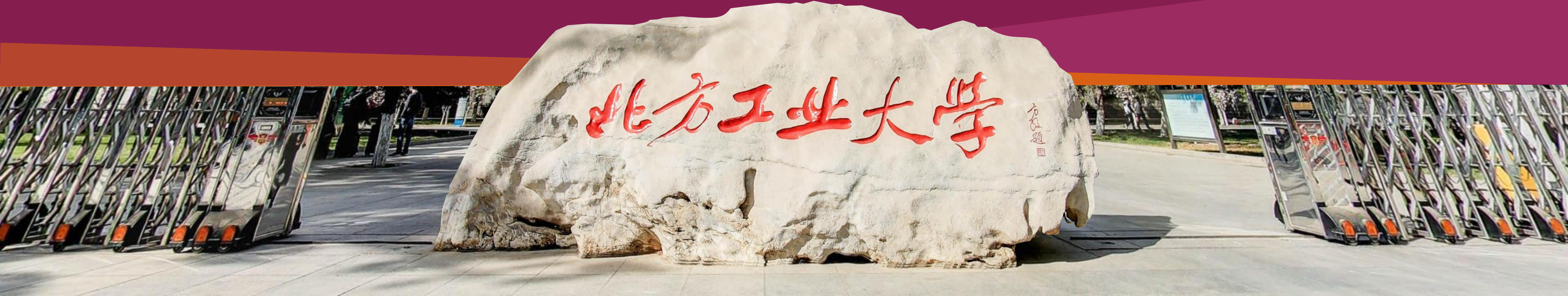
North China University of Technology
Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data

Rethinking Attention Mechanism for Spatio-Temporal Modeling: A Decoupling Perspective in Traffic Flow Prediction

Qi Yu, Weilong Ding*, Hao Zhang, Yang Yang, Tianpu Zhang



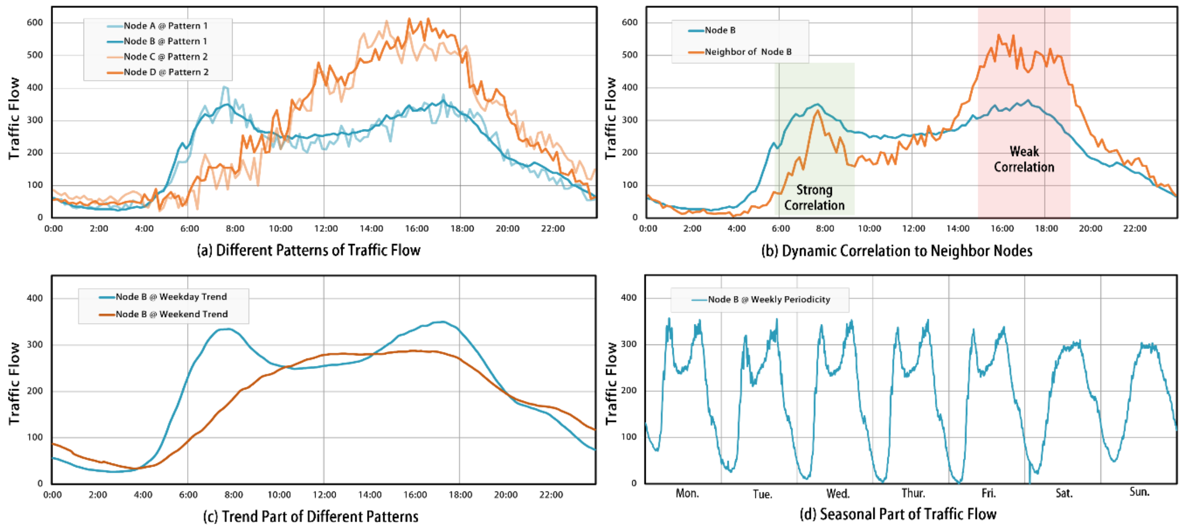
NORTH CHINA UNIVERSITY OF TECHNOLOGY

Attention in Spatio-Temporal Modeling

- **Spatial Attention:** Graph Attention Network (GAT)
- **Temporal Attention:** Multi-Head Attention
- **Advanced Attention Techniques:** Calculations in Fourier or wavelet space

Traffic flow prediction

- Neighbor nodes with different patterns
- Variations across different days
- Long-term periodic nature



Contributions

DEC-Former：A traffic flow prediction model from a DECoupling perspective

- **Decoupled Perspective**
  - *Decouple* the time series data into trend and seasonal parts
  - *Decouple* the geographical adjacency of road network
  - *Decouple* the classical encoder-decoder architecture
- **Efficient Attention Utilization**
  - Attention only for the seasonal part and a dynamic spatial attention module
- **Superior Predictive and Computational Performance**
  - Evaluate on four real-world datasets

Preliminary

Definition and Problem Statement

Traffic Topology Graph

- $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, $\mathcal{A} \in \mathbb{R}^{N \times N}$

Traffic Flow Tensor

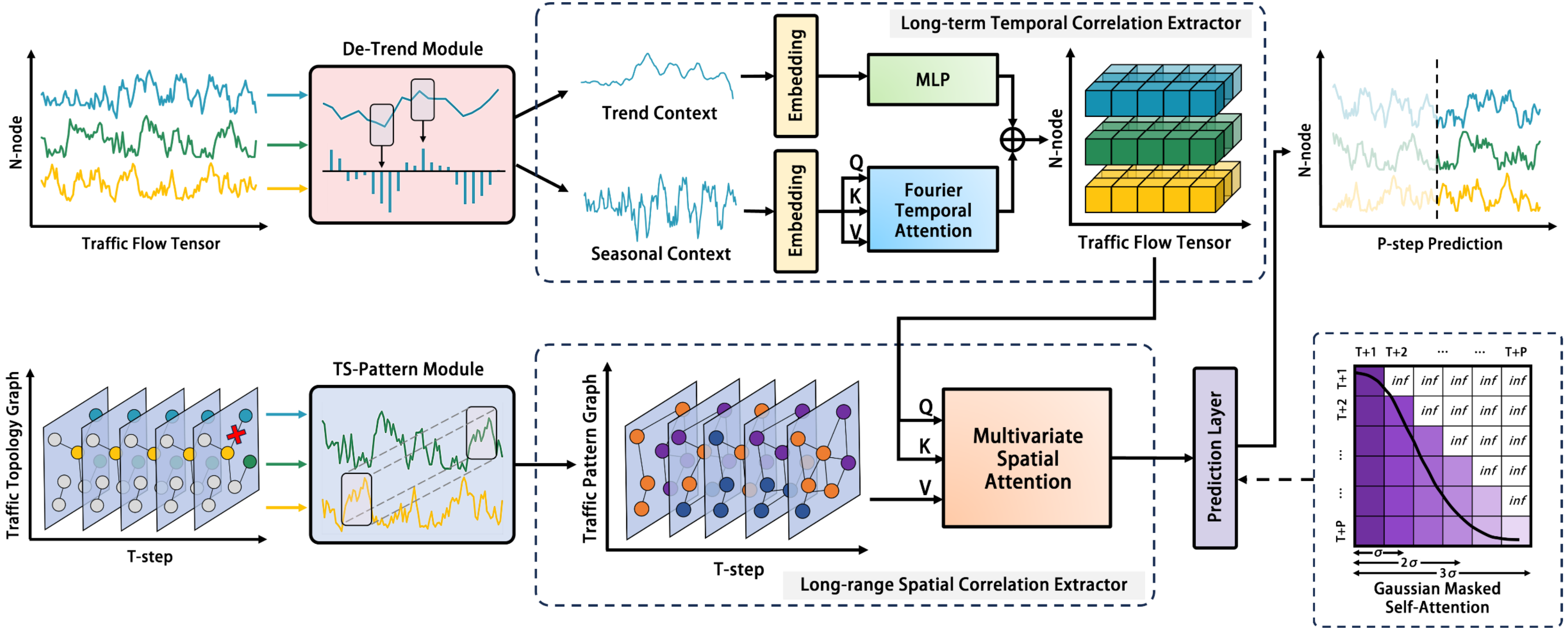
- $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_t, \ldots, \mathbf{X}_T\} \in \mathbb{R}^{N \times C \times T}$

Traffic Flow Prediction Problem

- $[\mathbf{X}_{t+1}, \cdots, \mathbf{X}_{t+P}] = \mathcal{F}_\theta([\mathbf{X}_{t-T+1}, \cdots, \mathbf{X}_t; \mathcal{G}])$

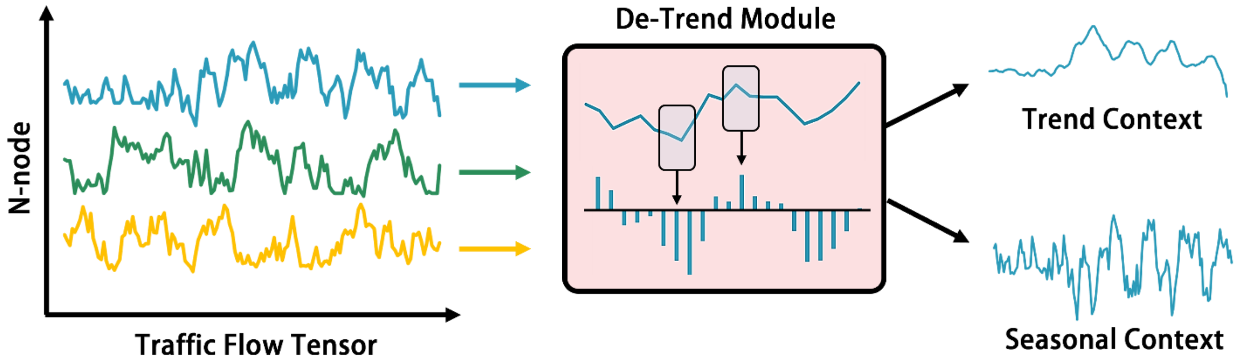
Methodology

DEC-Former



Trend Decomposition Module

- **Focus on the trend and seasonal parts**



- $\mathcal{X}_{input} \in \mathbb{R}^{N\times C\times T}$

$$\mathcal{X}_{tre} = AvgPool(padding(\mathcal{X}_{input}), m),$$

$$\mathcal{X}_{sea} = \mathcal{X}_{input} - \mathcal{X}_{tre}.$$

Methodology

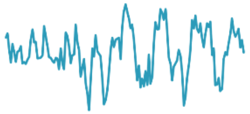
Long-term Temporal Correlation Extractor

MLP for Trend Context

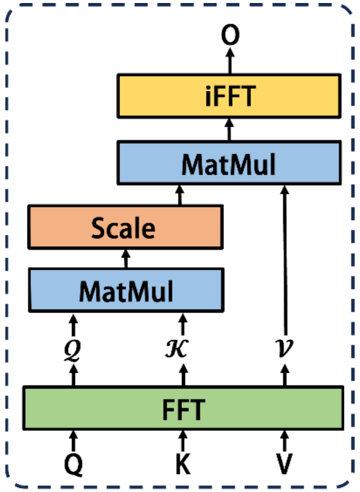
Trend Context

- $\hat{\mathcal{X}}_{tre} = ReLU(W_L^{(l)} \mathcal{X}_{tre}^{(l-1)} + b_L^{(l)}).$

Fourier Attention for Seasonal Context



Seasonal Context



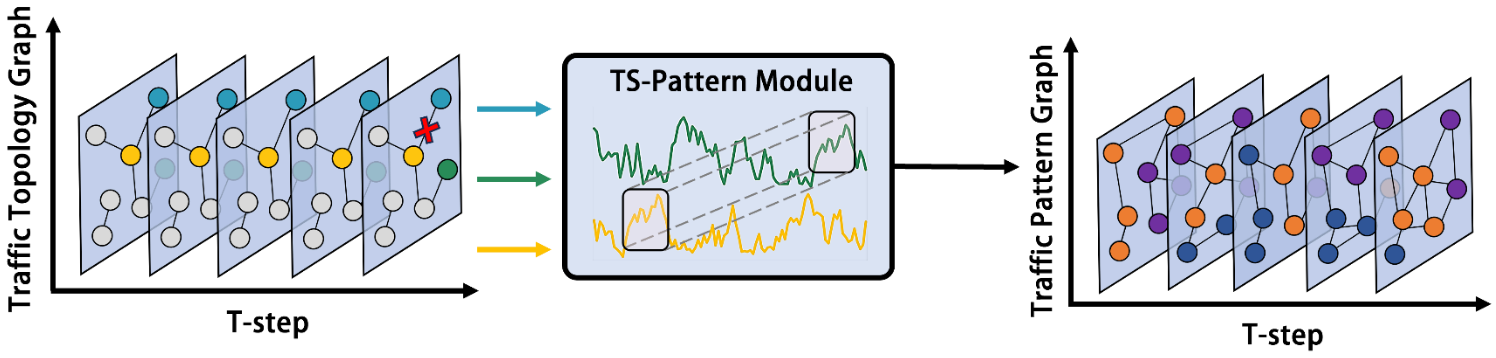
(a) Fourier Temporal Attention

- $$\begin{cases} \mathcal{Q}_f = FFT(Q_f) = FFT(\mathcal{X}_{sea} W_{F,q}), \\ \mathcal{K}_f = FFT(K_f) = FFT(\mathcal{X}_{sea} W_{F,k}), \\ \mathcal{V}_f = FFT(V_f) = FFT(\mathcal{X}_{sea} W_{F,v}), \end{cases}$$

- $\hat{\mathcal{X}}_{sea} = iFFT(softmax(\mathcal{Q}_f \mathcal{K}_f^T)\mathcal{V}_f).$

Long-range Spatial Correlation Extractor

Traffic Pattern Extraction



- K clusters as K traffic patterns

$$C = \{C_1, \cdots, C_K\}$$

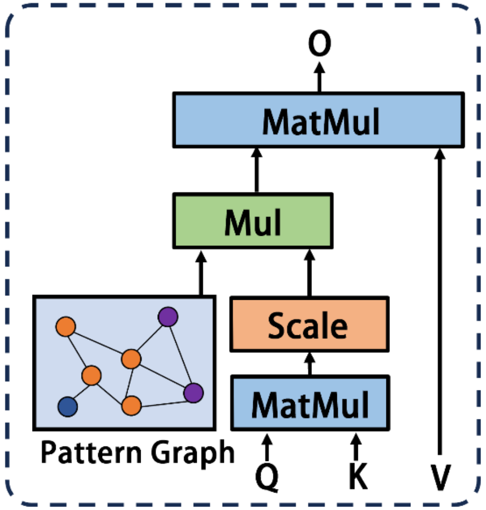
- $S_{m,n}^{(\tau)}$ as the pattern frequency

$$S_{m,n}^{(\tau)} = \sum (C_m^{\tau_d} == C_n^{\tau_d}) \quad > \varphi \quad \Longrightarrow \quad (A_P^{(\tau)})_{mn} = 1.$$

Methodology

Long-range Spatial Correlation Extractor

◆ Multivariate Spatial Attention



(b) Multivariate Spatial Attention

- spatial correlation weight matrix

$$A_S^h = softmax(\frac{\mathcal{Q}_m^h(\mathcal{K}_m^h)^T}{\sqrt{d_m}}) \in \mathbb{R}^{N\times N},$$

- incorporate the pattern correlation

$$\mathcal{L}_h = (A_S^h \odot A_P)\mathcal{V}_m^h.$$

- outputs from multiple attention heads

$$\hat{\mathcal{L}} = \oplus(\mathcal{L}_1, \cdots, \mathcal{L}_h)W_M.$$

Dataset

- **PeMS03, PeMS04, PeMS07, PeMS08**
  - Data aggregated at 5-minute intervals, i.e., 12 sample points per hour.

Datasets	#Node	#Time step	Time Range
PeMS03	358	26202	09/01/2018-11/30/2018
PeMS04	307	16992	01/01/2018-02/28/2018
PeMS07	883	28224	05/01/2017-08/31/2017
PeMS08	170	17856	07/01/2016-08/31/2016

Evaluation

Setting

- **Training : Validation : Test** = 6 : 2 : 2
- **Goal:** Predicting the next hour's data using the past day's data.
- **Evaluation Metrics**
  - $$\text{MAE} = \frac{1}{N}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|,$$
  - $$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2},$$
  - $$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{n}\frac{\left|y_i - \hat{y}_i\right|}{y_i} \times 100\%$$

Evaluation

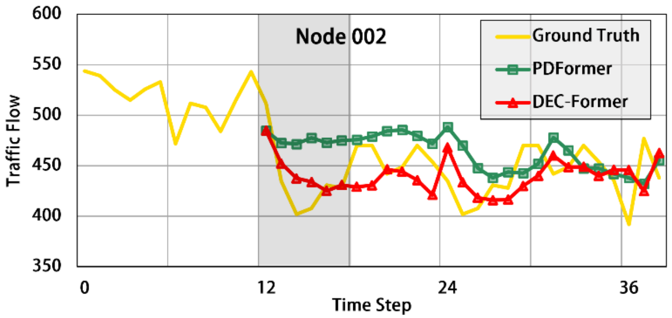
Baseline

- **Time series models**
  - VAR, SVR, FC-LSTM.
- **GNN-based models**
  - DCRNN, STGCN, Graph Wave Net, STSGCN.
- **Attention-based models**
  - ASTGCN, GMAN, DSTAGNN, STGSA,
  - **ISTNet** captures local correlations through CNN to supplement into transformer model,
  - **PDFormer** designs a self-attention mechanism through a graph mask and a delay-aware module.

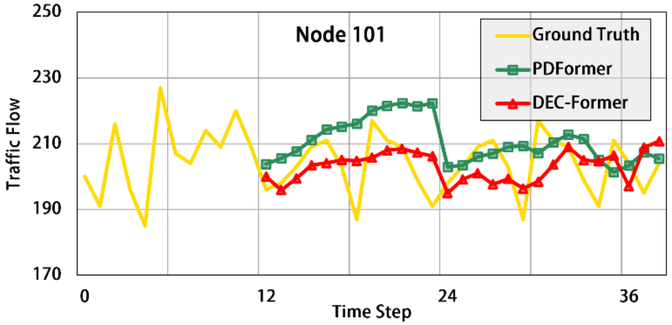
Result Analysis

Performance Comparison

Model	PeMS03			PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)
VAR	19.72	32.38	20.50	24.44	37.76	17.27	27.96	41.31	12.11	19.83	29.24	13.08
SVR	19.77	32.78	23.04	26.18	38.91	22.84	28.45	42.67	14.00	20.92	31.23	14.24
FC-LSTM	19.56	33.38	19.56	23.60	37.11	16.17	34.05	55.70	15.31	21.18	31.88	13.72
DCRNN	17.62	29.86	16.83	24.42	37.48	16.86	24.45	37.61	10.67	18.49	27.30	11.69
STGCN	19.76	33.87	17.33	23.90	36.43	13.67	26.22	39.18	10.74	18.79	28.20	10.55
GWNet	15.67	26.42	15.72	19.91	31.06	13.62	20.83	33.62	9.10	15.57	24.32	10.32
STSGCN	17.51	29.05	16.92	21.52	34.14	14.50	23.99	39.32	10.10	17.88	27.36	11.71
ASTGCN	18.67	30.71	19.85	22.90	33.59	16.75	28.13	43.67	13.31	18.72	28.99	12.53
GMAN	15.52	26.53	15.19	19.25	30.85	13.00	20.68	33.56	9.31	14.87	24.06	9.77
DSTAGNN	15.57	27.21	14.68	19.30	31.46	12.70	21.42	34.51	9.01	15.67	24.77	9.94
STGSA	15.36	27.89	14.45	19.32	31.30	12.90	20.80	34.30	8.86	15.26	24.28	9.81
ISTNet	15.12	25.14	15.43	18.54	30.46	12.52	19.79	33.06	8.77	14.13	23.39	9.43
PDFormer	14.73	24.54	15.42	18.51	30.24	12.38	20.65	34.36	8.68	14.34	23.68	9.88
DEC-Former	**14.33**	**23.55**	**14.27**	**18.23**	**29.24**	**12.04**	**19.48**	**33.04**	**8.54**	**13.23**	**23.06**	**9.12**



(a) Comparison of Predictive Curves on Trend Data

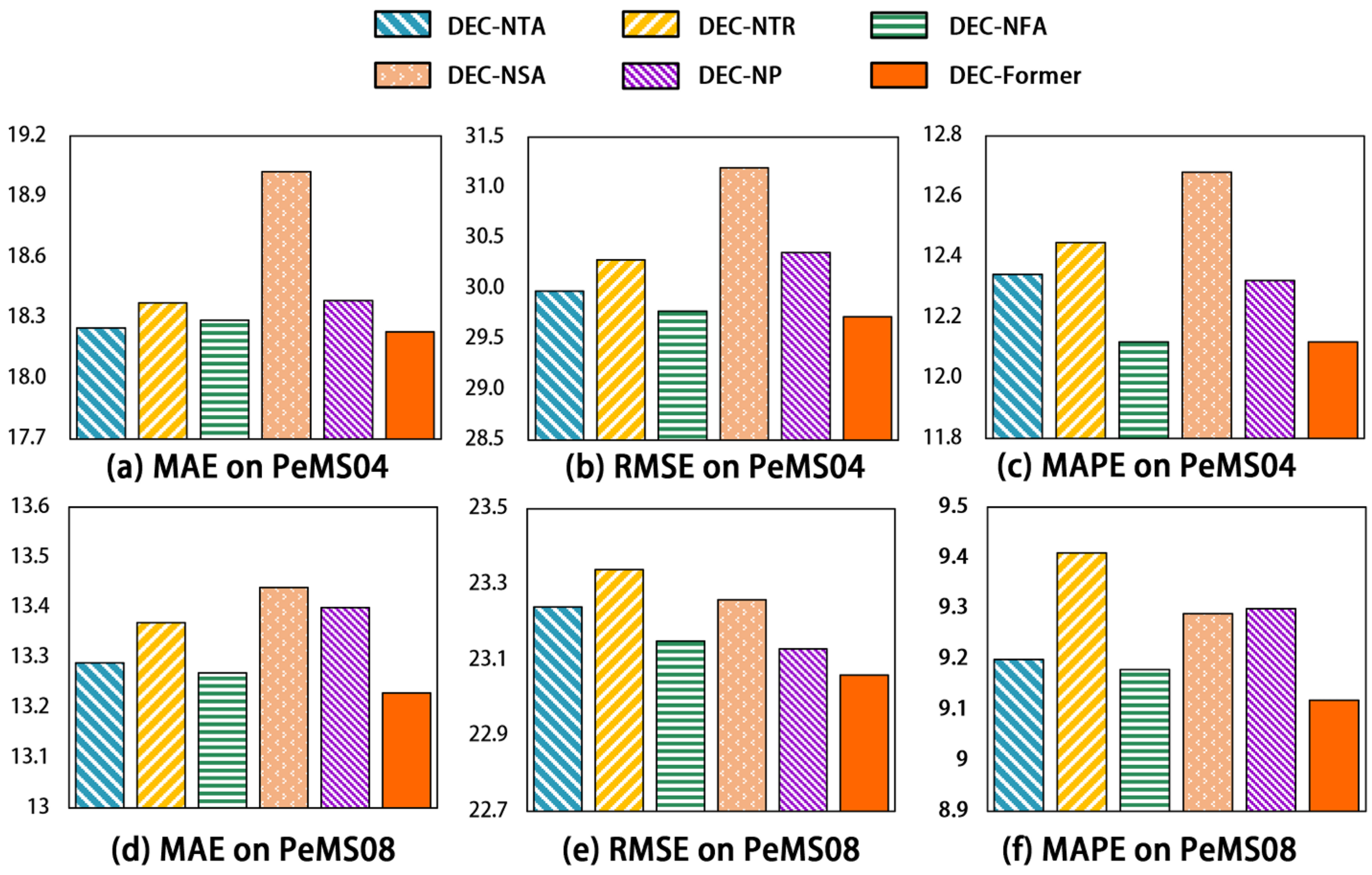


(b) Comparison of Predictive Curves on Seasonal Data

- MAE ↓3.04%
- RMSE ↓2.20%
- MAPE ↓2.23%

Result Analysis

Ablation Experiments



Result Analysis

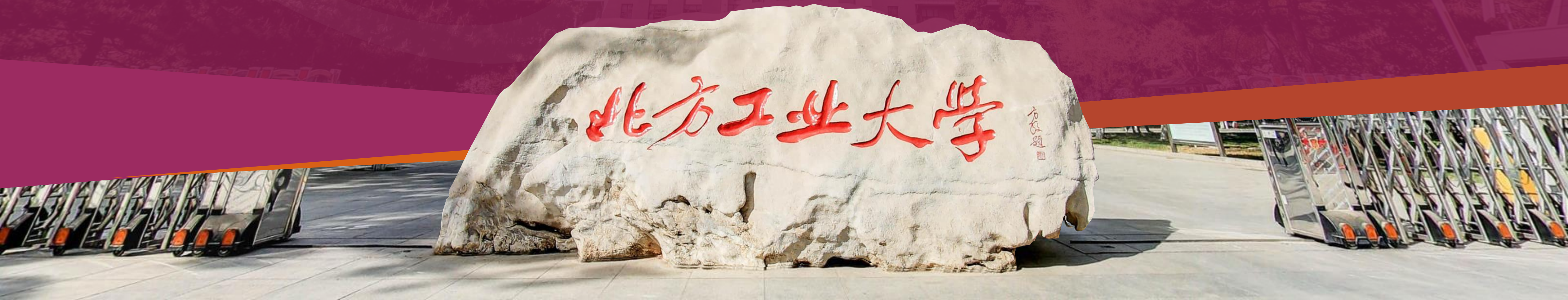
Computational Performance

Model	Training Time (s/epoch)	Inference Time on Testset (s)
STSGCN	848.08	532.80
DSTAGNN	1064.61	588.22
ISTNet	443.28 ↓ 16%	64.78 ↓ 30%
PDFormer	430.24 ↓ 14%	56.00 ↓ 18%
DEC-Former	371.74	45.56

Model	PeMS03			PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)
ISTNet	15.12	25.14	15.43	18.54	30.46	12.52	19.79	33.06	8.77	14.13	23.39	9.43
PDFormer	14.73	24.54	15.42	18.51	30.24	12.38	20.65	34.36	8.68	14.34	23.68	9.88
DEC-Former	**14.33**	**23.55**	**14.27**	**18.23**	**29.24**	**12.04**	**19.48**	**33.04**	**8.54**	**13.23**	**23.06**	**9.12**

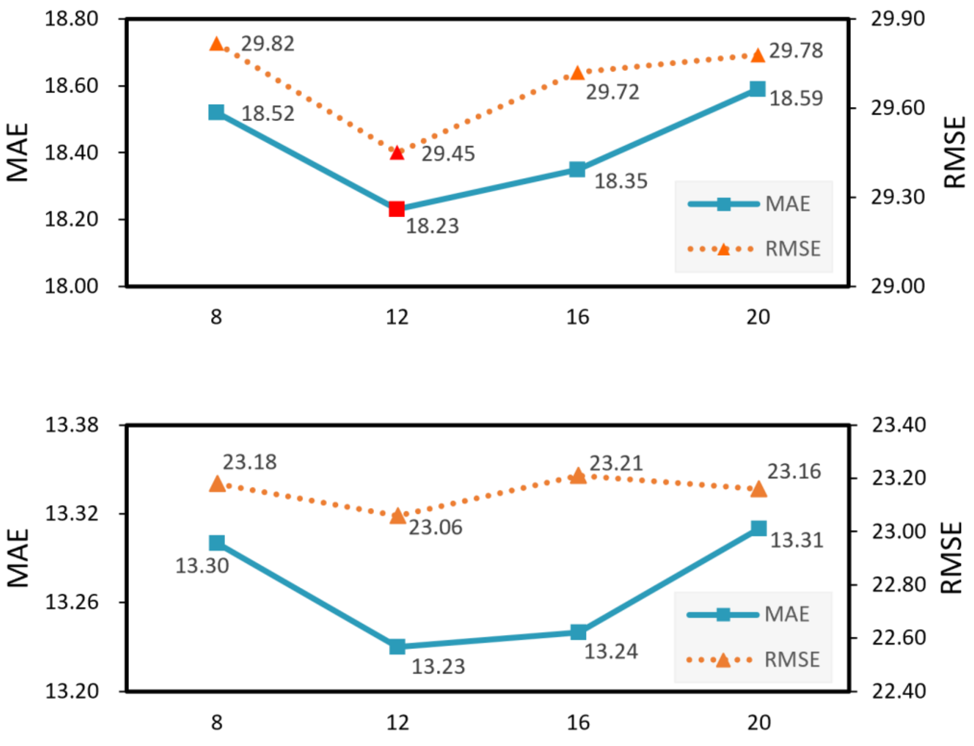
North China University of Technology
Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data

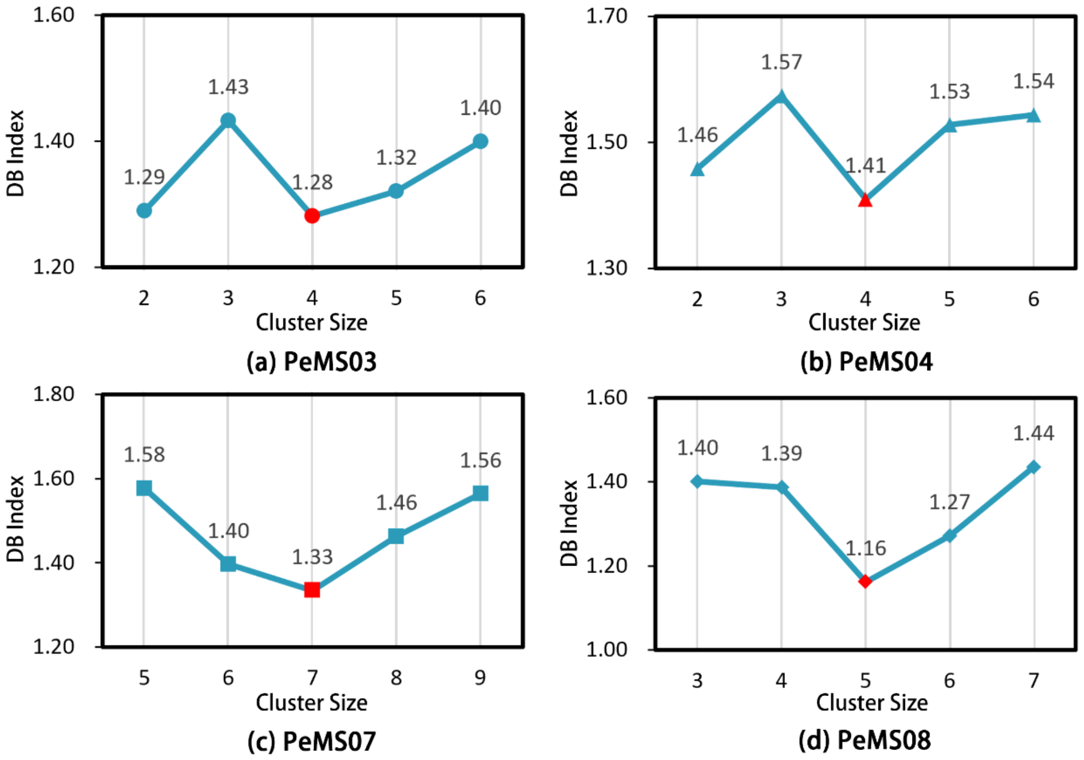
Thanks!



Result Analysis

Parameter Analysis





- The Impact of Window Size m.
- The Impact of the Cluster Size k.

Rethinking Attention Mechanism for Spatio-Temporal Modeling: A Decoupling Perspective in Traffic Flow Prediction

Qi Yu, Weilong Ding*, Hao Zhang, Yang Yang, Tianpu Zhang

Rethinking Attention Mechanism for Spatio-Temporal Modeling: A Decoupling Perspective in Traffic Flow Prediction

Qi Yu
North China University of Technology
Beijing, China
yuqi_april@mail.ncut.edu.cn

Weilong Ding*
North China University of Technology
Beijing, China
dingweilong@ncut.edu.cn

Hao Zhang
North China University of Technology
Beijing, China
2023322030127@mail.ncut.edu.cn

Yang Yang
North China University of Technology
Beijing, China
2023312120113@mail.ncut.edu.cn

Tianpu Zhang
SINOPEC Beijing Research Institute of Chemical Industry
Beijing, China
zhangtianpu@hotmail.com

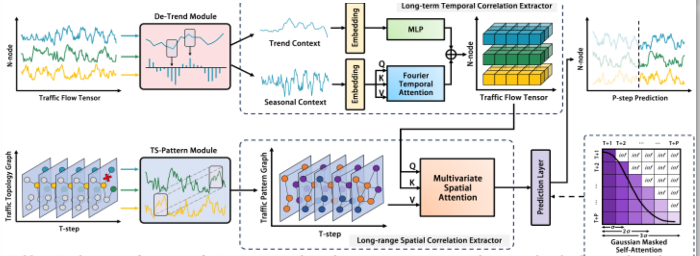


Table 2: Predictive Performance of DEC-Former and Baselines on PeMS Datasets. The optimal and suboptimal results are highlighted in bold and underlined respectively.

Model	PeMS03			PeMS04			PeMS07			PeMS08		
	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)	MAE	RMSE	MAPE(%)
VAR	19.72	32.38	20.50	24.44	37.76	17.27	27.96	41.31	12.11	19.83	29.24	13.08
SVR	19.77	32.78	23.04	26.18	38.91	22.84	28.45	42.67	14.00	20.92	31.23	14.24
FC-LSTM	19.56	33.38	19.56	23.60	37.11	16.17	34.05	55.70	15.31	21.18	31.88	13.72
DCRNN	17.62	29.86	16.83	24.42	37.48	16.86	24.45	37.61	10.67	18.49	27.30	11.69
STGCN	19.76	33.87	17.33	23.90	36.43	13.67	26.22	39.18	10.74	18.79	28.20	10.55
GWNet	15.67	26.42	15.72	19.91	31.06	13.62	20.83	33.62	9.10	15.57	24.32	10.32
STSGCN	17.51	29.05	16.92	21.52	34.14	14.50	23.99	39.32	10.10	17.88	27.36	11.71
ASTGCN	18.67	30.71	19.85	22.90	33.59	16.75	28.13	43.67	13.31	18.72	28.99	12.53
GMAN	15.52	26.53	15.19	19.25	30.85	13.00	20.68	33.56	9.31	14.87	24.06	9.77
DSTAGNN	15.57	27.21	14.68	19.30	31.46	12.70	21.42	34.51	9.01	15.67	24.77	9.94
STGSA	15.36	27.89	14.45	19.32	31.30	12.90	20.80	34.30	8.86	15.26	24.28	9.81
ISTNet	15.12	25.14	15.43	18.54	30.46	12.52	19.79	33.06	8.77	14.13	23.39	9.43
PDFormer	14.73	24.54	15.42	18.51	30.24	12.38	20.65	34.36	8.68	14.34	23.68	9.88
DEC-Former	**14.33**	**23.55**	**14.27**	**18.23**	**29.24**	**12.04**	**19.48**	**33.04**	**8.54**	**13.23**	**23.06**	**9.12**