Learning the Evolutionary and Multi-scale Graph Structure for Multivariate Time Series Forecasting

Junchen Ye[1*], Zihan Liu[1*], Bowen Du[1], Leilei Sun[1], Weimiao Li[1], Yanjie Fu[2], Hui Xiong[3]

[1]Beihang University [2]University of Central Florida [3] Hong Kong University of Science and Technology

Introduction

Time series forecasting is a ubiquitous problem in practical scenarios. By modeling the evolution of the states or events in the future, it enables decision-making and plays a vital role in numerous domains, such as traffic, healthcare, and finance.
Recent studies have shown great promise in applying graph neural networks for multivariate time series forecasting, where the interactions of time series are described as a graph structure and the variables are represented as the graph nodes. Along this line, existing methods usually assume that the graph structure (or the adjacency matrix), which determines the aggregation manner of graph neural network, is fixed either by definition or self-learning. However,

1. **The interactions of variables can be dynamic and evolutionary in real-world scenarios.**
2. **The interactions of time series are quite different if they are observed at different time scales.**

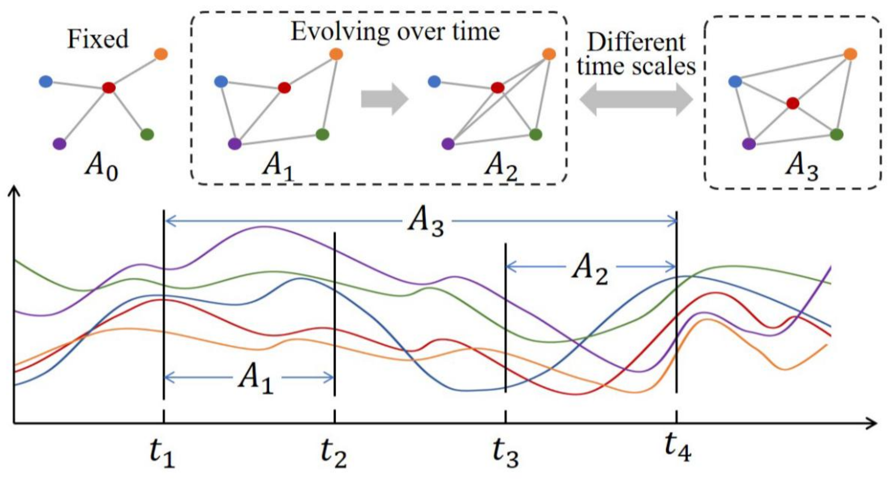
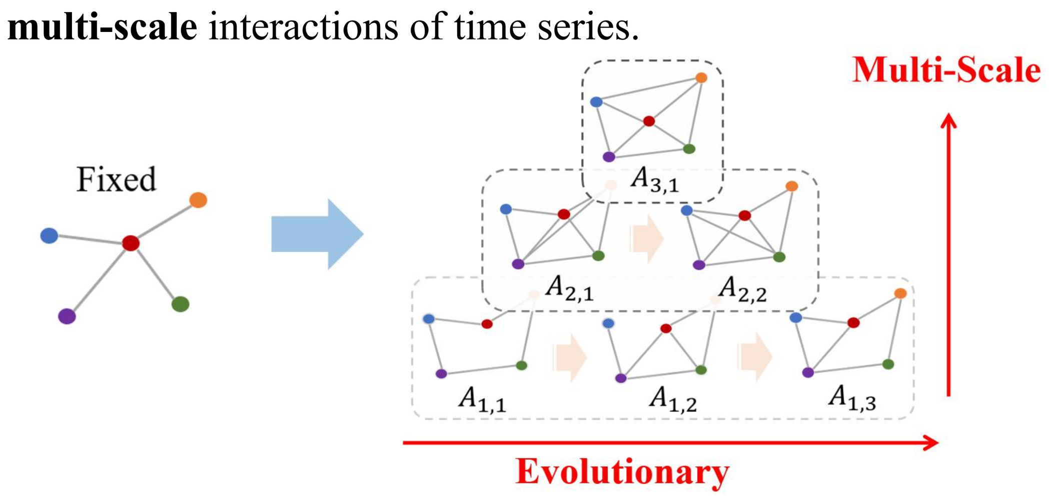


Figure 1: The possible interactions of variables in multivariate time series forecasting. Most existing works utilize the fixed correlation (A_0). However, the graph structure is evolving (A_1 and A_2) and varies in different observation scales (A_3).

To equip the graph neural network with a flexible and practical graph structure, in this paper, we investigate how to model the **evolutionary** and **multi-scale** interactions of time series.

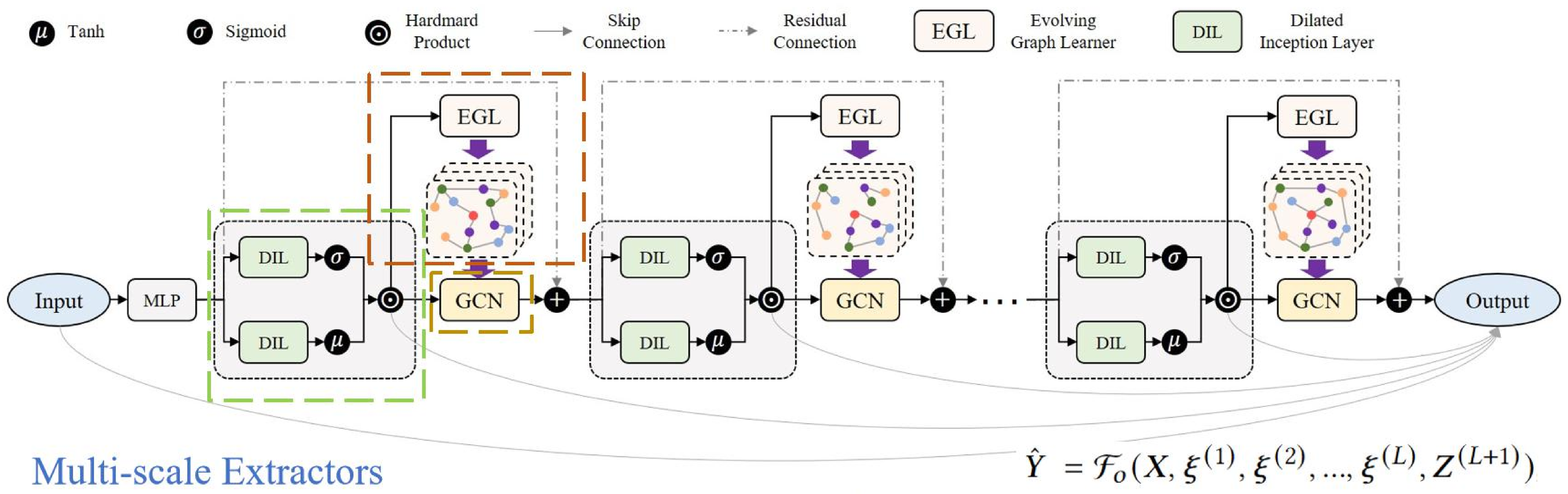


When we propose to take a further step and address the two problems above, three challenges are faced:

1. The evolving graph structure is not only influenced by the current input but also strongly correlated to itself at the previous time step. The recurrent construction manner has been rarely discussed.
2. Generating the graph structure for each time step to model the evolution through existing self-learned methods would bring too many parameters, which results in difficulty for model convergence.
3. It is a nontrivial endeavor to capture the scale-specific graph structure among nodes due to the excess information and messy relationship behind it.

Methodology

Figure 2: The framework of ESG.



Multi-scale Extractors

$$\hat{Y} = \mathcal{F}_o(X, \xi^{(1)}, \xi^{(2)}, ..., \xi^{(L)}, Z^{(L+1)})$$

Each extractor : temporal convolution module , $\xi^{(l)} = \mathcal{F}_t^{(l)}(Z^{(l)})$,

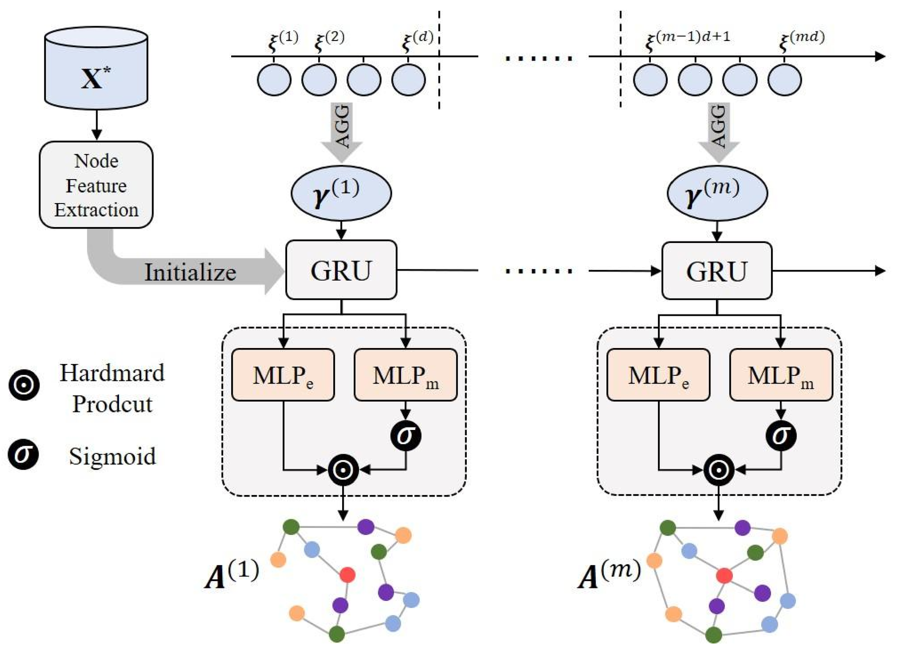
evolving graph structure learner, $\mathrm{A}^{(l)} = \mathcal{F}_a^{(l)}(\xi^{(l)})$,

graph convolution module $Z'^{(l+1)} = \mathcal{F}_g^{(l)}(\xi^{(l)}, \mathrm{A}^{(l)})$,

Overview: The multivariate time series is first fed into a MLP to obtain the initial representation, and the stacked multi-scale extractors follow. Each extractor is made up of three components. The temporal convolution module is utilized to capture multiscale representations on the temporal dimension. The output of the evolving graph structure learner is a series of adjacency matrices which are fed into the graph convolution module to model the evolutionary correlations among time series. skip connection is utilized to deliver the information to the final output.

Evolving Graph Structure Learner (EGL)

We design an EGL to extract dynamic correlations among variables, which is the highlight of our work. This module both considers the dependency with the current input values and the graph structure at last time step, which could be formulated under a recurrent manner.



- Aggregate the features of each segment:

$$\gamma^{(m)} = AGG(\xi^{((m-1)d+1:md)}) \in \mathbb{R}^{N \times C_\xi}$$

$$[\gamma^{(1)}, \gamma^{(2)}, ..., \gamma^{(m)}, ..., \gamma^{(M)}]$$

- Update the node representation $\boldsymbol{\alpha}$ with GRU:

$$\boldsymbol{r}^{(m)} = \sigma(\boldsymbol{W}_r[\boldsymbol{\gamma}^{(m)}, \boldsymbol{\alpha}^{(m-1)}] + \boldsymbol{b}_r),$$
$$\boldsymbol{u}^{(m)} = \sigma(\boldsymbol{W}_u[\boldsymbol{\gamma}^{(m)}, \boldsymbol{\alpha}^{(m-1)}] + \boldsymbol{b}_u),$$
$$\boldsymbol{o}^{(m)} = \mu(\boldsymbol{W}_o[\boldsymbol{\gamma}^{(m)}, (\boldsymbol{r}^{(m)} \odot \boldsymbol{\alpha}^{(m-1)})] + \boldsymbol{b}_o),$$
$$\boldsymbol{\alpha}^{(m)} = \boldsymbol{u}^{(m)} \odot \boldsymbol{\alpha}^{(m-1)} + (1 - \boldsymbol{u}^{(m)}) \odot \boldsymbol{o}^{(m)},$$

- Init hidden state of GRU:

$$\boldsymbol{\alpha}^{(0)} = \mathrm{MLP}_s(\boldsymbol{\alpha}_s)$$

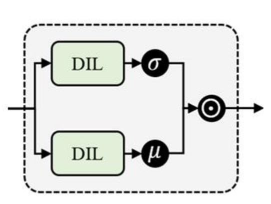
$$\boldsymbol{\alpha}_{s,i} = \mathcal{F}_s(X_i^*)$$

- Derive the graph structure:

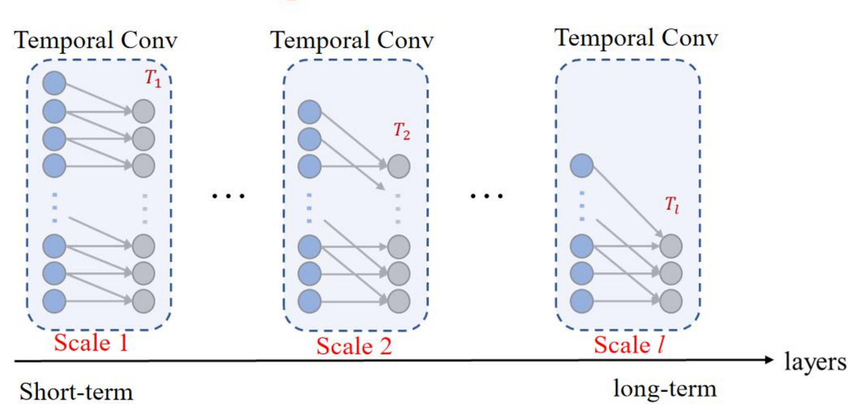
$$\hat{\mathrm{A}}_{ij}^{(m)} = \mathrm{MLP}_e(\boldsymbol{\alpha}_i^{(m)}, \boldsymbol{\alpha}_j^{(m)}),$$
$$\boldsymbol{M}_{ij}^{(m)} = \mathrm{MLP}_m(\boldsymbol{\alpha}_i^{(m)}, \boldsymbol{\alpha}_j^{(m)}),$$
$$\boldsymbol{A}^{(m)} = \hat{\mathrm{A}}^{(m)} \odot \sigma(\boldsymbol{M}^{(m)}),$$

Temporal Convolution Module

- Dilated inception layer (DIL):
  - Dilated causal convolution
  - Multiple filters with different sizes
- Gating mechanism:

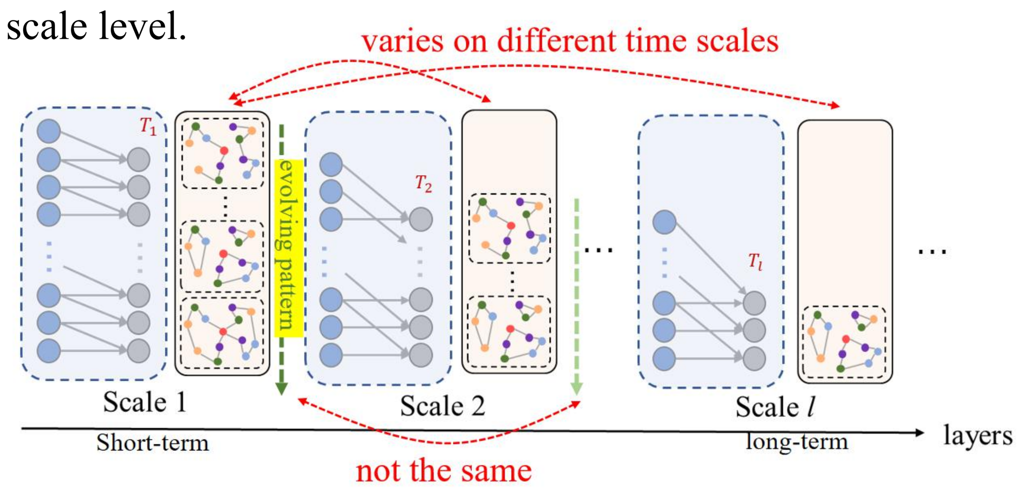


By Stacking multiple layers, the temporal convolution module capture temporal patterns at different temporal scales.



Scale-specific EGL

The dependency among variables not only evolves over time but also varies on different time scales. In addition, the evolving patterns of the graph structure are also not the same at different time scales. Thus, we utilize the scale-specific evolving graph structure learner to discover correlations among variables for the specific scale level.



Scale-specific evolving graph structure learner

$$[\boldsymbol{A}^{(l,1)}, \boldsymbol{A}^{(l,2)}, ..., \boldsymbol{A}^{(l,M^{(l)})}] = \boxed{\mathcal{F}_a^{(l)}}(\xi^{(l)}, d^{(l)}) \quad l \text{ -th EGL}$$

Evolving Graph Convolution Module

- Mix-hop propagation

information propagation: $\boldsymbol{H}_{(\psi)} = \beta\xi + (1-\beta)\boldsymbol{A}\boldsymbol{H}_{(\psi-1)}$ information selection: $Z' = \sum_{\psi=0}^{\Psi} \boldsymbol{H}_{(\psi)}\boldsymbol{W}_{(\psi)}$

- Fed into mix-hop propagation layer with its corresponding adjacency matrix

$$Z'^{(l,m)} = \mathcal{F}_g^{(l)}(\xi^{(l,(m-1)d^{(l)}+1:md^{(l)})}, \boldsymbol{A}^{(l,m)})$$

Formulation

- Multivariate Time Series Forecasting
  - The time series with N variables :

$$\mathrm{X} = \{X^{(1)}, X^{(2)}, \cdots, X^{(T)}\} \in \mathbb{R}^{N \times T \times C}$$

  - Given a look-back window P :

Single-step forecasting :

$$X^{(t-P+1:t)} \xrightarrow{\mathcal{F}_1} X^{(t+Q)}$$

Multi-step forecasting :

$$X^{(t-P+1:t)} \xrightarrow{\mathcal{F}_2} X^{(t+1:t+Q)}$$

Datasets & Setup

We conduct detailed experiments on six popular real-world datasets. Brief statistical information is listed in Table 1.

Table 1: The overall information for datasets.

Datasets	Nodes	Timesteps	Granularity	Task Types	Partition
Solar-Energy	137	52560	10min	Single-step	6/2/2
Electricity	321	26304	1hour		
Exchange Rate	8	7588	1day		
Wind	28	10957	1day		
NYC-Bike	250	4368	30min	Multi-step	7/1.5/1.5
NYC-Taxi	266	4368	30min		

We utilize two groups of evaluation metrics for the different forecasting tasks. For the single-step prediction, Root Relative Squared Error (RSE) and Empirical Correlation Coefficient (CORR) are selected. The multi-step prediction tasks are evaluated by Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Empirical Correlation Coefficient (CORR). The lower value indicates better performance for all evaluation metrics except CORR.

Results

Comparison With Baselines

Table 2: Comparison with baselines on single-step forecasting.

Dataset	Metrics	Solar-Energy 3	6	12	24	Electricity 3	6	12	24	Exchange Rate 3	6	12	24	Wind 3	6	12	24
AR	RSE	0.2435	0.3790	0.5911	0.8699	0.0995	0.1035	0.1050	0.1054	0.0228	0.0279	0.0353	0.0445	0.7161	0.7572	0.8076	0.9371
	CORR	0.9710	0.9263	0.8107	0.5314	0.8845	0.8632	0.8591	0.8595	0.9734	0.9656	0.9526	0.9357	0.6459	0.6046	0.5560	0.4633
GP	RSE	0.2259	0.3286	0.5200	0.7973	0.1500	0.1907	0.1621	0.1273	0.0239	0.0272	0.0394	0.0580	0.6689	0.6761	0.6772	0.6819
	CORR	0.9751	0.9448	0.8518	0.5971	0.8670	0.8334	0.8394	0.8818	0.8713	0.8193	0.8484	0.8278	0.6964	0.6877	0.6846	0.6781
VARMLP	RSE	0.1922	0.2679	0.4244	0.6841	0.1393	0.1620	0.1557	0.1274	0.0265	0.0394	0.0407	0.0578	0.7356	0.7769	0.8071	0.8334
	CORR	0.9829	0.9655	0.9058	0.7149	0.8708	0.8389	0.8192	0.8679	0.8609	0.8725	0.8280	0.7675	0.6415	0.5973	0.5724	0.5470
RNN-GRU	RSE	0.1932	0.2628	0.4163	0.4852	0.1102	0.1144	0.1183	0.1295	0.0192	0.0264	0.0408	0.0626	0.6131	0.6479	0.6573	0.6381
	CORR	0.9823	0.9675	0.9150	0.8823	0.8597	0.8623	0.8472	0.8651	0.9786	0.9712	0.9531	0.9223	0.7403	0.7089	0.6956	0.7173
LSTNet	RSE	0.1843	0.2559	0.3254	0.4643	0.0864	0.0931	0.1007	0.1007	0.0226	0.0280	0.0356	0.0449	**0.6079**	0.6262	0.6279	**0.6257**
	CORR	0.9843	0.9690	0.9467	0.8870	0.9283	0.9135	0.9077	0.9119	0.9735	0.9658	0.9511	0.9354	**0.7436**	0.7275	0.7249	**0.7284**
TPA-LSTM	RSE	0.1803	0.2347	0.3234	0.4389	0.0823	0.0916	0.0964	0.1006	**0.0174**	**0.0241**	**0.0341**	**0.0444**	0.6093	0.6292	0.6290	0.6335
	CORR	0.9850	0.9742	0.9487	0.9081	0.9439	0.9337	0.9250	0.9133	0.9790	0.9709	0.9564	0.9381	0.7433	0.7240	0.7235	0.7202
MTGNN	RSE	0.1778	0.2348	0.3109	0.4270	0.0745	0.0878	0.0916	**0.0953**	0.0194	0.0259	0.0349	0.0456	0.6204	0.6346	0.6363	0.6426
	CORR	0.9852	0.9726	0.9509	0.9031	0.9474	0.9316	0.9278	0.9234	0.9786	0.9708	0.9551	0.9372	0.7337	0.7209	0.7164	0.7134
StemGNN	RSE	0.1839	0.2612	0.3564	0.4768	0.0799	0.0909	0.0989	0.1019	0.0506	0.0674	0.0676	0.0685	0.6197	0.6358	0.6243	0.6379
	CORR	0.9841	0.9679	0.9395	0.8740	0.9490	**0.9397**	**0.9342**	0.9209	0.8871	0.8703	0.8499	0.8738	0.7282	0.7202	0.7228	0.7130
ESG	RSE	**0.1708**	**0.2278**	**0.3073**	**0.4101**	**0.0718**	**0.0844**	**0.0898**	0.0962	0.0181	0.0246	0.0345	0.0468	0.6118	**0.6250**	**0.6272**	0.6298
	CORR	**0.9865**	**0.9743**	**0.9519**	**0.9100**	**0.9494**	0.9372	0.9321	**0.9279**	**0.9792**	**0.9717**	**0.9564**	**0.9392**	0.7417	**0.7281**	**0.7258**	0.7225

Table 3: Comparison with baselines on multi-step forecasting.

Dataset	Method	Horizon 3 RMSE	MAE	CORR	Horizon 6 RMSE	MAE	CORR	Horizon 12 RMSE	MAE	CORR	All RMSE	MAE	CORR
NYC-Bike	XGBoost	3.7048	2.2167	0.5232	4.1747	2.5511	0.3614	4.3925	2.7091	0.2894	4.0494	2.4689	0.4107
	DCRNN	3.0172	1.7917	0.6967	3.2369	1.9078	0.6609	3.5100	2.0325	0.6196	3.2274	1.8973	0.6601
	STGCN	2.6256	1.6456	0.7539	3.8368	2.2827	0.6282	4.3713	2.6052	0.4521	3.7829	2.2076	0.5933
	STG2Seq	3.4669	2.0409	0.5999	3.9145	2.2630	0.5079	4.2373	2.5163	0.4443	3.7843	2.2055	0.5413
	STSGCN	2.7328	1.6973	0.7386	2.8861	1.7416	0.7179	3.0548	1.8224	0.6903	2.8846	1.7538	0.7126
	MTGNN	2.5962	1.5668	0.7626	2.7588	1.6525	0.7447	3.3068	1.7892	0.6931	2.7791	1.6595	0.7353
	CCRNN	2.6538	1.6565	0.7534	2.7561	1.7061	0.7411	2.9436	1.8040	0.7029	2.7674	1.7133	0.7333
	GTS	2.7628	1.7159	0.7248	2.9287	1.7769	0.7007	3.1649	1.8905	0.6622	2.9258	1.7798	0.6985
	ESG	**2.5529**	**1.5483**	**0.7638**	**2.6484**	**1.6026**	**0.7511**	**2.8778**	**1.7173**	**0.7152**	**2.6727**	**1.6129**	**0.7449**
NYC-Taxi	XGBoost	15.0372	8.4121	0.6862	21.3395	11.8491	0.4433	26.7073	15.7165	0.0452	21.1994	11.6806	0.4416
	DCRNN	12.3223	7.0655	0.7591	15.1599	8.6639	0.6634	17.8194	10.5095	0.5395	14.8318	8.4835	0.6671
	STGCN	11.2175	6.1441	0.8090	14.0360	7.6797	0.7470	18.7168	10.2211	0.5922	14.6473	7.8435	0.7257
	STG2Seq	14.0756	7.7274	0.7258	19.1757	10.5066	0.5429	24.5691	14.3603	0.2855	19.2077	10.4925	0.5389
	STSGCN	10.5381	5.6448	0.8370	10.8444	5.7634	0.8302	11.9443	6.3185	0.7988	10.9692	5.8299	0.8242
	MTGNN	10.3394	5.6775	0.8374	10.7534	5.8168	0.8312	12.5164	6.5285	0.7972	10.9472	5.9192	0.8249
	CCRNN	9.3033	5.4586	0.8529	9.7794	5.6362	0.8438	10.9585	6.1416	0.8186	9.8744	5.6636	0.8416
	GTS	10.7796	6.2337	0.7974	13.0215	7.3251	0.7299	14.9906	8.5328	0.6524	12.7511	7.2095	0.7348
	ESG	**8.5745**	**4.8750**	**0.8656**	**9.0125**	**5.0500**	**0.8592**	**9.7857**	**5.4019**	**0.8450**	**8.9759**	**5.0344**	**0.8592**

Ablation Study

Table 4 show that all components contribute to the final state-of-the-art results to a certain extent.

ESG outperforms the methods using only one scale information by a large margin, which indicates the superiority of fusing the multi-scale representations to make the final prediction.

Table 4: Ablation Study.

Method	RMSE	MAE	CORR
Static Graph Only	2.7439±0.0438	1.6302±0.0176	0.7388±0.0050
w/o Scale-Specific	2.8102±0.0433	1.6663±0.0150	0.7259±0.0047
Same Pattern of Evolution	2.7274±0.0177	1.6296±0.0036	0.7402±0.0024
ESG	**2.6727±0.0117**	**1.6129±0.0086**	**0.7449±0.0051**

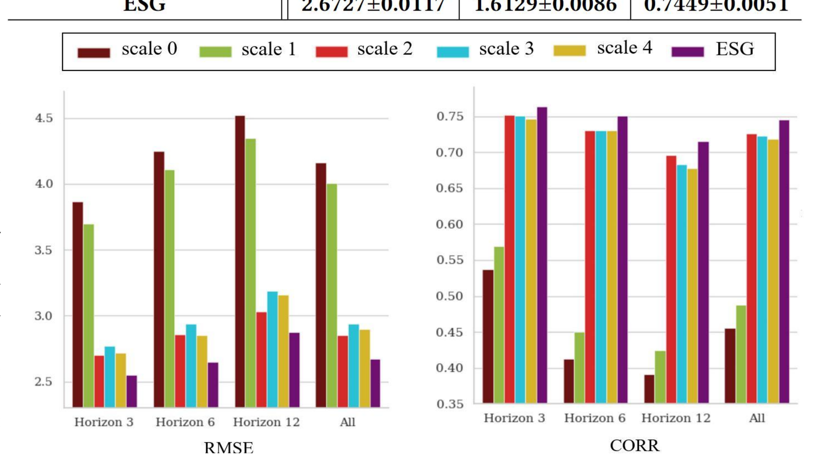


Figure 4: Utilizing the information at different scales.

Case Study: Verify the Effectiveness of EGL

- The evolutionary correlations

1) In Figure 5(c), before 16:30, station 166 and station 141 have a strong correlation with each other. However, after 16:30, station 141 remains stable but station 166 fluctuates dramatically. The fact that the correlations evolve from high to low is well captured by the adjacency matrices.

2)The evolutionary correlation captured by the adjacency matrices between station166 and station 217 rises in the beginning and falls in the end, which is also consistent with the fact shown in Figure 5(c).

- The correlations at different observation scales

In Figure 5(d) , the values in the adjacency matrix at the scale 1 $A^{(1,6)}$ tend to be highly polarized, which indicates the short-term dependency of the stations is more likely to differ from others. However, at the last scale, the more average values in the adjacency matrix $A^{(3,1)}$ clarify that the 4 time series possess the same pattern from the long-term view.

The two case studies above offer us a strong support to verify that the evolving and multi-scale correlations among multivariate time series are well captured by ESG.

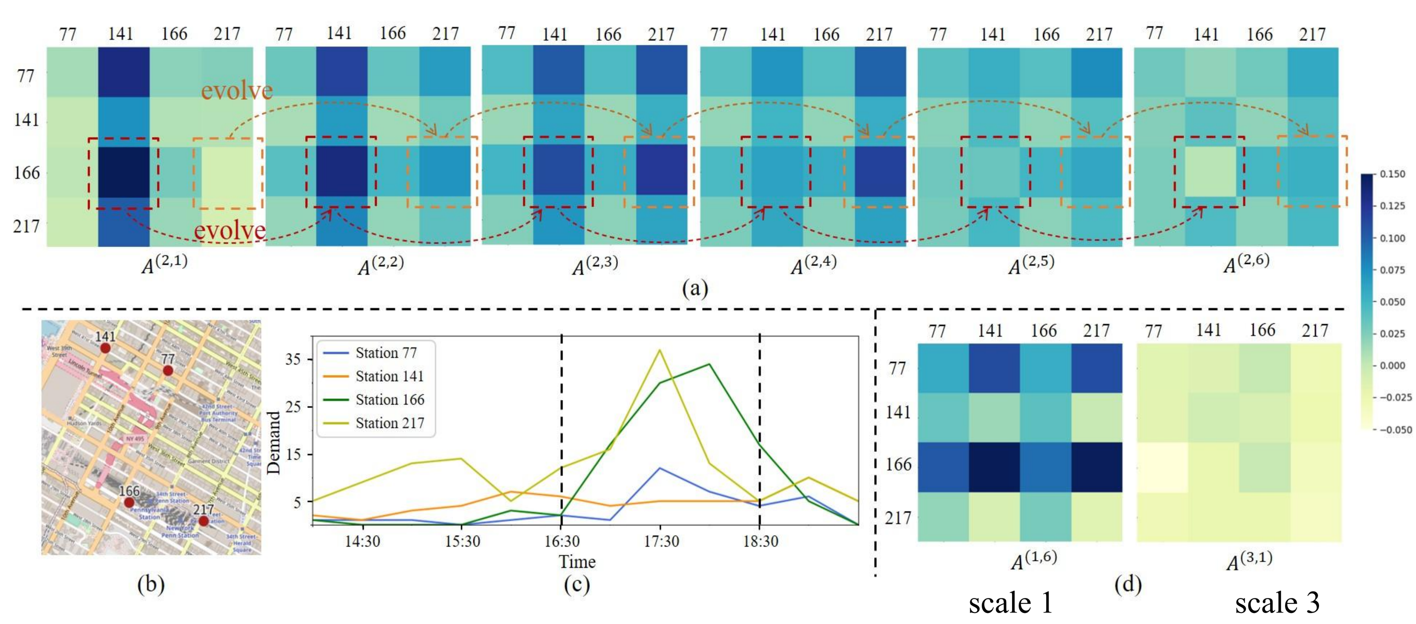


Figure 5: (a) A series of adjacency matrices in scale 2 on the NYC-Bike dataset, which reveals a strong evolving pattern. (b) The location of node 77, 141, 166 and 217 on the map. (c) The raw time series curves on 12 time steps, which corresponds to the adjacency matrices shown in (a) and (d). (d) Several adjacency matrices on scale 1 and 3.