# Spatial-Temporal-Decoupled Masked Pre-training for Spatiotemporal Forecasting

Haotian Gao[1,2], Renhe Jiang[1]*, Zheng Dong[2], Jinliang Deng[3], Yuxin Ma[2], Xuan Song[2]

[1]The University of Tokyo

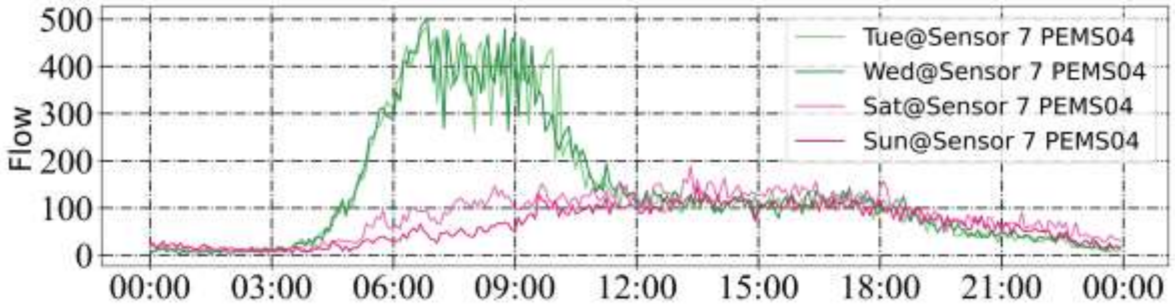
[2]Southern University of Science and Technology

[3]University of Technology Sydney

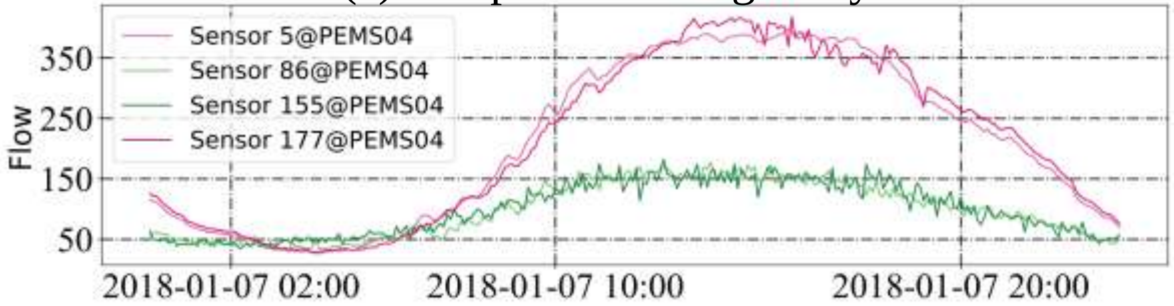
# Challenges: Heterogeneity and Mirage



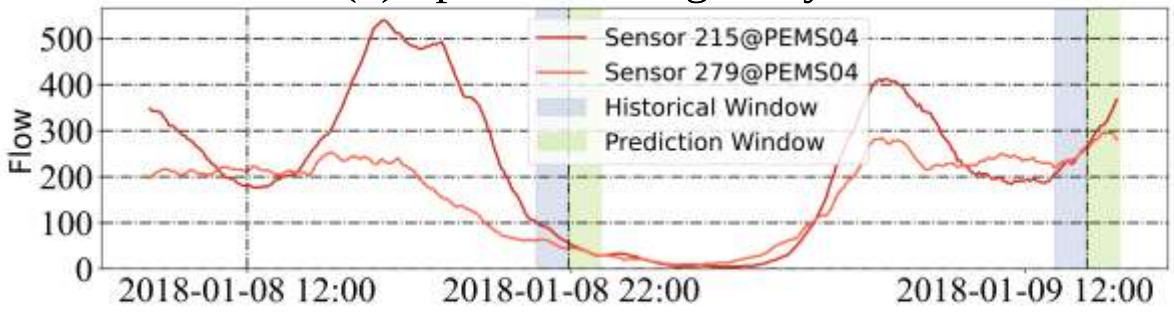
- **Spatiotemporal heterogeneity and mirage**
  - Spatiotemporal heterogeneity:
    Differences and inconsistencies in patterns or behaviors observed across different locations and times.
  - Spatiotemporal mirage:
    i) Dissimilar input time series followed by similar future values;
    ii) Similar input time series followed by dissimilar future values.



(a) Temporal Heterogeneity



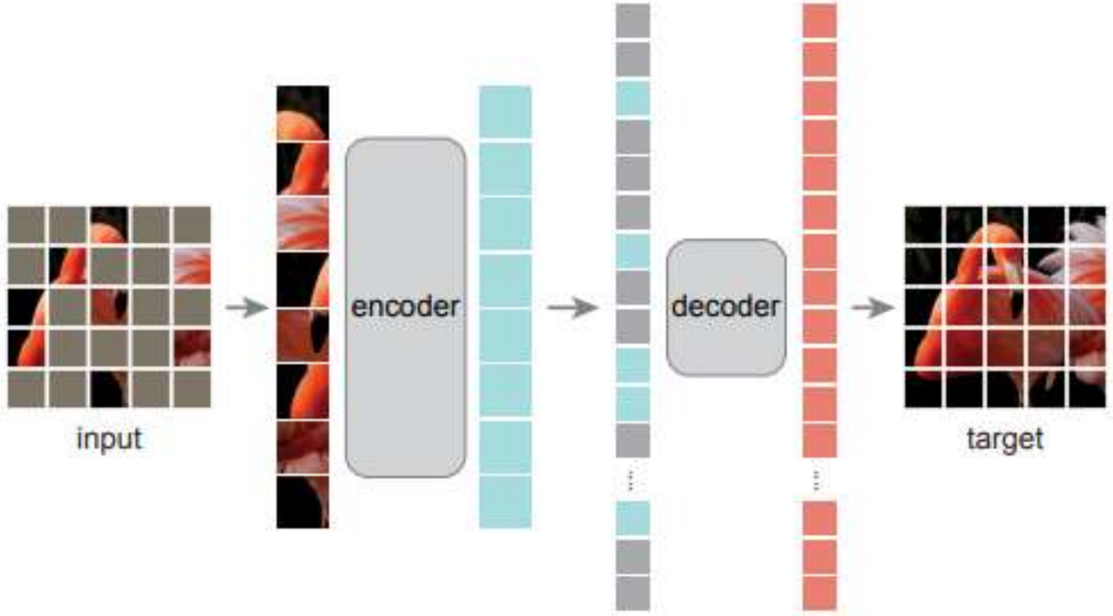
(b) Spatial Heterogeneity



(c) Spatiotemporal Mirage

# Masked Pre-training

- **Masked Autoencoder (MAE):**
  **Mask part of the input and reconstruct them from the visible part.**

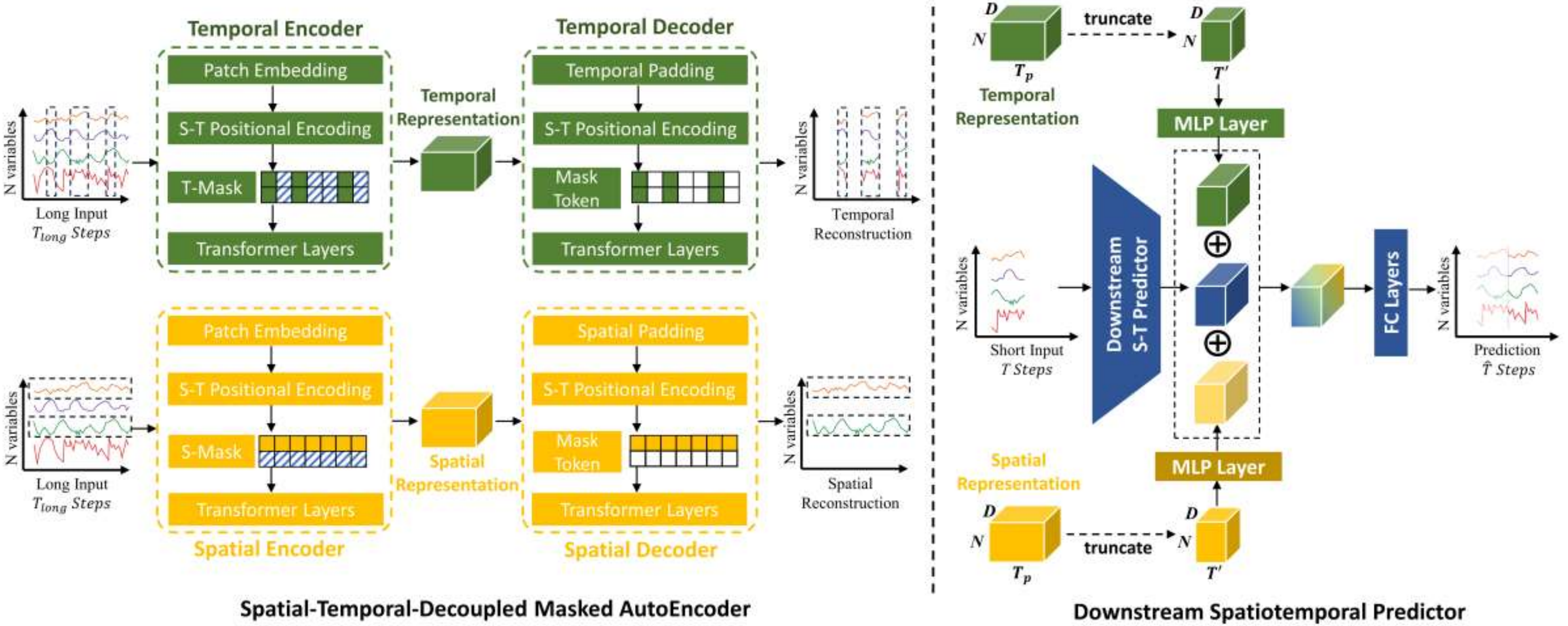



- **Basic Solution:**
  Apply MAE to learn temporal representation. (*STEP KDD 2022*)


He, K., Chen, X., Xie, S., Li, Y., Dollár, P. and Girshick, R., 2022. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000-16009)
Shao, Z., Zhang, Z., Wang, F. and Xu, Y., 2022, August. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 1567-1577)..

# STD-MAE Framework

- **Therefore, we aim to introduce spatial-temporal-decoupled masked pre-training (STD-MAE) to capture clear and complete spatiotemporal heterogeneity.**

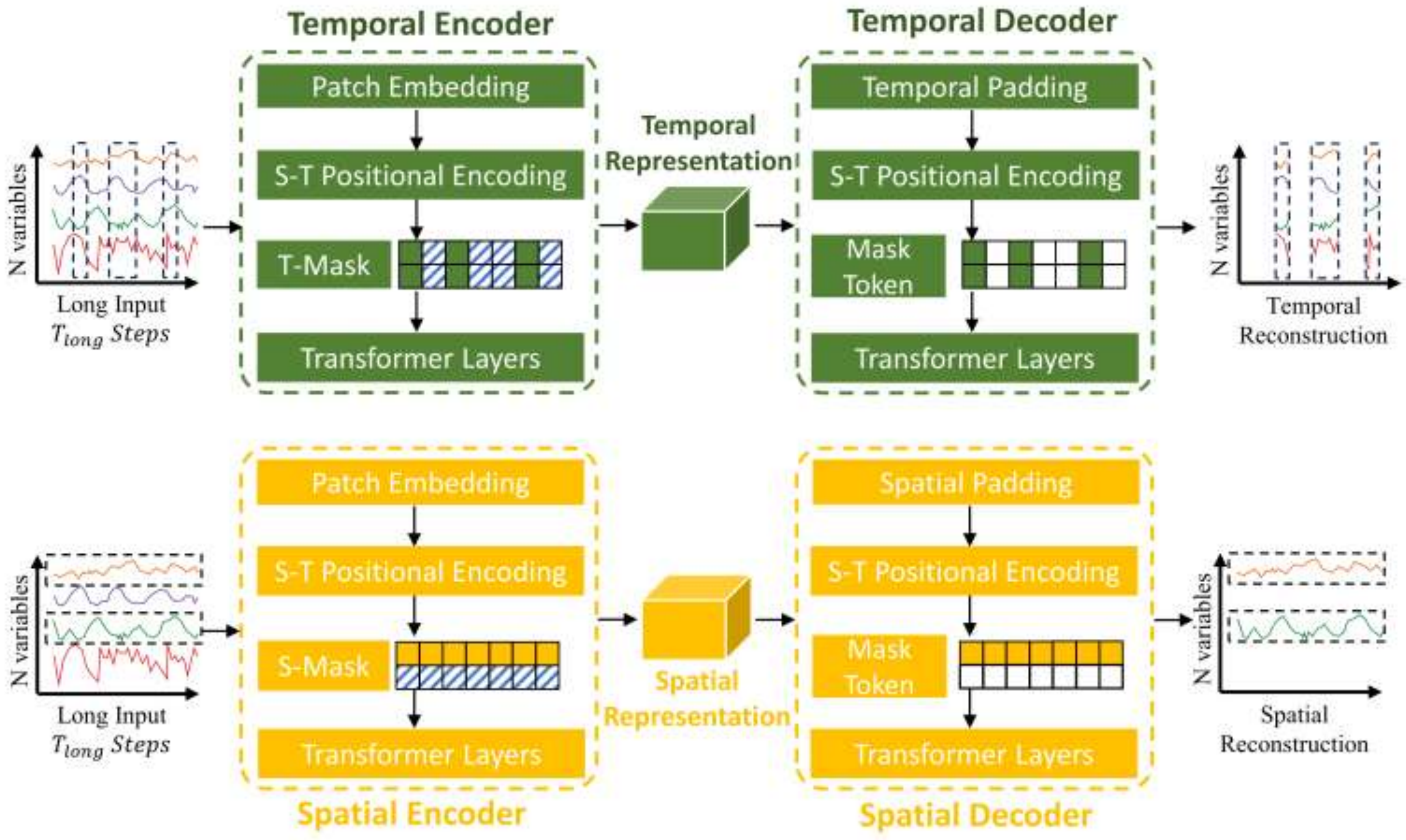


# Pre-training Phase



- Two **decoupled** masked autoencoders are trained to learn the temporal and spatial relation.

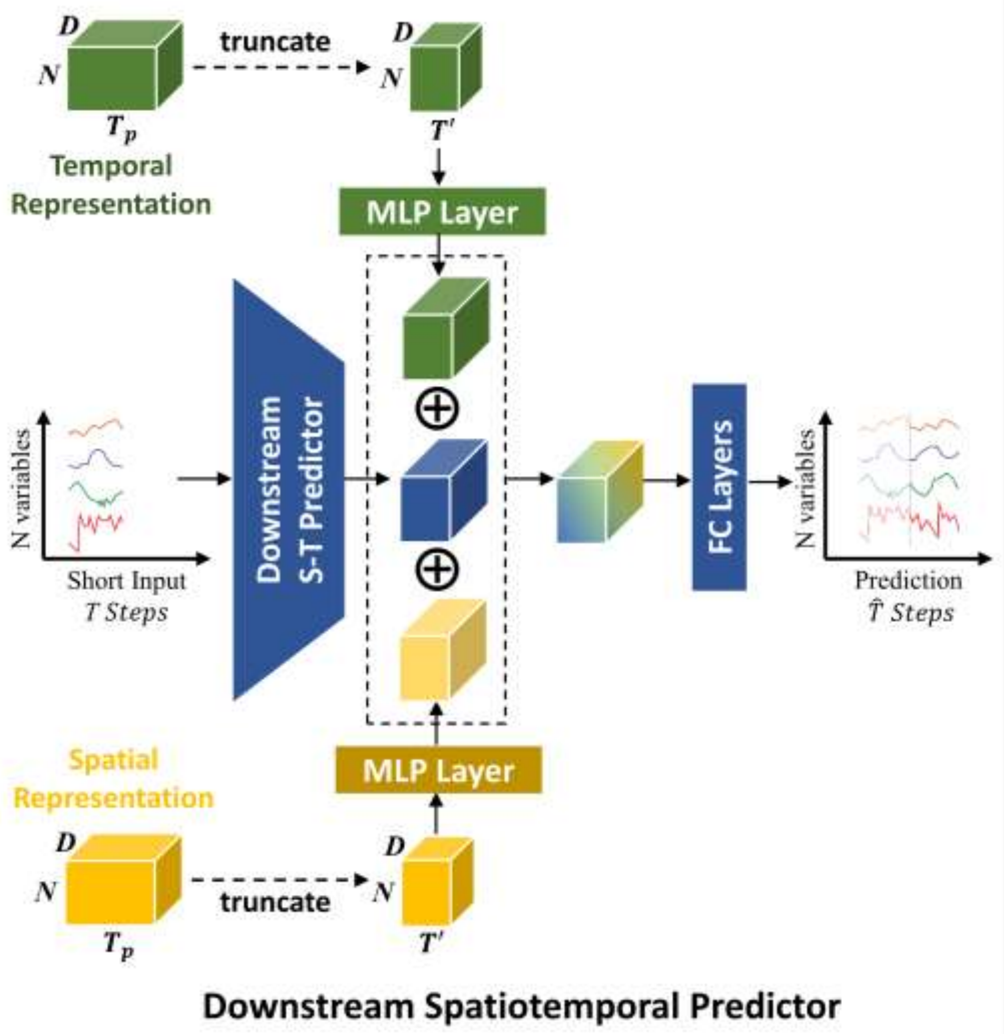
- **Decoupled Masking Strategies:**
  - Temporal Masking (T-Mask)
  - Spatial Masking (S-Mask)




Spatial-Temporal-Decoupled Masked AutoEncoder

# Forecasting Phase

- The learned spatial and temporal representations by two autoencoders through pre-training are fed to downstream predictors **without modifying original architecture**.



# Overall Performance



- **Datasets:** Four real-world datasets including **PEMS03, 04, 07,** and **08** collected from the California Transportation Performance Management System (**PEMS**) are used.

- STD-MAE has sustained **state-of-the-art** performance on these four spatiotemporal forecasting benchmarks for almost one year.

| Model | PEMS03 | | | PEMS04 | | | PEMS07 | | | PEMS08 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE | MAE | RMSE | MAPE |
| ARIMA [Fang *et al.*, 2021] | 35.31 | 47.59 | 33.78 | 33.73 | 48.80 | 24.18 | 38.17 | 59.27 | 19.46 | 31.09 | 44.32 | 22.73 |
| VAR [Song *et al.*, 2020] | 23.65 | 38.26 | 24.51 | 23.75 | 36.66 | 18.09 | 75.63 | 115.24 | 32.22 | 23.46 | 36.33 | 15.42 |
| SVR [Song *et al.*, 2020] | 21.97 | 35.29 | 21.51 | 28.70 | 44.56 | 19.20 | 32.49 | 50.22 | 14.26 | 23.25 | 36.16 | 14.64 |
| LSTM [Song *et al.*, 2020] | 21.33 | 35.11 | 23.33 | 27.14 | 41.59 | 18.20 | 29.98 | 45.84 | 13.20 | 22.20 | 34.06 | 14.20 |
| TCN [Lan *et al.*, 2022] | 19.31 | 33.24 | 19.86 | 31.11 | 37.25 | 15.48 | 32.68 | 42.23 | 14.22 | 22.69 | 35.79 | 14.04 |
| Transformer [Vaswani *et al.*, 2017] | 17.50 | 30.24 | 16.80 | 23.83 | 37.19 | 15.57 | 26.80 | 42.95 | 12.11 | 18.52 | 28.68 | 13.66 |
| DCRNN [Li *et al.*, 2018] | 18.18 | 30.31 | 18.91 | 24.70 | 38.12 | 17.12 | 25.30 | 38.58 | 11.66 | 17.86 | 27.83 | 11.45 |
| STGCN [Yu *et al.*, 2018] | 17.49 | 30.12 | 17.15 | 22.70 | 35.55 | 14.59 | 25.38 | 38.78 | 11.08 | 18.02 | 27.83 | 11.40 |
| ASTGCN [Guo *et al.*, 2019] | 17.69 | 29.66 | 19.40 | 22.93 | 35.22 | 16.56 | 28.05 | 42.57 | 13.92 | 18.61 | 28.16 | 13.08 |
| STG2Seq [Bai *et al.*, 2019] | 19.03 | 29.73 | 21.55 | 25.20 | 38.48 | 18.77 | 32.77 | 47.16 | 20.16 | 20.17 | 30.71 | 17.32 |
| GWNet [Wu *et al.*, 2019] | 19.85 | 32.94 | 19.31 | 25.45 | 39.70 | 17.29 | 26.85 | 42.78 | 12.12 | 19.13 | 31.05 | 12.68 |
| STSGCN [Song *et al.*, 2020] | 17.48 | 29.21 | 16.78 | 21.19 | 33.65 | 13.90 | 24.26 | 39.03 | 10.21 | 17.13 | 26.80 | 10.96 |
| STFGNN [Li and Zhu, 2021] | 16.77 | 28.34 | 16.30 | 19.83 | 31.88 | 13.02 | 22.07 | 35.80 | 9.21 | 16.64 | 26.22 | 10.60 |
| STGODE [Fang *et al.*, 2021] | 16.50 | 27.84 | 16.69 | 20.84 | 32.82 | 13.77 | 22.99 | 37.54 | 10.14 | 16.81 | 25.97 | 10.62 |
| DSTAGNN [Lan *et al.*, 2022] | 15.57 | 27.21 | 14.68 | 19.30 | 31.46 | 12.70 | 21.42 | 34.51 | 9.01 | 15.67 | 24.77 | 9.94 |
| ST-WA [Cirstea *et al.*, 2022] | 15.17 | 26.63 | 15.83 | 19.06 | 31.02 | 12.52 | 20.74 | 34.05 | 8.77 | 15.41 | 24.62 | 9.94 |
| ASTGNN [Guo *et al.*, 2021] | 15.07 | 26.88 | 15.80 | 19.26 | 31.16 | 12.65 | 22.23 | 35.95 | 9.25 | 15.98 | 25.67 | 9.97 |
| EnhanceNet [Cirstea *et al.*, 2021] | 16.05 | 28.33 | 15.83 | 20.44 | 32.37 | 13.58 | 21.87 | 35.57 | 9.13 | 16.33 | 25.46 | 10.39 |
| AGCRN [Bai *et al.*, 2020] | 16.06 | 28.49 | 15.85 | 19.83 | 32.26 | 12.97 | 21.29 | 35.12 | 8.97 | 15.95 | 25.22 | 10.09 |
| Z-GCNETs [Chen *et al.*, 2021] | 16.64 | 28.15 | 16.39 | 19.50 | 31.61 | 12.78 | 21.77 | 35.17 | 9.25 | 15.76 | 25.11 | 10.01 |
| STNorm [Deng *et al.*, 2021] | 15.32 | 25.93 | 14.37 | 19.21 | 32.30 | 13.05 | 20.59 | 34.86 | 8.61 | 15.39 | 24.80 | 9.91 |
| STEP [Shao *et al.*, 2022b] | 14.22 | 24.55 | 14.42 | 18.20 | 29.71 | 12.48 | 19.32 | 32.19 | 8.12 | 14.00 | 23.41 | 9.50 |
| PDFormer [Jiang *et al.*, 2023] | 14.94 | 25.39 | 15.82 | 18.32 | 29.97 | 12.10 | 19.83 | 32.87 | 8.53 | 13.58 | 23.51 | 9.05 |
| STAEformer [Liu *et al.*, 2023] | 15.35 | 27.55 | 15.18 | 18.22 | 30.18 | 11.98 | 19.14 | 32.60 | 8.01 | 13.46 | 23.25 | 8.88 |
| **STD-MAE (Ours)** | **13.80** | **24.43** | **13.96** | **17.80** | **29.25** | **11.97** | **18.65** | **31.44** | **7.84** | **13.44** | **22.47** | **8.76** |

Table 2: Performance Comparison with Baseline Models on PEMS03,04,07,08 Benchmarks

# Ablation Study

- **Ablation experiments on Masking Mechanism:**
  - T-MAE: Only **temporal** masking.
  - S-MAE: Only **spatial** masking.
  - STM-MAE: Masking on **mixed** spatial and temporal dimension.
  - w/o Mask: **Without** masking mechanism.

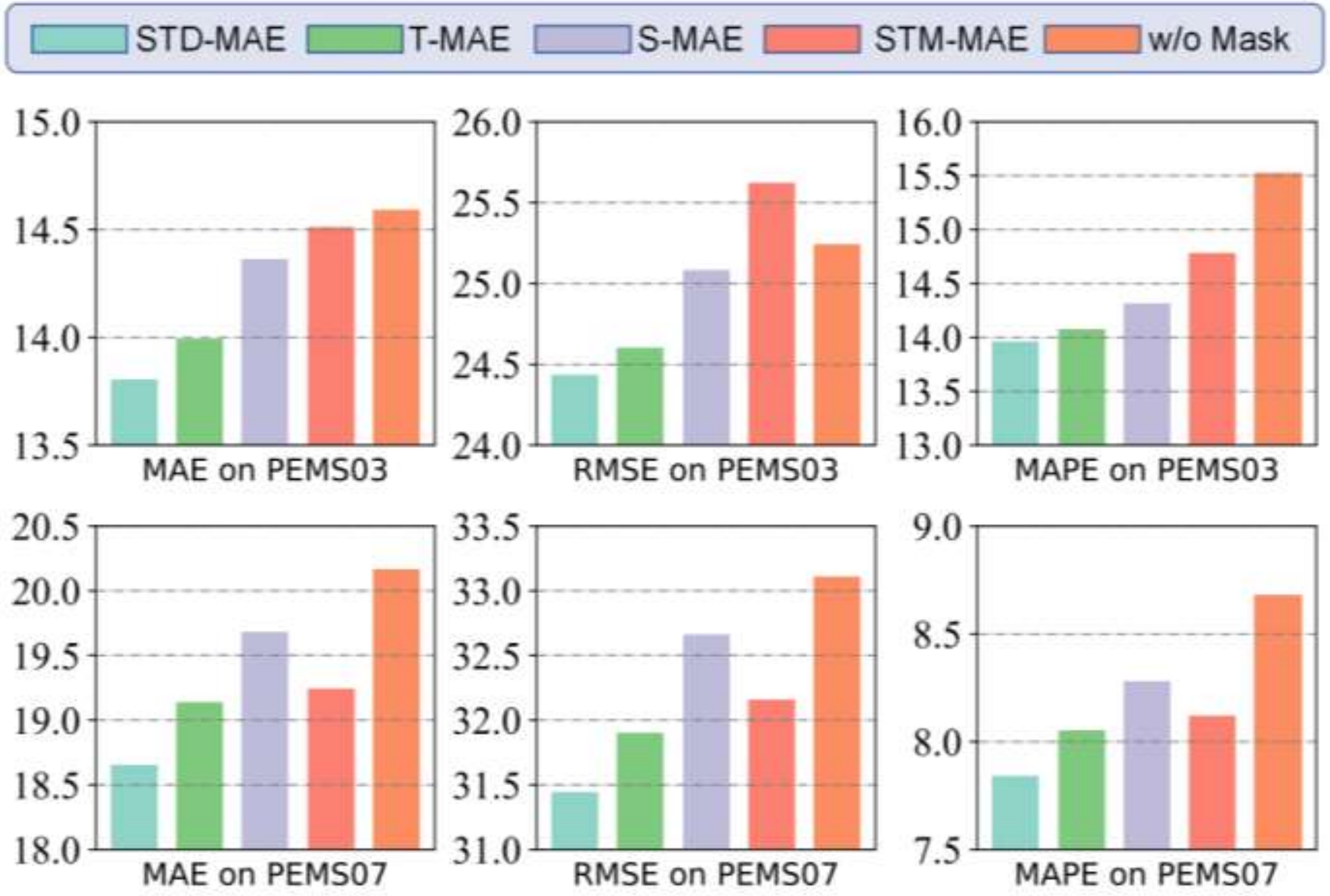



Figure 3: Masking Ablation on PEMS03 and PEMS07

IJCAI JEJU 2024

- **Hyper-parameter Study on Mask Ratio:** The **optimal mask ratio** is **0.25**, which indicates that time series data require a lower mask ratio compared to natural language data (mask ratio=**0.9**) and image data (mask ratio=**0.75**).

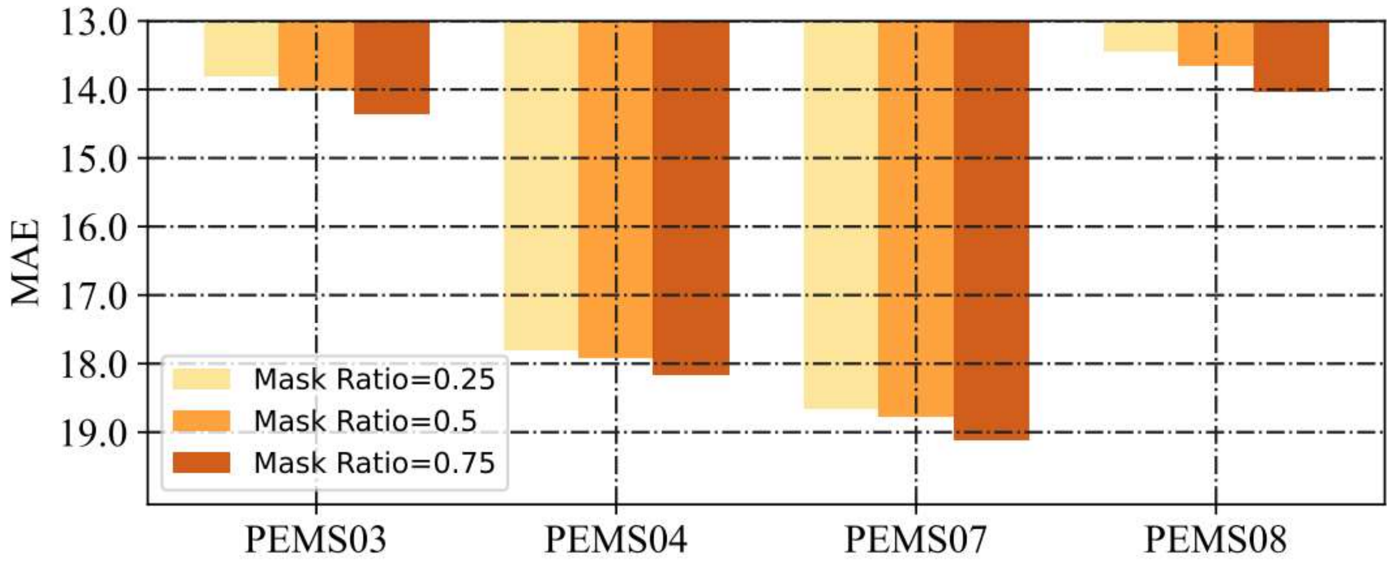
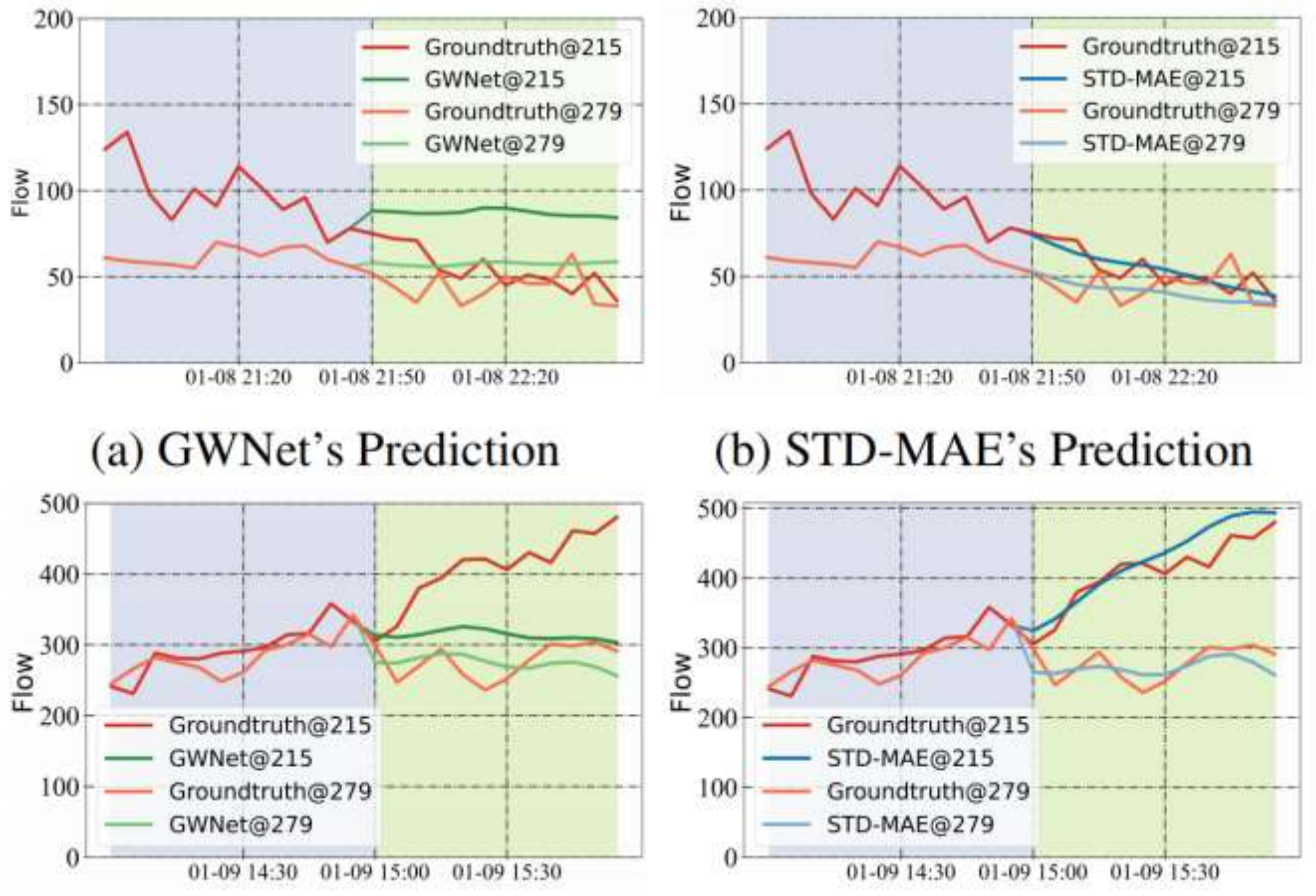



Figure 4: Hyper-parameter Study on Masking Ratio

- **Prediction under Spatiotemporal Mirage:**

  - GWNet: **Without** masked pre-trainin
  - STD-MAE: With **spatial-temporal-masked pre-training**.




(a) GWNet's Prediction (b) STD-MAE's Prediction

(c) GWNet's Prediction (d) STD-MAE's Prediction

Figure 6: Prediction under Spatiotemporal Mirage

# Conclusion



- **Conclusion:**

  1. We propose STD-MAE, a novel spatial-temporal-decoupled masked pre-training framework for spatiotemporal forecasting.

  2. We propose a novel spatial-temporal-decoupled masking strategy to effectively learn spatial and temporal heterogeneity.

  3. Comprehensive experiments and in-depth analyses conducted on four benchmark datasets demonstrate the superiority of STD-MAE.

arXiv

GitHub