Appendix

A Supplementary Dataset Analysis

In this section, we conduct a data analysis to verify the distribution dependencies between normal and overload scenarios on G56 and G60.

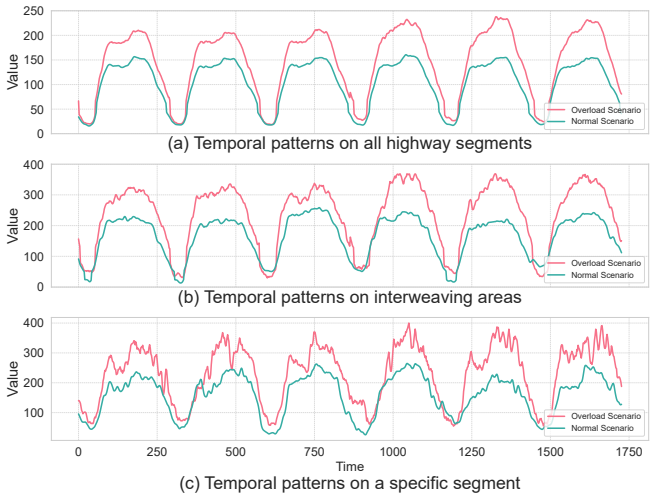

Figure 6: Traffic patterns under normal and overload scenarios on G56.

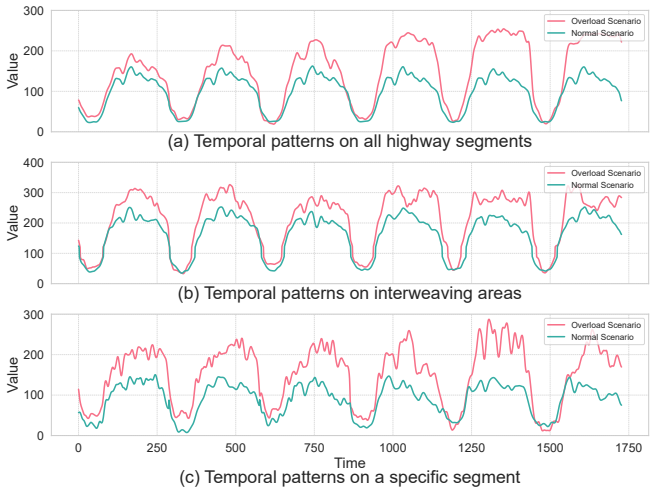

Figure 7: Traffic patterns under normal and overload scenarios on G60.

Figures 6 and 7 illustrate the average temporal patterns across all highway segments, interweaving areas, and a specific segment. We have the following observations: 1) Under overload scenarios, the average temporal patterns exhibit higher traffic peaks and sharper changes compared to normal scenarios (Figure 6 and 7 (a)), driven by exceptionally high demand. 2) In interweaving areas, frequent traffic interactions during overload scenarios lead to temporal patterns that differ significantly from those observed during normal periods (Figure 6 and 7 (b)). 3) On a specific highway segment, the temporal pattern under overload scenarios shows irregular variations with no clear periodic behavior (Figure 6 and 7 (c)).

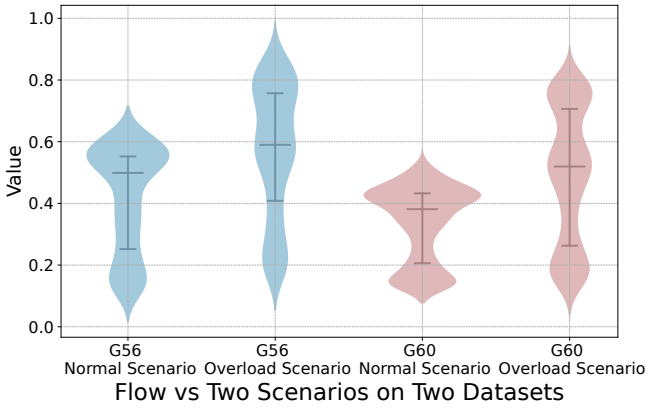

Figure 8: Comparative analysis of traffic flow on G56 and G60 datasets under Two Scenarios.

The violin Figure 8 shows the normalized traffic volume statistics on both datasets during the normal and overload scenarios. The traffic flow range under overload scenarios (the second violin for each dataset) is broader compared to normal scenarios, indicating increased variability in traffic flow (higher values and greater fluctuations). Additionally, in both datasets, the median (central line) and inter-quartile range are more pronounced under overload scenarios, which also confirms the complex traffic changes.

The above data analysis demonstrates the complex spatial-temporal patterns present in overload scenarios. Moreover, Table 1 reveals that traffic data under overload scenarios accounts for only about 5% of the total, highlighting the data sparsity issues.

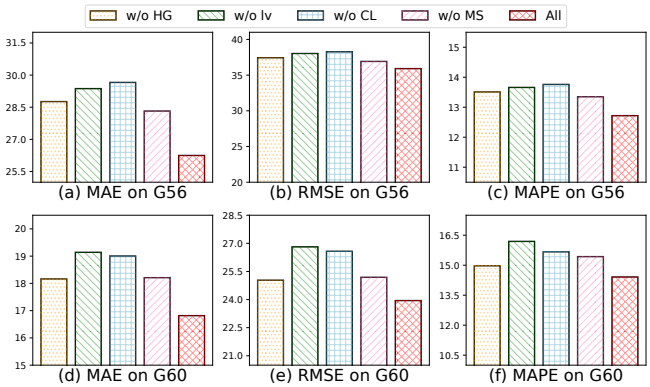

Figure 9: Effectiveness of each component on G56 and G60 for 20-minute-ahead prediction.

B Supplementary Experiments

B.1 Supplementary Ablation Studies

In this section, we provide additional ablation experiments with prediction ahead of 40 minutes and 60 minutes. The

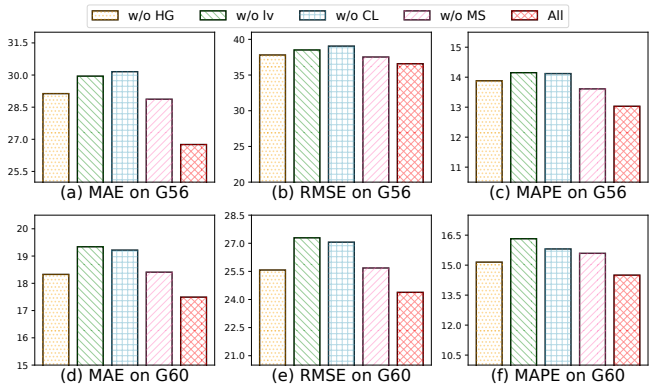

Figure 10: Effectiveness of each component on G56 and G60 for 40-minute-ahead prediction.

results are shown in Figures 9 and 10, which are consistent with those presented in Section 4.3 of the main paper.

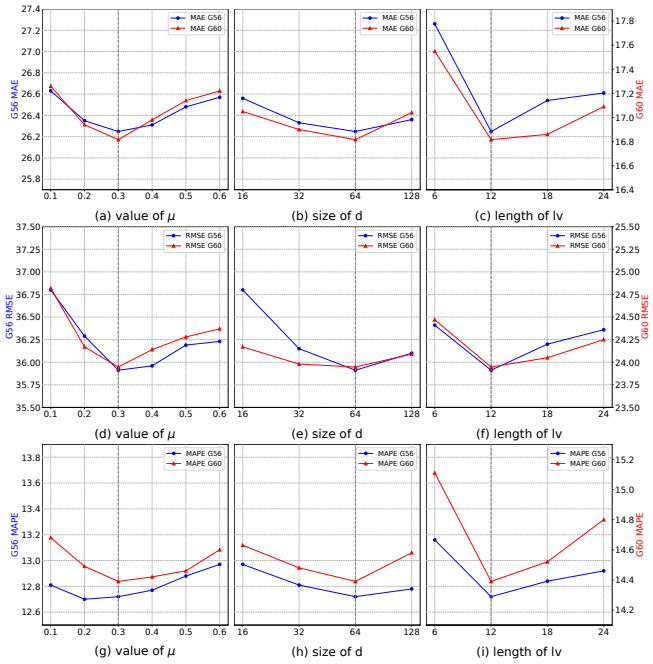

Figure 11: Hyper-parameter analysis on metrics MAE, RMSE, and MAPE on the G56 and G60 datasets for 20-minute-ahead prediction.

B.2 Supplementary Hyper-parameter Analysis

In this section, we provide supplementary hyper-parameter analysis experiments for all metrics with prediction ahead of 40 and 60 minutes. The results, as shown in Figures 11, 12, and 13, are consistent with the observations in Section 4.5 of the main paper.

B.3 Comparison Results on PeMS04

We simulate an overload scenario on the public dataset PeMS04 to further verify the effectiveness of HST-WAVE. For comparison, we select two best-performing benchmarks,

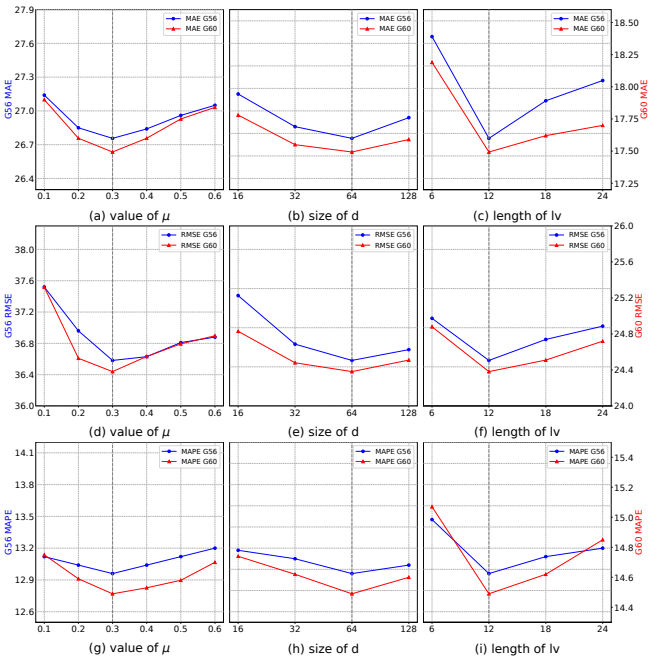

Figure 12: Hyper-parameter analysis on metrics MAE, RMSE, and MAPE on the G56 and G60 datasets for 40-minute-ahead prediction.

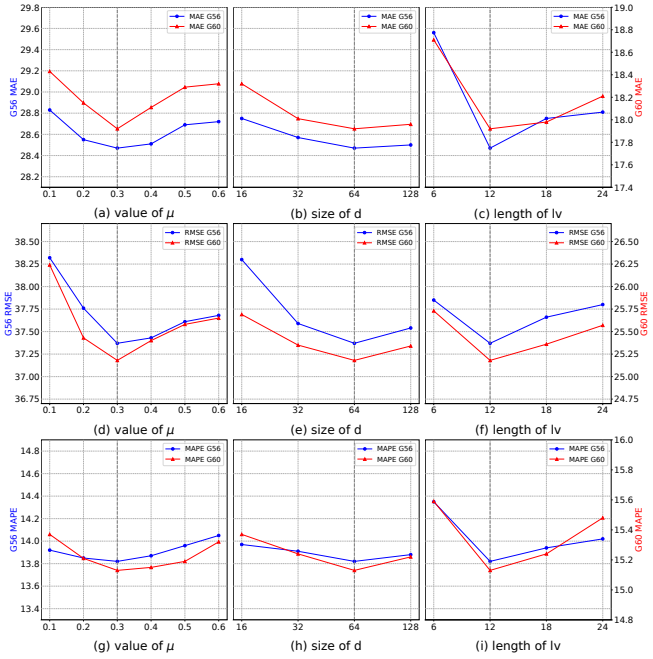

Figure 13: Hyper-parameter analysis on metrics MAE, RMSE, and MAPE on the G56 and G60 datasets for 60-minute-ahead prediction.

Data	Methods	20 min			40 min			60 min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
PEMS04	PDFormer *(AAAI'23)*	26.029	38.297	0.109	28.231	41.381	0.116	30.471	44.436	0.119
	STPGNN *(AAAI'24)*	28.223	40.913	0.109	30.024	43.410	0.114	31.973	45.898	0.121
	HST-WAVE (ours)	**24.892**	**37.164**	**0.092**	**25.514**	**38.183**	**0.095**	**25.986**	**38.781**	**0.097**

Table 3: Performance comparison evaluated by MAE, MAPE, RMSE (lower is better) on PEMS04 datasets. The best performances are highlighted in bold and underlined, respectively.

STPGNN [Kong *et al.*, 2024] and PDFormer [Jiang *et al.*, 2023]. PeMS04 is a homogeneous traffic network with 300 nodes and 400 edges collected from the Caltrans Performance Measurement System (PeMS). The dataset has a sampling frequency of 5 minutes and covers the period from January 1 to February 28, 2028.

For the entire dataset, we define the overload scenario data by truncating the temporal pattern from 12 noon to 6 pm each day, as this period includes the mid- and evening peaks with significantly high traffic flow. We then train the baseline models on datasets from other time periods to directly predict the temporal patterns during the truncated time window. In our HST-WAVE, we replace CHGAN with GCN to better suit the homogeneous network structure of PeMS04. The comparison results are presented in Table 3.

HST-WAVE achieves superior performance on PeMS04 for all metrics, demonstrating that our method is also effective in dealing with peak flow prediction on pure highways. This experimental result is consistent with the G56 and G60 datasets (as shown in Table 2).

B.4 Case study

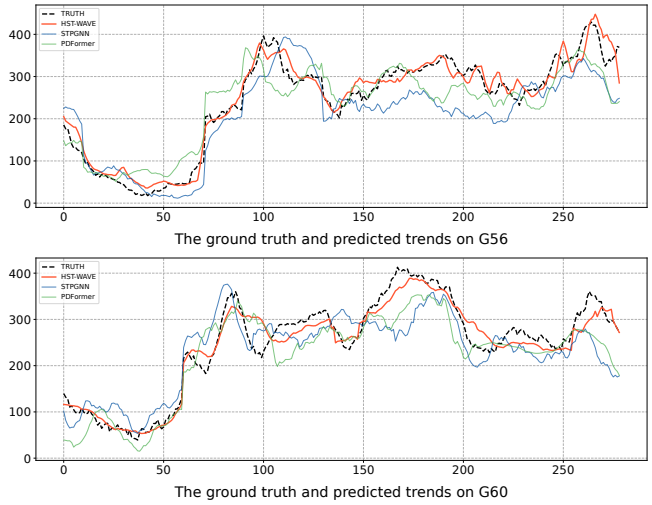

Figure 14: Prediction trends comparison between different methods for two interweaving areas on G56 and G60 datasets.

We plot the ground truth and predicted trends within an interweaving area on G56 and G60 over a specific day in Figure 14. As shown in this figure, our method accurately fits the ground truth more than STPGNN and PDFormer. In high traffic volume and intense fluctuation patterns, the predicted trends of STPGNN and PDFormer struggle to match with the ground truth. This discrepancy may stem from two factors: 1) their limited ability to capture the complex traffic behaviors and irregular temporal patterns inherent in overload scenarios and 2) their inability to address distribution shifts caused by data sparsity. In contrast, HST-WAVE exhibits superior prediction performance, benefiting from its ability to capture complex heterogeneous spatial-temporal information and its designed adaptive temporal enhancement strategy. The visualization experiment provides an intuitive validation of HST-WAVE's effectiveness in addressing the HIPO problem.

C Symbols

Symbol	Description
$\mathcal{V}$	the highway network.
$\mathcal{S}$	the highway segment set.
$\mathcal{P}$	the local segment set.
$\mathcal{S}^{\mathcal{P}}$	the interweaving area set.
$X_{t-T:t}$	the traffic flow on $\mathcal{V}$ with T steps.
$\mathcal{G}$	the heterogeneous traffic graph.
$\mathbf{h}^v, \mathbf{h}$	the traffic feature representation on a node v.
$\mathbf{H}$	the traffic feature representations on all nodes.
d	the dimension of node features.
w_i	the i-th kernel scale.
$\mathbf{h}_{w_i}$	the i-th scale kernel's output in MSWT.
$\hat{\mathbf{h}}_{w_i}$	the i-th scale Transformer's output in MSWT.
$\hat{\mathbf{h}}, \hat{\mathbf{H}}$	the temporal output of MSWT.
u, v	the target node and source node in $\mathcal{V}$.
$\hat{h}_{\phi(e_i)}$	the i-th edge's type feature.
$\beta_{(u,v)}$	the edge bias from u to v.
$\mathbf{h}^*$	the spatial output of CHGAN for one node.
$\mathcal{A}$	the augmentation operation set.
lv	the adaptive temporal pattern.

Table 4: The descriptions of symbols in this paper.