metrics	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE	WMAPE	MAE	RMSE
Channel-Independent	PatchTST	0.1079	4.132	13.38	0.1391	5.821	20.84	0.1507	6.249	20.37	0.1130	4.983	16.56
	PETformer	0.1078	4.124	13.35	0.1396	5.839	20.86	0.1520	6.303	20.42	0.1130	4.982	16.48
	TimeMixer	0.1113	4.261	13.47	0.1411	5.902	20.90	0.1548	6.420	20.42	0.1159	5.113	16.73
	STID	0.1122	4.299	13.48	0.1419	5.941	20.92	0.1542	6.397	20.22	0.1166	5.149	16.74
	DLinear	0.1171	4.483	13.48	0.1441	6.029	20.64	0.1545	6.403	19.78	0.1186	5.234	16.29
	TimeBase	0.1222	4.674	13.68	0.1437	6.008	20.38	0.1553	6.437	19.67	0.1209	5.342	16.19
	SegRNN	0.1116	4.275	13.60	0.1418	5.932	20.98	0.1533	6.360	20.48	0.1196	5.280	17.03
	RMLP	0.1135	4.344	13.51	0.1431	5.986	20.91	0.1562	6.476	20.39	0.1176	5.185	16.67
	SparseTSF	0.1134	4.342	13.60	0.1420	5.941	20.99	0.1553	6.442	20.53	0.1176	5.193	16.75
	PhaseFormer	0.1151	4.411	13.69	0.1432	5.993	21.03	0.1529	6.339	20.26	0.1226	5.423	17.07
Multivariate	PMDformer	0.1077	4.122	13.28	0.1384	5.790	20.70	0.1486	6.157	20.01	0.1125	4.962	16.39
	TimeFilter	0.1099	4.206	13.30	0.1390	5.812	20.67	0.1494	6.187	20.00	0.1137	5.012	16.42
	Crossformer	0.1096	4.199	13.42	0.1390	5.816	20.76	0.1533	6.355	20.46	0.1141	5.039	16.59
	SOFTS	0.1111	4.252	13.45	0.1408	5.888	20.84	0.1533	6.354	20.29	0.1159	5.112	16.76
	ModernTCN	0.1117	4.279	13.42	0.1416	5.925	20.86	0.1512	6.269	20.11	0.1171	5.171	16.78
	TQNet	0.1117	4.277	13.44	0.1413	5.910	20.83	0.1537	6.373	20.25	0.1161	5.121	16.66
	iTransformer	0.1117	4.276	13.47	0.1410	5.899	20.88	0.1518	6.290	20.16	0.1169	5.159	16.79
	Leddam	0.1117	4.280	13.47	0.1417	5.928	20.87	0.1534	6.361	20.30	0.1163	5.133	16.68
	CrossGNN	0.1125	4.309	13.59	0.1417	5.926	21.00	0.1542	6.392	20.45	0.1170	5.162	16.81
	DUET	0.1147	4.397	13.57	0.1437	6.011	20.97	0.1539	6.379	20.23	0.1193	5.267	16.86
Exogenous	TiDE	0.1043	3.996	12.43	0.1293	5.397	18.96	0.1282	5.305	17.29	0.1104	4.866	15.81
	CrossLinear	0.1051	4.032	12.65	0.1269	5.296	18.74	0.1328	5.505	18.43	0.1100	4.854	15.87
	XLinear	0.1075	4.124	12.86	0.1300	5.431	19.42	0.1326	5.493	18.52	0.1144	5.055	16.30
	CATS	0.1103	4.225	12.95	0.1321	5.510	19.53	0.1379	5.703	18.76	0.1159	5.106	16.13
	DAG	0.1090	4.174	13.30	0.1388	5.803	20.64	0.1476	6.115	19.91	0.1141	5.033	16.55
	TimeXer	0.1093	4.207	12.93	0.1388	5.804	20.14	0.1371	5.687	18.42	0.1172	5.177	16.40
	MeraTSF	**0.1027**	**3.943**	**12.01**	**0.1217**	**5.074**	**17.77**	**0.1261**	**5.216**	**16.65**	**0.1069**	**4.712**	**15.14**

A.7 Forecasting Visualization

Figures 12, 13, 14, and 15 visualize forecasting results on the Ride-Hailing dataset under holiday effect and large-scale event impact scenarios.

Table 14: Full results on 12 small-scale datasets with exogenous variables. The best results are **bold**, and the second best results are underline.

Method		MeraTSF		DAG		GCGNet		TimeXer		TFT		TiDE		DUET		CrossLinear		PatchTST	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.161**	**0.213**	0.202	0.237	0.208	0.237	0.236	0.266	0.219	0.249	0.284	0.301	0.246	0.287	0.210	0.266	0.249	0.294
	360	**0.482**	**0.430**	0.521	0.451	0.531	0.459	0.600	0.475	0.539	0.501	0.601	0.498	0.576	0.528	0.531	0.508	0.530	0.498
	Avg	**0.322**	**0.322**	0.362	0.344	0.370	0.348	0.418	0.371	0.379	0.375	0.443	0.400	0.411	0.408	0.371	0.387	0.390	0.396
PJM	24	**0.055**	0.145	0.057	**0.143**	0.060	0.150	0.075	0.166	0.095	0.195	0.106	0.214	0.072	0.166	0.088	0.191	0.116	0.239
	360	**0.105**	**0.215**	0.130	0.218	0.129	0.223	0.140	0.231	0.133	0.219	0.177	0.279	0.131	0.228	0.135	0.254	0.149	0.279
	Avg	**0.080**	**0.180**	0.093	**0.180**	0.095	0.187	0.108	0.198	0.114	0.207	0.142	0.246	0.102	0.197	0.112	0.223	0.133	0.259
BE	24	**0.348**	0.233	0.361	**0.229**	0.350	0.248	0.392	0.253	0.426	0.272	0.426	0.285	0.432	0.272	0.391	0.259	0.452	0.326
	360	**0.482**	0.321	0.485	0.330	0.511	0.340	0.512	0.327	**0.482**	**0.310**	0.571	0.364	0.597	0.436	0.568	0.416	0.702	0.538
	Avg	**0.415**	**0.277**	0.423	0.279	0.431	0.294	0.452	0.290	0.454	0.291	0.498	0.325	0.515	0.354	0.479	0.337	0.577	0.432
FR	24	**0.323**	0.173	0.355	**0.171**	0.347	0.188	0.366	0.208	0.543	0.253	0.418	0.255	0.384	0.251	0.390	0.226	0.518	0.368
	360	0.475	0.265	0.473	0.268	0.482	0.279	0.489	0.273	**0.465**	**0.261**	0.551	0.308	0.607	0.403	0.575	0.370	0.658	0.452
	Avg	**0.399**	**0.219**	0.414	**0.219**	0.415	0.234	0.427	0.241	0.504	0.257	0.484	0.281	0.496	0.327	0.483	0.298	0.588	0.410
DE	24	**0.251**	**0.314**	0.277	0.322	0.280	0.331	0.339	0.362	0.380	0.383	0.367	0.383	0.376	0.378	0.387	0.396	0.413	0.412
	360	**0.431**	**0.411**	0.462	0.418	0.523	0.447	0.610	0.474	0.599	0.509	0.630	0.511	0.589	0.482	0.583	0.507	0.589	0.498
	Avg	**0.341**	**0.363**	0.370	0.370	0.401	0.389	0.475	0.418	0.489	0.446	0.499	0.447	0.482	0.430	0.485	0.452	0.501	0.455
Energy	24	**0.051**	**0.174**	0.079	0.215	0.081	0.221	0.122	0.273	0.093	0.235	0.103	0.248	0.117	0.283	0.241	0.418	0.239	0.390
	360	**0.132**	**0.289**	0.169	0.320	0.182	0.334	0.204	0.357	0.167	0.331	0.202	0.355	0.288	0.452	0.237	0.385	0.214	0.363
	Avg	**0.092**	**0.232**	0.124	0.267	0.131	0.277	0.163	0.315	0.130	0.283	0.153	0.302	0.203	0.367	0.239	0.402	0.226	0.377
Sdwpfm1	24	**0.350**	0.401	0.351	**0.400**	0.376	0.415	0.558	0.533	0.366	0.421	0.474	0.488	0.551	0.564	0.355	0.473	0.374	0.454
	360	**0.445**	**0.497**	0.495	0.522	0.472	0.499	0.845	0.684	0.597	0.528	0.492	0.526	0.646	0.577	0.497	0.532	0.497	0.550
	Avg	**0.397**	**0.449**	0.423	0.461	0.424	0.457	0.701	0.609	0.482	0.474	0.483	0.507	0.599	0.570	0.426	0.502	0.435	0.502
Sdwpfm2	24	0.390	0.431	**0.372**	**0.414**	0.421	0.441	0.627	0.570	0.411	0.458	0.461	0.492	0.445	0.452	0.477	0.536	0.418	0.492
	360	0.492	0.526	0.583	0.556	0.529	0.531	0.978	0.736	0.541	**0.519**	**0.511**	0.540	0.584	0.528	0.589	0.611	0.602	0.603
	Avg	**0.441**	**0.479**	0.477	0.485	0.475	0.486	0.803	0.653	0.476	0.488	0.486	0.516	0.514	0.490	0.533	0.573	0.510	0.547
Sdwpfh1	24	**0.380**	0.440	0.408	**0.438**	0.435	0.486	0.651	0.587	0.401	0.460	0.434	0.489	0.527	0.513	0.548	0.585	0.422	0.505
	360	0.483	0.527	0.489	0.534	**0.465**	**0.514**	0.841	0.700	0.557	0.523	0.472	0.527	0.551	0.519	0.566	0.601	0.513	0.548
	Avg	**0.432**	**0.484**	0.448	0.486	0.450	0.500	0.746	0.643	0.479	0.491	0.453	0.508	0.539	0.516	0.557	0.593	0.468	0.527
Sdwpfh2	24	**0.433**	0.468	0.438	**0.465**	0.473	0.506	0.820	0.677	0.474	0.493	0.579	0.553	0.629	0.563	0.468	0.540	0.457	0.527
	360	**0.536**	0.564	0.608	0.595	0.566	0.565	0.962	0.761	0.657	**0.549**	0.619	0.614	0.665	0.569	0.608	0.609	0.641	0.632
	Avg	**0.484**	**0.516**	0.523	0.530	0.520	0.536	0.891	0.719	0.566	0.521	0.599	0.583	0.647	0.566	0.538	0.574	0.549	0.580
Colbun	10	**0.055**	**0.088**	0.061	0.094	0.065	0.108	0.113	0.172	0.092	0.135	0.089	0.131	0.089	0.134	0.071	0.102	0.062	0.122
	30	0.146	0.251	**0.135**	**0.215**	0.149	0.243	0.176	0.299	0.383	0.460	0.240	0.322	0.307	0.397	0.182	0.288	0.417	0.496
	Avg	0.101	0.169	**0.098**	**0.154**	0.107	0.175	0.145	0.235	0.238	0.297	0.164	0.227	0.198	0.266	0.126	0.195	0.239	0.309
Rapel	10	0.177	0.219	**0.151**	**0.203**	0.211	0.230	0.301	0.308	0.201	0.253	0.228	0.271	0.174	0.219	0.163	0.209	0.163	0.218
	30	0.317	0.385	**0.309**	0.408	0.401	**0.384**	0.387	0.416	0.409	0.414	0.411	0.432	0.365	0.432	0.340	0.417	0.374	0.445
	Avg	0.247	**0.302**	**0.230**	0.305	0.306	0.307	0.344	0.362	0.305	0.333	0.320	0.351	0.269	0.326	0.252	0.313	0.269	0.332

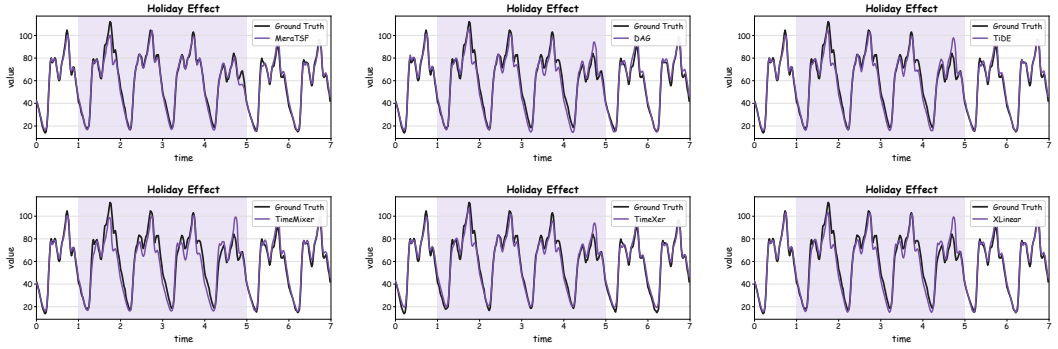


Figure 12: Forecasting visualization on the Ride-Hailing under holiday effects (Sample 1).

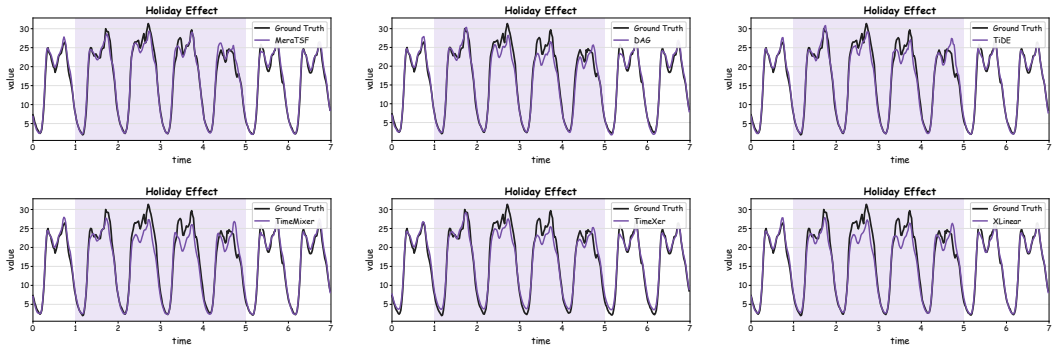


Figure 13: Forecasting visualization on the Ride-Hailing under holiday effects (Sample 2).

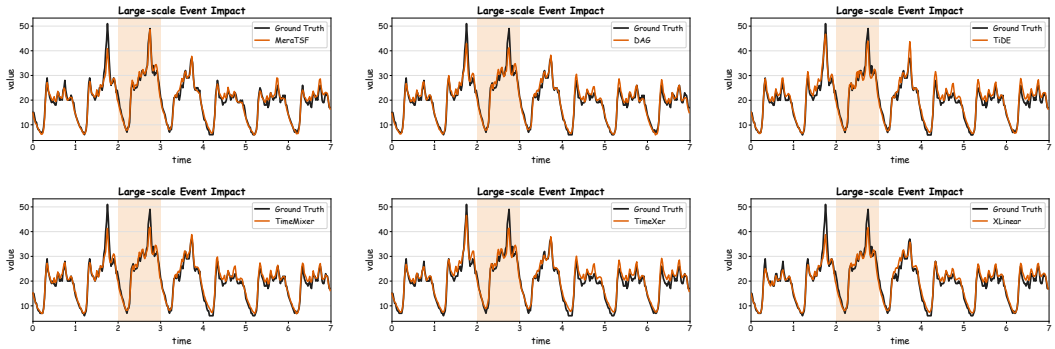


Figure 14: Forecasting visualization on the Ride-Hailing under Large-scale Event (Sample 1).

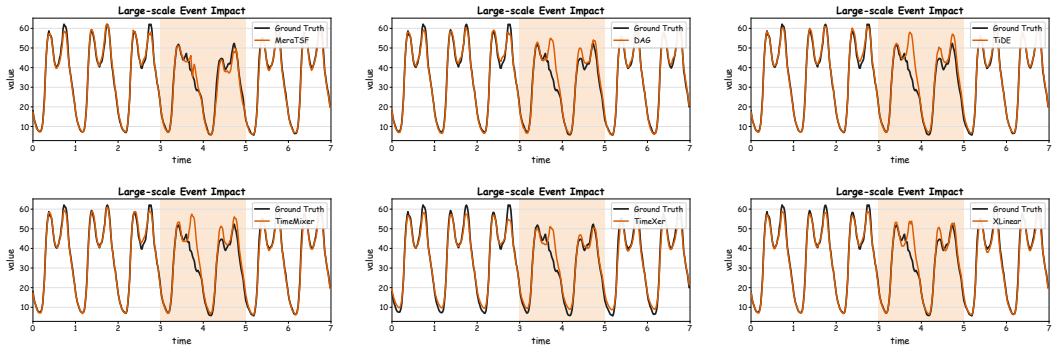


Figure 15: Forecasting visualization on the Ride-Hailing under Large-scale Event (Sample 2).