Supplementary Information

S1 Characteristics of Calm Episodes

Figure S1 summarizes the distributions of episode-level chemical and meteorological variables for the calm episodes identified during the heating seasons of 2016–2023. Overall, the selected episodes are characterized by low wind speeds, relatively shallow boundary layers, and moderate humidity, consistent with meteorological conditions under which local pollutant accumulation is favored and regional transport is comparatively reduced.

The mean wind speeds during the identified episodes were 1.4, 1.0, and 1.1 m s^{-1} for Suburban, UrbanNW, and UrbanSE, respectively, indicating generally weak horizontal advection across all three clusters. The corresponding mean boundary layer heights were 318–428 m, suggesting limited vertical dilution during the selected periods.

The distributions also show clear spatial heterogeneity in pollutant levels across the three clusters. UrbanSE had the highest mean $PM_{2.5}$ concentration (37.9 μg m^{-3}), followed by Suburban (33.7 μg m^{-3}) and UrbanNW (29.5 μg m^{-3}). NO_2 concentrations were substantially higher in the two urban clusters than in Suburban, with mean values of 34.3 and 34.1 μg m^{-3} in UrbanNW and UrbanSE, respectively, compared with 23.3 μg m^{-3} in Suburban. This contrast is consistent with stronger local anthropogenic emissions in urban areas. In contrast, O_3 concentrations were higher in Suburban, which is consistent with weaker titration by freshly emitted NO in less urbanized environments.

The distributions of SO_2 were concentrated at relatively low concentrations in all clusters, with mean values of 3.7–4.4 μg m^{-3}, reflecting the generally low ambient SO_2 levels during the study period. Temperature and pressure distributions were broadly similar across clusters, whereas relative humidity was slightly higher in the two urban clusters. Episode duration was generally short, with mean durations of 7.3–8.3 h, indicating that the selected calm episodes mainly capture transient accumulation periods.

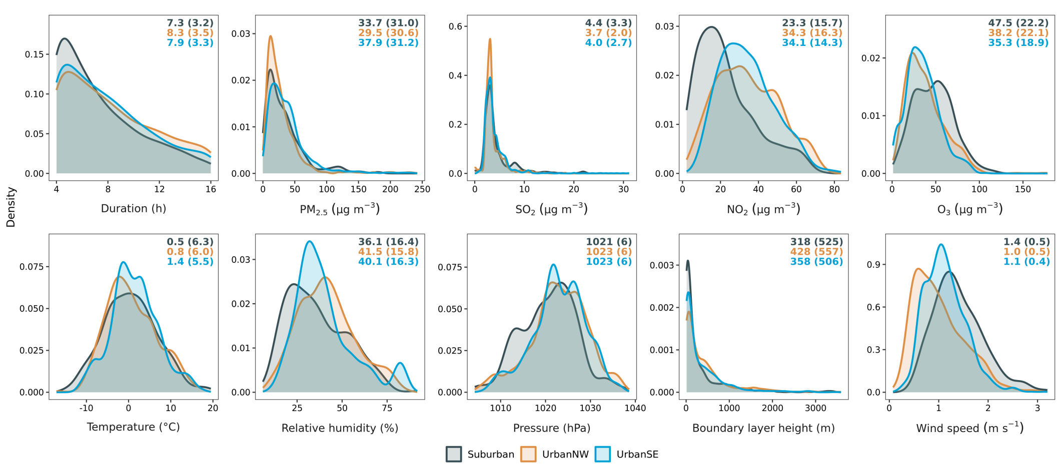

Figure S1. Distributions of episode-level variables during calm episodes in heating seasons (2016–2023). Colored labels in the upper-right corner of each panel indicate the mean and standard deviation for the corresponding variable within each site cluster.

To further characterize the meteorological context of the identified calm episodes, we examined the transition from calm to non-calm periods by extending each episode by five hours. As shown in Fig. S2, this transition is associated with systematic but non-uniform changes in key meteorological variables.

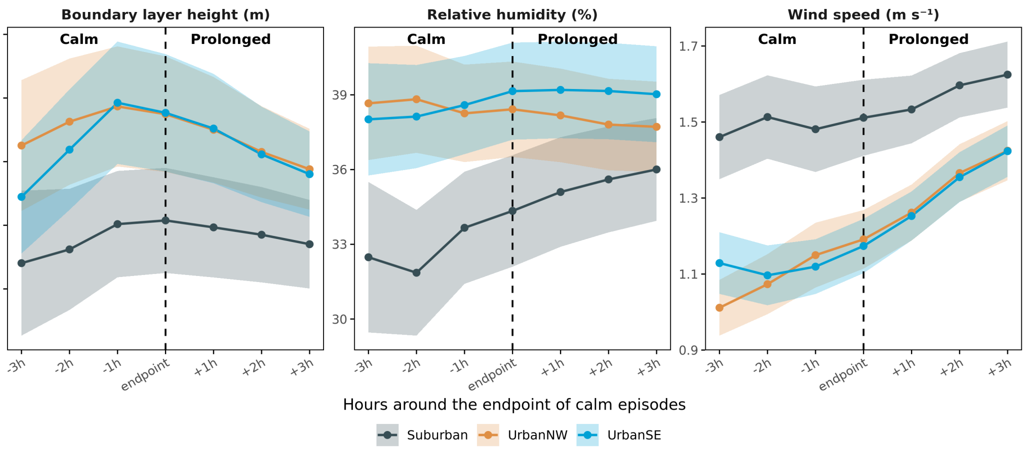

Figure S2. The boundary layer height (BLH), relative humidity (HUMI) and wind speed (WS) aligned relative to the endpoint of calm episodes (dashed vertical line), averaged across episodes for each site cluster. Negative hours denote the final three hours of the calm period, and positive hours denote the subsequent prolonged period. Shaded bands indicate 95% confidence intervals.

Near the episode endpoint, BLH exhibits a turning point, indicating a shift in vertical mixing conditions. Wind speed increases continuously across the transition, reflecting enhanced horizontal advection. In contrast, HUMI shows heterogeneous patterns across site clusters, with no consistent directional change in urban areas. According to the absolute values of the meteorological variables, calm episodes can be interpreted as periods characterized by low wind speeds and suppressed vertical mixing, which are conducive to local accumulation. Their termination, in contrast, reflects the onset of more complex dynamical regimes rather than a uniform enhancement of all dispersion-related processes.

Given this complexity, the validity of the calm-episode definition is evaluated based on its ability to isolate a regime in which precursor–$PM_{2.5}$ relationships can be estimated with reduced influence from unobserved transport processes. Supporting evidence is provided in Sect. S2.5, where we show that extending episodes to include non-calm periods introduces additional model instability and transport-induced bias consistent with theoretical expectations.

S2 Linear Model Framework

S2.1 Hourly Structural System During Calm Episodes and Episode-Level Approximation

Figure 2 prescribes that under calm atmospheric conditions the $PM_{2.5}$ concentration Y_t at time t is jointly influenced by several factors. One is its previous state $Y_{t\text{-}1}$, which provides the baseline $PM_{2.5}$ level and modulates both heterogeneous reaction rates and natural removal processes. Another is the lagged precursor trajectory $\overrightarrow{\mathbf{A}}_{t-\tau:t-1}$, representing the availability of SO_2 and NO_2 within a τ-hour effective exposure window that determines the burden of secondary aerosol formation through cumulative chemical conversion. Local emission intensity $U_{t\text{-}1}$ constitutes another key factor, contributing directly to contemporaneous $PM_{2.5}$ increments. Moreover, meteorological and oxidative conditions, summarized by $\mathbf{M}_{t\text{-}1}$ and $O_{3,t\text{-}1}$, jointly regulate secondary formation and accumulation processes of $PM_{2.5}$.

We assume that, within each calm episode that excludes regional transport confounding, $PM_{2.5}$ evolves according to a time-homogeneous first-order autoregressive process driven primarily by local primary emissions and secondary formation. Let i index years, and let $j = 1, \ldots, n_i$ index calm episodes within year i. Let t denote the hourly time index within episode j, with $t = 1, \ldots, L_{ij}$, where L_{ij} is the duration of episode j. Let $Y_{ij,t}$ denote the $PM_{2.5}$ concentration at hour t of episode j. For $t > \tau$, the hourly evolution of $PM_{2.5}$ within episode j is represented by

$$Y_{ij,t} = \psi_0 + \psi_Y Y_{ij,t-1} + \sum_{k=1}^{\tau} \boldsymbol{\psi}_{A,k}^\top \mathbf{A}_{ij,t-k} + \boldsymbol{\psi}_M^\top \mathbf{M}_{ij,t-1} + \psi_O O_{3,ij,t-1} + \psi_U U_{ij,t-1} + \sum_r \psi_r \mathbb{I}(\text{year}_{ij} = r) + \varepsilon_{Y,ij,t}, \tag{S1}$$

where the autoregressive coefficient ψ_Y captures persistence in $PM_{2.5}$ within an episode through the carryover of the previous-hour concentration $Y_{ij,t-1}$. We impose $|\psi_Y| < 1$ to ensure dynamic stability of the autoregressive system, which implies geometrically decaying lag weights and guarantees that the cumulative representation derived below is well-defined. This condition is consistent with the dissipative nature of atmospheric processes, where removal and dispersion mechanisms prevent unbounded accumulation of pollutants. The lag-specific vectors $\boldsymbol{\psi}_{A,k}$ quantify the contributions of precursor concentrations at lag k hours, $\mathbf{A}_{ij,t-k}$, to the current $PM_{2.5}$ level. These terms capture distributed-lag effects of precursor concentrations over an exposure window of length τ, which is selected using statistical criteria. The coefficients $\boldsymbol{\psi}_M$ and ψ_O capture the influence of meteorological state variables $\mathbf{M}_{ij,t-1}$ and the oxidative environment, proxied by ozone $O_{3,ij,t-1}$, on $PM_{2.5}$ evolution by modulating dispersion conditions and chemical production efficiency. The coefficient ψ_U links the latent emission intensity $U_{ij,t-1}$ to $PM_{2.5}$, thereby controlling for unobserved hour-to-hour variation in local primary emissions and co-emitted precursors that may confound observed precursor–$PM_{2.5}$ relationships. The year fixed effects ψ_r absorb unobserved interannual variation in emissions and atmospheric chemical regimes not captured by observed covariates, and $\varepsilon_{Y,ij,t}$ is an error term with zero conditional mean.

The causal pathway from NO_2 to O_3 is not explicitly modeled in our causal framework (Fig. 2). Under the selected calm conditions, the short-term co-variability between NO_2 and O_3 would reflect rapid NO–NO_2–O_3 cycling approaching a quasi-steady state (Mannschreck et al., 2004), such that their variation primarily represents differences in the prevailing oxidative environment across episodes rather than directional chemical production. Accordingly, the estimated NO_2 response should be interpreted as a conditional exposure–response relationship at comparable oxidative states, rather than a total effect that includes possible pathways mediated through changes in O_3.

The proxy variables $\mathbf{W}_{ij,t} = (\text{CO}_{ij,t}, \text{PM}_{2.5\text{-}10,ij,t})^\top$ evolve according to

$$\mathbf{W}_{ij,t} = \boldsymbol{\alpha}_W^\top \mathbf{W}_{ij,t-1} + \boldsymbol{\alpha}_M^\top \mathbf{M}_{ij,t-1} + \alpha_U U_{ij,t-1} + \sum_r \alpha_r \mathbb{I}(\text{year}_{ij} = r) + \varepsilon_{W,ij,t}, \tag{S2}$$

where $\varepsilon_{W,ij,t}$ is an error term with zero conditional mean. This auxiliary system characterizes the evolution of proxy variables used to construct control functions for the latent emission intensity $U_{ij,t-1}$, which helps restore conditional exogeneity of observed covariates.

Evaluating Eq. (S1) at time point $t > \tau$ and applying recursive substitution for each lagged $Y_{ij,t}$ yields

$$Y_{ij,t} = \psi_Y^t Y_{ij,0} + \sum_{\ell=1}^{t} \psi_Y^{t-\ell} \left[\sum_{k=1}^{\tau} \boldsymbol{\psi}_{A,k}^\top \mathbf{A}_{ij,\ell-k} + \boldsymbol{\psi}_M^\top \mathbf{M}_{ij,\ell-1} + \psi_O O_{3,ij,\ell-1} + \psi_U U_{ij,\ell-1} \right] + \sum_r \psi_r \mathbb{I}(\text{year}_{ij} = r) \sum_{\ell=0}^{t-1} \psi_Y^\ell + \eta_{ij,t}, \tag{S3}$$

where $\eta_{ij,t}$ collects the residuals. For notational convenience and to facilitate an episode-level representation, we define the τ-lagged local average

$$\overline{\mathbf{A}}_{ij,\ell,\tau} = \frac{1}{\tau} \sum_{k=1}^{\tau} \mathbf{A}_{ij,\ell-k}.$$

We impose a local homogeneity assumption on the distributed-lag effects, namely that $\boldsymbol{\psi}_{A,k} \approx \boldsymbol{\psi}_A$ for $k \le \tau^*$. This assumption is appropriate when precursor effects vary smoothly over short lag horizons, and allows the distributed-lag structure to be approximated by the local average exposure $\overline{\mathbf{A}}_{ij,\ell,\tau^*}$ (selected using statistical criteria). Under this assumption, the summation of lagged effects can be expressed in terms of local averages $\overline{\mathbf{A}}_{ij,\ell,\tau^*}$, which summarize the overall intensity of lagged precursor exposures. With these definitions, Eq. (S3) can be reparameterized in the following compact form:

$$Y_{ij,t} = \psi_Y^t Y_{ij,0} + \sum_{\ell=1}^{t} \psi_Y^{t-\ell} \left[\boldsymbol{\psi}_A^\top \overline{\mathbf{A}}_{ij,\ell,\tau^*} + \boldsymbol{\psi}_M^\top \mathbf{M}_{ij,\ell-1} + \psi_O O_{3,ij,\ell-1} + \psi_U U_{ij,\ell-1} \right] + \sum_r \psi_r \mathbb{I}(\text{year}_{ij} = r) \sum_{\ell=0}^{t-1} \psi_Y^\ell + \eta_{ij,t}, \tag{S4}$$

where $\eta_{ij,t}$ is redefined as a composite approximation error induced by the reparameterization.

Equation (S4) provides the dynamic foundation for the episode-level structural function (2) in the main text. For each episode ij with endpoint t, the cumulative contribution of any time-varying covariate entering Eq. (S4) takes the form

$$\sum_{\ell=1}^{t} \psi_Y^{t-\ell} V_{ij,\ell},$$

where $V_{ij,\ell}$ generically denotes the corresponding lagged or locally averaged process, including the lag-averaged precursor exposure $\overline{\mathbf{A}}_{ij,\ell,\tau^*}$, meteorological conditions $\mathbf{M}_{ij,\ell-1}$, and ozone $O_{3,ij,\ell-1}$. This expression represents a geometrically weighted sum with weights ψ_Y^{t-j}. For a given episode length L_{ij}, the weighting structure is fully determined by ψ_Y and is common across episodes, and therefore does not contribute to cross-episode variation. Identification thus relies on variation in the level and temporal evolution of covariate trajectories rather than on differences in weighting schemes.

For any covariate sequence $\{V_{ij,\ell}\}$, the weighted sum can be decomposed as

$$\sum_{\ell=1}^{t} \psi_Y^{t-\ell} V_{ij,\ell} = \left(\sum_{\ell=1}^{t} \psi_Y^{t-\ell} \right) \overline{V}_{ij} \; + \sum_{\ell=1}^{t} \psi_Y^{t-\ell} (V_{ij,\ell} - \overline{V}_{ij}),$$

where $\overline{V}_{ij} = L_{ij}^{-1} \sum_{\ell=1}^{L_{ij}} V_{ij,\ell}$ denotes the episode-level average, which serves as an approximation to the local mean over the effective time window. The first term depends on the mean level and episode length, while the second term captures within-episode deviations.

Calm episodes are typically short and exhibit relatively smooth temporal evolution. Under the assumption that $|\psi_Y| < 1$, the geometric weighting places more emphasis on recent observations, which are strongly correlated with the episode mean. Consequently, the deviation term becomes smaller, and the weighted sum is well approximated by the mean component.

Furthermore, since the cumulative weight $\sum_{\ell=1}^{t} \psi_Y^{t-\ell}$ increases with episode length indicated by t, the episode duration L_{ij} is included to capture variation in the overall scale of cumulative exposure. This inclusion allows the variation in cumulative weighting induced by episode length to be captured through L_{ij}, while its interaction with exposure level is reflected in the joint effects of $\overline{V}_{ij}$ and L_{ij}.

Our estimation further utilizes all hourly observations with $t \geq \tau^*$ within each episode. This enables each episode of length L_{ij} to contribute multiple effective subsamples associated with different endpoints t and overlapping histories. Through the repeated use of partial trajectories, within-episode temporal patterns are implicitly incorporated in the estimation, with the episode-level mean capturing the dominant source of cross-episode variation.

In this framework, the episode-level average precursor concentration $\overline{\mathbf{A}}_{ij,\tau^*}$ serves as a low-dimensional approximation of sustained precursor availability within episodes. As a result, the resulting estimates reflect the average causal response to sustained exposure levels under the prevailing trajectory shapes observed in the data.

Aggregating the hourly dynamics in Eq. (S1) over each episode under the sustained-exposure approximation, we define episode-level averages as

$$\overline{\mathbf{A}}_{ij,\tau} = \frac{1}{L_{ij}} \sum_{t=1}^{L_{ij}} \left(\frac{1}{\tau} \sum_{k=1}^{\tau} \mathbf{A}_{ij,t-k} \right), \quad \overline{\mathbf{M}}_{ij} = \frac{1}{L_{ij}} \sum_{t=1}^{L_{ij}} \mathbf{M}_{ij,t}, \quad \overline{O}_{3,ij} = \frac{1}{L_{ij}} \sum_{t=1}^{L_{ij}} O_{3,ij,t}.$$

The corresponding episode-level model is defined as:

$$Y_{ij} = \psi_0 + \psi_Y Y_{0,ij} + \boldsymbol{\psi}_A^\top \overline{\mathbf{A}}_{ij,\tau} + \boldsymbol{\psi}_M^\top \overline{\mathbf{M}}_{ij} + \psi_O \overline{O}_{3,ij} + \psi_U \overline{U}_{ij} + \psi_L L_{ij} + \sum_r \psi_r \mathbb{I}(\text{year}_{ij} = r) + \eta_{ij}, \tag{S5}$$

where L_{ij} captures episode-level variation in the cumulative temporal weighting induced by autoregressive persistence ψ_Y, $\overline{U}_{ij}$ denotes the episode-level latent emission implicitly controlled for by the proxy-based control function, and η_{ij} collects approximation errors.

S2.2 Proxy Control Function and Orthogonalization

At the episode level, the latent emission factor $\overline{U}_{ij}$ enters both the outcome equation and the proxy equation,

$$\begin{aligned} Y_{ij} &= \psi_0 + \boldsymbol{\psi}_A^\top \overline{\mathbf{A}}_{ij,\tau^*} + \boldsymbol{\psi}_M^\top \overline{\mathbf{M}}_{ij} + \psi_Y Y_{0,ij} + \psi_O \overline{O}_{3,ij} + \psi_U \overline{U}_{ij} \\ &\quad + \psi_L L_{ij} + \sum_r \psi_r \mathbb{I}(\text{year}_{ij} = r) + \varepsilon_{Y,ij}, \end{aligned} \tag{S6}$$

$$\mathbf{W}_{ij} = \alpha_0 + \boldsymbol{\alpha}_M^\top \overline{\mathbf{M}}_{ij} + \boldsymbol{\alpha}_W^\top \mathbf{W}_{0,ij} + \alpha_U \overline{U}_{ij} + \alpha_L L_{ij} + \sum_r \alpha_r \mathbb{I}(\text{year}_{ij} = r) + \varepsilon_{W,ij}, \tag{S7}$$

where Y_{ij} and $\mathbf{W}_{ij}$ denote the terminal $PM_{2.5}$ and proxy (CO and $PM_{2.5\text{-}10}$) concentrations, $\overline{\mathbf{A}}_{ij,\tau^*}$, $\overline{\mathbf{M}}_{ij}$, and $\overline{O}_{3,ij}$ are episode-level averages, L_{ij} is episode duration, and year indicators absorb unobserved temporal heterogeneity.

Since $\mathbf{W}_{ij}$ reflects both the latent confounder $\overline{U}_{ij}$ and observed covariates, directly including it in the outcome equation would introduce non-causal pathways influencing Y_{ij}. We therefore residualize $\mathbf{W}_{ij}$ with respect to $\overline{\mathbf{M}}_{ij}$, $\mathbf{W}_{0,ij}$, L_{ij}, and year effects. The resulting auxiliary projection,

$$\widetilde{\mathbf{W}}_{ij} = \mathbf{W}_{ij} - \widehat{\mathbb{E}}[\mathbf{W}_{ij} \mid \mathbf{W}_{0,ij}, \overline{\mathbf{M}}_{ij}, L_{ij}, \text{year}_{ij}], \tag{S8}$$

$$\widehat{\mathbb{E}}[\mathbf{W}_{ij} \mid \mathbf{W}_{0,ij}, \overline{\mathbf{M}}_{ij}, L_{ij}, \text{year}_{ij}] = \alpha_0^* + \boldsymbol{\alpha}_M^{*T} \overline{\mathbf{M}}_{ij} + \boldsymbol{\alpha}_W^{*T} \mathbf{W}_{0,ij} + \alpha_L^* L_{ij} + \sum_r \alpha_r^* \mathbb{I}(\text{year}_{ij} = r), \tag{S9}$$

is intentionally misspecified by omitting $\overline{U}_{ij}$. Under Assumptions S3 and S4, the residual $\widetilde{\mathbf{W}}_{ij}$ isolates the variation related to $\overline{U}_{ij}$ and provides the necessary information to construct a control function that restores conditional exogeneity of the regressors.

While CO and $PM_{2.5\text{-}10}$ serve as imperfect tracers for combustion and dust emissions respectively, ambient concentrations reflect complex mixtures from multiple sources. Combustion processes can generate both gaseous CO and particulate matters, while dust emissions may co-vary with certain combustion activities. To reduce multicollinearity and extract dominant variation associated with the latent confounder, orthogonal rotation transformation is applied to $\widetilde{\mathbf{W}}_{ij}$. Let $\mathbf{P}$ denote orthogonal rotation matrix, and the orthogonal rotated proxy scores are

$$\mathbf{e}_{ij} = \mathbf{P}^\top \widetilde{\mathbf{W}}_{ij}. \tag{S10}$$

The estimating equation used in the main text is

$$Y_{ij} = \psi_0^* + \boldsymbol{\psi}_A^{*\top} \overline{\mathbf{A}}_{ij,\tau^*} + \boldsymbol{\psi}_M^{*\top} \overline{\mathbf{M}}_{ij} + \psi_Y^* Y_{0,ij} + \psi_O^* \overline{O}_{3,ij} + \boldsymbol{\psi}_e^{*\top} \mathbf{e}_{ij} + \psi_L^* L_{ij} + \sum_r \psi_r^* \mathbb{I}(\text{year}_{ij} = r) + \nu_{ij}. \tag{S11}$$

Under Assumptions S1–S7 in Sect. S2.2.1, $\widehat{\boldsymbol{\psi}}_A^*$ is consistent for the causal parameter $\boldsymbol{\psi}_A$ in the structural model. A formal proof is provided in Sect. S2.2.2 Theorem S1.

S2.2.1 Identification Assumptions

Assumption S1 (Error Orthogonality)**.** *The error terms satisfy:*

$$\mathbb{E}[\varepsilon_{Y,ij} \mid \overline{\mathbf{A}}_{ij,\tau^*}, \overline{\mathbf{M}}_{ij}, Y_{0,ij}, \overline{O}_{3,ij}, \overline{U}_{ij}, L_{ij}, \mathit{year}_{ij}] = 0, \quad \mathbb{E}[\varepsilon_{W,ij} \mid \overline{\mathbf{M}}_{ij}, \mathbf{W}_{0,ij}, \overline{U}_{ij}, L_{ij}, \mathit{year}_{ij}] = 0.$$

Assumption S2 (Measurement Error Orthogonality)**.** *The measurement error $\varepsilon_{W,ij}$ in the proxy model is uncorrelated with the structural error $\varepsilon_{Y,ij}$:* $\mathbb{E}[\varepsilon_{W,ij}\varepsilon_{Y,ij}] = 0$.

Assumption S3 (Exclusion Restriction)**.** *The residualized proxy* $\widetilde{\mathbf{W}}_{ij}$ *does not have a direct causal effect on* Y_{ij} *conditional on* $\overline{U}_{ij}$ *and observed covariates.*

Assumption S4 (Control Function Representation)**.** *The latent confounder* $\overline{U}_{ij}$ *can be linearly represented using the residualized proxy:*

$$\overline{U}_{ij} = \boldsymbol{\gamma}^T \widetilde{\mathbf{W}}_{ij} + \zeta_{ij},$$

where $\mathbb{E}[\zeta_{ij} \mid \widetilde{\mathbf{W}}_{ij}, \overline{\mathbf{A}}_{ij,\tau^*}, \overline{\mathbf{M}}_{ij}, Y_{0,ij}, \overline{O}_{3,ij}, L_{ij}, \mathit{year}_{ij}] = 0$ *and* $\mathbb{E}[\zeta_{ij}\varepsilon_{Y,ij}] = 0$.

Assumption S5 (Orthogonal Rotation Approximation Quality)**.** *The orthogonally rotated proxy* $\mathbf{e}_{ij} = \mathbf{P}_k^\top \widetilde{\mathbf{W}}_{ij}$*, where* $\mathbf{P}_k$ *is an orthogonal matrix (*$\mathbf{P}^\top\mathbf{P} = \mathbf{I}$*), provides a sufficiently accurate representation such that:*

$$\psi_U \boldsymbol{\gamma}^T \widetilde{\mathbf{W}}_{ij} = \boldsymbol{\psi}_e^{*T} \mathbf{e}_{ij} + \delta_{ij},$$

where the approximation error δ_{ij} *satisfies:*

$$\frac{1}{n}\sum_i \sum_{j=1}^{n_i} \mathbf{Z}_{ij}\delta_{ij} \xrightarrow{p} 0 \quad \text{and} \quad \frac{1}{n}\sum_i \sum_{j=1}^{n_i} \mathbf{e}_{ij}\delta_{ij} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

with $\mathbf{Z}_{ij} = [1, \overline{\mathbf{A}}_{ij,\tau^*}^T, \overline{\mathbf{M}}_{ij}^T, Y_{0,ij}, \overline{O}_{3,ij}, L_{ij}, \mathbb{I}(\mathit{year}_{ij} = 1), \ldots, \mathbb{I}(\mathit{year}_{ij} = R)]^T$ *and* $n = \sum_i n_i$.

Assumption S6 (First-Stage Misspecification)**.** *The misspecified first-stage model* (S9) *produces residuals* $\widetilde{\mathbf{W}}_{ij}$ *that preserve the relevant variation for identifying* $\boldsymbol{\psi}_A$*, even if the estimated coefficients* $\boldsymbol{\alpha}_M^*$*,* $\boldsymbol{\alpha}_W^*$ *and* α_r^* *are biased. Specifically,* $\widetilde{\mathbf{W}}_{ij}$ *contains sufficient information about* $\overline{U}_{ij}$ *to construct a valid control function.*

Assumption S7 (Full Rank and Regularity)**.** *The regressor matrix*

$$\mathbf{X}_{ij} = [1, \overline{\mathbf{A}}_{ij,\tau^*}^T, \overline{\mathbf{M}}_{ij}^T, Y_{0,ij}, \overline{O}_{3,ij}, L_{ij}, \mathbb{I}(\mathit{year}_{ij} = 1), \ldots, \mathbb{I}(\mathit{year}_{ij} = R), \mathbf{e}_{ij}^T]^T$$

has full rank asymptotically, and $\frac{1}{n}\sum_i \sum_{j=1}^{n_i} \mathbf{X}_{ij}\mathbf{X}_{ij}^T$ *converges to a positive definite matrix* Q_{XX} *as* $n = \sum_i n_i \to \infty$.

S2.2.2 Consistency Theorem and Proof

Theorem S1 (Consistency of Orthogonal Rotation-Control Function Estimator)**.** *Under Assumptions S1-S7, the ordinary least squares estimator* $\widehat{\boldsymbol{\psi}}_A^*$ *from Eq.* (S11) *is consistent for the structural parameter* $\boldsymbol{\psi}_A$ *in Eq.* (S6)*:*

$$\widehat{\boldsymbol{\psi}}_A^* \xrightarrow{p} \boldsymbol{\psi}_A \quad \text{as} \quad n \to \infty.$$

Proof. The proof proceeds by constructing an asymptotically equivalent estimating equation and systematically verifying orthogonality conditions.

From Assumption S4, we express the latent confounder as:

$$\overline{U}_{ij} = \boldsymbol{\gamma}^T \widetilde{\mathbf{W}}_{ij} + \zeta_{ij},$$

where $\mathbb{E}[\zeta_{ij} \mid \widetilde{\mathbf{W}}_{ij}, \overline{\mathbf{A}}_{ij,\tau^*}, \overline{\mathbf{M}}_{ij}, Y_{0,ij}, \overline{O}_{3,ij}, L_{ij}, \text{year}_{ij}] = 0$ and $\mathbb{E}[\zeta_{ij}\varepsilon_{Y,ij}] = 0$. Substituting into Eq. (S6):

$$\begin{aligned} Y_{ij} =& \psi_0 + \boldsymbol{\psi}_A^T \overline{\mathbf{A}}_{ij,\tau^*} + \boldsymbol{\psi}_M^T \overline{\mathbf{M}}_{ij} + \psi_Y Y_{0,ij} + \psi_O \overline{O}_{3,ij} + \psi_L L_{ij} + \sum_{r=1} \psi_r \mathbb{I}(\text{year}_{ij} = r) + \psi_U(\boldsymbol{\gamma}^T \widetilde{\mathbf{W}}_{ij} + \zeta_{ij}) + \varepsilon_{Y,ij} \\ =& \psi_0 + \boldsymbol{\psi}_A^T \overline{\mathbf{A}}_{ij,\tau^*} + \boldsymbol{\psi}_M^T \overline{\mathbf{M}}_{ij} + \psi_Y Y_{0,ij} + \psi_O \overline{O}_{3,ij} + \psi_L L_{ij} + \sum_{r=1} \psi_r \mathbb{I}(\text{year}_{ij} = r) + \psi_U \boldsymbol{\gamma}^T \widetilde{\mathbf{W}}_{ij} + (\varepsilon_{Y,ij} + \psi_U \zeta_{ij}). \end{aligned} \tag{S12}$$

By Assumption S5, we have:

$$\psi_U \boldsymbol{\gamma}^T \widetilde{\mathbf{W}}_{ij} = \boldsymbol{\psi}_e^{*T} \mathbf{e}_{ij} + \delta_{ij},$$

where δ_{ij} is the approximation error satisfying $\frac{1}{n}\sum_i \sum_{j=1}^{n_i} \mathbf{Z}_{ij}\delta_{ij} \xrightarrow{p} 0$ and $\frac{1}{n}\sum_i \sum_{j=1}^{n_i} \mathbf{e}_{ij}\delta_{ij} \xrightarrow{p} 0$ as $n \to \infty$.

Substituting into Eq. (S12):

$$Y_{ij} = \psi_0 + \boldsymbol{\psi}_A^T \overline{\mathbf{A}}_{ij,\tau^*} + \boldsymbol{\psi}_M^T \overline{\mathbf{M}}_{ij} + \psi_Y Y_{0,ij} + \psi_O \overline{O}_{3,ij} + \psi_L L_{ij} + \sum_{r=1} \psi_r \mathbb{I}(\text{year}_{ij} = r) + \boldsymbol{\psi}_e^{*T} \mathbf{e}_{ij} + (\varepsilon_{Y,ij} + \psi_U \zeta_{ij} + \delta_{ij}). \tag{S13}$$

Comparing Eq. (S13) with the estimating Eq. (S11), we identify:

$$\psi_0^* = \psi_0, \quad \boldsymbol{\psi}_A^* = \boldsymbol{\psi}_A, \quad \boldsymbol{\psi}_M^* = \boldsymbol{\psi}_M, \quad \psi_Y^* = \psi_Y,$$
$$\psi_r^* = \psi_r, \quad \psi_O^* = \psi_O, \quad \psi_L^* = \psi_L, \quad \nu_{ij} = \varepsilon_{Y,ij} + \psi_U \zeta_{ij} + \delta_{ij}.$$

for $r = 1, \ldots, R$, crucially, we have $\boldsymbol{\psi}_A^* = \boldsymbol{\psi}_A$.

To establish consistency, we need to verify that the regressors in Eq. (S11) are asymptotically uncorrelated with the composite error term $\nu_{ij} = \varepsilon_{Y,ij} + \psi_U \zeta_{ij} + \delta_{ij}$. Let $\mathbf{X}_{ij} = [1, \overline{\mathbf{A}}_{ij,\tau^*}^T, \overline{\mathbf{M}}_{ij}^T, Y_{0,ij}, \overline{O}_{3,ij}, L_{ij}, \mathbb{I}(\text{year}_{ij} = 1), \ldots, \mathbb{I}(\text{year}_{ij} = R), \mathbf{e}_{ij}^T]^T$ denote the vector of all regressors. The OLS estimator for $\boldsymbol{\psi}_A^*$ is given by:

$$\widehat{\boldsymbol{\psi}}_A^* = \boldsymbol{\psi}_A + \left(\frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{X}_{ij} \mathbf{X}_{ij}^T \right)^{-1} \left(\frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{X}_{ij} \nu_{ij} \right). \tag{S14}$$

We analyze each component of ν_{ij} separately:

Structural error $\varepsilon_{Y,ij}$: By Assumption S1, $\mathbb{E}[\varepsilon_{Y,ij} \mid \overline{\mathbf{A}}_{ij,\tau^*}, \overline{\mathbf{M}}_{ij}, Y_{0,ij}, \overline{O}_{3,ij}, \overline{U}_{ij}, L_{ij}, \text{year}_{ij}] = 0$. Given the exclusion restriction (Assumption S3) and the construction of $\mathbf{e}_{ij}$ from $\widetilde{\mathbf{W}}_{ij}$, we have:

$$\frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{X}_{ij} \varepsilon_{Y,ij} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty.$$

Representation error $\psi_U \zeta_{ij}$: By Assumption S4, $\mathbb{E}[\zeta_{ij} \mid \widetilde{\mathbf{W}}_{ij}, \overline{\mathbf{A}}_{ij,\tau^*}, \overline{\mathbf{M}}_{ij}, Y_{0,ij}, \overline{O}_{3,ij}, L_{ij}, \text{year}_{ij}] = 0$, which implies ζ_{ij} is uncorrelated with all regressors in $\mathbf{X}_{ij}$. Therefore:

$$\frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{X}_{ij} (\psi_U \zeta_{ij}) \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty.$$

Orthogonal rotation approximation error δ_{ij}: By Assumption S5,

$$\frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{Z}_{ij} \delta_{ij} \xrightarrow{p} 0 \quad \text{and} \quad \frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{e}_{ij} \delta_{ij} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

which together imply:

$$\frac{1}{n} \sum_i \sum_{j=1}^{n_i} \mathbf{X}_{ij} \delta_{ij} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty.$$

Combining these results:

$$\frac{1}{n}\sum_{i}\sum_{j=1}^{n_i}\mathbf{X}_{ij}\nu_{ij} \xrightarrow{p} \frac{1}{n}\sum_{i}\sum_{j=1}^{n_i}\mathbf{X}_{ij}(\varepsilon_{Y,ij}+\psi_U\zeta_{ij}+\delta_{ij}) \xrightarrow{p} 0 \quad \text{as} \quad n\to\infty. \tag{S15}$$

By Assumption S7, $\frac{1}{n}\sum_i\sum_{j=1}^{n_i}\mathbf{X}_{ij}\mathbf{X}_{ij}^T$ converges to a positive definite matrix Q_{XX}. Therefore:

$$\widehat{\boldsymbol{\psi}}_A^* \xrightarrow{p} \boldsymbol{\psi}_A + Q_{XX}^{-1}\cdot 0 \xrightarrow{p} \boldsymbol{\psi}_A \quad \text{as} \quad n\to\infty. \tag{S16}$$

□

S2.3 Extension to the nonlinear GAM setting

The consistency result in Theorem S1 is proved under the linear structural models (S6)–(S7). In the main text, we further extend the outcome model to a generalized additive specification (Eq. (8)) that accommodates nonlinear and regime-dependent precursor effects. This extension raises the question whether the proxy-based confounding adjustment remains valid.

The core identification strategy of removing the influence of the latent confounder $\overline{U}_{ij}$ via the orthogonalised proxy indices $\mathbf{e}_{ij}$ carries over to the nonlinear setting under two conditions. First, the control function representation (Assumption S4) and the orthogonality of the approximation error (Assumption S5) are independent of the functional form of the outcome equation; they concern only the relationship between the latent confounder and the observed proxies. Second, under the standard identifiability constraints of generalized additive models (Wood, 2017), the additive components are uniquely determined up to an overall intercept. Once $\mathbf{e}_{ij}$ is included as an additive smooth term, it absorbs the residual emission-driven variability, leaving the precursor-specific response surfaces $\phi_{\mathrm{SO_2}}$ and $\phi_{\mathrm{NO_2}}$ to be estimated free of emission-related confounding.

Formally, suppose the true conditional expectation admits an additive structure

$$\begin{aligned}\mathbb{E}[Y_{ij}\mid \overline{\mathbf{A}}_{ij,\tau^*},\mathbf{Z}_{ij},\mathbf{e}_{ij},\mathrm{year}_{ij}] =& \beta_{\mathrm{year}_{ij}} + \phi^0_{\mathrm{SO_2}}(A_{\mathrm{SO_2},ij},E^*_{\mathrm{SO_2},ij}) + \phi^0_{\mathrm{NO_2}}(A_{\mathrm{NO_2},ij},E^*_{\mathrm{NO_2},ij}) \\ &+ s^0_{e_1}(e_{1,ij}) + s^0_{e_2}(e_{2,ij}) + \sum_V s^0_V(V_{ij}),\end{aligned} \tag{S17}$$

where each additive component belongs to a suitable smoothness class. Under the same proxy-variable assumptions and standard regularisation conditions on the spline basis (for instance, basis dimension $K_n\to\infty$ with $K_n/n\to 0$), the penalised spline estimates $\widehat{\phi}_{\mathrm{SO_2}}$ and $\widehat{\phi}_{\mathrm{NO_2}}$ are consistent for $\phi^0_{\mathrm{SO_2}}$ and $\phi^0_{\mathrm{NO_2}}$ in the L_2 sense (Claeskens et al., 2009; Wood, 2017). A rigorous proof is beyond the scope of this supplementary information but follows from standard arguments in nonparametric regression theory.

S2.4 Selection of Exposure Windows

The episode-level exposure variable requires specifying the time window over which recent precursor concentrations are aggregated. Atmospheric studies have shown that sulfate and nitrate formation can occur through multiple pathways, including gas-phase oxidation, heterogeneous uptake, and aqueous-phase reactions, with rates that vary strongly with oxidant availability, relative humidity, aerosol liquid water, particle acidity, and ammonia availability (Brown and Stutz, 2012; Cheng et al., 2016; Wang et al., 2016; Liu et al., 2021). Rather than attempting to infer a pathway-specific chemical lifetime, we used the observed time series to identify an effective exposure-response window: the duration over which lagged SO_2 and NO_2 concentrations retain additional explanatory power for hourly $PM_{2.5}$ evolution after adjustment for previous $PM_{2.5}$, meteorology, oxidative conditions, emission proxies, and more recent precursor levels.

To determine the temporal horizon over which precursor concentrations influence $PM_{2.5}$ growth, based on the dynamic model (S1) introduced in Sect. S2.1, we compared a sequence of nested hourly models with increasing maximum lag order. The selected lag order was then used to define the operational exposure window τ^* in the episode-level model. Starting from a baseline specification derived from Model (S1) excluding the precursor terms:

$$Y_{ij,t} = \psi_0 + \psi_Y Y_{ij,t-1} + \boldsymbol{\psi}_M^\top \mathbf{M}_{ij,t-1} + \psi_O O_{3,ij,t-1} + \boldsymbol{\psi}_e \mathbf{e}_{ij,t-1} + \sum_r \psi_r \mathbb{I}(\mathrm{year}_{ij}=r) + \varepsilon_{Y,ij,t}. \tag{S18}$$

We fitted a sequence of distributed-lag models with maximum lag order $K = 1, \ldots, 6$, where the lagged SO_2 and NO_2 terms up to hour K were included.

$$Y_{ij,t} = \psi_0 + \psi_Y Y_{ij,t-1} + \sum_{k=1}^{K} \beta_k^S \mathrm{SO}_{2,ij,t-k} + \sum_{k=1}^{K} \beta_k^N \mathrm{NO}_{2,ij,t-k}$$
$$+ \boldsymbol{\psi}_M^\top \mathbf{M}_{ij,t-1} + \psi_O O_{3,ij,t-1} + \boldsymbol{\psi}_e \mathbf{e}_{ij,t-1} + \sum_r \psi_r \mathbb{I}(\mathrm{year}_{ij} = r) + \varepsilon_{Y,ij,t}. \tag{S19}$$

Lag orders beyond 6 h were not pursued, because the autoregressive dynamic of $PM_{2.5}$ (Eq. (S3)) implies geometrically decaying contributions from more distant past exposures. When $|\psi_Y| < 1$, the influence of earlier precursor concentrations declines exponentially, so distant lags contribute progressively less to the current $PM_{2.5}$ level.

We employed an episode-based bootstrap resampling procedure to evaluate the stability of lag order selection. For each site cluster, 1,000 bootstrap samples were generated by resampling complete episodes with replacement, while preserving the original number of episodes and the internal temporal structure within each episode. For each bootstrap sample, the full lag-selection procedure was repeated over $K = 1, \ldots, 6$, and the optimal lag order was identified separately by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Selection frequencies were then computed as the proportions of bootstrap samples in which each lag order was selected as optimal.

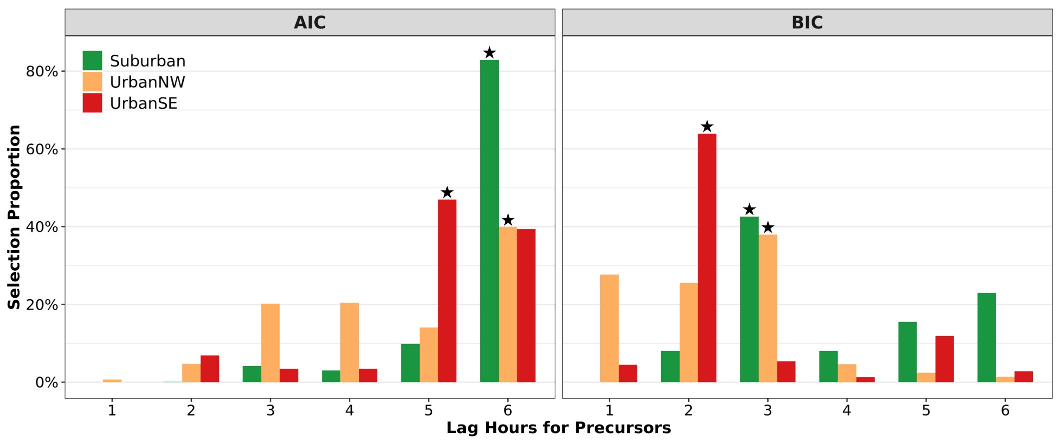

Figure S3. Distribution of selected lag orders across site clusters based on 1,000 bootstrap iterations. Stars indicate the optimal lag selected for each cluster, with Suburban and UrbanNW favoring a 3 h window and UrbanSE favoring a 2 h window.

Figure S3 summarizes the bootstrap distribution of selected precursor lag orders across site clusters under AIC and BIC. The precursor effects exhibit a clear multi-hour persistence alongside a short dominant response window. Across bootstrap samples, the most frequently selected lag under AIC reaches 5 h in UrbanSE (47.0%) and 6 h in UrbanNW (39.9%) and Suburban clusters (83.9%). Conditional on Y_{t-1}, meteorology, oxidants, emission proxies, and more recent precursor levels, earlier precursor concentrations retain additional explanatory power beyond proximal lags. This indicates that precursor contributions are not instantaneous but persist over multiple hours within the examined 6 h lag range.

Under BIC, the most frequently selected lag is 2 h in UrbanSE (63.9%) and 3 h in UrbanNW (38.0%) and Suburban clusters (42.6%). This pattern is consistent with the dynamic structure of the system, in which the influence of earlier precursor concentrations propagates forward but decays approximately exponentially, leading to rapidly diminishing marginal contributions from distant lags. The dispersion of bootstrap-selected lag orders further indicates that effective oxidation timescales vary across episodes.

For subsequent episode-level modeling, we adopt BIC-selected lag orders as the operational exposure window τ^*. In addition to its parsimony, this choice ensures compatibility with the episode-level modeling framework, where lag-specific effects are assumed to remain approximately constant within τ^*, thereby supporting stable identification of aggregated precursor effects. Accordingly, τ^* is set to 2 h for UrbanSE and 3 h for both UrbanNW and Suburban clusters, and is used consistently in all subsequent analyses.

S2.5 Transport-induced Bias under Prolonged Episode Definitions

To illustrate the impact of transport-induced confounding, we consider a linear representation of the structural model under prolonged episode definitions. In this setting, the outcome model (S6) can be extended to include an unobserved transport component R_{ij}:

$$Y_{ij} = \theta_0 + \boldsymbol{\theta}_A^\top \overline{\mathbf{A}}_{ij,\tau^*} + \boldsymbol{\theta}_M^\top \overline{\mathbf{M}}_{ij} + \theta_Y Y_{0,ij} + \theta_O \overline{O}_{3,ij} + \theta_U \overline{U}_{ij} + \theta_L L_{ij} + \theta_R R_{ij} + \sum_r \theta_r \mathbb{I}(\text{year}_{ij} = r) + \epsilon_{Y,ij}, \tag{S20}$$

where R_{ij} denotes the net contribution of regional transport to $PM_{2.5}$ during episode ij, and θ_R its associated effect. Because R_{ij} is unobserved and omitted from the estimating equation, the precursor effect estimates are subject to omitted-variable bias, whose direction depends on the sign of θ_R and the conditional association between $\overline{\mathbf{A}}_{ij,\tau^*}$ and R_{ij} given the observed covariates $\mathbf{Z}_{ij}$.

The qualitative bias patterns implied by this structure are summarized in Fig. S4. Depending on the joint configuration of θ_R and $\mathrm{Cov}(\mathbf{A}, R \mid \mathbf{Z})$, the resulting bias may be either positive or negative, and cannot be determined from observed data in the absence of direct measurements of R_{ij}.

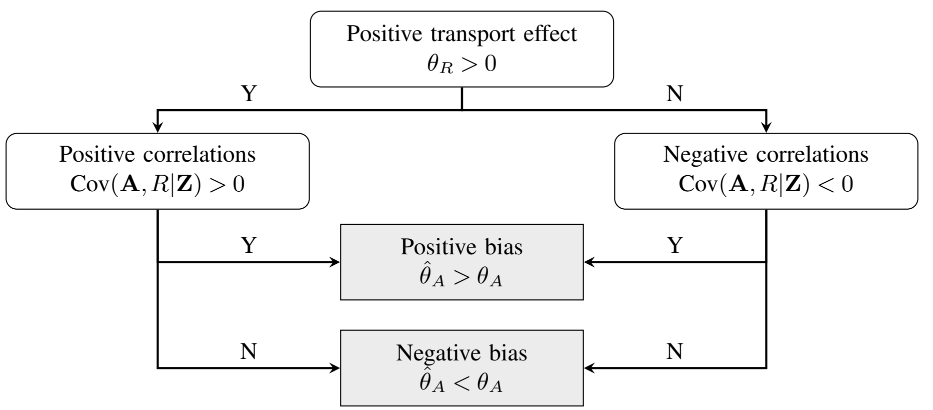

Figure S4. Schematic illustration of how the sign of the transport effect θ_R and the conditional association $\mathrm{Cov}(\mathbf{A}, R \mid \mathbf{Z})$ jointly determine the direction of omitted-variable bias in precursor effect estimates.

To examine this mechanism empirically, we construct a prolonged-episode dataset by extending each selected calm episode by 5 h. Total episode duration is capped at 16 h to maintain comparability with the calm-episode framework and to avoid introducing substantially longer event structures. Relative to the original calm episodes, these extended segments are more likely to include periods influenced by regional transport processes not represented in the assumed DAG (Fig. 2).

Under this setting, the proxy-calibrated coefficient estimates exhibit systematic shifts when prolonged episodes are included (Fig. S5). The direction and magnitude of changes in SO_2 and NO_2 coefficients vary across site clusters, indicating that transport-induced confounding is heterogeneous and depends on the underlying association between precursor levels and transport contributions.

Prolonged episodes also lead to degraded predictive performance, with RMSE increasing by 2.3 μg m^{-3} (19.6%) in the Suburban cluster, 0.6 μg m^{-3} (4.2%) in UrbanNW, and 1.8 μg m^{-3} (12.2%) in UrbanSE. The concurrent occurrence of coefficient instability and reduced model fit indicates that including transport-influenced periods introduces additional confounding that cannot be resolved through proxy calibration alone.

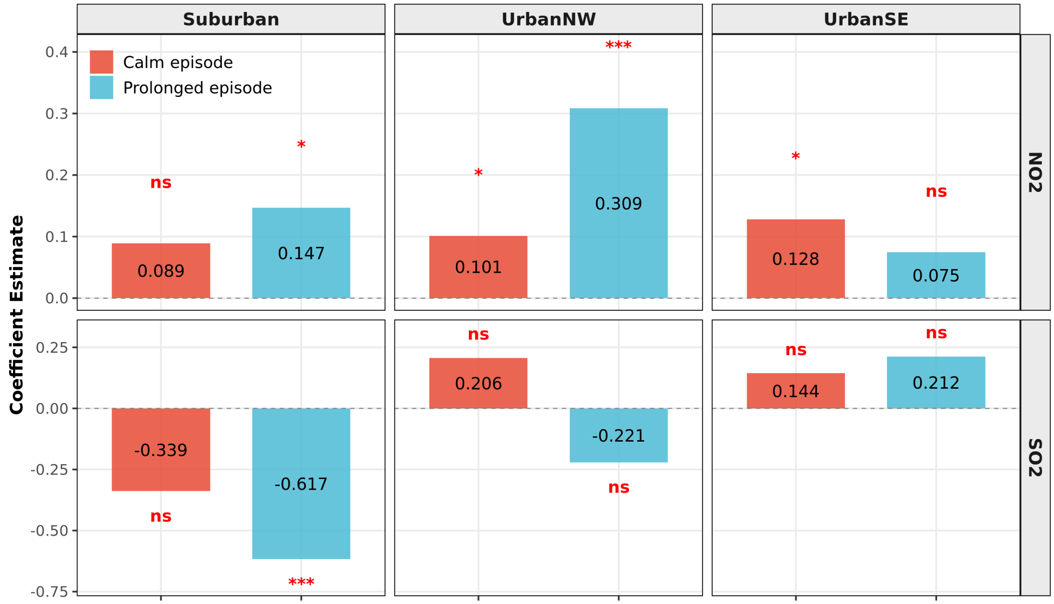

Figure S5. Estimated linear regression coefficients for SO_2 and NO_2 from the proxy-adjusted models under the calm-episode (red) and prolonged-episode (blue) definitions. Asterisks indicate statistical significance (* $p < 0.05$, *** $p < 0.001$).

These considerations motivate the restriction to calm episodes in the main analysis, for which the influence of regional transport is expected to be attenuated and the identifying assumptions of the causal framework are therefore more plausible.

S3 Nonlinear Modeling Specification

S3.1 Selection of Effect Modifiers

In the oxidation-neutralization pathways of secondary inorganic aerosol (SIA) formation, oxidation of gaseous SO_2 and NO_2 is largely governed by heterogeneous reactions involving aqueous and particle-phase processes, implying enhanced reaction efficiency under humid conditions and with greater available particle surface area, the latter proxied by the initial $PM_{2.5}$ concentration Y_0 (Wang et al., 2016; Ma et al., 2017; Dong et al., 2020). Ozone, as a key oxidant, regulates oxidation chemistry (Yang et al., 2020; Ren et al., 2021) and additionally modulates nocturnal nitrate formation through its role in NO_3/N_2O_5 chemistry (Ma et al., 2023). Precursor-to-particle conversion rates are further temperature dependent due to chemical thermodynamics and reversible reactions (Brown and Stutz, 2012; Wang et al., 2021).

Motivated by these mechanisms, we select $\{\overline{\mathrm{TEMP}}, \overline{\mathrm{HUMI}}, \overline{O}_3, Y_0\}$ from the covariates in Model (8) as candidate effect modifiers. In particular, Y_0 represents $PM_{2.5}$ conditions at episode onset following northerly cleansing events and is predetermined with respect to sustained SO_2 and NO_2 exposures within the episode, which supports its interpretation as an effect modifier rather than a downstream outcome of within-episode precursor accumulation.

To characterize heterogeneity in precursor effects across atmospheric regimes in $PM_{2.5}$ formation, we model nonlinear effect modification between precursors $\mathbf{A}$ and candidate modifiers $\mathbf{E}$ using bivariate smooths. For each precursor $p \in \{SO_2, NO_2\}$, the effect-modification function is parameterized using tensor-product splines:

$$\phi_p(A_p, E_p) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \theta^p_{k_1 k_2} B^p_{k_1}(A_p) B^p_{k_2}(E_p), \tag{S21}$$

where $B^p_{k_1}(A_p)$ are spline basis functions in the precursor dimension, with monotonicity enforced via shape-constrained P-splines, implying $\partial\phi_p/\partial A_p \geq 0$ under the working assumption that sustained precursor availability does not reduce $PM_{2.5}$ formation. The functions $B^p_{k_2}(E_p)$ are B-spline bases for candidate effect modifiers $E_p \in \{\overline{\mathrm{TEMP}}, \overline{\mathrm{HUMI}}, \overline{O}_3, Y_0\}$.

Let $\mathcal{C}_p = \{\overline{\mathrm{TEMP}}, \overline{\mathrm{HUMI}}, \overline{O}_3, Y_0\}$ denote the set of all candidate effect modifiers. For each precursor $p \in \{\mathrm{SO_2}, \mathrm{NO_2}\}$, one effect modifier E_p is selected from $\mathcal{C}_p$. The set of all candidate effect-modification structures is defined as

$$\mathcal{C} = \{(E_{\mathrm{SO_2}}, E_{\mathrm{NO_2}}) \mid E_{\mathrm{SO_2}} \in \mathcal{C}_{\mathrm{SO_2}}, E_{\mathrm{NO_2}} \in \mathcal{C}_{\mathrm{NO_2}}\} \cup \varnothing, \tag{S22}$$

where $\varnothing$ represents the case without effect modification. This yields $|\mathcal{C}| = 17$ candidate configurations.

To select the optimal set of effect modifiers, we compared candidate effect-modification structures using bootstrap-stabilized improvements in Akaike Information Criterion (AIC) relative to the no-modifier specification. For each site cluster $s \in \{\mathrm{UrbanNW}, \mathrm{Suburba}$ and each bootstrap replicate $b = 1, \ldots, N_b$ with $N_b = 500$, we fit Model (8) under each candidate configuration $E \in \mathcal{C}$ and compute the corresponding AIC:

$$\mathrm{AIC}^{(E)}_{h,s,b} = 2k - 2\ln(\hat{L}), \tag{S23}$$

where k is the number of model parameters and $\hat{L}$ is the likelihood. To ensure comparability across bootstrap samples and site clusters, model performance is evaluated in terms of the AIC improvement relative to the no-modifier specification:

$$\Delta\mathrm{AIC}^{(E)}_{s,b} = \mathrm{AIC}^{(\varnothing)}_{s,b} - \mathrm{AIC}^{(E)}_{s,b}, \tag{S24}$$

where $\varnothing$ denotes the model without effect modification. Positive values of ΔAIC indicate improved model fit after introducing effect modification.

The overall performance of each candidate configuration is summarized by its mean AIC improvement across all site clusters and bootstrap replicates:

$$\overline{\Delta\mathrm{AIC}}_E = \frac{1}{N_s N_b} \sum_s \sum_b \Delta\mathrm{AIC}^{(E)}_{s,b}, \tag{S25}$$

where $N_s = 3$ is the number of site clusters.

The optimal effect-modification configuration $\mathbf{E}^*$ is selected as the one that maximizes $\overline{\Delta\mathrm{AIC}}_E$. This criterion directly quantifies the average gain in model fit attributable to effect modification, while retaining robustness to sampling variability through bootstrap aggregation.

S3.2 Basis Dimension Selection and Nonlinear Model Estimation

In the two-stage GAM–SCAM framework, data from three site clusters were pooled to conduct basis-dimension diagnostics, under the working assumption that smooth components share comparable functional complexity across clusters.

In the *mgcv* and *scam* implementations, the parameter k specifies the upper bound on the number of basis functions used to represent a smooth term. For a given k, model parameters are estimated via penalized least squares under a Gaussian error assumption, solving

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \left\{ |\mathbf{Y} - \Phi\boldsymbol{\gamma}|^2 + \sum_{m=1}^{M} \lambda_m \boldsymbol{\gamma}^\top S_m \boldsymbol{\gamma} \right\}, \tag{S26}$$

where Φ is the spline basis matrix determined by k, S_m are quadratic penalty matrices controlling smoothness, and λ_m are smoothing parameters selected using generalized cross-validation (GCV)-based criteria. This penalization yields effective degrees of freedom (EDF) that are data-adaptive and typically smaller than k.

Smooth terms in the GAM component were represented using thin-plate regression splines (Wood, 2017), while the SCAM component employed shape-constrained P-splines (SCOP-splines) that impose linear inequality constraints on spline coefficients to enforce monotonicity (Pya and Wood, 2015).

Candidate basis dimensions for univariate smooth terms were evaluated over $k \in \{4, 5, 6, 7\}$. These ranges were chosen to provide sufficient flexibility for capturing nonlinear effects while avoiding overparameterization. Variables expected to exhibit relatively simple exposure–response relationships were assigned smaller candidate ranges ($k = 4$ to 5). This includes initial concentrations in the CO and $PM_{2.5\text{-}10}$ proxy models, as well as the precursor variables SO_2 and NO_2 in the $PM_{2.5}$ outcome model, whose effects are constrained to be non-decreasing. Smooth terms for the orthogonally rotated emission proxies in the $PM_{2.5}$ outcome model were evaluated over a moderately flexible range ($k = 4$ to 6). In contrast, meteorological variables, which may exhibit non-monotonic responses, threshold effects, and broader nonlinear variation, were evaluated over a wider range ($k = 5$ to 7). The episode duration variable L was assigned a fixed upper bound of $k = 4$, reflecting its expected low functional complexity.

For each candidate k, models were refitted while holding other smooth terms fixed, and the optimal basis dimension for each smooth was selected by minimizing the AIC. Cluster-specific models were then fitted using the selected basis dimensions, and EDF values were examined to assess adequacy. When the EDF approached the upper bound implied by k, the basis dimension was increased, subject to a maximum of $k = 7$.

S3.3 Model specifications for proxy, modifier and nonlinearity

This section provides the explicit formulas for the linear and nonlinear models compared in Table 1 of the main text. All models share the same set of covariates (episode-averaged meteorology $\overline{\mathbf{M}}_{ij}$, initial $PM_{2.5}$ $Y_{0,ij}$, episode duration L_{ij}, year fixed effects), but differ in whether the proxy indices $\mathbf{e}_{ij}$ are included and whether the precursor effects and their interactions are represented linearly or via smooth functions.

S3.3.1 Linear models

Let $\overline{\mathbf{M}}^*_{ij}$ denote the vector of episode-averaged meteorological covariates $\overline{\mathbf{M}}_{ij}$ excluding $\overline{\mathrm{TEMP}}_{ij}$.

Linear model without proxy: excludes the proxy indices $\mathbf{e}_{ij}$:

$$\begin{aligned} Y_{ij} = {} & \psi^*_0 + \psi^*_{\mathrm{SO_2}}\,\overline{\mathrm{SO}}_{2,ij} + \psi^*_{\mathrm{TEMP}}\,\overline{\mathrm{TEMP}}_{ij} + \psi^*_{\mathrm{SO_2:TEMP}}\left(\overline{\mathrm{SO}}_{2,ij} \times \overline{\mathrm{TEMP}}_{ij}\right) + \psi^*_{\mathrm{NO_2}}\,\overline{\mathrm{NO}}_{2,ij} + \psi^*_{O_3}\,\overline{O}_{3,ij} \\ & + \psi^*_{\mathrm{NO_2}:O_3}\left(\overline{\mathrm{NO}}_{2,ij} \times \overline{O}_{3,ij}\right) + \boldsymbol{\psi}^{*\top}_M \overline{\mathbf{M}}^*_{ij} + \psi^*_Y Y_{0,ij} + \psi^*_L L_{ij} + \sum_r \psi^*_r \mathbb{I}(\mathrm{year}_{ij} = r) + \nu_{ij}. \end{aligned} \tag{S27}$$

Linear full model: includes the proxy indices and linear precursor–modifier interactions:

$$\begin{aligned} Y_{ij} = {} & \psi^*_0 + \psi^*_{\mathrm{SO_2}}\,\overline{\mathrm{SO}}_{2,ij} + \psi^*_{\mathrm{TEMP}}\,\overline{\mathrm{TEMP}}_{ij} + \psi^*_{\mathrm{SO_2:TEMP}}\left(\overline{\mathrm{SO}}_{2,ij} \times \overline{\mathrm{TEMP}}_{ij}\right) + \psi^*_{\mathrm{NO_2}}\,\overline{\mathrm{NO}}_{2,ij} + \psi^*_{O_3}\,\overline{O}_{3,ij} \\ & + \psi^*_{\mathrm{NO_2}:O_3}\left(\overline{\mathrm{NO}}_{2,ij} \times \overline{O}_{3,ij}\right) + \boldsymbol{\psi}^{*\top}_M \overline{\mathbf{M}}^*_{ij} + \psi^*_Y Y_{0,ij} + \boldsymbol{\psi}^{*\top}_e \mathbf{e}_{ij} + \psi^*_L L_{ij} + \sum_r \psi^*_r \mathbb{I}(\mathrm{year}_{ij} = r) + \nu_{ij}. \end{aligned} \tag{S28}$$

S3.3.2 Nonlinear models

The nonlinear models are based on Eq. (8) in the main text, which is denoted as the full model in Table 1. They share the same additive structure but differ in which terms are included.

Nonlinear model without proxy: excludes the proxy indices $\mathbf{e}_{ij}$:

$$\mathbb{E}[Y_{ij} \mid \overline{\mathbf{A}}_{ij,\tau^*}, \mathbf{Z}_{ij}, \mathrm{year}_{ij}] = \beta_{\mathrm{year}_{ij}} + \phi_{\mathrm{SO_2}}(A_{\mathrm{SO_2},ij}, E^*_{\mathrm{SO_2},ij}) + \phi_{\mathrm{NO_2}}(A_{\mathrm{NO_2},ij}, E^*_{\mathrm{NO_2},ij}) + \sum_{V \in \mathbf{Z} \setminus \mathbf{E}^*} s_V(V_{ij}), \tag{S29}$$

where $\phi_{\mathrm{SO_2}}$ and $\phi_{\mathrm{NO_2}}$ are tensor-product smooths (non-decreasing in the precursor dimension), $E^*_{\mathrm{SO_2}} = \overline{\mathrm{TEMP}}$, $E^*_{\mathrm{NO_2}} = \overline{O}_3$, and $\mathbf{Z}_{ij} = (Y_{0,ij}, \overline{\mathbf{M}}_{ij}, \overline{O}_{3,ij}, L_{ij})$.

Nonlinear model without modifier: includes the proxy indices but replaces the bivariate tensor-product smooths with separate univariate shape-constrained smooths for the precursors and univariate smooths for the modifiers:

$$\mathbb{E}[Y_{ij} \mid \overline{\mathbf{A}}_{ij,\tau^*}, \mathbf{Z}_{ij}, \mathbf{e}_{ij}, \mathrm{year}_{ij}] = \beta_{\mathrm{year}_{ij}} + s_{\mathrm{SO_2}}(A_{\mathrm{SO_2},ij}) + s_{\mathrm{NO_2}}(A_{\mathrm{NO_2},ij}) + s_{e_1}(e_{1,ij}) + s_{e_2}(e_{2,ij}) + \sum_{V \in \mathbf{Z}} s_V(V_{ij}),$$

(S30)

where $s_{\mathrm{SO_2}}$, $s_{\mathrm{NO_2}}$, s_V, are univariate smooths, and the remaining terms are as in the previous specification.

S3.4 Standardization for estimating annual precursor contributions

To compare the contribution of precursor concentrations to $PM_{2.5}$ across years (2016–2023), we construct a standardized annual response. This response isolates year-to-year changes in the precursor distribution (SO_2 or NO_2) while eliminating two sources of confounding: differences in the distribution of episode duration L, and differences in the distribution of the effect modifier E^* (TEMP for SO_2, O_3 for NO_2). Specifically, the standardized response holds the duration and modifier distributions fixed at their pooled reference distributions (2016–2023) and allows only the precursor concentration distribution to vary with the year.

Let $p \in \{\mathrm{SO_2}, \mathrm{NO_2}\}$ index the target precursor, and let $\phi_p(A_p, E_p^*)$ denote the corresponding tensor-product smooth in Model (8). The estimation is performed within duration strata. Each distinct episode duration value (from 4 to 16 hours) defines a stratum L. Let $\mathcal{L}$ be the set of duration strata present in the pooled reference sample (2016–2023). For each stratum L, let n_L be the number of reference episodes, and define the reference weight $\hat{\omega}_L = n_L / \sum_{L \in \mathcal{L}} n_L$. For a given year r and precursor p, let $m_{p,r,L}$ be the number of episodes in that year falling into stratum L, and denote their observed precursor concentrations as a^p_{rLj} $(j = 1, \ldots, m_{p,r,L})$.

The standardized annual response is then computed as the weighted average over strata, where within each stratum we average the response ϕ_p over all combinations of the year-specific precursor concentrations and the reference modifier values:

$$\bar{\Phi}_{p,r} = \sum_{L \in \mathcal{L}} \hat{\omega}_L \left(\frac{1}{m_{p,r,L}\, n_L} \sum_{j=1}^{m_{p,r,L}} \sum_{i=1}^{n_L} \phi_p\left(a^p_{rLj}, E^{*,\mathrm{ref}}_{p,Li}\right) \right), \tag{S31}$$

where $E^{*,\mathrm{ref}}_{p,Li}$ is the modifier value of the ith reference episode in stratum L. This double average aligns each year-specific precursor distribution with the reference distribution of both the effect modifier and episode duration. For computational efficiency, an equivalent implementation draws for each a^p_{rLj} a random $E^{*,\mathrm{ref}}_{p,Li}$ from the same stratum, which yields the same expectation.

The change in the precursor-attributable $PM_{2.5}$ response relative to the 2016 baseline is then

$$\Delta\bar{\Phi}_{p,r} = \bar{\Phi}_{p,r} - \bar{\Phi}_{p,2016}, \qquad p \in \{\mathrm{SO_2}, \mathrm{NO_2}\}. \tag{S32}$$

Negative values indicate that the precursor distribution in year r leads to lower predicted $PM_{2.5}$ than the 2016 distribution.

S4 Tables and Figures

Figure S6 provides supporting information for the calm episode identification procedure described in Sect. 2.2. Table S1 reports the Pearson correlations between the proxy indices, residualized proxy variables, and observed pollutant concentrations, supporting the interpretation of the proxy-based emission adjustment in Sect. 3.1. Figures S7–S8 show the estimated univariate and bivariate smooth partial-effect components from the SCAM outcome models, respectively, complementing the model interpretation in Sect. 3.2. Figure S9 provides additional diagnostics for the causal analysis in Sect. 3.3.

Table S1. Pearson correlation coefficients between the proxy indices (e_1, e_2), residualized proxy variables, and observed pollutant concentrations across site clusters. All correlations are significant at 0.001.

Cluster	**Index**	**Residual CO**	**Residual $PM_{2.5\text{-}10}$**	**CO**	**$PM_{2.5\text{-}10}$**	**$PM_{2.5}$**
UrbanNW	e_1	0.79	0.77	0.61	0.56	0.49
	e_2	0.64	-0.61	0.53	-0.41	0.50
Suburban	e_1	0.77	0.77	0.65	0.59	0.51
	e_2	0.64	-0.64	0.53	-0.41	0.41
UrbanSE	e_1	0.78	0.73	0.56	0.53	0.41
	e_2	0.69	-0.62	0.62	-0.36	0.48

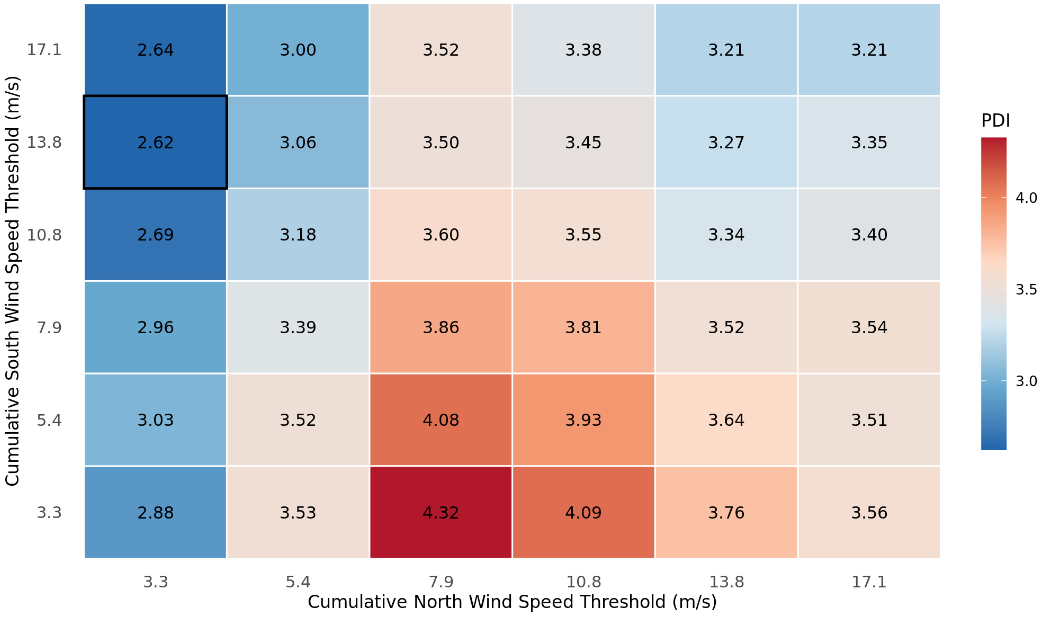

Figure S6. Pollution dispersion index (PDI) for calm episodes identified under different cumulative wind speed thresholds. The selected threshold pair corresponds to the minimum PDI value and is highlighted by the black rectangle.

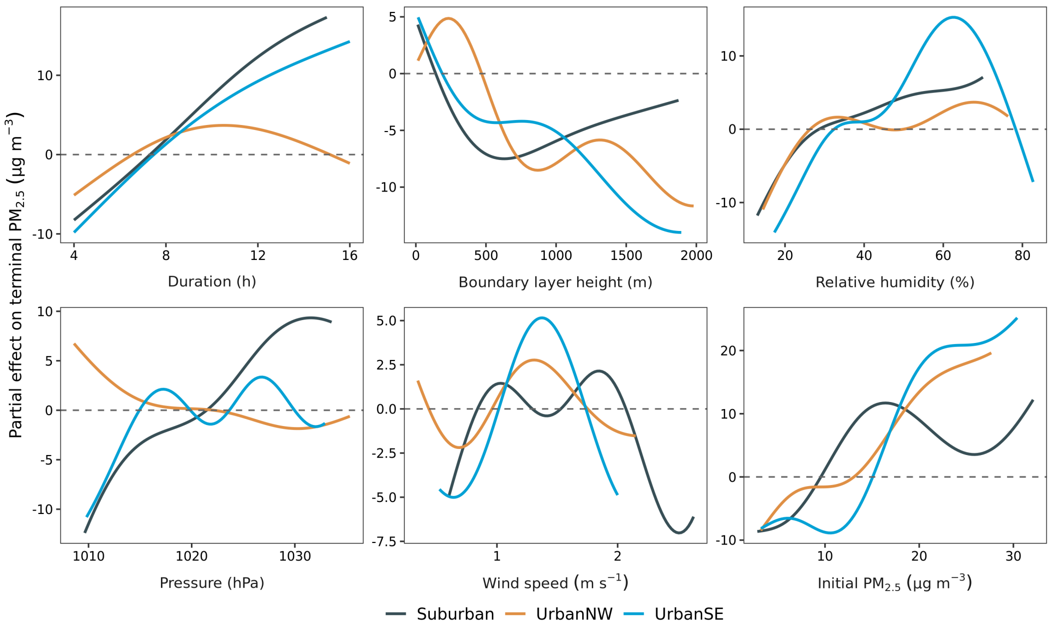

Figure S7. Estimated univariate smooth partial-effect curves from the outcome Model (8). The curves represent the estimated partial contribution of each univariate smooth term to terminal $PM_{2.5}$ in the additive predictor. The dashed horizontal line indicates zero partial effect.

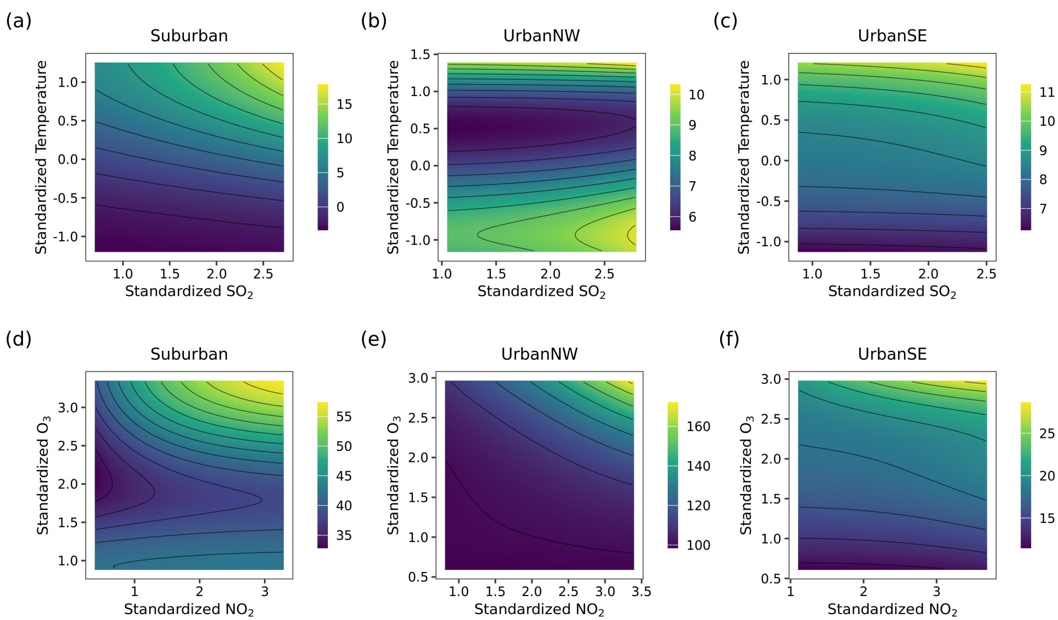

Figure S8. Estimated bivariate smooth partial-effect surfaces from the SCAM outcome model. Panels (a)–(c) show the smooth term for SO_2 and temperature, and panels (d)–(f) show the smooth term for NO_2 and O_3. Colour shading and contour lines represent the term-wise fitted values of the corresponding bivariate smooth in the additive predictor for terminal $PM_{2.5}$. All predictors are shown on the standardized scale.

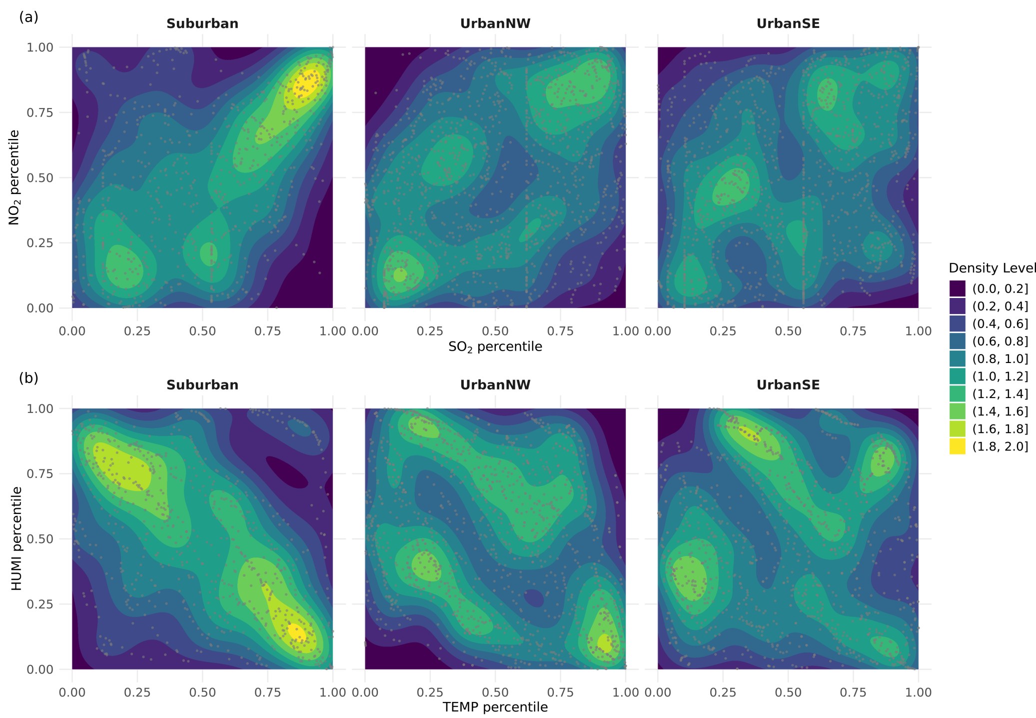

Figure S9. Joint percentile distributions of episode-level SO_2 and NO_2 (a) and temperature and relative humidity (b) during heating-season calm episodes from 2016 to 2023. Both variables were transformed to empirical percentile scales. The filled contours represent the estimated copula density, where values larger than unity indicate stronger joint concentration than expected under independence. Gray points denote individual episodes.

References

Brown, S. S. and Stutz, J.: Nighttime radical observations and chemistry, Chemical Society reviews, 41, 6405–6447, https://doi.org/10.1039/C2CS35181A, 2012.

Cheng, Y., Zheng, G., Wei, C., Mu, Q., Zheng, B., Wang, Z., et al.: Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China, Science Advances, 2, https://doi.org/10.1126/sciadv.1601530, 2016.

Claeskens, G., Krivobokova, T., and Opsomer, J. D.: Asymptotic properties of penalized spline estimators, Biometrika, 96, 529–544, https://doi.org/10.1093/biomet/asp035, 2009.

Dong, Z., Su, F., Zhang, Z., and Wang, S.: Observation of chemical components of $PM_{2.5}$ and secondary inorganic aerosol formation during haze and sandy haze days in Zhengzhou, China, Journal of Environmental Sciences, 88, 316–325, https://doi.org/10.1016/j.jes.2019.09.016, 2020.

Liu, Y., Feng, Z., Zheng, F., Bao, X., Liu, P., Ge, Y., et al.: Ammonium nitrate promotes sulfate formation through uptake kinetic regime, Atmospheric Chemistry and Physics, 21, 13 269–13 286, https://doi.org/10.5194/acp-21-13269-2021, 2021.

Ma, P., Quan, J., Dou, Y., Pan, Y., Liao, Z., Cheng, Z., et al.: Regime-Dependence of Nocturnal Nitrate Formation via N_2O_5 Hydrolysis and Its Implication for Mitigating Nitrate Pollution, Geophysical Research Letters, 50, e2023GL106 183, https://doi.org/10.1029/2023GL106183, 2023.

Ma, Q., Wu, Y., Zhang, D., Wang, X., Xia, Y., Liu, X., et al.: Roles of regional transport and heterogeneous reactions in the $PM_{2.5}$ increase during winter haze episodes in Beijing, The Science of the total environment, 599-600, 246–253, https://doi.org/10.1016/j.scitotenv.2017.04.193, 2017.

Mannschreck, K., Gilge, S., Plass-Duelmer, C., Fricke, W., and Berresheim, H.: Assessment of the applicability of NO-NO_2-O_3 photostationary state to long-term measurements at the Hohenpeissenberg GAW Station, Germany, Atmospheric Chemistry and Physics, 4, 1265–1277, https://doi.org/10.5194/acp-4-1265-2004, 2004.

Pya, N. and Wood, S. N.: Shape-constrained additive models, Statistics and Computing, 25, 543–559, https://doi.org/10.1007/s11222-013-9448-7, 2015.

Ren, C., Huang, X., Wang, Z., Sun, P., Chi, X., Ma, Y., et al.: Nonlinear response of nitrate to NOx reduction in China during the COVID-19 pandemic, Atmospheric Environment, 264, 118 715, https://doi.org/10.1016/j.atmosenv.2021.118715, 2021.

Wang, G., Zhang, R., Gomez, M. E., Yang, L., Zamora, M. L., Hu, M., et al.: Persistent sulfate formation from London Fog to Chinese haze, Proceedings of the National Academy of Sciences - PNAS, 113, 13 630–13 635, https://doi.org/10.1073/pnas.1616540113, 2016.

Wang, W., Liu, M., Wang, T., Song, Y., Zhou, L., Cao, J., et al.: Sulfate formation is dominated by manganese-catalyzed oxidation of SO_2 on aerosol surfaces during haze events, Nature communications, 12, 1993, https://doi.org/10.1038/s41467-021-22091-6, 2021.

Wood, S. N.: Generalized Additive Models: An Introduction with R, Chapman and Hall/CRC, New York, 2 edn., ISBN 9781315370279, https://doi.org/10.1201/9781315370279, 2017.

Yang, Y., Wang, Y., Zhou, P., Yao, D., Ji, D., Sun, J., Wang, Y., Zhao, S., Huang, W., Yang, S., Chen, D., Gao, W., Liu, Z., Hu, B., Zhang, R., Zeng, L., Ge, M., Petäjä, T., Kerminen, V.-M., Kulmala, M., and Wang, Y.: Atmospheric reactivity and oxidation capacity during summer at a suburban site between Beijing and Tianjin, Atmospheric chemistry and physics, 20, 8181–8200, https://doi.org/10.5194/acp-20-8181-2020, 2020.