Supplement of Atmos. Chem. Phys., 25, 14643–14668, 2025
https://doi.org/10.5194/acp-25-14643-2025-supplement
© Author(s) 2025. CC BY 4.0 License.

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Measurement report: Unraveling PM_{10} sources and oxidative potential across Chinese regions based on CNN-LSTM data preprocessing and receptor model

Qinghe Cai et al.

Correspondence to: Dongqing Fang (fangdongqing@163.com) and Yang Zhang (zhangyang@ucas.ac.cn)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Ablation Study of CNN-LSTM

To validate the contribution of each component in our hybrid architecture, we conducted a comprehensive ablation study, with results detailed in Table S1. The ablation experiments systematically removed or modified key components of the model to assess their individual contributions to the overall performance.

Table S1. Performance Comparison of MAE, RMSE, and R^2 across Model Variants

Model Type	MAE（μg·m^{-3}）	RMSE（μg·m^{-3}）	R^2
CNN	14.7668	19.2806	0.8591
LSTM	14.3056	21.1303	0.8308
CNN-LSTM	12.6705	17.4935	0.8840

The CNN-only model, which utilized convolutional layers without the LSTM component, achieved an MAE of 14.7668 μg·m^{-3}, RMSE of 19.2806 μg·m^{-3}, and R^2 of 0.8591. While this configuration effectively extracted spatial features from the input data, the absence of temporal modeling resulted in a 16.54% increase in MAE compared to the full CNN-LSTM model.

The LSTM-only configuration, utilizing only recurrent layers without convolutional preprocessing, demonstrated an MAE of 14.3056 μg·m^{-3}, RMSE of 21.1303 μg·m^{-3}, and R^2 of 0.8308. Although the LSTM component successfully captured temporal dependencies, the lack of spatial feature extraction capabilities led to suboptimal performance, with a 12.90% increase in MAE compared to the complete model.

The ablation study conclusively demonstrates that the synergistic combination of CNN and LSTM components is essential for achieving optimal performance in PM_{10} gap-filling applications.

S2 Comparison between Northern and Southern sites of OP_v (nmol H_2O_2 ·m^{-3})

China was divided into northern and southern regions using the Qinling Mountains-Huaihe River line as the boundary. This line represents the traditional geographical and climatic divide in China. Sites north of this boundary were classified as northern regions, while sites south were classified as southern regions. Lhasa, located on the Tibetan Plateau, was treated as a separate plateau region due to

its distinct high-altitude environmental characteristics. The detailed site classification is provided in Table S2.

Due to the non-normal distribution of the data, a non-parametric Mann-Whitney U test was conducted to examine regional differences in ROS concentrations. The analysis revealed that OP_v levels were significantly higher in the Northern sites compared to the Southern sites ($U = 168517.0$, $p = 0.025$), with the difference exceeding the threshold for statistical significance ($p < 0.05$).

Table S2. The geographical division corresponding to the station.

Geographic region	Station name
Northern sties	LFS, DL, GC, DH, XA, ZZ
Southern sites	LA, CD, JS, NN, CHD

S3 Source profiles from the PMF

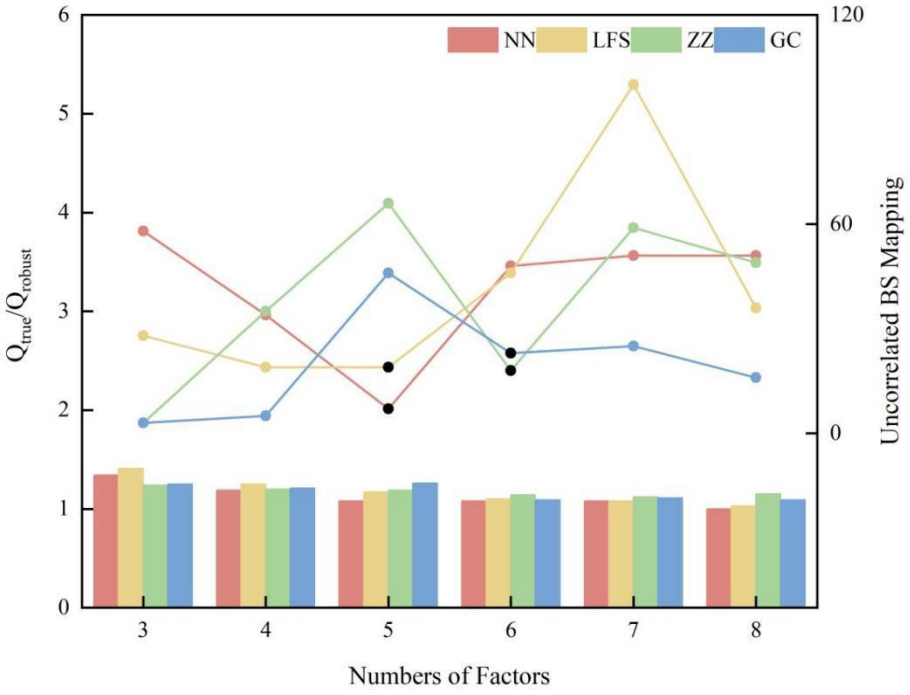

Figure S1. Changes in Q_{true}/Q_{robust} and uncorrelated bootstrap (BS) mapping calculated by the PMF model for varying numbers of factors (3-8). Black dots represent the optimum solution factors at each station.

Table S3. Summary of error estimation diagnostic with PMF at NN, LFS, ZZ and GC station.

Diagnostics		NN	LFS	ZZ	GC
	Number of base run	20	20	20	20
	Q_{robust}	2338	2658.86	2043.86	2066.75
	Q_{true}	2521.1	3160.89	2341.26	2249.16
	Q_{true}/Q_{robust}	1.08	1.18	1.14	1.09
DISP	% dQ	< 0.1 %	< 0.1 %	< 0.1 %	< 0.1 %
	swaps	0	0	0	0
BS	Number of run	100	100	100	100
	Extra modeling uncertainty (%)	0	0	0	0
	Min. Correlation R-Value	0.6	0.6	0.6	0.6
	Mapping	Dust: 94% Biomass burning: 99% Others: 100%	Secondary aerosol: 88% Traffic: 97% Agricultural activities:95% Others: 100%	Coal combustion: 87% Secondary aerosol: 97% Biomass burning: 98% Others: 100%	Agricultural activities: 90% Coal combustion: 85% Others: 100%
BS-DISP	#of Cases Accepted	89	78	80	71
	% cases with swaps	9%	18%	15%	25%