

## A APPENDIX

This appendix presents extended experimental results and methodological details to complement the main body of the paper.

### A.1 DISCUSSION ON DIFFERENCES WITH OTHER GRAPH GENERATION METHODS

In existing spatiotemporal prediction models, graph generation methods can be divided into two categories: (1) Prior Graphs: These are graph structures known to the model in advance, typically static and representing physical connections between nodes. (2) Dynamic Graphs: These are graph structures learned by the model based on input spatiotemporal data. Existing methods for generating dynamic graphs can be further divided into two types: Learnable Parameters: In models like GWNet (Wu et al., 2019) and PGCN (Shin & Yoon, 2024), learnable parameters (denoted as  $\mathbf{A} \in \mathbb{R}^{N \times N}$ ) are set as the adjacency matrix and updated through backpropagation. However, since  $\mathbf{A}$  cannot be adjusted based on test data during inference, it remains essentially static. Module Generation: In STIDGCN (Liu & Zhang, 2024), dynamic graphs are generated through a pattern bank that integrates attention mechanisms. Meanwhile, MegaCRN (Jiang et al., 2023) generates dynamic graphs through memory cells and updates them based on input data.

DPGNet also generates dynamic graphs through a modular generation approach. Unlike existing methods, DPGNet’s AGL leverages G-RNN to generate corresponding local dynamic graphs for each patch. Through a recursive process, the generation of the current dynamic graph considers information from the previous patch, allowing the model to retain important node information while discarding weak connections. As a result, the dynamic graphs generated by DPGNet through AGL can sense local changes in spatial relationships, enabling the model to promptly adjust the learned adjacency relationships. Additionally, these dynamic graphs are less affected by noise, enhancing the model’s ability to focus on relevant spatiotemporal patterns.

### A.2 DATASET

Table 5: Statistics of datasets. METR-LA, PEMS-Bay, and PEMS08 are widely used spatiotemporal series datasets, while Electricity and Weather are commonly used time series datasets.

Type	Dataset	Node	Edges	Time Step	Time Interval	Time Range
Speed	METR-LA	207	1722	34272	5 minutes	Mar 1st 2012 to Jun 30th 2012
Flow	PEMS-Bay	325	2694	52116	5 minutes	Jan 1st 2017 to June 30th 2017
	PEMS08	170	548	17856	5 minutes	July 1st 2018 to Aug 31th 2018
Time-Series	Electricity	321	-	26304	1 hour	July 1st 2016 to July 2nd 2019
	Weather	21	-	52695	10 minutes	Jan 1st 2020 to Jan 1st 2021

In this section, we present the statistical properties and the initial adjacency configuration of the datasets used in the paper for spatiotemporal prediction tasks. Table 5 presents the statistical properties of datasets. For PEMS-Bay, METR-LA, and PEMS08, we follow previous studies Lee & Ko (2024); Bai et al. (2020); Liu & Zhang (2024) and use the physical connectivity among nodes as the initial adjacency configuration. For Weather and Electricity, no explicit relationships among nodes are provided. Therefore, we use the identity matrix as the initial adjacency configuration.

### A.3 BASELINE MODELS

We conduct comparative experiments with five state-of-the-art spatialtemporal prediction models. Below is a brief introduction to each of these models.

- PGCN (Shin & Yoon, 2024) proposes a novel approach that enables the dynamic updating of graph structures during the training phase to capture complex, time-varying spatial correlations in transportation networks and adapt to online input data across both training and testing stages.

- PMC-GCN (Gao et al., 2024) introduces an innovative personalized-enhanced graph convolution that employs learnable diagonal matrices combined with multi-head self-attention modules to improve prediction robustness.
- STIDGCN (Liu & Zhang, 2024) designs a dynamic interactive learning framework, comprising spatial and temporal modules, to effectively capture spatiotemporal dependencies in traffic flow data.
- TESTAM (Lee & Ko, 2024) presents a pioneering hybrid model that integrates three specialized experts and incorporates time-aware attention mechanisms along with dynamic graph modeling to address both recurrent and non-recurrent traffic patterns.
- WAVGCRN (Qian & Mallick, 2024) develops a multi-scale spatiotemporal prediction framework that integrates wavelet transforms with graph convolutional recurrent networks. The model synergizes prior road network knowledge with data-driven graph learning to achieve dynamic graph structure adaptation, thereby enhancing spatiotemporal dependency modeling and forecasting performance.

#### A.4 EVALUATION METRICS

The calculation formula for MAE and MSE used in the experiment is as follows: Let the predictions be denoted as:  $\hat{\mathbf{X}} = (\hat{\mathbf{X}}_{L+1}, \hat{\mathbf{X}}_{L+2}, \dots, \hat{\mathbf{X}}_{L+O}) \in \mathbb{R}^{N \times O \times C}$ . For the time step  $t \in \{L+1, \dots, L+O\}$ , the corresponding ground truth and prediction are respectively expressed as  $\mathbf{X}_t = (\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, \dots, \mathbf{x}_{t,N}) \in \mathbb{R}^{C \times N}$  and  $\hat{\mathbf{X}}_t = (\hat{\mathbf{x}}_{t,1}, \hat{\mathbf{x}}_{t,2}, \dots, \hat{\mathbf{x}}_{t,N}) \in \mathbb{R}^{C \times N}$ , where  $\mathbf{x}_{t,n} \in \mathbb{R}^C$  denotes the feature of node  $n$  at time step  $t$ , and  $C$  represents the feature dimension, and  $\hat{\mathbf{x}}_{t,n} \in \mathbb{R}^C$  denotes the corresponding prediction. The formulations for Mean Absolute Error (MAE) and Mean Squared Error (MSE) are respectively defined as:

$$\text{MAE} := \frac{1}{O} \frac{1}{N} \sum_{t=L+1}^{L+O} \sum_{n=1}^N \frac{1}{C} \|\mathbf{x}_{t,n} - \hat{\mathbf{x}}_{t,n}\|_1,$$

$$\text{MSE} := \frac{1}{O} \frac{1}{N} \sum_{t=L+1}^{L+O} \sum_{n=1}^N \frac{1}{C} \|\mathbf{x}_{t,n} - \hat{\mathbf{x}}_{t,n}\|_2^2,$$

where,  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote the  $\ell_1$  and  $\ell_2$  norms, respectively. These formulations express the errors as the average norm of the discrepancy vectors across all time steps and nodes.

#### A.5 LONG-TERM SPATIAL TEMPORAL PREDICTION

In this section, we will present the complete experimental results to highlight the accuracy advantage of DPGNet in long-term spatial-temporal prediction tasks. Table 6 presents the complete results of DPGNet on the long-term spatiotemporal prediction task. As shown in the Table, DPGNet achieved optimal results in 79.17% of the long-term prediction tasks and suboptimal results in the remaining predictions.

#### A.6 AGL ANALYSIS

In this section, we will present the complete experimental results to validate the effectiveness of the AGL module. Table 7 presents the adjacency matrices learned by the baseline models and DPGNet on the METR-LA and PEMS08 datasets. For PMC-GCN and WAVGCRN, they only account for explicit relationships between nodes, without considering implicit relationships. Therefore, we did not include the adjacency matrices learned by these models in Table 7. As shown in the table, the adjacency matrix learned by DPGNet exhibits the fewest weak connections and most connections that contain significant information. Table 8 presents the complete results of the replacement experiment. For an introduction to the replacement experiment, please refer to Section 4.3. It can be observed that in 81.25% of the experimental scenarios, AGL outperforms the existing graph generation methods.

Table 6: Results of long-term spatiotemporal prediction. We report the average value and standard deviation of experiments. **Bolded values** indicate the best results, underlined values indicate the second-best results.

Dataset	Model	$O = 24$		$O = 36$		$O = 48$	
		MAE	MSE	MAE	MSE	MAE	MSE
METR-LA	PGCN	7.3111	<u>194.2460</u>	8.3067	237.6264	9.1226	<u>262.8521</u>
		$\pm 0.1075$	$\pm 1.8733$	$\pm 0.1487$	$\pm 0.6319$	$\pm 0.4191$	$\pm 2.8221$
	PMC-GCN	7.4773	195.8361	8.9720	239.7505	9.3148	267.5599
		$\pm 0.1788$	$\pm 1.8718$	$\pm 0.0865$	$\pm 1.9711$	$\pm 0.1417$	$\pm 1.9818$
	STIDGCN	7.5140	197.2141	8.6893	<u>235.4212</u>	9.9960	278.8592
		$\pm 0.1747$	$\pm 1.7023$	$\pm 0.1695$	$\pm 2.2484$	$\pm 0.0375$	$\pm 5.1389$
	TESTAM	7.5162	204.7470	8.9428	257.9806	9.4628	274.6246
		$\pm 0.2481$	$\pm 7.2873$	$\pm 0.2868$	$\pm 6.7583$	$\pm 0.1818$	$\pm 5.1078$
WAVGCRN	16.7661	509.1225	10.6540	283.3590	18.5267	541.1127	
	$\pm 0.3883$	$\pm 10.5507$	$\pm 1.0256$	$\pm 27.5473$	$\pm 1.8289$	$\pm 87.5866$	
DPGNet	<b>7.1034</b>	<b>191.6066</b>	<b>7.8676</b>	<b>223.1617</b>	<b>8.6342</b>	<b>249.3086</b>	
		$\pm 0.2800$	$\pm 4.4801$	$\pm 0.4021$	$\pm 5.2667$	$\pm 0.3057$	$\pm 2.6067$
PEMS08	PGCN	24.2191	2279.4170	25.6723	2611.3066	30.8520	3160.2815
		$\pm 0.3345$	$\pm 21.5158$	$\pm 0.5522$	$\pm 45.6822$	$\pm 1.4169$	$\pm 129.2806$
	PMC-GCN	26.4791	2530.5159	29.1199	3047.9412	30.1908	3196.2134
		$\pm 0.3335$	$\pm 34.0195$	$\pm 0.1475$	$\pm 13.5515$	$\pm 0.3271$	$\pm 25.2430$
	STIDGCN	23.8807	<u>2240.3438</u>	<u>25.4253</u>	<u>2526.1126</u>	<u>26.6498</u>	<u>2718.4949</u>
		$\pm 0.0020$	$\pm 0.7330$	$\pm 0.0013$	$\pm 0.1567$	$\pm 0.0017$	$\pm 0.6759$
	TESTAM	37.1592	3446.3430	42.3845	4308.4917	48.4166	5585.1382
		$\pm 3.5786$	$\pm 446.6575$	$\pm 3.1031$	$\pm 422.9279$	$\pm 4.1474$	$\pm 622.0002$
WAVGCRN	34.5698	3212.1919	44.2962	4718.5728	49.0788	5506.2710	
	$\pm 0.9612$	$\pm 123.5820$	$\pm 1.0149$	$\pm 161.8669$	$\pm 1.8447$	$\pm 249.7399$	
DPGNet	<b>23.0820</b>	<b>2052.7918</b>	<b>23.5623</b>	<b>2268.9053</b>	<b>24.6720</b>	<b>2463.2017</b>	
	$\pm 0.4422$	$\pm 33.791$	$\pm 0.0505$	$\pm 5.3981$	$\pm 0.1992$	$\pm 16.2884$	
PEMS-Bay	PGCN	2.0592	18.7733	2.2667	21.9986	2.4562	24.5206
		$\pm 0.0296$	$\pm 0.4506$	$\pm 0.0128$	$\pm 0.2535$	$\pm 0.0331$	$\pm 0.4006$
	PMC-GCN	2.3946	24.4524	2.6632	29.6252	2.9132	34.0172
		$\pm 0.0543$	$\pm 0.2096$	$\pm 0.0304$	$\pm 0.3363$	$\pm 0.0329$	$\pm 0.3319$
	STIDGCN	1.9394	<b>16.8079</b>	<b>2.0612</b>	<b>19.1240</b>	2.2206	21.0813
		$\pm 0.0164$	$\pm 0.1429$	$\pm 0.01262$	$\pm 0.1689$	$\pm 0.0103$	$\pm 0.0648$
	TESTAM	2.2037	20.8823	2.4821	25.0493	2.7603	29.6056
		$\pm 0.1150$	$\pm 0.9980$	$\pm 0.0833$	$\pm 0.9822$	$\pm 0.0908$	$\pm 1.3966$
WAVGCRN	2.6232	27.6689	3.7462	42.1594	4.2947	54.7341	
	$\pm 0.1028$	$\pm 1.1057$	$\pm 0.0162$	$\pm 0.9214$	$\pm 0.1673$	$\pm 3.0691$	
DPGNet	<b>1.9350</b>	17.2711	2.0806	19.3558	<b>2.1948</b>	<b>20.9597</b>	
	$\pm 0.0077$	$\pm 0.1273$	$\pm 0.01243$	$\pm 0.2204$	$\pm 0.0100$	$\pm 0.1162$	
Weather	PGCN	0.4522	0.4086	0.4924	0.4641	0.5955	0.6402
		$\pm 0.0283$	$\pm 0.0329$	$\pm 0.0400$	$\pm 0.0727$	$\pm 0.0830$	$\pm 0.1982$
	PMC-GCN	<b>0.1409</b>	<b>0.0969</b>	0.1879	0.1231	0.2198	0.1437
		$\pm 0.0139$	$\pm 0.0037$	$\pm 0.0033$	$\pm 0.0014$	$\pm 0.0209$	$\pm 0.0185$
	STIDGCN	0.1565	0.1011	0.1878	0.1229	0.2137	0.1390
		$\pm 0.0026$	$\pm 0.0008$	$\pm 0.0198$	$\pm 0.0158$	$\pm 0.0019$	$\pm 0.0008$
	TESTAM	0.2145	0.1445	0.2452	0.1693	0.3082	0.2938
		$\pm 0.0286$	$\pm 0.0133$	$\pm 0.0585$	$\pm 0.0777$	$\pm 0.0381$	$\pm 0.0552$
WAVGCRN	0.1971	0.1242	0.2285	0.1588	0.2932	0.2238	
	$\pm 0.0039$	$\pm 0.0027$	$\pm 0.0103$	$\pm 0.0112$	$\pm 0.0255$	$\pm 0.0288$	
DPGNet	0.1489	0.0988	<b>0.1771</b>	<b>0.1177</b>	<b>0.1990</b>	<b>0.1327</b>	
	$\pm 0.0056$	$\pm 0.0018$	$\pm 0.0183$	$\pm 0.0164$	$\pm 0.0280$	$\pm 0.0476$	

Table 7: The learned adjacency matrix of the model. Adjacency values within the range of (0, 0.1] are considered weak connections.

Datasets	Model	Similarity Range			
		0	(0-0.1]	(0.1-0.5]	(0.5-1]
METR-LA	DPGNet	90.94%	7.47%	1.31%	0.28%
	GWNet	0.00%	98.84%	1.07%	0.09%
	PGCN	0.00%	100.00%	0.00%	0.00%
	STIDGCN	20.29%	79.71%	0.00%	0.00%
	TESTAM	0.00%	100.00%	0.00%	0.00%
PEMS08	DPGNet	89.76%	8.95%	0.93%	0.37%
	GWNet	0.00%	99.03%	0.86%	0.11%
	PGCN	0.00%	100.00%	0.00%	0.00%
	STIDGCN	10.14%	89.86%	0.00%	0.00%
TESTAM	0.00%	100.00%	0.00%	0.00%	

### A.7 SEASON LEARNER ANALYSIS

In this section, we introduce the generation process of the simulated data used in Section 4.3 and provide a detailed description of the experimental setup. In the experiment, we set both the historical window length  $L$  and the prediction length  $O$  to 96, and generate 900 data windows. For the  $i$ -th

Table 8: Results of the replacement experiment, with shading indicating performance improvement after replacing the original graph generation method with AGL.

Dataset	Model	15 minutes		30 minutes		60 minutes		120 minutes	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
GWNNet	vanilla	3.4800	66.8846	4.4492	110.7876	5.6984	134.0552	7.3338	189.2081
	w/ AGL	<u>3.3575(3.32%↓)</u>	<u>64.7486(3.19%↓)</u>	<u>4.1814(6.02%↓)</u>	<u>98.6049(11.00%↓)</u>	<u>5.3846(5.31%↓)</u>	<u>131.3513(2.02%↓)</u>	<u>7.1034(3.14%↓)</u>	<u>191.6066(1.21%↑)</u>
PMC-GCN	vanilla	3.5768	66.5593	4.5233	92.7049	5.8151	139.9326	7.4774	195.8361
	w/ AGL	<u>3.5700(0.19%↓)</u>	<u>67.0761(0.78%↑)</u>	<u>4.6437(2.66%↑)</u>	<u>93.7024(1.08%↑)</u>	<u>5.7812(0.58%↓)</u>	<u>136.1889(2.68%↓)</u>	<u>7.9444(6.25%↑)</u>	<u>206.2663(5.33%↑)</u>
METR-LA	vanilla	3.7565	64.9473	4.3655	92.5479	5.4594	135.1364	7.5140	197.2141
	w/ AGL	<u>3.5665(5.06%↓)</u>	<u>64.4809(0.72%↓)</u>	<u>4.3366(0.66%↓)</u>	<u>91.3041(1.34%↓)</u>	<u>5.4104(0.90%↓)</u>	<u>131.5148(2.69%↓)</u>	<u>6.9571(7.41%↓)</u>	<u>184.8884(6.25%↓)</u>
STGCN	vanilla	3.6539	66.7832	4.4214	91.2254	5.6993	134.4666	7.4967	198.6619
	w/ AGL	<u>3.5311(3.36%↓)</u>	<u>66.2596(0.79%↓)</u>	<u>4.3790(0.96%↓)</u>	<u>95.7793(4.99%↑)</u>	<u>5.9036(3.58%↑)</u>	<u>138.7884(3.23%↑)</u>	<u>7.7486(3.36%↑)</u>	<u>198.0429(0.31%↓)</u>
WAVGCRN	vanilla	4.9907	82.6108	5.3720	111.7549	6.9338	160.7631	16.76613	509.1225
	w/ AGL	<u>3.6849(26.16%↓)</u>	<u>67.4041(18.41%↓)</u>	<u>4.6048(14.28%↓)</u>	<u>94.2841(15.63%↓)</u>	<u>6.7905(2.07%↓)</u>	<u>149.5729(6.96%↓)</u>	<u>7.2698(56.64%↓)</u>	<u>196.2354(61.47%↓)</u>
GWNNet	vanilla	15.8843	830.0247	18.3152	1196.2059	21.1694	1745.7639	23.8123	2278.5420
	w/ AGL	<u>15.5646(2.01%↓)</u>	<u>826.3616(0.44%↓)</u>	<u>17.9393(2.05%↓)</u>	<u>1197.1906(0.08%↑)</u>	<u>20.8184(1.66%↓)</u>	<u>1765.5440(1.13%↑)</u>	<u>23.0820(3.07%↓)</u>	<u>2052.7918(9.91%↓)</u>
PMC-GCN	vanilla	15.9641	840.1459	18.3765	1205.8416	22.0897	1808.6214	26.4791	2530.5159
	w/ AGL	<u>15.6584(1.91%↓)</u>	<u>817.4102(2.71%↓)</u>	<u>18.0830(1.60%↓)</u>	<u>1191.1047(1.22%↓)</u>	<u>21.6807(1.85%↓)</u>	<u>1778.2541(1.68%↓)</u>	<u>26.1885(1.10%↓)</u>	<u>2460.0098(2.79%↓)</u>
PEMS08	vanilla	15.3111	790.1659	17.5923	1201.1804	19.8736	1641.5082	23.8807	2240.3438
	w/ AGL	<u>15.2417(0.45%↓)</u>	<u>786.2132(0.50%↓)</u>	<u>17.3605(1.32%↓)</u>	<u>1200.6130(0.05%↓)</u>	<u>19.8284(0.27%↓)</u>	<u>1677.8730(2.22%↑)</u>	<u>22.1619(7.20%↓)</u>	<u>2168.7754(3.19%↓)</u>
STGCN	vanilla	18.1283	951.7638	20.3185	1306.2333	22.9451	1931.8759	28.2738	2874.3938
	w/ AGL	<u>18.1212(0.03%↓)</u>	<u>921.6303(3.17%↓)</u>	<u>20.0083(1.53%↓)</u>	<u>1304.6440(0.12%↓)</u>	<u>23.1401(0.85%↑)</u>	<u>1924.5240(0.38%↓)</u>	<u>28.8598(2.07%↑)</u>	<u>2779.5898(3.30%↓)</u>
WAVGCRN	vanilla	22.6586	1314.9849	21.0412	1359.0916	28.9435	2339.8194	34.5698	3212.1919
	w/ AGL	<u>18.2564(19.43%↓)</u>	<u>1002.9889(23.73%↓)</u>	<u>20.3240(3.41%↓)</u>	<u>1291.0087(5.01%↓)</u>	<u>20.3240(3.41%↓)</u>	<u>1897.6740(18.90%↓)</u>	<u>27.6968(19.89%↓)</u>	<u>2561.0410(20.27%↓)</u>

Table 9: Ablation study for core modules: AGL and ASL. Trend and Season, respectively, denote the trend feature processing and seasonal feature processing in ASL. **Bolded values** indicate the best results, underlined values indicate the second-best results.

Dataset	model		120 minutes		180 minutes		240 minutes		
	AGL	ASL		MAE	MSE	MAE	MSE	MAE	MSE
		Trend	Season						
METR-LA	✓	✓	✓	<b>7.1034</b>	<b>191.6066</b>	<b>8.1956</b>	<b>224.3926</b>	<b>8.6342</b>	<b>249.3086</b>
	✓	✓	×	7.7387	204.8815	8.8438	235.1586	9.4877	255.3363
	✓	×	✓	<u>7.6340</u>	<u>194.6127</u>	<u>8.6864</u>	<u>230.1645</u>	<u>9.2585</u>	<u>249.8750</u>
	✓	×	×	8.2673	213.8754	9.3672	253.5769	9.6361	278.6919
	×	✓	✓	8.0057	215.9630	9.1913	245.3591	9.9757	287.2614
	×	✓	×	8.3199	214.2154	9.4653	254.7740	10.3952	290.2376
GWNNet	×	×	✓	8.0788	212.7841	9.0339	223.5677	10.1710	280.9923
	×	×	×	8.4021	223.9778	9.5560	257.2167	10.4587	290.7861

window of the simulated node, the node sequence is generated using a cosine function (Woo et al., 2022):

$$Node_1^i = \alpha_1 \cos(2\pi f_1(\mathbf{t} + i)) + \alpha_2 \cos(2\pi f_2(\mathbf{t} + i)) + \epsilon_1, \quad (13)$$

$$Node_2^i = \cos\left(\frac{2\pi}{900}i + s\right) * Node_1^i + \epsilon_2, \quad (14)$$

where  $\alpha_1, \alpha_2 = 0.15$ ,  $f_1 = \frac{1}{10}$ ,  $f_2 = \frac{1}{13}$ , and  $s = \frac{\pi}{4}$ .  $\epsilon_1$  and  $\epsilon_2$  are noise terms sampled from a Gaussian distribution with mean 0 and variance 0.005.  $\mathbf{t} = [0, 1, \dots, L + O] \in \mathbb{R}^{L+O}$  denotes the input time sequence. Following this setting,  $Node_1$  and  $Node_2$  exhibit a clearly defined seasonal relationship.

## A.8 ABLATION STUDY

In this section, we perform ablation studies to assess the contribution of DPGNet’s core modules: Adaptive Graph Learner(AGL) and Adaptive Season Learner(ASL). Table 9 presents the results of the ablation study. In addition, we further assess the contribution of trend feature processing and seasonal feature processing in ASL. All core modules within DPGNet demonstrate their importance, as evidenced by a noticeable decline in prediction accuracy when any of these components is removed.

## A.9 SENSITIVITY ANALYSIS

In this section, we conduct a sensitivity analysis of the multi-scale parameter  $m$ . Table 10 presents the results of the sensitivity analysis for the multi-scale parameter  $m$ . As shown in the table, the multi-scale processing approach enhanced the prediction accuracy of DPGNet. Moreover, in the majority of cases, setting  $m = 1$  yielded the best overall performance of the model.

## A.10 LIMITATIONS

In the following, we briefly discuss the limitations of DPGNet’s core modules, AGL and ASL.

Table 10: Sensitivity analysis of the multi-scale parameter  $m$  on METR-LA and Weather. **Bolded values** indicate the best results, underlined values indicate the second-best results.

Dataset	Model	120 minutes		180 minutes		240 minutes	
		MAE	MSE	MAE	MSE	MAE	MSE
METR-LA	$m = 0$	8.0352	215.9630	8.7688	229.228	9.6037	259.0429
	$m = 1$	<b>7.1034</b>	<b>191.6066</b>	<u>8.1956</u>	<u>224.3926</u>	<b>8.6342</b>	<b>249.3086</b>
	$m = 2$	7.6167	194.5623	<b>7.8676</b>	<b>223.1617</b>	<u>9.1886</u>	<u>257.1401</u>
	$m = 3$	<u>7.2794</u>	<u>193.0593</u>	8.5363	222.5458	9.9809	269.3656
Weather	$m = 0$	0.1867	0.1119	0.2206	0.1443	0.2678	0.1824
	$m = 1$	<u>0.1653</u>	<u>0.1019</u>	<b>0.1969</b>	<u>0.1267</u>	<b>0.2136</b>	<b>0.1437</b>
	$m = 2$	0.1825	0.1120	0.2043	0.1347	0.2490	<u>0.1590</u>
	$m = 3$	<b>0.1624</b>	<b>0.1009</b>	0.2101	0.1377	<u>0.2400</u>	0.1601
	$m = 4$	0.1803	0.1124	0.2061	0.1400	<u>0.2474</u>	0.1708
	$m = 5$	0.1672	0.1058	<u>0.2031</u>	<b>0.1224</b>	0.2640	0.1716

For AGL, while the patch processing approach significantly reduces the number of G-RNNs, the number still increases with the length of the input sequence. An excessive number of G-RNNs can lead to issues such as gradient explosion and high computational cost. Additionally, although AGL can handle weak connections between nodes, as demonstrated by the experiments in Section 3, it lacks the capability to extract useful information from these weak connections.

For ASL, its multi-scale framework improves the model’s long-term prediction ability. However, ASL lacks an automated mechanism for selecting the multi-scale parameter  $m$ .

#### THE USE OF LARGE LANGUAGE MODELS

In this study, large language models (LLMs) were solely used for text refinement purposes. No other applications or functions of LLMs were utilized in the research.