# STEP:
# Pre-training Enhanced Spatial-temporal Graph Neural Network for Multivariate Time Series Forecasting

**Zezhi Shao**[1,2], Zhao Zhang , Fei Wang , Yongjun Xu
Institute of Computing Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences

中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY,CAS

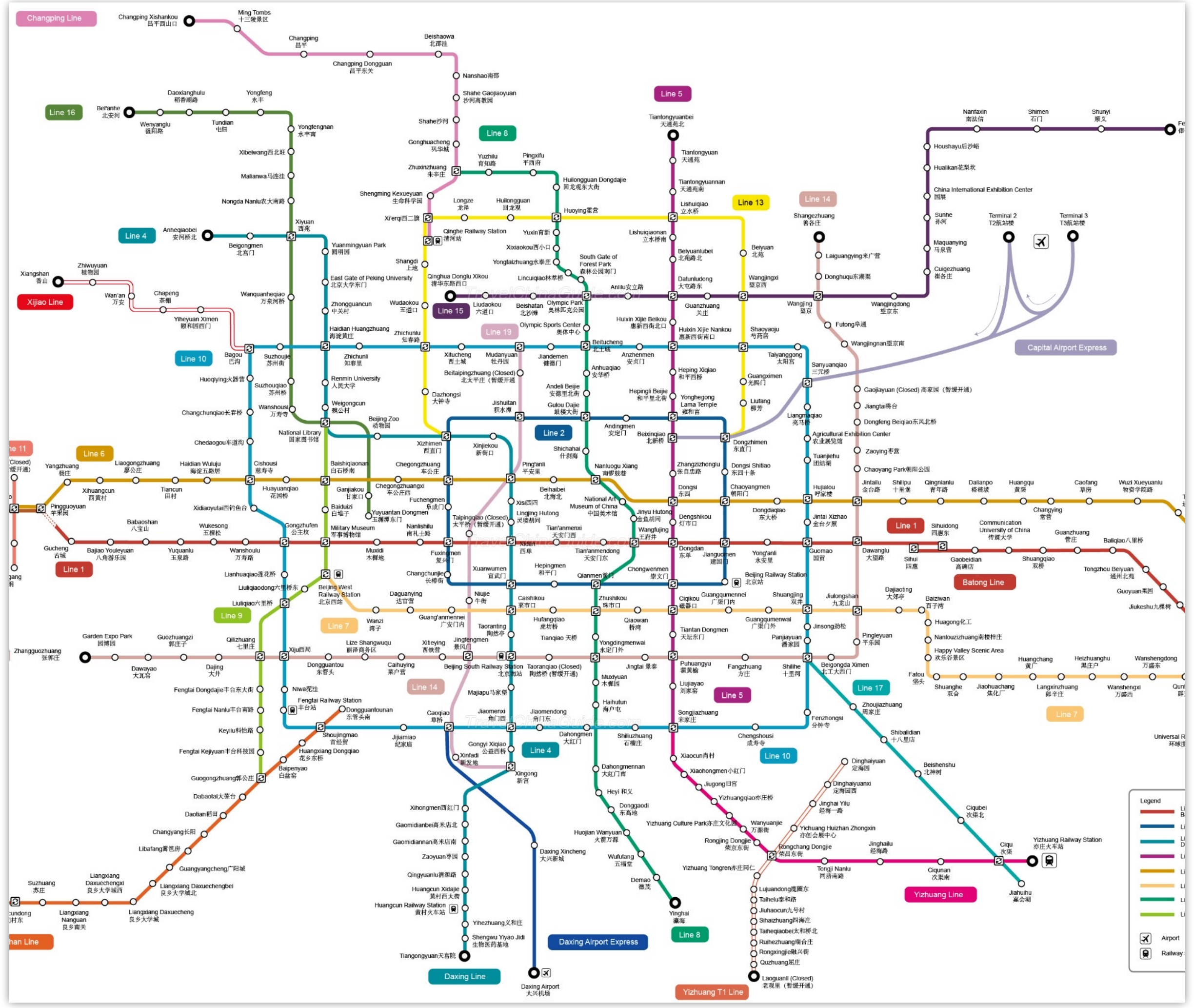
中国科学院
CHINESE ACADEMY OF SCIENCES

CONTENT

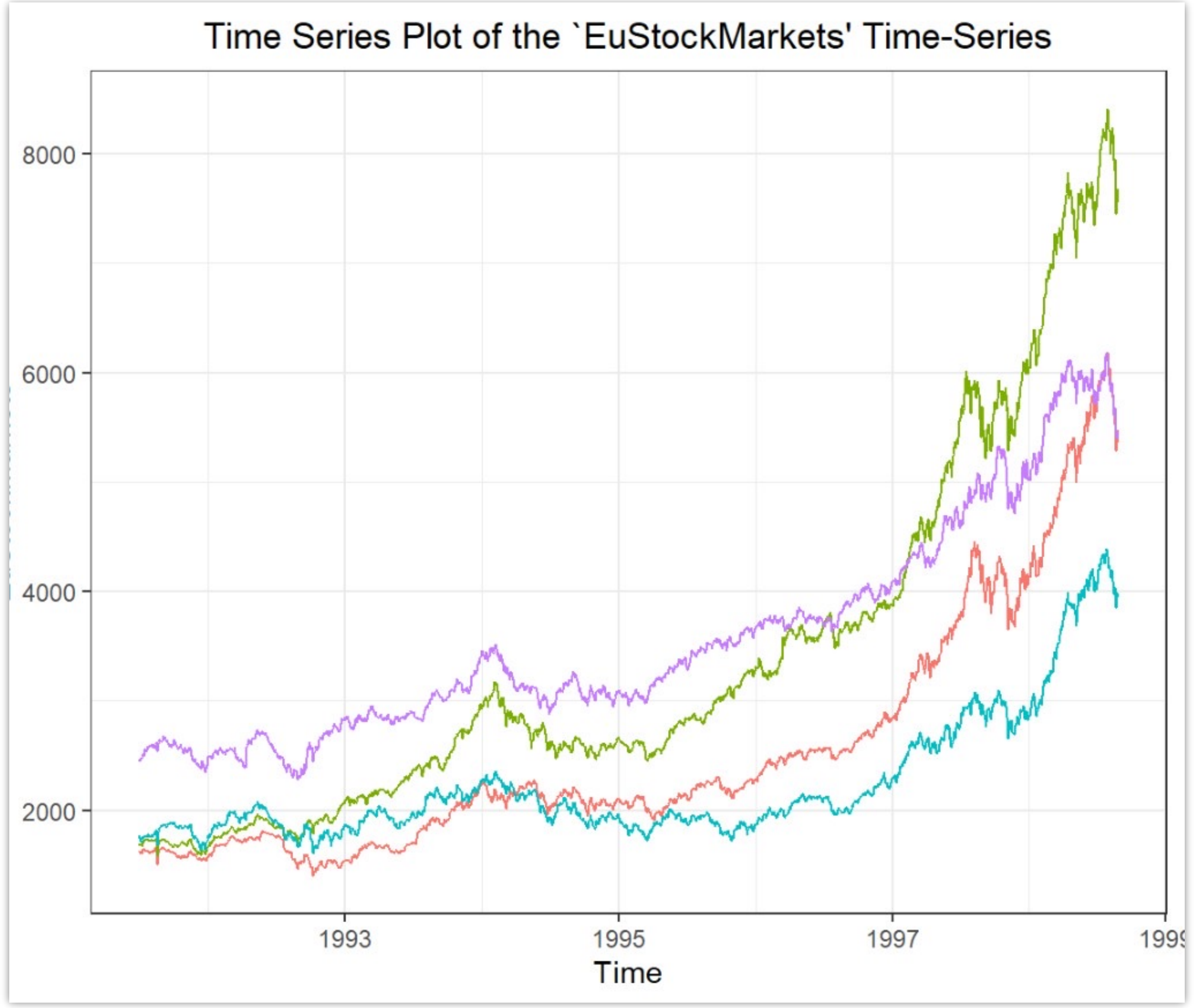
# CONTENT

## Multivariate Time Series (MTS)

- Multivariate time series data is ubiquitous in many systems.
- It contains time series from multiple interlinked variables.
- Each variable generates a time series.



Subway

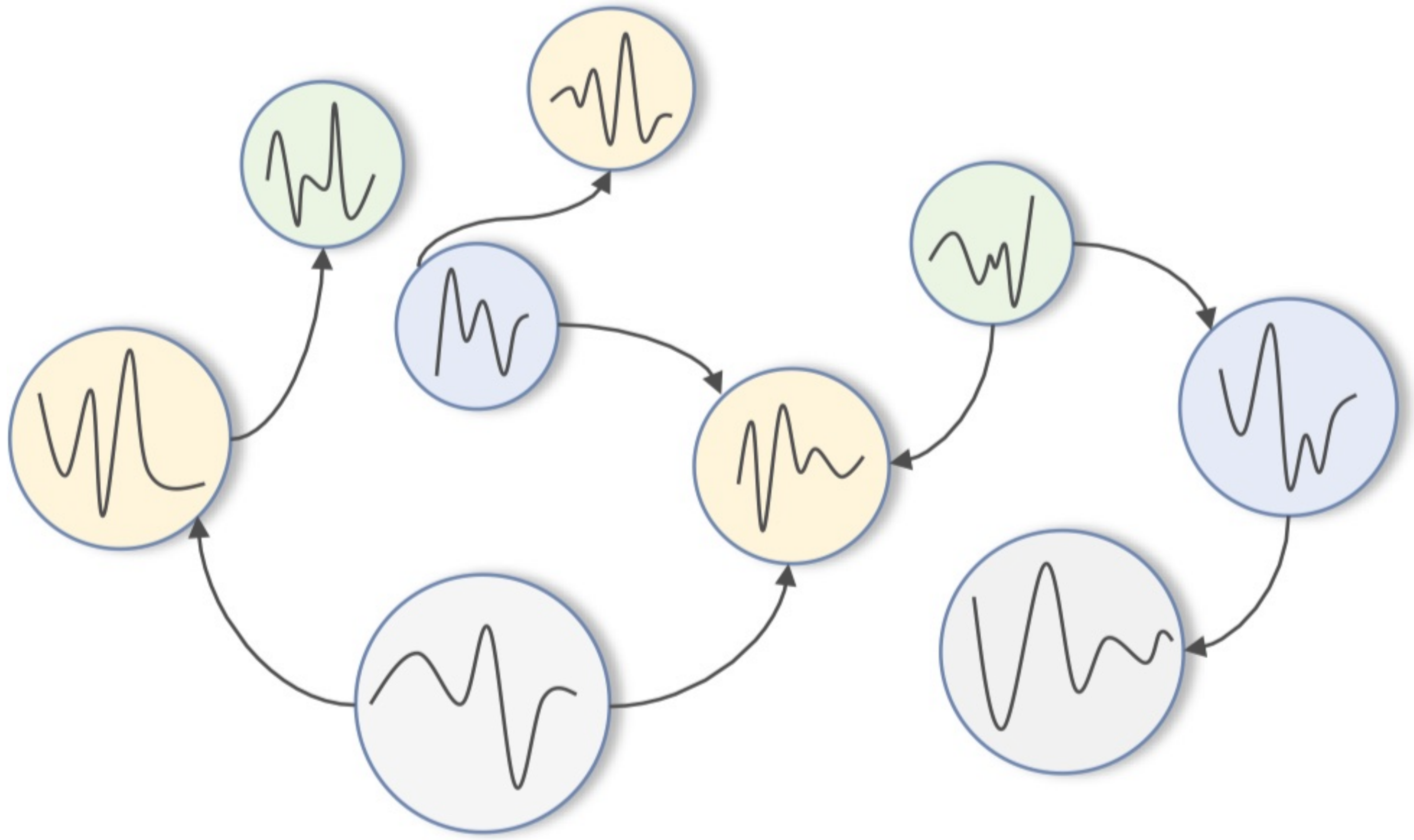



Stock

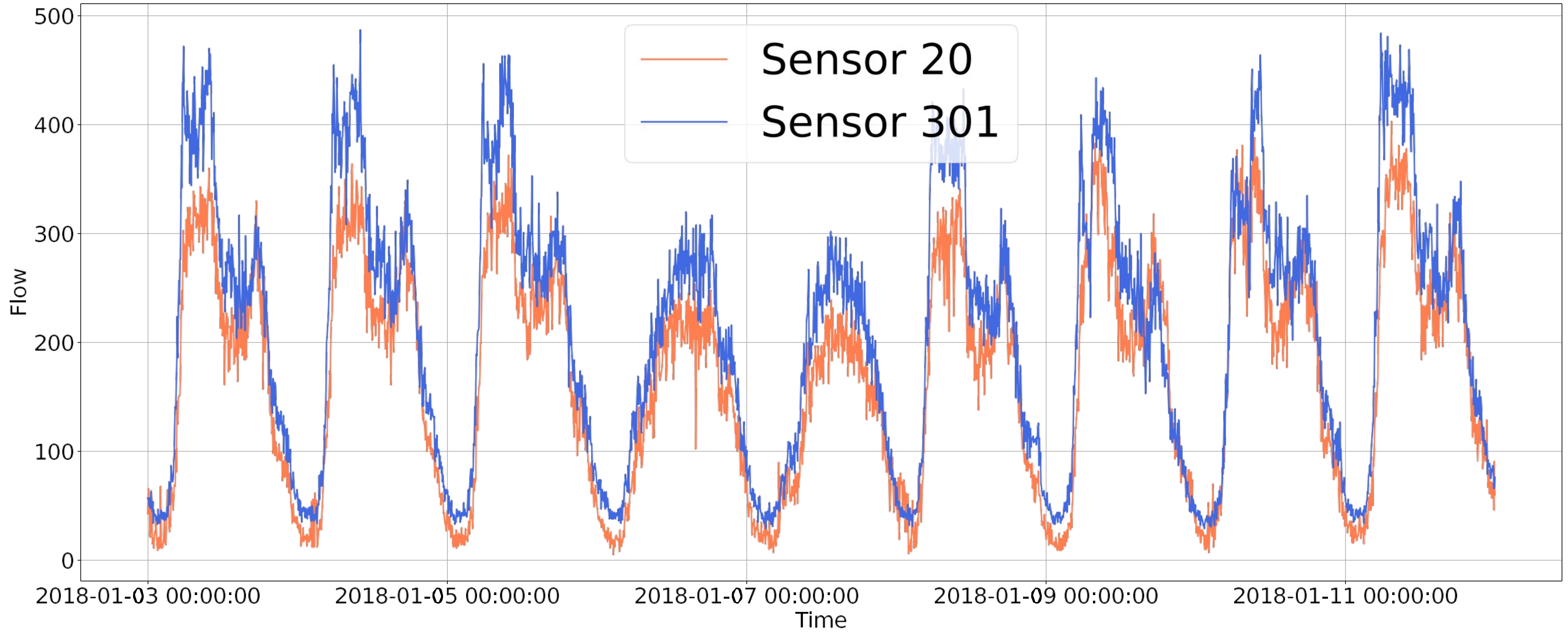
Electricity

# Spatial-Temporal Graph Data

- **Temporal:** Complex temporal patterns, *e.g.*, multiple periodicities.
- **Spatial:** Underlying interdependencies between variables, which is non-Euclidean and is reasonably modeled by the graph structure.



**Spatial-Temporal Graph Data**



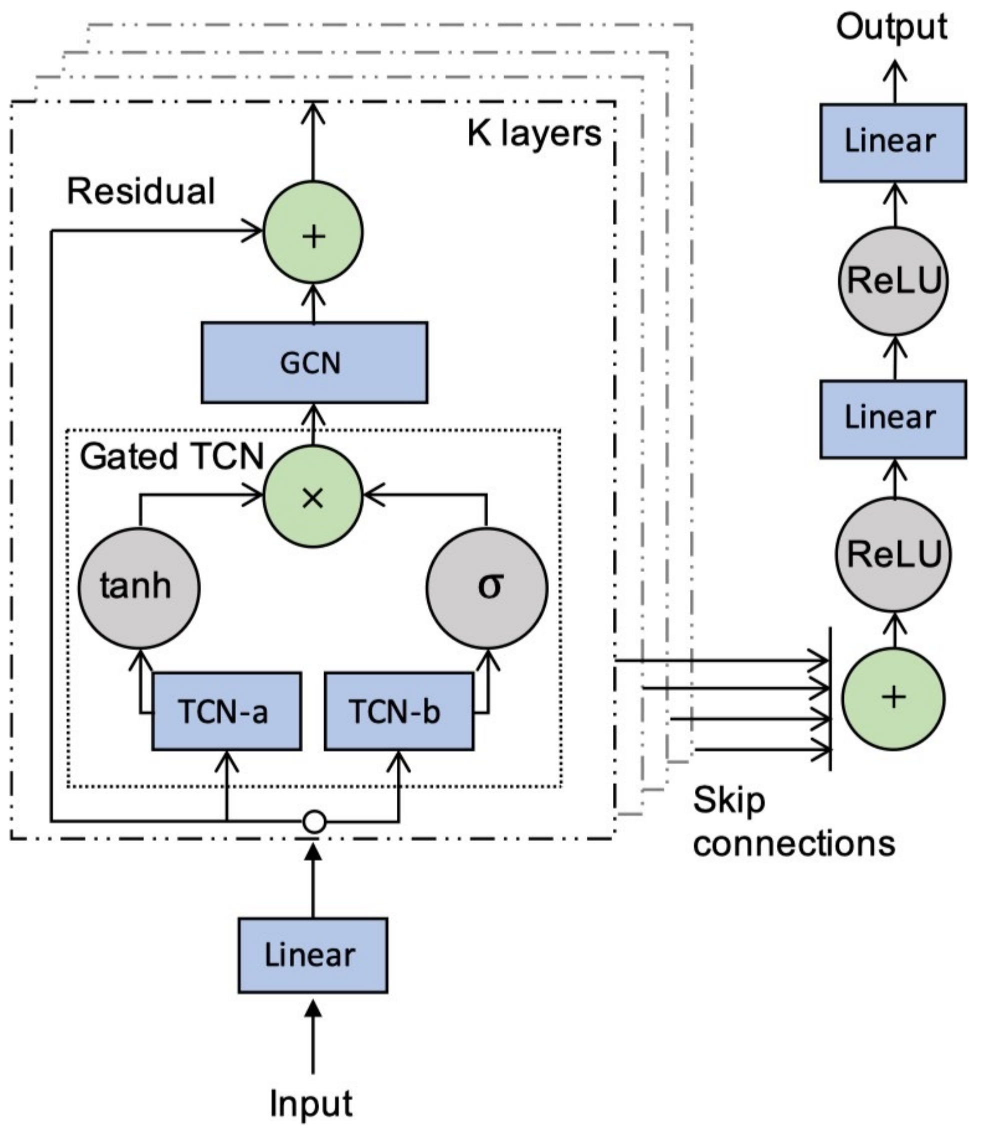

**MTS from Traffic Flow System**

## Spatial-Temporal Graph Neural Networks (STGNNs)

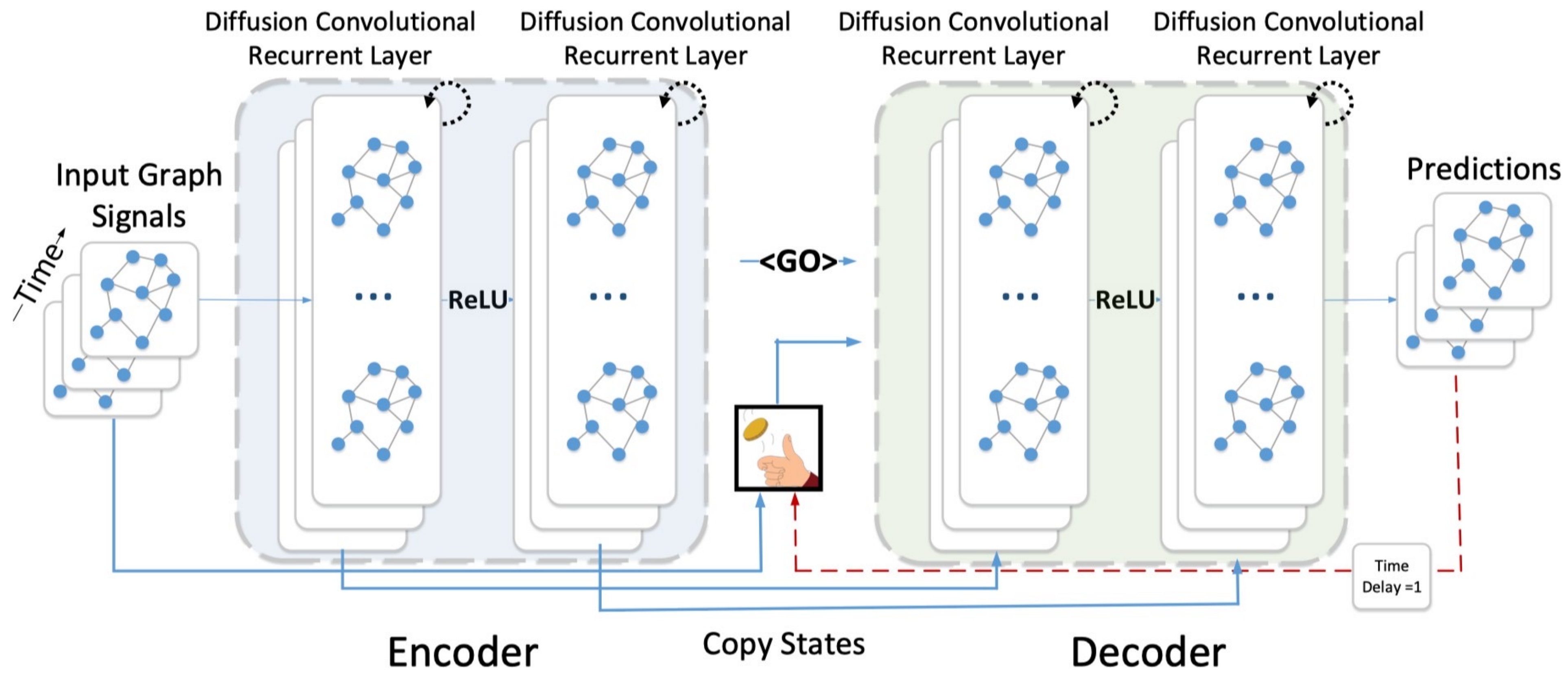
- Combine graph neural networks (spatial) and sequential models (temporal).

## Learning the Graph Structure

- The handcrafted dependency graph between time series is often biased and incorrect, even missing in many cases.




2019 IJCAI Graph WaveNet



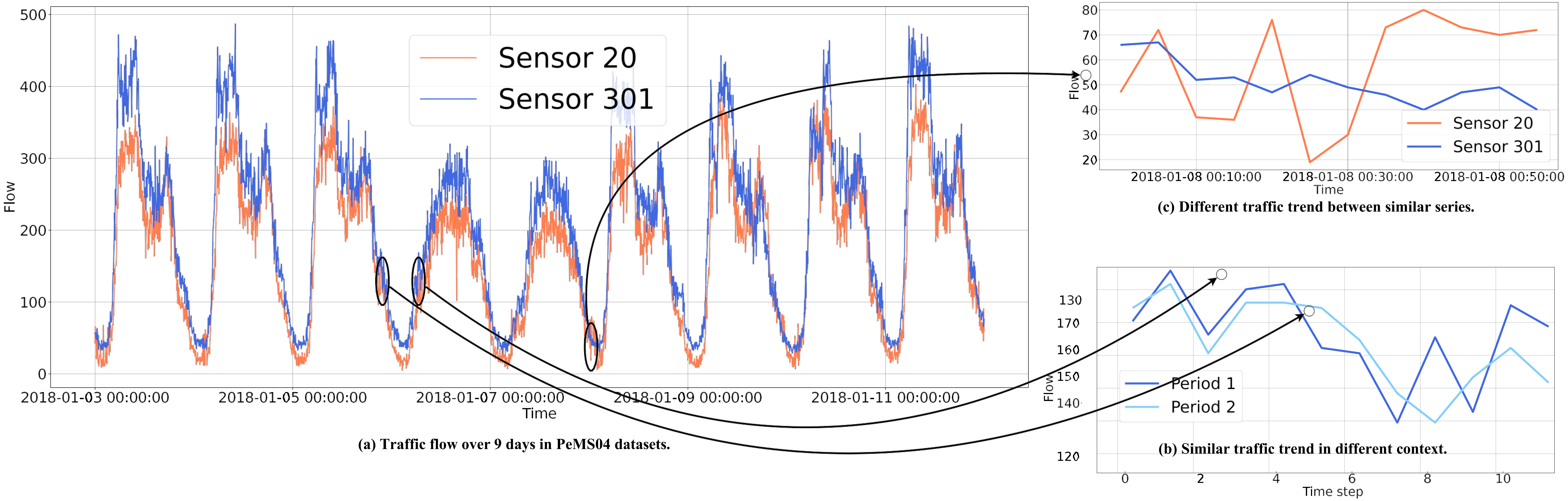

2018 ICLR DCRNN

## Existing STGNNs can not scale to very long-term history

- The computational complexity increases linearly or quadratically with the length and number of the input TS.
- The optimization of the model can also become problematic as the length of the input sequence increases.

## Long-term historical time series are crucial

- Beneficial for distinguishing short-term time series in different contexts.
- Beneficial for resisting noise, facilitating learning robust and accurate dependency graph.




(a) Traffic flow over 9 days in PeMS04 datasets.

(c) Different traffic trend between similar series.

(b) Similar traffic trend in different context.

## Challenges

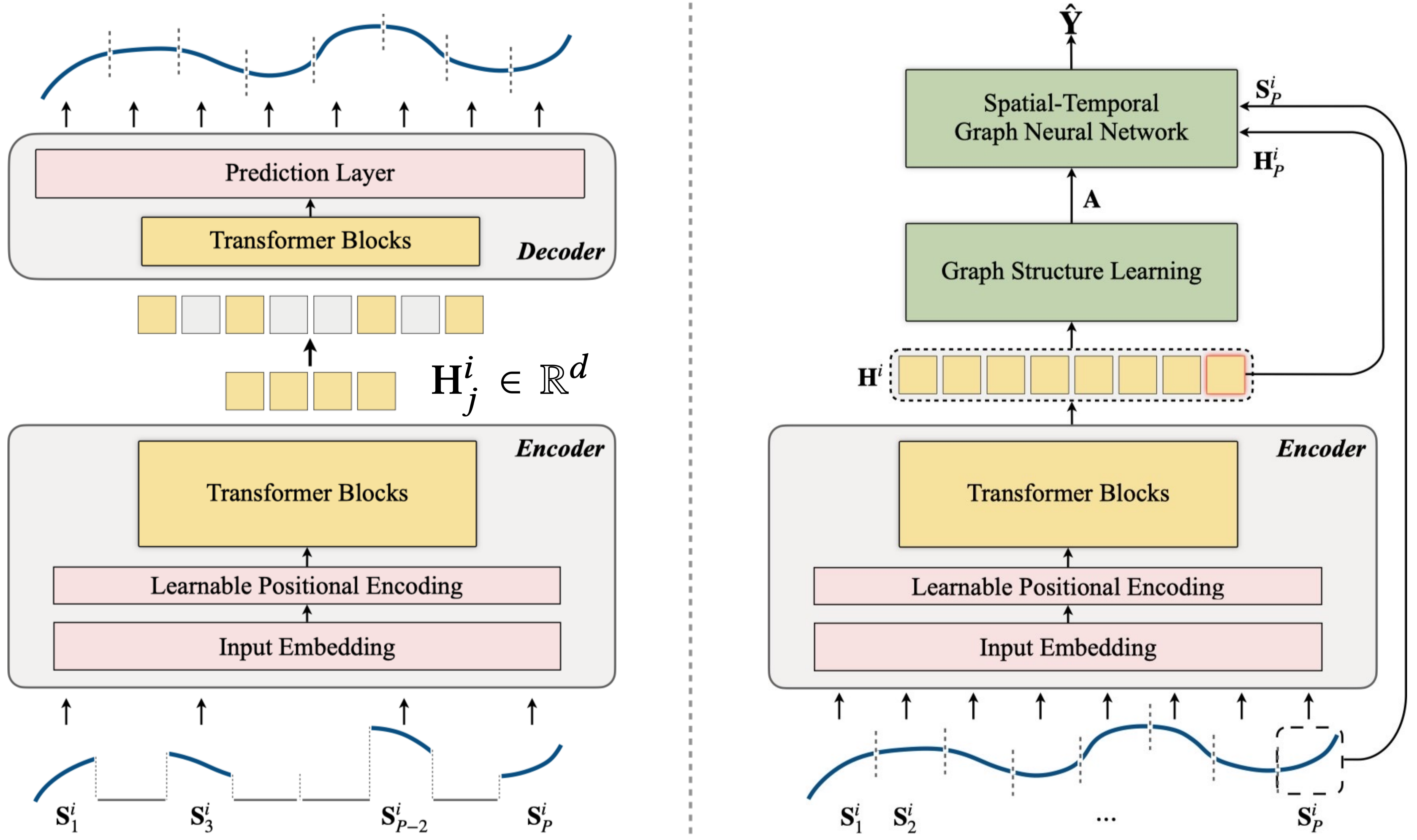
- How to learn from very long-term (*e.g.*, weeks) historical time series to enhance STGNNs?

## STEP Framework

- Instead of directly extending STGNNs to very long-term historical time series,
- we propose a novel framework, in which **ST**GNN is **E**nhanced by a scalable time series **P**re-training model (**STEP**).

# CONTENT

## STEP Framework

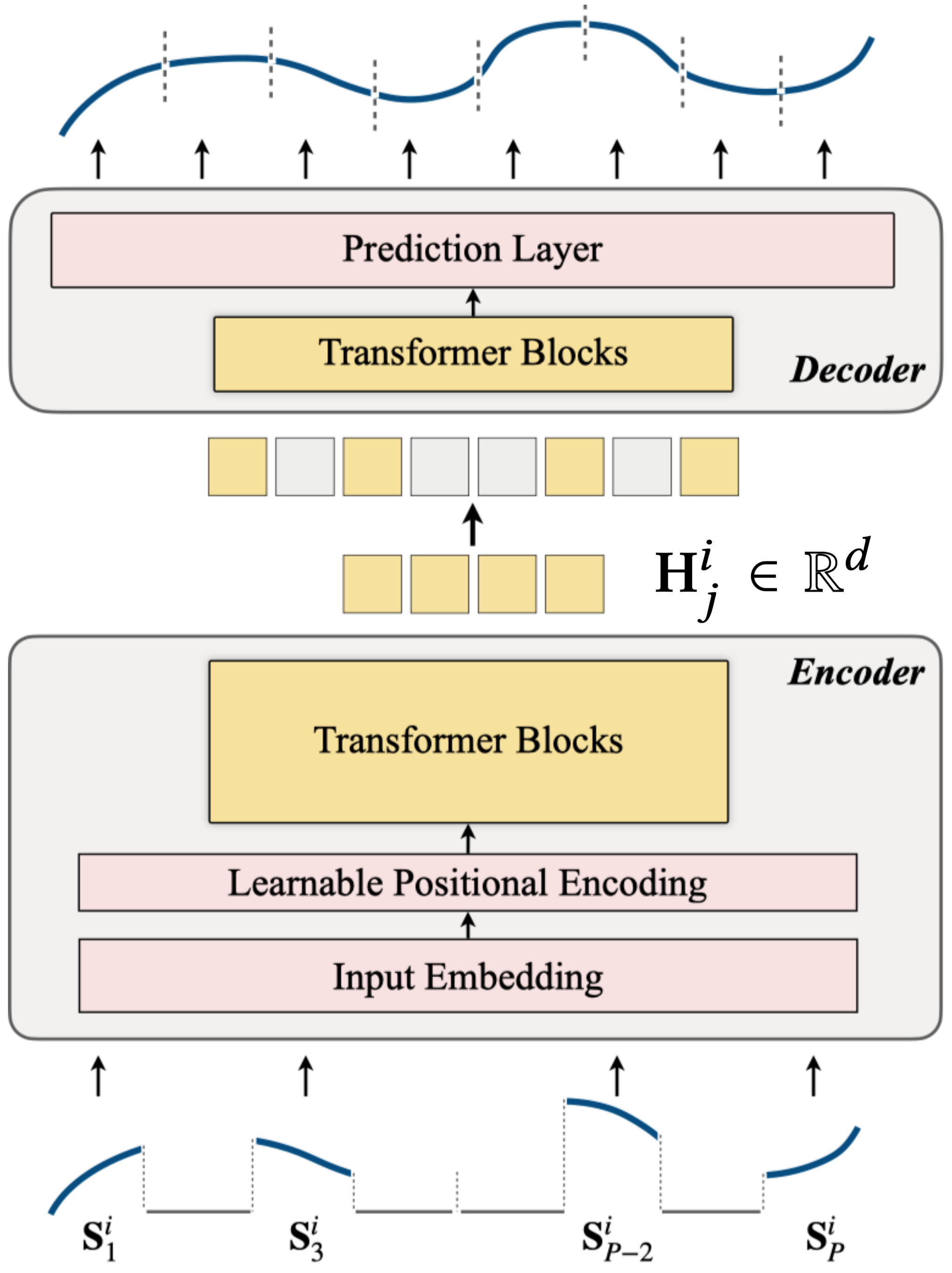



Pre-training Stage

Forecasting Stage

## Key Design Motivations

- **Time series information density is lower.**
  - Isolated data points in time series give less semantic information
  - Masked values in time series can often be trivially predicted by simple interpolation, making the pre-training model only focuses on low-level information.
- **Time series require longer sequences to learn the temporal patterns.**



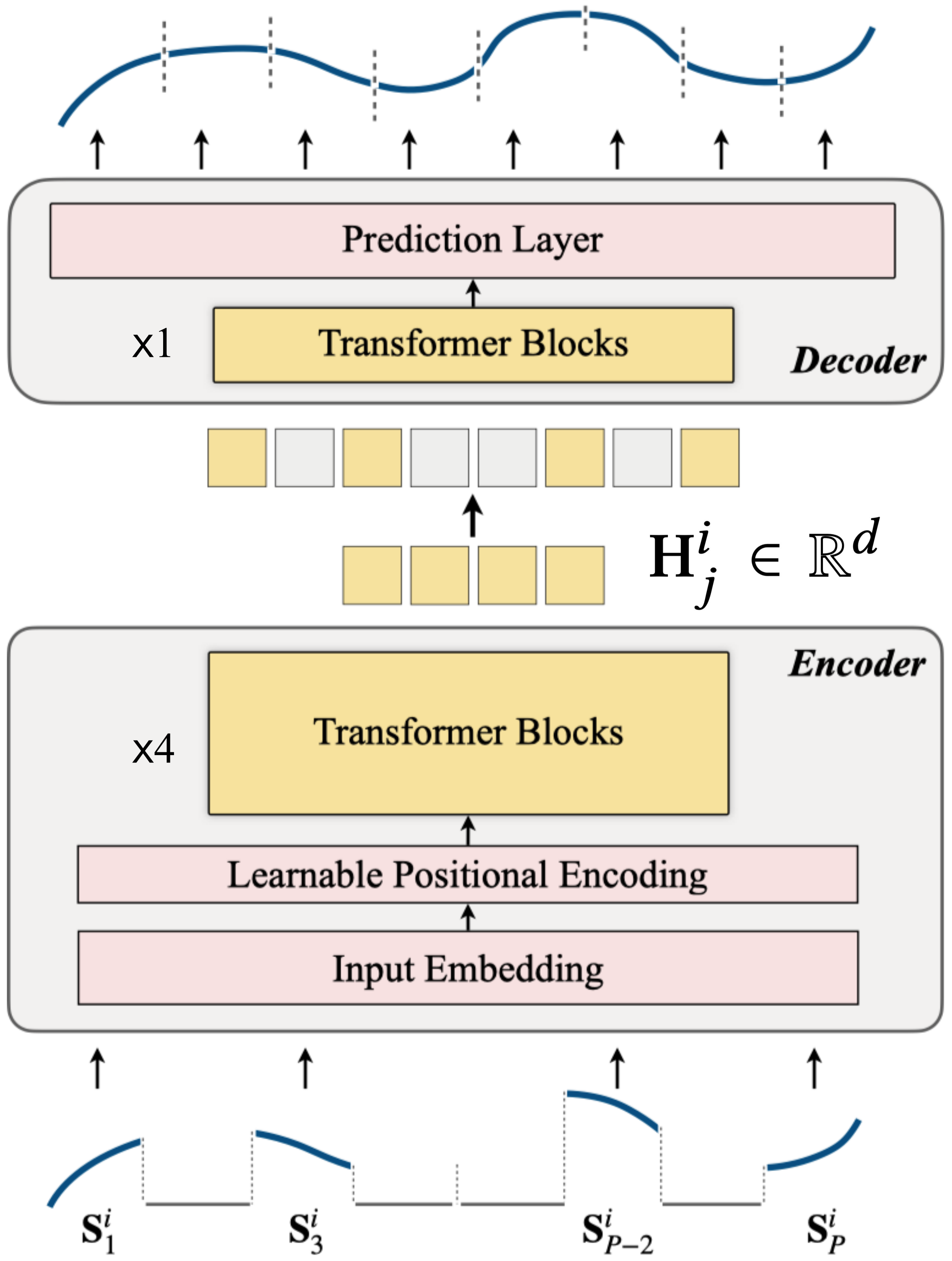

?.?.?

**Effective & High Efficiency**

## Key Design Motivations

- Time series information density is lower.
- Time series require longer sequences to learn the temporal patterns.

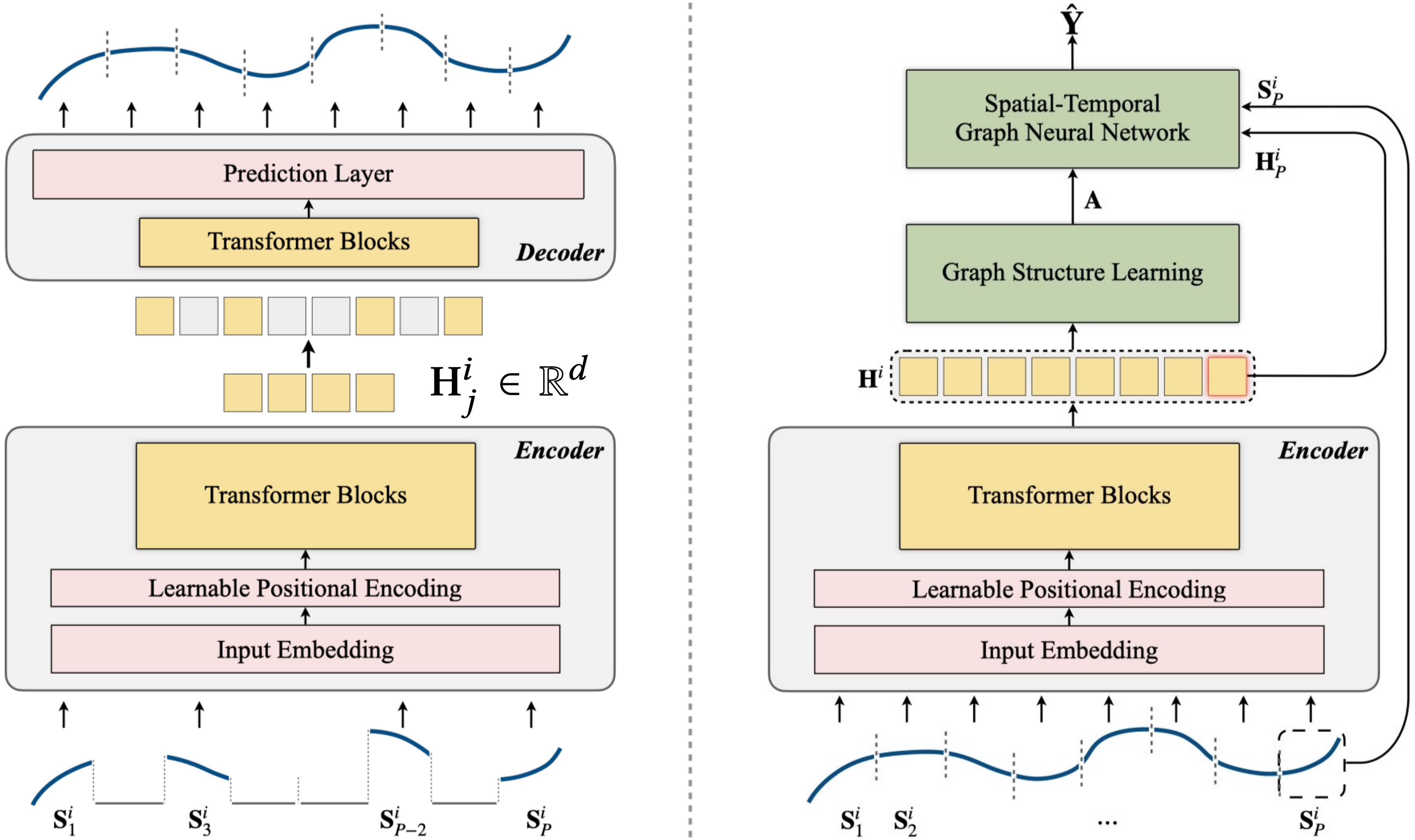
## Model: TSFormer




- **Segment-level representation**
- **High mask ratio**
- **Asymmetrical design**
- **Learnable positional encoding**

**Effective & High Efficiency**

## STEP Framework

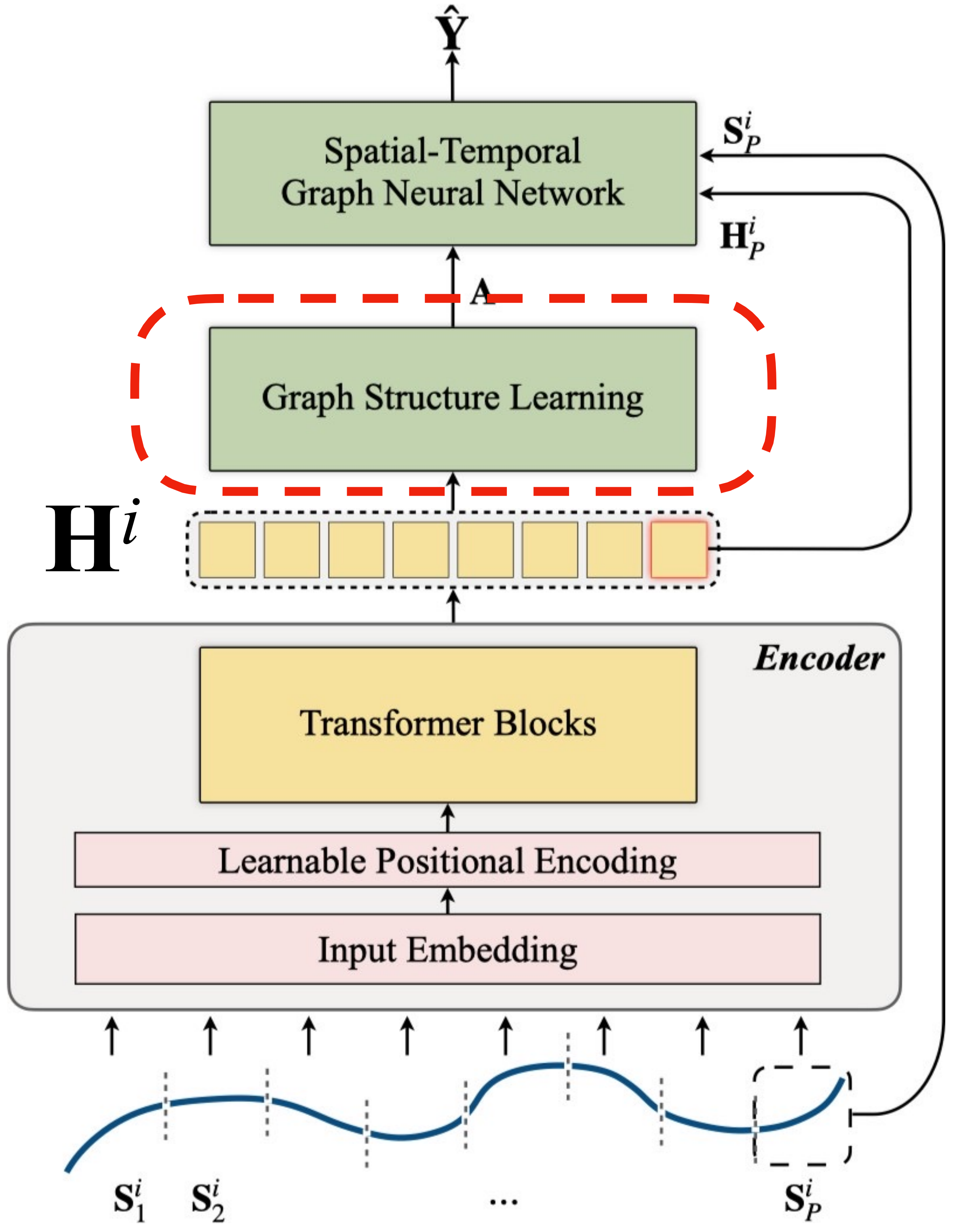



Pre-training Stage

Forecasting Stage

## Enhancing the STGNNs

- Graph structure learning.

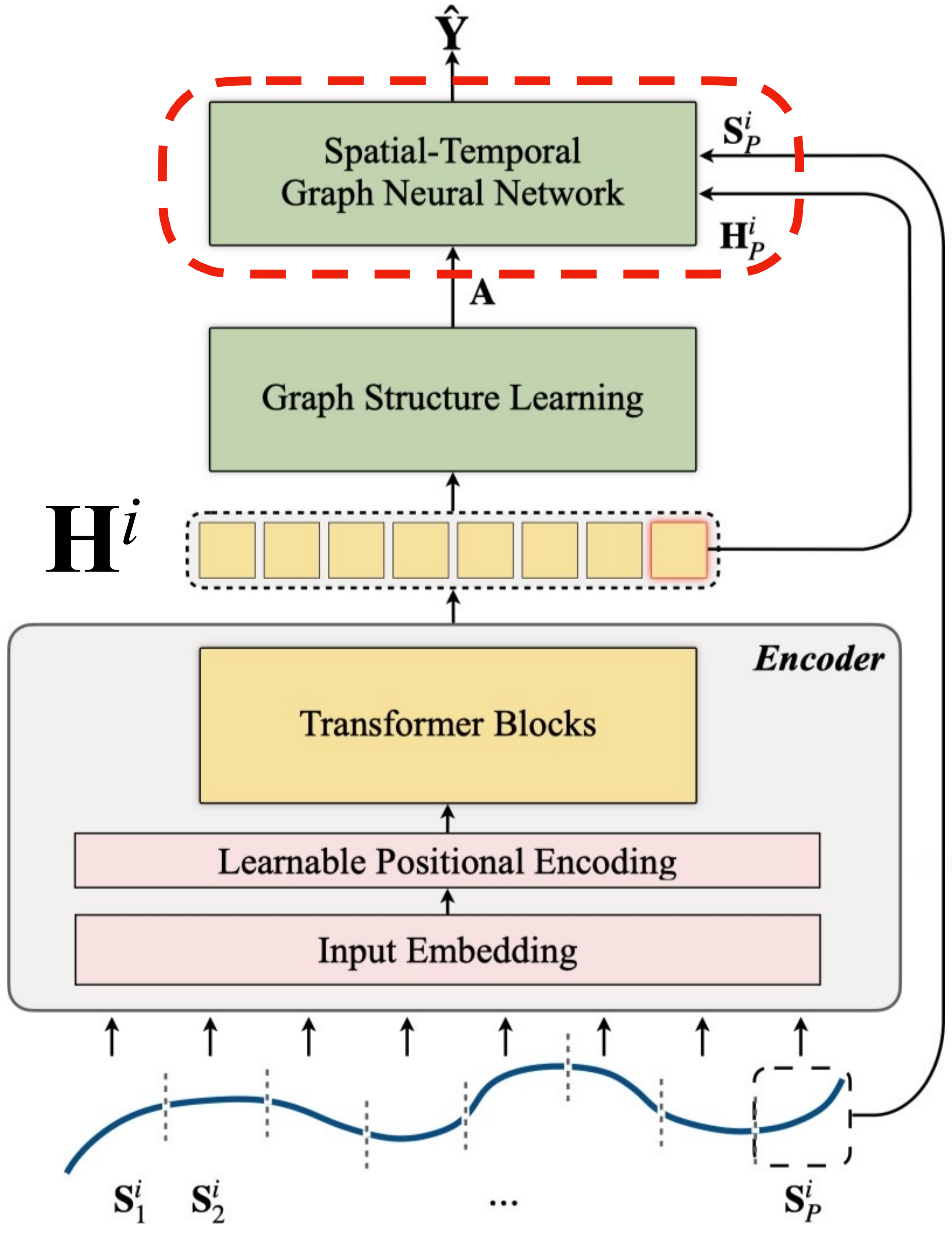



- **Calculate Bernoulli parameters based on $\mathbf{H}^i$**
- **Gumbel-Softmax: differentiable sampling**
- **Guide the Training of Graph Structure with a $k$NN graph $\mathbf{A}^a$**

**Robust Graph Structure Learning**

## Enhancing the STGNNs

- Downstream STGNNs.




- **Graph WaveNet[1] as an example backend**
- **Add context information $\mathbf{S}^i_P$**

$$\mathbf{H}_{final} = \mathrm{SP}(\mathbf{H}_P) + \mathbf{H}_{gw}$$

**Provide Long-Term Contextual Information**

- **Joint learning with the graph structure**

$$\mathcal{L} = \mathcal{L}_{regression} + \lambda \mathcal{L}_{graph}$$

[1] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In IJCAI.

CONTENT

1 Background

2 STEP

3 Experiments

4 Conclusion

## Baselines

- HA
- VAR
- SVR
- FC-LSTM
- DCRNN
- Graph WaveNet
- ASTGNN
- STSGCN
- GMAN
- MTGNN
- GTS

## Metrics

- MAE
- RMSE
- MAPE

## Hardware

- NVIDIA RTX3090

## Datasets

**Table 1: Statistics of datasets.**

| Dataset | # Samples | # Node | Sample Rate | Time Span |
|---|---|---|---|---|
| METR-LA | 34727 | 207 | 5mins | 4 months |
| PEMS-BAY | 52116 | 325 | 5mins | 6 months |
| PEMS04 | 16969 | 307 | 5mins | 2 months |

# 3 Experiments | Main Results

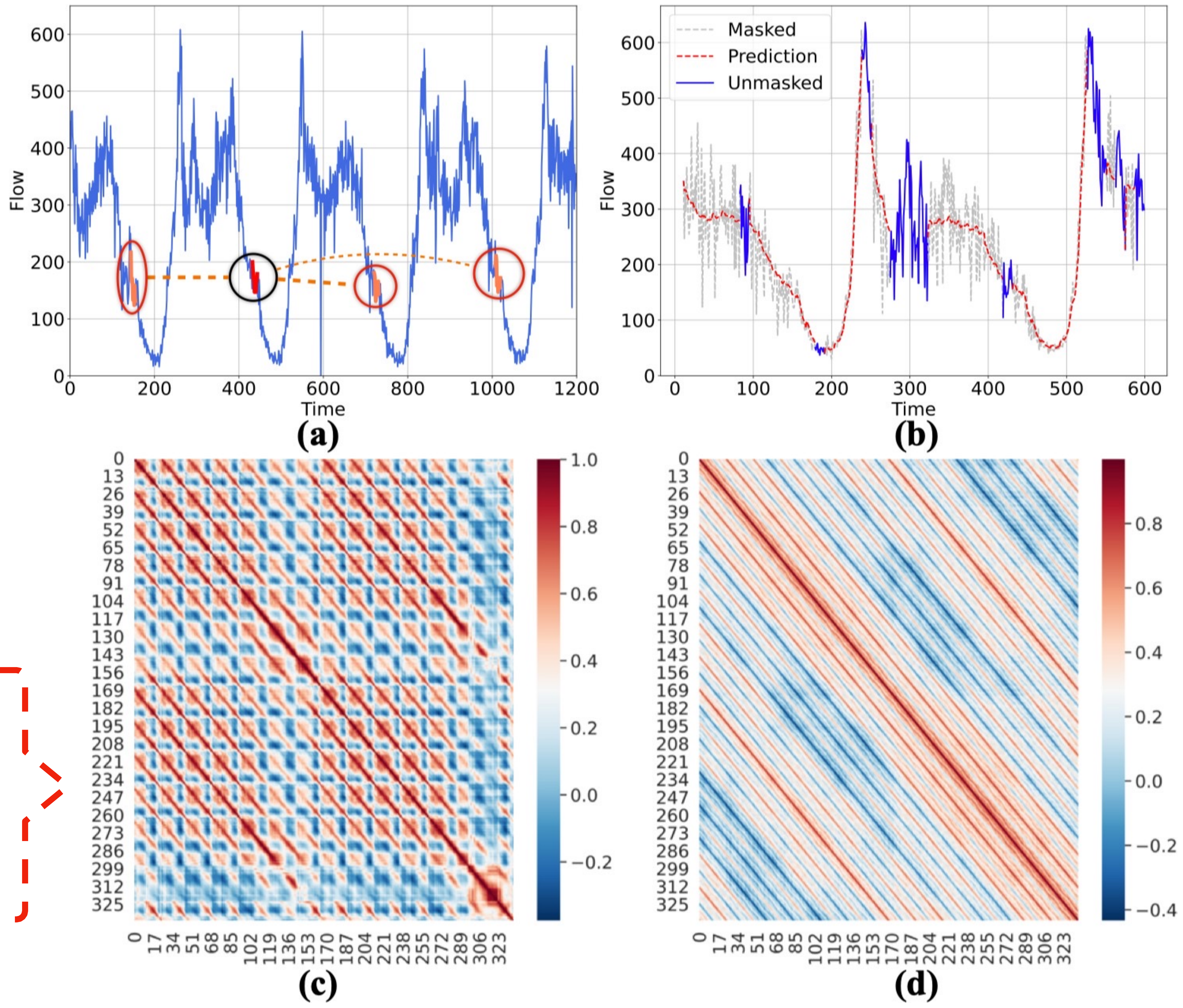
Table 2: Multivariate time series forecasting on the METR-LA, PEMS-BAY, and PEMS04 datasets. Numbers marked with * indicate that the improvement is statistically significant compared with the best baseline (t-test with p-value $< 0.05$).

| Datasets | Methods | Horizon 3 | | | Horizon 6 | | | Horizon 12 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| METR-LA | HA | 4.79 | 10.00 | 11.70% | 5.47 | 11.45 | 13.50% | 6.99 | 13.89 | 17.54% |
| | VAR | 4.42 | 7.80 | 13.00% | 5.41 | 9.13 | 12.70% | 6.52 | 10.11 | 15.80% |
| | SVR | 3.39 | 8.45 | 9.30% | 5.05 | 10.87 | 12.10% | 6.72 | 13.76 | 16.70% |
| | FC-LSTM | 3.44 | 6.30 | 9.60% | 3.77 | 7.23 | 10.09% | 4.37 | 8.69 | 14.00% |
| | DCRNN | 2.77 | 5.38 | 7.30% | 3.15 | 6.45 | 8.80% | 3.60 | 7.60 | 10.50% |
| | STGCN | 2.88 | 5.74 | 7.62% | 3.47 | 7.24 | 9.57% | 4.59 | 9.40 | 12.70% |
| | Graph WaveNet | 2.69 | 5.15 | 6.90% | 3.07 | 6.22 | 8.37% | 3.53 | 7.37 | 10.01% |
| | ASTGCN | 4.86 | 9.27 | 9.21% | 5.43 | 10.61 | 10.13% | 6.51 | 12.52 | 11.64% |
| | STSGCN | 3.31 | 7.62 | 8.06% | 4.13 | 9.77 | 10.29% | 5.06 | 11.66 | 12.91% |
| | GMAN | 2.80 | 5.55 | 7.41% | 3.12 | 6.49 | 8.73% | 3.44 | 7.35 | 10.07% |
| | MTGNN | 2.69 | 5.18 | 6.88% | 3.05 | 6.17 | 8.19% | 3.49 | 7.23 | 9.87% |
| | GTS | 2.67 | 5.27 | 7.21% | 3.04 | 6.25 | 8.41% | 3.46 | 7.31 | 9.98% |
| | STEP | **2.61*** | **4.98*** | **6.60%*** | **2.96*** | **5.97*** | **7.96%*** | **3.37*** | **6.99*** | **9.61%*** |
| PEMS-BAY | HA | 1.89 | 4.30 | 4.16% | 2.50 | 5.82 | 5.62% | 3.31 | 7.54 | 7.65% |
| | VAR | 1.74 | 3.16 | 3.60% | 2.32 | 4.25 | 5.00% | 2.93 | 5.44 | 6.50% |
| | SVR | 1.85 | 3.59 | 3.80% | 2.48 | 5.18 | 5.50% | 3.28 | 7.08 | 8.00% |
| | FC-LSTM | 2.05 | 4.19 | 4.80% | 2.20 | 4.55 | 5.20% | 2.37 | 4.96 | 5.70% |
| | DCRNN | 1.38 | 2.95 | 2.90% | 1.74 | 3.97 | 3.90% | 2.07 | 4.74 | 4.90% |
| | STGCN | 1.36 | 2.96 | 2.90% | 1.81 | 4.27 | 4.17% | 2.49 | 5.69 | 5.79% |
| | Graph WaveNet | 1.30 | 2.74 | 2.73% | 1.63 | 3.70 | 3.67% | 1.95 | 4.52 | 4.63% |
| | ASTGCN | 1.52 | 3.13 | 3.22% | 2.01 | 4.27 | 4.48% | 2.61 | 5.42 | 6.00% |
| | STSGCN | 1.44 | 3.01 | 3.04% | 1.83 | 4.18 | 4.17% | 2.26 | 5.21 | 5.40% |
| | GMAN | 1.34 | 2.91 | 2.86% | 1.63 | 3.76 | 3.68% | 1.86 | 4.32 | 4.37% |
| | MTGNN | 1.32 | 2.79 | 2.77% | 1.65 | 3.74 | 3.69% | 1.94 | 4.49 | 4.53% |
| | GTS | 1.34 | 2.83 | 2.82% | 1.66 | 3.78 | 3.77% | 1.95 | 4.43 | 4.58% |
| | STEP | **1.26*** | **2.73*** | **2.59%*** | **1.55*** | **3.58*** | **3.43%*** | **1.79*** | **4.20*** | **4.18%*** |
| PEMS04 | HA | 28.92 | 42.69 | 20.31% | 33.73 | 49.37 | 24.01% | 46.97 | 67.43 | 35.11% |
| | VAR | 21.94 | 34.30 | 16.42% | 23.72 | 36.58 | 18.02% | 26.76 | 40.28 | 20.94% |
| | SVR | 22.52 | 35.30 | 14.71% | 27.63 | 42.23 | 18.29% | 37.86 | 56.01 | 26.72% |
| | FC-LSTM | 21.42 | 33.37 | 15.32% | 25.83 | 39.10 | 20.35% | 36.41 | 50.73 | 29.92% |
| | DCRNN | 20.34 | 31.94 | 13.65% | 23.21 | 36.15 | 15.70% | 29.24 | 44.81 | 20.09% |
| | STGCN | 19.35 | 30.76 | 12.81% | 21.85 | 34.43 | 14.13% | 26.97 | 41.11 | 16.84% |
| | Graph WaveNet | 18.15 | 29.24 | 12.27% | 19.12 | 30.62 | 13.28% | 20.69 | 33.02 | 14.11% |
| | ASTGCN | 20.15 | 31.43 | 14.03% | 22.09 | 34.34 | 15.47% | 26.03 | 40.02 | 19.17% |
| | STSGCN | 19.41 | 30.69 | 12.82% | 21.83 | 34.33 | 14.54% | 26.27 | 40.11 | 14.71% |
| | GMAN | 18.28 | 29.32 | 12.35% | 18.75 | 30.77 | 12.96% | 19.95 | **30.21** | 12.97% |
| | MTGNN | 18.22 | 30.13 | 12.47% | 19.27 | 32.21 | 13.09% | 20.93 | 34.49 | 14.02% |
| | GTS | 18.97 | 29.83 | 13.06% | 19.29 | 30.85 | 13.92% | 21.04 | 34.81 | 14.94% |
| | STEP | **17.34*** | **28.44*** | **11.57%*** | **18.12*** | **29.81*** | **12.00%*** | **19.27*** | 31.33 | **12.78%*** |

## Inspecting the TSFormer

Temporal periodicity

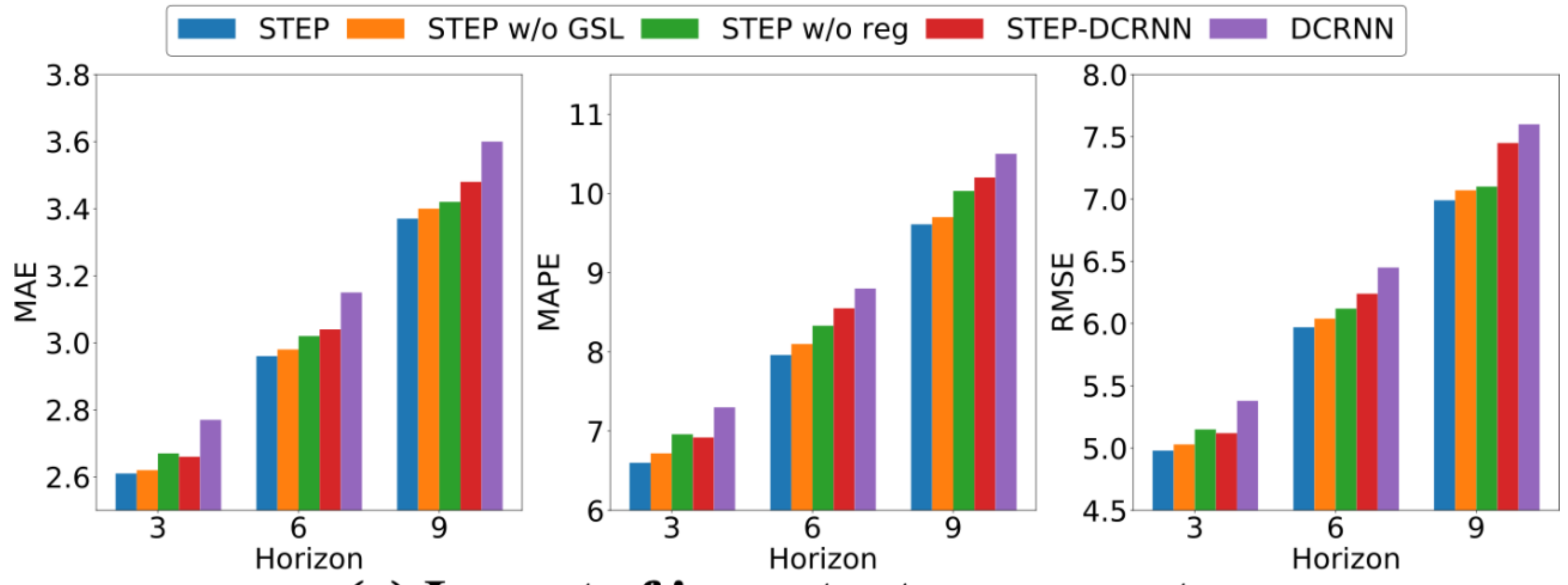
Reconstruction.



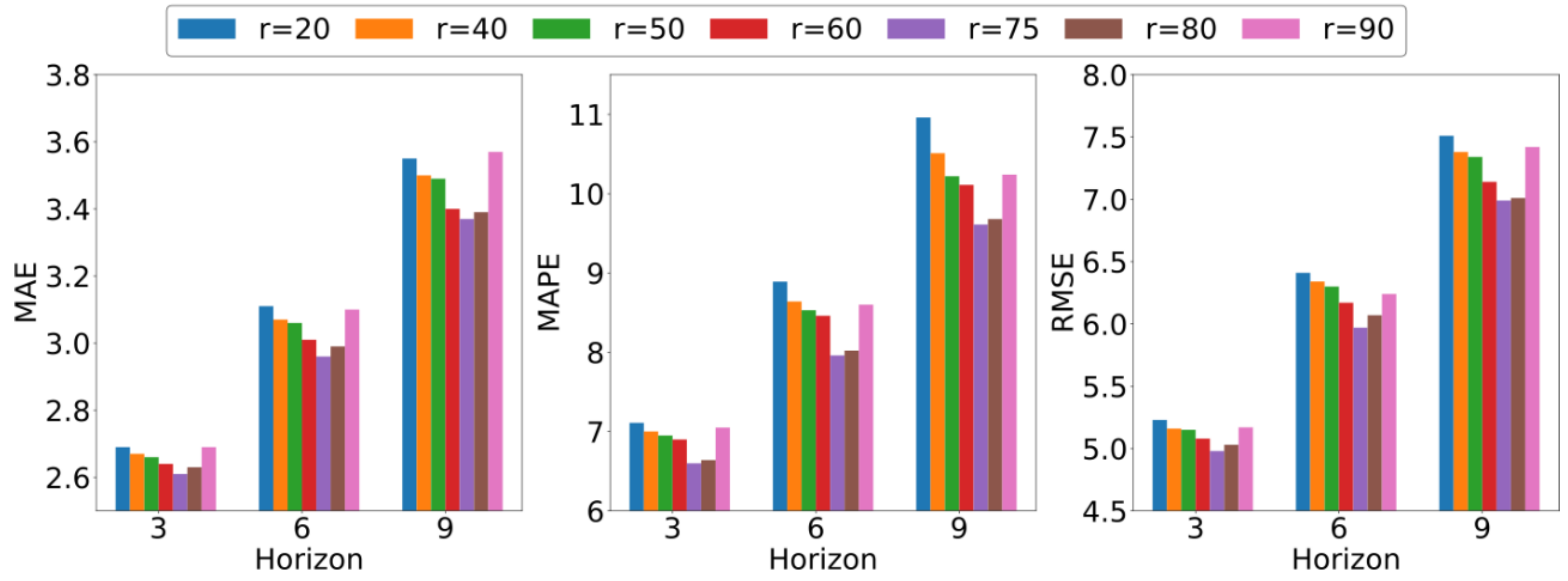

Similarity of latent representations among different patches.

Similarity of positional embeddings among different patches.

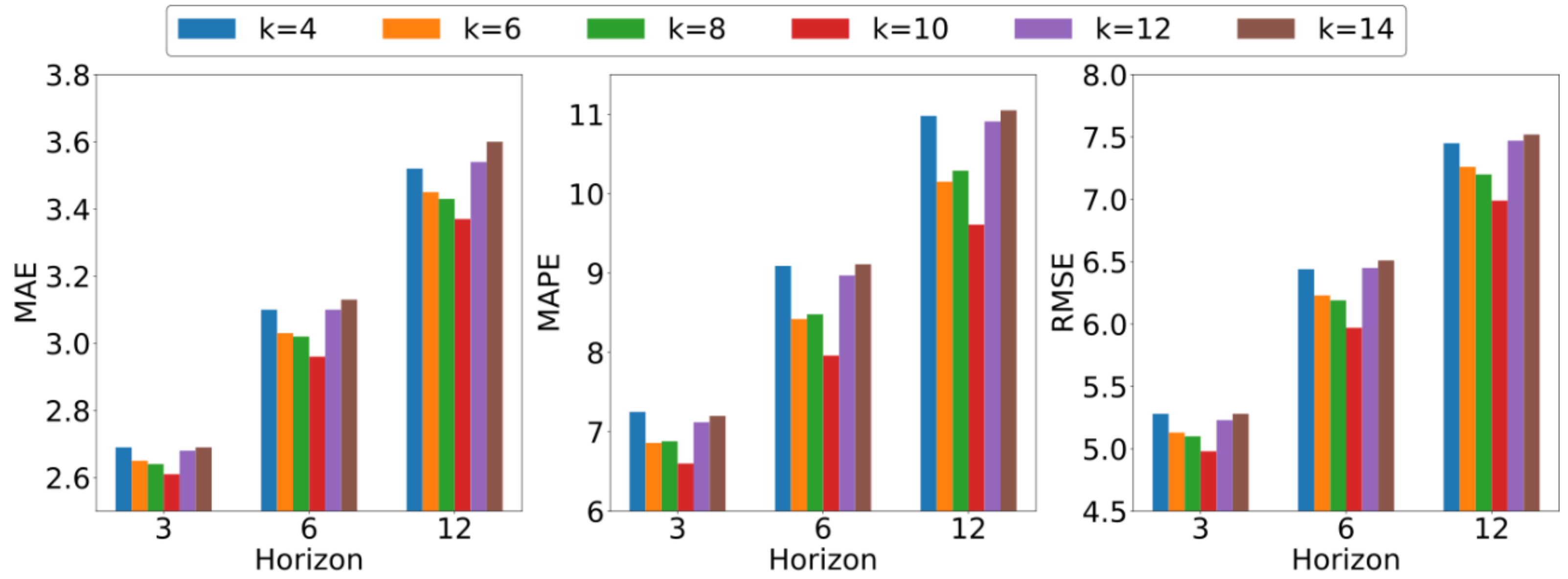
# 3 Experiments Ablation Study & Hyper-parameter Study




(a) Impact of important components.




(b) Impact of masking ratio $r$.




(c) Impact of $k$ of $k$NN graph.

**Figure 4: Ablation study and hyper-parameter study.**

## Ablation Study

- Graph structure learning module consistently plays a positive role.
- Segment-level representation plays a vital role.
- Long sequence representations of TSFormer is superior in improving the graph quality.
- STEP is a general framework.

## Hyper-parameter Study

- Best mask ratio: 75%
- Best $k$ of $k$NN graph: 10

CONTENT

## Conclusions

- Existing STGNNs can be improved by introducing more information from very long-term historical time series
- We design an efficient and effective pre-training model for time series, which generates segment-level representations and can be designed based on Transformer blocks and masked autoencoding strategy.

## Future Work

- Time series recovery using TSFormer
- Further improve the efficiency and effectiveness of TSFormer
- Provide contextual information more flexible
- …

# Thank You!

# Q & A

## More Materials:

Code of STEP: https://github.com/zezhishao/STEP

Fair comparison of all STGNNs: https://github.com/zezhishao/BasicTS

## Visualizations

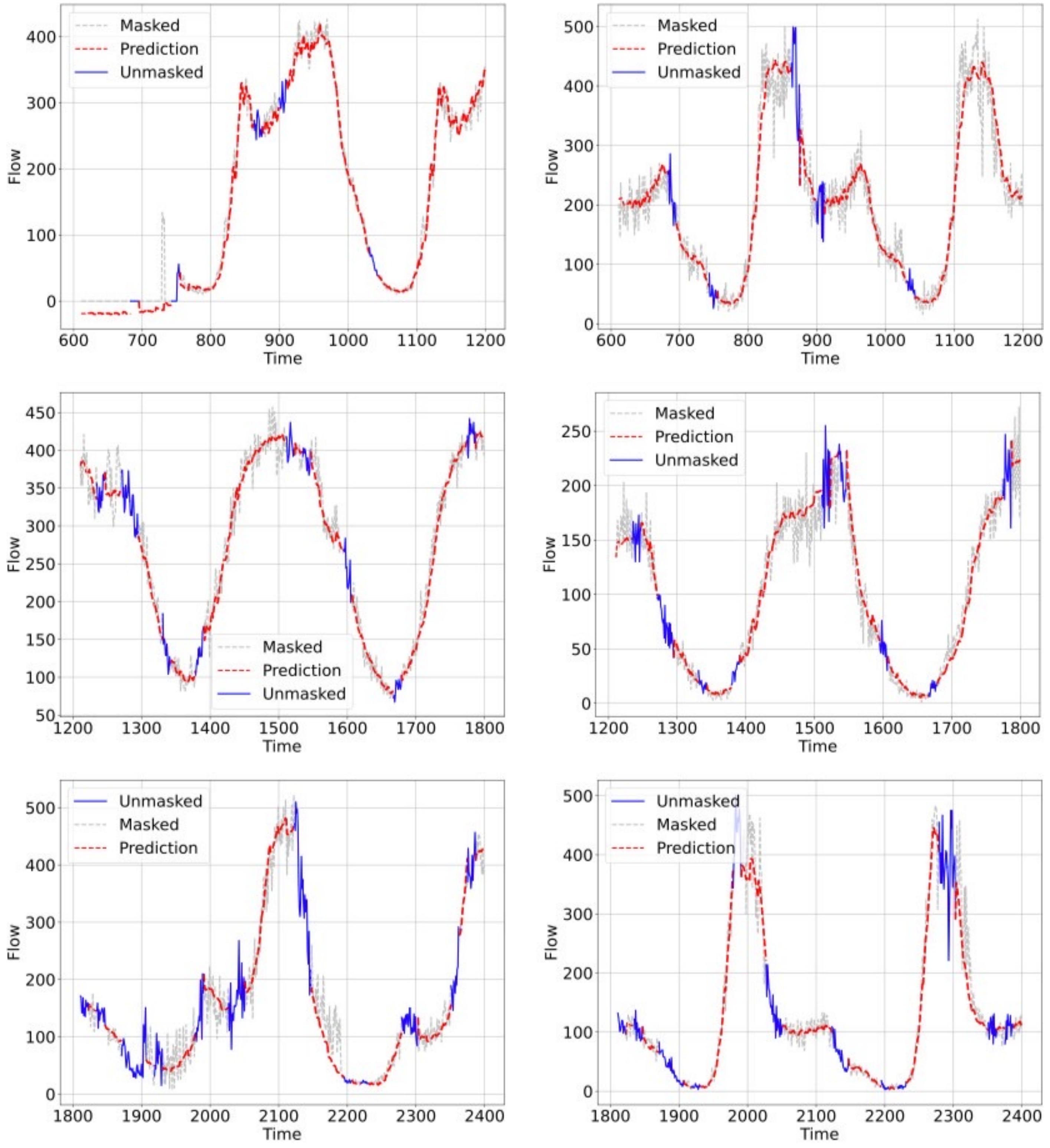



Figure 7: Reconstruction visualizations.

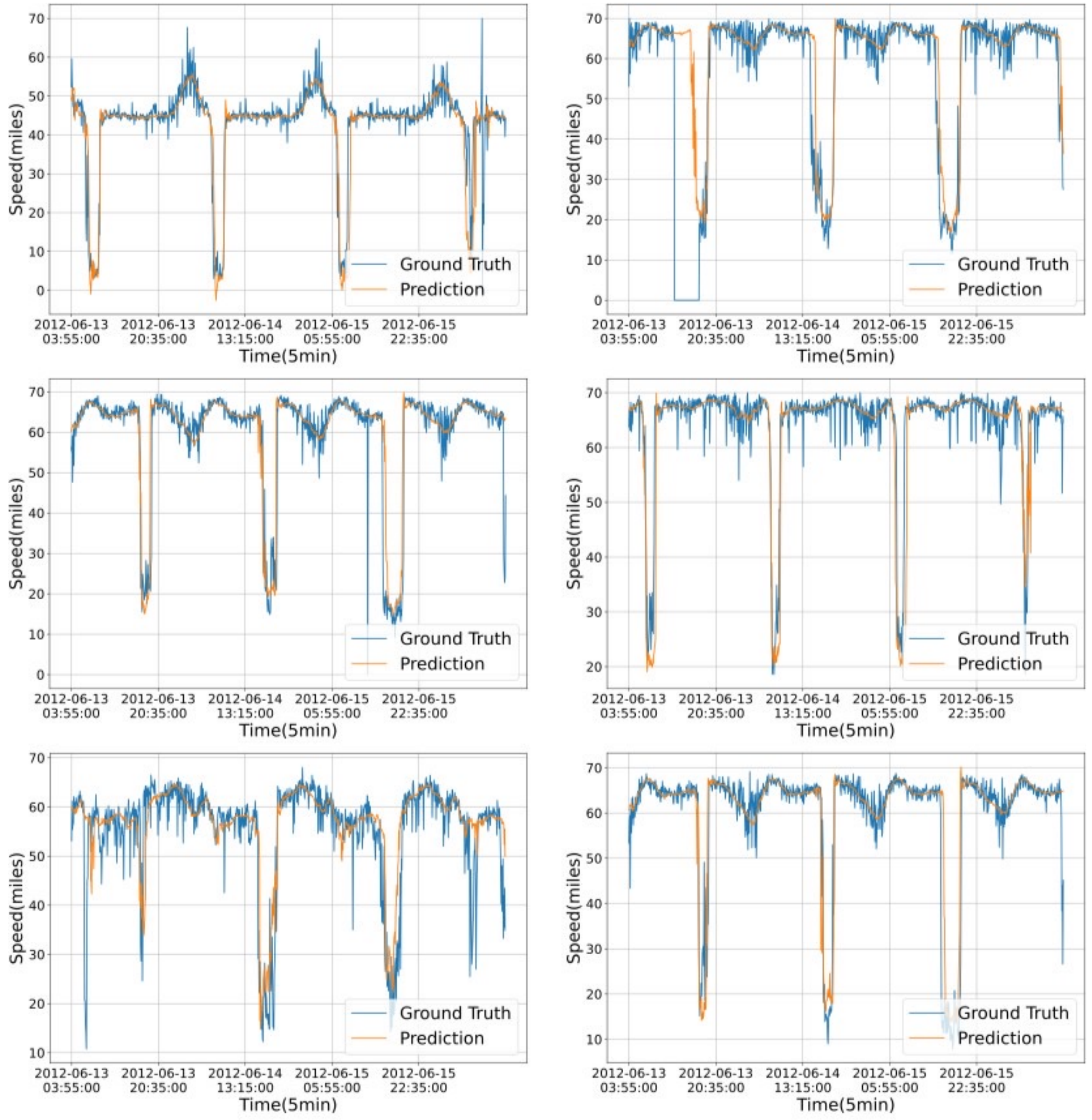



Figure 8: Forecasting visualizations.

# Efficiency & Speed

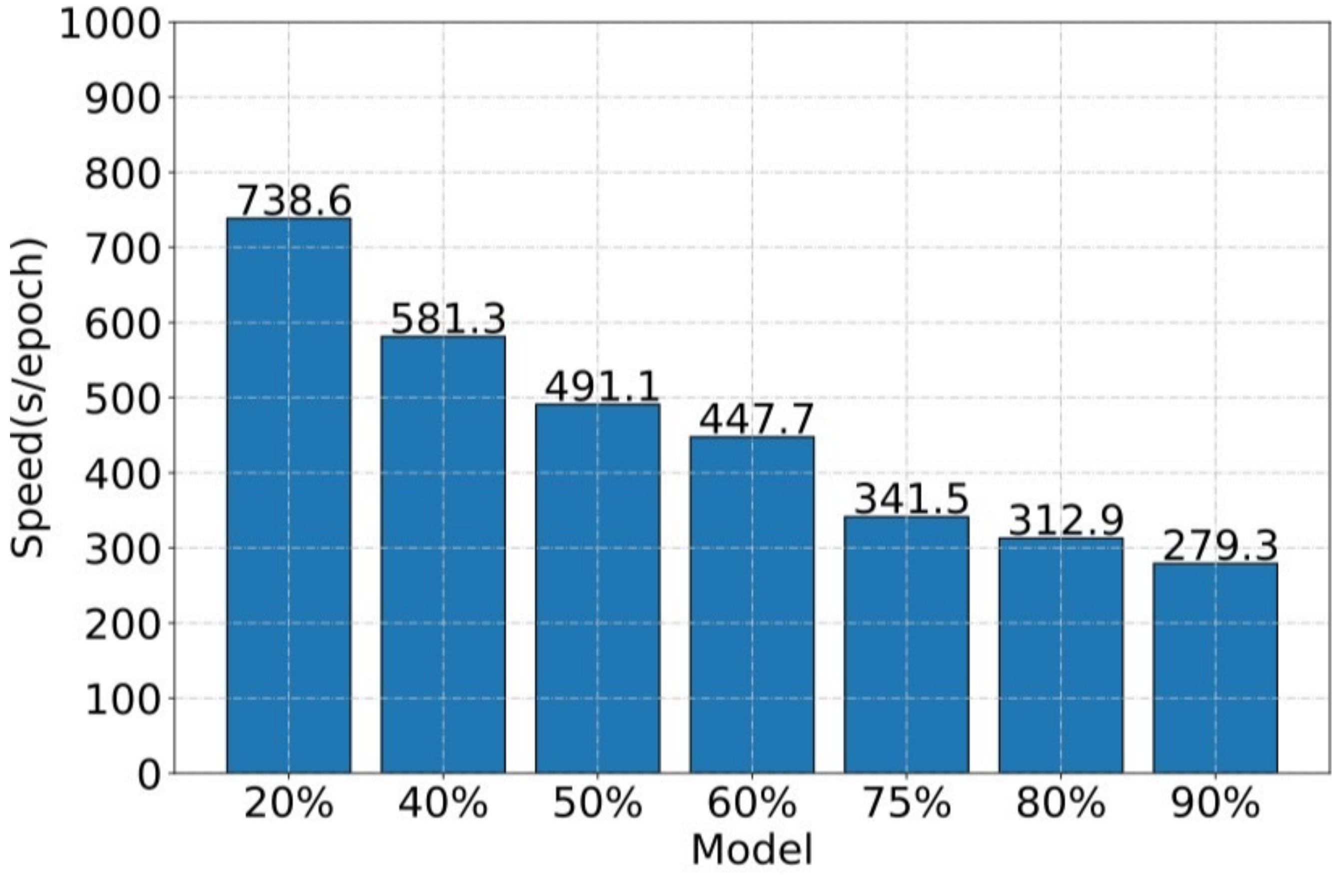



**Figure 5: Training speed of different masking ratio $r$.**

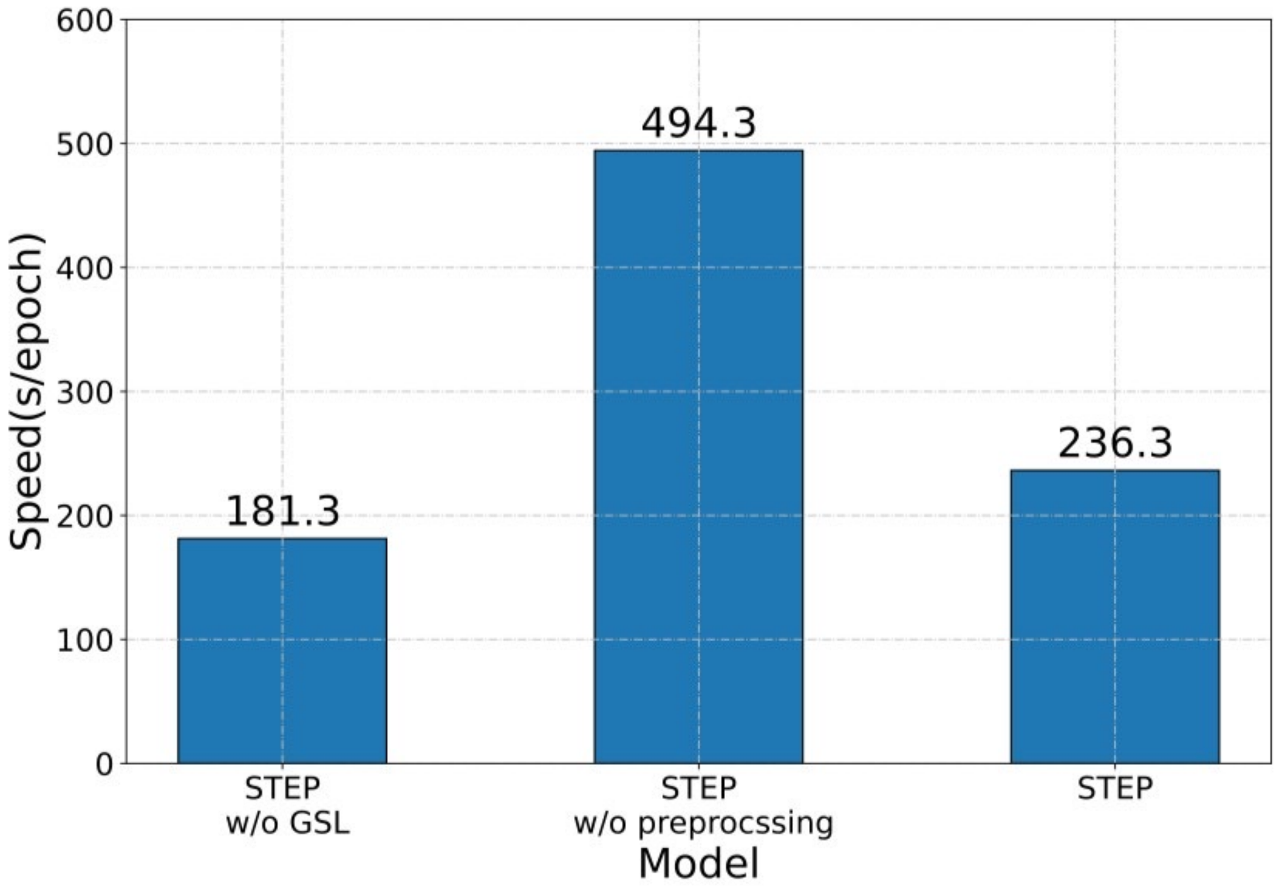



**Figure 6: Training speed of different methods.**