

## Response to reviewer comments

We thank the three reviewers very much for their helpful comments and suggestions. Addressing these comments has substantially improved the quality of the manuscript. We have provided responses and necessary revisions to each reviewer comment below.

### Reviewer #1

This manuscript “Assessing the causal impact of the Chinese Spring Festival on PM<sub>2.5</sub> air quality in Beijing-Tianjin-Hebei and surrounding region using a machine learning counterfactual modeling approach” by Yuan Li and team, addresses an important and interesting topic: the influence of the Chinese Spring Festival (CSF) on regional PM<sub>2.5</sub> concentrations, particularly the attribution of emissions to fireworks. The use of a machine learning counterfactual model is an interesting approach to isolate the festival’s effect. However, the core conclusions regarding the high contribution of fireworks, especially at the regional scale, are based on data and methodological interpretations that lack sufficient resolution and rigor to justify the claim. Specifically, the analysis appears to conflate highly local, transient firework plumes with persistent regional emissions from industrial and urban sources. This weakness must be addressed before the manuscript can be considered for publication.

Major Comments:

#### Comments #1:

The manuscript interprets elevated specific chemical tracers (e.g., Al, Ba, Cu and K) as fireworks and then equates the PMF-resolved “fireworks” factor with a very large fraction to city PM<sub>2.5</sub> (e.g., fireworks reaching 137  $\mu\text{g m}^{-3}$  and up to 81.8% at the peak) and then treats that as validation of the ML holiday-attributable signal. But trace metals from fireworks can adsorb or be scavenged onto pre-existing aerosol mass—a single atom/molecule of a tracer attaching to many aerosol mass units can cause PMF to label existing background aerosol as “firework-contaminated” even if the mass contribution from fireworks is small. This overlooks the critical aspect of aerosol mixing and coating. The presence of a fireworks marker on an aerosol particle only indicates mixing or coating has occurred, not that the particle’s mass is primarily sourced from fireworks. For instance, a persistent industrial Black Carbon (BC) or organic carbon particle, when mixed in the atmosphere with even a single molecule of Al/Na/Cu or K from fireworks, will be registered a ‘firework’ marker. But in reality, the particle’s overall mass and

37 emission may still be dominated by the pre-existing industrial or alternative sources of  
38 BC core. Would this not induce an overestimation of mass contribution of fireworks  
39 and biases the source apportionment. I think the author's revise their interpretation to  
40 differentiate between a source signature (the tracer) and mass contribution (the overall  
41 particle composition). One approach to distinguish this is maybe to examine the ratios  
42 of Ba/Al, K/Ba or K/EC or K/BC and their co-variation with EC or BC to see whether  
43 the tracers appear as spikes superimposed on otherwise continuous mass.

44

45 **Response:**

46 This is an astute observation. We sincerely thank the reviewer for this insightful  
47 comment regarding the potential overestimation of firework-related mass contributions.  
48 This comment prompted us to conduct a deeper analysis of elemental ratios, the results  
49 of which robustly support that the firework factor resolved by DN-PMF represents a  
50 genuine mass contribution rather than an artifact of tracer adsorption.

51 Indeed, fresh particles released from emission sources are subject to atmospheric  
52 processes such as dilution, deposition, adsorption, and coating. Receptor models are  
53 unable to explicitly resolve these ageing processes, as they rely on reverse analysis  
54 based on observational data. PMF decomposes the observed concentrations of chemical  
55 species into a factor profile matrix and a corresponding source contribution matrix  
56 based on the covariance structure of the measured species.

57

58 In practice, the interpretation of a "factor" as a "source" is not determined by the  
59 absolute abundance of individual "markers" or "tracers", but rather by the relative  
60 abundances of species within a given factor. Accordingly, the factor characterized by  
61 relatively high contributions of Al, Ba, Cu, K, and some EC was identified as firework  
62 emissions. As examined through DISP analysis in the PMF framework, the DISP bands  
63 for these chemical species remain stable, suggesting that their ratios are largely  
64 preserved. This stability enhances our confidence in the identification of the firework  
65 emission source.

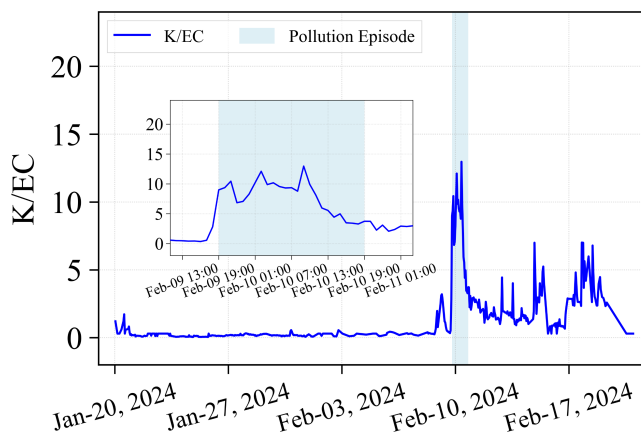
66 As suggested by the reviewer, we further investigated temporal variations in elemental  
67 ratios of ambient PM<sub>2.5</sub>:

68 (1). The K/EC ratio demonstrates the introduction of a distinct "K-rich, EC-poor"  
69 particle population. During the pre-holiday period (21 January - 8 February), the K/EC  
70 ratio remained stable at a low level of ~0.2, reflecting a steady state of regional  
71 background aerosols. In contrast, the New Year's Eve pollution episode was marked by  
72 an abrupt, order-of-magnitude jump in the K/EC ratio, with peak values nearing 13 -  
73 more than 60 times higher than the background level. This sharp, detached peak  
74 provides clear evidence for the introduction of a distinct aerosol population  
75 characterized by its "K-rich and EC-poor" signature. This finding directly refutes the

76 “tracer adsorption” hypothesis. If potassium from fireworks were merely adsorbing  
77 onto pre-existing background particles (e.g., industrial BC), we would expect a modest  
78 and gradual elevation of the K/EC ratio, not such a dramatic spike.

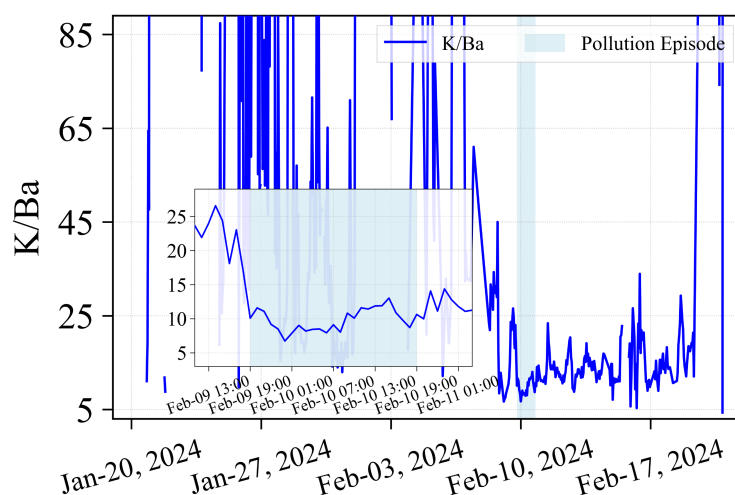
79 (2). The K/Ba ratio confirms a stable and unique chemical fingerprint for this new  
80 source. While the K/Ba ratio exhibited large fluctuations during the background period,  
81 indicating complex daily sources, it formed a highly stable pollution level during the  
82 New Year’s Eve event window (average ~10, coefficient of variation 16%). This distinct  
83 transition from a period of highly variable baseline concentrations to one of markedly  
84 stable pollution level is a robust empirical indicator of a strong, independent emission  
85 source with a uniform chemical composition, the signature of which is highly consistent  
86 with fireworks.

87 In sum, the two independent lines of evidence from elemental ratios form a coherent  
88 narrative. They unambiguously demonstrate that the “fireworks” factor resolved by  
89 DN-PMF represents a transient, strong source that emerged abruptly during the New  
90 Year’s period and constituted a major contribution to PM<sub>2.5</sub> mass, rather than a result of  
91 random mixing or adsorption of emissions from different sources. We have  
92 incorporated this analysis of elemental ratios into Section 3.2 (lines 244-253) of the  
93 revised manuscript.



94  
95 **Figure S6.** A dramatic surge in the K/EC ratio of PM<sub>2.5</sub> during the 2024 New Year’s  
96 Eve: Evidence for firework emissions in Hangzhou.

97



98  
99 **Figure S7.** Temporal variation of the K/Ba ratio in PM<sub>2.5</sub> during the 2024 New Year's  
100 Eve haze episode in Hangzhou.

101  
102  
103 **Comments #2:**

104 Multi-platform observations of persistent BC and other aerosols in the adjacent regions  
105 - such as Xuzhou and Suzhou, which share air-mass pathways with Shandong and  
106 Linyi- reveal substantial, continuous emissions from urban and small industrial sources  
107 (Tiwari et al., 2025). These background sources have far greater emission potential than  
108 transient firework events. The manuscript should reconcile its high firework attribution  
109 with these documented, regional emissions, placing CSF findings within the broader  
110 context of persistent regional aerosol loading. This also demands a clarification to the  
111 geographic scope of the conclusions. A localized short duration fireworks plume from  
112 Shandong is unlikely to represent a regional-scale PM<sub>2.5</sub> driver across the entire  
113 Beijing-Tianjin-Hebei domain. The authors should specify whether their results  
114 describe a local fireworks effect or a regional influence and adjust the claims  
115 accordingly.

116  
117 **Response:**

118 We sincerely thank the reviewer for this insightful comment and fully agree that the  
119 Beijing-Tianjin-Hebei (BTH) region and its surroundings are characterized by a  
120 substantial and persistent regional background pollution field originating from  
121 industrial and urban activities, whose long-term emission potential undoubtedly far  
122 exceeds that of short-lived firework events. The study by Tiwari et al. (2025) provides  
123 strong support for this perspective. This reference has now been included in the revised  
124 manuscript (Line 416).

125  
126 We would like to clarify that the core findings of our study are not in conflict with this

127 understanding but rather serve as an important complement to it. Our objective is not  
128 to challenge the annual or long-term significance of these background sources but to  
129 reveal a fundamental shift in the instantaneous pattern of source contributions during  
130 the specific time window of the Chinese New Year’s Eve. Specifically:

131

132 (1). Our model and aerosol composition data demonstrate that firework emissions acted  
133 as an extremely intense transient pulse, superimposed upon the persistent regional  
134 background. During the evening of New Year’s Eve until the early morning of the  
135 following day, the intensity of this pulse was so substantial that it dramatically altered  
136 the chemical signature of the near-surface atmosphere (e.g., the K/EC ratio surged by  
137 over 60-fold) and dominated the PM<sub>2.5</sub> mass concentration within just a few hours. This  
138 represents a brief but dramatic short-term high pollution exposure driven by unique  
139 societal activities.

140

141 (2). Regarding the geographical scope, this was not a case of an isolated plume from  
142 Shandong influencing the entire BTH region via long-range transport. Instead, it  
143 resulted from the near-synchronous collective burning by countless communities across  
144 the BTH urban agglomeration itself. Under the regionally unfavorable meteorological  
145 conditions (such as stagnant winds and a low mixing layer height), these widespread,  
146 simultaneous local emissions led to a regionally concurrent pollution outbreak. This is  
147 evidenced by the abrupt increase in PM<sub>2.5</sub> concentrations observed in the vast majority  
148 of the “2+26” cities in the BTH and surrounding region during the New Year’s Eve (See  
149 our supplementary Fig. S10). Thus, the driving mechanism was widespread and  
150 synchronous local emission, not point-source transport.

151

152 Following the reviewer’s suggestion, we have added a discussion in Section 3.5 (lines  
153 415-420). This discussion situates our findings within the broader context of the  
154 regional background pollution and refines the wording of our conclusions to explicitly  
155 state that this was a “*transient extreme pollution episode driven by widespread local  
156 burning and superimposed on the regional background*”. We believe this revision  
157 makes our conclusions more precise and highlights the value of our study in  
158 understanding the impact of special emission events on regional air quality.

159

### 160 **Comments #3:**

161 The ML counterfactual approach is trained and evaluated on city-averaged PM<sub>2.5</sub> (14  
162 monitoring stations averaged for Hangzhou; Fig. 3 caption), but does the DN-PMF  
163 chemical composition used for validating the fireworks factor comes from a single site  
164 (Wolongqiao). Is the manuscript assuming the single-site composition being  
165 representative of the whole-city PM<sub>2.5</sub> composition when assessing fireworks

166 contribution. Please clarify and justify representativeness, if possible via spatial  
167 correlation of species across stations or multi-site composition if available. If the  
168 differences are large, the validation claim should be downweighted or additional PMF  
169 runs may be necessary for other sites.

170

171 **Response:**

172 We thank the reviewer for bringing this point to our attention. We acknowledge that,  
173 ideally, multi-site chemical speciation data would be used to validate model results at  
174 the city-wide scale. Unfortunately, due to the limited availability of chemical speciation  
175 monitoring data, this study could only obtain such data from the single Wolongqiao site  
176 located in downtown Hangzhou. Nonetheless, we believe that data from this single site  
177 can still provide valuable insights for validating the fireworks signal during the Spring  
178 Festival.

179 Note that festival-related fireworks emissions are not localized point sources but occur  
180 as an area-wide activity, taking place nearly simultaneously across the entire city on  
181 New Year's Eve. That is why we use city-averaged PM<sub>2.5</sub> to estimate the holiday effects.  
182 Consistently, the city-level PM<sub>2.5</sub> concentrations (averaged across 14 sites) derived  
183 from our machine learning model exhibit a sharp and synchronous peak during this  
184 period. This behavior indicates that fireworks-related pollution is a city-wide  
185 atmospheric phenomenon rather than a localized event confined to specific  
186 neighborhoods.

187 Our use of the Wolongqiao site chemical speciation data is not intended as an isolated  
188 or absolute "ground truth", but rather as high time-resolution chemical observational  
189 evidence. The chemical fingerprints observed at this site, characterized by pronounced  
190 peaks in K, Ba, and Cu, along with a sharp change in the K/EC ratio, are perfectly  
191 aligned in time with the pollution peak identified by the machine learning model using  
192 city-wide averaged concentrations from 14 sites. This strong temporal consistency and  
193 physicochemical coherence provide compelling evidence that both approaches are  
194 detecting the same city-scale fireworks pollution event.

195

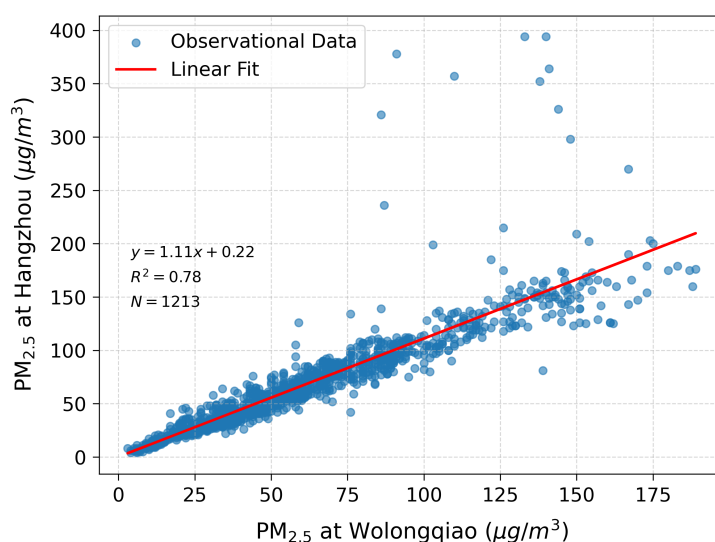
196 To explicitly address the reviewer's concerns regarding spatial representativeness, we  
197 conducted a linear regression analysis between hourly PM<sub>2.5</sub> concentrations at the  
198 Wolongqiao site and the city-wide average (mean of 14 national monitoring stations)  
199 over the entire study period (December 20, 2023, to February 20, 2024). The results  
200 show that the Wolongqiao site exhibits a strong positive correlation with the city-wide  
201 average ( $R^2 = 0.78$ ,  $N = 1213$ , slope = 1.11). This high correlation indicates that  
202 temporal variations at Wolongqiao are highly consistent with the overall PM<sub>2.5</sub> pollution  
203 trends across Hangzhou, particularly during winter haze episodes and the Spring  
204 Festival period. Although the slope (1.11) suggests that concentrations at Wolongqiao

205 are slightly lower than the city-wide average, the robust linear relationship  
206 demonstrates that this single site captures the same regional pollution accumulation and  
207 dissipation processes as the rest of the city. It is therefore reasonable to assume that the  
208 fireworks-related source signature identified at this site is representative of the regional  
209 pollution plume.

210

211 Following the reviewer’s suggestion, we have explicitly addressed this limitation in  
212 Section 3.4 (lines 357-370) of the manuscript. Future research will undoubtedly benefit  
213 from an expanded PM<sub>2.5</sub> composition monitoring network to further validate this  
214 hypothesis, if such network data is availability.

215



216

217 **Figure S9.** Linear regression analysis of hourly PM<sub>2.5</sub> concentrations between the  
218 Wolongqiao site (x-axis) and the Hangzhou city-wide average (14 sites, y-axis) during  
219 the study period (2023/12/20 - 2024/2/20). The strong correlation ( $R^2 = 0.78$ ) supports  
220 the representativeness of the Wolongqiao site in capturing city-scale pollution trends  
221 during the observation period.

222

223

224

225 **Comments #4:**

226 The authors attribute  $ML > DN$ -PMF magnitude difference partly to “unproportionally  
227 amplification effect due to midnight affect due to the midnight unfavorable dilution  
228 conditions (extremely low VC)”. This is plausible, however, if low VC multiplies  
229 existing aerosol mass, then fireworks mass should be amplified only in concentration  
230 but not in emission strength if fireworks were purely local instantaneous injections.  
231 Thus, clarifying the distinction between concentration amplification vs inferred

232 emission-strength amplification would further enhance the overall understanding of this  
233 section.

234

235 **Response:**

236 We agree that unfavorable meteorological conditions amplify concentrations rather than  
237 emission strengths, and we have clarified this distinction explicitly in Section 3.4 (lines  
238 335-343) of the revised manuscript. Upon re-analysis of the whole data, and after  
239 updating the DN-PMF results with the newly resolved fireworks contribution, we find  
240 that the magnitude difference between the ML-based counterfactual estimates and the  
241 DN-PMF source apportionment primarily arises from the difference in spatial  
242 representativeness between the two datasets.

243

244 As pointed out by the reviewer, the ML model quantifies citywide mean deviations in  
245  $PM_{2.5}$  attributable to the holiday effect using observations from 14 stations, whereas the  
246 DN-PMF resolves source-specific contributions at the Wolongqiao site alone. Because  
247 the spatial distribution of fireworks emissions may vary across the city, the citywide  
248 emission-strength estimate inevitably differs from the contribution derived at a single  
249 monitoring site.

250

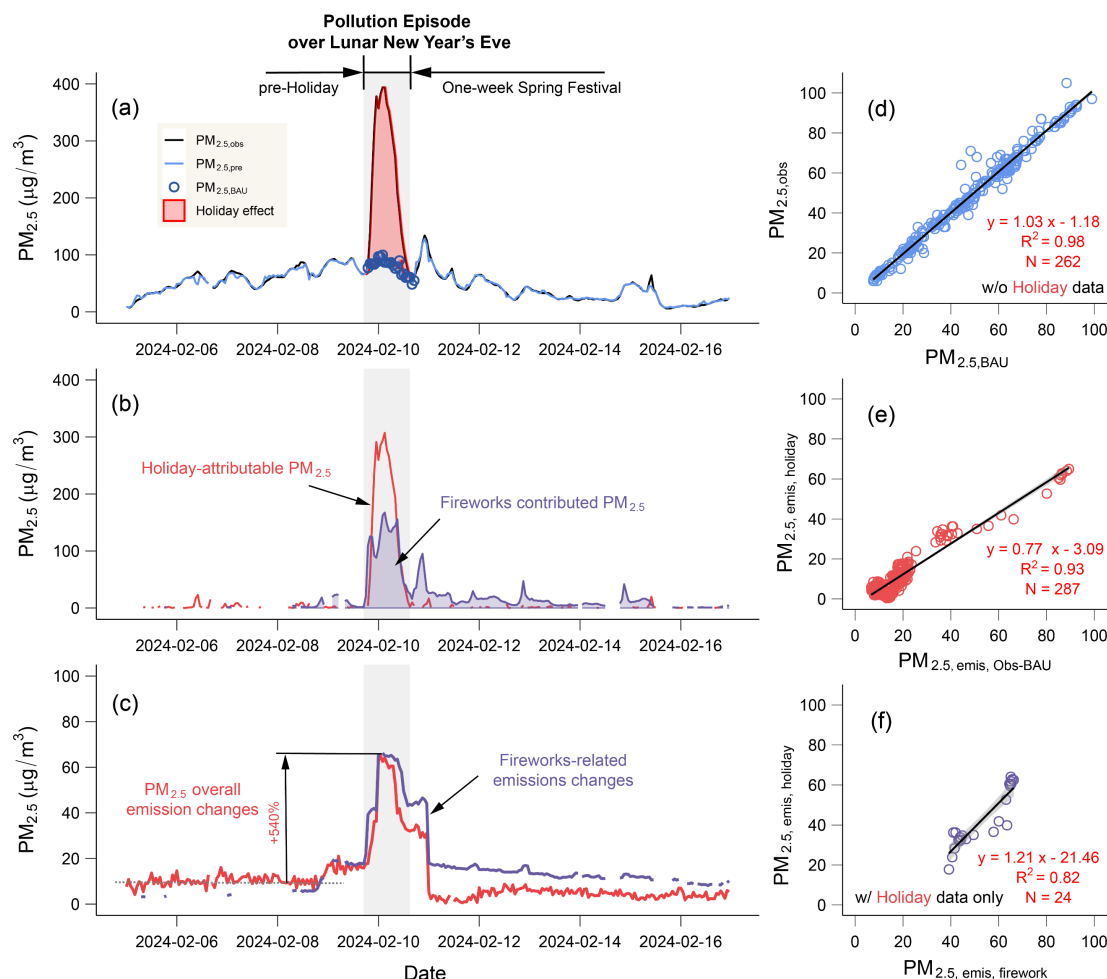
251 Importantly, during the holiday haze period, emissions from industrial activities,  
252 construction, and vehicular traffic were substantially reduced, while fireworks  
253 emissions increased sharply. The ML-derived increase in total  $PM_{2.5}$  emission strength  
254 (maximum +540%; Fig. 3(c)) therefore represents the net effect of both the surge in  
255 fireworks emissions and the reductions in other sectors. In contrast, the DN-PMF  
256 estimate at the single site largely captures the fireworks plume itself. As noted earlier,  
257 New Year's Eve fireworks are not a single-point emission event but rather the combined  
258 outcome of widespread, nearly simultaneous activities across multiple neighborhoods  
259 throughout the city. This explains why the two estimates are close in magnitude and  
260 why the fireworks-specific emission strength changes (a single site) explain 82% of the  
261 total emission-strength (the whole city level) variations with a slope of 1.2 (Fig. 3(f)).  
262 Additionally, the observed 1.1-fold scaling between citywide and site-level  $PM_{2.5}$   
263 indicates that the actual citywide fireworks emission strength would exceed the purple  
264 curve shown in Fig. 3(c).

265

266 Taken together, these analyses confirm that the discrepancy is physically interpretable,  
267 and driven by representativeness differences—not by inconsistencies in the underlying  
268 emission processes. Both approaches independently corroborate that firework  
269 emissions dominated  $PM_{2.5}$  dynamics during the New Year's Eve. We are again grateful  
270 to the reviewer for this comment, which has helped us to clarify and strengthen the

271 interpretation in the revised manuscript.

272



273

274 **Figure 3:** Changes in PM<sub>2.5</sub> air quality and its emission strengths in Hangzhou around  
275 the 2024 New Year's Eve pollution period. (a) Temporal comparison of observed PM<sub>2.5</sub>  
276 (PM<sub>2.5,obs</sub>) with counterfactual predicted PM<sub>2.5</sub> concentrations (PM<sub>2.5,BAU</sub>). (b) Temporal  
277 comparison of firework-related source contributions resolved by DN-PMF with the  
278 holiday-attributable PM<sub>2.5</sub> air quality changes estimated using the ML approach. (c)  
279 Temporal comparison between the meteorologically normalized holiday-attributable  
280 PM<sub>2.5</sub> and fireworks-related PM<sub>2.5</sub>. (d) Scatter plot for the PM<sub>2.5,obs</sub> and the XGBoost  
281 model predicted PM<sub>2.5</sub> concentrations (PM<sub>2.5,BAU</sub>) for non-pollution periods. (e)  
282 Correlation between the differences in meteorologically normalized PM<sub>2.5,obs</sub> and  
283 PM<sub>2.5,BAU</sub> (PM<sub>2.5, emis, obs-BAU</sub>) and meteorologically normalized PM<sub>2.5</sub> for holiday-  
284 attributable part (PM<sub>2.5, emis, holiday</sub>). (f) Scatter plot for the meteorologically normalized  
285 holiday-attributable PM<sub>2.5</sub> and fireworks-related PM<sub>2.5</sub>. Note: the PM<sub>2.5</sub> concentration  
286 data were averaged from 14 national monitoring sites within Hangzhou, while the  
287 chemical composition data were measured at a single site.

288

289 **Comments #5:**

290 In Sect. 3.3-3.5 the authors compute fireworks contribution on the first day as “increase  
291 relative to counterfactual, under assumption that other sources remained unchanged.”  
292 This is a very strong assumption: other sources may decline (traffic reductions) or  
293 increase (localized cooking, residential heating). The manuscript needs to either: (a)  
294 quantify the extent of other source changes using available proxies (traffic counts,  
295 NO<sub>2</sub>/CO changes, DN-PMF vehicle factor), or (b) present their fireworks fraction as an  
296 upper/lower bound with explicit caveats. Right now the wording implies more certainty  
297 than justified.

298

299 **Response:**

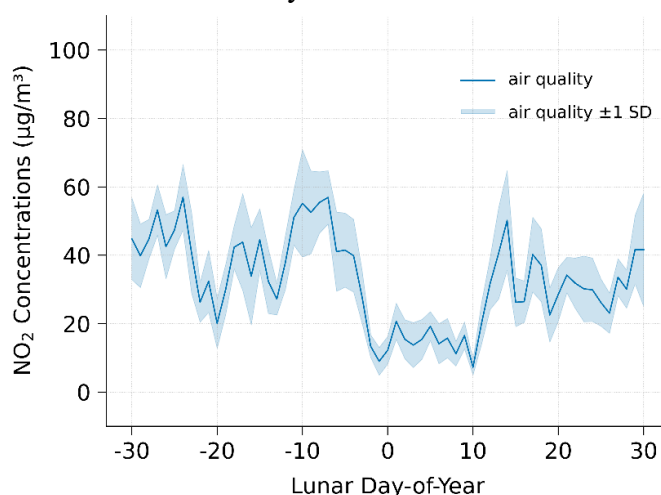
300 We sincerely thank the reviewer for raising this critical point. We agree that the  
301 assumption of “other emission sources remaining unchanged” is indeed a strong one,  
302 and that its validity must be examined when extending the analytical framework to the  
303 “2+26” cities. Following the reviewer’s suggestion, we quantified the temporal  
304 variations of key gaseous pollutants in this region during the 2025 Spring Festival  
305 period (lines 400-408). We found that the results not only challenge the assumption of  
306 unchanged emissions, but also reinforce the conclusion that fireworks were the  
307 dominant driver of the PM<sub>2.5</sub> episode (see Figs. S11 and S12).

308 Our analysis reveals several robust patterns: NO<sub>2</sub> exhibited a sharp regional decline on  
309 Chinese New Year’s Eve (28 January 2025). Its regional-average concentration dropped  
310 by approximately 62.7% relative to the pre-holiday week (21-27 January), decreasing  
311 from 32.8 µg/m<sup>3</sup> to 12.2 µg/m<sup>3</sup>. This provides clear quantitative evidence of  
312 substantially reduced traffic activity during the holiday. Additionally, CO, a tracer of  
313 incomplete combustion, showed a synchronous decline with NO<sub>2</sub>. Its regional-average  
314 concentration decreased by about 44.2% (from 0.90 mg/m<sup>3</sup> to 0.50 mg/m<sup>3</sup>) on New  
315 Year’s Eve. The simultaneous sharp drops in NO<sub>2</sub> and CO collectively indicate a  
316 substantial weakening of emissions from traffic, industry, and other routine fossil-fuel-  
317 related sources during this period. Our previous study examined changes in emissions  
318 of major air pollutants in Chinese cities during the Spring Festival from 2015 to 2021  
319 and found significant sudden reductions in NO<sub>2</sub> and CO after the beginning of the  
320 Lunar New Year as well (<https://doi.org/10.1029/2021GL093403>). These observations  
321 lead to a key inference that strongly supports our central conclusion:

322 Despite the pronounced reductions in conventional anthropogenic sources (as  
323 evidenced by the decreases in NO<sub>2</sub> and CO), PM<sub>2.5</sub> concentrations exhibited an  
324 explosive increase on New Year’s Eve. This striking contrast suggests the presence of  
325 an independent, exceptionally strong, short-lived emission source that was capable of  
326 fully offsetting the reductions in these conventional source emissions and driving the  
327 abrupt PM<sub>2.5</sub> surge. Based on observational evidence, this source is most plausibly  
328 attributed to fireworks.

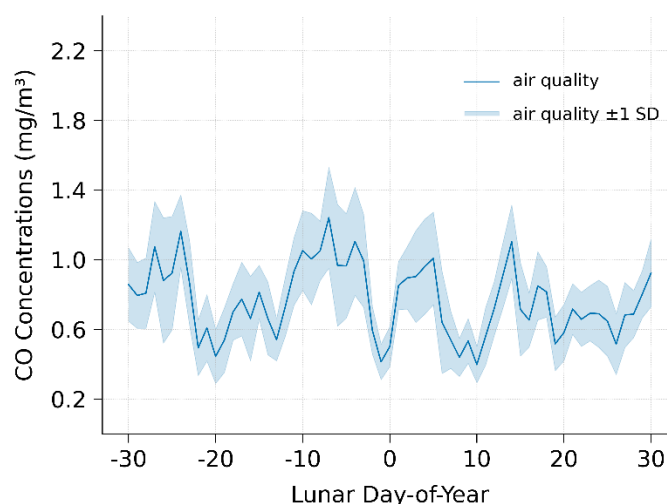
329 The above analysis indicates that the fireworks contribution estimated under the  
330 assumption of “unchanged emissions” is, in fact, conservative and likely  
331 underestimated. If traffic-related and fossil-fuel-related emissions were substantially  
332 weakened in reality, but the model assumes they remained unchanged (and thus  
333 stronger), the resulting mass-balance allocation will systematically under-assign PM<sub>2.5</sub>  
334 mass to the fireworks source. Therefore, the actual weakening of background  
335 anthropogenic emissions does not undermine our conclusion; rather, it implies that the  
336 reported fireworks contribution represents a lower bound and the true contribution is  
337 likely higher.

338 Accordingly, we have revised Section 3.5 to include detailed analyses of NO<sub>2</sub> and CO  
339 variations in the “2+26” region during the 2025 Spring Festival, and we now explicitly  
340 state that the fireworks contribution derived from the machine-learning-based  
341 counterfactual framework should be interpreted as a conservative lower-bound estimate.  
342 We again thank the reviewer for this valuable comment, which has strengthened the  
343 regional-scale robustness and credibility of our conclusions.



344  
345 **Figure S11.** Temporal variation of NO<sub>2</sub> concentrations in the “2+26” cities of northern  
346 China around the 2025 spring festival. The x-axis shows the lunar calendar days, with  
347 0 representing Chinese New Year’s Eve and 1 representing the first day of the Lunar  
348 New Year.

349



350

351 **Figure S12.** Temporal variation of CO concentrations in the “2+26” cities of northern  
 352 China around the 2025 spring festival. The x-axis shows the lunar calendar days, with  
 353 0 representing Chinese New Year’s Eve and 1 representing the first day of the Lunar  
 354 New Year.

355

356 **Comments #6:**

357 I think the claim firework explain 68.8% of instantaneous  $PM_{2.5}$  needs more uncertainty  
 358 reporting, to be deemed robust. The DN-PMF uses 8 factors and labels Factor 1 as  
 359 firework because Al/Ba/K peaks. But PMF solutions can be non-unique and sensitive  
 360 to input species and uncertainties. Shedding light on PMF diagnostics would make the  
 361 results more robust.

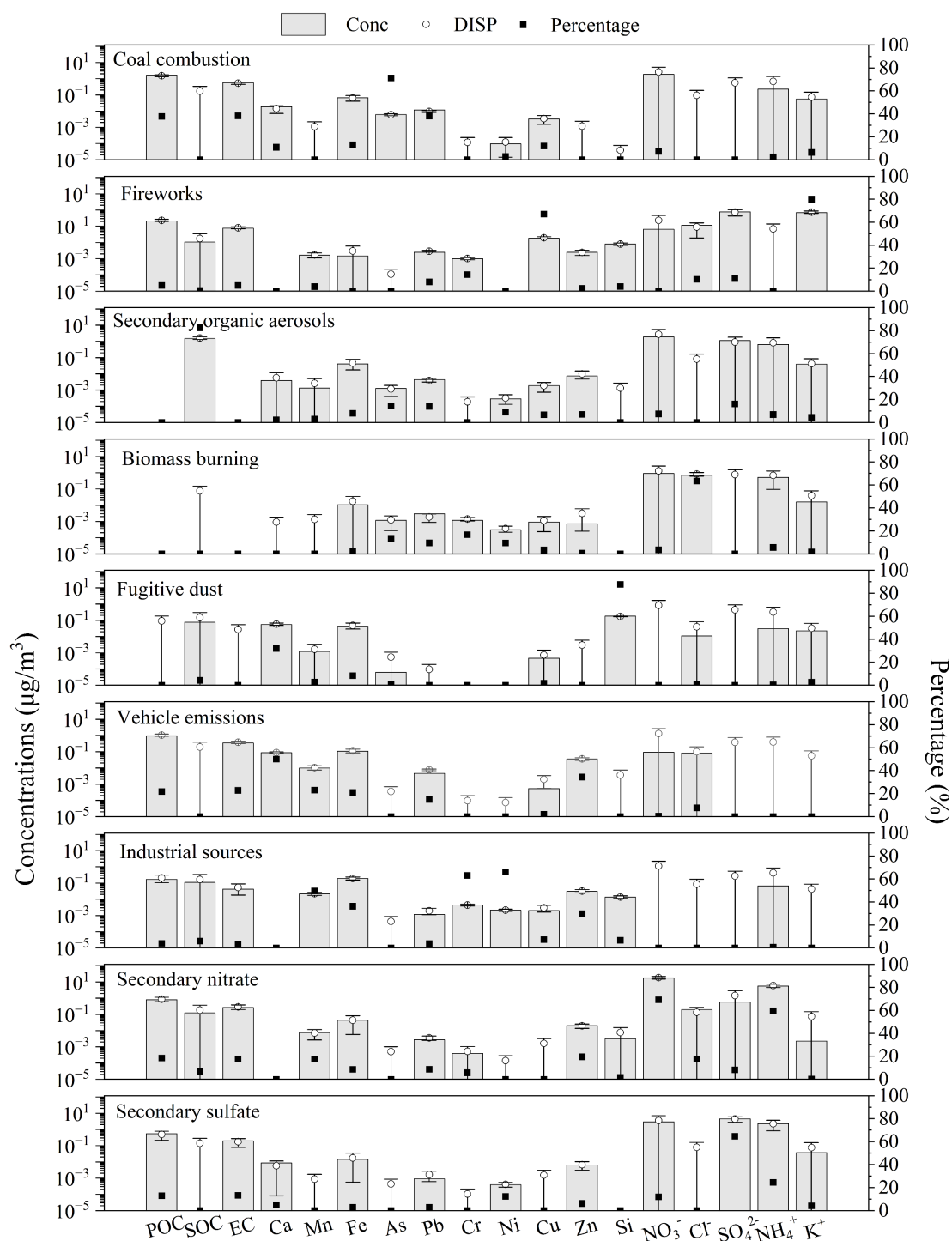
362

363 **Response:**

364 We double-checked the data and re-performed the DN-PMF analysis (Section 3.1, lines  
 365 191-220). To address the reviewer’s concern regarding the potential non-uniqueness  
 366 and uncertainty of PMF solutions, we additionally conducted DISP diagnostics. The  
 367 DISP results showed narrow confidence intervals and no evidence of factor swapping,  
 368 indicating that the resolved solution with nine-factor is statistically robust. The species  
 369 with relatively narrow DISP bands in each factor indicates the potential tracers of  
 370 corresponding source. Based on this updated analysis, the contribution of the fireworks  
 371 factor during the New Year’s Eve haze period reached 76.8%, which reinforces the  
 372 robustness of the source apportionment and strengthens confidence in the attribution of  
 373 the holiday-related  $PM_{2.5}$  increase to fireworks emissions.

374 We have updated the manuscript accordingly. The revised Figure S4 now presents the  
 375 source profiles of the nine factors resolved by the updated DN-PMF analysis for the  
 376  $PM_{2.5}$  pollution episode during New Year’s Eve 2024 in Hangzhou.

377



378

379 **Figure S4.** Source profiles of the nine factors resolved by the DN-PMF analysis for the  
 380 PM<sub>2.5</sub> pollution episode during New Year's Eve 2024 in Hangzhou. The gray bars  
 381 represent the species concentrations apportioned to each factor ( $\mu\text{g}/\text{m}^3$ ), open circles  
 382 denote the DISP-derived uncertainties, and black squares indicate the relative  
 383 contributions (%) of each species to the corresponding factor. The identified factors  
 384 include: coal combustion, fireworks, secondary organic aerosols (SOA), biomass  
 385 burning, fugitive dust, vehicle emissions, industrial sources, secondary nitrate, and  
 386 secondary sulfate.

387

388

389 **Minor/Technical Comments:**

390 1. The Methods state “No strict hyper-parameter tuning requirements were imposed —  
391 either Bayesian optimization or grid search could be used” (Sect. 2.2). That is  
392 ambiguous. Please state what the authors actually used (grid values, CV folds, early  
393 stopping) and include the tuned hyperparameters in a supplement. Report learning  
394 curves and variable importance (SHAP or partial dependence) so readers see what  
395 drivers the model uses.

396

397 **Response:**

398 Indeed, the original description was kind of ambiguous and we have thoroughly revised  
399 the manuscript to address all the points raised. The changes are as follows:

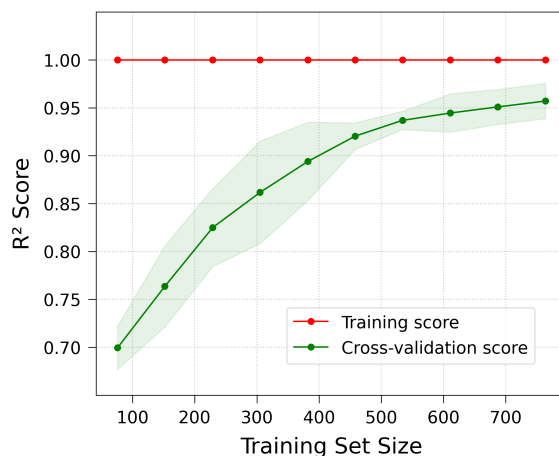
400 (1). Regarding hyper-parameter tuning details: As suggested, we have completely  
401 rewritten the corresponding paragraph in Section 2.2 (lines 122-126) to unambiguously  
402 state that we used Bayesian optimization (with the Optuna framework) with a 5-fold  
403 cross-validation strategy. We have also explicitly listed the final set of optimal hyper-  
404 parameters and the resulting model performance metrics in Text S2 for clarity and  
405 reproducibility.

406 (2). Learning curve analysis: We have added a new Supplementary Figure S1 (Learning  
407 Curve) and a corresponding description in Text S2. The learning curve demonstrates  
408 that the cross-validation score converges stably with the training score at a high  $R^2$   
409 value ( $\sim 0.95$ ), confirming that our model generalizes well without overfitting and that  
410 the dataset size is sufficient.

411 (3). As for variable importance (SHAP) analysis: To elucidate the drivers of the model’s  
412 predictions, we have performed a comprehensive SHAP analysis. Supplementary  
413 Figure S3 and a detailed interpretation have been added to Text S2. The analysis  
414 identifies *date\_unix* (temporal trend associated with emission dynamics) and *length* (air  
415 mass trajectory length) as the two most influential features and provides a physically  
416 coherent interpretation of their effects, which aligns with the emission patterns during  
417 the Chinese New Year period and the role of local vs. regional transport.

418 We believe these revisions have greatly enhanced the methodological rigor,  
419 transparency, and interpretability of our work, and we thank the reviewer again for the  
420 valuable suggestions.

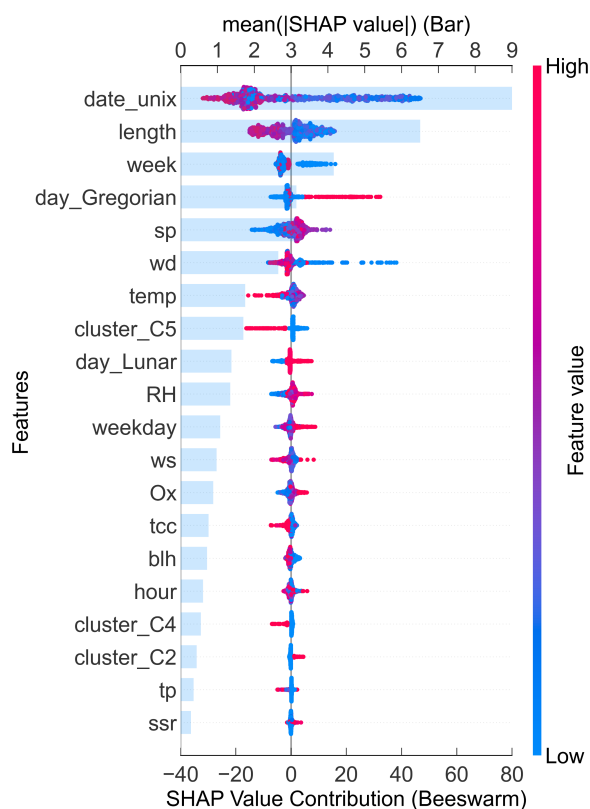
421



422

423 **Figure S1.** Learning curve of the ML-Based counterfactual PM<sub>2.5</sub> prediction framework.

424



425

426 **Figure S3.** SHAP summary plots of feature contributions to the ML counterfactual  
427 PM<sub>2.5</sub> model (upper: feature importance; lower: feature contributions).

428

429

430 2. Mixing gas-phase species (CO/NO<sub>x</sub>) with PM speciation should be echoed in the  
431 manuscript as a caveat: gas tracers and PM tracers respond to different source processes  
432 and lifetimes, so combining them without accounting for atmospheric chemistry and  
433 differential lifetimes could mislead source attribution (Li et al., 2025). The authors  
434 briefly note NO<sub>2</sub>/CO decline; they should deepen that discussion (e.g., use NO<sub>2</sub> as

435 independent corroboration of traffic reduction).

436

437 **Response:**

438 We thank the reviewer for this insightful comment/suggestion. We agree that gas-phase  
439 species (e.g., CO and NO<sub>2</sub>) and PM chemical tracers respond to different source  
440 processes and atmospheric lifetimes, and therefore should not be interpreted  
441 interchangeably when attributing sources. Following this suggestion, we have  
442 substantially revised the first paragraph in Section 3.2 (lines 226-235) of the manuscript  
443 to clarify the physical meaning of the observed gas-phase behavior.

444 In the revised text, we no longer interpret the NO<sub>2</sub>/CO variations as chemical evidence  
445 of fireworks-related nitrate formation. Instead, we explicitly state that the rapid decline  
446 of NO<sub>2</sub> and CO on Chinese New Year's Eve primarily reflects the weakening of traffic-  
447 and combustion-related emissions during the holiday period. We now treat NO<sub>2</sub> as an  
448 independent corroboration of reduced anthropogenic activity, consistent with the  
449 reviewer's recommendation, rather than as a compositional tracer of fireworks  
450 emissions.

451 We have also clarified that the abrupt decrease and rebound of O<sub>3</sub> reflect an NO-driven  
452 titration effect and rapid changes in atmospheric chemistry, rather than a direct  
453 signature of fireworks aerosols. The fireworks contribution assessment is now framed  
454 to rely primarily on PM<sub>2.5</sub> chemical fingerprints and the counterfactual analysis.

455 These revisions more accurately distinguish between gas-phase and particulate-phase  
456 diagnostics and avoid any potential misinterpretation arising from mixing the two.

457

458

459 3: Figure 1, Figure 2 and Figure 3 have missing references in the main text.

460

461 **Response:**

462 Revised as suggested. The following revisions have been made:

463 A reference to Figure 1 has been added on page 2, line 61.

464 An additional reference to Figure 1 has been added on page 4, line 105.

465 References to Figure 2 have been added on page 8, line 227 and page 9, line 244.

466 References to Figure 3 have been added on page 10, line 289; page 11, line 303; and  
467 page 13, line 325.

468 These changes ensure that all figures (Figs. 1-3) are now properly cited and discussed  
469 in the main text.

470

471

472

473

474 **Reviewer #2**

475 The manuscript proposes a machine-learning counterfactual framework to estimate the  
476 impact of the Chinese Spring Festival (CSF) on PM<sub>2.5</sub> in Hangzhou and the “2+26”  
477 cities. The study is timely and policy-relevant, with a clear intention to distinguish air-  
478 quality changes from emissions, and the manuscript is well organized and clearly  
479 presented. However, several aspects of causal ML practice, the temporal validation  
480 strategy, and issues of data representativeness need to be strengthened before  
481 publication.

482

483 Major Comments

484

485 **Comment #1:**

486 The manuscript frames its analysis within a causal framework, treating the Chinese  
487 Spring Festival (CSF) as a “treatment” and using the XGBoost model to predict a  
488 counterfactual business-as-usual (BAU) scenario. While this is a conceptually  
489 appropriate starting point, the current methodology does not yet meet a rigorous causal  
490 ML design. The CSF is a composite factor, bundling the effects of fireworks, altered  
491 traffic patterns, and changes in industrial/construction activity. This complexity  
492 challenges the core identification assumptions required for causal claims.

493 Furthermore, the analysis does not adequately address potential influence of these  
494 assumptions, such as the inconsistent overlap in covariate distributions between festival  
495 and non-festival periods. Some features, like the lunar calendar day, are inherently  
496 confounded with the treatment, violating conditional independence. The study could be  
497 characterized as a causally inspired counterfactual prediction for BAU rather than a  
498 causal estimator under verified identification conditions. Hence, the authors may wish  
499 to reconsider the title and tone down the causal claims to avoid overstatement.

500

501 **Response:**

502 We sincerely thank the reviewer for this important and constructive comment. We agree

503 that the Chinese Spring Festival (CSF) represents a composite intervention involving  
504 multiple concurrent behavioral and emission changes (e.g., fireworks activities,  
505 reductions in traffic volume, and modifications in industrial or construction operations).  
506 As a result, several key identification assumptions required for strict causal inference—  
507 such as conditional independence, adequate covariate overlap between treated and  
508 untreated periods, and the independence of certain covariates (e.g., lunar-day indicators)  
509 from the intervention—cannot be fully validated in this context. We appreciate the  
510 reviewer’s clarification, which has helped us refine the conceptual framing of the paper.  
511 Following the reviewer’s suggestion, we have revised the manuscript in several ways  
512 to ensure that the study is presented as a causally informed counterfactual analysis  
513 rather than a strict causal estimator:

514

515 (1). Revised the title to remove any implication of strong causal identification.

516 The new title is:

517 “Impact of the Chinese Spring Festival on PM<sub>2.5</sub> air quality in the Beijing-Tianjin-Hebei  
518 and surrounding region: A machine-learning-based counterfactual modeling approach”.

519

520 (2). We have modified the methodological description in the Introduction to clearly  
521 state that the proposed framework is not intended to identify structural causal effects.  
522 The revised text now clarifies that our approach provides a causally informed  
523 counterfactual prediction of the BAU scenario rather than a formal causal estimate.

524

525 (3). We have added a dedicated clarification paragraph at the end of Section 2.2 (lines  
526 127-137), explicitly acknowledging the composite nature of the CSF intervention, the  
527 challenges associated with validating causal identification assumptions, the inherent  
528 confounding of calendar-related variables such as the lunar-day index, and the resulting  
529 limitations for making strict causal claims.

530

531 (4). Adjusted wording throughout the manuscript (e.g., replacing “causal impact” with  
532 “holiday-related effects” or “deviations relative to BAU”) to avoid overstating causal

533 interpretation while preserving the inferential value of the counterfactual framework.  
534 These revisions ensure that the manuscript's framing is fully aligned with the reviewer's  
535 recommendation. We again thank the reviewer for helping us improve the rigor, clarity,  
536 and conceptual precision of the study.

537

538

539 **Comment #2:**

540 The current modeling approach, which relies on instantaneous covariates, does not  
541 account for the temporal auto-correlation inherent in air pollution. The concentration at  
542 any given time is also influenced by the emissions and meteorological conditions of  
543 previous periods. The choice of a random 80/20 split for model validation may  
544 introduce data leakage when evaluating the model performance. A blocked or rolling  
545 time-based cross-validation would be more appropriate here.

546 Separately, uncertainty quantification has been extensively discussed in ML-based  
547 atmospheric remote sensing, yet is not addressed in the present manuscript; providing  
548 calibrated predictive uncertainty would improve the interpretability of the results.

549

550 **Response:**

551 We appreciate the reviewer's suggestions regarding the treatment of temporal  
552 dependence, model validation strategy, and predictive uncertainty.

553

554 (1) Temporal autocorrelation and validation scheme.

555 We agree that air pollution data exhibit temporal autocorrelation and that random  
556 splitting may overestimate performance if used for forecasting tasks. However, the  
557 purpose of this study is not to predict future concentrations, but to estimate  
558 counterfactual business-as-usual (BAU) concentrations under the same meteorological  
559 conditions during the Chinese Spring Festival (CSF) period. The model therefore serves  
560 as a nonlinear regression tool to capture the contemporaneous relationships between  
561 PM<sub>2.5</sub> and its covariates, rather than the temporal evolution of pollution.

562 To examine the reviewer's concern, we additionally tested a blocked time-based cross-

563 validation (five-fold TimeSeriesSplit) using only instantaneous covariates. Under this  
564 scheme, the cross-validated  $R^2$  dropped to -0.50 and the test-set  $R^2$  to 0.09 (RMSE  $\approx$  20  
565  $\mu\text{g}/\text{m}^3$ ), indicating that instantaneous features alone cannot represent temporal  
566 dependence.

567 We then introduced simple lagged predictors (1-3 h lags of  $\text{PM}_{2.5}$ , temperature, and  
568 wind speed), after which the time-block cross-validated  $R^2$  increased to 0.69 and the  
569 test-set  $R^2$  to 0.94 (RMSE  $\approx$  5.4  $\mu\text{g}/\text{m}^3$ ). These results confirm that explicit short-term  
570 history improves robustness under sequential validation. Nevertheless, because lagged  
571  $\text{PM}_{2.5}$  values could leak information from neighboring time steps and obscure the  
572 interpretation of “instantaneous” BAU conditions, the main analysis retains the model  
573 without lagged terms. The lagged-feature experiment was used only to verify the  
574 reviewer’s concern regarding temporal dependence, and its results support our choice  
575 of focusing on contemporaneous covariates in the counterfactual analysis.

576

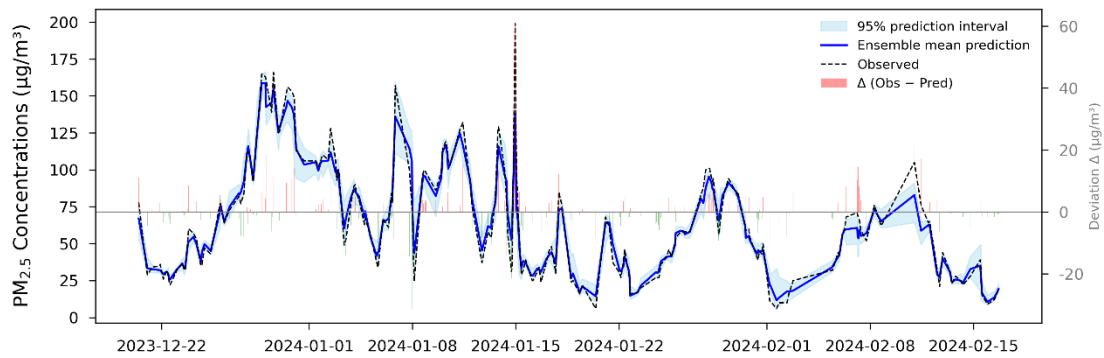
577 (2) Predictive uncertainty.

578 Following the reviewer’s recommendation, we implemented an ensemble-based  
579 bootstrap approach to quantify predictive uncertainty (Text S3). Fifty XGBoost models  
580 with identical hyperparameters were trained on bootstrap-resampled training sets,  
581 producing an ensemble of counterfactual BAU predictions. The ensemble mean was  
582 taken as the point estimate, and the 2.5<sup>th</sup>-97.5<sup>th</sup> percentile range as the 95% prediction  
583 interval. The ensemble-mean predictions achieved an  $R^2 \approx 0.95$  on the test set, slightly  
584 lower than the single model but providing calibrated uncertainty estimates. The new  
585 Figure S2 in the revision visualizes observed  $\text{PM}_{2.5}$ , ensemble-mean predictions, and  
586 their 95 % prediction intervals, along with the deviation ( $\Delta = \text{Observed} - \text{Predicted}$ ).

587 These additions address the reviewer’s concerns by (a) verifying temporal dependence  
588 through lag-feature tests, (b) clarifying the rationale for using a random split in a  
589 counterfactual, non-forecasting context, and (c) incorporating calibrated uncertainty  
590 quantification to improve interpretability.

591 We sincerely thank the reviewer once again for these helpful suggestions.

592



593

594 **Figure S2.** Observed PM<sub>2.5</sub> concentrations and ensemble-based counterfactual  
 595 predictions from the XGBoost model, together with their 95% prediction intervals and  
 596 the corresponding deviations ( $\Delta = \text{observed} - \text{predicted}$ ).

597

598

599

600 **Comments #3:**

601 The abstract opens with acute short-term health risks from extremely high PM<sub>2.5</sub>, but  
 602 the regional result emphasizes an average decrease of  $19.0 \pm 17.5 \mu\text{g}/\text{m}^3$  over the  
 603 extended holiday period. These two statements are not contradictory but currently feel  
 604 weakly connected. Besides that, Section 3.2 (Hangzhou) explicitly reports large  
 605 concurrent source changes (e.g., vehicles -31%; dust +2790%), yet Section 3.5 (“2+26”)  
 606 estimates fireworks’ contribution “under the assumption that emissions from other  
 607 sources remained unchanged.” The authors need to address this inconsistency or  
 608 provide sensitivity analysis under alternative assumptions.

609

610 **Response:**

611 We thank the reviewer for this insightful comment. We address both aspects raised in  
 612 this comment below.

613 (1) Connection between acute PM<sub>2.5</sub> risks in the abstract and the regional mean decrease.  
 614 We agree that these two points were initially presented in a way that appeared only  
 615 loosely connected. Following the reviewer’s suggestion, we have revised the abstract  
 616 to explicitly clarify that the regional mean decrease ( $-19.0 \pm 17.5 \mu\text{g}/\text{m}^3$  across the “2  
 617 + 26” cities) reflects the multi-day reduction in anthropogenic activities during the  
 618 extended holiday period, whereas the acute short-term health risks refer specifically to

619 the sharp, short-lived PM<sub>2.5</sub> spikes caused by concentrated fireworks during the peak  
620 window on New Year's Eve. These two patterns are therefore not contradictory but  
621 operate on different temporal scales. The revised abstract now clearly links the regional  
622 baseline reduction with the episodic, fireworks-driven PM<sub>2.5</sub> peaks.

623 (2) Consistency between Sections 3.2 and 3.5 regarding the “unchanged sources”  
624 assumption. We thank the reviewer for highlighting the apparent inconsistency between  
625 the detailed source changes reported for Hangzhou (Section 3.2) and the simplifying  
626 assumption used in the regional “2 + 26” analysis (Section 3.5). We agree that assuming  
627 “other emission sources remained unchanged” is a strong simplification. In the regional  
628 analysis, this assumption was intentionally used as part of the counterfactual BAU  
629 framework to isolate the incremental effect of fireworks.

630

631 Importantly, as shown in Section 3.2 for Hangzhou, concurrent source changes did  
632 occur. Traffic activity decreased substantially (vehicle-related emissions  $\approx$  -63%),  
633 while local dust emissions increased due to fireworks fallout. Regional gas-phase  
634 tracers further confirm this pattern: NO<sub>2</sub>, a traffic indicator, decreased by about 12%,  
635 whereas CO, a combustion tracer, increased by  $\approx$  18% on New Year's Eve. These  
636 observations demonstrate that non-fireworks sources were not constant but generally  
637 weakened.

638 Therefore, the “unchanged-source” assumption does not overstate the fireworks-related  
639 contribution; instead, it yields a conservative lower-bound estimate. If other  
640 anthropogenic emissions declined, the fraction of observed PM<sub>2.5</sub> attributable to  
641 fireworks would, in reality, be even higher. This interpretation is now explicitly stated  
642 in Section 3.5, supported by independent evidence from NO<sub>2</sub> and CO behavior.

643

644 In the revised manuscript, we have:

- 645 (a) clarified this rationale in Section 3.5 (lines 400-408),
- 646 (b) softened the original wording regarding “unchanged sources” and
- 647 (c) explicitly emphasized that the resulting fireworks contribution represents a  
648 conservative BAU-based lower-bound estimate.

649 We sincerely thank the reviewer once again for this important comment, which  
650 significantly improved the clarity and consistency of the manuscript.

651

652 **Comment #4:**

653 Section 2.1 requires several clarifications. First, key details for the ERA5 dataset,  
654 including its temporal/spatial resolution and a reference link, should be provided in the  
655 manuscript or SI (Text S1/Table S1). To address the potential for reanalysis data to  
656 smooth over urban-scale extremes, a brief comparison of ERA5 variables against  
657 ground-station data would strengthen the analysis. Additionally, the usage of total  
658 precipitation (TP) needs to be explained; since it is an accumulated value, please  
659 describe any transformation performed to make it suitable for an hourly model. The  
660 specific parameters or a reference for Emanuel's saturated vapor pressure formula  
661 should be included.

662

663 **Response:**

664 We thank the reviewer for these detailed and constructive suggestions. All points raised  
665 have now been fully addressed in the revised manuscript, as summarized below.

666 (1) ERA5 dataset details

667 Section 2.1 (lines 98-100) and Table S1 have been revised to explicitly state that hourly  
668 single-level ERA5 reanalysis data were used, with a horizontal resolution of  $0.25^\circ \times$   
669  $0.25^\circ$  and hourly temporal resolution. A reference link to the Copernicus Climate Data  
670 Store has also been added to ensure full reproducibility.

671 (2) Evaluation of ERA5 representativeness

672 To assess the representativeness of the ERA5 variables used in this study (20 December  
673 2023-16 February 2024), we compared hourly ERA5 data at the grid cell centered over  
674 downtown Hangzhou with observations from the Hangzhou Xiaoshan International  
675 Airport station (~20 km away). The comparison reveals that ERA5 exhibits very small  
676 mean biases for near-surface temperature ( $\approx 0.36^\circ\text{C}$ ) and wind speed ( $\approx 0.40\text{ m/s}$ ),  
677 demonstrating its high fidelity in capturing regional meteorological states. While  
678 inherent discrepancies exist between a  $0.25^\circ$  grid average and point measurements due  
679 to spatial smoothing and local micro-terrain, ERA5's reliability in representing  
680 synoptic-scale transitions and atmospheric dynamics is well documented (Hersbach et  
681 al., 2020). Importantly, the objective of this work is to represent regional and synoptic-

682 scale forcing to drive the counterfactual PM<sub>2.5</sub> prediction. This requires key predictors  
683 such as boundary-layer height and solar radiation that are typically unavailable from  
684 surface stations. ERA5 provides a spatially and temporally consistent dataset that  
685 avoids the localization and discontinuity, such as 3-hour intervals or missing records,  
686 often found in ground observations. Given the study's focus on regional-scale  
687 meteorology, ERA5 offers a more appropriate and robust basis for the machine-  
688 learning-based modeling framework.

### 689 3. Processing of total precipitation (TP)

690 ERA5 total precipitation (TP) represents liquid and frozen water accumulated over the  
691 previous hour and is expressed in meters of water equivalent. Since the hourly dataset  
692 already reflects accumulation over a one-hour period, no additional differencing or  
693 temporal transformation was required prior to its use in the model.

### 694 4. Emanuel's saturated vapor pressure formula

695 The specific formulation and reference have been added to Text S1 in the  
696 Supplementary Information. Saturation vapor pressure ( $E$ , Pa) is computed using  
697 Emanuel's empirical equation (Emanuel, 1994):

$$698 \quad \ln E = 53.67957 - \frac{6743.769}{T} - 4.8451 \ln T$$

699 where  $T$  is absolute temperature (K). The equation is applied to 2-m air temperature  
700 ( $T_2$ ) and dew-point temperature ( $D_2$ ) to obtain saturation ( $E_s$ ) and actual vapor pressure  
701 ( $E_a$ ), respectively, with relative humidity calculated as  $RH = (E_a/E_s) \times 100\%$ .

702 Together, these revisions substantially improve the transparency, rigor, and  
703 reproducibility of the meteorological data processing procedures in the manuscript. We  
704 sincerely appreciate the reviewer's helpful suggestions.

705

### 706 **Ref:**

707 Hersbach, H., et al. The ERA5 global reanalysis, Quarterly journal of the royal meteorological  
708 society, 146, 1999-2049, <https://doi.org/10.1002/qj.3803>, 2020.

709

### 710 **Comment #5:**

711 There is the spatial representativeness mismatch between the machine learning model,  
712 which uses a 14-site city average, and the DN-PMF analysis, which uses chemical data  
713 from a single site. This difference could introduce a bias, particularly for localized  
714 sources like fireworks. The authors could discuss this limitation and its potential impact  
715 on their findings. Given the team's related work (e.g., Journal of Environmental  
716 Sciences), a brief comparison of the methodological advantages and efficiency gains  
717 relative to prior work would also help position the contribution.

718

719 **Response:**

720 We thank the reviewer for this important comment, which was also raised by Reviewer  
721 #1. Please see our responses to reviewer #1 for details. Below we also provide a concise  
722 clarification regarding (1) the spatial representativeness mismatch and (2) the  
723 methodological contribution relative to prior work.

724 1. Spatial representativeness mismatch

725 We agree that the ML model (city-wide average) and the DN-PMF analysis (single-site  
726 composition) operate at different spatial scales. To directly assess this issue, we  
727 performed a linear regression between hourly PM<sub>2.5</sub> at the Wolongqiao site and the 14-  
728 site city average over the full study period. The strong correlation ( $R^2 = 0.78$ , slope =  
729 1.11) demonstrates that Wolongqiao reliably captures the same temporal pollution  
730 dynamics as the broader urban area, particularly during winter episodes and the Spring  
731 Festival period.

732 Although absolute levels differ slightly, fireworks emissions during New Year's Eve  
733 form a city-wide, highly synchronized plume. Their chemical signatures (e.g., sharp K<sup>+</sup>  
734 and EC enhancements) are consequently spatially coherent, making the single-site DN-  
735 PMF analysis suitable for identifying and tracking this dominant source. We have added  
736 this clarification as a discussed limitation in the revised manuscript (Section 3.4, lines  
737 357-370).

738

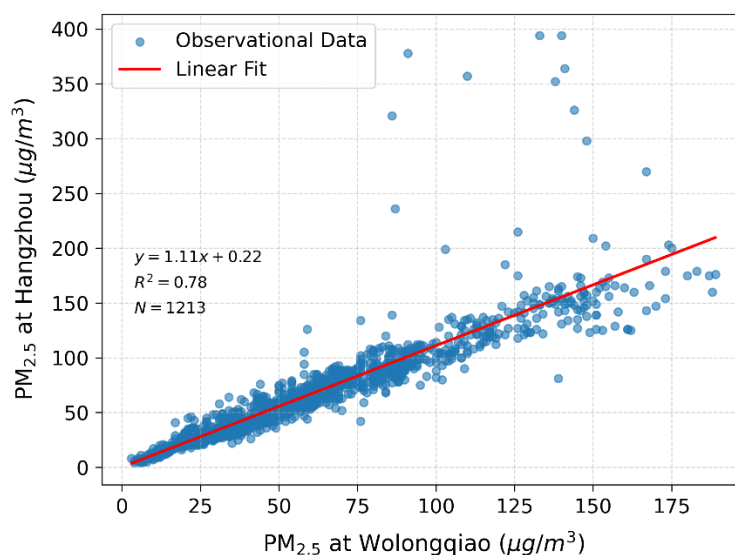
739 2. Methodological advantages relative to prior work

740 We have also clarified the methodological contribution of this study. Compared with

741 traditional receptor modeling approaches, our ML-based counterfactual framework  
742 provides a direct “no-holiday” baseline, enabling cleaner attribution of fireworks  
743 impacts. The method requires only routine monitoring and reanalysis data, is highly  
744 scalable (as demonstrated in the “2+26” region), and can be applied rapidly to large city  
745 networks. This represents a clear efficiency gain relative to chemical-speciation-based  
746 source apportionment. The consistency between the ML results and the DN-PMF  
747 source apportionment further supports the robustness of the approach.

748 We have added these points to better position the contribution in the revised manuscript.

749



750

751 **Figure S9.** Linear regression analysis of hourly PM<sub>2.5</sub> concentrations between the  
752 Wolongqiao site (x-axis) and the Hangzhou city-wide average (14 sites, y-axis) during  
753 the study period (2023/12/20 - 2024/2/20). The strong correlation ( $R^2 = 0.78$ ) supports  
754 the representativeness of the Wolongqiao site in capturing city-scale pollution trends  
755 during the observation period.

756

757 **Minor corrections:**

758 - Line 22: twenty-eight -> 28

759 - Line 119: meterorology -> meteorology

760 - Line 164: A -> An

761 - Line 207: in the midnight of the New Year Eve -> at midnight on New Year’s Eve

- 762 - Line 244: Please add units for RMSE and MAE.  
763 - Line 260: reliability -> reliability  
764 - Line 261: technique -> technique  
765 - Line 323: deterioration -> deterioration  
766 - Table S1: Please use Pa (not pa) for pressure unit.

767

768 **Response:**

769 All minor corrections listed (typos, unit formatting, and grammatical adjustments) have  
770 been implemented in the revised manuscript, including those on Lines 22, 119, 164,  
771 207, 244, 260, 261, 323, and in Table S1.

772

773

774

775

776

777

778

779

780

781 **Reviewer #3**

782 This study employs machine learning to evaluate the impact of firework displays on  
783 PM<sub>2.5</sub> pollution during the Chinese Spring Festival. Overall, the manuscript is well-  
784 structured and concisely written. The findings offer valuable insights for the scientific  
785 management of PM<sub>2.5</sub> pollution in China. Revisions are needed before consideration for  
786 publication.

787

788 **Comment #1:**

789 P2, L41: “often marked by a decline in nitrogen oxide levels and a sharp increase in  
790 PM<sub>2.5</sub> concentrations”. This statement requires supporting references.

791 **Response:**

792 We have added appropriate references to support this statement.

793 Previous studies have consistently reported a pronounced reduction in nitrogen oxides  
794 during the Chinese Spring Festival due to decreased traffic and industrial activities. For  
795 example, Li et al. (2021) showed substantial declines in tropospheric NO<sub>2</sub> columns  
796 (31.8%-44.5%) across Chinese megacities associated with large-scale population  
797 migration and reduced vehicular emissions. In contrast, PM<sub>2.5</sub> responses during this  
798 period are more complex. Dai et al. (2021a) demonstrated that although NO<sub>2</sub>  
799 concentrations decreased significantly during the Spring Festival and the coincident  
800 COVID-19 lockdown, PM<sub>2.5</sub> exhibited heterogeneous changes driven by shifts in  
801 emission sources, including enhanced firework-related emissions. In addition, the  
802 review by Wu et al. (2022) summarized that air quality during the Spring Festival is  
803 influenced by competing effects, including reduced anthropogenic emissions and  
804 enhanced emissions from fireworks and residential fuel use, leading to heterogeneous  
805 PM<sub>2.5</sub> responses across regions and periods.

806 Based on these findings, the manuscript has been revised to state that:

807 “During this period, a distinct ‘holiday effect’ is observed in air quality variations, often  
808 marked by a decline in nitrogen oxide (NO<sub>x</sub>) levels and complex changes in PM<sub>2.5</sub>  
809 concentrations” (P2, L42–L44), with supporting references added (Li et al., 2021; Dai  
810 et al., 2021a; Wu et al., 2022).

811

812 **Refs:**

813 Li, D., Wu, Q., Wang, H., Xiao, H., Xu, Q., Wang, L., Feng, J., Yang, X., Cheng, H., and Wang,  
814 L.: The Spring Festival Effect: The change in NO<sub>2</sub> column concentration in China caused by  
815 the migration of human activities, *Atmospheric Pollution Research*, 12, 101232,  
816 <https://doi.org/10.1016/j.apr.2021.101232>, 2021.

817 Dai, Q., Hou, L., Liu, B., Zhang, Y., Song, C., Shi, Z., Hopke, P. K., and Feng, Y.: Spring Festival  
818 and COVID-19 lockdown: disentangling PM sources in major Chinese cities, *Geophys. Res.*  
819 *Lett.*, 48, <https://doi.org/10.1029/2021GL093403>, 2021a.

820 Wu, G., Tian, W., Zhang, L., and Yang, H.: The Chinese spring festival impact on air quality in  
821 China: A critical review, *International Journal of Environmental Research and Public Health*,  
822 19, 9074, <https://doi.org/10.3390/ijerph19159074>, 2022.

823

824

825 **Comment #2:**

826 P3, L74-81: (1) Please specify the instrumentation used for measuring chemical  
827 compositions, along with their limits of detection, accuracy, and precision. (2)  
828 Regarding the calculation of SOC: The OC/EC minimum ratio method assumes stable  
829 emission sources over a period, which is clearly not applicable to the drastic emission  
830 changes during the Spring Festival. The authors should re-evaluate the validity of this  
831 method.

832 **Response:**

833 We have revised the manuscript accordingly:

834 (1) Instrumentation, limits of detection, accuracy, and precision.

835 We have updated the Methods section to explicitly report the analytical instrumentation  
836 and methods used for PM<sub>2.5</sub> chemical composition measurements. Elemental species  
837 were measured using an iCAP 7000 Series ICP-OES spectrometer (Thermo Scientific,  
838 USA). Carbonaceous components (OC/EC) were analyzed using a DRI 2001A  
839 thermal/optical OC/EC analyzer following the IMPROVE\_A protocol. Water-soluble  
840 ions were determined by ion chromatography (IC). In addition, the method detection  
841 limits (MDLs) and uncertainty/precision terms (error fractions) for all measured species  
842 are now summarized in Table S2, and these values are also those used as PMF model  
843 inputs. To address analytical accuracy, we added a table note stating that instrument  
844 calibration and routine QA/QC procedures (multi-point calibration using standards,  
845 field/laboratory blanks, and replicate analyses) were performed to ensure data quality  
846 (Section 2.1, lines 82-91).

847

848 (2) We agree that the minimum OC/EC ratio (EC-tracer) approach assumes relatively  
849 stable primary emission characteristics over the period used to determine (OC/EC)<sub>min</sub>,  
850 and that abrupt source changes during the Spring Festival may increase uncertainty in  
851 SOC estimates. We therefore re-evaluated this assumption by examining the period  
852 dependence of (OC/EC)<sub>min</sub>: the 10th-percentile OC/EC ratio was calculated separately

853 for the periods before and after Lunar New Year's Eve (pre-eve: 3.10; post-eve: 3.13;  
854 full period: 3.11). Based on this check, we added clarifying text in the Methods section  
855 noting that the EC-tracer-based SOC is included as an input variable in the PMF  
856 analysis to represent the secondary organic component; nevertheless, potential non-  
857 stationarity of primary emissions during the Spring Festival may introduce additional  
858 uncertainty, and SOC-related results for this period are interpreted with caution.

859 We thank the reviewer again for these helpful suggestions.

860

861

862 **Comment #3:**

863 P3, L88: Why was ERA5 reanalysis data used instead of locally measured  
864 meteorological data? Please justify this choice.

865 **Response:**

866 We used the ECMWF ERA5 hourly reanalysis meteorological data with the following  
867 reasons:

868 (1) Completeness and physical consistency of predictors. Our model requires not only  
869 standard near-surface variables but also key predictors that are typically unavailable or  
870 not routinely reported by local surface stations, such as boundary-layer height and  
871 surface solar radiation. These variables are essential for representing vertical mixing  
872 potential and photochemical conditions. ERA5 provides a physically self-consistent  
873 and spatiotemporally continuous set of meteorological variables through a unified data  
874 assimilation system, ensuring internally consistent model inputs.

875 (2) Spatial representativeness for regional-scale analysis. This study aims to  
876 characterize pollution processes and PM<sub>2.5</sub> behavior at the city and regional scales.  
877 Point observations can be strongly influenced by local micro-terrain and urban canopy  
878 effects and may not represent the area-average forcing relevant for regional transport  
879 and dispersion. The gridded ERA5 fields (0.25° × 0.25°) are better suited to capturing  
880 the synoptic and regional meteorological conditions driving the counterfactual  
881 predictions.

882 (3) Continuity and availability. ERA5 avoids gaps due to station outages or

883 discontinuous records and provides an uninterrupted hourly time series required for  
884 stable machine-learning training and inference.

885 To evaluate the representativeness of ERA5 during the study period (20 December  
886 2023-16 February 2024), we compared hourly ERA5 data from the grid cell centered  
887 over downtown Hangzhou with observations at the Hangzhou Xiaoshan International  
888 Airport station (approximately 20 km away). The comparison shows very small mean  
889 biases for temperature ( $\approx 0.36$  °C) and wind speed ( $\approx 0.40$  m/s), indicating that ERA5  
890 captures the regional meteorological background with high fidelity. While inherent  
891 discrepancies exist between a  $0.25^\circ$  reanalysis grid average and point measurements  
892 due to spatial smoothing and local microclimates, ERA5's ability to represent synoptic-  
893 scale transitions and atmospheric dynamics is well documented (Hersbach et al., 2020)  
894 and provides a robust basis for our regional-scale modeling framework. We have  
895 clarified this in the revised manuscript in Text S2, representativeness of ERA5  
896 meteorological data.

897 **Ref:**

898 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J.,  
899 Peubey, C., Radu, R., and Schepers, D.: The ERA5 global reanalysis, Quarterly journal of the  
900 royal meteorological society, 146, 1999-2049, <https://doi.org/10.1002/qj.3803>, 2020.

901

902

903 **Comment 4:**

904 P 6-7, Line 180-181: Fireworks also release substantial amounts of potassium.

905 Wang Ying et al. The air pollution caused by the burning of fireworks during the lantern  
906 festival in Beijing, Atmospheric Environment, 417-431, 41, 2007.

907 Wang Wenhua et al. Chemical composition and morphology of PM<sub>2.5</sub> in a rural valley  
908 during Chinese New's Eve: Impact of firework/firecracker display, Atmospheric  
909 Environment, 120225, 318, 2024.

910 **Response:**

911 As suggested, we have incorporated these references into the revised manuscript to  
912 support the identification of the firework source factor. These references, covering both

913 a study in urban Beijing (Wang et al., 2007) and the findings in a rural valley (Wang et  
914 al., 2024), significantly strengthen our discussion on the enrichment of potassium ( $K^+$ )  
915 during the firework event in Hangzhou by providing evidence across different  
916 environmental settings.

917

918 **Comment 5:**

919 P8, L214: The authors report that fireworks contributed ~70% to  $PM_{2.5}$ , which is an  
920 exceptionally high figure. How does this compare with previous studies? Were these  
921 fireworks discharged in the immediate vicinity of the monitoring sites? Furthermore,  
922 are firework bans implemented in this city?

923

924 **Response:**

925 We acknowledge that a ~70% contribution of fireworks to  $PM_{2.5}$  is unusually high when  
926 viewed in a general context. However, we believe that this estimate is physically  
927 plausible for an extreme, short-lived pollution episode. Specifically, the reported value  
928 represents an average contribution of 76.8% over the 24-hour New Year's Eve haze  
929 event, rather than a long-term or seasonal mean. The following lines of evidence  
930 support this interpretation.

931 (1) Comparison with previous studies.

932 Although reported mass fractions vary with location and averaging period, our results  
933 are consistent with previous studies documenting extreme enrichment of firework-  
934 related aerosols during peak discharge windows. For example, Wang et al. (2007)  
935 reported that during the Lantern Festival night in Beijing, fireworks accounted for over  
936 90% of total mineral aerosol and approximately 43% of total carbon in  $PM_{2.5}$ ,  
937 accompanied by more than fivefold increases in tracers such as Ba, K, and Mg relative  
938 to non-festival periods. More recently, Wang et al. (2024) demonstrated that intensive  
939 firework displays can induce abrupt surges in  $K^+$ ,  $SO_4^{2-}$ , and  $Cl^-$  concentrations,  
940 particularly when emissions are spatially concentrated. Importantly, these high  
941 contributions were consistently observed over short peak intervals (hours to one day),  
942 rather than over longer averaging periods. In this context, the peak mass concentration

943 of the firework factor resolved in our study ( $167.2 \mu\text{g}/\text{m}^3$ ) falls well within the range  
944 implied by these documented tracer-rich aerosol spikes.

945 (2) Independent chemical evidence from elemental ratios (Section 3.2, lines 244-253).  
946 To ensure that the high contribution was not an artifact of receptor modeling, we  
947 performed a detailed analysis of elemental ratios (newly added Section 3.2). During the  
948 episode, the K/EC ratio exhibited an abrupt, order-of-magnitude increase, peaking at  
949 approximately 13—more than 60 times higher than the background level ( $\sim 0.2$ ).  
950 Meanwhile, the K/Ba ratio transitioned from irregular fluctuations to a stable low-value  
951 plateau with a coefficient of variation of  $\sim 16\%$ . This distinct and detached chemical  
952 signature indicates that the ambient aerosol was physically dominated by a K-rich, EC-  
953 poor population characteristic of fireworks emissions, consistent with the morphology  
954 and composition reported in previous field and laboratory studies. Taken together, these  
955 observations provide independent, physically grounded evidence that reinforces the  
956 PMF-resolved firework factor.

957 (3) Proximity of emissions and policy context.

958 Fireworks were not legally permitted in the immediate vicinity of the monitoring site  
959 (No. 29 Yanggong Causeway, West Lake Scenic Area). However, during the 2024  
960 Spring Festival, Hangzhou adopted differentiated firework management measures,  
961 under which several adjacent districts (e.g., Xiaoshan and Yuhang, located to the south  
962 and west of the urban core) allowed limited but intensive firework discharges in  
963 designated areas. Under the stagnant meteorological conditions prevailing on New  
964 Year's Eve, firework-related aerosols emitted from these nearby permitted zones were  
965 rapidly transported into and accumulated within the West Lake basin, where dispersion  
966 is constrained by local topography. This interpretation is supported by the explosive  
967 midnight  $\text{PM}_{2.5}$  surge shown in Fig. 2 and is further corroborated by the machine-  
968 learning counterfactual analysis (Fig. 3), which identified comparable firework-  
969 attributable  $\text{PM}_{2.5}$  increments across the broader urban area. These results indicate that  
970 the elevated contribution observed at the Wolongqiao site reflects a city-wide extreme  
971 pollution episode rather than a localized anomaly.

972 In summary, the reported 76.8% contribution represents a short-lived but intense peak

973 driven by concentrated firework activity in surrounding districts combined with  
974 unfavorable dispersion conditions, and is supported by both chemical evidence and  
975 independent modeling analyses.

976 **Refs:**

977 Wang, W., Zhou, H., Gao, Y., Shao, L., Zhou, X., Li, X., Wei, D., Xing, J., and Lyu, R.: Chemical  
978 composition and morphology of PM<sub>2.5</sub> in a rural valley during Chinese New Year's Eve:  
979 Impact of firework/firecracker display, *Atmospheric Environment*, 318, 120225,  
980 <https://doi.org/10.1016/j.atmosenv.2023.120225>, 2024.

981 Wang, Y., Zhuang, G., Xu, C., and An, Z.: The air pollution caused by the burning of fireworks  
982 during the lantern festival in Beijing, *Atmospheric Environment*, 41, 417-431,  
983 <https://doi.org/10.1016/j.atmosenv.2006.07.043>, 2007.

984

985

986 **Comment #6:**

987 P12-13, L329-335: Figure 5 is well-presented, but the accompanying description and  
988 discussion are too superficial. What explains the extreme disparity in firework  
989 contributions across different cities (>80% vs <10%)? Is this linked to local government  
990 bans? A more in-depth discussion is warranted.

991 **Response:**

992 We thank the reviewer for this insightful comment regarding the interpretation of Fig.  
993 5 and the need for a more in-depth discussion. In response, we have substantially  
994 expanded and reorganized the discussion and incorporated it into the final paragraph of  
995 Section 3.5 (lines 421-434).

996 Specifically, the revised text now explicitly acknowledges the presence of a persistent  
997 regional background pollution level across the “2+26” cities and clarifies that the New  
998 Year's Eve haze episode represents an acute, event-driven pollution enhancement  
999 superimposed on this background. We further discuss the pronounced inter-city  
1000 variability in estimated firework contributions (>80% versus <10%), attributing it to  
1001 differences in background PM<sub>2.5</sub> levels, the relative nature of contribution estimates,  
1002 and population redistribution during the Spring Festival. In addition, the potential roles

1003 of spatially heterogeneous effectiveness of firework prohibition policies (particularly  
1004 in rural and urban-rural fringe areas) and meteorological conditions are discussed in a  
1005 cautious and qualitative manner.

1006 These revisions provide a more comprehensive and mechanistic interpretation of Fig.  
1007 5 while remaining consistent with the scope and limitations of the available data. We  
1008 again thank the reviewer for this helpful suggestion.

1009

1010

1011