

Response to the comment on “Meteorological influence on surface ozone trends in China: Assessing uncertainties caused by multi-dataset and multi-method” by X. Wang et al. (Ms. Ref. No.: EGUSPHERE-2025-1880)

Response to Referee #1

I thank the authors for their time in responding to my comments. Regarding their response to the question about the $R^2 > 0.5$ criterion, I appreciate the need to balance model skill and spatial coverage. However, it would strengthen the justification if the authors could show how results change when this threshold is tightened or relaxed. For example, a table in the supplementary could include attribution results when the currently excluded stations are reinstated or when a higher bar such as $R^2 > 0.6$ is applied (more in line with the skill of models in Weng et al., 2022). This would allow assessment of whether lower-skill stations systematically bias the meteorology–emissions separation, dampen trends, or increase variability in the CV metric. Such a sensitivity analysis would also align with best practice in attribution studies, where robustness to model-skill thresholds is important for ensuring that conclusions are not an artefact of an arbitrary cutoff.

Response:

Thanks for your insightful suggestion. We fully agree that evaluating the robustness to different model-skill thresholds is essential for ensuring the reliability of our conclusions. Following the reviewer’s suggestion, we have conducted a comprehensive sensitivity analysis by implementing two additional R^2 thresholds ($R^2 \geq 0.6$ and $R^2 \geq 0.4$) to filter unreliable stations across all seasons and regions. The complete results are presented in Table S3 and summarized as follows:

- 1) Tightening the threshold to $R^2 \geq 0.6$ reduced the number of available stations by 13.5% (autumn) to 39.4% (summer) compared to $R^2 \geq 0.5$, while relaxing to $R^2 \geq 0.4$ increased station inclusion by 4.1% (autumn) to 15.5% (summer).
- 2) Meteorology-driven MDA8 O₃ trends derived from the MLR, RF, and GC models exhibited minor variations across the three thresholds. The maximum difference was observed in the MLR model for summer in China, with trends of +0.23 ppb yr⁻¹ ($R^2 \geq 0.5$), +0.31 ppb yr⁻¹ ($R^2 \geq 0.6$), and +0.20 ppb yr⁻¹ ($R^2 \geq 0.4$). The overall directional consistency of meteorology-driven O₃ trends across thresholds confirms that our primary conclusions are not artifacts of an arbitrary cutoff.
- 3) The uncertainty metric (CV) for multi-method spread showed minor changes

across R^2 thresholds, indicating that methodological uncertainties are robustly quantified regardless of the station inclusion criteria.

We have added the following statements to the manuscript:

"To evaluate the robustness of the $R^2 \geq 0.5$ criterion, we performed sensitivity analyses using thresholds of $R^2 \geq 0.6$ and $R^2 \geq 0.4$, to ensure that our conclusions are not an artifact of an arbitrary cutoff (Table S3)." *[Lines 171–173 in the tracked-changes version of the revised manuscript]*

"Meteorology-driven MDA8 O₃ trends exhibited minor variations across different R^2 thresholds (Table S3), indicating that the trends are not an artifact of an arbitrary cutoff." *[Lines 335–336 in the tracked-changes version of the revised manuscript]*

"The low uncertainties are further corroborated by consistent CV estimates derived under different RF's R^2 thresholds (Table S3). " *[Lines 344–345 in the tracked-changes version of the revised manuscript]*

Table S3. Comparison of meteorology-driven MDA8 O₃ trends (ppb yr⁻¹) derived from multiple linear regression (MLR), random forest (RF), and GEOS-Chem (GC) models using different model performance criteria (the R² of the RF model ≥ 0.5, 0.6, and 0.4) across seasons and regions.

Season	R ²	Available stations ^a	China				BTH				YRD				PRD			
			MLR	RF	GC	CV ^b	MLR	RF	GC	CV	MLR	RF	GC	CV	MLR	RF	GC	CV
Spring	R ² ≥ 0.5	1162	+0.48	+0.15	+0.14	0.75	+0.26	-0.01	+0.12	1.06	+0.80	+0.25	+0.37	0.61	+0.94	+0.78	+0.77	0.11
	R ² ≥ 0.6	882	+0.50	+0.14	+0.14	0.81	+0.26	-0.00	+0.12	1.07	+0.86	+0.25	+0.37	0.66	+0.89	+0.75	+0.77	0.10
	R ² ≥ 0.4	1250	+0.47	+0.16	+0.14	0.73	+0.26	-0.01	+0.12	1.06	+0.77	+0.25	+0.36	0.60	+0.89	+0.75	+0.77	0.09
Summer	R ² ≥ 0.5	1059	+0.23	+0.01	-0.03	2.00	+0.03	-0.16	-0.18	1.08	+0.45	+0.25	+0.40	0.29	+0.07	+0.13	+0.10	0.32
	R ² ≥ 0.6	642	+0.31	+0.02	-0.01	1.69	+0.06	-0.20	-0.17	1.40	+0.51	+0.21	+0.34	0.42	+0.08	+0.13	+0.10	0.25
	R ² ≥ 0.4	1223	+0.20	+0.00	-0.04	2.39	+0.04	-0.15	-0.19	1.24	+0.43	+0.25	+0.39	0.27	+0.06	+0.13	+0.09	0.34
Autumn	R ² ≥ 0.5	1203	+0.15	+0.34	+0.03	0.88	-0.27	+0.19	-0.13	3.38	+0.37	+0.53	+0.24	0.38	+0.83	+0.81	+0.53	0.23
	R ² ≥ 0.6	1041	+0.18	+0.36	+0.04	0.85	-0.27	+0.19	-0.14	3.40	+0.40	+0.55	+0.24	0.39	+0.85	+0.83	+0.54	0.23
	R ² ≥ 0.4	1252	+0.15	+0.33	+0.03	0.90	-0.27	+0.19	-0.13	3.38	+0.36	+0.52	+0.24	0.38	+0.83	+0.81	+0.53	0.23
Winter	R ² ≥ 0.5	1094	+0.30	+0.12	+0.25	0.40	+0.26	+0.13	+0.09	0.55	+0.19	+0.06	+0.27	0.59	+0.72	+0.64	+0.72	0.07
	R ² ≥ 0.6	738	+0.33	+0.13	+0.24	0.41	+0.26	+0.13	+0.10	0.53	+0.20	+0.06	+0.27	0.59	+0.70	+0.66	+0.73	0.05
	R ² ≥ 0.4	1217	+0.28	+0.12	+0.25	0.40	+0.26	+0.13	+0.10	0.55	+0.17	+0.06	+0.27	0.62	+0.70	+0.63	+0.72	0.07

^a “Available stations” denotes the number of state-controlled monitoring stations with the R² of the Random Forest model ≥ 0.5, 0.6, and 0.4.

^b The absolute value of the coefficient of variation (CV) is calculated by the standard deviation of the trends derived from MLR, RF, and GC models divided by the mean.

Minor note: the reference “Wang 2024c” is still incomplete.

Response:

Thanks for your suggestion. We have corrected the reference “Wang 2024c” as follows:

Wang, X., Zhu, J., Li, K., Chen, L., Yang, Y., Zhao, Y., Yue, X., Gu, Y., and Liao, H.:
Meteorology-driven trends in PM_{2.5} concentrations and related health burden over
India, Atmos. Res., 308, 107548, <https://doi.org/10.1016/j.atmosres.2024.107548>,
2024c.

[Lines 623–624 in the tracked-changes version of the revised manuscript]