Supplementary information for

Wind-Aware, Time-Scale Adaptive $PM_{2.5}$ Spatiotemporal Forecasting with Spatial Regularization

Guyu Zhao[a,b] , Shuo Zhang[a,b] , QiMing Liu[a,b] and Hongdou He[a,*]
[a]School of Information Science and Engineering, Yanshan University, Qinhuangdao
[b]The Key Laboratory for Software Engineering of Hebei Province, Qinhuangdao

*Corresponding Author
He Hongdou, E-mail: hhd@ysu.edu.cn

Table S1. Optimal parameter values used in the experiments.

Hyperparameters	Values
Training epoch	100
Input sequence length	168
Hidden dimension	64
Mid-layer channel	32
Number of heads	4
Window size	8
Wind-bias scale	0.5
Number of experts	4
Top-U temperature	0.5
Number of prototypes	50
Cluster-softmax temperature	1.0
Dynamic-adj temperature	0.5
Static–dynamic fusion α init	0.1
Laplacian-reg weight	0.1
Learning rate	1e-4
Weight decay	1e-3

Table S2. Air Quality Level Classification Based on the "Technical Regulations for Ambient Air Quality Index"

Level	PM2.5 concentration range (μg/m³)
Excellent	0-35
Good	36-75
Light Pollution	76-115
Moderate Pollution	116-150
Heavy Pollution	151-250
Severe Pollution	>250

Table S3. Sensitivity Analysis of the Neighborhood Size k in kNN Graph Construction

K	1h		6h		12h		24h	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
2	4.637	7.300	8.932	13.568	11.056	16.985	13.057	19.677
4	**4.614**	**7.181**	**8.387**	**13.065**	**10.769**	**16.152**	**12.896**	**19.333**
6	4.635	7.204	8.654	13.548	11.036	16.835	13.202	19.712
8	4.659	7.259	8.854	13.882	11.532	17.002	13.636	20.051

Fig.S1. Distribution of Stations in the Beijing-Tianjin-Hebei Region

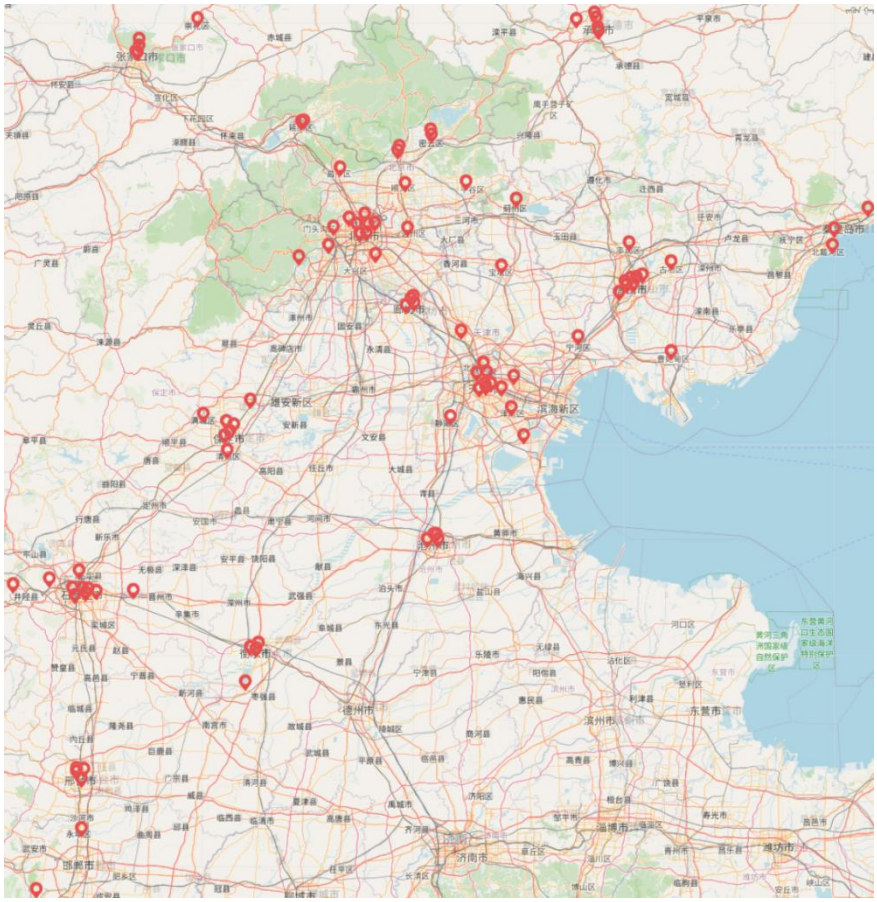

Fig.S2. Time Convolution Block structure

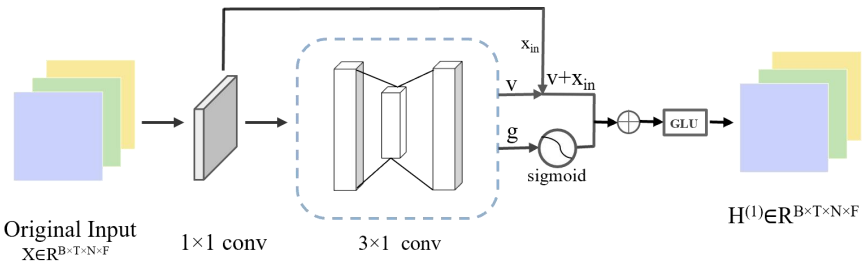

Fig.S3. Contour maps comparing $PM_{2.5}$ forecasts at different lead times with ground-truth observations. Results are shown for 1h, 6h, 12h, and 24h forecasts across cities in the BTH region under varying pollution levels.