

Author Comment on egusphere-2025-6276

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

Authors: Yuhuan Zhu et al.

Manuscript No.: egusphere-2025-6276

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

Dear Editor and Referees,

We sincerely thank the Editor and the Referees for their careful evaluation of our manuscript and for their constructive comments and suggestions. We especially appreciate the comments concerning the updates to the MCM v3.3.1 mechanism, particularly the selection and justification of reaction rate coefficients and kinetic parameters related to stabilized Criegee intermediates (sCIs), as well as the suggestions for improving the interpretation and presentation of the machine learning results. These comments have been very helpful in guiding us to further refine the mechanism updates, improve the clarity and interpretability of the machine learning results, and strengthen the overall scientific rigor of the manuscript. Specifically, we further revised the sCI-related gas-phase mechanism, expanded the AtChem-MCM simulations, reconstructed the XGBoost-based diagnostic framework, and reinterpreted the relative roles of sCI+SO₂ pathways in H₂SO₄ formation. These revisions have changed several quantitative results compared with the discussion manuscript, but the main qualitative conclusion remains unchanged: the revised mechanism still supports an important, regime-dependent role of sCI chemistry in atmospheric H₂SO₄ formation.

We have carefully considered all comments and revised the manuscript accordingly. Below, we provide a point-by-point response to each comment. For clarity, the Referees' comments are shown in black, our responses are shown in blue, and the corresponding revisions to the manuscript are summarized in red. Because these revisions required substantial new simulations and a near-complete rewriting of several sections, we are finalizing the revised manuscript and will submit it together with a complete "final author reply to the editor" as soon as permitted by the editor. In this public author comment, we provide detailed responses to the reviewers' major concerns and summarize the main changes and new results that will be incorporated into the revised manuscript.

Note: As the revised manuscript is still being finalized, the section, figure, table, and line numbers referred to in this public author comment should be regarded as provisional.

Reviewer #1

General Comment:

The authors present the results of a study to investigate the significance of stabilized Criegee intermediate (sCI) chemistry for the production of sulfuric acid in the atmosphere using model calculations. While investigation of the role of such chemistry does warrant attention, there are a number of significant flaws in the study that should

be addressed before publication can be considered.

In general, the paper is poorly written and details of the aims, methods, and results are unclear. The authors acknowledge the use of ChatGPT “to proofread and refine the English expression of the manuscript”, but the result is a manuscript that lacks specific detail relating to the study undertaken. The aims and use of the machine learning method and the analysis of the data needs to be explained much more clearly, with the descriptions specifically linked to the study in question. The descriptions given are too generic to be useful. It is not clear what analysis has been performed and what is being assessed.

The manuscript refers to uncertainties in the mechanisms used in models, Examples or details of the uncertainties should be given. Do the main uncertainties relate to rate coefficients, product yields, or missing reactions in the mechanisms? Do the results of the study help to clarify what we need to know better in order to better understand the production of sulfuric acid in the atmosphere?

Importantly, the study described in the manuscript is an incomplete assessment of sCI chemistry. The authors compare the results from model calculations using the MCM v3.3.1 with those from model calculations using an updated mechanism with rate coefficient recommendations taken from Cox et al. which was published in 2020. However, there have been a number of studies since the Cox et al. recommendation that should be considered. The authors note the potential impact of temperature, but all calculations have been performed using rate coefficients obtained at ~298 K. Measurements of the temperature and pressure dependence of sCI reactions with SO₂ and water have been reported in the literature since 2020 and these should be considered in the study to enable a more complete assessment. The study also neglects Z-isomers for sCIs where E/Z isomers are possible, and studies have shown that sCI reactivity is very different for E and Z isomers.

In addition, the study assumes that all sCI reactions with SO₂ produce SO₃ and thus lead to production of sulfuric acid. The reaction of the sCI CH₂OO with SO₂ has been shown to produce SO₃, but this is not the case for other sCIs at atmospheric pressure and there is evidence that larger sCIs may produce secondary ozonides at atmospheric pressure rather than SO₃. The impact of the assumption that all sCI reactions with SO₂ produce SO₃ and lead to sulfuric acid production should be investigated.

Author Response:

We sincerely thank the reviewer for the thorough and highly constructive evaluation of our manuscript. The comments have substantially helped us improve the clarity, rigor, and completeness of our work. Below we provide a detailed, point-by-point response organized into two parts: (1) Reply to Comments on Manuscript Writing, and (2) Reply to Comments on Mechanism Revision.

(1) Reply to Comments on Manuscript writing:

We sincerely apologize for the lack of clarity in the original manuscript. We have thoroughly revised the manuscript to address the issues of unclear research objectives, vague methodology, and insufficient logical coherence among the three components of the study.

Regarding research objectives and logical framework:

Since the laboratory discovery that sCIs can react with SO₂ in the gas phase, extensive experimental and theoretical studies have been conducted to elucidate the relevant reaction pathways, and field observations together with atmospheric modeling studies have confirmed and quantified the important contribution of sCIs to gas-phase SO₂ oxidation. From the perspective of atmospheric pollution science, “the oxidation of SO₂ by sCIs to produce H₂SO₄” can be framed as a problem of “secondary pollutant formation governed by atmospheric oxidative capacity”. The current consensus for addressing such secondary pollution issues is to establish quantitative relationships between precursors and secondary pollutants, and to elucidate the roles of key intermediate products (e.g., free radicals) within these relationships, thereby enabling a comprehensive understanding of pollution formation from both mechanistic and control perspectives. Existing studies on the contribution of sCIs to atmospheric H₂SO₄ or SO₄²⁻ aerosol formation have primarily focused on quantifying their contributions in specific regions and identifying key precursor alkene species through their production pathways. While these studies are highly valuable for characterizing sCI contributions and identifying the dominant alkene species in specific locations and time periods, their conclusions are not necessarily transferable to other regions, due to the site-specific nature of environmental conditions and precursor compositions. Consequently, a systematic understanding of the key drivers governing sCI contributions remains lacking. Furthermore, the production and loss processes of OH and sCIs are deeply coupled. This coupling implies that when evaluating the drivers controlling the sCI+SO₂ pathway contribution, one cannot ignore the concomitant effects on the OH+SO₂ pathway and the overall SO₂ oxidation rate. In particular, how the dominant controlling factors and the direction of their influence change as the gas-phase SO₂ oxidation regime transitions from being strictly OH-dominated to being co-driven by OH and sCIs remains an unresolved question.

Building on these motivations, our study presents a systematic assessment of the role of sCIs in atmospheric SO₂ gas-phase oxidation, structured around three progressive components:

Section 3.1 – Diagnosing key drivers of sCI-mediated SO₂ oxidation: We developed an interpretable machine-learning framework (XGBoost-SHAP) to elucidate the dependence of the sCI+SO₂ reaction rate on key controlling variables, including O₃, specific alkene species, and major competing sinks (H₂O, (H₂O)₂, and NO_x). Particular attention was given to quantifying the distinct effects of different alkene precursors and to examining how updates to the CI mechanism modify their relative importance.

Section 3.2 – Surrogate modeling of sCI contributions and regime analysis: Building on the diagnostic results in Section 3.1, we designed simulation scenarios covering perturbations to all control factors. An atmospheric box model coupled with the updated mechanism (MCM v3.3.1g) was then used to generate a comprehensive dataset of sCI contributions under diverse environmental conditions, from which an XGBoost-based surrogate model was constructed. Sobol sensitivity analysis and Partial Dependence Plots (PDPs) were integrated to quantify the magnitude and direction of each factor's influence. By stratifying the dataset based on the median sCI contribution, we further

examined how the sensitivity of the total SO₂ oxidation rate differs between high- and low-sCI regimes, and identified the key reaction pathways responsible for these contrasts through mechanistic budget analysis.

Section 3.3 – Validation against ambient observations: The predictive performance of the surrogate model was evaluated by comparing its predictions with box model simulations for an actual ambient case. Additionally, direct XGBoost modeling of hourly observational data was performed to evaluate the sensitivity of SO₄²⁻ to various factors under different sCI contribution levels, thereby assessing whether the regime-dependent sensitivity relationships predicted by our model framework are consistent with ambient observations.

Regarding the Methods section:

In the revised manuscript, we have substantially expanded and reorganized Section 2 so that each method is explicitly linked to the corresponding scientific question addressed in this study.

In Section 2.1, “Updating the sCI gas-phase chemical mechanism,” we have added a more detailed description of the mechanism revision. In particular, we now describe how the concentration of water dimers was calculated using a temperature-dependent equilibrium constant and the square of the water monomer concentration, rather than prescribing it as a fixed fraction of water vapor. We also provide more detailed information on the updated kinetic parameters and branching treatments used in the revised mechanism, including the implementation of temperature-dependent rate coefficients for relevant sCI+H₂O/(H₂O)₂ reactions.

In Section 2.2, “Model setup and observation data,” we have expanded the description of the AtChem simulations and chemical diagnostics. The revised text now distinguishes the different types of box-model simulations used in the study, explains how the rates of SO₂ oxidation by sCIs and OH were calculated, and clarifies how the observational dataset was used in different parts of the analysis. For the observation data, we no longer simply list the measured variables; instead, we now explain their specific roles in constraining the box model, constructing observationally based machine-learning models, and evaluating whether the sensitivity relationships inferred from the box-model surrogate are reflected in the real atmosphere.

We have also added a new Section 2.3, “Machine learning surrogate and interpretation framework,” to clearly describe the machine-learning methods used in the study. This section explains the purpose of each model, the target variable, the input features, the training dataset, and the interpretation method. In this way, the machine-learning analysis is now directly connected to the scientific objectives of the study, rather than being presented as a generic data-driven method.

Compared with the original manuscript, we have also substantially increased the size and representativeness of the modeling datasets. For Section 3.1, the number of box-model simulations used to train the XGBoost-SHAP model was increased from 768 to 3898. For Section 3.2, the number of simulation cases was increased from 243 to 1352. In this section, we also added two additional explanatory variables: ARI, representing the alkene-specific sCI production potential, and RH, representing the influence of atmospheric water vapor and water dimers on sCI loss. These additions improve the

connection between Sections 3.1 and 3.2 and allow the surrogate model to better capture both sCI production and competitive loss processes. In addition, we replaced the ANOVA-based interpretation with Sobol sensitivity analysis, which is more appropriate for quantifying the contribution of individual variables and their interactions to the variance of the model output. In addition, we added a new analysis in Section 3.2 to examine whether the sensitivity of the total gas-phase SO₂ oxidation rate differs between conditions with relatively high and low sCIs contributions. Specifically, the simulation dataset was divided into high- and low- μ_{sCIs} subsets according to the median value of μ_{sCIs} . We then evaluated and compared the sensitivity relationships between the explanatory variables and the total SO₂ oxidation rate in these two subsets. This analysis allows us to assess how the relative importance of the sCI pathway modifies the response of overall SO₂ oxidation to changes in O₃, alkenes, NO_x, RH, and other environmental variables. Therefore, the revised Section 3.2 now more directly addresses the question of how the transition from an OH-dominated regime to a regime co-influenced by OH and sCIs changes the controlling factors of sulfuric acid production. These additions improve the connection between Sections 3.2 and 3.3. For Section 3.3, we replaced the Random Forest model used in the original manuscript with XGBoost to maintain methodological consistency across the study. The observational period used to evaluate the sensitivity relationships was also extended from 18 days, 1-18 July 2021, to 45 days, 1 June-15 July 2021, thereby improving the robustness of the observational analysis.

Regarding the Results and Discussion:

we have substantially revised Section 3 because the revised treatment of key kinetic processes changed the box-model simulations. Specifically, the rate coefficients for relevant sCI reactions with SO₂ and H₂O/(H₂O)₂ were updated from fixed 298 K values to temperature-dependent expressions where recommended, and the water dimer concentration was recalculated using the temperature-dependent equilibrium constant and the square of the water monomer concentration. Together with the increased number of simulations, the expanded feature set, the longer observational period, and the replacement of the interpretation method in Section 3.2, these changes led to updated results throughout the manuscript.

In revising Section 3, we have also strengthened the interpretation of the results. Rather than only reporting numerical rankings from SHAP values, Sobol indices, or PDPs, the revised text now focuses on the physical and chemical meaning behind these diagnostics. In the original manuscript, we discussed the impact of the mechanism revision in both Sections 3.1 and 3.3 and emphasized these comparisons in the Abstract. Our original intention was to demonstrate the necessity and reasonableness of the mechanism revision by showing the magnitude of the resulting changes. However, this presentation blurred the focus of the study and may have given the impression that the manuscript was primarily an assessment of the completeness of sCI chemistry. In the revised manuscript, we have therefore refocused the discussion. The comparison between the original and updated mechanisms is now retained only in Section 3.1, where it serves a specific and focused purpose. This analysis is used to quantify how the relative importance of individual alkene precursors to $L_{\text{SO}_2, \text{sCIs}}$ changes after

updating the sCI mechanism, and to identify which revised kinetic processes are mainly responsible for these changes. The results indicate that one of the dominant sources of uncertainty in MCM v3.3.1 lies in the rate coefficients that govern the competition between the sCI + SO₂ pathway and the major water sinks (H₂O and (H₂O)₂). This revised treatment keeps the mechanism comparison directly connected to the main research question: which chemical processes control the ability of sCIs to oxidize SO₂ and contribute to sulfuric acid formation? It also provides practical guidance for future mechanism development, particularly for deciding which CI kinetic processes should be prioritized when incorporating CI chemistry into reduced chemical mechanisms used in three-dimensional atmospheric models. At the same time, we have removed unnecessary mechanism-comparison statements from Section 3.3 and the Abstract, so that the later sections focus more clearly on the environmental conditions under which sCIs can effectively compete with OH and on whether the inferred sensitivity relationships are supported by ambient observations.

(2) Reply to Comments on Mechanism Revision:

Selection of Basis for Mechanism Revision:

We appreciate the reviewer's concern and would like to clarify our rationale for selecting (Cox et al., 2020) as the primary basis for mechanism revision. Our study is not intended as a comprehensive re-evaluation of sCI chemistry per se, but rather investigates the importance of sCIs in gas-phase SO₂ oxidation and identifies the key controlling factors. For this purpose, the mechanism update serves as a prerequisite to minimize the propagation of kinetic uncertainties into our conclusions. Our selection of the IUPAC Task Group evaluation (Cox et al., 2020) as the revision basis was guided by the following considerations:

- **Scientific consensus and reliability:** The IUPAC evaluation critically assesses and synthesizes multiple experimental and theoretical studies to provide recommended kinetic parameters that have undergone extensive peer scrutiny. The description in the original manuscript was imprecise: we stated that the mechanism was updated using “the latest research findings”. This has now been accurately rephrased as “the most recent systematically evaluated kinetic data by IUPAC”.
- **Benchmarking against MCM v3.3.1:** The MCM v3.3.1 is one of the mainstream mechanisms widely coupled in atmospheric models. Rate coefficients for reactions of O₃ with alkenes within MCM v3.3.1 were previously reviewed by (Saunders et al., 2003), (Jenkin et al., 1997) and (Jenkin et al., 2015) with recommendations. Adopting the 2020 IUPAC evaluation — which represents the same category of authoritative, community-vetted assessment — provides robust basis for updating this mechanism while maintaining the credibility that the international modeling community widely accepts.
- **Practical requirements for mechanism implementation:** Mechanisms intended for integration into computational models require precise functional expressions. For example, if temperature dependence is to be considered, Arrhenius expressions with well-defined parameters are needed. The IUPAC evaluation provides such expressions where available. In contrast, some frontier studies, while scientifically valuable, do not report explicit kinetic parameters and therefore cannot be directly

integrated into the MCM v3.3.1 mechanism.

Compared with the original MCM v3.3.1 parameters, the IUPAC-evaluated values reveal order-of-magnitude differences in sCI bimolecular reaction rate coefficients, and the original mechanism entirely omits the unimolecular decomposition of sCIs, reactions with (H₂O)₂, and the distinct reactivities of different sCI stereoisomers. These fundamental deficiencies fully justify the mechanism revision. Because the IUPAC-evaluated data sufficiently addresses these critical gaps, we opted for a more conservative approach: rather than incorporating newer studies published post-2020, we strictly adopted these comprehensively evaluated recommendations.

Temperature and Pressure Dependence:

We acknowledge that the original manuscript insufficiently addressed the temperature dependence of sCI reactions. In the revised manuscript, we have made the following changes:

- Temperature dependence of $k_{(sCI+H_2O)}$ and $k_{(sCI+(H_2O)_2)}$: The IUPAC evaluation provides both 298 K preferred values and temperature-dependent expressions for the reactions of C₁-C₄ sCIs with H₂O and (H₂O)₂. In the original manuscript, we adopted the 298 K values for consistency with the sCI yields and other bimolecular rate coefficients, which are predominantly derived from direct kinetic studies at 298 K. However, we recognize that the reactions of sCIs with H₂O/(H₂O)₂ exhibit significant negative temperature dependence (e.g., the CH₂OO + (H₂O)₂ reaction), whereas the sCIs+SO₂ reactions show weaker temperature sensitivity. This differential temperature dependence means that using only 298 K values does not adequately capture the competitive balance between these pathways. In the revised mechanism (MCM v3.3.1g), we have updated all $k_{(sCI+H_2O)}$ and $k_{(sCI+(H_2O)_2)}$ for C₁-C₄ sCIs to their recommended temperature-dependent expressions $k(T)$.
- Pressure dependence of $k_{(sCI+SO_2)}$: The IUPAC evaluation indicates that for C₁-C₄ sCIs, there is no significant pressure dependence for reactions with SO₂ within the range relevant to atmospheric applications.

E/Z Isomer Consideration:

We sincerely apologize that our treatment of E/Z isomers was not clearly communicated in the original manuscript. In fact, our mechanism update accounts for the distinct reactivities of E- and Z-isomers, but this was not adequately described. We have added explicit descriptions in Section 2.1 and the Supplementary Material (Table S1) of the revised manuscript. The key points are as follows:

Among the sCIs covered in our revision, CH₃CHOO (from propene, cis/trans-but-2-ene ozonolysis), C₂H₅CHOO (from but-1-ene ozonolysis) and the C₄ intermediates from isoprene ozonolysis exist as E- and Z-conformers. E- and Z-conformers exhibit markedly different reactivities, particularly in their propensity for unimolecular decomposition versus bimolecular reactions:

- For CH₃CHOO: Z-[CH₃CHOO]* and Z-CH₃CHOO undergo extremely rapid unimolecular decomposition via the well-established 1,4 H-shift isomerization mechanism, producing vinyl hydroperoxide intermediates that decompose to OH. This rapid decomposition means that bimolecular reactions of Z-CH₃CHOO with SO₂, H₂O, and other species are kinetically uncompetitive under atmospheric

conditions. In our revised mechanism, the unimolecular decomposition of Z-CH₃CHOO is combined with the prompt decomposition of excited CIs(Z-[CH₃CHOO]*), with the branching ratio determined by the recommended OH yield. The sCI yield and bimolecular reaction rate coefficients listed for CH₃CHOO in Tables 1 and 2 specifically refer to E-CH₃CHOO, which is the conformer that participates in bimolecular chemistry with SO₂ and H₂O/(H₂O)₂. A small contribution from E-[CH₃CHOO]* unimolecular decomposition is also accounted for and deducted from the total sCI yield.

- For C₂H₅CHOO (from but-1-ene): Similarly, Z-C₂H₅CHOO is assumed to undergo primarily unimolecular decomposition, with its contribution reflected in the OH yield, and the reported sCI yield represents predominantly the E-conformer.
- For C₄ sCIs from isoprene: Z-(CH=CH₂)(CH₃)COO and Z-(C(CH₃)=CH₂)CHOO undergo rapid 1,5 ring-closure reactions with rate coefficients of ~2,800 s⁻¹ and ~14,000 s⁻¹, respectively. Their unimolecular decomposition processes are similarly accounted for in the OH yield, and their branching ratios are deducted from the total sCI yields. Only the E-conformers are treated as participating in bimolecular chemistry.

In summary: The sCI yields(CH₃CHOO, C₂H₅CHOO,and C₄ sCIs) reported in Table 2 of our revised manuscript have been adjusted to account for the rapid decomposition of Z-isomers, and thus represent only the E-isomers that are available for bimolecular reactions. This is now stated explicitly in the table caption and in Section 2.1: “In MCM v3.3.1g, the species CH₃CHOO and C₂H₅CHOO specifically denote the E-isomers. The unimolecular decomposition of Z-isomer sCIs is integrated with the prompt decomposition of excited CIs; consequently, their branching ratios are deducted from the total sCI yields.”

SO₃ Yield from sCI + SO₂ Reactions:

We thank the reviewer for raising this important point. In the revised manuscript, we have added an explicit discussion of this assumption and its implications.

Following the IUPAC evaluation, the reaction of small sCIs (C₁-C₄) with SO₂ proceeds via the barrierless formation of a chemically activated secondary ozonide (SOZ) intermediate. For CH₂OO+SO₂, the dominant product is SO₃ (+ HCHO), which is well-established by experimental evidence. There is evidence that larger sCIs may favor secondary ozonide formation at atmospheric pressure. However, the product branching ratios for sCIs + SO₂ reactions remain poorly constrained experimentally. In our mechanism, we follow the IUPAC recommendation that, in the absence of quantitative experimental branching ratio data for the primary (SO₃-forming) and secondary (SOZ-forming) channels, the sCI + SO₂ reaction is treated as predominantly producing SO₃ (and the corresponding carbonyl compound).

We now explicitly acknowledge this as an assumption in Section 2.1 of the revised manuscript and discuss its implications: By assuming 100% SO₃ yield, our calculated sCI contributions to H₂SO₄ production represent an upper bound. If a fraction of sCI + SO₂ reactions produces SOZs instead of SO₃, the actual contribution to H₂SO₄ would be proportionally reduced. It should be noted, however, that the SO₃ yield essentially acts as a linear scaling factor on our results: while it may overestimate the absolute

H₂SO₄ production rate via the sCI pathway as well as the sCI fractional contribution, this assumption does not alter the functional relationships between the controlling factors and the sCI contribution, nor does it affect our identification of the conditions under which their contribution becomes significant. Consequently, these core insights remain qualitatively robust despite uncertainties in the exact SO₃ branching ratio.

Specific Comments

Comment 1:

line 15: sCIs are not free radicals and ‘oxidated’ should be changed to ‘capable of oxidising’.

Response 1:

We thank the reviewer for this clarification and for correcting our terminology. sCIs are reactive zwitterionic intermediates (R₁R₂C=O⁺-O⁻), rather than traditional free radicals. We have replaced the term ‘oxidated’ with ‘capable of oxidising’ to ensure accurate expression, and similar inaccuracies have been carefully reviewed and corrected throughout the manuscript.

Changes in Manuscript:

Modified version: “...stabilized Criegee intermediates (sCIs) are recognized as important atmospheric intermediates capable of oxidising sulfur dioxide (SO₂)...”

Comment 2

Line 17: No apostrophe in sCIs.

Response 2

We thank the reviewer for catching this error. The apostrophe has been removed from all instances of "sCIs" throughout the manuscript.

Changes in Manuscript:

The apostrophe has been removed from all occurrences of “sCIs” throughout the revised manuscript.

Comment 3

Line 18: Clarify the meaning of the factor reported.

Response 3

We thank the reviewer for this helpful comment. In the original manuscript, the reported factor referred to the ratio of the mean absolute SHAP values, mean(|SHAP|), obtained from the XGBoost models trained with MCM v3.3.1g and MCM v3.3.1, respectively. Thus, this 1.97- to 10.75-fold increase demonstrates that the updated mechanism significantly amplifies the importance of these precursors in predicting L_{SO₂,sCIs}. Because the Abstract has been substantially rewritten, this potentially confusing statement has been removed from the Abstract.

Changes in Manuscript:

This sentence has been removed from the Abstract.

Comment 4

Line 32: ‘Fine particles ... are ...’ rather than ‘fine particles ... is’, and ‘its’ to ‘their’.

Response 4

We thank the reviewer for pointing out these grammatical errors. We have thoroughly checked the manuscript for similar expression inaccuracies.

Changes in Manuscript:

Corrected to "Fine particles ... are ..." and "their".

Comment 5

Line 37: Which of the formation routes described is the primary route?

Response 5

We thank the reviewer for raising this important question. The primary formation route of sulfate depends on atmospheric conditions. During severe haze events, high humidity and aerosol loadings strongly promote aqueous-phase and heterogeneous SO₂ oxidation, making these multiphase processes the dominant pathways for secondary sulfate formation. As the atmosphere becomes progressively cleaner, conditions favorable for aqueous-phase and heterogeneous sulfate formation weaken, and the relative importance of the gas-phase pathway (i.e., oxidation of SO₂ by OH and sCIs) is expected to increase significantly. Furthermore, the gas-phase pathway remains fundamentally important regardless of the dominant sulfate formation pathway because its end product, sulfuric acid (H₂SO₄), is a key precursor for new particle formation (NPF). NPF governs aerosol number concentrations and is frequently observed even in highly polluted urban environments with strong condensation sinks. We have clarified this dependence on atmospheric conditions and the growing importance of the gas-phase route in the revised manuscript.

Changes in Manuscript:

The relevant paragraph in the Introduction has been revised

Comment 6

Line 45: The reference to ‘anon’ should be changed to reference the WHO

Response 6

We thank the reviewer for pointing out these citation errors. We have corrected. We thank the reviewer for pointing out this citation issue. The reference to “anon” has been corrected to the appropriate (WHO, 2021) citation in the revised manuscript and similar citation issues have been carefully checked and corrected throughout the manuscript.

Changes in Manuscript:

The citation has been corrected to (WHO, 2021), and the reference list has been updated accordingly.

Comment 7

Line 50: The reference to ‘anon’ should be changed to Mauldin et al. 2012

Response 7

We thank the reviewer for pointing out these citation errors. We have corrected. We thank the reviewer for pointing out this citation issue. The reference to “anon” has been corrected to Mauldin et al. and similar citation issues have been carefully checked and corrected throughout the manuscript.

Changes in Manuscript:

The citation has been corrected to (Mauldin et al., 2012), and the reference list has been updated accordingly.

Comment 8

Line 62: There are earlier references to the production of Criegee intermediates.

Response 8

We sincerely thank the reviewer for pointing out this historical detail. We have updated the citation to include the pioneering work by Criegee (1949), which first proposed the zwitterionic mechanism and the production of Criegee intermediates. The reference list has also been updated accordingly.

Changes in Manuscript:

The Introduction has been revised to include the suggested references.

Comment 9

Line 69: Clarify the description of gas phase sulfate ions.

Response 9

We thank the reviewer for catching this imprecise expression. We intended to describe the gas-phase formation pathway of sulfate (i.e., the gas-phase oxidation of SO₂ to H₂SO₄, which subsequently contributes to particulate sulfate), rather than implying the existence of SO₄²⁻ ions in the gas phase.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 10

Line 70: Ozone is not formed through direct reactions between VOCs and NO_x. The statement should be clarified.

Response 10

We thank the reviewer for catching this imprecise expression. O₃ is formed through the combination of O₂ and O generated from NO₂ photolysis. This sequence is continuously driven by the complex photochemical cycles of its precursors, VOCs and NO_x.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 11

Line 101: Subscript in O₃.

Response 11

We thank the reviewer for pointing out this formatting issue. We have corrected the O₃ notation and have carefully checked and revised the chemical notation throughout the manuscript.

Changes in Manuscript:

All instances have been checked and corrected.

Comment 12

Line 106: Clarify the meaning of ‘bimolecular water’. Should this refer to water dimers, (H₂O)₂? If the statement is intended to refer to water dimers the following statement regarding the relative concentration to H₂O (monomers) is incorrect. How were water dimer concentrations calculated?

Response 12

We thank the reviewer for pointing out these inaccuracies. First, we clarify that the term "bimolecular water" was indeed intended to refer to the water dimer, (H₂O)₂. We have corrected this non-standard terminology throughout the manuscript. Second, we acknowledge the writing error regarding the concentration ratio in the original manuscript. The intended value was 10⁻⁴ (not 10⁴), which reflects the typical atmospheric relative abundance of water dimers to monomers (10³~10⁴) (Tretyakov et al., 2014).

However, addressing the valid question regarding the actual calculation of water dimer concentrations, treating this ratio as a constant introduces significant uncertainty. The actual dimer concentration scales quadratically with water monomer abundance and is highly temperature-dependent. To accurately represent the kinetics of the sCIs + (H₂O)₂ reaction in the box model, we employed a pre-equilibrium approximation. The dimer is assumed to be in steady-state equilibrium with the monomer. We rigorously parameterized the temperature-dependent equilibrium constant (K_{eq}^D) using high-accuracy thermochemical data from the Active Thermochemical Tables (ATcT) (Ruscic, 2013), allowing us to calculate an apparent third-order rate constant ($k_{eff} = k_{sCIs+(H_2O)_2} \cdot K_{eq}^D$). The complete mathematical derivation and parameter details have now been explicitly added to Section 2.1.

Changes in Manuscript:

The term “bimolecular water” has been replaced by "water dimer, (H₂O)₂”.

Modified version:

To accurately represent sCIs+(H₂O)₂ in MCM v3.3.1 within the box model, a pre-equilibrium approximation was employed. Given the rapid exchange between water monomers and dimers, the concentration of (H₂O)₂ is assumed to be in steady-state equilibrium with the water monomer:

$$K_{eq}^D = \frac{[(H_2O)_2]}{[H_2O]^2}$$

where K_{eq}^D is the temperature-dependent equilibrium constant (Scribano et al., 2006). The reaction of sCIs with water dimers was parameterised using an apparent third-order rate constant (Lade et al., 2024b), k_{eff} , such that the total reaction rate (r) is expressed as:

$$r = k_b [sCI] [(H_2O)_2] = k_b \times K_{eq}^D [sCI] [H_2O]^2 = k_{eff} [sCI] [H_2O]^2$$

The apparent rate constant is defined as the product of the bimolecular rate constant for the sCIs+(H₂O)₂ reaction (k_b) and the equilibrium constant (K_{eq}^D). To ensure the highest thermodynamic accuracy in calculating K_{eq}^D , thermochemical data were retrieved from the Active Thermochemical Tables (ATcT). Standard Gibbs free energies of formation

$(\Delta_r G_T^\circ)$ for both the water monomer and the water dimer were extracted from Table 1 and Table 3 of (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 covers the typical conditions of the troposphere. The temperature dependence of K_{eq}^D was then obtained via a linear regression of K_{eq}^D against $\frac{1}{T}$, resulting in the following Arrhenius-type parameterisation used in the model:

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

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

Comment 13

Line 110: Some of the rate coefficients given are for unimolecular processes.

Response 13

We thank the reviewer for pointing out this inaccuracy. To accurately reflect that the table includes both unimolecular and bimolecular reactions, the title has been revised to refer generally to the loss processes of sCIs.

Changes in Manuscript:

The title of Table 1 has been revised to:

"Table 1. Comparison of rate coefficients for sCIs loss processes before and after the update."

Comment 14

Line 112: No rate coefficients are given in the table.

Response 14

We thank the reviewer for catching this omission and inaccuracy. We have added the missing rate coefficients for the alkene ozonolysis reactions ($\text{O}_3 + \text{alkenes}$) to Table 2. Accordingly, the table title has been revised to accurately reflect that it now presents both the kinetic rate coefficients and the corresponding OH/sCIs yields ($Y_{i,\text{OH}}$ and $Y_{i,\text{sCIs}}$).

Changes in Manuscript:

Tables 1 and 2 have been revised.

Comment 15

Line 130: 'Elements ... were determined ...' should be changed to 'elemental analysis ... was performed ...', but it's not clear that the elemental analysis is relevant.

Response 15

We thank the reviewer for pointing out the phrasing issue and for prompting us to clarify the relevance of these measurements. We have revised the text to "Elemental analysis was performed," as suggested. To avoid confusion, we have removed references to elements that were not used in the model, except for iron (Fe). Fe was retained because it is an important variable in this study and was included as a key input feature in the XGBoost model. As discussed in Section 3.3, the SHAP values of Fe were used as an indicator to assess the role of transition-metal-catalyzed aqueous-phase reactions in

sulfate formation. This has now been clarified in the revised methodology section.

Changes in Manuscript:

Modified version: “Hourly observations from the Wuhai City Atmospheric Environment Super Monitoring Station were used for the observation-constrained simulations and observation-based machine learning analysis. The variables used in this study include trace gases (PM_{2.5}, CO, SO₂, NO₂, O₃, VOCs), meteorological parameters (WS, T, P, RH), photolysis frequencies, inorganic ions including SO₄²⁻ and NO₃⁻, and Fe as an indicator of transition-metal-related aqueous-phase processes.”. Variables measured at the station but not used in the present analysis have been removed from the revised description to avoid ambiguity.

Comment 16

Line 186 (and elsewhere): Check the units given for rates, is ‘mole’ rather than ‘molecule’ correct?

Response 16

We thank the reviewer for catching this unit error. In the manuscript, the intended unit was molecule (e.g., molecule·cm⁻³·s⁻¹), but we incorrectly used “mole” as a shorthand for “molecule”. We have now systematically checked and corrected all corresponding units throughout the manuscript and figures.

Changes in Manuscript:

All units corrected throughout.

Comment 17

Line 200: Subscripts in CH₃CHOO and CH₂OO.

Response 17

We thank the reviewer for pointing out this formatting issue. We have corrected the CH₃CHOO/CH₂OO notation and have carefully checked and revised the chemical notation throughout the manuscript.

Changes in Manuscript:

All chemical formulae have been checked and corrected for proper subscripts.

Comment 18

Line 206 (and elsewhere): The text in the figures is too small to read.

Response 18

We thank the reviewer for pointing out this issue. We have increased the font size and improved the resolution of all figures in the manuscript to enhance readability.

Changes in Manuscript:

All figures revised with improved readability.

Comment 19

Line 237: Why were these values chosen?

Response 19

We thank the reviewer for this helpful comment. In the revised manuscript, the factor levels are no longer defined as fixed percentages (e.g., 10%, 100%, and 190% of the

baseline) for all variables. Instead, the range of each variable was determined separately based on its three-year hourly observational distribution in the Wuhai region. Specifically, the lower and upper bounds were selected with reference to the 5th and 95th percentiles, so that the design could cover a wide range of pollution conditions while avoiding unrealistic extreme scenarios. We have clarified this rationale in the revised manuscript and added the corresponding variable ranges in the figure showing the values at the normalized levels of 0, 0.5, and 1.

Changes in Manuscript:

Justification for the scenario design added in Section 2 or 3.2.

Comment 20

Line 321: Subscript in O₃.

Response 20

We thank the reviewer for pointing out this formatting issue. We have corrected the O₃ notation and have carefully checked and revised the chemical notation throughout the manuscript.

Changes in Manuscript:

The O₃ notation has been corrected throughout the revised manuscript.

Comment 21

Line 328: It would be helpful to provide a summary of the input data. As a minimum, the mean, standard deviation, and median values should be reported for each input parameter.

Response 21

We thank the reviewer for this helpful suggestion. We have added a summary table reporting the mean, median, and standard deviation of the major constrained variables during the simulation period (2021.06.01-2021.07.15). We also added a time-series figure to show the temporal evolution of the major constrained species and meteorological parameters.

Changes in Manuscript:

A summary table and a time-series figure have been added to the Supplement.

Reviewer #2

General Comment:

This study is relevant to the field of atmospheric chemistry and addresses an important topic: the contribution of the stabilized Criegee intermediates (sCI) reactions with SO₂ to the formation of H₂SO₄ in the atmosphere, which subsequently contributes to the production of sulfate aerosols. The paper provides new findings coupling updates of the Master Chemical Mechanism (MCM), a benchmark mechanism within the atmospheric chemistry community, and machine learning technique. I recommend publication after addressing the comments below.

The text lacks clarity at times and there is need to provide more detail all throughout the manuscript, particularly about the application of machine learning method. The

terms used in the description of the machine learning method should be clearly defined, given that the intended readership may not be familiar with this specialised terminology and the associated methodology. In general, the figure captions should be expanded to more clearly describe the plotted variables and symbols.

At the end of the Introduction the authors should state that they used field observations obtained at the Wuhai City Super Monitoring Station and explain why this site was chosen as representative for their study.

Author Response:

We sincerely thank the reviewer for the positive evaluation of the scientific relevance of our study and for the constructive suggestions regarding the clarity of the manuscript, the description of the machine learning methods, the figure captions, and the justification of the observational site. In the original manuscript, the role of machine learning and the connection between the different modelling steps were not explained with sufficient clarity. In the revised manuscript, we have therefore substantially rewritten the relevant methodological sections and revised the figure captions to make the analysis more transparent to readers who may not be familiar with machine learning terminology.

First, we have substantially revised and expanded Sect. 2.3, now entitled “**Machine learning surrogate and interpretation framework**”, to clarify why machine learning was used and how each machine learning analysis serves a specific research objective. In the revised text, we now explicitly distinguish three related but different applications of machine learning in this study. The first XGBoost model, used in Sect. 3.1, was trained on AtChem scenario simulations to emulate the absolute sCI-driven SO₂ oxidation rate, L_{SO_2,sCI_s} , and to diagnose how the updated MCM v3.3.1g mechanism changes the sensitivity of this rate to O₃, individual alkenes, relative humidity, and NO₂. The second XGBoost surrogate model, used in Sect. 3.2, was developed to predict μ_{sCI_s} , the fractional contribution of the sCI pathway to total gas-phase SO₂ oxidation, under a broad range of atmospheric conditions. This model was further used to examine how the sensitivity of the total gas-phase SO₂ oxidation rate differs between low- and high- μ_{sCI_s} regimes. The third machine learning analysis, used in Sect. 3.3, was based on ambient observations and was designed to examine whether the sensitivity relationships inferred from the AtChem-derived surrogate analysis are reflected in observed sulfate variability. This revision makes clear that machine learning was as an efficient surrogate and interpretation tool linked directly to the chemical questions addressed in each results section.

Second, we have added clearer definitions of the machine learning terminology used throughout the manuscript. In particular, we now define “features” as the input variables used by the model and “target variable” as the output variable to be predicted. We also clarify that the feature variables differ among the three analyses. In Sect. 3.1, the features are the prescribed initial values of O₃, individual alkenes, RH, and NO₂ used as inputs for AtChem, while the target variable is the AtChem-simulated L_{SO_2,sCI_s} . In Sect. 3.2, the features are normalized scenario descriptors, including NO_x, NO₂ fraction, O₃, VOCs, alkene fraction (alkene%), RH, and the Aggregated Reactivity Index(ARI),

while the target variables are diagnosed from the constrained AtChem simulations. In Sect. 3.3, the features are observed pollutant, meteorological, photolysis, and source-related variables, while the target variable is the observed sulfate concentration. We also define “hyperparameters” as model settings specified before training, such as learning rate, tree depth, number of trees, subsampling ratio, and regularization strength. The L1 and L2 penalties are now described as regularization terms used to reduce overfitting and improve model generalization.

Third, we have expanded the explanation of the interpretation methods, including SHAP, Sobol sensitivity analysis, and partial dependence plots. In the revised manuscript, SHAP values are defined as the contribution of each feature to an individual model prediction relative to the average model prediction. We now explicitly state that a positive SHAP value means that a feature increases the predicted target variable, whereas a negative SHAP value means that it suppresses the prediction. The mean absolute SHAP value, $\text{mean}(|\text{SHAP}|)$, is used as a measure of global feature importance. We also clarify that partial dependence plots show the marginal response of the predicted target variable to one or two selected features while averaging over the remaining features. Sobol sensitivity analysis is now described as a variance-based global sensitivity method, in which the first-order index S_1 represents the independent contribution of a feature and the total-order index S_T includes both the independent contribution and all interaction effects.

Fourth, we have revised the captions of the machine-learning-related figures to define the plotted variables, axes, color scales, and symbols more explicitly. For example, in the SHAP summary plots, the revised captions now state that the x-axis represents the SHAP value, which indicates the direction and magnitude of the feature’s influence on the predicted target variable, while the color scale represents the actual feature value from low to high.

Finally, following the reviewer’s suggestion, we have revised the end of the Introduction to state explicitly that field observations from the Wuhai City Atmospheric Environment Super Monitoring Station were used in this study. We have also added a justification for selecting this site in Sect. 2.2.2(Observation data). Wuhai is a semi-arid coal-chemical industrial city in Inner Mongolia, China, characterized by relatively high SO_2 emissions, abundant anthropogenic VOCs including alkene precursors, frequent O_3 pollution events, and comparatively dry atmospheric conditions. These characteristics make Wuhai particularly suitable for examining gas-phase SO_2 oxidation under conditions where both sCI precursors, such as alkenes and O_3 , and the sulfur precursor SO_2 are present at elevated levels. The semi-arid environment is also relevant because lower humidity can reduce the competitive loss of sCIs to H_2O and water dimers, thereby providing favorable conditions for assessing when sCIs may effectively compete with OH in SO_2 oxidation.

Overall, these revisions improve the clarity of the manuscript, define the machine-learning terminology more explicitly, better connect each modelling method to the corresponding scientific question, make the figure captions more informative, and provide a clearer rationale for the use of the Wuhai observations.

Specific Comments

Comment 1:

line 15: ‘...are recognised to be atmospheric intermediates responsible for the oxidation of sulfur dioxide’ instead of ‘...are recognised to be one of the free radicals oxidated sulfur dioxide’. It is well known that Criegee intermediates have a zwitterionic character and, thus referring to them as ‘intermediates’ rather than ‘free radicals’ is more appropriate.

Response 1:

We thank the reviewer for this clarification and for correcting our terminology. sCIs are reactive zwitterionic intermediates ($R_1R_2C=O^+-O^-$), rather than traditional free radicals. We have therefore revised the description accordingly throughout the manuscript. We have revised and have been carefully reviewed and corrected similar inaccuracies throughout the manuscript.

Changes in Manuscript:

The sentence has been revised in Abstract.

Comment 2:

line 15: It is not clear what the word ‘dominant’ means in here. OH radicals are not the most abundant radicals in the atmosphere. The authors should clarify that the intended meaning is that OH is typically the dominant oxidant of SO₂.

Response 2:

We thank the reviewer for this helpful comment. Here, we intended to indicate that OH is typically the dominant oxidant of SO₂, rather than the most abundant radical in the atmosphere.

Changes in Manuscript:

The sentence has been revised in Abstract.

Comment 3:

line 40: add the word ‘Earth’ before ‘surface’, i.e. ‘near the Earth surface’.

Response 3:

We thank the reviewer for this suggestion. In the revised manuscript, the Introduction has been substantially reorganized, and the original sentence containing this wording has been removed. Nevertheless, we have carefully checked the manuscript and corrected similar expressions.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 4:

line 50: The reference ‘Anon, n.d.’ (line 423 in the reference list) should include the source of the pdf document, such as a webpage and the date the webpage was accessed by the authors.

Response 4:

We thank the reviewer for pointing out this reference issue. The incorrect citation

“Anon, n.d.” has been replaced with the appropriate reference to (Mauldin et al., 2012), and the full reference information has been corrected in the reference list. We have also carefully checked the manuscript for similar citation and reference errors and corrected them.

Changes in Manuscript:

Modified version: 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), and urban Beijing (Guo et al., 2021) have consistently supported the substantial role of sCIs in SO₂ oxidation and subsequent sulfuric acid (H₂SO₄) formation.

Comment 5:

line 54: the cited reference, Boy et al., 2013, is not in the list of references given at the end of the manuscript

Response 5:

We thank the reviewer for pointing this out. The missing reference, (Boy et al., 2013), has now been added to the reference list. We have also cross-checked the in-text citations and reference list throughout the manuscript to ensure consistency.

Changes in Manuscript:

Modified version: 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), and urban Beijing (Guo et al., 2021) have consistently supported the substantial role of sCIs in SO₂ oxidation and subsequent sulfuric acid (H₂SO₄) formation.

Comment 6:

line 58-59: lack of references for: ‘Existing studies reported ...environments’.

The authors should comment on a number of other relevant previous studies addressing the contribution of the sCI + SO₂ reactions to the production of H₂SO₄ and sulfate aerosols in the atmosphere, such as: Mauldin et al. Nature 2012; Kim, S et al. Environ. Sci. Technol. 2015; Kukui, A. Atmos. Chem. Phys. 2021; Sarwar, G. et al Atmos. Environ. 2014.

Response 6:

We thank the reviewer for pointing out these important and highly relevant studies. The original manuscript did not provide sufficient references to support the statement that the importance of sCI + SO₂ chemistry varies across atmospheric environments. In the revised Introduction, we have expanded the literature review to include both field-based and modeling studies addressing the contribution of sCIs to H₂SO₄ and sulfate formation. Specifically, we have added (Mauldin et al., 2012), (Kim et al., 2015), and (Kukui et al., 2021) as field-based studies showing that sCIs can provide an additional, environment-dependent source of H₂SO₄ under VOC-rich or biogenically influenced conditions. (Mauldin et al., 2012) showed that OH oxidation alone could not fully explain the observed sulfuric acid budget in boreal forests, whereas inclusion of sCI chemistry improved the closure of H₂SO₄ formation. (Kim et al., 2015) evaluated sCI-

derived H₂SO₄ production downwind of Dallas-Fort Worth and showed that the sCI pathway can provide an additional H₂SO₄ source, although it remains smaller than the OH pathway during midday. (Kukui et al., 2021) further estimated that sCIs contributed approximately 10% of daytime and 40% of nighttime H₂SO₄ formation at a Mediterranean site influenced by biogenic emissions. We have also added (Sarwar et al., 2014), which incorporated explicit SCI chemistry into the CMAQ model and demonstrated that SO₂ oxidation by sCIs can enhance summertime sulfate in biogenically active regions, with strong sensitivity to the assumed SCI+SO₂ and SCI + H₂O/(H₂O)₂ kinetics. These added studies further highlight the strong regional and environmental variability in the importance of sCI chemistry, and they reinforce the need to incorporate updated CI chemical kinetics when quantitatively evaluating the contribution of sCIs to SO₂ oxidation.

Changes in Manuscript:

The Introduction has been revised to include the suggested references.

Comment 7:

lines 63-64: The authors should change ‘...unimolecular decomposition as well’ to ‘...unimolecular decomposition/isomerisation as well’

Response 7:

We thank the reviewer for pointing out this imprecise description of the reaction pathways. In the revised manuscript, the Introduction has been substantially reorganized, and the original sentence containing this expression has been removed. Nevertheless, we have carefully checked the manuscript and corrected similar expressions.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 8:

line 67 states that there is an ‘intense competition’ between H₂O/(H₂O)₂ and SO₂ for reaction with sCI’. The authors should clarify which type of Criegee intermediates they are referring to here as the intensity of this competition depends on the sCI structure; for example CH₂OO reacts much more rapidly with water vapour than other sCIs.

Response 8: We thank the reviewer for this important comment. In the revised manuscript, the Introduction has been substantially reorganized, and the original sentence containing this expression has been removed. Nevertheless, we have checked the manuscript carefully and specify the relevant sCI species where appropriate when discussing their reactions with H₂O/(H₂O)₂ and SO₂.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 9:

lines 70-71: It is not a direct reaction of VOCs with NO_x producing O₃. Therefore,

please change the statement about the O₃ formation in the troposphere at day time to one such as ‘...O₃ is primarily formed through chemistry involving VOCs and NO_x, where NO₂ produced in these reactions photolyses to generate ozone...’.

Response 9:

We thank the reviewer for catching this imprecise expression. O₃ is produced through the reaction of O₂ with O generated from NO₂ photolysis. This process is continuously driven by the complex photochemical cycles of its precursors, VOCs and NO_x. In the revised manuscript, the Introduction has been substantially reorganized, and the original sentence containing this expression has been removed. Nevertheless, we have carefully checked the manuscript and corrected similar expressions.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 10:

line 72: The O₃ photolysis in the presence of water vapour is an important source of OH at day time and should be mentioned: ‘...through the photolysis of precursors such as HONO and O₃ in the presence of water vapour...’

Response 10:

We thank the reviewer for this helpful suggestion. The photolysis of O₃ in the presence of water vapour is an important daytime source of OH and should be acknowledged when introducing major OH sources. In the revised manuscript, the Introduction has been substantially reorganized, and the original sentence containing this expression has been removed. Nevertheless, we have carefully checked the manuscript and corrected similar expressions.

Changes in Manuscript:

The Introduction has been substantially revised, and the original sentence containing this expression has been removed.

Comment 11:

lines 74-75: The statement ‘OH and sCIs are intermediate products generated via different reaction pathways’ is not completely true. The authors state just before this that alkene ozonolysis is an important source of OH at night time; the decomposition of sCI formed following alkene ozonolysis can be a significant source of OH.

Response 11:

We thank the reviewer for pointing out this inaccurate statement. OH and sCIs are not strictly generated through independent reaction pathways, because the decomposition of sCIs formed during alkene ozonolysis can also contribute to OH production. In the revised manuscript, the Introduction has been substantially reorganized, and the original sentence containing this expression has been removed. Nevertheless, we have carefully checked the manuscript and revised related descriptions to better reflect the chemical coupling between OH and sCIs chemistry.

Changes in Manuscript:

The original sentence in the Introduction has been removed during revision. Modified

version: Moreover, the production and loss processes of OH and sCIs are not mutually independent but are deeply coupled through the underlying reaction mechanisms (Zhu et al., 2023).

Comment 12:

line 85: Reference for MCM v3.3.1 is missing

Response 12:

We thank the reviewer for pointing this out. The citation to the MCM website has now been added to the manuscript.

Changes in Manuscript:

Modified version: As a prerequisite for all subsequent analyses, we first updated the gas-phase kinetics of Criegee intermediates in the Master Chemical Mechanism (MCM v3.3.1, via website: www.mcm.york.ac.uk) using the latest evaluated rate coefficients recommended by the International Union of Pure and Applied Chemistry (IUPAC), thereby reducing the propagation of mechanistic uncertainties into our conclusions.

Comment 13:

lines 106-107: The statement ‘the concentration of bimolecular water is 10^4 times the concentration of H_2O in the atmosphere’ is wrong. I think the authors meant the other way around. There is no explanation why [water monomer] was considered 10^4 times larger than [water dimer]. From the equilibrium constant for the dimerisation, $K = [\text{dimer}]/[\text{monomer}]^2$, it follows that at $T = 25^\circ\text{C}$ $[\text{dimer}] = 10^{-4} [\text{monomer}]$ if RH is around 13%. How much was the relative humidity and temperature at the observation site (Wuhai city)? A clear explanation about the choice of the 10^4 factor is needed, in both the manuscript and the supplement, where this factor is included in the rate coefficients for the sCI + water dimer reactions.

Response 13:

We thank the reviewer for this careful and important comment and for providing the precise thermodynamic context and guidance. First, we acknowledge that the relationship between water monomer and dimer concentrations was incorrectly stated in the original manuscript. In fact, in our study, $[\text{H}_2\text{O}]$ is considered to equal $1 \times 10^4 [\text{H}_2\text{O}]_2$.

The reviewer is absolutely correct that, strictly based on the equilibrium constant at 25°C , a $[(\text{H}_2\text{O})_2]/[\text{H}_2\text{O}]$ ratio of 10^{-4} corresponds to a relatively dry environment (e.g., RH around 13%). Previous studies have established that in typical Earth's atmospheric conditions, the relative concentration of water dimer to water monomer spans from 10^{-3} to 10^{-4} (Tretyakov et al., 2014). Even at a high relative humidity of 85% at 25°C , the dimer concentration is roughly three orders of magnitude lower than the monomer (Ryzhkov and Ariya, 2006). Regarding the meteorological conditions at our observation site: Wuhai city is located in an semi-arid region of Inner Mongolia, China. During our observation period, the average temperature was $28 \pm 4^\circ\text{C}$ and the average relative humidity was relatively low, around $39 \pm 14\%$.

However, regarding the valid question on the actual calculation of water dimer concentrations, treating this ratio as a constant introduces significant uncertainty. The

actual dimer concentration scales quadratically with water monomer abundance and is highly temperature-dependent. To accurately represent the kinetics of the sCIs + (H₂O)₂ reaction in the box model, we employed a pre-equilibrium approximation. The dimer is assumed to be in steady-state equilibrium with the monomer. We rigorously parameterized the temperature-dependent equilibrium constant (K_{eq}^D) using high-accuracy thermochemical data from the Active Thermochemical Tables (ATcT) (Ruscic, 2013), allowing us to calculate an apparent third-order rate constant ($k_{eff} = k_{sCIs+(H_2O)_2} \cdot K_{eq}^D$). The complete mathematical derivation and parameter details have now been explicitly added to Section 2.1 of the revised manuscript.

Changes in Manuscript:

Modified version:

To accurately represent sCIs+(H₂O)₂ in MCM v3.3.1 within the box model, a pre-equilibrium approximation was employed. Given the rapid exchange between water monomers and dimers, the concentration of (H₂O)₂ is assumed to be in steady-state equilibrium with the water monomer:

$$K_{eq}^D = \frac{[(H_2O)_2]}{[H_2O]^2}$$

where K_{eq}^D is the temperature-dependent equilibrium constant (Scribano et al., 2006). The reaction of sCIs with water dimers was parameterised using an apparent third-order rate constant (Lade et al., 2024b), k_{eff} , such that the total reaction rate (r) is expressed as:

$$r = k_b[sCI][(H_2O)_2] = k_b \times K_{eq}^D[sCI][H_2O]^2 = k_{eff}[sCI][H_2O]^2$$

The apparent rate constant is defined as the product of the bimolecular rate constant for the sCIs+(H₂O)₂ reaction (k_b) and the equilibrium constant (K_{eq}^D). To ensure the highest thermodynamic accuracy in calculating K_{eq}^D , thermochemical data were retrieved from the Active Thermochemical Tables (ATcT). Standard Gibbs free energies of formation ($\Delta_r G_T^\circ$) for both the water monomer and the water dimer were extracted from Table 1 and Table 3 of (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 covers the typical conditions of the troposphere. The temperature dependence of K_{eq}^D was then obtained via a linear regression of K_{eq}^D against $\frac{1}{T}$, resulting in the following Arrhenius-type parameterisation used in the model:

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

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

Comment 14:

Tables 1 and 2: Remove the word ‘bimolecular’ from the title as the tables show pseudo-first rate coefficients for the sCI decomposition/isomerisation too. Suggest adding notes under the tables showing the units of the rate coefficients (Please consult how tables are presented in other papers published in Atmos. Chem. Phys.) The errors associated with the rate coefficient values should be included, as well as the temperature, pressure,

and a reference for MCM v3.3.1.

Response 14:

We thank the reviewer for this helpful suggestion. The tables have been comprehensively revised to include: (a) corrected titles; (b) units of rate coefficients in table notes; (c) associated uncertainties for all values; (d) temperature and pressure conditions; and (e) MCM v3.3.1 references.

Changes in Manuscript:

Tables 1 and 2 have been revised.

Comment 15:

line 119: Please add ‘see Section 2.2.2 after ‘(...and PAN)’

Response 15:

We thank the reviewer for this suggestion. The manuscript has been revised accordingly.

Changes in Manuscript:

Modified version: In this mode, hourly observations(see Section 2.2.2) of trace gases (NO_x, CO, SO₂, HONO, NMHCs, and OVOCs), meteorological parameters (temperature, T; relative humidity, RH; and pressure, P), and photolysis frequencies act as time-varying constraints for the simulation.

Comment 16:

lines 122-133 (Observation data): The key observations time series as well as a chart showing the percentage contributions of the alkenes shown in Table 2 to the total alkene concentration during the campaign should be included in the supplement.

Were all the observations used in the present study? I suggest removal of the ones which were not used.

Response 16:

We thank the reviewer for this helpful suggestion. The observational constraints used in the simulations should be presented more clearly. In the revised Supplement, we have added a summary table of the major constrained variables during the target simulation period (1 June to 15 July 2021), including their mean, median, and standard deviation, as well as time-series figures showing the temporal variation of the major constrained species and meteorological parameters. We have further revised Section 2.2.2 (Observation data) to remove observations that were not used in the present study. In the original manuscript, this section partly described the measurement capability of the instruments, which may have caused ambiguity regarding which observations were actually used in the analysis. This has now been clarified.

Changes in Manuscript:

Modified version: “Hourly observations from the Wuhai City Atmospheric Environment Super Monitoring Station were used for the observation-constrained simulations and observation-based machine learning analysis. The variables used in this study include trace gases (PM_{2.5}, CO, SO₂, NO₂, O₃, VOCs), meteorological parameters (WS, T, P, RH), photolysis frequencies, inorganic ions including SO₄²⁻ and NO₃⁻, and Fe as an indicator of transition-metal-related aqueous-phase processes.”. Variables measured at the station but not used in the present analysis have been removed from

the revised description to avoid ambiguity. The Supplement now includes time series of the key constrained variables and a chart showing the relative contributions of the representative alkenes listed in Table 2 to total alkene abundance during the campaign.

Comment 17:

line 135, regarding ‘We treated the AtChem inputs as features’. Is the meaning that part of the AtChem outputs represented input variables (‘features’) in the machine learning method? Please re-write the sentence to clarify. The term ‘feature’ should be explained here.

Response 17:

We thank the reviewer for pointing out this ambiguity. The original wording could be misread as implying that AtChem outputs were used as input variables in the machine-learning model. This was not our intention. In the revised manuscript, we now define the term “feature” at first use. In machine-learning terminology, a feature refers to an input predictor used by the model. In the AtChem-based machine learning analyses, the features are the prescribed chemical, meteorological, or scenario variables used as AtChem inputs or used to define the simulation scenarios, such as O₃, alkenes, NO_x, RH, temperature, and photolysis rates. The quantities calculated from the AtChem simulations, including LSO_{2,sCl_s}, LSO₂, and μ_{sCl_s}, are used as target variables rather than features. This revision clarifies that the machine learning model was trained to map prescribed AtChem input/scenario variables to AtChem-calculated diagnostic outputs, rather than treating AtChem outputs as input features.

Changes in Manuscript:

In the revised manuscript, we have added further clarifications in Section 2.3 (Machine learning surrogate and interpretation framework).

Comment 18:

line 142: Please explain what is meant by ‘L1 and L2 penalties’

Response 18:

We thank the reviewer for this helpful suggestion. The terms “L1 and L2 penalties” were not sufficiently explained in the original manuscript. In the revised manuscript, we now define them when they are first mentioned. Specifically, we explain that L1 and L2 penalties are regularization terms added to the XGBoost objective function to reduce overfitting and improve model generalization. In XGBoost, the model prediction is obtained from an ensemble of decision trees, and the terminal leaves of these trees are assigned numerical weights. The L1 penalty penalizes the absolute magnitude of these leaf weights and can shrink small weights toward zero, thereby promoting a simpler and more sparse model. The L2 penalty penalizes the squared magnitude of the leaf weights and discourages excessively large weights, thereby producing smoother and more stable predictions. Both penalties help prevent the model from fitting noise in the training data.

Changes in Manuscript:

We have incorporated a concise version of this explanation into Section 2.3 (Machine learning surrogate and interpretation framework). Modified version: “The L1 and L2

penalties are regularization terms used to reduce overfitting in XGBoost. The L1 penalty penalizes the absolute magnitude of the tree leaf weights and can promote sparsity, whereas the L2 penalty penalizes the squared magnitude of the leaf weights and discourages overly large weights, leading to smoother and more stable predictions.”

Comment 19:

line 143: Please add a reference after Python xgboost library.

Response 19:

We thank the reviewer for this helpful suggestion, which improves the reproducibility of our study. We have updated the manuscript to explicitly cite both the foundational algorithm and the specific Python library used. The revised sentence now cites Chen and Guestrin (2016) and specifies the use of the XGBoost Python package (Version 2.1.3).

Changes in Manuscript:

Modified version: “In this study, the XGBoost model was implemented using the XGBoost Python package (Version 2.1.3; XGBoost Developers; <https://github.com/dmlc/xgboost>).”

Comment 20:

line 144: The term ‘hyperparameter’ should be explained.

Response 20:

We thank the reviewer for this helpful suggestion. In the revised manuscript, we now explain that hyperparameters are model settings specified before training. Specifically, in XGBoost, hyperparameters control the learning process and model complexity. Examples include the learning rate, maximum tree depth, number of trees, subsampling ratio, and regularization strengths. These settings determine how fast the model learns, how complex each tree can be, how many trees are included in the ensemble, and how strongly overfitting is penalized. In this study, these hyperparameters were tuned using cross-validation to improve model performance and generalization.

Changes in Manuscript:

This definition has been added to Sect. 2.3. Modified version: “Hyperparameters refer to model settings specified before training which include the learning rate, maximum tree depth, number of trees, subsampling ratio, and regularization strengths in XGBoost. These hyperparameters control the learning process, model complexity, and the degree of regularization, and were tuned by cross-validation to reduce overfitting and improve generalization.”

Comment 21:

line 152: Suggest to replace ‘xi and xc constitute...’ with ‘the sum of features xi and xc constitute...’

Response 21:

We thank the reviewer for this helpful wording suggestion. We have revised the sentence as suggested to make the notation clearer.

Changes in Manuscript:

This definition has been added to Sect. 2.3. Modified version: “Here, x_i denotes the feature or feature subset of interest, and x_c denotes all remaining complementary features. Together, x_i and x_c constitute the full set of input features used by the model.”.

Comment 22:

line 153: Add ‘(equation 1)’ at the end of the sentence, i.e. ‘...model (equation 1)’

Response 22:

We thank the reviewer for the careful reading. We have added "(equation 1)" at the end of the sentence as suggested.

Changes in Manuscript:

Modified version: “Here, x_i denotes the feature or feature subset of interest, and x_c denotes all remaining complementary features. Together, x_i and x_c constitute the full set of input features used by the model.”. The equation reference has also been added in the revised manuscript.

Comment 23:

line 154: What does E represent?

Response 23:

We thank the reviewer for pointing out this omission. In the Partial Dependence Plot (PDP) equation, E represents the expectation operator. Specifically, it denotes the average model prediction over the distribution of the complementary features x_c . In practical terms, the PDP is calculated by fixing the feature of interest x_i at a given value and averaging the model predictions over all samples of the remaining features x_c . We have added this explanation after the equation.

Changes in Manuscript:

This definition has been added to Sect. 2.3. Modified version: “ E represents the mathematical expectation operator, which calculates the average model prediction over the marginal distribution of x_c .”

Comment 24:

line 161: The authors should provide examples of the specific fields they are referring to in: ‘...adopted across a broad range of fields’.

Response 24:

We thank the reviewer for this helpful suggestion. In the revised manuscript, we have rewritten this sentence to explicitly mention the fields in which SHAP has been applied.

Changes in Manuscript:

This definition has been added to Sect. 2.3. Modified version: “SHAP has been increasingly applied in environmental and atmospheric sciences, including urban climate and remote-sensing studies that quantify the effects of urban morphology on land surface temperature (Chen et al., 2024), ecosystem-scale environmental response analysis (Yi and Wu, 2023), and atmospheric studies that separate aerosol effects from meteorological co-variability in cloud-water observations (Zhang et al., 2025).

Comment 25:

lines 175-176 states that ‘Figures 1a and 1b displays the global SHAP values for each feature, ranked from top to bottom by their mean |SHAP| values.’ However, the high to low axes in those figures are labelled ‘feature value’ instead of SHAP. The meaning of SHAP in the present study should be explained.

Response 25:

We thank the reviewer for pointing out this lack of clarity. To clarify, the SHAP summary plot (beeswarm plot) incorporates three distinct dimensions of information:

The Y-axis: The y-axis represents the individual features, which are arranged from top to bottom in descending order of their global importance (mean |SHAP|).

The X-axis: This represents the actual SHAP value, which shows the impact of a feature on the model's output for individual data points.

The Color Scale (High to Low): The color bar labeled "feature value" strictly represents the actual numerical magnitude of the input features (e.g., whether the initial concentration of O₃ or alkenes is high (red) or low (blue)).

Changes in Manuscript:

We have thoroughly revised the figure caption to ensure this multi-dimensional information of SHAP summary plot are clearly communicated to the readers.

Comment 26:

lines 200-201: The authors should provide references—such as Onel et al, Phys.Chem.Chem.Phys. 2021 and Lade et al. J. Phys. Chem. A 2024—that discuss the dominant removal of E-CH₃CHOO and CH₂OO by reaction with water vapour, in comparison with their losses via reaction with SO₂ under tropospherically relevant conditions.

Response 26:

We thank the reviewer for these relevant references. We have added (Onel et al., 2021) and (Lade et al., 2024a) to the discussion of the dominant removal pathways of E-CH₃CHOO and CH₂OO.

Changes in Manuscript:

We have added the citations for (Onel et al., 2021) and (Lade et al., 2024a) to Section 3.1.

Comment 27:

Figures 1(a) and 1(b): The meaning of the x axis labels are confusing and should be explained.

Response 27:

We thank the reviewer for pointing out this. The x-axis represents the SHAP value, which quantifies the contribution of each feature to the predicted L_{SO₂,sCIs} relative to the mean model prediction. Positive SHAP values indicate that the feature increases the predicted SO₂ loss rate via sCIs, whereas negative values indicate a decrease.

Changes in Manuscript:

The revised caption now explains that the x-axis shows SHAP values and that the color scale represents feature values.

Comment 28:

Figure 2: The unit of SHAP value is mole cm^{-3} (unit of concentration) while in Figure 1 is mole $\text{cm}^{-3} \text{ s}^{-1}$ (unit of rate). Why is this difference? The authors should clarify why moles were used rather than number of molecules, the latter being more commonly used in atmospheric chemistry. What are D1(n=120) and D2(n=128) in the legend of top left figure? There are no units for ‘chemical feature concentrations’. The authors should clarify what is meant by ‘chemical feature concentrations’ in both the main text and Figure 2 capture. Do these represent the initial alkene concentration inputs in the machine learning model?

Response 28:

We thank the reviewer for pointing out these ambiguities in Fig. 2. First, the SHAP values should have the same physical unit as the target variable of the XGBoost regression model. In this analysis, the target variable is the sCI-mediated SO_2 loss rate, $L_{\text{SO}_2, \text{sCIs}}$; therefore, the SHAP values represent contributions to the predicted reaction rate relative to the model baseline. The original unit label “molecule cm^{-3} ” was incorrect and has been corrected. To avoid overloading the figure, the revised axis is labeled “SHAP value (impact on $L_{\text{SO}_2, \text{sCIs}}$)”, and the caption states that the SHAP values correspond to the same unit as the target reaction rate, i.e., molecules $\text{cm}^{-3} \text{ s}^{-1}$.

Second, in the manuscript, the intended unit was molecule (e.g., molecule $\cdot \text{cm}^{-3} \cdot \text{s}^{-1}$), but we incorrectly used “mole” as a shorthand for “molecule”. We have now systematically checked and corrected all corresponding units throughout the manuscript and figures.

Third, the labels “D1” and “D2” in the original legend referred to the datasets generated using MCM v3.3.1 and MCM v3.3.1g, respectively, and the numbers in parentheses indicated the number of plotted samples. During the revision, this figure was removed and replaced with a revised analysis focused on the RH-dependent responses of key alkenes under MCM v3.3.1g mechanism. Therefore, the ambiguous D1/D2 notation no longer appears. The revised figure caption now explicitly defines the plotted variables, the model target, and the meaning of the SHAP axis.

Changes in Manuscript:

The original Fig. 2 has been replaced. The revised figure and caption now define the SHAP axis, the target variable $L_{\text{SO}_2, \text{sCIs}}$, the meaning of the feature values, and the corrected unit convention.

Comment 29:

-line 235: The authors should specify which ‘specific region’ they are referring to

Response 29:

We thank the reviewer for this helpful suggestion. The “specific region” refers to the Wuhai City area in Inner Mongolia, China, where the monitoring station is located. We have replaced the vague phrasing with the specific location name.

Changes in Manuscript:

The vague phrase “specific region” has been replaced by “Wuhai City, Inner Mongolia, China”.

Comment 30:

lines 244 - 245 states: that ‘the strongest pairs’ are ‘O₃ × alkene%, O₃ × VOCs, and VOCs × alkene%’. However, the Top 10 Interaction Effects in Figure 3b shows that the contribution of NO_x × NO₂% is larger than the contribution of VOCs × alkene%. Why NO_x × NO₂% is not listed in ‘the strongest pairs’?

Response 30

We thank the reviewer for pointing out this inconsistency. The original wording did not accurately reflect the ranking shown in the ANOVA interaction plot. In the revised manuscript, the original ANOVA-based figure has been removed and replaced by Sobol sensitivity analysis, which more appropriately quantifies both first-order effects and interaction-related contributions. The revised daytime analysis shows that ARI, O₃, alkene%, and VOCs dominate the variance in μ_{SCIS} , with first-order Sobol indices (S_1) of approximately 0.18, 0.16, 0.15, and 0.11, respectively. The revised text therefore no longer makes the unsupported statement about the "strongest pairs" in the original ANOVA figure.

Changes in Manuscript:

The original ANOVA figure and the corresponding text describing the “strongest pairs” have been removed. The revised manuscript now reports Sobol first-order and total-order indices and explains the dominant controls on μ_{SCIS} using the updated sensitivity framework.

Comment 31:

Figure 3b: I recommend including error bars in the contribution values showed in the Main Effects Contribution and Top 10 Interaction Effects figures. The authors should describe more clearly the methodology used to generate the pie chart in in both the main text and Figure 3 capture. What is the meaning of the numbers on the right vertical axis of the figure in the bottom right corner? The legend includes only the text ‘(b) ANOVA effect analysis for all features’, which is not sufficiently explanatory.

Response 31

We thank the reviewer for this helpful suggestion. The original Fig. 3b did not provide sufficient methodological information and that some graphical elements, including the contribution values and right-axis labels, were not clearly explained. In the revised manuscript, we have removed the ANOVA-based plot and replaced it with Sobol sensitivity analysis. This avoids the ambiguity associated with the previous pie/bar visualization and provides a clearer variance-based interpretation of main and interaction effects. The revised caption now defines S_1 as the independent contribution of each factor and S_T as the total contribution including interactions.

Changes in Manuscript:

The original ANOVA effect plot has been removed. The revised figure presents Sobol sensitivity indices (S_1 and S_T), and the caption now explains the method, plotted quantities, and interpretation of first-order and total-order effects.

Comment 32:

line 258, regarding: ‘The three factors with the largest main effects ...in the order O₃ > alkene% > VOCs.’ This sentence refers to Figure 4 where the Main Effects Contribution

plot shows that the order is O₃ > VOCs > alkene%.

Response 32

We thank the reviewer for identifying this mismatch between the text and Fig. 4. The original sentence incorrectly described the order of the main effects. In the revised manuscript, the original nighttime ANOVA figure has been removed and replaced by Sobol sensitivity analysis. The revised nighttime results show that VOCs, alkene%, and O₃ are important controls, with first-order Sobol indices of approximately 0.13, 0.11, and 0.08, respectively. The revised text has therefore been rewritten according to the updated Sobol results rather than the previous ANOVA ranking.

Changes in Manuscript:

The inconsistent sentence and the original ANOVA figure have been removed. The revised nighttime discussion now reports Sobol sensitivity indices.

Comment 33:

line 258: The authors should explain why the NO_x × NO₂% interaction is not listed in ‘the strongest pairs’ because the Top 10 Interaction Effects plot shows that its contribution is the largest (see similar comment about Figure 3 above).

Response 33

We thank the reviewer for raising this point. In the original manuscript, VOCs, O₃, and alkene% were discussed separately from NO_x and NO₂%, and the importance of the NO_x×NO₂% interaction was not stated consistently with the ANOVA interaction ranking. We agree that this was an error in presentation. In the revised manuscript, the original ANOVA-based interaction plot has been removed and replaced by Sobol sensitivity analysis. The revised text now discusses interaction effects based on the difference between total-order (S_T) and first-order (S₁) Sobol indices.

Changes in Manuscript:

The original interaction ranking has been removed. The revised analysis now uses Sobol sensitivity indices to quantify independent and interaction-related contributions, and the discussion of NO_x and NO₂% has been rewritten accordingly.

Comment 34:

line 263: ‘the promoting potential of O₃, VOCs, and alkene% on μ_{sCIs}% was unlocked’ is confusing and should be clarified.

Response 34

We thank the reviewer for pointing out this unclear phrasing. The original sentence aimed to describe that under high NO_x/NO₂ conditions, increases in O₃, VOCs, and the fraction of alkenes lead to a larger increase in μ_{sCIs} (the relative contribution of sCIs to SO₂ oxidation). Mechanistically, at night, elevated NO₂ promotes OH termination, reducing the OH-mediated SO₂ oxidation pathway. As a result, μ_{sCIs} becomes more sensitive to the availability of O₃, VOCs, and alkenes. We have revised the manuscript text.

Changes in Manuscript:

The “Summary and conclusion” has been substantially revised, and the original sentence containing this expression has been removed.

Comment 35:

Figure 4 capture: The word ‘assessing’ should be replaced by ‘assessment of’. Regarding both Figures 3 and 4 captures: Please include how LSO₂(g) was generated and what machine learning method was used to generate the plots in (a).

Response 35

We thank the reviewer for this suggestion. In the revised manuscript, the original Figs. 3 and 4 have been replaced because we expanded the simulation dataset and replaced the ANOVA-based interpretation with Sobol sensitivity analysis. The revised captions now explicitly state that μ_{sClS} and LSO₂ were diagnosed from AtChem simulations coupled with the updated MCM v3.3.1g mechanism. The XGBoost surrogate model was trained on these AtChem-derived outputs, and the PDPs show the predicted marginal responses from the trained model across the prescribed feature space. The revised figure captions also define each input feature, the target variable, and the meaning of S_I and S_T in the Sobol panels.

Changes in Manuscript:

The revised captions now describe how the AtChem outputs were generated, which XGBoost surrogate model was used, and how the PDP and Sobol sensitivity panels should be interpreted.

Comment 36:

lines 273-284: The authors should clarify what is meant by low – high NO₂%. The entire paragraph is somewhat confusing and should be reorganised for better clarity.

Response 36

We thank the reviewer for pointing out this lack of clarity. We thank the reviewer for pointing out the ambiguity in this paragraph. In the revised manuscript, we have clarified that “low NO₂%” and “high NO₂%” refer to the relative fraction of NO₂ within total NO_x, which modulates the partitioning between NO and NO₂. Specifically, low NO₂% indicates that most NO_x is present as NO, favoring radical propagation and OH regeneration, whereas high NO₂% indicates a higher proportion of NO₂, promoting OH termination. We have reorganized the paragraph to improve clarity and readability, explaining how NO₂% and NO_x alter the relative contribution of sClS to SO₂ oxidation (μ_{sClS}).

Changes in Manuscript:

The paragraph has been reorganized.

Comment 37:

Figure 5: There are no explanations for the numbers shown in any of the schematics (a-f), making their meaning unclear. Please clarify in both the main text and the figure capture.

Response 37

We thank the reviewer for pointing out this lack of clarity. The numbers in the original Fig. 5 represent the reaction rate coefficients (in molecules·cm⁻³·s⁻¹) for the key reaction pathways shown in the schematics. To avoid confusion, the revised manuscript now

explicitly defines these numbers in both the main text and the figure caption. Where the figure was replaced during the structural revision, the new figure captions now define all plotted quantities and units.

Changes in Manuscript:

The relevant figure caption and main-text description have been revised to define the numbers and their units.

Comment 38:

Figure 6: What are the differences between SA-sCI and SA-sCIg and between SA-OH and SA-OHg? The main text should clearly state what SA-sCI and SA-OH represent and the figure capture should explain the meaning of all 4 notations (SA-sCI, SA-sCIg, SA-OH and SA-OHg) and which version of MCM corresponds to each of the plot line (black and purple).

Response 38

We thank the reviewer for pointing out this lack of clarity. In the original manuscript, SA-sCI and SA-OH referred to sulfuric acid production rates from the sCI- and OH-mediated oxidation pathways calculated with MCM v3.3.1, whereas SA-sCIg and SA-OHg referred to the corresponding rates calculated with the updated MCM v3.3.1g mechanism. We agree that this notation was confusing and that the comparison between the original and updated mechanisms in this later section distracted from the main purpose of the observational evaluation. In the revised manuscript, this figure has been removed. The revised observational section now focuses on (i) the comparison between simulated H₂SO₄-related diagnostics and observed SO₄²⁻, and (ii) the comparison between AtChem-simulated μ_{sCl_2} and the μ_{sCl_2} predicted by the XGBoost surrogate model. Therefore, the unclear SA-sCI/SA-sCIg notation no longer appears in the revised figure.

Changes in Manuscript:

The original Fig. 6 and the SA-sCI/SA-sCIg notation have been removed. The revised figure now focuses on the observational evaluation of the updated mechanism and surrogate model, with all plotted variables defined explicitly in the caption.

Comment 39:

line 342: Please state the meaning of WS.

Response 39

We thank the reviewer for this suggestion. WS stands for wind speed (m/s). We clarified this in Section 2.2.2 (Observation data).

Comment 40:

line 348: The authors should clarify the rationale for including Fe in the features contributing to the sulfate formation.

Response 40

We thank the reviewer for this helpful suggestion. Fe was included as a feature because transition metals, especially Fe(III) and Mn(II), have been widely recognized as important indicators of aqueous-phase SO₂ oxidation processes in cloud/fog water and

aerosol liquid water. In our XGBoost model, Fe served as an indicator to evaluate whether transition-metal-catalyzed processes were associated with sulfate formation during the study period. The SHAP results showed that Fe was associated with a negative effect on SO_4^{2-} during the 2021.06.01-2021.07.15 episode, which does not support a dominant role of aqueous-phase oxidation under these conditions. This has now been clarified in the 3.3 section.

Changes in Manuscript:

The rationale for including Fe has been added to Sect. 2.2.2 and Sect. 3.3, and Fe is now described as an indicator of transition-metal-related aqueous-phase oxidation.

Comment 41:

Figure 7(a): See comments on Figure 1a and b above.

Response 41

We have applied the same revisions as described for Figures 1a and 1b: improved readability, proper axis labels, and expanded figure captions.

Changes in Manuscript:

The figure has been revised with clearer axis labels, larger text, and an expanded caption.

Comment 42:

Figure 7(b): Explain ‘high-sCIs and low-sCIs datasets’.

Response 42

We thank the reviewer for pointing out that the definitions of the high-sCIs and low-sCIs datasets were not sufficiently clear. In the revised manuscript, we have clarified that these datasets were defined according to the fractional contribution of the sCI pathway, μ_{sCIs} , simulated by AtChem coupled with the updated MCM v3.3.1g mechanism. Specifically, in Sect. 3.2, we calculated μ_{sCIs} for all designed simulation scenarios and used the median values of the scenario-based μ_{sCIs} distribution as the classification thresholds. The median threshold was 1.6% for daytime and 10% for nighttime. These thresholds were then applied consistently in Sect. 3.3 to classify the observation-constrained simulation period from 1 June to 15 July 2021. For this period, μ_{sCIs} was calculated at each time step from the AtChem-MCM v3.3.1g reaction rates, and daily mean values were calculated separately for daytime and nighttime. Daytime days with mean μ_{sCIs} above 1.6% were classified as the high- μ_{sCIs} dataset, while those below this threshold were classified as the low- μ_{sCIs} dataset. Similarly, nighttime periods with mean μ_{sCIs} above 10 % were classified as high- μ_{sCIs} , while those below this threshold were classified as low- μ_{sCIs} .

Changes in Manuscript:

We have revised the text and the caption these figures to define the high- μ_{sCIs} and low- μ_{sCIs} datasets explicitly.

Comment 43:

line 354: A reference for the methodology used to generate comparative ‘beeswarm’ plots should be given.

Response 43

We thank the reviewer for this helpful suggestion. The comparative beeswarm plot used in this study is not a separate interpretation method; rather, it is a customized grouped visualization of standard SHAP values. To avoid confusion, we have revised the text and figure caption to explain how the comparative beeswarm plots were generated. Specifically, after calculating SHAP values for all samples, we divided the data into high- and low- μ_{sCl_2} subsets. For each selected feature, the SHAP values from the two subsets were plotted on the same x-axis, with the high- μ_{sCl_2} subset shown in the upper half of each feature row and the low- μ_{sCl_2} subset shown in the lower half. The color scale represents the corresponding feature value from low to high, following the convention of standard SHAP beeswarm plots. Therefore, the plot is used only to visually compare the distribution and magnitude of feature contributions between the two regimes.

Changes in Manuscript:

Modified caption text: “The comparative beeswarm plots are grouped SHAP beeswarm visualizations. For each feature, points in the upper half of the row represent the high- μ_{sCl_2} subset, whereas points in the lower half represent the low- μ_{sCl_2} subset. The x-axis shows the SHAP value, indicating the contribution of the feature to the predicted SO_4^{2-} , and the color indicates the corresponding feature value from low to high.”

Comment 44:

line 359: Replace the word ‘indispensable’ with ‘significant’.

Response 44

We thank the reviewer for this helpful suggestion. We have carefully checked the manuscript and corrected similar expressions.

Changes in Manuscript:

The word “indispensable” has been replaced with “significant” and similar overstatements have been checked throughout the manuscript.

Comment 45:

line 378: Explain the word ‘paradoxically’.

Response 45

We thank the reviewer for pointing this out. The word “paradoxically” was not sufficiently precise. What we intended to convey is that elevated NO_x decreased the overall SO_2 oxidation rate (H_2SO_4 formation) by reducing OH, whereas the sCl₂ pathway is much less affected. As a result, the fractional contribution of sCl₂ increases in relative terms, even though this does not imply an absolute enhancement of the sCl₂ pathway itself. We have revised the relevant text to make this point clearer.

Changes in Manuscript:

The “Summary and conclusion” has been substantially revised, and the original sentence containing this expression has been removed.

Comment 46:

line 383: Replace the word ‘indispensable’ with ‘important’.

Response 46

We thank the reviewer for this helpful suggestion. We have carefully checked the manuscript and corrected similar expressions.

Changes in Manuscript:

The “Summary and conclusion” has been substantially revised, and the original sentence containing this expression has been removed.

Comment 47:

lines 386-387 state: ‘While our box model simulations ...they are limited by the exclusion of meteorological factors..’. Please clarify what meteorological factors were excluded as line 118 states: ‘the model was constrained by the observed meteorological parameters (T, RH and p)’.

Response 47

We thank the reviewer for identifying this ambiguity. Our original wording lacked precision. The AtChem box model was indeed constrained by observed meteorological parameters, including temperature, relative humidity, and pressure, as stated in Sect. 2.2.1. These variables were used to calculate water vapor concentrations and to account for their effects on gas-phase chemical reaction rates and photochemical conditions. What we intended to emphasize in the original sentence was not that meteorological variables were excluded, but that a zero-dimensional box model does not explicitly represent meteorology-driven physical and dynamical processes, such as horizontal advection, vertical mixing, boundary-layer evolution, dilution, and deposition. We have clarified this and revised the relevant text in the manuscript accordingly.

Changes in Manuscript:

Modified version: While our box-model simulations provide detailed mechanistic insight into gas-phase SO₂ oxidation, they remain limited by the zero-dimensional framework. Although the simulations were constrained by observed temperature, relative humidity, pressure, and photolysis frequencies, they do not explicitly represent meteorology-driven physical processes, such as horizontal transport, vertical mixing, boundary-layer evolution, turbulent dilution, and deposition. In addition, multiphase chemistry was not included. Future studies using three-dimensional chemical transport models are therefore needed to quantify the role of sCIs in complex polluted atmospheres more comprehensively.

References

Boy, M., Mogensen, D., Smolander, S., Zhou, L., Nieminen, T., Paasonen, P., Plass-Dülmer, C., Sipilä, M., Petäjä, T., Mauldin, L., Berresheim, H., and Kulmala, M.: Oxidation of SO₂ by stabilized Criegee intermediate (sCI) radicals as a crucial source for atmospheric sulfuric acid concentrations, *Atmos. Chem. Phys.*, 13, 3865–3879, <https://doi.org/10.5194/acp-13-3865-2013>, 2013.

Chen, J., Wang, K., Du, P., Zang, Y., Zhang, P., Xia, J., Chen, C., and Yu, Z.: Quantifying the main and interactive effects of the dominant factors on the diurnal cycles of land surface temperature in typical urban functional zones, *Sustainable Cities and Society*, 114, 105727, <https://doi.org/10.1016/j.scs.2024.105727>, 2024.

Cox, R. A., Ammann, M., Crowley, J. N., Herrmann, H., Jenkin, M. E., McNeill, V. F., Mellouki, A., Troe, J., and Wallington, T. J.: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume VII – Criegee intermediates, *Atmos. Chem. Phys.*, 20, 13497–13519, <https://doi.org/10.5194/acp-20-13497-2020>, 2020.

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

Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol 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.

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

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

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

Lade, R. E., Onel, L., Blitz, M. A., Seakins, P. W., and Stone, D.: Kinetics of the Gas-Phase Reactions of *syn*- and *anti*-CH₃CHOO Criegee Intermediate Conformers with SO₂ as a Function of Temperature and Pressure, *J. Phys. Chem. A*, 128, 2815–2824, <https://doi.org/10.1021/acs.jpca.4c00199>, 2024a.

Lade, R. E., Blitz, M. A., Rowlinson, M., Evans, M. J., Seakins, P. W., and Stone, D.: Kinetics of the reactions of the Criegee intermediate CH₂OO with water vapour: experimental measurements as a function of temperature and global atmospheric modelling, *Environ. Sci.: Atmos.*, 4, 1294–1308, <https://doi.org/10.1039/D4EA00097H>, 2024b.

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

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

Ruscic, B.: Active Thermochemical Tables: Water and Water Dimer, *J. Phys. Chem. A*, 117, 11940–11953, <https://doi.org/10.1021/jp403197t>, 2013.

Ryzhkov, A. B. and Ariya, P. A.: The importance of water clusters (H_2O) ($n=2, \dots, 4$) in the reaction of Criegee intermediate with water in the atmosphere, *Chemical Physics Letters*, 419, 479–485, <https://doi.org/10.1016/j.cplett.2005.12.016>, 2006.

Sarwar, G., Simon, H., Fahey, K., Mathur, R., Goliff, W. S., and Stockwell, W. R.: Impact of sulfur dioxide oxidation by Stabilized Criegee Intermediate on sulfate, *Atmospheric Environment*, 85, 204–214, <https://doi.org/10.1016/j.atmosenv.2013.12.013>, 2014.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, *Atmos. Chem. Phys.*, 3, 161–180, <https://doi.org/10.5194/acp-3-161-2003>, 2003.

Scribano, Y., Goldman, N., Saykally, R. J., and Leforestier, C.: Water Dimers in the Atmosphere III: Equilibrium Constant from a Flexible Potential, *J. Phys. Chem. A*, 110, 5411–5419, <https://doi.org/10.1021/jp056759k>, 2006.

Tretyakov, M. Y., Koshelev, M. A., Serov, E. A., Parshin, V. V., Odintsova, T. A., and Bubnov, G. M.: Water dimer and the atmospheric continuum, 2014.

WHO: WHO Global Air Quality Guidelines: Particulate Matter (PM_{2.5} and PM₁₀), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide, 1st ed., World Health Organization, Geneva, 1 pp., 2021.

Yi, Z. and Wu, L.: Identification of factors influencing net primary productivity of terrestrial ecosystems based on interpretable machine learning --evidence from the county-level administrative districts in China, *Journal of Environmental Management*, 326, 116798, <https://doi.org/10.1016/j.jenvman.2022.116798>, 2023.

Zhang, H., Zhang, Y., Li, Z., and Zheng, Y.: Aerosol Influences on Cloud Water: Insights From ARM EPCAPE Observations With Explainable Machine Learning, *Geophysical Research Letters*, 52, e2025GL115163, <https://doi.org/10.1029/2025GL115163>, 2025.

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, 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, Y., Tang, X., Wang, T., Wang, T., Wang, W., Wang, X., Wang, Z., Yin, Y., Zhang, Q., Zhang, W.,

Zhang, Y., Zhang, Y., Zhao, Y., Zheng, M., Zhu, B., and Zhu, J.: Recent Progress in Atmospheric Chemistry Research in China: Establishing a Theoretical Framework for the “Air Pollution Complex,” *Adv. Atmos. Sci.*, 40, 1339–1361, <https://doi.org/10.1007/s00376-023-2379-0>, 2023.