

Revisiting the critical role of stabilized Criegee intermediates (sCIs) in sulfuric acid formation: coupling mechanistic updates with interpretable machine learning

Yuhuan Zhu¹, Qiang Chen^{1,2}, Luyan He¹, Chunlin Shang³, Li Jiang⁴, Donghong Guan⁴, Guirong Yao⁴, Wenkai Guo⁵

¹Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, College of Atmospheric Sciences, Lanzhou University, Lanzhou, 730000, China

²Lanzhou University Applied Technology Research Institute Co., Ltd, Lanzhou, 730000, China

³Inner Mongolia Autonomous Region Environmental Monitoring Center, Wuhai Branch, Wuhai, 016000, China

⁴Gansu Provincial Ecological and Environmental Engineering Assessment Center, Lanzhou 730000, China

⁵Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, 611756, China

Correspondence to: Qiang Chen(chenqqh@lzu.edu.cn)

Abstract. Sulfuric acid, formed via gas-phase SO₂ oxidation, drives new particle formation and secondary sulfate aerosol under low relative humidity (RH). Beyond the well-established OH-mediated pathway, stabilized Criegee intermediates (sCIs) from alkene ozonolysis provide an additional SO₂ oxidation route whose rate ($L_{\text{SO}_2, \text{sCI}s}$) and fractional contribution ($\mu_{\text{sCI}s}$) vary markedly across environments. Although prior studies have documented this spatiotemporal heterogeneity, quantifying the individual factor contributions governing these differences remains lacking. Furthermore, because sCI and OH formation share common precursors yet differ mechanistically, $\mu_{\text{sCI}s}$ can alter the sensitivity of total SO₂ oxidation (L_{SO_2}) to precursor variations, but this influence's magnitude remains unclear. Here, we updated Criegee-related reactions in MCM v3.3.1 following recent IUPAC recommendations, conducted extensive AtChem box-model simulations, and trained interpretable XGBoost surrogate models to attribute $L_{\text{SO}_2, \text{sCI}s}$ and $\mu_{\text{sCI}s}$ to their controlling factors and examine how $\mu_{\text{sCI}s}$ regimes modulate L_{SO_2} sensitivity. $L_{\text{SO}_2, \text{sCI}s}$ was primarily controlled by RH and isoprene. The dominant drivers of $\mu_{\text{sCI}s}$ differed between day and night: daytime factors included alkene speciation, O₃, and the alkene fraction in VOCs; nighttime factors included alkene speciation, RH, and NO_x. Under high- $\mu_{\text{sCI}s}$ regimes, positive L_{SO_2} responses to O₃, VOCs, and alkene fraction were consistently amplified relative to low- $\mu_{\text{sCI}s}$ regimes. Field observation-based modeling of a summer episode in Wuhai supported these regime-dependent patterns. These findings indicate that reducing VOC emissions provides a more effective mechanism for regulating sulfuric acid formation under conditions with active alkene ozonolysis.

1 Introduction

The environmental and climatic impacts of fine particulate matter (PM_{2.5}) are strongly influenced by sulfate aerosol (SO₄²⁻), which plays a key role in regulating aerosol acidity (Yao et al., 2018), hygroscopicity (Gunthe et al., 2011), and direct

radiative forcing (Quaas et al., 2022). Although a minor fraction of particulate sulfate is directly emitted from anthropogenic sources such as coal and biomass combustion (Dai et al., 2019), industrial activities (Song et al., 2025), and transportation (Cao et al., 2025), it is formed predominantly from the atmospheric oxidation of sulfur dioxide (SO_2) via gas-phase, aqueous-phase, and heterogeneous pathways (Ye et al., 2023a). During severe haze events, elevated relative humidity and aerosol loadings significantly enhance aqueous-phase and heterogeneous SO_2 oxidation. These multiphase processes dominate sulfate formation and have therefore been the focus of extensive mechanistic investigations (Cao et al., 2021). Nevertheless, the gas-phase oxidation of SO_2 remains critically important because its product, sulfuric acid (H_2SO_4), is a key precursor for atmospheric new particle formation (NPF). Atmospheric nucleation and the subsequent growth of newly formed particles represent a major source of aerosol particles in terms of number concentration, and NPF has been frequently observed even if in highly polluted urban environments characterized by strong condensation sinks (Yao et al., 2018). Moreover, over the past decade, stringent emission-control policies in East Asia, particularly in China, have substantially reduced $\text{PM}_{2.5}$ concentrations (Geng et al., 2021), thereby weakening conditions favorable for aqueous-phase and heterogeneous sulfate formation. In this progressively cleaner atmosphere, the relative importance of gas-phase pathways is expected to increase (Wang et al., 2025). Therefore, under increasingly stringent air pollution control targets (WHO, 2021), the role of SO_2 gas-phase oxidation pathways in secondary aerosol production is expected to become more important and warrants further investigation.

Hydroxyl radicals (OH) have long been recognized as the dominant oxidant in gas-phase SO_2 oxidation. Since the pioneering work of Cox and Penkett (1971) first demonstrated that Criegee intermediates (CIs, zwitterionic carbonyl oxides) react with SO_2 in the gas phase, extensive laboratory and theoretical studies (Taatjes et al., 2013; Vereecken et al., 2012) have advanced understanding of the formation, stabilization, and fate of CIs. Stabilized Criegee intermediates (sCIs) are therefore increasingly considered an important complementary pathway for the gas-phase SO_2 oxidation and a potentially significant contributor to H_2SO_4 formation. Multiple field observation studies conducted in the boreal forests in Finland (Mauldin et al., 2012), the SMEAR II and Hohenpeissenberg stations in Germany (Boy et al., 2013), in urban Beijing (Guo et al., 2021), and at a remote Mediterranean site strongly influenced by biogenic VOC emissions (Kukui et al., 2021a) suggest that sCIs can provide an additional, environment-dependent pathway for SO_2 oxidation. Notably, as the direct measurement of atmospheric sCIs concentrations remains a formidable challenge, observational studies have primarily relied on the analysis of precursors and OH radical concentrations to provide qualitative evidence. Consequently, numerical simulations grounded in CIs chemical kinetics are indispensable for quantifying the contribution of sCIs to the gas-phase SO_2 oxidation rate.

To date, numerical modeling studies have extensively evaluated the contribution of sCIs to SO_2 oxidation across a broad range of spatial scales and environmental conditions. For instance, utilizing a steady-state approximation, Khan et al. (2017) estimated that the OH-driven SO_2 oxidation rate in urban London (0.64 Tg/year) was approximately 17 times that of the sCI pathway (37.6 Gg/year). However, proxy calculations of H_2SO_4 concentrations have revealed pronounced regional disparities: at the semi-pristine boreal forest site in Hyytiälä, the photochemical pathway ($\text{SO}_2 + \text{OH}$) and the alkene ozonolysis pathway ($\text{SO}_2 + \text{sCIs}$) accounted for approximately 34% and 60% of H_2SO_4 production, respectively, whereas in

65 heavily polluted Beijing, the corresponding proportions shifted to 28% and 22% (Dada et al., 2020). Vereecken et al. (2017)
66 predicted that in equatorial regions, the contribution of sCIs to the gas-phase conversion of SO₂ to sulfuric acid could reach
67 up to 75%. Furthermore, simulations using the STOCHEM-CRI global atmospheric chemistry and transport model
68 demonstrated (Khan et al., 2018) that the SO₂ oxidation contribution from sCIs significantly exceeds that of OH radicals in
69 terrestrial rainforests and high-latitude boreal forests. This pronounced variability is primarily driven by two intertwined
70 factors. First, considerable uncertainty in numerical simulations arises from inherent limitations in our understanding of CIs
71 chemical kinetics. With continuous breakthroughs in monitoring technologies, the scientific understanding of CI evolution
72 processes is dynamically evolving, resulting in discrepancies in the reaction rate coefficients incorporated into models across
73 different periods. Notably, the unimolecular decomposition rates of CIs/sCIs and the bimolecular reaction rates of sCIs with
74 water monomers and dimers (H₂O/(H₂O)₂) exert a decisive influence on numerical calculations (Sarwar et al., 2014). For
75 example, a modeling study under summer conditions in the eastern United States demonstrated that once the competitive
76 reaction between sCIs and H₂O/(H₂O)₂ was introduced, the sCI-driven enhancement in surface sulfate plummeted from 18%
77 to less than 0.5% (Li et al., 2013). Similarly, in simulations of nocturnal power plant plumes in the southeastern US, whether
78 CI thermal decomposition was considered led to a seven-fold difference in the estimated contribution to secondary sulfate
79 aerosol (SSA) formation (Meidan et al., 2019). Second, pronounced regional heterogeneity in precursor composition, such as
80 biogenic isoprene and monoterpenes in forested regions and complex anthropogenic alkene mixtures in urban areas, together
81 with differences in relative humidity and NO_x levels, leads to highly variable sCI production and scavenging fluxes. While
82 existing studies have made progress in quantifying the contribution of sCIs to gas-phase SO₂ oxidation in specific
83 environments and identifying their precursors, the relationship between sCIs and their precursors remains unclear.
84 Consequently, a systematic understanding of the critical driving factors dictating the relative contribution of sCIs is still
85 lacking. Moreover, the production and loss processes of OH and sCIs are not mutually independent; rather, they are deeply
86 coupled within the underlying chemical mechanisms (Zhu et al., 2023). This coupled nature dictates that when evaluating the
87 drivers controlling the contribution of the sCIs+SO₂ pathway, their concurrent impacts on the OH+SO₂ pathway and the
88 overall SO₂ oxidation rate cannot be overlooked. Fundamentally, the interplay between precursor abundances and
89 competitive sinks determines the fractional contribution of the sCIs pathway. Yet, how this interplay shapes the sensitivity of
90 the total gas-phase SO₂ oxidation to environmental drivers has received remarkably little attention. In particular, it remains
91 elusive how the dominant controlling factors and their directional impacts shift as the atmospheric gas-phase SO₂ oxidation
92 regime transitions from being strictly OH-dominated to being co-driven by both OH and sCIs.

93 To this end, this study presents a systematic assessment of the role of sCIs in atmospheric SO₂ gas-phase oxidation. As a
94 prerequisite for all subsequent analyses, we first updated the gas-phase kinetics of Criegee intermediates in the Master
95 Chemical Mechanism (MCM v3.3.1, via website: www.mcm.york.ac.uk) using the latest evaluated rate coefficients
96 recommended by the International Union of Pure and Applied Chemistry (IUPAC), thereby minimizing the propagation of
97 mechanistic uncertainties into our conclusions. Utilizing an atmospheric box model coupled with this revised mechanism, we
98 integrated interpretable machine learning techniques to quantify the importance of key controlling factors on sCI-mediated

99 SO₂ oxidation. Building upon this insight, we employed a robust experimental design to construct a machine learning
100 surrogate model capable of predicting sCI contributions across diverse and highly variable environmental conditions.
101 Furthermore, analytical methods, including SHAP, Sobol sensitivity analysis, Partial Dependence Plots (PDPs), and
102 individual conditional expectation (ICE) plots, were utilized to quantify how these factors dictate the relative contribution of
103 sCIs, and to elucidate the differential sensitivities of total SO₂ oxidation under high- and low-sCI contribution regimes.
104 Finally, leveraging online hourly observational data of atmospheric pollutants, we assessed the disparities in their sensitivity
105 relationships with SO₄²⁻ across varying sCI contribution levels, ultimately validating the reliability of our box model-derived
106 surrogate model.

107 **2 Data and methodology**

108 **2.1 Updates to alkene ozonolysis and Criegee intermediate chemistry in MCM v3.3.1**

109 The Master Chemical Mechanism (MCM v3.3.1; available via <http://mcm.york.ac.uk>) is a near-explicit chemical mechanism
110 that describes the gas-phase oxidation of 143 volatile organic compounds (VOCs) to carbon dioxide (CO₂) and water (H₂O).
111 In MCM v3.3.1, the rate coefficients and associated reaction mechanisms for alkene ozonolysis reactions follow the
112 recommendations of Jenkin et al. (1997, 2015) and Saunders et al. (2003). Although this mechanism provides a detailed
113 representation of CI formation, loss, and subsequent chemical evolution, several key rate coefficients and parameters differ
114 substantially from the latest IUPAC recommendations (Cox et al., 2020). In addition, several important processes are not
115 explicitly represented, including the specific fates of CIs, the unimolecular decomposition of sCIs, the reactions of sCIs with
116 water dimers ((H₂O)₂), and the distinct reactivities of sCI stereoisomers. To address these limitations, we updated the gas-
117 phase alkene + O₃ chemistry in MCM v3.3.1 primarily in accordance with current IUPAC recommendations. The updates
118 include the alkene ozonolysis rate coefficients ($k_{O_3+\text{alkene}}$), OH yields ($Y_{i,\text{OH}}$), sCI yields ($Y_{i,\text{sCI}}$), and sCI bimolecular reaction
119 rate coefficients (k_b), as well as accounting for sCI unimolecular decomposition pathways and adding sCI + (H₂O)₂ reactions.
120 Where IUPAC provides temperature-dependent kinetic expressions, these were adopted; in addition, a temperature-
121 dependent rate expression was incorporated for the CH₂OO + SO₂ reaction, based on (Onel et al., 2021). Specifically, the
122 branching ratios for the unimolecular decomposition or isomerization pathways of CIs and sCIs were constrained using
123 reported OH yields and the yields of other relevant products. For sCIs exhibiting stereoisomerism, including CH₃CHOO,
124 C₂H₅CHOO, and the C₄ intermediates formed during isoprene ozonolysis, the Z-stereoisomers were assumed to undergo
125 predominantly rapid unimolecular decomposition under tropospheric conditions, such that their contribution to bimolecular
126 reactions was considered negligible (Newland et al., 2015). Accordingly, in the updated mechanism, hereafter denoted MCM
127 v3.3.1g, stereoisomeric sCIs participating in bimolecular reactions are represented explicitly by their E-stereoisomers. The
128 unimolecular decomposition pathways of sCIs were considered together with the decomposition or isomerization channels of
129 their corresponding chemically activated CIs. The kinetic parameters and product yields for the alkene ozonolysis reactions
130 before and after the updates are summarized in Tables S1 and S2. The complete set of updated reaction equations, together

with the corresponding rate coefficients and explicit branching ratios, is provided in Table S3. These updates were applied to C₁–C₄ Criegee intermediates, which are widely recognized to play a critical role in tropospheric chemistry. The concentration of (H₂O)₂ is determined by assuming rapid thermal equilibrium between water monomers and dimers. Given this fast atmospheric exchange, the dimer concentration is considered to be in thermodynamic equilibrium with the monomer, as expressed in Eq. (1):

$$K_{\text{eq}}^{\text{D}} = \frac{[(\text{H}_2\text{O})_2]}{[\text{H}_2\text{O}]^2}, \quad (1)$$

where K_{eq}^{D} is the temperature-dependent equilibrium constant for dimer formation (Scribano et al., 2006). Following the methodology of Lade et al. (2024), the reaction of sCIs with water dimers is parameterized using an effective third-order rate coefficient, $k_{\text{eff}}(\text{cm}^6 \text{ molecule}^{-2} \text{ s}^{-1})$, and the rate of sCI loss via reaction with water dimers, $r_{\text{sCI}+(\text{H}_2\text{O})_2}$, is given by Eq. (2):

$$r_{\text{sCI}+(\text{H}_2\text{O})_2} = k_{\text{b,dimer}}[\text{sCI}][(\text{H}_2\text{O})_2] = k_{\text{b,dimer}}K_{\text{eq}}^{\text{D}}[\text{sCI}][\text{H}_2\text{O}]^2 = k_{\text{eff}}[\text{sCI}][\text{H}_2\text{O}]^2, \quad (2)$$

where k_{eff} is defined as the product of the bimolecular rate coefficient for the sCI + (H₂O)₂ reaction ($k_{\text{b,dimer}}$, cm³ molecule⁻¹ s⁻¹) and K_{eq}^{D} . To calculate K_{eq}^{D} accurately, thermochemical data were retrieved from the Active Thermochemical Tables (ATcT). Standard Gibbs free energies of formation ($\Delta_f G_T^\circ$) for both the water monomer and dimer were extracted from Tables 1 and 3 in Ruscic (2013). These discrete data points were then used to calculate the reaction Gibbs free energy ($\Delta_r G_T^\circ$) over the temperature range of 200–360 K, which spans typical tropospheric conditions. The temperature dependence of K_{eq}^{D} was subsequently obtained via a linear regression of $\ln(K_{\text{eq}}^{\text{D}})$ as a function of $\frac{1}{T}$, yielding the following empirical van’t Hoff–type parameterization in Eq. (3):

$$K_{\text{eq}}^{\text{D}}(T) = A \exp\left(\frac{B}{T}\right), \quad (3)$$

where the coefficients A and B represent the intercept and slope, respectively, derived from the fitting procedure, with $A = 1.15 \times 10^{-23} \text{ cm}^3 \text{ molecule}^{-1}$ and $B = 1549.32 \text{ K}$.

2.2 Box model configuration and observation data

2.2.1 AtChem model setup and chemical diagnostics

AtChem (Sommariva et al., 2020), an open-source zero-dimensional atmospheric box model, was used in this study to simulate the gas-phase chemical evolution of SO₂, CIs, and their precursors. The AtChem model is coupled with MCM v3.3.1 and MCM v3.3.1g (see Sect. 2.1). H₂O concentrations for sCI loss reactions were calculated from relative humidity, temperature, and atmospheric pressure.

Two types of AtChem simulations were performed in this study. The first used an unconstrained configuration, in which the initial concentrations of selected chemical species were prescribed and allowed to evolve according to the chemical

159 mechanism without time-varying constraints. This configuration was used in Sect. 3.1 to quantify the oxidation capacity of
 160 sCIs derived from the ozonolysis of specific alkenes, including ethene, propene, but-1-ene, *trans*-but-2-ene, *cis*-but-2-ene, 2-
 161 methylpropene, and isoprene. The second used a constrained configuration, in which selected trace gases, meteorological
 162 parameters, and photolysis frequencies were prescribed as hourly time-varying inputs from observational data or designed
 163 perturbation scenarios. This configuration was used in Sects. 3.2 and 3.3 to evaluate the role of sCIs in gas-phase SO₂
 164 oxidation under atmospherically relevant conditions. A spin-up period of 2–3 days was implemented to initialize reactive
 165 free radicals and intermediates, such as OH, HO₂, RO₂, and sCIs, allowing them to reach realistic steady-state concentrations.
 166 Chemical production and loss rates were quantified using the Rate of Production/Destruction Analysis (ROPA/RODA)
 167 within AtChem. The gas-phase SO₂ oxidation rate by sCIs was defined according to Eq. (4):

$$168 \quad L_{\text{SO}_2, \text{sCIs}} = \sum_i k_{\text{SO}_2 + \text{sCI}} [\text{sCI}_i] [\text{SO}_2] , \quad (4)$$

169 where i denotes an individual sCI species. The OH-driven SO₂ oxidation rate was defined as in Eq. (5):

$$170 \quad L_{\text{SO}_2, \text{OH}} = k_{\text{SO}_2 + \text{OH}} [\text{OH}] [\text{SO}_2] , \quad (5)$$

171 The total gas-phase SO₂ oxidation rate and the fractional sCI contribution were then calculated as in Eqs. (6) and (7):

$$172 \quad L_{\text{SO}_2} = L_{\text{SO}_2, \text{sCIs}} + L_{\text{SO}_2, \text{OH}} , \quad (6)$$

$$173 \quad \mu_{\text{sCIs}} = \frac{L_{\text{SO}_2, \text{sCIs}}}{L_{\text{SO}_2}} , \quad (7)$$

174 Here, $L_{\text{SO}_2, \text{sCIs}}$ and μ_{sCIs} provide complementary perspectives on the absolute rate and fractional contribution, respectively,
 175 of the sCI pathway.

176 2.2.2 Observation data

177 The observational data used in this study were obtained from the automatic monitoring system at the Wuhai City
 178 Atmospheric Environment Super Monitoring Station from 9 October 2019 to 30 June 2022. Wuhai is a semi-arid coal-
 179 chemical industrial city in Inner Mongolia, China, characterized by relatively high emissions of SO₂ and anthropogenic
 180 VOCs (including alkenes), frequent O₃ pollution, and comparatively dry atmospheric conditions. The dataset includes air
 181 pollutants (PM_{2.5}, carbon monoxide (CO), SO₂, nitric oxid (NO), nitrogen dioxide (NO₂), and ozone (O₃)), meteorological
 182 variables (wind speed (WS), temperature (T), pressure (P), and relative humidity (RH)), water-soluble inorganic ions
 183 (including NO₃⁻ and SO₄²⁻), elemental components (including Fe), VOCs, and photolysis frequencies. PM_{2.5}, NO, NO₂, SO₂,
 184 CO, and O₃ were measured using a Metone BAM-1020 PM_{2.5} Monitor, an API T201 NH₃–NO_x Analyzer, an API T100 SO₂
 185 Analyzer, an API T300 CO Analyzer, and an API T400 O₃ Analyzer, respectively. VOCs were measured using an ENTECH
 186 BCT-7800 VOCs Analyzer. Water-soluble inorganic ions were measured using a Metrohm MARGA ADI 2080 online ion
 187 chromatograph, and elemental components were measured using a CES Xact-625 atmospheric heavy metal analyzer.

188 Meteorological variables (T, RH, P, WS) were measured using a Lufft WS601-UMB six-parameter meteorological sensor.
189 Photolysis frequencies were measured using a Metcon UF-CCD photolysis spectroradiometer. The observational data had an
190 hourly time resolution and were screened for invalid values. Short missing intervals were linearly interpolated when
191 appropriate, whereas longer gaps were excluded from the analysis.

192 **2.3 Machine learning model and interpretation framework**

193 An interpretable machine learning (ML) framework was developed in this study to systematically evaluate the role of sCIs in
194 SO₂ oxidation, encompassing model development, validation, and interpretation. The learning algorithm used throughout
195 was Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), which constructs an additive ensemble of regression
196 trees by iteratively minimizing a regularized loss function, with each successive tree fitted to the negative gradients (i.e.,
197 pseudo-residuals) of the current ensemble's predictions. This formulation enables nonlinear responses and predictor
198 interactions to be represented far more flexibly than a single regression tree. Regularization, subsampling, cross-validation,
199 and early stopping were applied to suppress overfitting, and hyperparameters were tuned via random search or Bayesian
200 optimization. The same sampling–training–testing workflow was applied to every model in this study. Two classes of
201 XGBoost models were constructed (using the XGBoost Python package, version 2.1.3; <https://github.com/dmlc/xgboost>):
202 surrogate models trained on AtChem-simulated chemical evolution, and an observation-based model trained on ambient
203 measurements. Four complementary interpretation methods were used to explain the XGBoost model outputs: SHAP
204 (SHapley Additive exPlanations), Sobol sensitivity analysis, partial dependence plots (PDPs), and individual conditional
205 expectation (ICE) plots.

206 **2.3.1 Model Development Details**

207 Three surrogate models were developed based on SO₂ gas-phase oxidation simulations from AtChem to successively
208 quantify the key controls on sCI-mediated SO₂ oxidation, identify the atmospheric conditions under which the sCI pathway
209 becomes important, and examine how this importance affects the sensitivity of total SO₂ oxidation. An observation-based
210 model was developed to assess whether the sensitivities inferred from the surrogate models are also reflected in ambient
211 measurements. These XGBoost models are introduced in turn below.

212 First, a surrogate model was developed to emulate the sCI-mediated SO₂ oxidation rate ($L_{\text{SO}_2, \text{sCIs}}$), with SHAP analysis
213 applied to diagnose the contributions of its key controlling factors, namely the principal sCI precursors and competing sinks
214 (most notably NO₂, H₂O, and (H₂O)₂). The feature set therefore comprises O₃, RH, NO₂, and six alkene precursors: ethene,
215 propene/but-1-ene, *trans*-but-2-ene, *cis*-but-2-ene, 2-methylpropene, and isoprene, where propene and but-1-ene are merged
216 into a single feature owing to the close similarity of their ozonolysis mechanisms. To capture individual perturbations and
217 interactions across all features, 3898 simulation scenarios were designed spanning the prescribed feature ranges. In each
218 scenario, the feature values prescribed the initial input concentrations for a 1-h unconstrained AtChem simulation, with SO₂

219 held constant; the mean $L_{\text{SO}_2, \text{sCI}}$ over the simulated hour, obtained under both MCM v3.3.1 and v3.3.1g, served as the target
 220 output. The resulting feature matrix and corresponding simulation outputs together constitute the training dataset for the
 221 surrogate model. The prescribed feature ranges, full scenario design, training details, and model performance are provided in
 222 Text S1. Note that in all simulation scenarios, temperature and pressure were held constant at $T = 290 \text{ K}$ and $P = 888 \text{ hPa}$.
 223 Temperature-dependent results for the full scenario set are provided in Text S4.

224 Second, to identify the atmospheric conditions that favour sCI contribution to total gas-phase SO_2 oxidation, a surrogate
 225 model was developed to predict the fractional importance of the sCI pathway (μ_{sCI}), combined with Sobol sensitivity
 226 analysis, PDPs, and ICE plots. Unlike $L_{\text{SO}_2, \text{sCI}}$, which depends solely on sCI production and loss, μ_{sCI} is also governed by
 227 the parallel OH-mediated oxidation pathway and therefore requires a feature set that captures both sCI and OH chemistry.
 228 Accordingly, the feature set comprises O_3 , NO_x , NO_2 fraction ($\text{NO}_2\% = \text{NO}_2/\text{NO}_x$), total VOC concentration (VOCs), alkene
 229 fraction ($\text{alkene}\% = \sum \text{alkenes}/\text{VOC}$), RH, and the Aggregated Reactivity Index (ARI). The NO_x – $\text{NO}_2\%$ and VOC–alkene%
 230 pairs jointly characterize the oxidant state and reactive VOC composition while minimizing multicollinearity among features.
 231 ARI is an ordinal variable encoding the sCI-oxidation potential of the prevailing alkene mixture, derived directly from the
 232 importance ranking of alkene features established by the $L_{\text{SO}_2, \text{sCI}}$ surrogate model. The training samples generated for this
 233 surrogate differ from those of the $L_{\text{SO}_2, \text{sCI}}$ model. Rather than prescribing initial concentrations, the feature values here
 234 represent normalized scaling factors applied to time-varying constraints, as the AtChem model was run in a constrained
 235 configuration. The baseline constraints are the three-year mean diurnal cycles of hourly observations at the Wuhai supersite
 236 (Sect. 2.1), and the perturbation range of each feature spans the 5th–95th percentile range of the observed concentrations
 237 (normalized range is $[0,1]$); ARI, as an ordinal discrete variable, is the sole exception and takes integer values of 0, 1, or 2,
 238 corresponding to the three tiers of sCI-oxidation potential. Because photolysis frequencies play a decisive role in OH
 239 production and radical cycling, daytime and nighttime conditions were treated as entirely separate modeling problems, with
 240 target time periods of 08:00–18:00 LT and 22:00–05:00 LT, respectively. For each scenario, the constraints of the target
 241 period were adjusted according to the prescribed feature values, while all remaining constraints (non-feature variables) were
 242 held at baseline (Figure S2); a constrained AtChem simulation under MCM v3.3.1g then yielded the mean μ_{sCI} and L_{SO_2}
 243 over the target period as the target outputs. The resulting feature matrix and simulated μ_{sCI} values together constitute the
 244 training datasets for the daytime and nighttime μ_{sCI} surrogate models. These daytime and nighttime datasets were then each
 245 divided into low- and high- μ_{sCI} subsets based on the median μ_{sCI} value, and a further pair of surrogate models was
 246 developed to predict L_{SO_2} separately in each regime, combined with partial dependence analysis to assess whether the
 247 sensitivity of L_{SO_2} to each feature depends on the relative importance of the sCI pathway. The baseline constraints, the
 248 mapping from normalized feature values to constraint magnitudes, full scenario design, training details, and model
 249 performance are provided in Text S2.

250 Finally, to assess whether the regime-dependent sensitivities inferred from the surrogate models are also expressed in the real
 251 atmosphere, we developed an observation-based model trained on ambient observations from the Wuhai supersite (9 October

252 2019 to 30 June 2022, hourly resolution) to predict SO_4^{2-} . The features, including $\text{PM}_{2.5}$, NO_3^- , SO_2 , O_3 , total VOCs, alkene%,
253 NO_x , $\text{NO}_2\%$, RH, wind speed, and Fe, were selected to represent sulfate aerosol precursor abundance, photochemical
254 oxidation capacity, aerosol loading, secondary inorganic aerosol formation, meteorological conditions, and gas-phase and
255 aqueous-phase oxidation pathways. SHAP analysis was applied to evaluate the differential contributions of key features to
256 SO_4^{2-} under contrasting μ_{sClS} regimes, enabling comparison with the regime-dependent sensitivities diagnosed from the
257 AtChem-based surrogate models. Training details and model performance are provided in Text S3.

258 2.3.2 Model Interpretation Methods

259 In this study, XGBoost was not used merely as a black-box surrogate for predicting chemical reaction rates or target
260 pollutant concentrations. Instead, it served as a computationally efficient emulator of the box-model simulations, enabling
261 systematic sensitivity analysis and interpretation of the chemical response within the prescribed parameter space. To
262 interpret the trained XGBoost models, we combined SHAP analysis, Sobol sensitivity analysis, partial dependence plots, and
263 individual conditional expectation plots. These four approaches were selected to provide complementary perspectives
264 spanning global versus local attribution, main effects versus interaction effects, and average versus sample-level response
265 behaviour. Specifically, SHAP values were used to quantify the relative importance and direction of predictor effects at both
266 individual-sample and aggregate levels, while SHAP interaction values were used to diagnose key two-factor interactions.
267 Sobol sensitivity analysis was used to provide a variance-based decomposition of the model output, separating the
268 independent contribution of each input factor from interaction-driven contributions. PDPs were used to visualize the average
269 nonlinear response of the model output to selected predictors and to identify potential thresholds, whereas ICE plots were
270 used to examine whether these average responses were consistent across individual samples or instead varied among
271 different chemical regimes, revealing heterogeneity that PDPs alone cannot capture. For models trained on controlled
272 numerical experiments, the feature values were prescribed through independent, design-of-experiments-based sampling, and
273 the target outputs were generated by a deterministic, mechanism-based chemical model (AtChem). Because the training data
274 are thus not subject to uncontrolled confounding, and because XGBoost was trained to approximate the fixed relationships
275 embedded in the chemical mechanism, the resulting SHAP, Sobol, PDP, and ICE analyses were interpreted as model-
276 estimated chemical sensitivities that reasonably reflect the causal structure inherent in chemical mechanism, within the
277 sampled parameter space. For the model trained on ambient observations, by contrast, the input features are inherently
278 correlated, and the observed feature–outcome relationships may be subject to unknown confounding. Consequently, SHAP-
279 based interpretations for this model were treated as statistical associations used to evaluate consistency with the chemical
280 sensitivities inferred from the AtChem-based surrogate models, rather than as evidence of direct causal effects.

281 (1) SHAP (SHapley Additive exPlanations) is a game-theoretic approach for interpreting machine learning model
282 predictions (Lundberg and Lee, 2017; Lundberg et al., 2018). Each input feature is treated as a “player” in a cooperative
283 game, and the model prediction for any individual sample is expressed as an additive decomposition of per-feature
284 contributions. Specifically, for a sample x_i with N features, the predicted value $f(x_i)$ is decomposed as Eq. (8):

$$f(x_i) = \phi_0 + \sum_{j=1}^N \phi_j(f, x_i), \quad (8)$$

where $\phi_0 = E[f(x)]$ is the base value, i.e., the expected model output averaged over the entire dataset, and $\phi_j(f, x_i)$ is the SHAP value of feature j for sample x_i , quantifying the contribution of feature j to the deviation of the prediction from the base value ϕ_0 . The SHAP value $\phi_j(f, x_i)$ is computed as the weighted average marginal contribution of feature j across all possible subsets of the remaining features, as shown in Eq. (9):

$$\phi_j(f, x_i) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{j\}) - f(S)], \quad (9)$$

where F is the full set of features, S denotes a subset of features that excludes feature j , and $f(S)$ denotes the conditional expectation of the model output given the features in S . The sign of $\phi_j(f, x_i)$ indicates the direction of the feature effect for a specific sample (positive: the feature j increases the predicted value relative to the baseline; negative: it decreases it), and the magnitude indicates the strength of that effect. For XGBoost models, SHAP values are computed exactly and efficiently using the TreeExplainer algorithm (Lundberg et al., 2020).

The global importance of feature j is summarized by its mean absolute SHAP value (Eq. (10)):

$$|\overline{\phi_j}| = \frac{1}{n} \sum_{i=1}^n |\phi_j(f, x_i)|, \quad (10)$$

where n is the total number of samples. $|\overline{\phi_j}|$ represents the average magnitude of feature j 's contribution to model predictions across the dataset and is used to rank features by their global importance.

Pairwise feature interactions are further quantified via SHAP interaction values (Lundberg et al., 2019). The mean absolute interaction value between features j and p is defined as Eq. (11):

$$|\overline{\phi_{jp}}| = \frac{1}{n} \sum_{i=1}^n |\phi_{jp}(f, x_i)|, \quad (11)$$

Where $\phi_{jp}(f, x_i)$ is the SHAP interaction value between features j and p for sample x_i , quantifying the portion of their joint effect on the prediction.

- (2) A variance-based global sensitivity analysis was performed using the Sobol method (Saltelli et al., 2010; Sobol, 2001). Sobol analysis decomposes the variance of the model output into contributions from individual input factors and their interactions over a prescribed input space and probability distribution. It thereby provides a global measure of sensitivity that is not determined by the empirical distribution of the training data.

For a model output $Y = f(x)$, where $x_1, x_2, \dots, x_N$ denote N mutually independent input factors, the total output variance, $\text{Var}(Y)$, can be decomposed into contributions from individual factors and their interactions (Eq. (12)):

$$\text{Var}(Y) = \sum_{j=1}^N V_j + \sum_{j < k} V_{jk} + \dots, \quad (12)$$

where $V_j = \text{Var}_{x_j} [E_{x_{\sim j}}(Y|x_j)]$ is the variance attributable to the main effect of x_j , whereas V_{jk} represents the additional variance attributable to the interaction between x_j and x_k . Higher-order terms represent interactions involving three or more input factors.

Two Sobol sensitivity indices were computed for each input factor. The first-order index, S_j , quantifies the proportion of output variance attributable to the main effect of x_j (Eq. (13)):

$$S_j = \frac{V_j}{\text{Var}(Y)}, \quad (13)$$

The total-order index, S_{T_j} , quantifies the contribution of x_j to the output variance, including its main effect and all interaction effects involving x_j (Eq. (14)):

$$S_{T_j} = \frac{E_{x_{\sim j}}[\text{Var}_{x_j}(Y|x_{\sim j})]}{\text{Var}(Y)} = 1 - \frac{\text{Var}_{x_{\sim j}}[E_{x_j}(Y|x_{\sim j})]}{\text{Var}(Y)}, \quad (14)$$

where $x_{\sim j}$ denotes all input factors except x_j . Under the assumed independent input distributions, both indices range from 0 to 1. The difference, $S_{T_j} - S_j$, quantifies the cumulative contribution of interaction effects involving x_j . Thus, values of S_{T_j} substantially larger than S_j indicate that x_j participates strongly in non-additive interactions with other predictors, an effect that cannot be identified from main-effect rankings alone.

Input samples for the Sobol sensitivity analysis were generated using the Saltelli sampling scheme based on Sobol' quasi-random sequences (Saltelli, 2002). For each continuous input factor, the sampling bounds were defined by the minimum and maximum values in the training dataset, and samples were generated independently over these bounds. A base sample size of 32,768 was used. Following the Saltelli sampling scheme with second-order effects enabled, this resulted in 524,288 surrogate-model evaluations for the seven input factors. This large evaluation budget was made computationally efficient by the trained XGBoost surrogate model, which evaluates each scenario several orders of magnitude faster than the corresponding box-model simulation. For ARI, which was encoded as an ordinal variable with integer values of 0, 1, and 2, the continuous Saltelli samples generated over [0,2] were rounded to the nearest integer and clipped to the valid range to match the representation used for model training. Sobol indices, including first-order and total-order indices, were estimated using the SALib library (Herman and Usher, 2017). Bootstrap 95 % confidence intervals were estimated for all Sobol indices.

- (3) Partial dependence plots (PDPs) were used to visualize the model-predicted average relationship between selected factors and the target variable (Friedman, 2001). For a factor of x_p and the complementary factor set x_c , the one-dimensional partial dependence function is defined in Eq. (15):

$$\hat{f}_p(x_p) = E_{x_c}[f(x_p, x_c)] \approx \frac{1}{n} \sum_{i=1}^n f(x_p, x_c^{(i)}), \quad (15)$$

where the expectation is approximated by averaging the surrogate model predictions ($\hat{f}_p$) over all n training samples, while fixing x_p at each point of a prescribed grid.

Two-dimensional PDPs were further constructed for selected factor pairs to visualize their joint dependence on the target variable. For a factor pair (x_p, x_q) , the two-dimensional partial dependence function is defined in Eq. (16):

$$\hat{f}_{pq}(x_p, x_q) = E_{x_c}[f(x_p, x_q, x_c)] \approx \frac{1}{n_{MC}} \sum_{i=1}^{n_{MC}} f(x_p, x_q, x_c^{(i)}), \quad (16)$$

where n_{MC} denotes a Monte Carlo subsample of the training data used to reduce computational cost over the two-dimensional grid.

(4) Individual conditional expectation (ICE) plots were used to complement the PDP analysis by examining the model response at the individual-sample level (Goldstein et al., 2015). Whereas a PDP describes the average model-predicted relationship between a factor and the target variable across samples, an ICE plots displays a separate conditional response curve for each sample. For each curve, the factor of interest is varied across a prescribed grid, while all remaining factors are held at their sample-specific values.

3 Results and Discussion

3.1 Effects of key controlling factors on sCI-mediated SO₂ oxidation

sCIs are short-lived reactive intermediates generated from the ozonolysis of alkenes. Their contribution to gas-phase SO₂ oxidation is governed jointly by their atmospheric steady-state concentration and their bimolecular rate coefficient with SO₂. The steady-state concentration is itself determined by the initial ozonolysis rate coefficient, the total sCI yield, and the structure-dependent reactivities of individual sCI species toward other competing sinks—most notably H₂O, (H₂O)₂, and NO₂—which collectively regulate the fraction of sCIs available to react with SO₂. We therefore introduce the sCI-mediated SO₂ oxidation rate, $L_{\text{SO}_2, \text{sCIs}}$, as the target variable, which simultaneously accounts for sCI production and the effective SO₂ oxidation capacity. Figures 1a and 1b show that, in both MCM v3.3.1 and MCM v3.3.1g, alkene and O₃ concentrations are positively correlated with $L_{\text{SO}_2, \text{sCIs}}$, whereas RH and NO_x exhibit marked negative correlations. This pattern is consistent with alkenes and O₃ driving sCI production, while H₂O/(H₂O)₂ and NO₂ act as competing sinks that reduce the fraction of sCIs available to react with SO₂. Notably, updating the mechanism substantially increases the magnitude of the absolute SHAP (|SHAP|) values across all features (Fig. 1c), with the average |SHAP| value ranging from 211.9 to 11,177.3 molecule cm⁻³ s⁻¹ in MCM v3.3.1, versus 694.0 to 30,429.5 molecule cm⁻³ s⁻¹ in MCM v3.3.1g. This indicates that the updated mechanism yields a substantially stronger sensitivity of $L_{\text{SO}_2, \text{sCIs}}$ to perturbations in environmental and precursor variables, such that the same change in a given feature produces a larger response in $L_{\text{SO}_2, \text{sCIs}}$. Importantly, the update does not amplify the influence of all features to the same degree; rather, it also reshapes the relative importance attributed to individual alkenes (Fig. 1d). In MCM v3.3.1, the alkene relative importance follows the order: *trans*-but-2-ene (TBUT2ENE, 21.5%) > *cis*-but-2-ene (CBUT2ENE, 17.6%) > isoprene (C₅H₈, 9.5%) > 2-methylpropene (MEPROPENE, 4.3%) > ethene (C₂H₄, 0.7%) > propene/but-1-ene (C₃C₄, 0.5%). In MCM v3.3.1g, isoprene becomes the most important alkene feature (23.8%), followed by *trans*-but-2-ene (11.0%), *cis*-but-2-ene (9.7%), and 2-methylpropene (6.0%), while ethene (0.9%), and propene/but-1-ene (0.7%) remain minor contributors. The approximately sixfold increase in the average isoprene |SHAP| value, rising from 3773.1 to 24,500.7 molecule cm⁻³ s⁻¹ after the mechanism update, is markedly larger than the corresponding shift observed for any other alkene. Indeed, prior studies have shown that regions with intense biogenic emissions tend to exhibit elevated

376 sCI-mediated SO₂ oxidation rates, an effect attributed primarily to sCIs derived from isoprene (Hata et al., 2023; Kukui et al.,
377 2021a). It should be noted, however, that the relative importance ranking obtained here reflects each alkene's attributed
378 contribution when all alkenes are perturbed over an identical concentration range (0.1–20 ppb); by contrast, isoprene's
379 prominence in previous field and modeling studies may partly reflect its typically higher ambient concentration relative to
380 other alkenes, rather than solely an intrinsically higher sCI reactivity. Within the present perturbation ranges (O₃: 2–90 ppb;
381 each alkene: 0.1–20 ppb; RH: 10%–80%), RH emerges as the most influential feature overall, exceeding the attributed
382 importance of every alkene (Fig. 1d). This is consistent with prior kinetic evidence that H₂O/(H₂O)₂ dominates the
383 bimolecular loss of sCIs, such that most sCIs have limited capacity to survive under high humidity (Cox et al., 2020; Lade et
384 al., 2024).

385

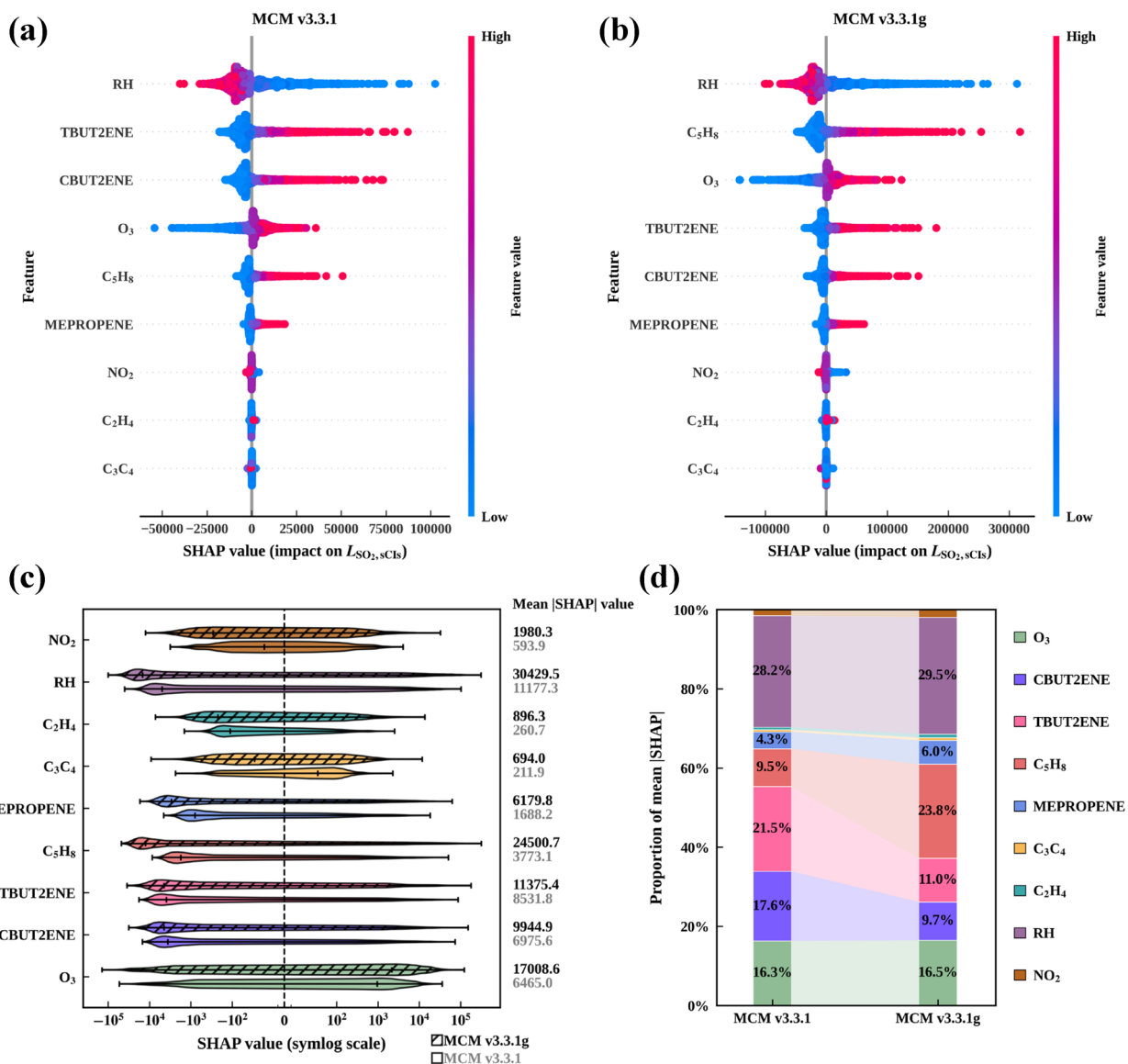
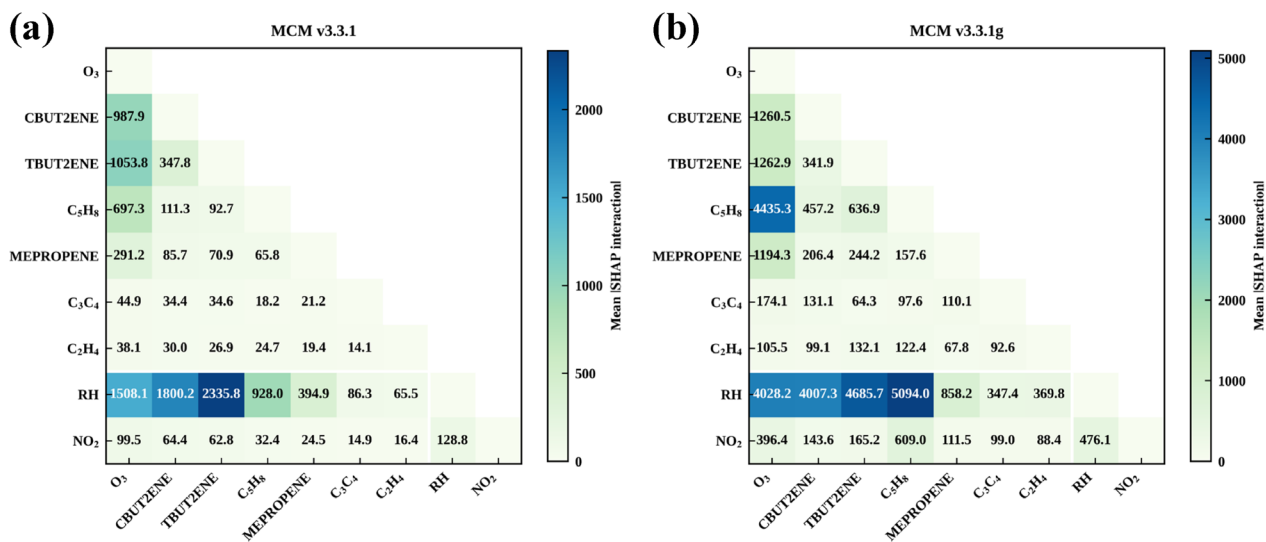


Figure 1: SHAP beeswarm summary plots for the XGBoost surrogate models of $L_{\text{SO}_2, \text{sClS}}$ under (a) MCM v3.3.1 and (b) MCM v3.3.1g. Each point represents one simulation scenario, positioned according to its SHAP value (molecule $\text{cm}^{-3} \text{s}^{-1}$) for the corresponding feature. Point colour denotes the feature value from low to high. Features are ordered by mean absolute SHAP value, so that those at the top exert the strongest overall influence on the predicted sCI-driven SO_2 oxidation rate. (c) Comparison of SHAP-value distributions for individual features between the XGBoost surrogate models trained on MCM v3.3.1 and MCM v3.3.1g. Boxplots summarize the full distribution of SHAP values across all simulation scenarios. Differences between mechanism versions reflect how the kinetic update changes both the magnitude of feature effects. (d) Relative feature contributions to the XGBoost surrogate models of $L_{\text{SO}_2, \text{sClS}}$ under MCM v3.3.1 and MCM v3.3.1g, quantified as the normalized mean absolute SHAP value of each feature. Contributions are expressed as fractions of the total sum of mean absolute SHAP values across all features for each model. This comparison highlights how the mechanism update redistributes the overall importance among ozonolysis precursors and competing sinks.

398 The changes in mean absolute SHAP interaction values before and after the update (Fig. 2) further clarify the mechanistic
 399 origin of the differential amplification magnitudes in alkene attribution. In MCM v3.3.1g, the elevated SHAP interaction
 400 values involving RH with both O₃ and the individual alkenes indicate that MCM v3.3.1 substantially underestimated the
 401 competitive scavenging of sCIs by H₂O/(H₂O)₂, thereby masking the strong regulatory role of humidity in sCI-mediated SO₂
 402 oxidation. In parallel, the interaction effects between O₃ and the individual alkenes are also enhanced in MCM v3.3.1g,
 403 indicating that MCM v3.3.1 likewise underestimated the sCI+SO₂ oxidation capacity. Critically, the relative degree of
 404 underestimation in these two competing removal pathways differs across alkene systems. For isoprene, the RH–C₅H₈
 405 interaction increased from 928.0 to 5094.0 molecule cm⁻³ s⁻¹ (a 5.5-fold increase), whereas the O₃–C₅H₈ interaction increased
 406 from 697.3 to 4435.3 molecule cm⁻³ s⁻¹ (a 6.4-fold increase). This indicates that MCM v3.3.1 underestimated the reactivity of
 407 its derived sCIs toward SO₂ to a greater extent than their reactivity toward H₂O/(H₂O)₂; a comparable pattern is observed for
 408 2-methylpropene. For the remaining alkenes, by contrast, the underestimation of the sCI+H₂O/(H₂O)₂ pathway predominates.
 409



410
 411 **Figure 2: Heat maps of pairwise SHAP interaction values among the nine input features for the XGBoost surrogate models of**
 412 **$L_{\text{SO}_2, \text{sCIs}}$ under MCM v3.3.1 and MCM v3.3.1g. Colours denote the mean absolute SHAP interaction value for each feature pair**
 413 **across all simulation scenarios. Larger values indicate stronger non-additive interactions in the surrogate prediction, thereby**
 414 **highlighting the precursor–sink combinations that most strongly modulate sCI-driven SO₂ oxidation.**

415 Based on the MCM v3.3.1g results, isoprene contributes more than twice as much as *cis*-/*trans*-but-2-ene, 3.8 times as much
 416 as 2-methylpropene, and roughly thirty times more than ethene and propene/but-1-ene in terms of their attributed influence
 417 on $L_{\text{SO}_2, \text{sCIs}}$ (Fig. 1c). At 298.15 K, the ozonolysis rate coefficients span more than two orders of magnitude (Table S1),
 418 following the order: *trans*-but-2-ene (2.0×10^{-16} cm³ molecule⁻¹ s⁻¹) > *cis*-but-2-ene (1.3×10^{-16} cm³ molecule⁻¹ s⁻¹) >
 419 isoprene (1.28×10^{-17} cm³ molecule⁻¹ s⁻¹) $\approx$ 2-methylpropene (1.15×10^{-17} cm³ molecule⁻¹ s⁻¹) $\approx$ propene (1.05×10^{-17} cm³
 420 molecule⁻¹ s⁻¹) $\approx$ but-1-ene (1.00×10^{-17} cm³ molecule⁻¹ s⁻¹) > ethene (1.56×10^{-18} cm³ molecule⁻¹ s⁻¹). Total sCI yields follow

the sequence (Tables S1 and S3): isoprene (0.66; 0.55 CH₂OO + 0.08 (CH=CH₂)(CH₃)COO + 0.03 (C(CH₃)=CH₂)CHOO) > ethene (0.42; CH₂OO) > propene (0.28; 0.20 CH₂OO + 0.08 CH₃CHOO) > but-1-ene (0.23; 0.15 CH₂OO + 0.08 C₂H₅CHOO) ≈ *trans*-but-2-ene (0.22; CH₃CHOO) ≈ *cis*-but-2-ene (0.20; CH₃CHOO) ≈ 2-methylpropene (0.20; 0.13 CH₂OO + 0.07 (CH₃)₂COO). Beyond overall yield, sCI speciation varies considerably: aside from *cis*- and *trans*-but-2-ene, CH₂OO is the dominant sCI across all remaining alkene systems. Critically, the bimolecular reactivities of sCIs are strongly structure-dependent. With respect to SO₂ oxidation, CH₃CHOO, C₂H₅CHOO, (CH₃)₂COO, and (C(CH₃)=CH₂)CHOO exhibit rate coefficients three to four times higher than those of CH₂OO and (CH=CH₂)(CH₃)COO. The reactivity order toward H₂O/(H₂O)₂ is markedly different, however, broadly following CH₃CHOO ≈ C₂H₅CHOO > CH₂OO > (CH=CH₂)(CH₃)COO ≈ (CH₃)₂COO >> (C(CH₃)=CH₂)CHOO. These contrasts in sCI fate, including ozonolysis kinetics, effective sCI yield (i.e., the fraction of sCIs that persist to undergo bimolecular reactions), and the relative importance of sCI+SO₂ vs. sCI+H₂O/(H₂O)₂ pathways, collectively shape the relationship between alkenes and $L_{\text{SO}_2, \text{sCIs}}$. Hence, the capacity of different alkene ozonolysis systems to oxidize SO₂ cannot be reduced to ozonolysis kinetics and sCI yield alone; the competitive removal of sCIs by H₂O/(H₂O)₂ should also be considered (Fig. 2b). A previous H₂SO₄ proxy study similarly noted that the apparent rate constant describing the alkene ozonolysis source spans nearly three orders of magnitude in the real atmosphere, reflecting the underlying complexity of sCI chemistry (Yang et al., 2021).

Fig. 3 further shows the effects of various alkenes on $L_{\text{SO}_2, \text{sCIs}}$ at different RH under MCM v3.3.1g, illustrating the regulatory role of the sCI+H₂O/(H₂O)₂ pathway. Consistent with the average |SHAP| ranking, isoprene, *trans*-but-2-ene, *cis*-but-2-ene, and 2-methylpropene all exhibit substantially positive impacts of varying degrees as their concentrations increase, whereas ethene and propene/but-1-ene display near-zero or weakly positive dependence across the investigated concentration range. The substantial vertical dispersion at fixed alkene concentrations, observed for all alkenes, indicates that identical alkene abundances can yield markedly different impacts on $L_{\text{SO}_2, \text{sCIs}}$, signifying the non-linear responses arising from interaction effects. For most alkenes (excluding propene/but-1-ene), this vertical spread increases with concentration, indicating that higher alkene concentrations expand the range of possible $L_{\text{SO}_2, \text{sCIs}}$ values, while the realized magnitude remains strongly contingent on ambient conditions. Propene/but-1-ene, by contrast, shows large vertical dispersion across the full concentration range without a significant concentration dependence, suggesting that variability in its SHAP values is driven almost entirely by interaction effects. Ethene similarly shows a significant degree of dispersion in its SHAP values at low concentrations. Taken together, the relationship between alkene concentration and $L_{\text{SO}_2, \text{sCIs}}$ is nonlinear for all alkenes; however, the degree of nonlinearity differs among alkenes, reflecting differences in the relative strength of their interactions with other feature variables.

We use RH coloring in Fig. 3 to visualize the strength of humidity's interactive effects. Across all alkenes, higher RH consistently suppresses the contributions of alkenes to $L_{\text{SO}_2, \text{sCIs}}$. The degree of suppression, however, varies markedly with alkene identity, and can be rationalized directly from the kinetic properties of their produced sCIs. Under high-humidity conditions, the contribution of propene/but-1-ene to the variation of $L_{\text{SO}_2, \text{sCIs}}$ is weak. This is attributable to the relatively low

rate coefficients for the reactions of propene/but-1-ene with O_3 , which result in significantly lower sCI production than that from ethene and isoprene. The primary sCIs generated from propene/but-1-ene ozonolysis are CH_2OO , CH_3CHOO , and C_2H_5CHOO , all of which react rapidly with $H_2O/(H_2O)_2$. Additionally, CH_2OO , which constitutes the largest proportion, has relatively low reactivity toward SO_2 . Consequently, the limited sCIs produced from these two alkenes make a further restricted contribution to sCI-mediated SO_2 oxidation. Ethene, despite having the slowest ozonolysis rate coefficient, produces CH_2OO with a substantially higher yield (0.42) than propene or but-1-ene, partially compensating for its sluggish kinetics. However, since CH_2OO is efficiently scavenged by $H_2O/(H_2O)_2$, ethene's contribution remains small and sensitive to humidity, consistent with the mixed positive and negative SHAP values observed at low concentrations in Fig. 3a. In contrast, *cis*-but-2-ene and *trans*-but-2-ene benefit from both the highest ozonolysis rate coefficients among the investigated alkenes and the exclusive production of CH_3CHOO , which reacts rapidly with SO_2 . These advantages translate into strong positive contributions to $L_{SO_2,sCIs}$; however, because CH_3CHOO is also highly reactive toward $H_2O/(H_2O)_2$, the realized contribution is strongly RH-dependent, as evidenced by the pronounced RH stratification in the dependence plots. 2-methylpropene generates both CH_2OO and $(CH_3)_2COO$. $(CH_3)_2COO$ reacts rapidly toward SO_2 but comparatively slowly toward $H_2O/(H_2O)_2$, resulting in weak competition from $H_2O/(H_2O)_2$. This explains why 2-methylpropene shows the smallest vertical dispersion growth with concentration among the more influential alkenes, and why its relationship with $L_{SO_2,sCIs}$ is more nearly linear, since $(CH_3)_2COO$ maintains its reactivity toward SO_2 even under elevated humidity. Isoprene exhibits the strongest positive contribution to $L_{SO_2,sCIs}$ across the full humidity range, arising from the favourable combination of relatively fast ozonolysis, the highest total sCI yield (0.66), and a diverse sCI speciation that includes not only CH_2OO but also $(CH=CH_2)(CH_3)COO$ and $(C(CH_3)=CH_2)CHOO$. $(C(CH_3)=CH_2)CHOO$ reacts rapidly with SO_2 while undergoing only slow scavenging by $H_2O/(H_2O)_2$, sustaining efficient SO_2 oxidation even at high RH. The interaction between isoprene and RH is thus primarily attributable to the CH_2OO fraction, whereas $(CH=CH_2)(CH_3)COO$ and $(C(CH_3)=CH_2)CHOO$ enable isoprene to remain a dominant driver of $L_{SO_2,sCIs}$ across a wider humidity range. Taken together, when the investigated alkenes are perturbed over same concentration ranges, ethene and propene/but-1-ene exert negligible direct control on $L_{SO_2,sCIs}$ relative to the remaining alkenes. Isoprene and the C_4 alkenes, *cis*-/*trans*-but-2-ene and 2-methylpropene, represent the dominant alkene precursors, but their effective impacts are jointly shaped by RH. This finding underscores the need for surrogate modeling to resolve these complex, nonlinear atmospheric reactions into a clear relationship between precursors and $L_{SO_2,sCIs}$. Such modeling enables the identification of key alkene precursors, quantification of the dependence of their contributions on ambient environmental conditions, and elucidation of the key reactions governing their roles in sCI-mediated SO_2 oxidation.

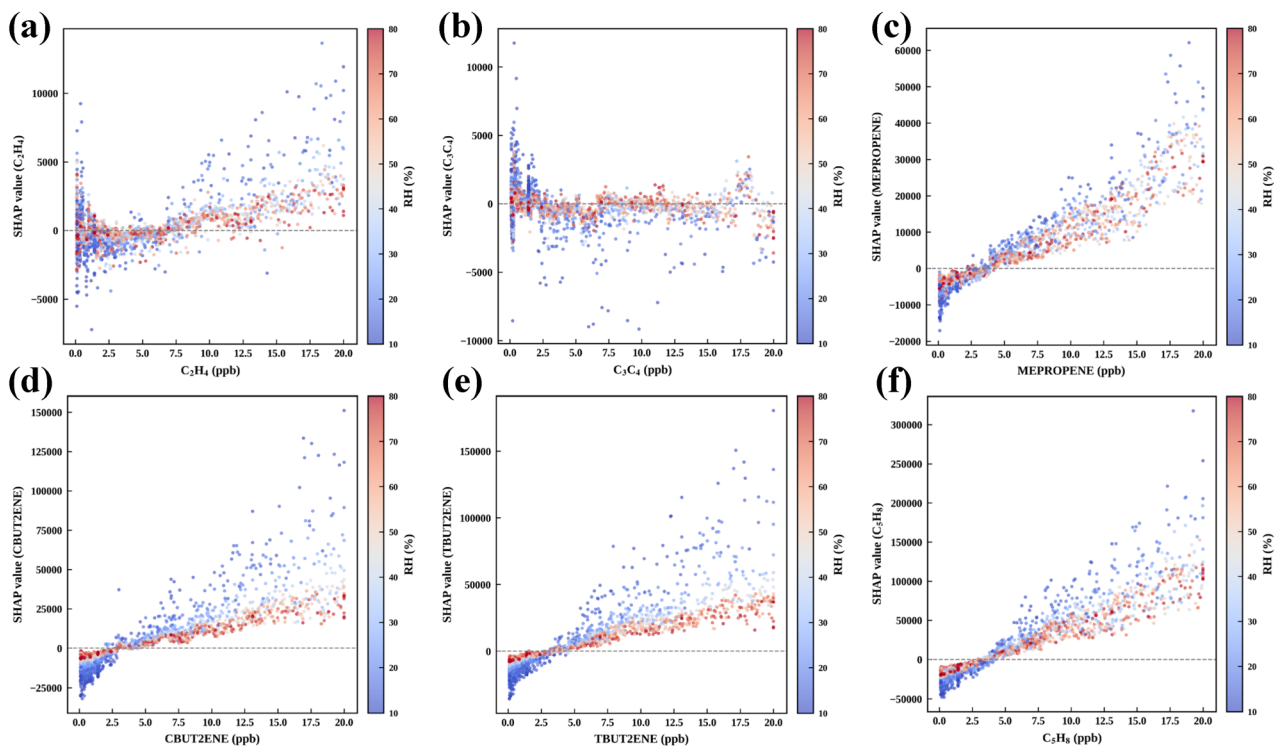


Figure 3: SHAP dependence plots for O_3 and the six alkene precursors—such as ethene (a), propene/but-1-ene (b), 2-methylpropene (c), *cis*-but-2-ene (d), *trans*-but-2-ene (e), and isoprene (f)—in the XGBoost surrogate models of $L_{\text{SO}_2, \text{sCIs}}$. The x axis shows the feature value, the y axis shows the SHAP value of the same feature, and point colour represents RH. Each point corresponds to one simulation scenario. These plots illustrate the marginal effect of each alkene precursor on the predicted sCI-driven SO_2 oxidation rate and how this effect varies with humidity.

3.2 Controls on the sCIs fractional contribution and regime-dependent sensitivities of SO_2 oxidation

The surrogate model developed in Sect. 3.1 quantifies how key factors govern $L_{\text{SO}_2, \text{sCIs}}$, the absolute rate of SO_2 oxidation by sCIs. A large $L_{\text{SO}_2, \text{sCIs}}$, however, does not necessarily imply that sCIs are important for SO_2 oxidation, because their importance also depends on the strength of the parallel OH-driven oxidation pathway. OH and sCIs share certain precursors, but their formation pathways are not entirely identical: alkene ozonolysis simultaneously produces both, so alkenes and O_3 drive both pathways, yet OH is additionally sustained through routes that do not involve sCIs, including $\text{RO}_x\text{--NO}_x$ radical cycling and the direct photolysis of O_3 , HONO, and other species. OH is therefore governed by a broader set of factors, including NO_x , $\text{NO}_2\%$, and the wider VOC composition, several of which exert little or no direct influence on sCI chemistry. As a result, the two oxidation pathways do not respond in tandem to environmental perturbations. Variations in the relative importance of sCIs therefore shape the sensitivity of L_{SO_2} to its controlling factors. To systematically diagnose the atmospheric conditions under which sCIs play a significant role in SO_2 oxidation, and to elucidate how this role modulates the precursor sensitivity of L_{SO_2} , we constructed dedicated surrogate models for μ_{sCIs} and L_{SO_2} . Their feature sets capture the

key controls on both pathways (see Sect. 2.3), and the models were trained on a large ensemble of constrained AtChem simulations spanning atmospheric conditions representative of anthropogenically and biogenically influenced environments. Because photolysis exerts decisive control over OH production, daytime (08:00–18:00 LT) and nighttime (22:00–05:00 LT) conditions were analyzed separately.

Sobol sensitivity analysis was applied to identify the dominant atmospheric controls on μ_{sClIs} and to quantify the extent to which each feature exerts its influence independently or through interaction with other features during the daytime period (Fig. 4a). Here, the first-order index S_1 measures the independent main-effect contribution of a single feature, whereas the total-order index S_T additionally encompasses all higher-order interactions involving that feature; a markedly larger S_T relative to S_1 therefore indicates that a feature operates predominantly through coupling with other features rather than through its own variation alone. The analysis identifies ARI, O_3 , alkene%, and VOCs as the dominant controls on μ_{sClIs} , with independent contributions ranked as ARI ($S_1 = 0.17$) > alkene% ($S_1 = 0.16$) > O_3 ($S_1 = 0.15$) > VOCs ($S_1 = 0.13$). RH ranks immediately behind ($S_1 = 0.09$), while NO_x and $\text{NO}_2\%$ register the weakest direct contributions ($S_1 = 0.03$ and 0.04 , respectively). The low-ARI tier (ARI = 0) is dominated by ethene and propene, representative of typical urban anthropogenic emissions (Zhao et al., 2020); the mid-ARI tier (ARI = 1) features an elevated proportion of but-2-ene, characteristic of industrial source environments or marine ecosystems (Giorio et al., 2022); and the high-ARI tier (ARI = 2) is driven primarily by isoprene, representative of biogenic (Rhew et al., 2017) or petrochemical-influenced conditions (Guo et al., 2022). The alkene mixture proportions at each ARI level are detailed in Fig. S1. Comparing S_1 and S_T across all features reveals a clear differentiation in how each feature exerts its influence. For RH, $S_T \approx S_1$ ($S_T = 0.11$), indicating negligible interaction effects; RH influences μ_{sClIs} almost entirely through its independent main effect. For ARI ($S_T = 0.27$), alkene% ($S_T = 0.26$), O_3 ($S_T = 0.22$), and VOCs ($S_T = 0.23$), S_T substantially exceeds S_1 , indicating that although their independent main effects dominate, non-negligible interaction effects are also present. Among these, ARI exhibits both the largest S_1 and S_T , reinforcing that alkene speciation not only exerts the strongest independent control on μ_{sClIs} but also acts as a key moderator of the effects of O_3 , alkene%, and VOCs. The contrast is most pronounced for NO_x and $\text{NO}_2\%$: their S_1 values account for less than half of their respective S_T values (NO_x : $S_T = 0.09$; $\text{NO}_2\%$: $S_T = 0.09$), indicating that their influence on μ_{sClIs} is overwhelmingly mediated through interactions with other features rather than through direct action. This strong interaction dependence reflects the fact that the influence of NO_x and $\text{NO}_2\%$ on μ_{sClIs} is conditional on the values of other features.

The directional effects of these features are further illustrated by the 2D PDPs (Fig. 4b-i). Hereafter, feature levels denote normalized values rather than the actual concentrations; corresponding time-varying concentrations are shown in Fig. S1. Among the features that directly govern $L_{\text{SO}_2, \text{sClIs}}$, ARI, O_3 , VOCs, and alkene% exhibit mutually reinforcing positive effects on μ_{sClIs} , suggesting that conditions favorable for alkene ozonolysis simultaneously enhance the relative competitiveness of the sCI pathway. RH negatively affects μ_{sClIs} by suppressing the sCI+ SO_2 pathway through enhanced sCI loss to $\text{H}_2\text{O}/(\text{H}_2\text{O})_2$. Despite NO_2 being a known sCI scavenger, $\text{NO}_2\%$ exhibits a synergistic co-enhancement with O_3 and VOCs on μ_{sClIs} . These synergistic effects are expected to arise from the influence of $\text{NO}_2\%$ on OH, through its role in governing how NO_x is partitioned between radical termination and cycling (Sillman and He, 2002). A higher NO_2 fraction, at fixed total NO_x , is

536 expected to reduce the NO available for HO₂/RO₂→OH conversion, thereby suppressing the SO₂+OH oxidation channel
 537 more strongly than it scavenges sCIs, particularly under high-O₃ and high-VOCs conditions. Correspondingly, the higher the
 538 NO₂%, the more strongly μ_{sCIs} increase with rising O₃ and VOCs. NO_x, by contrast, exerts predominantly suppressive
 539 interactions with O₃ and VOCs: μ_{sCIs} peaks at the lowest NO_x levels and declines sharply as NO_x increases only slightly from
 540 its minimum, before leveling off or increasing slightly at higher concentrations, such that μ_{sCIs} at the highest NO_x level
 541 remains well below the peak observed at the lowest NO_x level. The sensitivity of μ_{sCIs} to O₃ and VOCs is modulated by NO_x
 542 levels. At the lowest NO_x level, increases in O₃ and VOCs produce the largest increments in μ_{sCIs} . Across the remainder of
 543 the NO_x range, raising O₃ or VOCs from their minimum to maximum levels increases μ_{sCIs} by only 0.03–0.04. The effect of
 544 NO_x on μ_{sCIs} arises primarily from its control over the steady-state OH concentration: at extremely low NO_x level, radical
 545 recycling efficiency is greatly diminished, leading to suppressed OH; at excessively high NO_x levels, OH is likewise
 546 suppressed, but through accelerated termination reactions between radicals and NO_x. Under both conditions, the attenuation
 547 of OH-mediated SO₂ oxidation shifts a greater share of the total oxidation burden to sCIs, resulting in elevated μ_{sCIs} .
 548 Taken together, daytime μ_{sCIs} tends to be elevated under atmospheric conditions characterized by high O₃ concentrations,
 549 abundant VOCs, a large alkene fraction, and relatively low RH. With respect to NO_x, μ_{sCIs} is only appreciably enhanced
 550 under extremely low NO_x levels; across the remainder of the NO_x range, its influence on μ_{sCIs} is comparatively weak. This
 551 finding is consistent with previous studies reporting that sCI chemistry plays a more prominent role in SO₂ oxidation in
 552 biogenic-rich forested and rural environments (Kim et al., 2015) than in polluted urban settings (Dada et al., 2020). Among
 553 the identified controls, O₃, VOCs, alkene%, and ARI, which collectively characterize conditions favorable for sCI formation,
 554 not only directly promote $L_{\text{SO}_2, \text{sCIs}}$ but also exhibit pronounced synergistic positive effects on μ_{sCIs} . This pattern implies that,
 555 as these variables increase, OH increases less strongly than sCI production and may even decrease modestly. Consequently,
 556 although O₃ and alkenes jointly initiate the chemical processes leading to the formation of both OH and sCIs, their effects on
 557 the OH-mediated and sCI-mediated SO₂ oxidation pathways differ in both magnitude and direction. Under most daytime
 558 conditions, the OH pathway dominates SO₂ oxidation, such that the sensitivity of L_{SO_2} to precursor species is primarily
 559 shaped by how those precursors affect OH. However, as shown by the NO_x dependence of μ_{sCIs} , the response of OH to
 560 precursor perturbations is itself condition-dependent, exhibiting a non-monotonic relationship with NO_x. This gives rise to
 561 distinct OH response patterns under high- and low- μ_{sCIs} conditions and, by extension, distinct sensitivities of L_{SO_2} to its
 562 precursors. To quantify the magnitude of these differences, the dataset was partitioned into high- and low- μ_{sCIs} subsets based
 563 on the median daytime μ_{sCIs} value (1.1%), and the relationship between each feature and L_{SO_2} was characterized separately
 564 for each subset.
 565

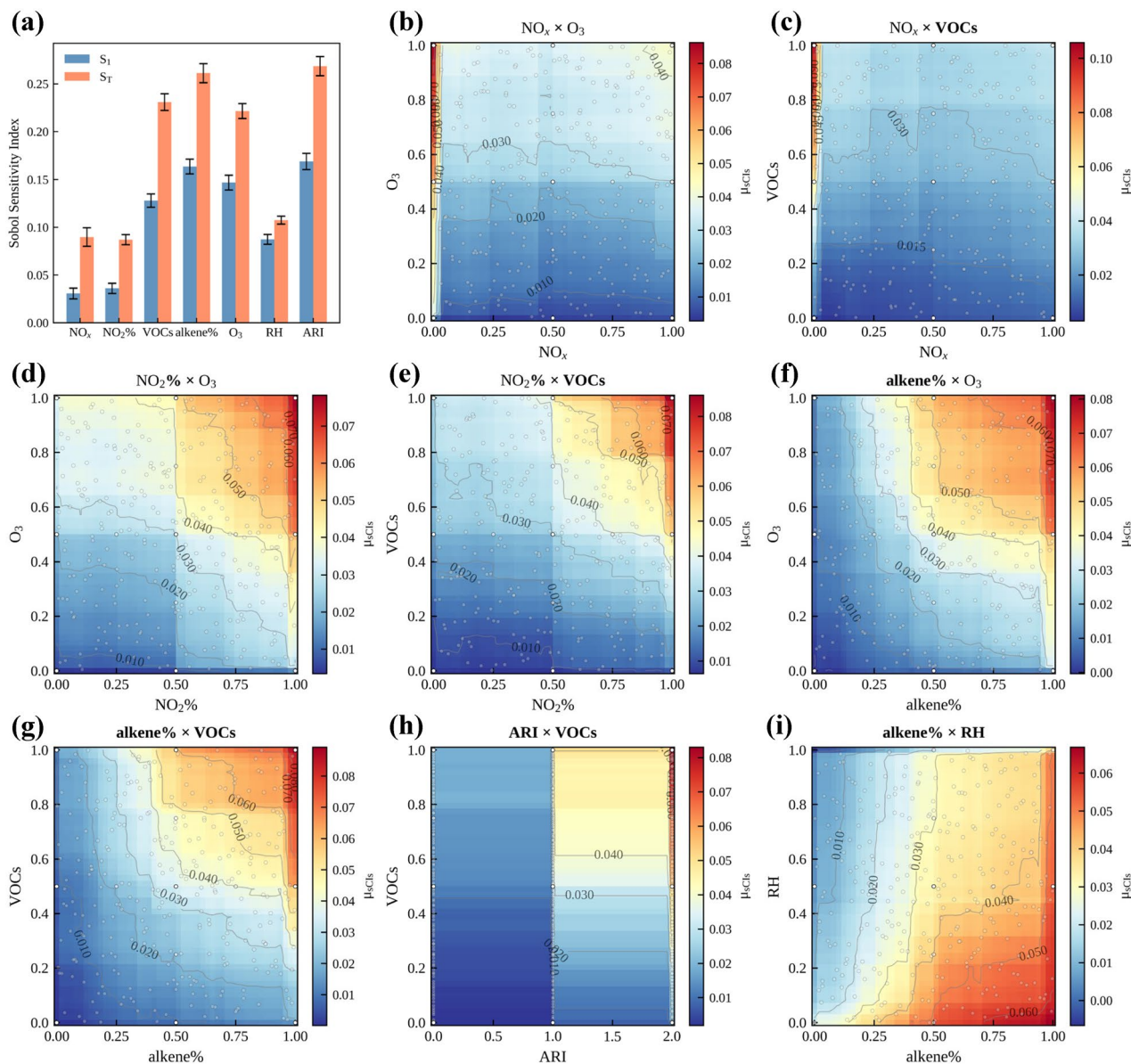


Figure 4: (a) First-order (S_1) and total-order (S_T) Sobol sensitivity indices for the seven input features in the XGBoost surrogate model of μ_{sClis} during daytime. S_1 quantifies the main effect of each feature acting alone, whereas S_T includes both the main effect and all interaction effects involving that feature. The difference between S_T and S_1 therefore reflects the extent to which a feature participates in non-additive interactions. (b-i) Two-dimensional partial dependence plots showing the joint effect of $\text{O}_3\text{norm} \times \text{NO}_x\text{norm}$, $\text{VOCsnorm} \times \text{NO}_x\text{norm}$, $\text{O}_3\text{norm} \times \text{NO}_2\%\text{norm}$, $\text{VOCsnorm} \times \text{NO}_2%\text{norm}$, $\text{O}_3\text{norm} \times \text{alkene}\%\text{norm}$, $\text{VOCsnorm} \times \text{alkene}\%\text{norm}$, $\text{VOCsnorm} \times \text{ARIcode}$, and $\text{RHnorm} \times \text{alkene}\%\text{norm}$ on the surrogate-model prediction of μ_{sClis} during daytime. The surface represents the average model response after marginalizing over the remaining features. Warmer colours indicate larger predicted μ_{sClis} .

Figure 5 illustrates how total gas-phase SO_2 oxidation rate varies with O_3 , VOCs, and alkene% across the high- and low- μ_{sClis} regimes. The mean ICE curves show that L_{SO_2} increases monotonically with all three features under both regimes; however,

the response is systematically amplified under high- μ_{sCl_2} conditions, with the mean curve exhibiting a steeper and more sustained increase across the full feature range. Beyond the mean response, the spread of individual ICE curves provides insight into response heterogeneity, namely, the extent to which the effect of a given feature on L_{SO_2} varies across different atmospheric conditions. Notably, negative L_{SO_2} responses to increases in VOCs and alkene% are observed under extremely low NO_x level. To quantify this directional heterogeneity, we computed the fraction of individual ICE curves exhibiting a net negative slope between the minimum and maximum feature values, termed the “neg-slope fraction”. For O_3 , positive associations with L_{SO_2} are nearly universal in both regimes, with a neg-slope fraction of 0% under the high- μ_{sCl_2} regime and 9% under the low- μ_{sCl_2} regime. For alkene%, the neg-slope fraction increases from 7% under the high- μ_{sCl_2} regime to 21% under the low- μ_{sCl_2} regime. A similar pattern is evident for VOCs, with the neg-slope fraction rising from 21% under high- μ_{sCl_2} to 35% under low- μ_{sCl_2} regimes. This regime contrast warrants further examination. Under extremely low NO_x level, high μ_{sCl_2} is associated with abundant alkene concentrations. Although higher alkene concentrations may promote radical termination and thereby suppress OH and reduce L_{SO_2} , the high- μ_{sCl_2} subset nevertheless exhibits a smaller neg-slope fraction than the low- μ_{sCl_2} subset. This suggests that the alkene ozonolysis pathway offsets part of the OH suppression caused by enhanced radical termination, thereby dampening the negative L_{SO_2} response. At moderate-to-high NO_x levels, positive L_{SO_2} responses to increases in VOCs, alkene%, and O_3 dominate and are considerably stronger under the high- μ_{sCl_2} regime than under the low- μ_{sCl_2} regime. This amplified sensitivity likely reflects the fact that high- μ_{sCl_2} regime is associated with abundant alkenes, which simultaneously promote OH propagation through RO_x cycling pathway and drive alkene ozonolysis.

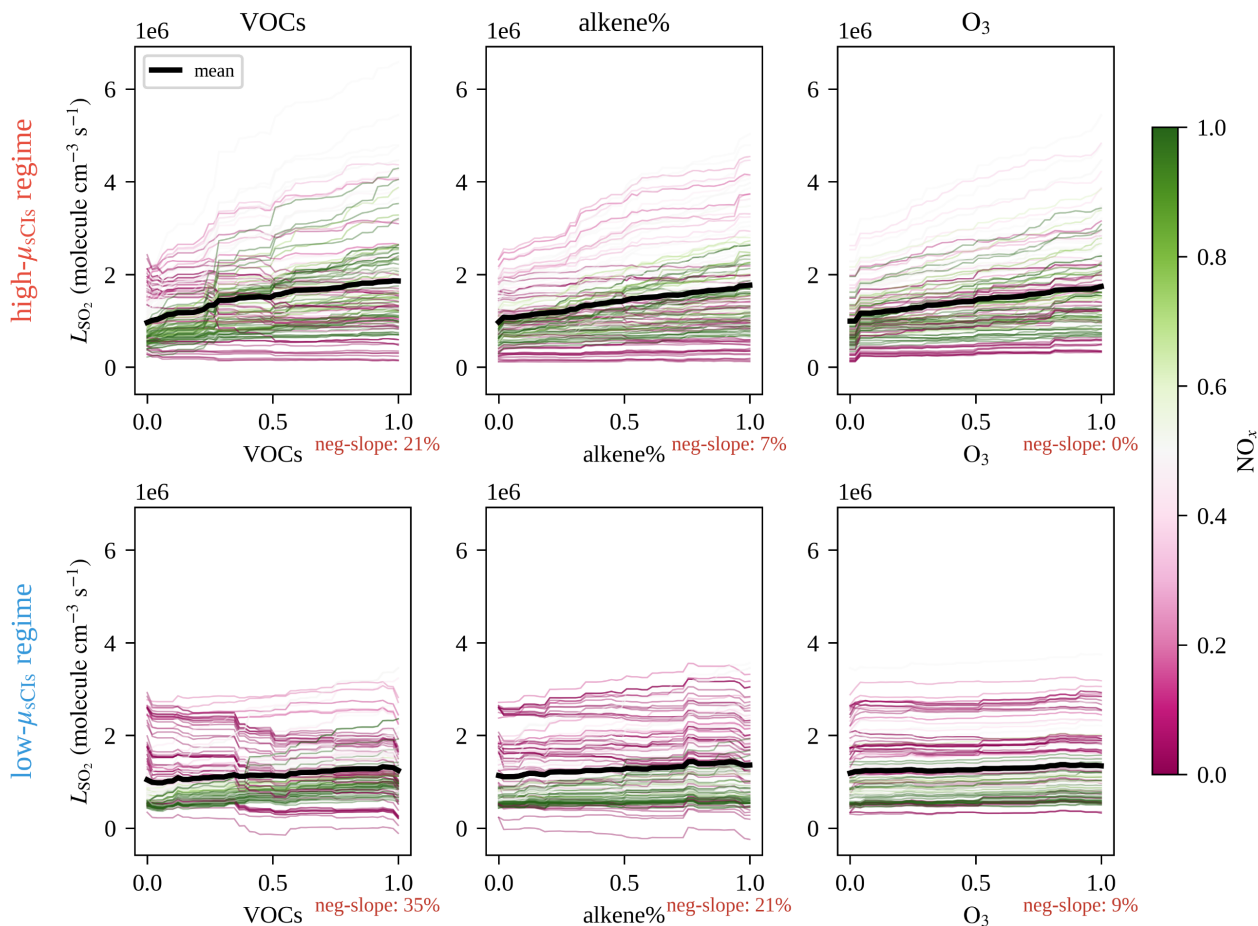


Figure 5: Individual conditional expectation (ICE) plots for the regime-specific L_{SO_2} models during daytime, showing the responses of predicted total gas-phase SO_2 oxidation to VOCsnorm, alkene%norm, and O_3 norm in the low- and high- μ_{sCl_s} regimes. The two rows correspond to the two μ_{sCl_s} regimes, and the three columns correspond to the three selected predictors. Each thin curve represents one sampled scenario, coloured by normalized NO_x . The curves are displayed in absolute prediction space (i.e. without centering), so both the slope and vertical spread reflect heterogeneity in the conditional model responses across different chemical backgrounds.

In summary, L_{SO_2} exhibits both positive and negative responses to VOCs and alkene% across the full scenario space, with the key distinction being that under the high- μ_{sCl_s} regime, negative responses are less frequent and positive responses are stronger. To clarify the mechanistic basis of this contrast, we examined the OH and sCI budgets for two representative scenarios at normalised NO_x levels of 0 and 0.5 (Fig. 6a, b). This analysis was used to explain why negative L_{SO_2} responses to increases in VOCs and alkene% are more frequent in the low- μ_{sCl_s} regime. OH production is partitioned into initiation ($\sum \text{OH}_{\text{new}}$), which encompasses all primary OH production via photolysis and alkene ozonolysis (Sheehy et al., 2010), and propagation ($\text{OH}_{\text{propag}}$), defined as the rate of $\text{HO}_2 \rightarrow \text{OH}$ conversion. Under extremely low NO_x level (NO_x , norm = 0), the ozonolysis pathway accounts for 39.65% of $\sum \text{OH}_{\text{new}}$ (see Text S4 for detailed calculations), and the ratio $\text{OH}_{\text{propag}}/\sum \text{OH}_{\text{new}}$ is

3.67, indicating that radical propagation contributes approximately 3.7 times as much OH as primary initiation. Under moderate NO_x level (NO_x , norm = 0.5), the ozonolysis share of $\sum \text{OH}_{\text{new}}$ decreases to 29.28%, whereas the $\text{OH}_{\text{propag}}/\sum \text{OH}_{\text{new}}$ ratio rises to 6.96, reflecting substantially enhanced radical cycling efficiency, as evidenced by the $\text{HO}_2 \rightarrow \text{OH}$ conversion rate reaching $7.04 \times 10^8 \text{ molecule cm}^{-3} \text{ s}^{-1}$. At extremely low NO_x level, the RO_x recycling is severely suppressed, thus primary OH initiation contributes a correspondingly larger fraction to the total OH budget, and the ozonolysis pathway in turn accounts for a larger share of OH production. Under this condition, the contribution of alkene ozonolysis to OH production can partly offset the suppressive effect of enhanced radical termination on OH, which explains why the neg-slope fraction for L_{SO_2} is smaller in the high- μ_{sClS} subset than in the low- μ_{sClS} subset. At moderate NO_x level, OH is sustained predominantly through radical cycling rather than through photolysis or ozonolysis, and the increase in NO_x additionally enhances HONO photolysis, further diminishing the relative contribution of ozonolysis to OH generation. Under these conditions, perturbations in VOCs and alkene% influence L_{SO_2} primarily through their modulation of OH propagation. Because high- μ_{sClS} conditions at this NO_x level are associated with abundant alkenes, both the RO_x cycling and ozonolysis pathways are strongly activated, and L_{SO_2} therefore exhibits a pronounced positive response to increases in precursor concentrations.

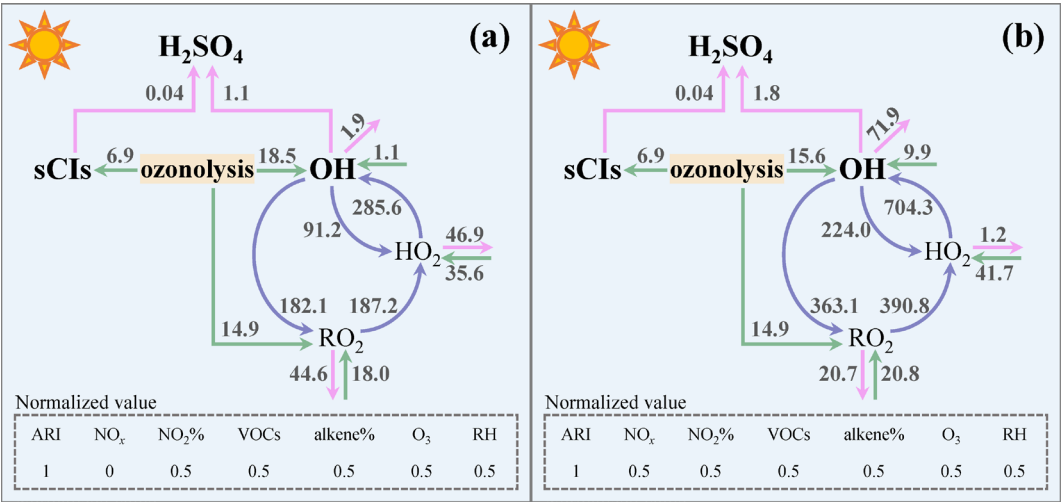


Figure 6: Two representative daytime scenarios illustrating the budgets of OH and sClS. Numbers denote reaction rates ($10^6 \text{ molecule cm}^{-3} \text{ s}^{-1}$). Pink, green, and purple lines indicate the termination, initiation, and propagation pathways of OH and sClS, respectively.

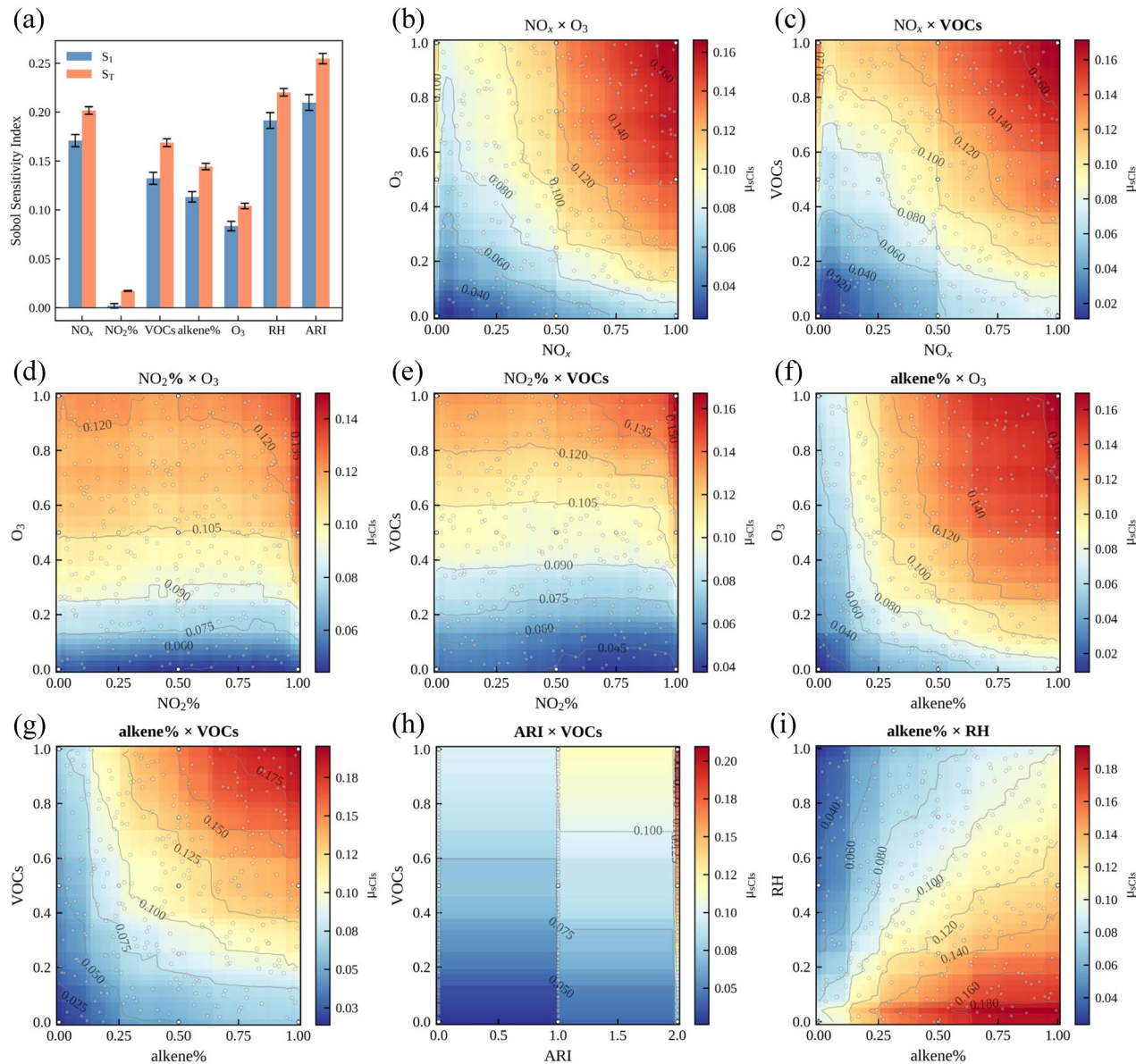
The nighttime Sobol sensitivity indices (Fig. 7a) reveal controls on μ_{sClS} that differ markedly from those identified during the day. ARI, RH and NO_x emerge as the dominant controls, with independent contributions ranked as ARI ($S_1 = 0.21$) > RH ($S_1 = 0.19$) > NO_x ($S_1 = 0.17$). VOCs ($S_1 = 0.13$), alkene% ($S_1 = 0.11$) and O_3 ($S_1 = 0.08$) rank immediately behind, while $\text{NO}_2\%$ exerts a negligible direct influence ($S_1 = 0.002$). Comparing S_1 and S_T across all features, S_T only slightly exceeds S_1 for most features, indicating that independent main effects dominate and interaction effects are limited, with the total-order ranking

634 following ARI ($S_T = 0.25$) > RH ($S_T = 0.22$) > NO_x ($S_T = 0.20$) > VOCs ($S_T = 0.17$) > alkene% ($S_T = 0.14$) > O₃ ($S_T = 0.10$).
635 NO₂% retains the smallest S_T ($S_T = 0.02$), reinforcing that its regulatory effect on μ_{sClIs} at night is negligible. The close
636 agreement between S_T and S_1 across most features at night stands in sharp contrast to the daytime results, where S_T
637 substantially exceeded S_1 for several key features. This collapse of interaction effects reflects a fundamental mechanistic
638 shift: in the absence of photolysis, the primary OH initiation pathway is alkene ozonolysis, which is simultaneously the sole
639 sCI formation pathway. Perturbations to precursors therefore affect both oxidation pathways in the same direction and with
640 similar magnitude, suppressing the conditional interactions that characterize the daytime sensitivity structure. This finding is
641 broadly consistent with previous studies demonstrating that Criegee intermediate chemistry constitutes a non-negligible OH
642 source under dark or low-light conditions (Khan et al., 2018). The rise of RH from a minor factor during the day to the
643 second-ranked control at night reflects the substantially lower OH concentrations characteristic of nocturnal conditions,
644 which elevate the relative contribution of sCIs to gas-phase SO₂ oxidation (Kukui et al., 2021b) and thereby amplify the
645 suppressive effect of H₂O/(H₂O)₂ on μ_{sClIs} . The contrasting behavior of NO_x and NO₂% between day and night also warrants
646 attention. At night, photolytic regeneration of NO from NO₂ is effectively absent; instead, NO within NO_x is oxidized to NO₂
647 through reactions with O₃ and peroxy radicals, leading to progressive NO₂ accumulation. This enhances the termination
648 reaction between NO₂ and OH, suppressing steady-state OH and indirectly elevating μ_{sClIs} . As a result, nighttime μ_{sClIs} is
649 governed primarily by total NO_x abundance rather than by the NO/NO₂ partitioning, which explains why NO₂% exerts a
650 negligible effect at night.

651 The directional effects of these features are further elucidated by the 2D-PDPs (Fig. 7b-i). ARI, O₃, VOCs, and alkene%
652 exhibit mutually reinforcing positive effects on μ_{sClIs} , consistent with their roles as sCI precursors: conditions favorable for
653 $L_{\text{SO}_2, \text{sClIs}}$ simultaneously enhance the relative competitiveness of the sCI oxidation pathway. RH exerts a pronounced negative
654 effect on μ_{sClIs} . With respect to NO_x, μ_{sClIs} reaches a locally elevated value (approximately 0.1) at the lowest NO_x level, then
655 drops sharply as NO_x increases only slightly from its minimum, before rising significantly and monotonically across the
656 remainder of the range to reach its global maximum (approximately 0.16) at the highest NO_x level, such that a strong positive
657 relationship between NO_x and μ_{sClIs} prevails across most of the range. This near-monotonic behavior reflects the role of NO_x
658 in modulating steady-state OH through radical termination, as described above.

659 Taken together, nighttime μ_{sClIs} tends to be elevated under atmospheric conditions characterized by high O₃ concentrations,
660 abundant VOCs, a large alkene fraction, and relatively low RH. With respect to NO_x, the highest μ_{sClIs} values occur under
661 elevated NO_x levels, although a smaller enhancement is also evident under extremely low NO_x conditions. This is consistent
662 with a previous study reporting that elevated NO_x levels in power plant environments at night are highly conducive to a large
663 sCI contribution fraction (Meidan et al., 2019). Furthermore, ARI, O₃, VOCs, and alkene%, which collectively characterize
664 conditions favorable for sCI formation, not only strongly promote $L_{\text{SO}_2, \text{sClIs}}$ but also exert pronounced synergistic effects on
665 μ_{sClIs} . This pattern indicates that O₃ and alkenes do not affect the OH- and sCI-mediated SO₂ oxidation pathways to the same
666 extent. Notably, unlike during the day, when OH is sustained primarily by radical propagation, nocturnal OH production is
667 more strongly linked to primary initiation, alkene ozonolysis, because the absence of photolysis limits OH production from

668 photolytic sources and radical recycling. Alkene ozonolysis is also the sole formation pathway for sCIs. Therefore,
 669 differences in the sensitivity of L_{SO_2} across μ_{sCIs} regimes are expected to reflect differences in the strength of the alkene
 670 ozonolysis pathway. To examine these differences, the dataset was partitioned into high- and low- μ_{sCIs} subsets based on the
 671 median nighttime μ_{sCIs} value (8.0%), and the relationship between each feature and L_{SO_2} was characterized separately for
 672 each subset.
 673



674

675 **Figure 7: (a) First-order (S_1) and total-order (S_T) Sobol sensitivity indices for the seven input features in the XGBoost surrogate**
676 **model of μ_{sCl_2} during nighttime. S_1 quantifies the main effect of each feature acting alone, whereas S_T includes both the main effect**
677 **and all interaction effects involving that feature. The difference between S_T and S_1 therefore reflects the extent to which a feature**
678 **participates in non-additive interactions. (b-i) Two-dimensional partial dependence plots showing the joint effect of**
679 **$\text{O}_3\text{norm} \times \text{NO}_x\text{norm}$, $\text{VOCsnorm} \times \text{NO}_x\text{norm}$, $\text{O}_3\text{norm} \times \text{NO}_2\%\text{norm}$, $\text{VOCsnorm} \times \text{NO}_2\%\text{norm}$, $\text{O}_3\text{norm} \times \text{alkene}\%\text{norm}$,**
680 **$\text{VOCsnorm} \times \text{alkene}\%\text{norm}$, $\text{VOCsnorm} \times \text{ARIcode}$, and $\text{RHnorm} \times \text{alkene}\%\text{norm}$ on the surrogate-model prediction of μ_{sCl_2} during**
681 **nighttime. The surface represents the average model response after marginalizing over the remaining features. Warmer colours**
682 **indicate larger predicted μ_{sCl_2} .**

683 Figure 8 shows how L_{SO_2} responds to precursor perturbations across the high- μ_{sCl_2} and low- μ_{sCl_2} regimes at night. The mean
684 ICE curves reveal that L_{SO_2} increases monotonically with all three features under both regimes; however, the response is
685 systematically enhanced under the high- μ_{sCl_2} regime, with the mean curve exhibiting a steeper and more sustained increase
686 across the full feature range. Beyond the mean response, the spread of individual ICE curves provides insight into response
687 heterogeneity across different atmospheric conditions. For O_3 , near-universal positive associations with L_{SO_2} are observed in
688 both regimes, with neg-slope fractions of 1% and 3% under the high- and low- μ_{sCl_2} regimes, respectively, indicating that O_3
689 consistently promotes L_{SO_2} regardless of the μ_{sCl_2} regime. For alkene%, the neg-slope fraction increases from 1% under the
690 high- μ_{sCl_2} regime to 10% under the low- μ_{sCl_2} regime. The contrast is most pronounced for VOCs, for which the neg-slope
691 fraction rises from 5% under the high- μ_{sCl_2} regime to 29% under the low- μ_{sCl_2} regime, with the excess negative responses in
692 the low- μ_{sCl_2} regime arising predominantly from high- NO_x scenarios.

693

694

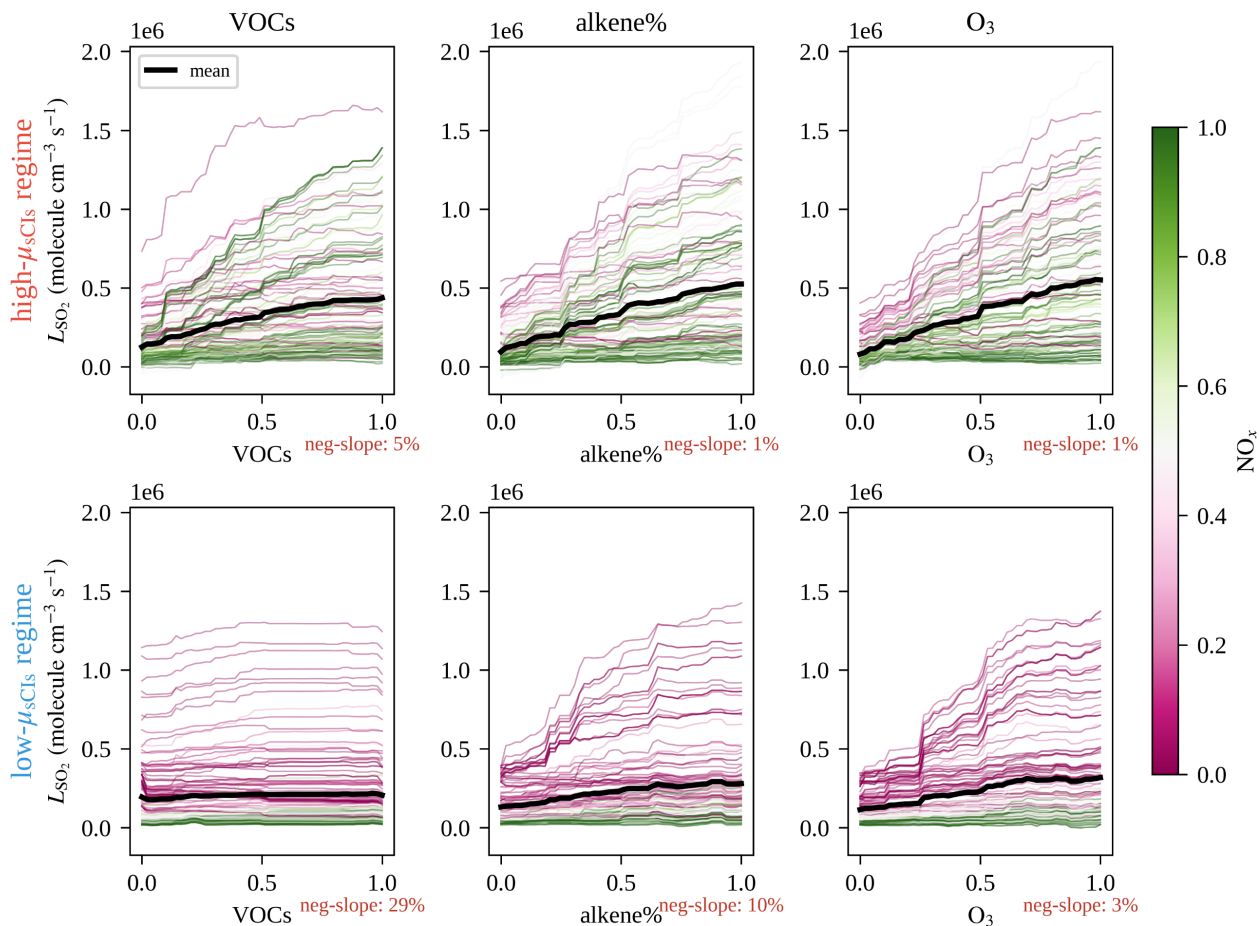
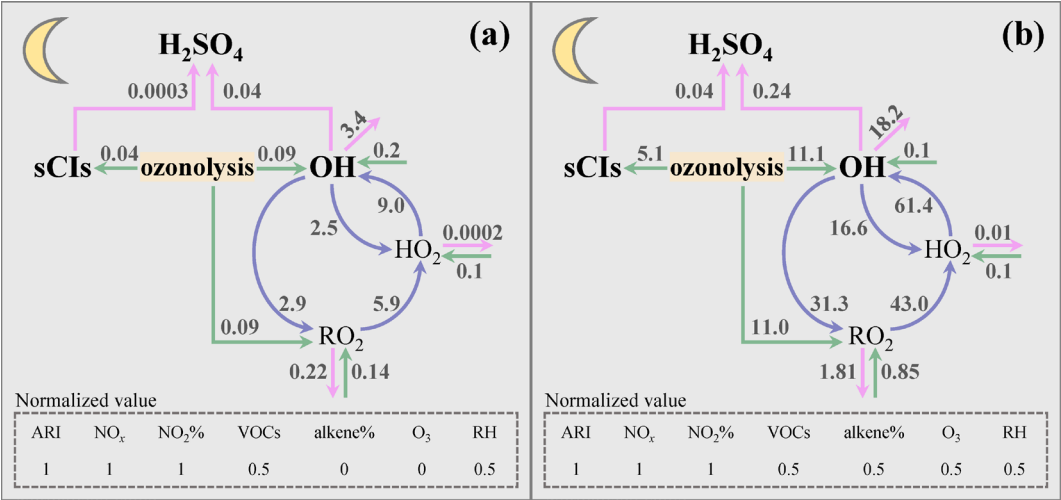


Figure 8: Individual conditional expectation (ICE) plots for the regime-specific L_{SO_2} models during nighttime, showing the responses of predicted total gas-phase SO_2 oxidation to VOCsnorm, alkene%norm, and O_3 norm in the low- and high- μ_{sCl} regimes. The two rows correspond to the two μ_{sCl} regimes, and the three columns correspond to the three selected predictors. Each thin curve represents one sampled scenario, coloured by normalized NO_x . The curves are displayed in absolute prediction space (i.e. without centering), so both the slope and vertical spread reflect heterogeneity in the conditional model responses across different chemical backgrounds.

To elucidate the role of alkene ozonolysis in driving these response differences, we examined the OH and sCI budgets for two representative high- NO_x scenarios corresponding to the low- μ_{sCl} ($\mu_{\text{sCl}} = 0.008$) and high- μ_{sCl} ($\mu_{\text{sCl}} = 0.14$) regimes, respectively (Fig. 9a, b). Under the low- μ_{sCl} scenario, alkene ozonolysis accounts for 30.18% of $\sum \text{OH}_{\text{new}}$, which is $5.56 \times 10^5 \text{ molecule cm}^{-3} \text{ s}^{-1}$. In contrast, under the high- μ_{sCl} scenario, alkene ozonolysis contributes 94.24% of $\sum \text{OH}_{\text{new}}$, which reaches $2.29 \times 10^7 \text{ molecules cm}^{-3} \text{ s}^{-1}$. These results show that, when alkene ozonolysis is efficient at night, it acts not only as the sole formation pathway for sCIs but also as the dominant OH initiation pathway. When alkene ozonolysis is weak, the ratio $\text{OH}_{\text{propag}}/\sum \text{OH}_{\text{new}}$ is 16.23, indicating that OH is sustained predominantly through radical cycling. In this regime, OH initiation is supplied primarily by photolysis in the early morning hours, with a smaller contribution from ozonolysis.

Reducing total VOCs does not affect the photolytic OH initiation source, but decreases the OH-to-RO₂ flux and thus weakens OH loss through VOC oxidation. Provided that radical recycling efficiency changes little, this reduction in OH reactivity can slightly increase steady-state OH and L_{SO₂,OH}. Because the alkene-derived sCI flux is negligible in this regime, ΔL_{SO₂,sCIs} is close to zero. The OH-mediated response can therefore dominate, producing a net increase in L_{SO₂} as VOCs decrease (a negative slope). Conversely, when alkene ozonolysis is strong, it simultaneously serves as the dominant OH initiation source and the principal sCI formation pathway. The ratio OH_{propag}/ΣOH_{new} decreases to 2.68, far below that under weak ozonolysis, indicating a much greater relative contribution of initiation to OH. Under these conditions, VOC perturbations influence the OH- and sCI- mediated SO₂ oxidation pathways in the same direction, thereby reducing the occurrence of negative L_{SO₂} responses. Between the two scenarios, the rates of ozonolysis-driven sCI and OH formation increase by factors of 127.5 and 123.3, respectively. Accordingly, the rates of sCI to forming H₂SO₄ increases by a factor of 133, whereas the rates of OH to forming H₂SO₄ increases by only a factor of six. This difference arises because OH reacts with alkenes and is incorporated into the RO_x cycle, such that part of the additional OH production supports radical cycling rather than directly increasing steady-state OH. The contrasting neg-slope fractions for VOCs and alkene% can thus be explained by their distinct chemical roles. Reducing alkene% simultaneously weakens a primary OH initiation source and the sCI formation pathway, such that L_{SO₂} declines almost universally. By contrast, reducing total VOCs primarily lowers OH reactivity rather than suppressing OH initiation under alkene-poor conditions corresponding to the low-μ_{sCIs} regime. The resulting increase in OH-mediated oxidation can outweigh the negligible change in the sCI pathway, producing a negative L_{SO₂} response (neg-slope = 29%).

728



729

730 **Figure 9: Two representative nighttime scenarios illustrating the budgets of OH and sCIs. Numbers denote reaction rates (10⁶**
 731 **molecule cm⁻³ s⁻¹). Pink, green, and purple lines indicate the termination, initiation, and propagation pathways of OH and sCIs,**
 732 **respectively.**

3.3 Observational–model corroboration for regime-dependent sensitivities of SO₂ oxidation

To validate the predictive capability of the surrogate model developed in Section 3.2 for μ_{sClS} , and to evaluate how the sensitivity of SO₂ oxidation to its precursors depends on the relative contributions of sCIs in ambient air, we selected a summer episode in Wuhai City as a real-atmosphere validation case. Figure 10 presents the constrained simulation results derived from AtChem–MCM v3.3.1, including time series of sulfuric acid concentrations (H₂SO₄–AtChem) and the daytime- and nighttime-averaged fractional contributions of sCIs to total SO₂ oxidation (μ_{sClS} –AtChem) over the period 1 June to 15 July 2021. The corresponding pollutant mixing ratios and meteorological parameters are shown in Figure S3. SO₄²⁻ denotes the ambient particulate sulfate mass concentrations measured concurrently by online ion chromatography. The term “ μ_{sClS} –surrogate model” represents the daytime- and nighttime-averaged μ_{sClS} predictions generated by the surrogate model described in Section 3.2. The strong agreement ($R^2 = 0.75$) between the surrogate predictions (μ_{sClS} –surrogate model) and the explicit AtChem simulations (μ_{sClS} –AtChem) demonstrates the robust predictive performance of the constructed machine learning surrogate for μ_{sClS} .

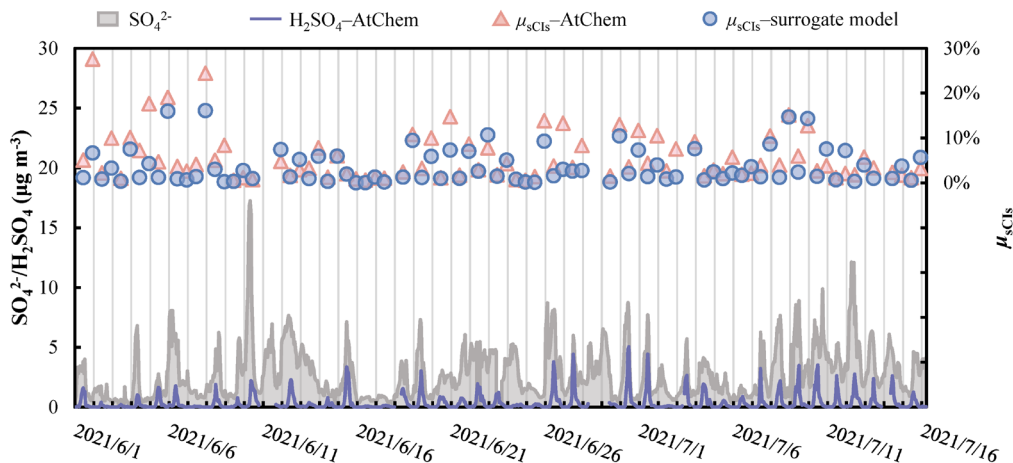


Figure 10: Constrained AtChem–MCM v3.3.1 simulations for a summer episode in Wuhai City (1 June–15 July 2021), showing time series of H₂SO₄–AtChem (AtChem-simulated sulfuric acid concentration), μ_{sClS} –AtChem (AtChem-simulated fractional contribution of sCIs to total SO₂ oxidation), SO₄²⁻ (concurrently measured ambient particulate sulfate mass concentration), and μ_{sClS} –surrogate model (daytime- and nighttime-averaged μ_{sClS} predicted by the surrogate model).

To evaluate whether ambient SO₄²⁻ concentrations serve as a reliable proxy for gas-phase H₂SO₄ production during the target episode, an XGBoost model was trained on a comprehensive long-term observational dataset spanning 9 October 2019 to 30 June 2022 (summary statistics of all variables are provided in Table S4), with hourly SO₄²⁻ concentration as the target variable. The feature set included PM_{2.5}, NO₃⁻, SO₂, O₃, VOCs, alkene%, NO_x, NO₂%, RH, wind speed (WS), and Fe. Training over this extended period ensures that the model captures the full range of variability in sulfate formation, including seasonal cycles, episodic pollution events, and varying meteorological regimes, thereby establishing a representative and

well-generalized mapping between precursor conditions and SO_4^{2-} . SHAP values were then extracted exclusively for the target episode (1 June to 15 July 2021) to attribute feature contributions during this specific period. As shown in Figure 11a, the three highest-ranked features are NO_3^- , SO_2 , and $\text{PM}_{2.5}$, all of which exhibit clear positive correlations with SO_4^{2-} . The elevated NO_3^- co-occurs with SO_4^{2-} , together with the concurrent increase in $\text{PM}_{2.5}$, consistent with the hallmark of secondary inorganic aerosol formation (Gao et al., 2021). The strong positive SHAP contribution of SO_2 directly corroborates its role as the primary precursor for SO_4^{2-} production. Notably, WS also exerts a discernible positive contribution, suggesting a degree of regional transport influence during this episode, although its impact remains secondary to in situ chemical production. By contrast, RH shows no significant positive correlation with SO_4^{2-} and ranks relatively low in feature importance. Fe, a key catalyst for aqueous-phase SO_2 oxidation via Fenton-type chemistry (Ye et al., 2023b), shows no clear positive association with SO_4^{2-} , with elevated ambient Fe concentrations not corresponding to enhanced sulfate formation. Taken together, these results indicate that aqueous-phase oxidation was not the dominant SO_4^{2-} formation pathway during the target episode. Meanwhile, O_3 , VOCs, and alkene%, which serve as key proxies of gas-phase oxidant pathways, exhibit positive correlations with SO_4^{2-} , providing evidence that gas-phase oxidation constituted the primary sulfate formation pathway during this period. On this basis, ambient SO_4^{2-} concentrations can be regarded as a reliable observational proxy for the variability in total gas-phase H_2SO_4 production during the target episode.

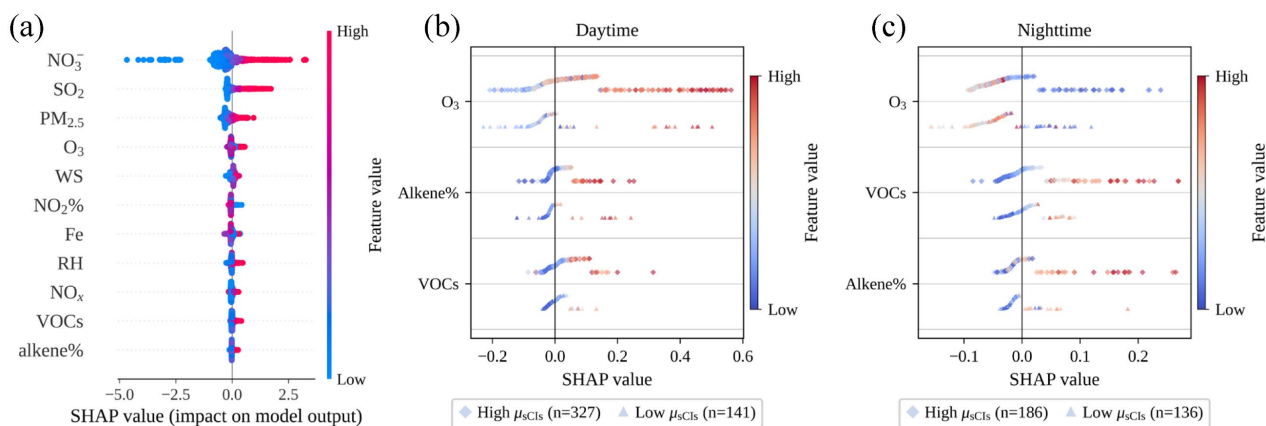


Figure 11: (a) SHAP beeswarm summary plot for the observation-based XGBoost model of ambient SO_4^{2-} . Each point represents one hourly observation, positioned according to the SHAP value of the corresponding predictor. Positive SHAP values indicate that the predictor increases the model-predicted SO_4^{2-} concentration, whereas negative values indicate a suppressing effect. Point colour denotes the predictor magnitude from low to high. Features are ordered by mean absolute SHAP value, so the plot summarizes both the relative importance and the directionality of the physically interpretable features. (b) Comparison of SHAP beeswarm plots for the observation-based XGBoost model under daytime low- and high- μ_{sCIs} regimes. Each point represents one hourly observation in the corresponding subset. The horizontal position gives the SHAP value, and the colour indicates the feature value. Differences between the two panels illustrate how the relative importance and directional effects of observed predictors on ambient SO_4^{2-} vary between daytime chemical regimes with low and high inferred sCI contributions. (c) Comparison of SHAP beeswarm plots for the observation-based XGBoost model under nighttime low- and high- μ_{sCIs} regimes. Each point represents one hourly observation in the corresponding subset. The horizontal position gives the SHAP value, and the colour indicates the feature value. Differences between the two panels illustrate how the relative importance and directional effects of observed predictors on ambient SO_4^{2-} vary between nighttime chemical regimes with low and high inferred sCI contributions.

To further examine whether the sensitivity of SO_4^{2-} to key precursors is consistent with the regime-dependent behavior of L_{SO_2} identified in Sect. 3.2, the target episode was partitioned into high- and low- μ_{sCI} subsets for daytime and nighttime periods separately, using the same classification thresholds applied in Sect. 3.2 (daytime: 1.1%; nighttime: 8.0%). SHAP values for O_3 , VOCs, and alkene% were extracted for each subset and compared. During the daytime (Fig. 11b), the high- μ_{sCI} subset exhibits systematically amplified SHAP responses relative to the low- μ_{sCI} subset across all three features. O_3 shows the most pronounced regime contrast: in the high- μ_{sCI} subset, high feature values (red) are associated with distinctly larger positive SHAP contributions and a considerably wider spread, whereas the corresponding responses in the low- μ_{sCI} subset are substantially attenuated. VOCs and alkene% follow the same pattern, with high feature values driving larger positive SHAP contributions under high- μ_{sCI} conditions. This daytime regime contrast is consistent with the finding in Sect. 3.2 that the sensitivity of SO_2 oxidation to O_3 , VOCs, and alkene% is amplified under high- μ_{sCI} conditions, providing observational corroboration of that mechanistic result. At night (Fig. 11c), VOCs and alkene% exhibit stronger positive SHAP contributions to SO_4^{2-} in the high- μ_{sCI} subset, consistent with their roles in promoting both sCIs and OH formation through nocturnal alkene ozonolysis. By contrast, O_3 exhibits negative SHAP contributions in both subsets, which can be attributed to the distinct behavior of O_3 at night: nocturnal O_3 is primarily of advective origin rather than in situ photochemical production, and elevated O_3 concentrations are often associated with suppressed local NO levels due to rapid O_3 -NO titration. Taken together, the regime-dependent SO_4^{2-} sensitivity patterns observed during the target episode (Fig. 11b, c) provide observational validation for the mechanistic findings derived from the AtChem-based surrogate model simulations in Sect. 3.2, confirming that these findings are manifested in ambient atmospheric measurements. This cross-validation between model-diagnosed sensitivity and observation-based attribution reinforces the conclusion that under high- μ_{sCI} regimes, alkene-targeted emission control strategies can effectively reduce sCIs and suppress the total SO_2 oxidation rate, while rarely triggering a rebound in OH-driven oxidation pathways.

4 Summary and conclusion

Sulfuric acid plays a central role in atmospheric aerosol formation, yet effective mitigation requires quantifying the sensitivity of H_2SO_4 formation to its precursors which varies across atmospheric conditions as the relative contributions of different gas-phase SO_2 oxidation pathways shift. By treating sCIs as reactive intermediate species linking precursor emissions to H_2SO_4 formation, this study deploys an interpretable multi-target XGBoost modelling framework trained on ensemble box-model simulations incorporating updated CI chemistry. This framework determines the controls on sCI-mediated SO_2 oxidation ($L_{\text{SO}_2, \text{sCI}}$) and on the fractional sCI contribution (μ_{sCI}), while establishing how the regime of μ_{sCI} itself reshapes the sensitivity of total SO_2 oxidation (L_{SO_2}) to its precursors. $L_{\text{SO}_2, \text{sCI}}$ is dominated by RH (29.5%) and isoprene (23.8%), revealing that low humidity and high biogenic emissions primarily promote the sCI-mediated oxidation pathways. For μ_{sCI} , the dominant controlling factors exhibit a pronounced diurnal shift: daytime control resides in alkene reactivity (ARI), O_3 , alkene fraction, and total VOCs, whereas nighttime

control transitions to ARI, RH, and NO_x . High- μ_{sCI} regimes uniformly amplify positive L_{SO_2} sensitivities to O_3 , VOCs, and alkene%, while decreasing the frequency of occasional negative L_{SO_2} responses to VOCs or alkene%. Observation-based modeling of a Wuhai summer episode with SHAP-based attribution of ambient sulfate provided support for this regime dependence.

Previous studies have extensively reported the contribution of the $\text{SO}_2 + \text{sCI}$ reaction to total sCI loss (Cox et al., 2020), and the contribution of sCIs to overall SO_2 oxidation across diverse atmospheric conditions (Khan et al., 2018). Building on the understanding that sCI chemistry constitutes one pathway for SO_2 oxidation and that sCIs are secondarily formed atmospheric intermediates, this study no longer treats them as terminal oxidants but rather takes seriously their role as an intermediate species and establishes quantitative relationships linking controlling factors to both the sCI-mediated SO_2 oxidation rate and its fractional contribution to total SO_2 oxidation. The resulting picture goes beyond estimating the sCIs contribution for an individual scenario, extending instead to a predictive understanding of the role sCIs may play across a wider range of atmospheric conditions. By further analyzing how the sensitivity of total SO_2 oxidation to factor perturbations varies among internal oxidation regimes characterized by different μ_{sCI} values, this work provides a scientific basis for developing strategies to mitigate H_2SO_4 formation. Beyond these mechanistic-sensitivity insights, the interpretable surrogate modeling framework developed here offers a transferable approach for distilling quantitative relationships between reactants and products from complex reaction networks. By linking targeted box-model scenario design to machine-learning-based sensitivity diagnostics, this methodology enables efficient identification of dominant kinetic controls—providing a basis for prioritizing kinetic processes for explicit representation when developing lumped chemical mechanisms for chemical transport models.

It should be noted that, applying these findings to real-world atmospheric phenomena requires recognizing specific applicability conditions. While key feature variables were explicitly modeled, capturing full atmospheric variability also requires future integration of transport, deposition, multiphase processes, temperature dependence, and other factors. Additionally, our 100% SO_3 yield assumption for the $\text{sCI} + \text{SO}_2$ reaction establishes an upper bound for absolute sCI-mediated H_2SO_4 production and fractional contributions. Because this yield functions as a linear scaling factor, absolute magnitudes may be overestimated, but the underlying functional relationships between controlling factors and sCI chemistry remain unchanged.

Data availability

The dataset used in this study is publicly available at: <https://doi.org/10.5281/zenodo.17946348>.

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

849 **Author contributions**

850 YZ developed the methodology, performed data curation, visualization, formal analysis, and wrote the original draft. QC
851 was responsible for conceptualization, formal analysis, funding acquisition, and reviewed and edited the manuscript. LH and
852 WG participated in the revision of the manuscript. CS, DG, LJ and GY contributed to data curation.

853 **Competing interests**

854 The contact author has declared that none of the authors has any competing interests.

855 **Disclaimer**

856 Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional
857 affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to
858 include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the
859 authors and do not necessarily reflect the views of the publisher.

860 **Acknowledgements**

861 This research was supported by the Lanzhou science and technology bureau, Project No. 2022-2-15, Research and
862 Development of Key technologies for photocatalysis of low concentration organic waste Gas and the National Natural
863 Science Foundation of China (grant no. 42305118). We thank the editors and anonymous reviewers for their valuable
864 comments on the manuscript. During the preparation of this work, we used ChatGPT to proofread and refine the English
865 expression of the manuscript. This included correcting grammatical errors, enhancing clarity, and improving the overall flow
866 and coherence of the text. After using this tool, we reviewed and edited the content and take full responsibility for the
867 content of the published article.

868 **Financial support**

869 The Lanzhou science and technology bureau, Project No. 2022-2-15, Research and Development of Key technologies for
870 photocatalysis of low concentration organic waste Gas. The National Natural Science Foundation of China (grant no.
871 42305118).

872 References

- 873 Boy, M., Mogensen, D., Smolander, S., Zhou, L., Nieminen, T., Paasonen, P., Plass-Dülmer, C., Sipilä, M., Petäjä, T.,
874 Mauldin, L., Berresheim, H., and Kulmala, M.: Oxidation of SO₂ by stabilized Criegee intermediate (sCI) radicals as a
875 crucial source for atmospheric sulfuric acid concentrations, *Atmos. Chem. Phys.*, 13, 3865–3879,
876 <https://doi.org/10.5194/acp-13-3865-2013>, 2013.
- 877 Cao, J., Qiu, X., Gao, J., Wang, F., Wang, J., Wu, J., and Peng, L.: Significant decrease in SO₂ emission and enhanced
878 atmospheric oxidation trigger changes in sulfate formation pathways in China during 2008–2016, *Journal of Cleaner*
879 *Production*, 326, 129396, <https://doi.org/10.1016/j.jclepro.2021.129396>, 2021.
- 880 Cao, J., Zheng, B., Chen, J., and Liu, Y.: Assessment of sulfate and nitrate variations in China during 1990–2020: Insights
881 into source contributions and formation pathways, *Journal of Hazardous Materials*, 489, 137600,
882 <https://doi.org/10.1016/j.jhazmat.2025.137600>, 2025.
- 883 Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD*
884 *International Conference on Knowledge Discovery and Data Mining, KDD '16: The 22nd ACM SIGKDD International*
885 *Conference on Knowledge Discovery and Data Mining, San Francisco California USA*, 785–794,
886 <https://doi.org/10.1145/2939672.2939785>, 2016.
- 887 Cox, R. A. and Penkett, S. A.: Oxidation of Atmospheric SO₂ by Products of the Ozone–Olefin Reaction, *Nature*, 230, 321–
888 322, <https://doi.org/10.1038/230321a0>, 1971.
- 889 Cox, R. A., Ammann, M., Crowley, J. N., Herrmann, H., Jenkin, M. E., McNeill, V. F., Mellouki, A., Troe, J., and
890 Wallington, T. J.: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume VII – Criegee intermediates,
891 *Atmos. Chem. Phys.*, 20, 13497–13519, <https://doi.org/10.5194/acp-20-13497-2020>, 2020.
- 892 Dada, L., Ylivinkka, I., Baalbaki, R., Li, C., Guo, Y., Yan, C., Yao, L., Sarnela, N., Jokinen, T., Daellenbach, K. R., Yin, R.,
893 Deng, C., Chu, B., Nieminen, T., Wang, Y., Lin, Z., Thakur, R. C., Kontkanen, J., Stolzenburg, D., Sipilä, M., Hussein, T.,
894 Paasonen, P., Bianchi, F., Salma, I., Weidinger, T., Pikridas, M., Sciare, J., Jiang, J., Liu, Y., Petäjä, T., Kerminen, V.-M.,
895 and Kulmala, M.: Sources and sinks driving sulfuric acid concentrations in contrasting environments: implications on proxy
896 calculations, *Atmos. Chem. Phys.*, 20, 11747–11766, <https://doi.org/10.5194/acp-20-11747-2020>, 2020.
- 897 Dai, Q., Bi, X., Song, W., Li, T., Liu, B., Ding, J., Xu, J., Song, C., Yang, N., Schulze, B. C., Zhang, Y., Feng, Y., and
898 Hopke, P. K.: Residential coal combustion as a source of primary sulfate in Xi'an, China, *Atmospheric Environment*, 196,
899 66–76, <https://doi.org/10.1016/j.atmosenv.2018.10.002>, 2019.
- 900 Friedman, J. H.: Greedy function approximation: A gradient boosting machine., *Ann. Statist.*, 29,
901 <https://doi.org/10.1214/aos/1013203451>, 2001.
- 902 Gao, J., Li, Y., Li, J., Shi, G., Liu, Z., Han, B., Tian, X., Wang, Y., Feng, Y., and Russell, A. G.: Impact of Formation
903 Pathways on Secondary Inorganic Aerosol During Haze Pollution in Beijing: Quantitative Evidence From High-Resolution
904 Observation and Modeling, *Geophysical Research Letters*, 48, e2021GL095623, <https://doi.org/10.1029/2021GL095623>,
905 2021.
- 906 Geng, G., Zheng, Y., Zhang, Q., Xue, T., Zhao, H., Tong, D., Zheng, B., Li, M., Liu, F., Hong, C., He, K., and Davis, S. J.:
907 Drivers of PM_{2.5} air pollution deaths in China 2002–2017, *Nat. Geosci.*, 14, 645–650, [https://doi.org/10.1038/s41561-021-](https://doi.org/10.1038/s41561-021-00792-3)
908 00792-3, 2021.

909 Giorio, C., Doussin, J., D'Anna, B., Mas, S., Filippi, D., Denjean, C., Mallet, M. D., Bourrienne, T., Burnet, F., Cazaunau,
910 M., Chikwililwa, C., Desboeufs, K., Feron, A., Michoud, V., Namwoonde, A., Andreae, M. O., Piketh, S. J., and Formenti,
911 P.: Butene Emissions From Coastal Ecosystems May Contribute to New Particle Formation, *Geophysical Research Letters*,
912 49, e2022GL098770, <https://doi.org/10.1029/2022GL098770>, 2022.

913 Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E.: Peeking Inside the Black Box: Visualizing Statistical Learning With
914 Plots of Individual Conditional Expectation, *Journal of Computational and Graphical Statistics*, 24, 44–65,
915 <https://doi.org/10.1080/10618600.2014.907095>, 2015.

916 Gunthe, S. S., Rose, D., Su, H., Garland, R. M., Achtert, P., Nowak, A., Wiedensohler, A., Kuwata, M., Takegawa, N.,
917 Kondo, Y., Hu, M., Shao, M., Zhu, T., Andreae, M. O., and Pöschl, U.: Cloud condensation nuclei (CCN) from fresh and
918 aged air pollution in the megacity region of Beijing, *Atmos. Chem. Phys.*, 11, 11023–11039, [https://doi.org/10.5194/acp-11-](https://doi.org/10.5194/acp-11-11023-2011)
919 11023-2011, 2011.

920 Guo, W., Yang, Y., Chen, Q., Zhu, Y., Zhang, Y., Zhang, Y., Liu, Y., Li, G., Sun, W., and She, J.: Chemical reactivity of
921 volatile organic compounds and their effects on ozone formation in a petrochemical industrial area of Lanzhou, Western
922 China, *Science of The Total Environment*, 839, 155901, <https://doi.org/10.1016/j.scitotenv.2022.155901>, 2022.

923 Guo, Y., Yan, C., Li, C., Ma, W., Feng, Z., Zhou, Y., Lin, Z., Dada, L., Stolzenburg, D., Yin, R., Kontkanen, J., Daellenbach,
924 K. R., Kangasluoma, J., Yao, L., Chu, B., Wang, Y., Cai, R., Bianchi, F., Liu, Y., and Kulmala, M.: Formation of nighttime
925 sulfuric acid from the ozonolysis of alkenes in Beijing, *Atmos. Chem. Phys.*, 21, 5499–5511, [https://doi.org/10.5194/acp-21-](https://doi.org/10.5194/acp-21-5499-2021)
926 5499-2021, 2021.

927 H. Khan, M. A., Morris, W. C., Galloway, M., A. Shallcross, B. M., Percival, C. J., and Shallcross, D. E.: An Estimation of
928 the Levels of Stabilized Criegee Intermediates in the UK Urban and Rural Atmosphere Using the Steady-State
929 Approximation and the Potential Effects of These Intermediates on Tropospheric Oxidation Cycles: AN ESTIMATION OF
930 THE LEVELS OF STABILIZED CRIEGEE INTERMEDIATES IN THE UK, *Int. J. Chem. Kinet.*, 49, 611–621,
931 <https://doi.org/10.1002/kin.21101>, 2017.

932 Hata, H., Hoshino, S., Fujita, M., and Tonokura, K.: Atmospheric impact of isoprene-derived Criegee intermediates and
933 isoprene hydroxy hydroperoxide on sulfate aerosol formation in the Asian region, *Atmospheric Environment: X*, 20, 100226,
934 <https://doi.org/10.1016/j.aeaoa.2023.100226>, 2023.

935 Herman, J. and Usher, W.: SALib: An open-source Python library for Sensitivity Analysis, *JOSS*, 2, 97,
936 <https://doi.org/10.21105/joss.00097>, 2017.

937 Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol
938 for mechanism development, *Atmospheric Environment*, 31, 81–104, [https://doi.org/10.1016/S1352-2310\(96\)00105-7](https://doi.org/10.1016/S1352-2310(96)00105-7), 1997.

939 Jenkin, M. E., Young, J. C., and Rickard, A. R.: The MCM v3.3.1 degradation scheme for isoprene, *Atmos. Chem. Phys.*, 15,
940 11433–11459, <https://doi.org/10.5194/acp-15-11433-2015>, 2015.

941 Khan, M. A. H., Percival, C. J., Caravan, R. L., Taatjes, C. A., and Shallcross, D. E.: Criegee intermediates and their impacts
942 on the troposphere, *Environ. Sci.: Processes Impacts*, 20, 437–453, <https://doi.org/10.1039/C7EM00585G>, 2018.

943 Kim, S., Guenther, A., Lefer, B., Flynn, J., Griffin, R., Rutter, A. P., Gong, L., and Cevik, B. K.: Potential Role of Stabilized
944 Criegee Radicals in Sulfuric Acid Production in a High Biogenic VOC Environment, *Environ. Sci. Technol.*, 49, 3383–3391,
945 <https://doi.org/10.1021/es505793t>, 2015.

946 Kukui, A., Chartier, M., Wang, J., Chen, H., Dusanter, S., Sauvage, S., Michoud, V., Locoge, N., Gros, V., Bourrianne, T.,
 947 Sellegri, K., and Pichon, J.-M.: Role of Criegee intermediates in the formation of sulfuric acid at a Mediterranean (Cape
 948 Corsica) site under influence of biogenic emissions, *Atmos. Chem. Phys.*, 21, 13333–13351, [https://doi.org/10.5194/acp-21-](https://doi.org/10.5194/acp-21-13333-2021)
 949 13333-2021, 2021a.

950 Kukui, A., Chartier, M., Wang, J., Chen, H., Dusanter, S., Sauvage, S., Michoud, V., Locoge, N., Gros, V., Bourrianne, T.,
 951 Sellegri, K., and Pichon, J.-M.: Role of Criegee intermediates in the formation of sulfuric acid at a Mediterranean (Cape
 952 Corsica) site under influence of biogenic emissions, *Atmos. Chem. Phys.*, 21, 13333–13351, [https://doi.org/10.5194/acp-21-](https://doi.org/10.5194/acp-21-13333-2021)
 953 13333-2021, 2021b.

954 Lade, R. E., Blitz, M. A., Rowlinson, M., Evans, M. J., Seakins, P. W., and Stone, D.: Kinetics of the reactions of the
 955 Criegee intermediate CH_2OO with water vapour: experimental measurements as a function of temperature and global
 956 atmospheric modelling, *Environ. Sci.: Atmos.*, 4, 1294–1308, <https://doi.org/10.1039/D4EA00097H>, 2024.

957 Li, J., Ying, Q., Yi, B., and Yang, P.: Role of stabilized Criegee Intermediates in the formation of atmospheric sulfate in
 958 eastern United States, *Atmospheric Environment*, 79, 442–447, <https://doi.org/10.1016/j.atmosenv.2013.06.048>, 2013.

959 Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, 2017.

960 Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K.-W., Newman, S.-
 961 F., Kim, J., and Lee, S.-I.: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery, *Nat*
 962 *Biomed Eng*, 2, 749–760, <https://doi.org/10.1038/s41551-018-0304-0>, 2018.

963 Lundberg, S. M., Erion, G. G., and Lee, S.-I.: Consistent Individualized Feature Attribution for Tree Ensembles,
 964 <https://doi.org/10.48550/arXiv.1802.03888>, 7 March 2019.

965 Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee,
 966 S.-I.: From local explanations to global understanding with explainable AI for trees, *Nat Mach Intell*, 2, 56–67,
 967 <https://doi.org/10.1038/s42256-019-0138-9>, 2020.

968 Mauldin, R. L., Berndt, T., Sipilä, M., Paasonen, P., Petäjä, T., Kim, S., Kurtén, T., Stratmann, F., Kerminen, V.-M., and
 969 Kulmala, M.: A new atmospherically relevant oxidant of sulphur dioxide, *Nature*, 488, 193–196,
 970 <https://doi.org/10.1038/nature11278>, 2012.

971 Meidan, D., Holloway, J. S., Edwards, P. M., Dubé, W. P., Middlebrook, A. M., Liao, J., Welti, A., Graus, M., Warneke, C.,
 972 Ryerson, T. B., Pollack, I. B., Brown, S. S., and Rudich, Y.: Role of Criegee Intermediates in Secondary Sulfate Aerosol
 973 Formation in Nocturnal Power Plant Plumes in the Southeast US, *ACS Earth and Space Chemistry*, 2019.

974 Newland, M. J., Rickard, A. R., Alam, M. S., Vereecken, L., Muñoz, A., Ródenas, M., and Bloss, W. J.: Kinetics of
 975 stabilised Criegee intermediates derived from alkene ozonolysis: reactions with SO_2 , H_2O and decomposition under
 976 boundary layer conditions, *Phys. Chem. Chem. Phys.*, 17, 4076–4088, <https://doi.org/10.1039/C4CP04186K>, 2015.

977 Onel, L., Lade, R., Mortiboy, J., Blitz, M. A., Seakins, P. W., Heard, D. E., and Stone, D.: Kinetics of the gas phase reaction
 978 of the Criegee intermediate CH_2OO with SO_2 as a function of temperature, *Phys. Chem. Chem. Phys.*, 23, 19415–19423,
 979 <https://doi.org/10.1039/D1CP02932K>, 2021.

980 Quaas, J., Jia, H., Smith, C., Albright, A. L., Aas, W., Bellouin, N., Boucher, O., Doutriaux-Boucher, M., Forster, P. M.,
 981 Grosvenor, D., Jenkins, S., Klimont, Z., Loeb, N. G., Ma, X., Naik, V., Paulot, F., Stier, P., Wild, M., Myhre, G., and Schulz,
 982 M.: Robust evidence for reversal of the trend in aerosol effective climate forcing, *Atmos. Chem. Phys.*, 22, 12221–12239,
 983 <https://doi.org/10.5194/acp-22-12221-2022>, 2022.

- 984 Rhew, R. C., Deventer, M. J., Turnipseed, A. A., Warneke, C., Ortega, J., Shen, S., Martinez, L., Koss, A., Lerner, B. M.,
 985 Gilman, J. B., Smith, J. N., Guenther, A. B., and De Gouw, J. A.: Ethene, propene, butene and isoprene emissions from a
 986 ponderosa pine forest measured by relaxed eddy accumulation, *Atmos. Chem. Phys.*, 17, 13417–13438,
 987 <https://doi.org/10.5194/acp-17-13417-2017>, 2017.
- 988 Ruscic, B.: Active Thermochemical Tables: Water and Water Dimer, *J. Phys. Chem. A*, 117, 11940–11953,
 989 <https://doi.org/10.1021/jp403197t>, 2013.
- 990 Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications*, 145,
 991 280–297, [https://doi.org/10.1016/S0010-4655\(02\)00280-1](https://doi.org/10.1016/S0010-4655(02)00280-1), 2002.
- 992 Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S.: Variance based sensitivity analysis of
 993 model output. Design and estimator for the total sensitivity index, *Computer Physics Communications*, 181, 259–270,
 994 <https://doi.org/10.1016/j.cpc.2009.09.018>, 2010.
- 995 Sarwar, G., Simon, H., Fahey, K., Mathur, R., Goliff, W. S., and Stockwell, W. R.: Impact of sulfur dioxide oxidation by
 996 Stabilized Criegee Intermediate on sulfate, *Atmospheric Environment*, 85, 204–214,
 997 <https://doi.org/10.1016/j.atmosenv.2013.12.013>, 2014.
- 998 Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical
 999 Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, *Atmos. Chem. Phys.*,
 1000 3, 161–180, <https://doi.org/10.5194/acp-3-161-2003>, 2003.
- 1001 Scribano, Y., Goldman, N., Saykally, R. J., and Leforestier, C.: Water Dimers in the Atmosphere III: Equilibrium Constant
 1002 from a Flexible Potential, *J. Phys. Chem. A*, 110, 5411–5419, <https://doi.org/10.1021/jp056759k>, 2006.
- 1003 Sheehy, P. M., Volkamer, R., Molina, L. T., and Molina, M. J.: Oxidative capacity of the Mexico City atmosphere – Part 2:
 1004 A RO_x radical cycling perspective, *Atmos. Chem. Phys.*, 10, 6993–7008, <https://doi.org/10.5194/acp-10-6993-2010>, 2010.
- 1005 Sillman, S. and He, D.: Some theoretical results concerning O₃ -NO_x -VOC chemistry and NO_x -VOC indicators, *J.-*
 1006 *Geophys.-Res.*, 107, <https://doi.org/10.1029/2001JD001123>, 2002.
- 1007 Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and*
 1008 *Computers in Simulation*, 55, 271–280, [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6), 2001.
- 1009 Sommariva, R., Cox, S., Martin, C., Borońska, K., Young, J., Jimack, P. K., Pilling, M. J., Matthaios, V. N., Nelson, B. S.,
 1010 Newland, M. J., Panagi, M., Bloss, W. J., Monks, P. S., and Rickard, A. R.: AtChem (version 1), an open-source box model
 1011 for the Master Chemical Mechanism, *Geosci. Model Dev.*, 13, 169–183, <https://doi.org/10.5194/gmd-13-169-2020>, 2020.
- 1012 Song, X., Yi, J., Chen, Y., Su, Y., Wang, H., Liu, A., Wu, D., and Li, Q.: Condensable particulate matter emissions regulated
 1013 by flue gas desulfurization technologies in typical industrial plants, *Journal of Hazardous Materials*, 489, 137527,
 1014 <https://doi.org/10.1016/j.jhazmat.2025.137527>, 2025.
- 1015 Taatjes, C. A., Welz, O., Eskola, A. J., Savee, J. D., Scheer, A. M., Shallcross, D. E., Rotavera, B., Lee, E. P. F., Dyke, J. M.,
 1016 Mok, D. K. W., Osborn, D. L., and Percival, C. J.: Direct Measurements of Conformer-Dependent Reactivity of the Criegee
 1017 Intermediate CH₃ CHOO, *Science*, 340, 177–180, <https://doi.org/10.1126/science.1234689>, 2013.
- 1018 Vereecken, L., Harder, H., and Novelli, A.: The reaction of Criegee intermediates with NO, RO₂, and SO₂, and their fate in
 1019 the atmosphere, *Phys. Chem. Chem. Phys.*, 14, 14682, <https://doi.org/10.1039/c2cp42300f>, 2012.

- 1020 Vereecken, L., Novelli, A., and Taraborrelli, D.: Unimolecular decay strongly limits the atmospheric impact of Criegee
1021 intermediates, *Phys. Chem. Chem. Phys.*, 19, 31599–31612, <https://doi.org/10.1039/C7CP05541B>, 2017.
- 1022 Wang, Z., Nie, W., Liu, Y., Yang, L., Liu, C., Zhang, Y., Yan, C., Zha, Q., Ge, D., Qi, X., Zhou, C., Chen, L., Zhou, X.,
1023 Wang, L., Huang, D. D., Chi, X., and Ding, A.: Gaseous Sulfuric Acid Contributions to Sulfate Formation in Urban
1024 Atmosphere, *Environ. Sci. Technol. Lett.*, 12, 189–195, <https://doi.org/10.1021/acs.estlett.4c01028>, 2025.
- 1025 WHO: WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur
1026 Dioxide and Carbon Monoxide, 1st ed., World Health Organization, Geneva, 1 pp., 2021.
- 1027 Yang, L., Nie, W., Liu, Y., Xu, Z., Xiao, M., Qi, X., Li, Y., Wang, R., Zou, J., Paasonen, P., Yan, C., Xu, Z., Wang, J., Zhou,
1028 C., Yuan, J., Sun, J., Chi, X., Kerminen, V.-M., Kulmala, M., and Ding, A.: Toward Building a Physical Proxy for Gas-
1029 Phase Sulfuric Acid Concentration Based on Its Budget Analysis in Polluted Yangtze River Delta, East China, *Environ. Sci.*
1030 *Technol.*, 55, 6665–6676, <https://doi.org/10.1021/acs.est.1c00738>, 2021.
- 1031 Yao, L., Garmash, O., Bianchi, F., Zheng, J., Yan, C., Kontkanen, J., Junninen, H., Mazon, S. B., Ehn, M., Paasonen, P.,
1032 Sipilä, M., Wang, M., Wang, X., Xiao, S., Chen, H., Lu, Y., Zhang, B., Wang, D., Fu, Q., Geng, F., Li, L., Wang, H., Qiao,
1033 L., Yang, X., Chen, J., Kerminen, V.-M., Petäjä, T., Worsnop, D. R., Kulmala, M., and Wang, L.: Atmospheric new particle
1034 formation from sulfuric acid and amines in a Chinese megacity, *Science*, 361, 278–281,
1035 <https://doi.org/10.1126/science.aao4839>, 2018.
- 1036 Ye, C., Lu, K., Song, H., Mu, Y., Chen, J., and Zhang, Y.: A critical review of sulfate aerosol formation mechanisms during
1037 winter polluted periods, *Journal of Environmental Sciences*, 123, 387–399, <https://doi.org/10.1016/j.jes.2022.07.011>, 2023a.
- 1038 Ye, C., Lu, K., Song, H., Mu, Y., Chen, J., and Zhang, Y.: A critical review of sulfate aerosol formation mechanisms during
1039 winter polluted periods, *Journal of Environmental Sciences*, 123, 387–399, <https://doi.org/10.1016/j.jes.2022.07.011>, 2023b.
- 1040 Zhao, Q., Bi, J., Liu, Q., Ling, Z., Shen, G., Chen, F., Qiao, Y., Li, C., and Ma, Z.: Sources of volatile organic compounds
1041 and policy implications for regional ozone pollution control in an urban location of Nanjing, East China, *Atmos. Chem.*
1042 *Phys.*, 20, 3905–3919, <https://doi.org/10.5194/acp-20-3905-2020>, 2020.
- 1043 Zhu, T., Tang, M., Gao, M., Bi, X., Cao, J., Che, H., Chen, J., Ding, A., Fu, P., Gao, J., Gao, Y., Ge, M., Ge, X., Han, Z., He,
1044 H., Huang, R.-J., Huang, X., Liao, H., Liu, C., Liu, H., Liu, J., Liu, S. C., Lu, K., Ma, Q., Nie, W., Shao, M., Song, Y., Sun,
1045 Y., Tang, X., Wang, T., Wang, T., Wang, W., Wang, X., Wang, Z., Yin, Y., Zhang, Q., Zhang, W., Zhang, Y., Zhang, Y.,
1046 Zhao, Y., Zheng, M., Zhu, B., and Zhu, J.: Recent Progress in Atmospheric Chemistry Research in China: Establishing a
1047 Theoretical Framework for the “Air Pollution Complex,” *Adv. Atmos. Sci.*, 40, 1339–1361, [https://doi.org/10.1007/s00376-](https://doi.org/10.1007/s00376-023-2379-0)
1048 023-2379-0, 2023.
- 1049