



Supplementary materials for

Yusong ZHOU, Xiaoyu JIANG, Shu SUN, Xinmin ZHANG, Yuanqiu MO, Zhihuan SONG, 2025. MltAuxTSPP: a unified benchmark for deep learning-based traffic state prediction with multi-source auxiliary data. *Front Inform Technol Electron Eng*, 26(10):1984-1999.

<https://doi.org/10.1631/FITEE.2500169>

1 Related works

Table S1 Multi-source data fusion in traffic forecasting

Model	Auxiliary information	Extraction method	Data type	Spatial format	Fusion method	Fusion stage
DST-GCNN (Chen et al., 2020)	Time, Date, Weather	One-hot, FC	Categories	-	Concatenation	Input
ST-ResNet (Zhang et al., 2017)	Date, Holidays, Weather	One-hot, FC	Categories	-	Addition	Output
ResLSTM (Zhang et al., 2021)	Temperature, Wind Speed	FC	Continuous	-		
	Weather, Air Quality	FC, LSTM	Continuous	-	Learnable Weighted Add	Double
	Query Impact	LSTM	Categories	Grid		
Seq2Seq (Liao et al., 2018)	Road Network	GCN	Continuous	Graph	Concatenation	Double
	Date, Holidays, Peak hours	-	Categories	-		
LSTM (Lee et al., 2020)	Road Network	GE	Continuous	Graph	Concatenation	Input
	Peak hour	-	Categories	-		
DxNat (Sun et al., 2017)	Time, Date	One-hot, FC	Categories	-	Concatenation	Output
STDN (Yao et al., 2019)	Weather	-	Categories	-	Concatenation	Input
TCP-D (Huang et al., 2019)	Weather, Peak hours	One-hot	Categories	-	Concatenation	Input
ACFM (Liu et al., 2018)	Date, Holidays, Weather	One-hot, FC	Categories	-	Learnable Weighted Add	Output
	Temperature, Wind Speed	FC	Continuous	-		
TRECK (Zheng et al., 2024)	Event stamp	Representation Learner	Categoricies	-	Concatenation	Double
MT-STNet (Zou et al., 2024)	Time, Date, Station	One-hot	Categories	-	Concatenation	Input

One-hot: One-hot Encoding. FC: Fully Connected Layer. GE: Graph Embedding Module.

While spatiotemporal models excel at learning from traffic data—commonly sourced from sensor networks like PeMS¹ (Yu et al., 2018; Song et al., 2020; Guo et al., 2019)—their performance is inherently limited by overlooking external factors. Consequently, although many studies now incorporate auxiliary data such as weather and time features (see Table S1), their fusion methodologies exhibit critical bottlenecks. Specifically, current methods often rely on simplistic techniques like concatenation, failing to capture complex interdependencies. They also tend to treat auxiliary data as global variables, thereby ignoring crucial spatial heterogeneity (e.g., localized weather effects). Furthermore, these approaches typically lack the scalability and generalizability required to handle diverse or novel auxiliary data streams. These challenges underscore that effective multi-source data utilization requires more than just data availability; advanced, scalable, and spatially-aware fusion strategies are paramount.

2 Variables in the experiments

Our study leverages two standard traffic benchmark datasets, METR-LA and PEMS-BAY (Li et al., 2018), augmented with external data to investigate their impact. The auxiliary data includes three global weather datasets (ERA5-Land², ERA5-HEAT³ and GSH⁴) and three discrete temporal features: minute-of-hour (MoH), hour-of-day (HoD), and day-of-week (DoW). To unify the spatio-temporal heterogeneity of

¹<https://pems.dot.ca.gov/>

²<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-land?tab=overview>

³<https://cds.climate.copernicus.eu/datasets/derived-utci-historical?tab=overview>

⁴<https://www.ncei.noaa.gov/products/land-based-station/integrated-surface-database#tab-336>

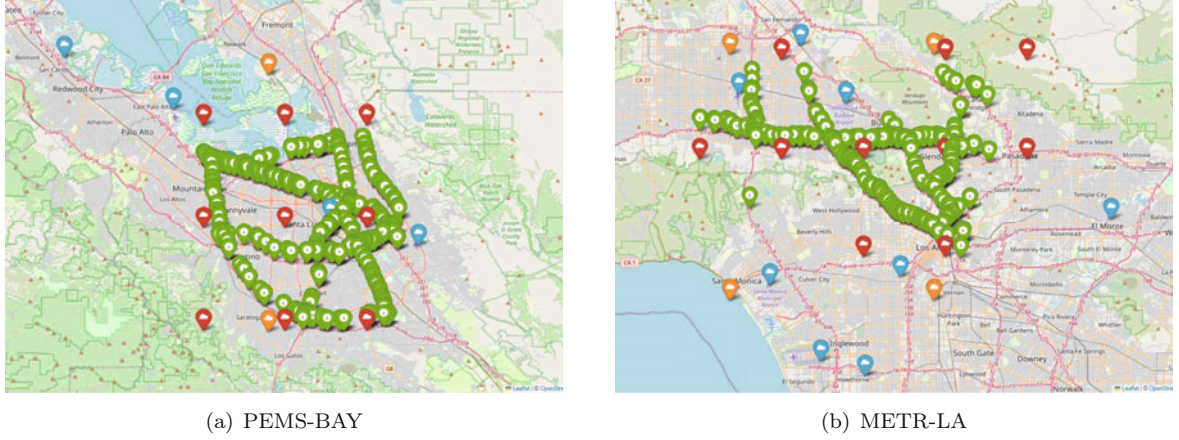


Fig. S1 Spatial distribution of data sampling points for traffic and weather datasets

Table S2 Summary of datasets

Source	Interval	METR-LA			PEMS-BAY		
		steps	nodes	variables	steps	nodes	variables
Traffic	5 min		207	2		325	2
ERA5-Land		34272	10	10	52116	9	10
ERA5-HEAT			4	2		2	2
GSH			8	1		5	1

these sources, we developed a preprocessing pipeline. First, all weather data are temporally interpolated to a 5-minute frequency to match the traffic data. Second, spatially, all datasets are converted to point-based representations, with weather data aligned to the nearest traffic sensor. Spatial distribution of data sampling points for traffic and weather datasets is illustrated in Figure S1. Green markers are traffic data collection points, red markers are ERA5-Land data points, orange markers are ERA5-HEAT data points, and blue markers are GSH data points. Finally, we got two enriched datasets, with detailed variable descriptions and statistics provided in Table S2 and Table S3.

In the experiment, we used the PEMS-BAY and METR-LA dataset. S4 lists the variables from these datasets along with their sources and the corresponding Universal Data Container (UDC) components. The normalization was performed using the Z-score method, as shown below:

$$Z(\mathbf{x}, \bar{\mathbf{x}}, \sigma) = \begin{cases} (\mathbf{x} - \bar{\mathbf{x}}) / \sigma & \text{if } \sigma \neq 0 \\ 0 & \text{if } \sigma = 0 \end{cases} \quad (\text{S1})$$

Here, $\mathbf{x}$ is the original data, $\bar{\mathbf{x}}$ is the mean, and σ is the standard deviation. Table S5 and Table S7 provide details on the normalization process for traffic and meteorological variables in PEMS-BAY. N_n is the number of nodes, N_f is the number of features. The adjacency matrix in Table S5 only needs to be normalized when used as a covariate, not when used as a convolution kernel for T-GCN.

3 Hyperparameter tuning

In this section, we provide additional details regarding the hyperparameter tuning of model parameters. A summary of the hyperparameters for the Seq2Seq, Informer, and T-GCN models is presented in Table S9. Comma-separated values indicate explicit choices. The optimal hyperparameter settings for each model are also listed in the last column of the table.

Table S3 Summary of variables

Source	Var.	Description
Traffic	ts	traffic speed
	adj	adjacency matrix (0 or 1)
ERA5-Land	d2m	2m dewpoint temperature (K)
	rsn	mass of snow per cubic metre in the snow layer (kg m^{-3})
	sf	accumulated total snow that has fallen to the Earth's surface (m)
	skt	skin temperature (K)
	smlt	melting of snow averaged over the grid box (kg m^{-3})
	sp	surface pressure (Pa)
	t2m	2m temperature (K)
	tp	total precipitation (m)
	u10	10m u-component of wind (m s^{-1})
	v10	10m v-component of wind (m s^{-1})
ERA5-HEAT	mrt	mean radiant temperature (K)
	utci	universal thermal climate index (K)
GSH	vd	visibility distance (m)
Time	MoH	minute of the hour ($5i, i \in \{0, 1, \dots, 11\}$)
	Hod	hour of the day (0 - 23)
	dow	day of the week (0 - 6)

Table S4 Variables and sources for the unified data container components

Component	Source	Variable
t_{tgt}	PEMS-BAY	ts
s_{cont}	PEMS-BAY	adj
$t_{k,cat}$	Time	MoH, Hod, dow
$t_{o,cont}$	ERA5-Land	d2m, rsn, sf, skt, smlt, sp, t2m, tp, u10, v10
$t_{o,cont}$	ERA5-HEAT	mrt, utci
$t_{o,cont}$	GSH	vd

Table S5 Normalization for traffic variables in PEMS-BAY

Nodes	N_n	ts			adj		
		N_f	$\bar{x}$	σ	N_f	$\bar{x}$	σ
1-1	1	1	67.569	8.119	1	1.000	0.000
4-1	4	1	61.966	11.178	4	0.25	0.433
16-1	16	1	62.179	11.031	16	0.137	0.344
16-2	16	1	63.235	9.272	16	0.160	0.367
16-4	16	1	63.506	10.740	16	0.133	0.339
16-8	16	1	62.098	10.867	16	0.133	0.339
all	325	1	62.620	9.594	325	0.079	0.270

Table S6 Normalization for traffic variables in METR-LA

Nodes	N_n	ts			adj		
		N_f	$\bar{x}$	σ	N_f	$\bar{x}$	σ
1-1	1	1	54.956	16.844	1	1.000	0.000
4-1	4	1	48.108	20.706	4	1.000	0.000
16-1	16	1	49.272	20.860	16	0.672	0.470
16-2	16	1	49.962	20.979	16	0.680	0.467
16-4	16	1	51.825	20.759	16	0.367	0.482
16-8	16	1	52.752	20.937	16	0.297	0.457
all	207	1	53.719	20.261	207	0.274	0.446

Table S7 Normalization for weather covariates in PEMS-BAY

Source	Variable	N_n	N_f	$\bar{x}$	σ
ERA5-Land	d2m	9	1	281.465	3.372
	rsn	9	1	102.185	11.874
	sf	9	1	0.000	0.000
	skt	9	1	287.721	9.192
	smlt	9	1	0.000	0.000
	sp	9	1	99858.414	1312.915
	t2m	9	1	286.410	5.519
	tp	9	1	0.002	0.006
	u10	9	1	0.844	1.313
	v10	9	1	0.178	1.674
ERA5-HEAT	mrt	2	1	291.711	17.040
	utci	2	1	284.753	8.639
GSH	vd	5	1	15516.044	2206.688

Table S8 Normalization for weather covariates in METR-LA

Source	Variable	N_n	N_f	$\bar{x}$	σ
ERA5-Land	d2m	10	1	280.302	5.666
	rsn	10	1	107.554	18.230
	sf	10	1	0.000	0.000
	skt	10	1	292.412	10.646
	smlt	10	1	0.000	0.000
	sp	10	1	97286.622	2258.767
	t2m	10	1	289.513	6.251
	tp	10	1	0.000	0.002
	u10	10	1	0.287	1.206
	v10	10	1	0.581	1.464
ERA5-HEAT	mrt	4	1	297.150	19.890
	utci	4	1	289.505	8.879
GSH	v	8	1	16684.955	7082.730

Table S9 Hyperparameter tuning for models considered in this work

Model	Parameter	Space	Optimal
Seq2Seq	d_model	16, 48, 128, 256	128
	dropout	0.0, 0.1, 0.3	0.1
	num_layers	1, 2, 3	2
	bidirectional	True,False	True
	rnn_type	LSTM, GRU, RNN	GRU
Informer	d_ff	128,256,512,1024	512
	d_layers	1, 2, 3	1
	d_model	32, 128, 256,512	128
	dropout	0.0, 0.1, 0.3	0.1
	e_layers	1,2,3,4	3
	factor	2,3,5	5
	n_heads	4,6,8	8
T-GCN	d_model	16, 32, 48	48
	dropout	0.0, 0.1, 0.3	0.1

4 Complete experimental results

We evaluated the models’ performance across various conditions, conducting 10 experiments per condition, each with a unique random seed. All experiments used a batch size of 512, a learning rate of 0.001, and ran for 15 epochs with early stopping. Some experiments adjusted the batch size and the use of AMP to ensure successful execution. The conditions for the experiments are summarized in Table S10.

Table S10 Experimental conditions for model evaluation

Condition	Settings
Number of Nodes (N_s)	1-1, 4-1, 16-x, all($n - i$ denotes n nodes with index intervals i , and $x \in \{1, 2, 4, 8\}$)
Time Windows (L_w)	$L_w \in \{12, 24, 48, 96\}$ (input and output windows are identical)
Auxiliary Variables (Co.)	Traffic-only (<i>covariates</i> = <i>No</i>); All auxiliary variables (<i>covariates</i> = <i>All</i>)
Embedding Modules (E.M.)	FlattenCat, TFT, TFT-GWS
Model Perception of Future Window	With or without future window information
Model After Fusion Embedding (M.)	Seq2Seq, Informer, T-GCN

The following metrics are used to evaluate each model:

- Mean Absolute Error (MAE):

$$MAE(\mathbf{X}^*, \hat{\mathbf{X}}^*) = \frac{\sum \sum \sum |X_{i,j,k}^* - \hat{X}_{i,j,k}^*|}{h * N * F}$$

- Mean Squared Error (MSE):

$$MSE(\mathbf{X}^*, \hat{\mathbf{X}}^*) = \frac{\sum \sum \sum \left(X_{i,j,k}^* - \hat{X}_{i,j,k}^* \right)^2}{h * N * F}$$

Here, $\mathbf{X}^*, \hat{\mathbf{X}}^* \in \mathbb{R}^{h \times N \times F}$ are the predicted and ground truth data, h is the sequence length, N is the number of nodes, and F is the feature count.

4.1 Basic performance of forecasting

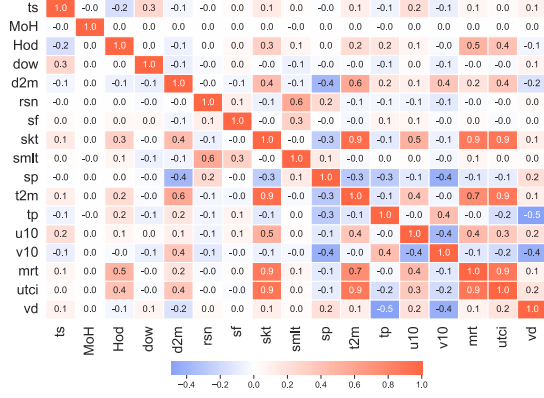
The baseline performance of models on the PEMS-BAY and METR-LR datasets, using only traffic data, is presented in Table S11.

Table S11 Forecast performance without covariates or future information

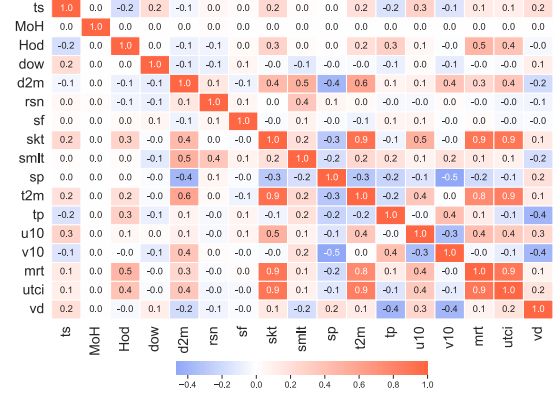
Dataset	Model	E.M. ^a L_w^c N_d^d	FlattenCat								TFT								TFT-GWS							
			12		24		48		96		12		24		48		96		12		24		48		96	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS-BAY	SegSeq	1-1	0.2967	0.2702	0.3750	0.3148	0.5221	0.3949	0.6479	0.4398	0.3065	0.2673	0.3663	0.3105	0.5422	0.4524	0.6146	0.4245	0.2910	0.2629	0.3645	0.3078	0.5069	0.3850	0.5767	0.3964
		4-1	0.4565	0.2476	0.3042	0.2883	0.3725	0.3526	0.3789	0.3452	0.4588	0.2434	0.2970	0.2826	0.3421	0.3210	0.3486	0.3242	0.4527	0.2375	0.2997	0.2808	0.3506	0.3303	0.3599	0.3344
		16-1	0.2864	0.2681	0.4540	0.3360	0.5574	0.3812	0.5541	0.3880	0.3067	0.2766	0.4266	0.3234	0.5638	0.3879	0.5541	0.3851	0.3096	0.2772	0.4507	0.3295	0.5516	0.3833	0.5266	0.3687
		16-2	0.3034	0.2584	0.4347	0.3110	0.5991	0.3692	0.6011	0.3651	0.3333	0.2663	0.4551	0.3154	0.6272	0.3879	0.5886	0.3511	0.3119	0.2630	0.4497	0.3141	0.5928	0.3640	0.5960	0.3602
	Informer	16-4	0.1875	0.2050	0.2624	0.2421	0.3250	0.2810	0.3782	0.3083	0.1964	0.2065	0.2645	0.2430	0.3378	0.2882	0.3873	0.3132	0.1974	0.2094	0.2602	0.2418	0.3217	0.2778	0.3698	0.3055
		16-8	0.2974	0.2341	0.3145	0.2882	0.3703	0.3209	0.3921	0.3294	0.2145	0.2394	0.3188	0.2878	0.3731	0.3220	0.3833	0.3339	0.2143	0.2393	0.3431	0.3017	0.3683	0.3199	0.3889	0.3285
		all	0.2841	0.2762	0.3178	0.2905	0.3810	0.3281	0.3867	0.3271	0.2890	0.2780	0.3194	0.2917	0.3817	0.3296	0.3803	0.3224	0.2916	0.2767	0.3216	0.2912	0.3814	0.3392	0.3857	0.3247
		1-1	0.3140	0.2770	0.4252	0.3352	0.5391	0.4249	0.6496	0.4424	0.3267	0.2879	0.4361	0.3421	0.5885	0.4160	0.6416	0.4393	0.3129	0.2775	0.4245	0.3369	0.5708	0.4161	0.6294	0.4319
		4-1	0.2400	0.2663	0.3336	0.3313	0.3946	0.3601	0.3511	0.3414	0.2562	0.2799	0.3401	0.3348	0.3774	0.3584	0.3682	0.3513	0.2459	0.2714	0.3437	0.3364	0.3841	0.3676	0.3744	0.3589
		16-1	0.3215	0.2936	0.4859	0.3569	0.5775	0.3990	0.5835	0.4530	0.3509	0.3066	0.4943	0.3543	0.5993	0.4580	0.5940	0.4574	0.4576	0.3147	0.5071	0.3612	0.5663	0.3951	0.5714	0.3950
T-GCN	16-2	0.3424	0.2798	0.4086	0.3361	0.6611	0.3948	0.6487	0.3881	0.3439	0.2844	0.5187	0.3777	0.6762	0.3933	0.6655	0.3949	0.3850	0.2926	0.5402	0.3418	0.6448	0.3864	0.6338	0.3798	
	16-4	0.2136	0.4534	0.2885	0.2605	0.3921	0.3184	0.4573	0.3359	0.4505	0.4575	0.2880	0.2587	0.4117	0.3247	0.4179	0.3412	0.4542	0.2321	0.2902	0.2618	0.3488	0.2950	0.4520	0.3283	
	16-8	0.2412	0.2600	0.3558	0.3136	0.4210	0.3506	0.4304	0.3527	0.2461	0.2604	0.3682	0.3203	0.4658	0.3678	0.4903	0.3777	0.2641	0.2741	0.3644	0.3195	0.4261	0.3452	0.4193	0.3475	
	all	0.3061	0.2908	0.3481	0.3118	0.3901	0.3342	0.3847	0.3300	0.3106	0.2914	0.3386	0.3043	0.4220	0.3494	0.3841	0.3318	0.3175	0.2964	0.3464	0.3120	0.3912	0.3357	0.3814	0.3291	
METR-LA	GWNet	1-1	0.3035	0.2777	0.3730	0.2959	0.5411	0.3951	0.7115	0.4678	0.2892	0.2667	0.3531	0.2927	0.5205	0.3748	0.6151	0.3995	0.2839	0.2586	0.3554	0.2946	0.5209	0.3731	0.6579	0.4447
		4-1	0.2698	0.2610	0.4586	0.3512	0.5736	0.4407	0.7665	0.5334	0.2637	0.2605	0.3855	0.3383	0.3356	0.4568	0.6542	0.4764	0.2678	0.2672	0.3918	0.3429	0.5567	0.4225	0.7227	0.5241
		16-1	0.2880	0.2677	0.4683	0.3677	0.7011	0.4772	0.9316	0.5682	0.2614	0.2534	0.4213	0.3456	0.6394	0.4411	0.8651	0.5424	0.2636	0.2512	0.4308	0.3482	0.6521	0.4644	0.8797	0.5495
		16-2	0.2948	0.2634	0.4678	0.3545	0.7075	0.4494	0.9393	0.5161	0.2849	0.2530	0.4746	0.3465	0.7150	0.4343	0.9173	0.4996	0.2832	0.2540	0.4818	0.3581	0.7069	0.4490	0.9194	0.5048
	Informer	16-4	0.2454	0.4565	0.4509	0.3071	0.6237	0.4503	0.8116	0.4885	0.4542	0.2153	0.3771	0.2894	0.5850	0.3710	0.7847	0.4747	0.2313	0.2159	0.3867	0.2943	0.5994	0.3922	0.7958	0.4853
		16-8	0.2744	0.2656	0.4598	0.3652	0.7143	0.4910	0.8856	0.5633	0.2436	0.2382	0.3947	0.3215	0.6205	0.4346	0.7870	0.5004	0.2488	0.2410	0.4114	0.3326	0.6648	0.4587	0.8324	0.5324
		all	0.3002	0.2633	0.4449	0.3372	0.5529	0.4501	0.6642	0.4382	0.3278	0.2889	0.4142	0.3178	0.5237	0.3938	0.6281	0.4163	0.3274	0.3075	0.4355	0.3341	0.5874	0.4215	0.6572	0.4244
		1-1	0.4520	0.2310	0.3203	0.2893	0.3907	0.3228	0.4184	0.3637	0.4562	0.2306	0.3228	0.2863	0.3749	0.3184	0.4418	0.3782	0.2352	0.2545	0.3315	0.3018	0.4247	0.3375	0.4377	0.3705
		16-1	0.4599	0.4583	0.3449	0.2892	0.4979	0.3468	0.5751	0.4511	0.2386	0.2331	0.3379	0.2897	0.5202	0.3614	0.5566	0.3867	0.2493	0.2564	0.3869	0.3133	0.5337	0.3673	0.3710	0.3948
		16-2	0.2472	0.4565	0.4504	0.3007	0.5775	0.3659	0.6215	0.3637	0.2614	0.2342	0.3275	0.2796	0.5641	0.3539	0.6373	0.3739	0.2676	0.2548	0.3955	0.3045	0.5627	0.3650	0.4477	0.3923
16-4	0.1794	0.1897	0.2383	0.4542	0.3592	0.2845	0.4367	0.3344	0.1933	0.1943	0.2422	0.4522	0.4740	0.2920	0.4395	0.3277	0.1998	0.2134	0.2604	0.2391	0.3676	0.2919	0.4366	0.3292		
16-8	0.1526	0.1925	0.2516	0.4361	0.3267	0.3897	0.3321	0.3897	0.1955	0.2112	0.2679	0.2528	0.4678	0.3382	0.3886	0.2263	0.2097	0.2340	0.2871	0.2689	0.4579	0.3372	0.3757	0.3224		
all	0.3881	0.3525	0.5782	0.4287	0.7490	0.5257	1.0124	0.6175	0.3663	0.3216	0.5355	0.3595	0.7282	0.5966	0.9459	0.5651	0.3755	0.3319	0.5572	0.4126	0.7497	0.4797	0.9192	0.9680	0.5700	
T-GCN	SegSeq	1-1	0.2602	0.2961	0.3976	0.3086	0.5142	0.4721	0.6913	0.5363	0.2607	0.2947	0.3857	0.3727	0.5015	0.4509	0.7112	0.5288	0.2550	0.2877	0.3779	0.3758	0.4895	0.4513	0.7419	0.5483
		16-1	0.3761	0.3945	0.5250	0.4646	0.6711	0.5519	0.9382	0.6370	0.3784	0.3941	0.5418	0.4743	0.7173	0.5686	1.0131	0.6516	0.3755	0.3927	0.5361	0.4627	0.7166	0.5575	0.9797	0.6422
		16-2	0.3555	0.3421	0.5311	0.4353	0.7000	0.5232	0.9132	0.6069	0.3519	0.3387	0.5433	0.4377	0.7047	0.5205	0.8658	0.5893	0.3569	0.3421	0.5524	0.4432	0.7355	0.5239	0.9582	0.6091
		16-4	0.4584	0.3436	0.5768	0.4303	0.7305	0.5197	1.0103	0.6356	0.4204	0.3475	0.5996	0.4410	0.7852	0.5462	1.0944	0.6739	0.4168	0.3469	0.5998	0.4408	0.7758	0.5383	0.9519	0.6304
	16-8	0.3346	0.3189	0.4866	0.4585	0.6674	0.5164	0.8848	0.6050	0.3351	0.3224	0.4781	0.3984	0.6950	0.5197	0.9548	0.6200	0.3380	0.3215	0.4958	0.4125	0.7006	0.5193	0.9162	0.6082	
	all	0.4801	0.3755	0.6294	0.4458	0.8284	0.5386	0.9998	0.6091	0.4848	0.3840	0.4944	0.3215	0.7900	0.5234	0.9684	0.5996	0.4847	0.3803	0.6342	0.4571	0.8149	0.5310	0.9364	0.6003	
	Informer	1-1	0.4507	0.3489	0.6010	0.4434	0.7934	0.5342	1.0929	0.6138	0.4110	0.3557	0.6035	0.4415	0.7761	0.5292	1.1473	0.6382	0.3923	0.3459	0.5693	0.4268	0.7613	0.5206	1.0432	0.5945
		4-1	0.2709	0.3074	0.4573	0.3919	0.5275	0.4683	0.7345	0.5412	0.2737	0.3023	0.4110	0.3998	0.5663	0.4838	0.7598	0.5550	0.2688	0.3092	0.3968	0.3886	0.5229	0.4665	0.7243	0.5375
		16-1	0.3875	0.4556	0.5463	0.4995	0.6819	0.5856	0.8656	0.6344	0.4180	0.4232	0.5905	0.5015	0.7022	0.5910	0.9130	0.6556	0.4393	0.4179	0.5332	0.5023	0.6906	0.5924	0.8237	0.6255
		16-2	0.3903	0.3629	0.5695	0.4635	0.7373	0.5587	0.8997	0.6391	0.4105	0.3712	0.5790	0.4706	0.7724	0.5789	0.9960	0.6568	0.4531	0.3724	0.5554	0.4565	0.7278	0.5574	0.9293	0.6309
16-4		0.4295	0.3675	0.5978	0.4633	0.7868	0.5671	1.0114	0.6722	0.4300	0.3873	0.6182	0.4821	0.8309	0.6132	1.1031	0.7063	0.4385	0.3736	0.6105	0.4770	0.7805	0.5844	1.015	0.7091	
16-8		0.3565	0.3413	0.5070	0.4317	0.6898	0.5443	0.9042	0.6287	0.3741	0.3476	0.5445	0.4499	0.7273	0.5690	0.9882	0.6603	0.3716	0.3499	0.5161	0.4339	0.7063	0.5586	0.9135	0.6310	
all	0.5076	0.3991	0.6505	0.4764	0.8283	0.5854	1.0957	0.6598	0.5064	0.4574	0.6601	0.4809	0.9085	0.6169	1.1292	0.6660	0.5120	0.4541	0.6691	0.4816	0.8699	0.5817	1.0470	0.6455		
T-GCN	GWNet	1-1	0.3737	0.3365	0.5413	0.4103	0.7396	0.5163	1.0954	0.6693	0.3729	0.3835	0.5592	0.4159	0.7181	0.4931	0.9753	0.6165	0.3785	0.3388	0.5400	0.4520	0.7221	0.5004	1.0344	0.6397
		4-1																								

4.2 Improvement with covariates

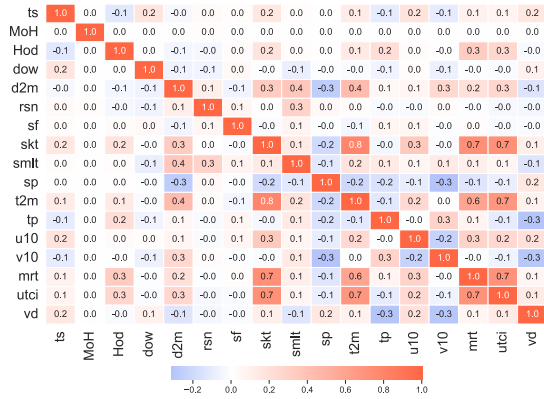
We evaluated the impact of covariates by comparing prediction performance with and without them under identical experimental configurations (subsection 4.1). As shown in Table S12, covariates significantly improve accuracy for long-term forecasts and in low-complexity, single-node settings, highlighting their value for extended prediction horizons.



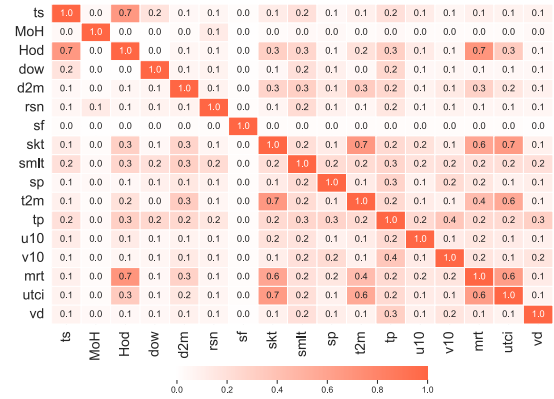
(a) Pearson



(b) Spearman



(c) Kendall



(d) MIC

Fig. S2 Correlation between traffic state and covariates

4.3 Impact of covariates in the future window

We conduct a comparative analysis of prediction tasks that consider future window information with those that do not. The analysis in Table S13 shows that incorporating future window information, particularly with known covariates, significantly improves prediction performance.

4.4 Variable importance

We evaluate the importance of variables using correlation. We calculate four types of correlation: Pearson Correlation, Spearman Correlation, Kendall Correlation, and Maximum Information Coefficient (MIC). As shown in Figure S2, the correlations between traffic speed and covariates in the first three methods are very low, but the MIC between traffic state and Hod is high. Notably, temperature-related

variables (mrt, utci, skt, t2m) show a high MIC with Hod, which is similar to that between traffic state and Hod.

We collect and analyze the variable selection weights generated by fusion embedding modules TFT-GWS and TFT across various tasks. Figure S3 show how the models focus on the target variables and covariates. The selection weights of static variables are not presented, as the experiment involves only a single static variable, and its weight is determined solely within the set of static variables.

As depicted in Figure S4, the PEMS-BAY and METR-LA datasets exhibit distinctly different traffic speed distribution patterns between weekdays and weekends, where speed is the spatial average over all sensor nodes.

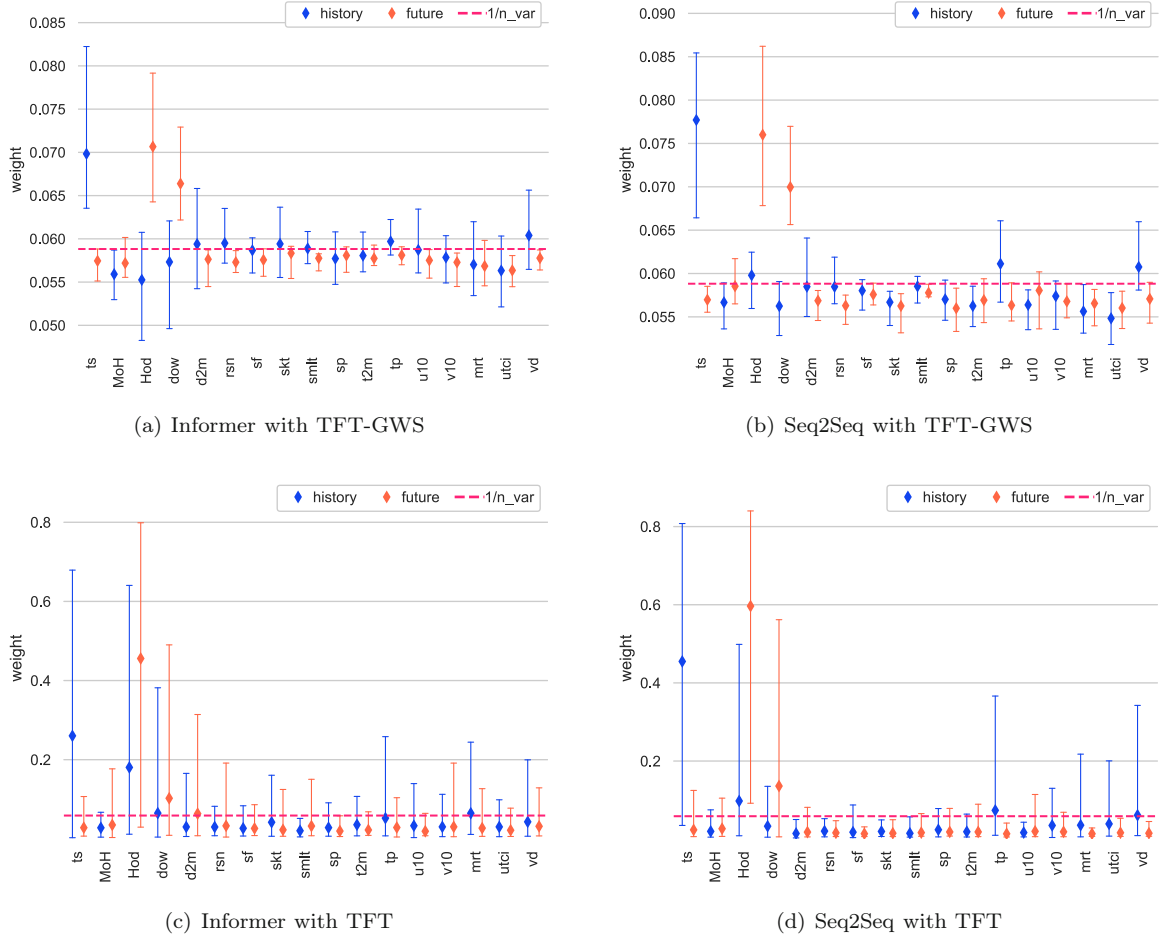


Fig. S3 Variable selection weights

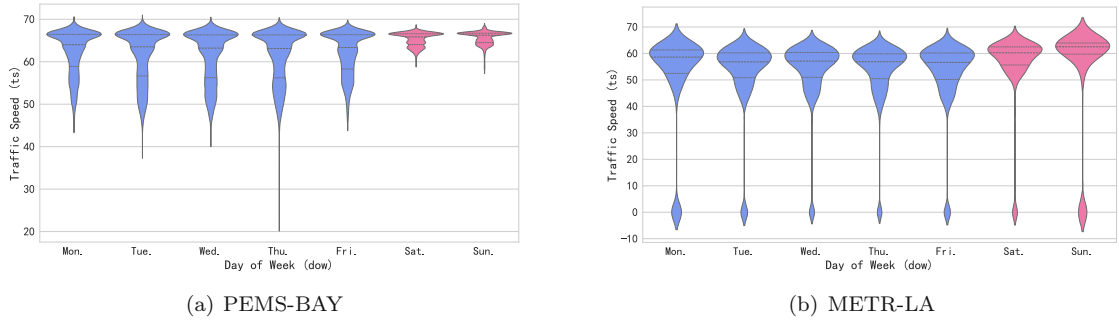


Fig. S4 Average traffic speed distribution by DoW

Table S13 Relative forecast performance gain with future information

Dataset	CoV ^a	E.M. ^c L_w^d Model	FlattenCat								TFT								TFT-GWS							
			12		24		48		96		12		24		48		96		12		24		48		96	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS-BAY	All	Seq2Seq	2.22	1.21	2.79	1.49	3.41	1.81	9.16	6.74	9.19	5.52	12.85	8.18	14.27	11.23	10.29	9.50	4.87	3.09	10.46	6.61	12.93	10.17	13.41	11.30
		Informr	-0.34	-0.17	3.57	1.75	4.50	2.21	-11.24	-4.55	7.95	3.84	4.85	3.74	6.42	5.43	12.00	10.14	0.46	0.21	1.50	1.82	2.53	3.17	10.45	8.56
		T-GCN	4.78	3.51	5.26	3.64	8.93	6.46	19.95	12.49	10.72	6.73	15.04	10.73	20.45	15.94	21.89	16.90	8.76	5.79	15.02	10.58	21.64	16.38	31.37	22.21
		GWNet	5.26	0.07	3.19	-0.70	0.26	-3.09	-9.11	-3.62	-3.44	-1.91	4.91	-0.36	7.41	1.71	15.75	10.76	7.14	4.26	1.56	-1.10	5.67	1.02	11.21	6.31
	No	Seq2Seq	-0.61	-0.18	-0.10	-0.09	0.78	0.18	1.32	1.22	-1.28	-1.87	-0.98	-1.00	3.35	2.20	1.58	2.01	-6.10	-3.16	-2.23	-1.58	0.44	0.30	-2.64	-1.46
		Informr	1.19	0.65	0.17	-0.24	-0.54	-0.62	-0.64	-0.53	3.38	1.43	-2.02	-1.76	-5.73	-4.23	-5.53	-4.37	4.80	1.99	0.52	0.03	-2.16	-1.29	-1.00	-1.32
		T-GCN	0.02	0.07	-0.71	-0.79	-0.07	0.12	0.89	0.14	-0.69	-0.39	1.43	1.01	0.82	-0.57	4.60	1.07	0.02	0.86	0.03	0.66	-0.53	-0.76	4.99	3.03
		GWNet	-4.25	-5.32	-1.61	-1.86	0.45	-0.07	0.53	0.71	2.11	-0.93	1.40	-3.88	-2.56	-1.57	-0.64	-0.74	-0.84	2.83	3.00	0.44	1.54	0.04	-2.64	-3.20
METR-LA	All	Seq2Seq	-1.75	-1.38	0.36	0.12	4.06	2.91	2.83	2.68	0.25	4.03	5.63	5.43	7.33	7.23	7.23	7.47	2.28	1.81	1.72	0.63	2.43	1.94	2.78	3.12
		Informr	-1.61	0.27	-1.51	-0.47	-5.14	-1.48	-3.56	-0.67	-1.34	1.27	1.04	3.79	6.05	5.27	1.56	2.59	4.04	4.61	0.36	1.32	2.61	1.95	4.89	3.25
		T-GCN	7.58	5.79	10.91	6.71	3.67	4.10	6.52	7.79	-0.37	0.72	4.52	2.73	12.38	7.51	6.12	9.61	7.70	6.36	11.06	10.33	13.29	12.37	10.26	12.75
		GWNet	-23.20	-5.74	-25.55	-6.15	-28.44	-8.98	-23.27	-6.32	-4.96	-0.18	-6.37	0.21	-12.77	-2.38	-16.20	-5.40	-10.94	-1.07	-8.54	0.51	-11.51	-3.27	-17.26	-7.81
	No	Seq2Seq	-0.60	-0.72	-0.49	-0.42	-0.78	0.49	-0.98	-0.70	0.39	0.66	0.57	0.69	-0.86	-0.44	-1.37	-0.33	0.06	0.08	-0.57	0.58	-0.14	0.84	-1.10	-0.03
		Informr	0.02	-0.89	1.18	0.67	-1.20	0.09	-3.11	-0.83	2.31	1.33	-0.17	0.13	-1.01	-0.01	-0.16	0.50	1.39	0.55	-0.28	0.35	0.29	1.02	-4.29	-1.47
		T-GCN	-0.13	0.10	-0.05	0.24	0.69	0.89	0.51	0.75	1.74	1.98	2.75	1.47	2.47	1.47	5.25	6.62	1.09	-0.09	2.35	2.09	5.46	4.86	8.26	8.26
		GWNet	-14.98	-6.51	-7.69	-2.09	-14.29	-9.38	-13.81	-6.25	3.33	1.55	2.06	0.54	-19.30	-16.62	-14.84	-8.16	-5.56	0.67	-4.27	-0.59	-13.75	-9.82	-15.73	-7.70

^a The percentage improvement is calculated as $\rho = -(m_{w/o} - m_{w/o}) / m_{w/o} * 100.0\%$, where $m_{w/o}$ and $m_{w/o}$ are metrics with and without future information, respectively. Positive improvements are **bolded**.

^{bcd} These abbreviations are defined in Table S10.

^c The results for the METR-LA dataset are provided in the Supplementary Material.

References

- Chen K, Chen F, Lai B, et al., 2020. Dynamic spatio-temporal graph-based cnns for traffic flow prediction. *IEEE Access*, 8:185136-185145.
<https://doi.org/10.1109/ACCESS.2020.3027375>
- Guo SN, Lin Yf, Feng N, et al., 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proc 39th AAAI Conf on Artificial Intelligence, :922-929.
<https://doi.org/10.1609/aaai.v33i01.3301922>
- Huang Z, Xia J, Li F, et al., 2019. A peak traffic congestion prediction method based on bus driving time. *Entropy*, 21(7):709.
<https://doi.org/10.3390/e21070709>
- Lee C, Kim Y, Jin S, et al., 2020. A visual analytics system for exploring, monitoring, and forecasting road traffic congestion. *IEEE Transactions on Visualization and Computer Graphics*, 26(11):3133-3146.
<https://doi.org/10.1109/TVCG.2019.2922597>
- Li Y, Yu R, Shahabi C, et al., 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting.
<https://openreview.net/forum?id=SJiHxGWAZ>
- Liao B, Zhang J, Wu C, et al., 2018. Deep sequence learning with auxiliary information for traffic prediction. Proc 24th ACM SIGKDD Inter Conf on Knowledge Discovery & Data Mining, p.537-546.
<https://doi.org/10.1145/3219819.3219895>
- Liu L, Zhang R, Peng J, et al., 2018. Attentive crowd flow machines. Proc 26th ACM Inter Conf on Multimedia, p.1553-1561.
<https://doi.org/10.1145/3240508.3240681>
- Song C, Lin Yf, Guo Sn, et al., 2020. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. Proc 39th AAAI Conf on Artificial Intelligence, :914-921.
<https://doi.org/10.1609/aaai.v33i01.5438>
- Sun F, Dubey A, White J, 2017. Dxnat—deep neural networks for explaining non-recurring traffic congestion. 2017 IEEE Inter Conf on Big Data, p.2141-2150.
<https://doi.org/10.1109/BigData.2017.8258162>
- Yao H, Tang X, Wei H, et al., 2019. Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction. Proc AAAI Conf on Artificial Intelligence, 33(01):5668-5675.
<https://doi.org/10.1609/aaai.v33i01.33015668>
- Yu B, Yin Ht, Zhu Zx, 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. Proc 27th Int Joint Conf on Artificial Intelligence, p.3634-3640.
<https://doi.org/10.24963/ijcai.2018/505>
- Zhang J, Chen F, Cui Z, et al., 2021. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Transactions on Intelligent Transportation Systems*, 22(11):7004-7014.
<https://doi.org/10.1109/TITS.2020.3000761>
- Zhang J, Zheng Y, Qi D, 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. Proc AAAI Conf on Artificial Intelligence, 31(1).
<https://doi.org/10.1609/aaai.v31i1.10735>
- Zheng X, Bagloee SA, Sarvi M, 2024. Treck: long-term traffic forecasting with contrastive representation learning. *IEEE Transactions on Intelligent Transportation Systems*, 25(11):16964-16977.
<https://doi.org/10.1109/TITS.2024.3421328>
- Zou G, Lai Z, Wang T, et al., 2024. Mt-stnet: a novel multi-task spatiotemporal network for highway traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 25(7):8221-8236.
<https://doi.org/10.1109/TITS.2024.3411638>