

Emulating CMAQ Simulations with a Conditional U-Net: A Data-Driven Approach for PM2.5 Concentration Prediction

Supplementary materials

Yohan Lee¹, Junghyun Park¹, Jinseok Kim², Jung-Hun Woo², and Jong Hyeon Lee^{1,*}

¹EH Research & Consulting Co., Ltd., E TechHive, 410 Jeongseojin-ro, Seo-gu, 22689, Incheon, Republic of Korea

²Civil and Environmental Engineering, College of Engineering, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, 05029, Seoul, Republic of Korea

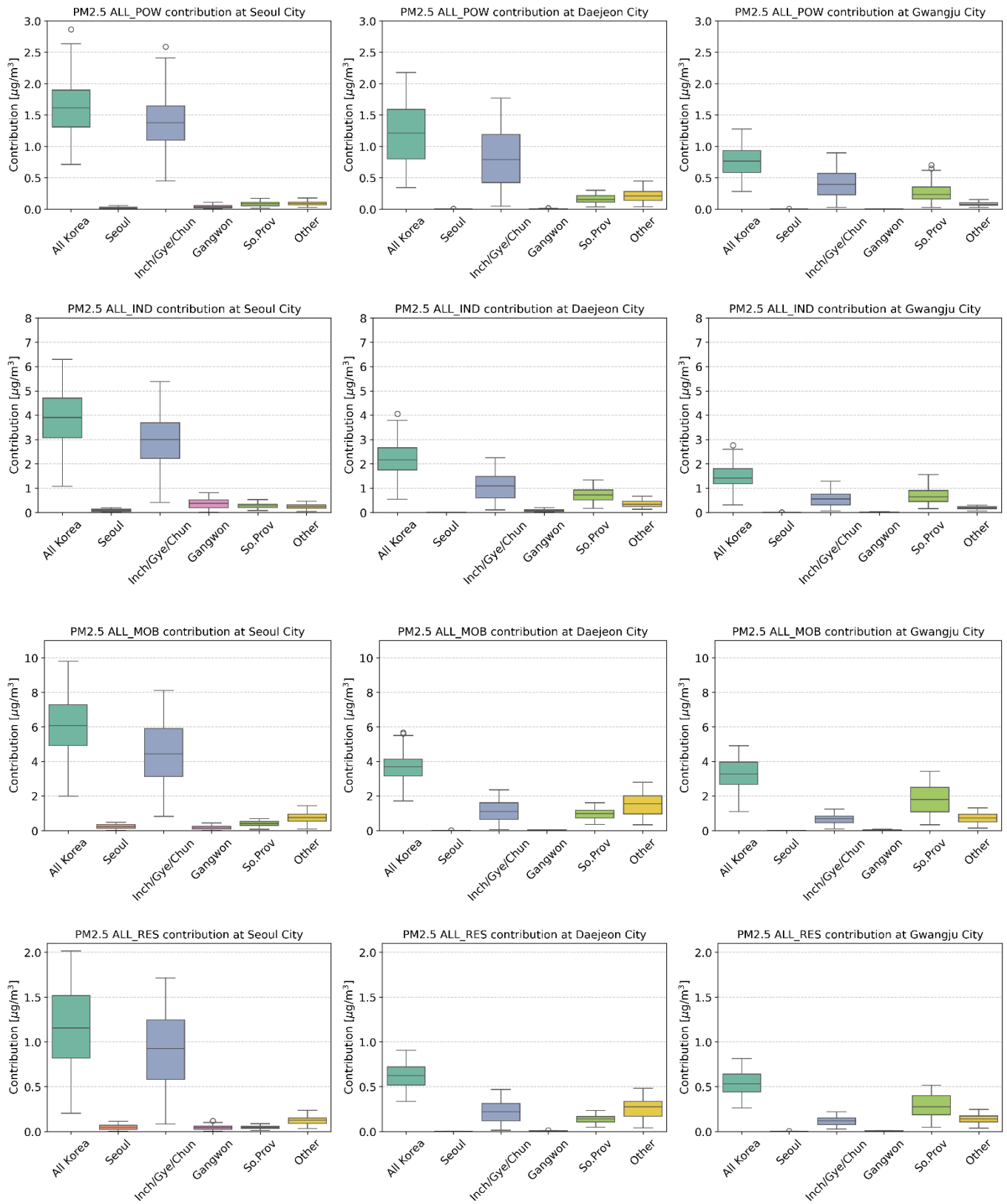
*corresponding author jh.lee@ehrnc.com

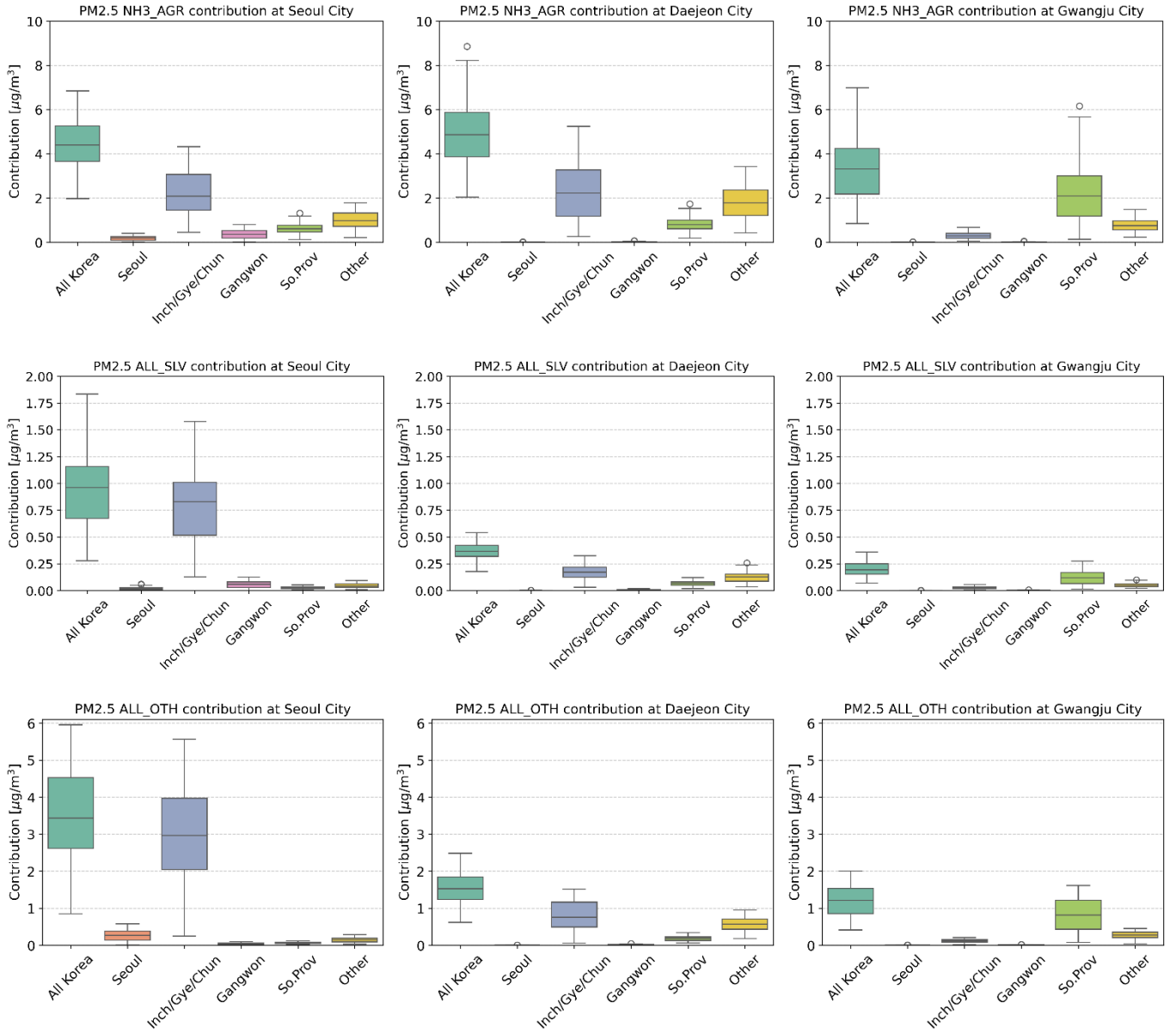
Contents

- S 1.** Table of input control matrix
- S 2.** Contribution analysis of Seoul, Daejeon, Gwangju region for all region and all sector
- S 3.** Emulation results of hourly changes in average PM2.5 at Seoul
- S 4.** Emulation results of hourly changes in average PM2.5 at Daejeon
- S 5.** Emulation results of hourly changes in average PM2.5 at Ulsan
- S 6.** Emulation results of hourly changes in average PM2.5 at Gwangju
- S 7.** Emulation results of random scenario for every grid and every time
- S 8.** Visual comparison of emulation results for April and October season
- S 9.** Video for hourly PM2.5 for every season

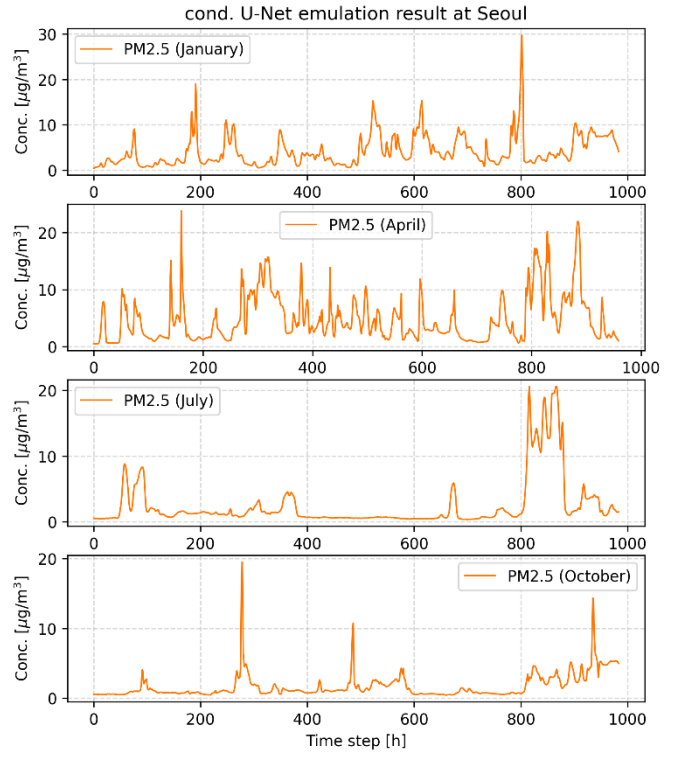
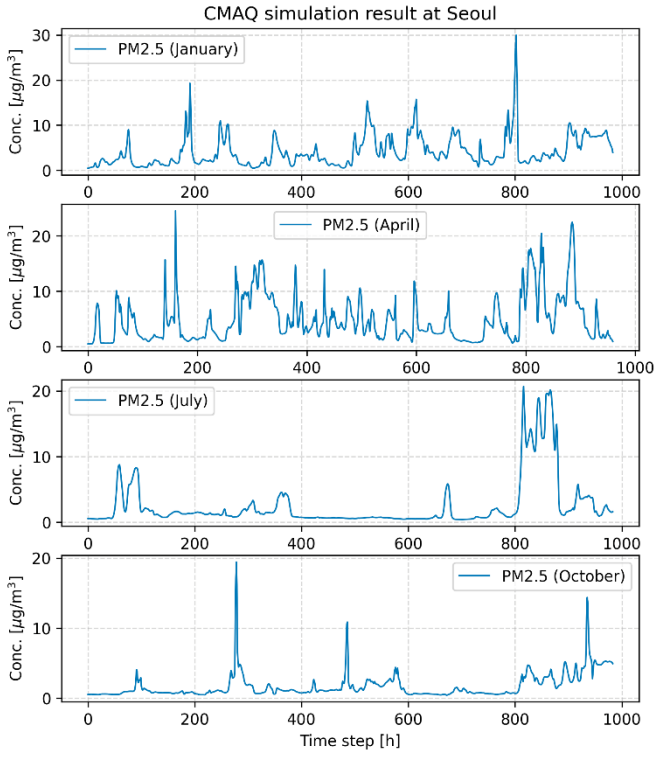
S 1. An actual view of the control matrix used as input to the conditional U-Net. Some values are omitted due to page limitations. S1, S2, ... S119 refers to 119 precursor emission scenarios in the order S1, S2, ... S119. Each row represents the activity of a sector by region, with the first letter representing the region and ALL_POW, ALL_IND, etc. representing the sector. For example, A_ALL_POW means the activity of the power plant (ALL_POW) in Seoul (A). For S1 it means the baseline emissions scenario in 2013. From these scenarios, 59 scenarios are randomly selected for training and the remaining 60 scenarios are used for testing.

	S1	S2	S3	...	S117	S118	S119
A_ALL_POW	1.000	0.548	0.514	...	0.839	0.576	0.858
A_ALL_IND	1.000	1.258	0.928	...	1.031	1.121	0.901
A_ALL_MOB	1.000	0.897	0.945	...	0.550	1.011	1.217
A_ALL_RES	1.000	0.591	0.692	...	0.824	0.554	1.191
A_NH3_AGR	1.000	1.398	1.109	...	1.404	1.000	1.413
A_ALL_SLV	1.000	1.102	1.314	...	1.027	0.853	0.898
A_ALL_OTH	1.000	0.716	0.721	...	1.083	1.390	0.587
B_ALL_POW	1.000	1.267	0.612	...	0.529	1.101	0.675
B_ALL_IND	1.000	1.255	1.087	...	1.128	0.730	0.800
B_ALL_MOB	1.000	1.372	0.949	...	1.494	1.440	0.669
B_ALL_RES	1.000	0.835	0.546	...	1.336	1.328	1.187
B_NH3_AGR	1.000	0.610	0.778	...	1.466	1.107	1.209
B_ALL_SLV	1.000	0.963	1.038	...	0.521	1.175	1.303
B_ALL_OTH	1.000	1.145	0.698	...	1.489	1.338	0.987
C_ALL_POW	1.000	1.073	0.505	...	0.683	1.141	0.974
C_ALL_IND	1.000	1.086	1.364	...	1.333	1.288	1.102
...	...	...	...	...	...	...	...
Q_ALL_POW	1.000	0.958	0.997	...	1.415	1.400	0.799
Q_ALL_IND	1.000	0.583	1.395	...	0.884	0.788	1.309
Q_ALL_MOB	1.000	1.315	0.550	...	1.209	0.752	0.907
Q_ALL_RES	1.000	1.379	0.787	...	0.768	1.184	0.840
Q_NH3_AGR	1.000	1.184	1.343	...	1.432	0.880	1.348
Q_ALL_SLV	1.000	1.116	1.209	...	1.269	1.142	1.300
Q_ALL_OTH	1.000	1.273	0.513	...	0.782	1.398	0.534

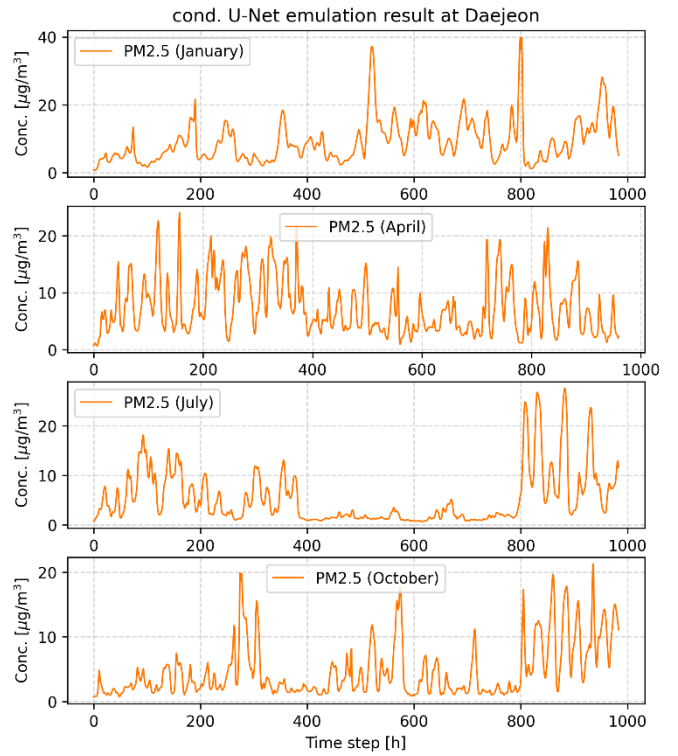
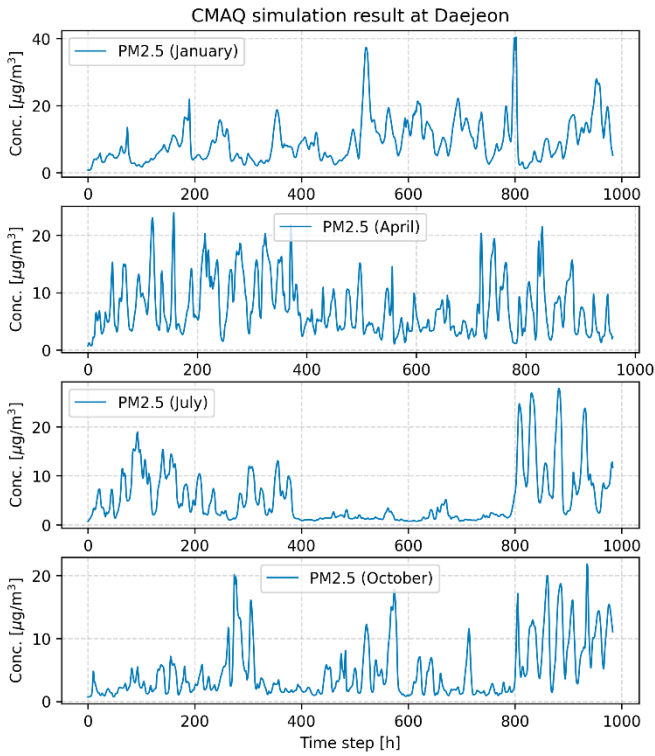




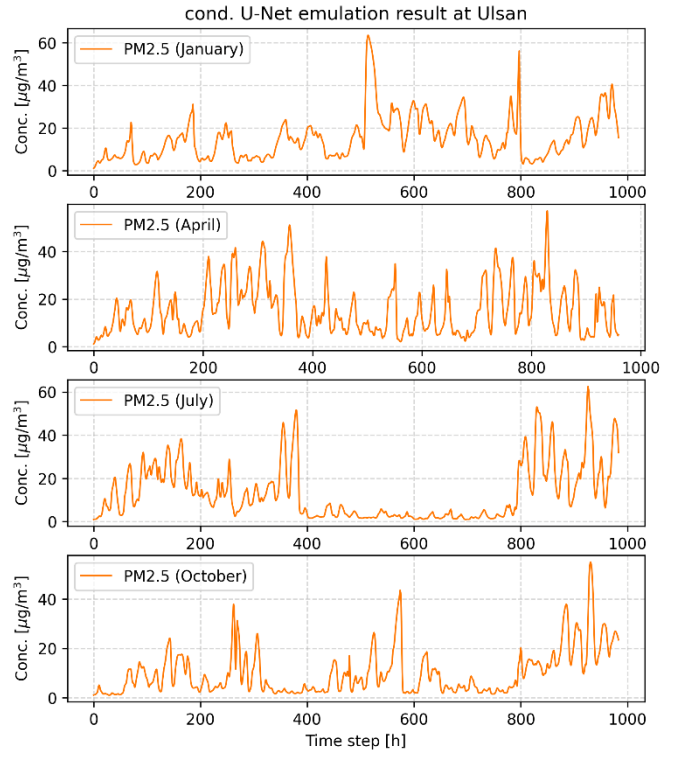
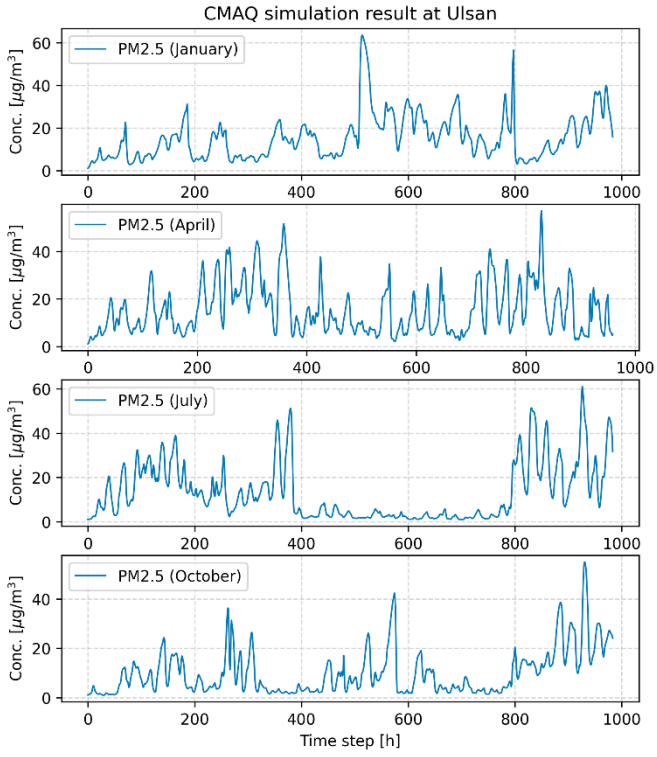
S 2. Figure comparing the contribution (SHAP value) of different sectors to daily PM2.5 concentrations in Seoul, Daejeon and Gwangju. Only the figure for power plants is shown here, but the comparison is also shown for other sectors for reference. Note that the scale of the contribution axis is different.



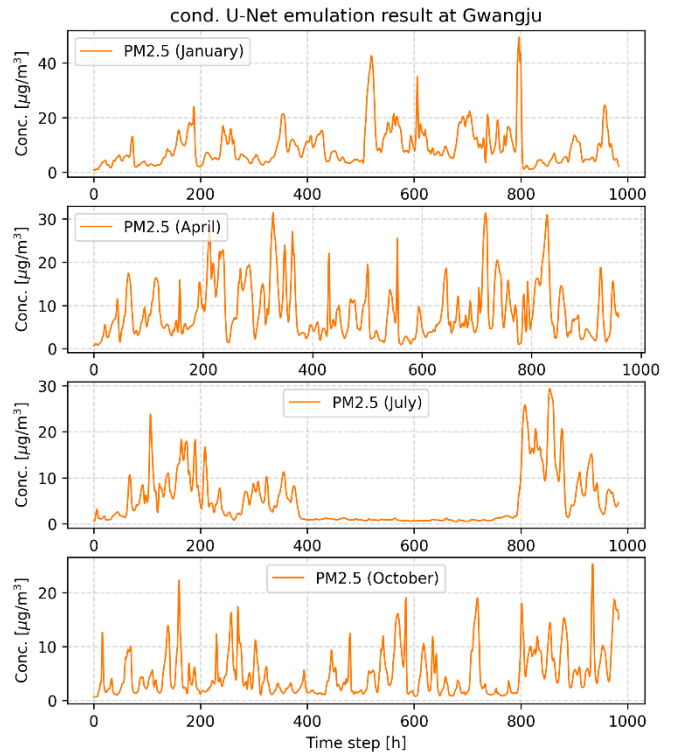
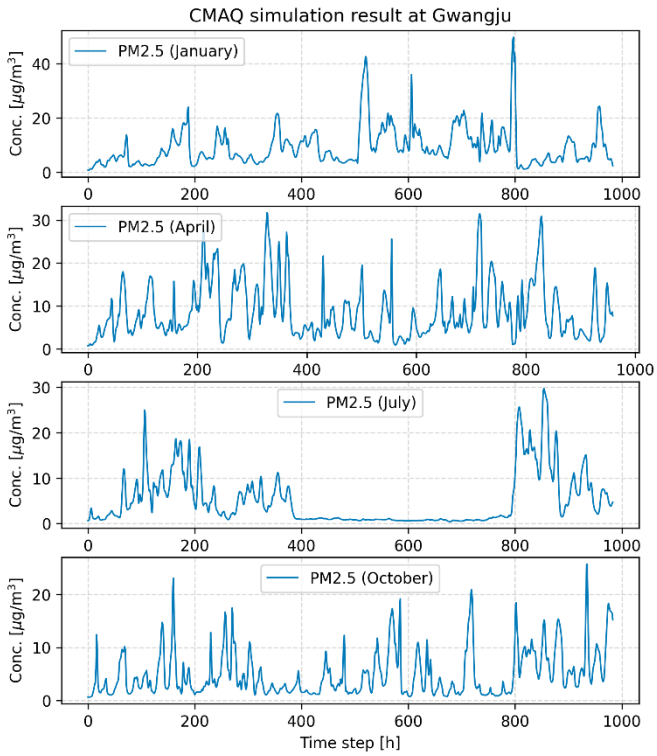
S3. Comparison of emulation results of hourly changes in average PM2.5 in the Seoul area grid for each season.



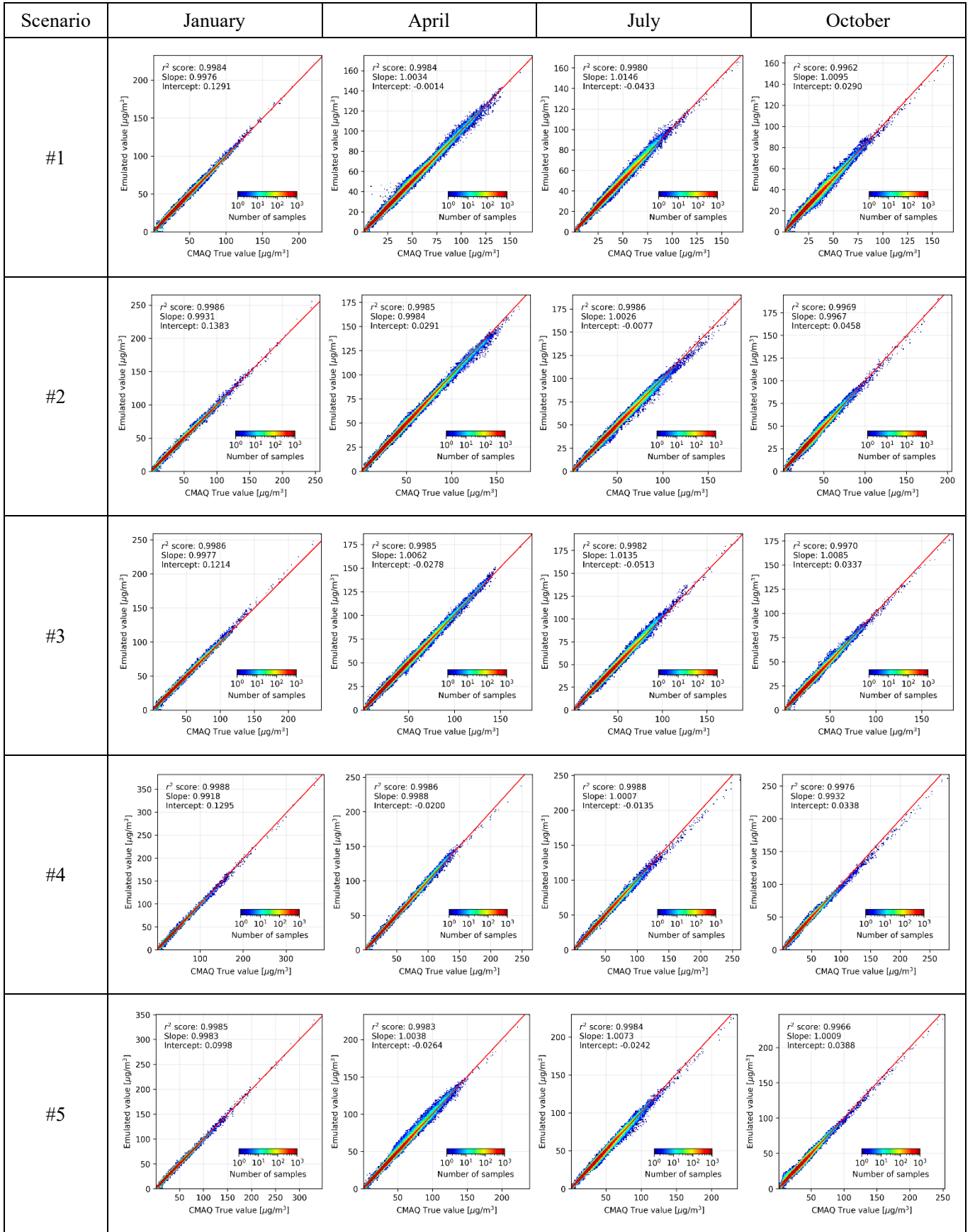
S4. Comparison of emulation results of hourly changes in average PM2.5 in the Daejeon area grid for each season.



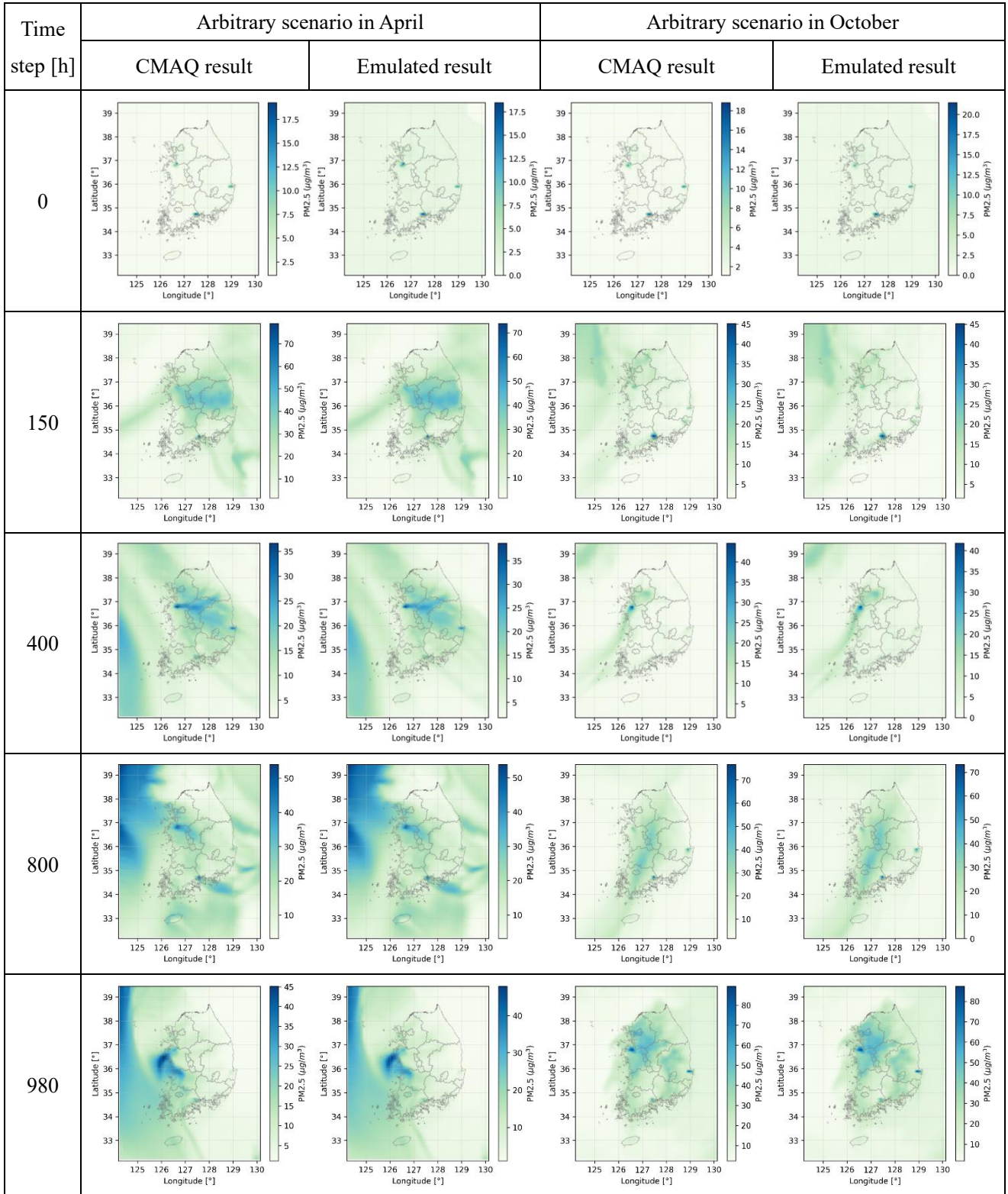
S 5. Comparison of emulation results of hourly changes in average PM2.5 in the Ulsan area grid for each season.



S 6. Comparison of emulation results of hourly changes in average PM2.5 in the Gwangju area grid for each season.



S 7. Emulation results of the conditional U-Net model for five PM_{2.5} concentration scenarios randomly selected from the test data. While there are slight differences in prediction performance for each scenario, the average r^2 performance is around 0.99, indicating that the model predicts the calculation of CMAQ with high accuracy.



S 8. Figure comparing CMAQ simulation results with our model's emulate results for an arbitrary scenario. April and October results are plotted, with the CMAQ results (true value) on the left and the conditional U-Net predictions (predicted value) on the right of each seasonal column. Each row represents a time step in hours, comparing PM2.5 concentrations at the 0, 150, 400, 800, and 980th hour.