Supplement of Atmos. Chem. Phys., 26, 12479–12503, 2026
https://doi.org/10.5194/acp-26-12479-2026-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

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

Yuhuan Zhu et al.

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

Section S1: Construction and evaluation of the XGBoost surrogate model for sCI-driven SO_2 gas-phase oxidation

To ensure that the surrogate model was trained on a representative and diverse sample of atmospheric conditions, we generated a large scenario matrix spanning the major controls on sCI formation and loss. Nine physicochemical variables were selected as input features: O_3, RH, NO_2, and six representative alkenes, namely ethene (C_2H_4), propene/but-1-ene (C_3C_4), trans-but-2-ene, cis-but-2-ene, 2-methylpropene, and isoprene. Propene and but-1-ene were merged into a single feature because of the close similarity of their ozonolysis mechanisms. O_3 and the alkenes jointly determine sCI production through ozonolysis, RH regulates sCI removal through reactions with water vapor and water dimers, and NO_2 provides an additional sink. The prescribed concentration ranges were O_3: 2–90 ppb, each alkene: 0.1–20 ppb, NO_2: 2–60 ppb, and RH: 10%–80%, chosen to span polluted near-surface atmospheric conditions.

The simulation matrix was assembled using four complementary design strategies. First, 3,000 scenarios were generated by scrambled Latin hypercube sampling (LHS) across the full nine-dimensional feature space to provide the main training sample. Second, one-factor-at-a-time (OFAT) simulations were performed by varying each of the nine variables over nine evenly spaced relative levels while holding all other variables at their midpoint values, yielding 81 scenarios in total. Third, RH-focused two-factor interaction scenarios were constructed for seven variable pairs, namely RH × O_3 and RH × each of the six alkenes. Each pair was sampled on an 11 × 11 full-factorial grid, giving 847 scenarios in total. This design was included because RH-dependent suppression of sCIs is mechanistically important and can interact strongly with ozonolysis precursors. Fourth, anchor scenarios were added at boundary and near-boundary locations of the predictor space, including global minima, midpoints, maxima, single-variable extremes, and RH-focused low-mid-high combinations, in order to improve model stability and reduce edge extrapolation errors. After removal of duplicate cases, the final scenario matrix contained 3,898 unique simulations. Each scenario prescribed the initial mixing ratios of the nine input features, while SO_2 was fixed at 10 ppb in all runs so that variability in the response primarily reflected changes in sCI abundance and reactivity rather than SO_2 availability. Temperature (T) and pressure (P) were likewise held constant at 290 K and 888 hPa, respectively. Chemical evolution was then simulated for 1 h using unconstrained AtChem box-model runs coupled separately to MCM v3.3.1 and MCM v3.3.1g, thereby producing two parallel datasets. For each run, the 1 h mean sCI-driven SO_2 oxidation rate, $L_{SO2,sCIs}$ (molec. cm^{-3} s^{-1}), was extracted as the target variable for surrogate modeling. The full scenario matrix and model outputs are archived in the Zenodo repository referenced at the end of this paper.

The same machine-learning workflow was applied to both mechanism versions. Each dataset was randomly divided into training (64%), validation (16%), and test (20%) subsets, corresponding to 2338, 780, and 780 samples, respectively. The test subset was withheld throughout model development and used only for final independent evaluation. XGBoost (Chen and Guestrin, 2016) was adopted as the surrogate model. Hyperparameters were optimized by randomized search (150 iterations) with five-fold cross-validation on the training subset over the following search ranges: number of estimators, 300-1200; maximum tree depth, 3-8; learning rate, 0.01-0.16; subsample ratio, 0.60-1.00; column subsampling ratio per tree, 0.50-1.00; minimum child weight, 1-9; L1 regularization coefficient α, 0-1; and L2 regularization coefficient λ, 0.5-2.5. After the optimal hyperparameter combination was identified, the final model was retrained on the training subset, with the independent validation subset used for early stopping (patience = 50 rounds) to suppress overfitting.

For MCM v3.3.1, early stopping selected 1129 trees, and the five-fold cross-validation score on the training subset was $R^2 = 0.9504 \pm 0.0186$. The final model achieved $R^2 = 0.9976$, RMSE = 1.31×10^3, and MAE = 9.22×10^2 molec. cm^{-3} s^{-1} on the training set; $R^2 = 0.9742$, RMSE = 4.17×10^3, and MAE = 2.15×10^3 molec. cm^{-3} s^{-1} on the validation set; and $R^2 = 0.9709$, RMSE = 4.34×10^3, and MAE = 2.24×10^3 molec. cm^{-3} s^{-1} on the fully independent test set. For MCM v3.3.1g, early stopping selected 1103 trees, and the five-fold cross-validation score on the training subset was

$R^2 = 0.9334 \pm 0.0109$. The final model achieved $R^2 = 0.9989$, RMSE = 2.44×10^3, and MAE = 1.72×10^3 molec. cm^{-3} s^{-1} on the training set; $R^2 = 0.9741$, RMSE = 1.15×10^4, and MAE = 6.13×10^3 molec. cm^{-3} s^{-1} on the validation set; and $R^2 = 0.9729$, RMSE = 1.21×10^4, and MAE = 6.34×10^3 molec. cm^{-3} s^{-1} on the independent test set. Overall, the close agreement between validation and test performance indicates that the selected XGBoost surrogates generalized well beyond the training data without detectable overfitting.

Section S2: Construction and evaluation of the XGBoost surrogate models for μ_{sCIs} and L_{SO2}

In total, seven features were used in the surrogate model: NO_x, NO_2%, O_3, total VOCs, alkene%, RH, and ARI. The first six were treated as continuous variables, whereas ARI was treated as a discrete ordinal variable. The ranges of the six continuous variables were determined from the 5th–95th percentiles of three years of hourly observations at the Wuhai supersite (Sect. 2.2.2). For scenario design and model training, each continuous feature was linearly normalized to the interval [0, 1], where 0 and 1 denote the lower and upper bounds of the prescribed range, respectively. The corresponding atmospheric values are shown in Fig. S1. The simulation scenario matrix was constructed separately within each ARI class using four complementary design strategies. First, 100 optimized Latin hypercube sampling (LHS) points were generated in the six-dimensional continuous parameter space. To ensure maximum space-filling coverage, the sampling was optimized using the centered L2-discrepancy (CD) criterion with a random start. Second, one-factor-at-a-time (OFAT) scenarios were generated by varying each continuous feature across four non-baseline normalized levels (0, 0.25, 0.75, and 1), while fixing all remaining features at the baseline value of 0.5; because the baseline configuration (all features = 0.5) is shared across factors, it was retained only once, yielding $6 \times 4 + 1 = 25$ scenarios per ARI class. Third, an interaction-focused design was introduced to better resolve nonlinear coupling among key predictors. This component included seven physically motivated two-factor grids, namely (NO_x, O_3), (NO_x, VOCs), (NO_x, alkene%), (NO_2%, VOCs), (NO_2%, alkene%), (NO_2%, O_3), and (alkene%, O_3), each sampled on a 3×3 full-factorial grid at normalized levels of 0, 0.5, and 1, together with a 3^5 full-factorial grid spanning NO_x, NO_2%, VOCs, alkene%, and O_3 (RH fixed at 0.5) to resolve higher-order couplings among the main oxidation drivers. After removing duplicate points arising from overlap between the two-factor grids, the five-factor grid, and the shared baseline configuration, this interaction component contributed 232 unique scenarios per ARI class. Fourth, anchor scenarios were added to reinforce the boundaries of the design space, including all-low, all-mid, and all-high conditions, together with single-factor low and high perturbations for each continuous feature while all other features were held at 0.5. Because the single-factor perturbations coincide exactly with the extreme levels already sampled by the OFAT design, and the all-mid anchor coincides with the shared baseline configuration, only the all-low and all-high anchor points remained unique after deduplication, yielding 2 scenarios per ARI class. In total, 359 unique scenarios were retained for each ARI class (100 LHS + 25 OFAT + 232 interaction + 2 anchor), resulting in a primary training dataset of 1,077 scenarios. An independent validation dataset was generated separately to provide a rigorous held-out test of surrogate-model generalization. For each ARI class, 30 additional LHS points were generated using a different random seed (seed = 999) to ensure independence from the training LHS design, supplemented by OFAT scenarios evaluated at quarter-point levels (0.125 and 0.875, with the shared baseline at 0.5 excluded as a duplicate), yielding 12 OFAT-derived scenarios per class. An additional random subsample of the five-factor full-factorial grid was originally included in the validation design but was entirely removed after an explicit train/validation overlap check, since these points duplicated combinations already present in the training set. After concatenation, deduplication, and overlap removal (62 overlapping scenarios removed in total), the independent validation dataset comprised 126 scenarios (42 per ARI class: 30 LHS + 12 OFAT). Each record contains the normalized feature values, the ARI class, the scenario type, and the data-allocation flag.

For each scenario, μ_{sCIs} was simulated using the AtChem box model in constrained mode coupled with the updated MCM v3.3.1g mechanism (Sect. 2.2.1). Among all constrained variables described in Sect. 2.2.1, the seven

surrogate-model features were assigned scenario-specific values, whereas all remaining constraints were fixed at the baseline conditions. Because photolysis strongly regulates OH production and radical cycling, daytime and nighttime were treated as separate modeling problems. Accordingly, the target variable was defined as the time-averaged fractional contribution of sCIs to total gas-phase SO_2 oxidation during daytime (08:00–18:00 LT) and nighttime (22:00–05:00 LT), respectively. Separate XGBoost surrogate models were trained for the daytime and nighttime datasets. Hyperparameter optimization was performed using Optuna with a Tree-structured Parzen Estimator sampler over 100 trials. Each candidate configuration was evaluated by five-fold cross-validation using RMSE as the objective function. The final model was then retrained on the full 1,077-scenario training set, while withholding 10% of that set as an internal early-stopping monitor (early stopping rounds=50). Model generalization was assessed using the fully independent 126-scenario validation dataset. For the daytime model, μ_{sCIs} achieved a five-fold cross-validated R^2 of 0.9645 ± 0.0144. On the training set, R^2 = 0.9967, RMSE = 0.0033, and MAE = 0.002; on the independent validation set, R^2 = 0.8788, RMSE = 0.008, and MAE = 0.005. For the nighttime model, μ_{sCIs} achieved a five-fold cross-validated R^2 of 0.9776. On the training set, R^2 = 0.9969, RMSE = 0.005, and MAE = 0.003; on the independent validation set, R^2 = 0.9646, RMSE = 0.013, and MAE = 0.009. These results indicate that the surrogate models reproduced the AtChem response surfaces with high fidelity and showed no evidence of substantial overfitting.

To further examine whether the controls on total SO_2 oxidation depend on the relative importance of the sCI pathway, the full dataset for each diurnal period was stratified into low- and high-μ_{sCIs} regimes using the median μ_{sCIs} value of the corresponding training set as the regime boundary. The resulting thresholds were 0.011 for daytime and 0.079 for nighttime. Separate XGBoost models were then trained to predict total gas-phase SO_2 oxidation loss, L_{SO2}, within the low- and high-μ_{sCIs} subsets, using the same seven environmental predictors. Here μ_{sCIs} was used only for regime classification and was not included as an input feature, thereby avoiding definitional coupling between predictor and target. Hyperparameters for each regime-specific model were optimized independently with Optuna using the same five-fold cross-validation framework. The regime-specific L_{SO2} models also showed strong predictive skill. For daytime, the global model achieved a five-fold cross-validated R^2 of 0.9645 ± 0.0070 (RMSE = 163,395 ± 16,651) and a hold-out R^2 of 0.9388 (RMSE = 194,238), while the low- and high-μ_{sCIs} models achieved cross-validated R^2 of 0.9554 ± 0.0096 and 0.9456 ± 0.0208, and hold-out R^2 of 0.9037 (RMSE = 197,828) and 0.9548 (RMSE = 175,929), respectively. Given the reduced hold-out sample sizes within each regime (N = 50 for low-μ_{sCIs}, N = 76 for high-μ_{sCIs}), these hold-out estimates carry a wider margin of uncertainty than the global hold-out estimate, but remain consistent with the corresponding cross-validated performance, indicating no systematic overfitting within either regime. For nighttime, the global model achieved a five-fold cross-validated R^2 of 0.9742 ± 0.0061 (RMSE = 51,202 ± 6,680) and a hold-out R^2 of 0.9542 (RMSE = 61,912), while the low- and high-μ_{sCIs} models achieved cross-validated R^2 of 0.9350 ± 0.0276 and 0.9688 ± 0.0122, and hold-out R^2 of 0.8970 (RMSE = 60,718) and 0.9583 (RMSE = 64,926), respectively.

Section S3: Observation-based XGBoost model

To provide an independent observational check on the response patterns inferred from the box-model-based surrogate analysis, an observation-based XGBoost regression model was developed using hourly measurements at the Wuhai supersite from 9 October 2019 to 30 June 2022. The raw dataset comprised 23,880 hourly records and included $PM_{2.5}$, O_3, SO_2, NO_3^-, Fe, total VOCs, alkene fraction, wind speed (WS), RH, NO_x, NO_2 fraction, SO_4^{2-}, and μ_{sCIs}-related diagnostic information. The observed SO_4^{2-} concentration was used as the target variable. The predictor set was designed to represent the principal controls on sulfate formation, including precursor abundance, oxidation environment, aerosol loading, and meteorological conditions. Specifically, O_3, SO_2, total VOCs, alkene fraction, NO_x, and NO_2% characterize the availability of sulfate precursors and the prevailing oxidation regime; WS,

RH, and $PM_{2.5}$ represent meteorological and pollution conditions affecting dilution, transport, aerosol liquid water, and gas-particle partitioning; NO_3^- was included as an indicator of secondary inorganic aerosol formation; and Fe was included because transition metals are widely recognized as markers of, and potential catalysts for, aqueous-phase SO_2 oxidation. Prior to model training, the observational dataset was quality controlled to reduce the influence of missing values and outliers. Rows with missing target SO_4^{2-} values were first removed, reducing the sample size from 23,880 to 23,499. Short missing gaps (up to 3 h) in predictor variables were then filled by linear interpolation. Outliers were subsequently excluded using a combined criterion consisting of a physically reasonable upper bound and a 3× interquartile range filter, leaving 22,385 records. To account for short-term temporal persistence in atmospheric sulfate, a 2 h rolling mean of SO_4^{2-} based on previous observations was added as a predictor. After feature engineering and removal of remaining rows with missing predictor values, 18,766 hourly samples were retained for modeling. The final predictor set therefore comprised 12 features: $PM_{2.5}$, O_3, SO_2, NO_3^-, Fe, VOCs, alkene%, wind speed (WS), RH, NO_x, NO_2%, and the 2 h rolling mean of SO_4^{2-}. Because the observational dataset is a time series, samples were split chronologically rather than randomly. The first 80% of the processed data (15,012 samples) was used for model development, and the remaining 20% (3,754 samples) was retained as an independent test set. Hyperparameters of the XGBoost regressor were optimized on the training subset using randomized search combined with five-fold TimeSeriesSplit cross-validation, thereby preserving temporal ordering and minimizing temporal leakage. The search space included the number of trees, maximum tree depth, learning rate, subsampling ratio, column-sampling ratio, minimum child weight, gamma, and L1/L2 regularization terms. The corresponding cross-validated performance on the training period was $R^2 = 0.80$. After retraining with the optimal settings, the model achieved $R^2 = 0.8895$, RMSE = 1.0286, and MAE = 0.6089 on the training set. On the independent chronological test set, performance remained strong, with $R^2 = 0.8689$, RMSE = 1.0954, and MAE = 0.6552. These results indicate that the observation-based XGBoost model reproduced the major temporal variability in ambient sulfate reasonably well and is therefore suitable as an empirical reference for evaluating whether the feature-response structures inferred from the surrogate-model analysis are consistent with real atmospheric observations.

Section S4: Temperature dependence of sCI-driven SO_2 oxidation

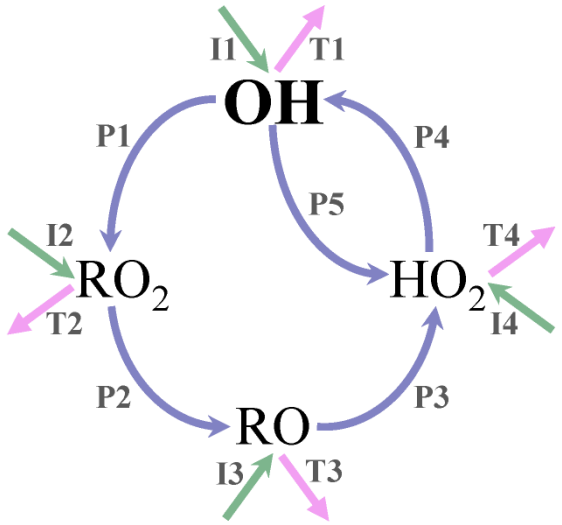
To evaluate the sensitivity of sCI-driven SO_2 oxidation to temperature, we conducted three additional sets of AtChem box-model simulations under identical precursor and meteorological perturbation schemes as in Sect. 3.1, but with the ambient temperature fixed at 280, 290, and 320 K, respectively (the main text results correspond to a reference temperature of 290 K). This range was chosen to bracket typical tropospheric conditions from cold to warm seasons/regions. All rate coefficients for sCI-related bimolecular reactions retain their IUPAC-recommended temperature dependence. In addition, because relative humidity (RH) rather than absolute water vapor concentration was held fixed across scenarios, temperature also modulates the absolute concentrations of H_2O and its dimer $(H_2O)_2$ through the temperature dependence of saturation vapor pressure—an important consideration given that reaction with $H_2O/(H_2O)_2$ is a dominant loss pathway for sCIs. The resulting differences in $L_{SO2,sCIs}$ therefore reflect the combined effect of (i) the intrinsic temperature dependence of sCI bimolecular rate coefficients and (ii) the temperature-driven change in absolute water vapor (and dimer) abundance at fixed RH, both of which act to suppress sCI-driven SO_2 oxidation as temperature increases.

The mean $L_{SO2,sCIs}$ decreases by approximately a factor of four as temperature increases from 280 to 320 K (from 8.8×10^4 to 2.2×10^4 molec. cm^{-3} s^{-1}), with the median showing a comparable reduction (from 4.8×10^4 to 8.6×10^3 molec. cm^{-3} s^{-1}). This decline is monotonic across the entire distribution, including the 5th, 25th, 75th, and 95th percentiles, indicating that the suppressive effect of higher temperature on sCI-driven oxidation is systematic rather than confined to a subset of scenarios. The standard deviation decreases proportionally with the mean (from 1.17×10^5 at 280 K to 3.47×10^4 at 320 K), so the relative variability (coefficient of variation) remains broadly stable across

the three temperatures, suggesting that temperature rescales the overall magnitude of $L_{SO2,sCIs}$ without fundamentally altering the shape of its distribution or, by extension, the relative ranking of precursor sensitivities identified in Sect. 3.1–3.3 at the reference temperature. This finding implies that the absolute magnitude of $L_{SO2,sCIs}$ and μ_{sCIs} will vary substantially across different thermal regimes (e.g., diurnal, seasonal, or latitudinal). The qualitative controls and regime-dependent sensitivity patterns identified at the reference temperature in this study should therefore be interpreted as representative of warm conditions, and quantitative extrapolation to substantially colder environments should be made with caution.

Section S5: Definition and calculation of OH_{propag} and $\sum OH_{new}$

The definitions and calculations of OH_{propag} and $\sum OH_{new}$ follow the RO_x cycling framework of Sheehy et al. (2010). A schematic of RO_x cycling is shown below. Only the predominant species involved in radical sources, sinks, and cycling are included. Radical pathways are labeled as initiation (I), propagation (P), or termination (T).



The term $\sum OH_{new}$ is defined as the sum of all new RO_x radical production (i.e., initiation), weighted by conversion factors that account for how efficiently each initiated RO_x species is converted to OH. The initiation pathways include the photolysis of HONO, O_3, and OVOCs, alkene ozonolysis reactions, and the reaction of VOCs with NO_3. Mathematically, $\sum OH_{new}$ = I(OH) + I(HO2)γHO2 + I(RO2)γROγRO2 = I1 + I4×[P4/(P4+T4)] + I2×[P2/(P2+T2)]×[P3/(P3+T3)]. The term OH_{propag} represents the OH formation associated with RO_x recycling (i.e., the propagation step in the RO_x chain rather than direct initiation). It is calculated as OH_{propag} = P4.

Table S1. Comparison of rate coefficients and total yields of sCIs and OH for O_3 + alkene reactions before and after the mechanism update.

Alkene	$k_{O3+alkene}(T)$ [c] cm^3 molec.$^{-1}$ s^{-1}	$Y_{i,sCI}$ [d]	$Y_{i,OH}$ [d]	Mechanism	Comments
ethene	6.82×10^{-15} exp(-2500/T)	0.42	0.17	MCM v3.3.1g	(a)
	9.1×10^{-15} exp(-2580/T)	0.37	0.13	MCM v3.3.1	(b)
propene	5.77×10^{-15} exp(-1880/T)	0.28	0.36	MCM v3.3.1g	(a)
	5.5×10^{-15} exp(-1880/T)	0.24	0.36	MCM v3.3.1	(b)
but-1-ene	3.55×10^{-15} exp(-1750/T)	0.23	0.38	MCM v3.3.1g	(a)
	3.55×10^{-15} exp(-1745/T)	0.24	0.36	MCM v3.3.1	(b)

cis-but-2-ene	3.37×10^{-15} exp(-970/T)	0.20	0.33	MCM v3.3.1g	(a)
	3.22×10^{-15} exp(-968/T)	0.18	0.57	MCM v3.3.1	(b)
trans-but-2-ene	7.0×10^{-15} exp(-1060/T)	0.22	0.60	MCM v3.3.1g	(a)
	6.64×10^{-15} exp(-1059/T)	0.18	0.57	MCM v3.3.1	(b)
2-methylpropene	2.92×10^{-15} exp(-1650/T)	0.20	0.67	MCM v3.3.1g	(a)
	2.7×10^{-15} exp(-1630/T)	0.18	0.82	MCM v3.3.1	(b)
isoprene	1.05×10^{-14} exp(-2000/T)	0.66[d]	0.26	MCM v3.3.1g	(a)
	1.03×10^{-14} exp(-1995/T)	0.56	0.25	MCM v3.3.1	(b)

[a] Based on Cox et al. (2020). [b] Based on Jenkin et al. (1997, 2015) and Saunders et al. (2003). [c] The applicable temperature ranges of $k_{O3+alkene}$ are 180–360 K for O_3 + ethene, 230–370 K for O_3 + propene, 220–370 K for O_3 + but-1-ene, cis-but-2-ene, trans-but-2-ene and 2-methylpropene, and 240–360 K for O_3 + isoprene. The uncertainty in *E/R*, Δ(*E/R*), is ±100 K for O_3 + ethene and O_3 + propene, and ±200 K for O_3 + but-1-ene, cis-but-2-ene, trans-but-2-ene, 2-methylpropene and isoprene. [d] Recommended total sCI yields ($Y_{i,sCI}$) and OH yields ($Y_{i,OH}$) from O_3 + alkene reactions are reported for atmospheric conditions at 298 K and 1 bar.

Table S2. Comparison of rate coefficients for bimolecular reactions of sCIs before and after the update.

	k_b (298K) cm^3 molec.$^{-1}$ s^{-1}	$k_b(T)$ cm^3 molec.$^{-1}$ s^{-1}	Mechanism	Comments
CH_2OO +SO_2		3.7×10^{-11} (T/298) $^{-2.05}$	MCM v3.3.1g	(c)
	7.0×10^{-14}		MCM v3.3.1	(b)
CH_2OO +NO_2	3×10^{-12}		MCM v3.3.1g	(a)
	1.0×10^{-15}		MCM v3.3.1	(b)
CH_2OO + H_2O	2.8×10^{-16}		MCM v3.3.1g	(a)
	1.0×10^{-17}		MCM v3.3.1	(b)
CH_2OO + $(H_2O)_2$		7.35×10^{-18} exp(4076/T)	MCM v3.3.1g	(a)
	/	/	MCM v3.3.1	(b)
CH_3CHOO [d] + SO_2	1.4×10^{-10}		MCM v3.3.1g	(a)
	7.0×10^{-14}		MCM v3.3.1	(b)
CH_3CHOO +NO_2	2.0×10^{-12}		MCM v3.3.1g	(a)
	1.0×10^{-15}		MCM v3.3.1	(b)
CH_3CHOO + H_2O	1.3×10^{-14}		MCM v3.3.1g	(a)
	1.0×10^{-17}		MCM v3.3.1	(b)
CH_3CHOO + $(H_2O)_2$	4.4×10^{-11}		MCM v3.3.1g	(a)
	/	/	MCM v3.3.1	(b)
C_2H_5CHOO [d] +SO_2	1.4×10^{-10}		MCM v3.3.1g	(a)
	7.0×10^{-14}		MCM v3.3.1	(b)
C_2H_5CHOO +NO_2	2.0×10^{-12}		MCM v3.3.1g	(a)
	1.0×10^{-15}		MCM v3.3.1	(b)
C_2H_5CHOO + H_2O	1.3×10^{-14}		MCM v3.3.1g	(a)
	1.0×10^{-17}		MCM v3.3.1	(b)
C_2H_5CHOO + $(H_2O)_2$	4.4×10^{-11}		MCM v3.3.1g	(a)
	/	/	MCM v3.3.1	(b)

$(CH_3)_2COO + SO_2$	1.55×10^{-10}		MCM v3.3.1g	(a)
	7.0×10^{-14}		MCM v3.3.1	(b)
$(CH_3)_2COO + NO_2$	2.1×10^{-12}		MCM v3.3.1g	(a)
	1.0×10^{-15}		MCM v3.3.1	(b)
$(CH_3)_2COO + H_2O$	6.0×10^{-17}		MCM v3.3.1g	(a)
	6.0×10^{-18}		MCM v3.3.1	(b)
$(CH_3)_2COO + (H_2O)_2$	6×10^{-14}		MCM v3.3.1g	(a)
	/	/	MCM v3.3.1	(b)
$(CH=CH_2)(CH_3)COO$ [d] $+SO_2$	4.2×10^{-11}		MCM v3.3.1g	(a)
	7.0×10^{-14}		MCM v3.3.1	(b)
$(CH=CH_2)(CH_3)COO + NO_2$	1.0×10^{-15}		MCM v3.3.1g	(a)
	1.0×10^{-15}		MCM v3.3.1	(b)
$(CH=CH_2)(CH_3)COO + H_2O$		$7.07 \times 10^{-19} T^{1.46} \exp(-3132/T)$	MCM v3.3.1g	(a)
	6.0×10^{-18}		MCM v3.3.1	(b)
$(CH=CH_2)(CH_3)COO + (H_2O)_2$		$7.63 \times 10^{-19} T^{1.45} \exp(-675/T)$	MCM v3.3.1g	(a)
	/	/	MCM v3.3.1	(b)
$(C(CH_3)=CH_2)CHOO$ [d] $+SO_2$	1.4×10^{-10}		MCM v3.3.1g	(a)
	7.0×10^{-14}		MCM v3.3.1	(b)
$(C(CH_3)=CH_2)CHOO + NO_2$	1.0×10^{-15}		MCM v3.3.1g	(a)
	1.0×10^{-15}		MCM v3.3.1	(b)
$(C(CH_3)=CH_2)CHOO + H_2O$		$2.93 \times 10^{-19} T^{1.66} \exp(-973/T)$	MCM v3.3.1g	(a)
	1.0×10^{-17}		MCM v3.3.1	(b)
$(C(CH_3)=CH_2)CHOO + (H_2O)_2$		$3.24 \times 10^{-19} T^{1.65} \exp(-1271/T)$	MCM v3.3.1g	(a)
	/	/	MCM v3.3.1	(b)

[a] Based on Cox et al. (2020). [b] Based on Jenkin et al. (1997, 2015) and Saunders et al. (2003). [c] Based on Onel et al. (2021). d In MCM v3.3.1g, CH_3CHOO, C_2H_5CHOO, $(CH=CH_2)(CH_3)COO$, and $(C(CH_3)=CH_2)CHOO$ specifically denote the E-stereoisomers.

Table S3. Detailed information about MCM v3.3.1 mechanism update.

MCM v3.3.1 [a]	**MCM v3.3.1g** [b]
O_3 + alkene → products	
ethene	
% 9.1D-15*EXP(-2580/TEMP): C2H4 + O3 = HCHO + CH2OOA ; % KDEC*0.37: CH2OOA = CH2OO ; % KDEC*0.50: CH2OOA = CO ; % KDEC*0.13: CH2OOA = HO2 + CO + OH ;	% 6.82D-15*EXP(-2500/TEMP): C2H4 + O3 = HCHO + CH2OOA; % KDEC*0.42: CH2OOA = CH2OO; % KDEC*0.18: CH2OOA = CO; % KDEC*0.17: CH2OOA = HO2 + CO + OH; % KDEC*0.18: CH2OOA = H2; % KDEC*0.05: CH2OOA = HO2 + HO2;

Mechanism Update Notes for ethene ozonolysis [c]: CH2OOA represents the chemically activated Criegee intermediate [CH_2OO]*. The yield of the stabilized Criegee intermediate, CH_2OO, has a preferred value of 0.42 ± 0.1 and is used to determine the branching ratio for the reaction CH2OOA = CH2OO. The preferred yields of OH and HO_2 are 0.17 ± 0.05 and approximately 0.27, respectively; these values are used to determine the branching ratios of the remaining CH_2OOA reactions. The products formed from the unimolecular decomposition of CH_2OO and [CH_2OO]* are considered to be broadly similar, including OH, HO_2, CO, and other related products. Therefore, the assigned branching ratios for the product channels (CH2OOA = CO, CH2OOA = HO2 + CO + OH, CH2OOA = H2, and CH2OOA = HO2 + HO2) are used to represent the combined unimolecular decomposition contributions from both [CH_2OO]* and CH_2OO.

propene

% 5.5D-15*EXP(-1880/TEMP)*0.5 : O3 + C3H6 = CH2OOB + CH3CHO ; % 5.5D-15*EXP(-1880/TEMP)*0.5 : O3 + C3H6 = CH3CHOOA + HCHO ; % KDEC*0.24 : CH2OOB = CH2OO ; % KDEC*0.40 : CH2OOB = CO ; % KDEC*0.36 : CH2OOB = HO2 + CO + OH ; % KDEC*0.24 : CH3CHOOA = CH3CHOO ; % KDEC*0.36 : CH3CHOOA = CH3O2 + CO + OH ;	% 5.77D-15*EXP(-1880/TEMP)*0.38 : O3 + C3H6 = CH2OOB + CH3CHO ; % 5.77D-15*EXP(-1880/TEMP)*0.62 : O3 + C3H6 = CH3CHOOA + HCHO ; % KDEC*0.53: CH2OOB = CH2OO ; % KDEC*0.13: CH2OOB = CO ; % KDEC*0.17: CH2OOB = HO2 + CO + OH ; % KDEC*0.12: CH2OOB = H2 ; % KDEC*0.05: CH2OOB = HO2 + HO2 ; % KDEC*0.08: CH3CHOOA = CH3OH + CO ; % KDEC*0.13: CH3CHOOA = CH3CHOO ; % KDEC*0.48: CH3CHOOA = HCOCH2O2 + OH ; % KDEC*0.16: CH3CHOOA = CH4 ;

Mechanism Update Notes for propene ozonolysis [c]: To distinguish the fate of [CH2OO]* in different alkene ozonolysis systems, CH2OOB is used here to represent the chemically activated Criegee intermediate [CH_2OO]*. CH3CHOOA represents the chemically activated Criegee intermediates [Z-CH_3CHOO]* and [E-CH_3CHOO]*. The reaction of propene with O_3 initially forms an energy-rich primary ozonide (POZ), which rapidly decomposes via two primary channels to produce either CH_3CHO + [CH_2OO]* or HCHO + [CH_3CHOO]*. Based on reported primary yields of CH_3CHO and HCHO, the branching ratios for these two channels are assigned as 0.38 and 0.62, respectively. The preferred total yield of stabilized Criegee intermediates, including CH_2OO, E-CH_3CHOO, and Z-CH_3CHOO, is 0.30 ± 0.10. CH_2OO accounts for approximately two-thirds of this total yield. Accordingly, the branching ratio for CH2OOB = CH2OO is assigned as 0.53 (estimated as 0.30×2/3/0.38 = 0.53). The combined yield of Z-CH_3CHOO and E-CH_3CHOO is estimated to be 0.1 (0.30 × 1/3 = 0.10). Under tropospheric conditions, Z-CH_3CHOO is expected to undergo mainly unimolecular decomposition, whereas E-CH_3CHOO is expected to be removed primarily through bimolecular reactions. Therefore, the [CH_3CHOO]* stabilization channel is used mainly to represent the formation of E-CH_3CHOO, with a minor contribution from Z-CH_3CHOO. Assuming that the combined yield of E-CH_3CHOO and a minor fraction of Z-CH_3CHOO available for bimolecular reactions is 0.08, the branching ratio for CH3CHOOA = CH3CHOO reaction is estimated as 0.08/0.62 = 0.13. The preferred total OH yield is 0.36 ± 0.04, with contributions from both the [CH_2OO]*/CH_2OO and [Z-CH_3CHOO]*/Z-CH_3CHOO pathways. The branching ratio for OH formation from CH2OOB is assumed to be the same as that used for CH2OOA.

Therefore, the OH contribution from the [CH_2OO]*/ CH_2OO pathway is 0.06, estimated as 0.36×0.17=0.06. The remaining OH yield, mainly attributed to decomposition of the [Z-CH_3CHOO]*/Z-CH_3CHOO, is 0.3, estimated as 0.36-0.06 = 0.3. This value is used to determine the branching ratio of CH3CHOOA = HCOCH2O2 + OH, giving 0.3/0.62=0.48. In addition, the decomposition pathways of [E-CH_3CHOO]*/E-CH_3CHOO are represented by the reactions CH3CHOOA = CH4 and CH3CHOOA = CH3OH + CO, with branching ratios of 0.16 and 0.08, respectively. These values are determined by reported yields of decomposition products such as CH_4 and CH_3OH.

but-1-ene

% 3.55D-15*EXP(-1745/TEMP)*0.5 : BUT1ENE + O3 = C2H5CHOOA + HCHO ; % 3.55D-15*EXP(-1745/TEMP)*0.5 : BUT1ENE + O3 = CH2OOB + C2H5CHO ; % KDEC*0.24 : C2H5CHOOA = C2H5CHOO ; % KDEC*0.36 : C2H5CHOOA = C2H5O2 + CO + OH ;	% 3.55D-15*EXP(-1750/TEMP)*0.65 : BUT1ENE + O3 = C2H5CHOOA + HCHO ; % 3.55D-15*EXP(-1750/TEMP)*0.35 : BUT1ENE + O3 = CH2OOA + C2H5CHO ; % KDEC*0.13 : C2H5CHOOA = C2H5CHOO ; % KDEC*0.49: C2H5CHOOA = PROPALO2 + OH ; % KDEC*0.08 : C2H5CHOOA = C2H5OH + CO ; % KDEC*0.15 : C2H5CHOOA = C2H6 ;

Mechanism Update Notes for but-1-ene ozonolysis [c]: The fate of [CH_2OO]* in but-1-ene ozonolysis is expected to follow the pathways in the ethene ozonolysis system; therefore, CH2OOA is used here to represent the chemically activated Criegee intermediate [CH_2OO]*. C2H5CHOOA represents the chemically activated Criegee intermediates [Z-C_2H_5CHOO]* and [E-C_2H_5CHOO]*. The reaction of but-1-ene with O_3 initially forms an energy-rich primary ozonide (POZ), which rapidly decomposes via two primary channels to produce either C_2H_5CHO + [CH_2OO]* or HCHO + [C_2H_5CHOO]*. Based on reported primary yields of C_2H_5CHO and HCHO, the branching ratios for these two channels are assigned as 0.35 and 0.65, respectively. The preferred total yield of stabilized Criegee intermediates, including CH_2OO, E-C_2H_5CHOO, and Z-C_2H_5CHOO, is 0.27. The yield of CH_2OO is estimated to be 0.35 × 0.42 = 0.15. The combined yield of Z-C_2H_5CHOO and E-C_2H_5CHOO is estimated to be 0.12, estimated as 0.27 - 0.15 = 0.12. Under tropospheric conditions, Z-C_2H_5CHOO is expected to undergo mainly unimolecular decomposition, whereas E-C_2H_5CHOO is expected to be removed primarily through bimolecular reactions. Therefore, the [C_2H_5CHOO]* stabilization channel is used mainly to represent the formation of E-C_2H_5CHOO, with a minor contribution from Z-C_2H_5CHOO. Assuming that the combined yield of E-C_2H_5CHOO and a minor fraction of Z-C_2H_5CHOO available for bimolecular reactions is 0.09, respectively, the branching ratio for the C2H5CHOOA = C2H5CHOO reaction is estimated as 0.09/0.65 ≈ 0.13. The preferred total OH yield is 0.38 ± 0.18, with contributions from both the [CH_2OO]*/CH_2OO and [Z-C_2H_5CHOO]*/Z-C_2H_5CHOO pathways. Therefore, the OH contribution from the [CH_2OO]*/ CH_2OO pathway is 0.06, estimated as 0.35 × 0.17 = 0.06. The remaining OH yield, mainly attributed to decomposition of the [Z-C_2H_5CHOO]*/Z-C_2H_5CHOO, is 0.32, estimated as 0.38 - 0.06 = 0.32. This value is used to determine the branching ratio of C2H5CHOOA = PROPALO2 + OH, giving 0.32/0.65 = 0.49. In addition, the decomposition pathways of [E-C_2H_5CHOO]*/E-C_2H_5CHOO are represented by the reactions C2H5CHOOA = C2H6 and C2H5CHOOA = C2H5OH + CO, with branching ratios of 0.15 and 0.08, respectively.

***cis*-but-2-ene**	
% 3.22D-15*EXP(-968/TEMP) : CBUT2ENE + O3 = CH3CHO + CH3CHOOB ; % KDEC*0.18 : CH3CHOOB = CH3CHOO ; % KDEC*0.57 : CH3CHOOB = CH3O2 + CO + OH ; % KDEC*0.125 : CH3CHOOB = CH3O2 + HO2 ; % KDEC*0.125 : CH3CHOOB = CH4 ;	% 3.37D-15*EXP(-970/TEMP) : CBUT2ENE + O3 = CH3CHO + CH3CHOOB ; % KDEC*0.20 : CH3CHOOB = CH3CHOO ; % KDEC*0.33 : CH3CHOOB = HCOCH2O2 + OH ; % KDEC*0.10 : CH3CHOOB = CH3OH + CO ; % KDEC*0.30 : CH3CHOOB = CH4 ;
Mechanism Update Notes for *cis*-but-2-ene ozonolysis [c]: To distinguish the fate of [CH_3CHOO]* in different alkene ozonolysis systems, CH3CHOOB is used here to represent the chemically activated Criegee intermediate [CH_3CHOO]*. The preferred total yield of stabilized Criegee intermediates, including E-CH_3CHOO and Z-CH_3CHOO, is 0.38 ± 0.10. Under tropospheric conditions, Z-CH_3CHOO is expected to undergo mainly unimolecular decomposition, whereas E-CH_3CHOO is expected to be removed primarily through bimolecular reactions. Therefore, the [CH_3CHOO]* stabilization channel is used mainly to represent the formation of E-CH_3CHOO, with a minor contribution from Z-CH_3CHOO. Based on the reported CH_3CHO yields in the presence of H_2O, the combined yield of E-CH_3CHOO and a minor fraction of Z-CH_3CHOO available for bimolecular reactions is estimated to be 0.20. Accordingly, the branching ratio for the CH3CHOOB = CH3CHOO reaction is assigned as 0.20. The preferred total OH yield is 0.33 ± 0.07, arising mainly from decomposition of [Z-CH_3CHOO]*/Z-CH_3CHOO. This value is used to determine the branching ratio of CH3CHOOB = HCOCH2O2 + OH, giving 0.33. In addition, the decomposition pathways of [E-CH_3CHOO]*/E-CH_3CHOO are represented by the reactions CH3CHOOB = CH4 and CH3CHOOB = CH3OH + CO, with branching ratios of 0.30 and 0.10, respectively. These values are determined by reported yields of decomposition products such as CH_4 and CH_3OH.	
***trans*-but-2-ene**	
% 6.64D-15*EXP(-1059/TEMP) : TBUT2ENE + O3 = CH3CHO + CH3CHOOB ; % KDEC*0.18 : CH3CHOOB = CH3CHOO ; % KDEC*0.57 : CH3CHOOB = CH3O2 + CO + OH ; % KDEC*0.125 : CH3CHOOB = CH3O2 + HO2 ; % KDEC*0.125 : CH3CHOOB = CH4 ;	%7.0D-15*EXP(-1060/TEMP) : TBUT2ENE + O3 = CH3CHO + CH3CHOOC; % KDEC*0.22 : CH3CHOOC = CH3CHOO ; % KDEC*0.60 : CH3CHOOC = HCOCH2O2 + OH ; % KDEC*0.07 : CH3CHOOC = CH3OH + CO ; % KDEC*0.11 : CH3CHOOC = CH4 ;
Mechanism Update Notes for *trans*-but-2-ene ozonolysis [c]: To distinguish the fate of [CH_3CHOO]* in different alkene ozonolysis systems, CH3CHOOC is used here to represent the chemically activated Criegee intermediate [CH_3CHOO]*. The preferred total yield of stabilized Criegee intermediates, including E-CH_3CHOO and Z-CH_3CHOO, is 0.43 ± 0.10. Under tropospheric conditions, Z-CH_3CHOO is expected to undergo mainly unimolecular decomposition, whereas E-CH_3CHOO is expected to be removed primarily through bimolecular reactions. Therefore, the [CH_3CHOO]* stabilization channel is used mainly to represent the formation of E-CH_3CHOO, with a minor contribution from Z-CH_3CHOO. Based on the reported CH_3CHO yields in the presence of H_2O, the combined yield of E-CH_3CHOO and a minor fraction of Z-CH_3CHOO available for bimolecular reactions is estimated to be 0.22. Accordingly, the branching ratio for the CH3CHOOC = CH3CHOO reaction is assigned as 0.22. The preferred total OH yield is 0.60 ± 0.06, arising mainly from decomposition of [Z-CH_3CHOO]* and Z-CH_3CHOO. This value is used to determine the branching ratio of CH3CHOOC = HCOCH2O2 + OH, giving 0.60. In	

addition, the decomposition pathways of [E-CH_3CHOO]*/E-CH_3CHOO are represented by the reactions CH3CHOOC = CH4 and CH3CHOOC = CH3OH + CO, with branching ratios of 0.11 and 0.07, respectively. These values are determined by reported yields of decomposition products such as CH_4 and CH_3OH.

2-methylpropene	
%2.7D-15*EXP(-1630/TEMP)*0.5 : MEPROPENE + O3 = CH2OOC + CH3COCH3 ; %2.7D-15*EXP(-1630/TEMP)*0.5 : MEPROPENE + O3 = CH3CCH3OOA + HCHO ; % KDEC*0.18: CH2OOC = CH2OO ; % KDEC*0.82 : CH2OOC = HO2 + CO + OH ; % KDEC*0.18:CH3CCH3OOA = CH3CCH3OO ; % KDEC*0.82:CH3CCH3OOA = CH3COCH2O2 + OH ;	% 2.92D-15*EXP(-1650/TEMP)*0.32 : MEPROPENE + O3 = CH2OOA + CH3COCH3 ; % 2.92D-15*EXP(-1650/TEMP)*0.68 : MEPROPENE + O3 = CH3CCH3OOA + HCHO ; % KDEC*0.10: CH3CCH3OOA = CH3CCH3OO ; % KDEC*0.90: CH3CCH3OOA = CH3COCH2O2 + OH ;
Mechanism Update Notes for 2-methylpropene ozonolysis [c]: The fate of [CH_2OO]* formed in 2-methylpropene ozonolysis is expected to follow the pathways in the ethene ozonolysis system; therefore, CH2OOA is used here to represent the chemically activated Criegee intermediate [CH_2OO]*. CH3CCH3OOA is used here to represent the chemically activated Criegee intermediates [$(CH_3)_2COO$]*. The reaction of 2-methylpropene with O_3 initially forms an energy-rich primary ozonide (POZ), which rapidly decomposes via two primary channels to produce either $CH_3C(O)CH_3$ + [CH_2OO]* or HCHO + [$(CH_3)_2COO$]*. Based on reported primary yields of $CH_3C(O)CH_3$ and HCHO, the branching ratios for these two channels are assigned as 0.32 and 0.68, respectively. The preferred total yield of stabilized Criegee intermediates, including CH_2OO and $(CH_3)_2COO$, is 0.21 ± 0.05. The yield of CH_2OO is estimated as 0.32 × 0.42 = 0.13. The yield of $(CH_3)_2COO$, accounting for about 30 % of the total sCI yield, is estimated to be 0.068. Accordingly, the branching ratio for the CH3CCH3OOA = CH3CCH3OO reaction is estimated as 0.068/0.68 = 0.10. [$(CH_3)_2COO$]* is expected to undergo mainly unimolecular decomposition, with only a minor fraction stabilized to form $(CH_3)_2COO$. The preferred total OH yield is 0.69, with contributions from both the [CH_2OO]*/CH_2OO and [$(CH_3)_2COO$]*/$(CH_3)_2COO$ pathways. The minor OH contribution from the [CH_2OO]*/ CH_2OO pathway is 0.05, estimated as 0.32 × 0.17 = 0.05. The remaining OH yield, 0.64, is mainly attributed to decomposition of [$(CH_3)_2COO$]* and $(CH_3)_2COO$. This value is used to determine the branching ratio of CH3CCH3OOA = CH3COCH2O2 + OH, giving 0.64/0.68 ≈ 0.90.	
isoprene	
% 1.03D-14*EXP(-1995/TEMP)*0.3 : O3 + C5H8 = CH2OOE + MACR ; % 1.03D-14*EXP(-1995/TEMP)*0.2 : O3 + C5H8 = CH2OOE + MVK ; % 1.03D-14*EXP(-1995/TEMP)*0.3 : O3 + C5H8 = HCHO + MACROOA ;	% 1.05D-14*EXP(-2000/TEMP)*0.41 : O3 + C5H8 = CH2OOE + MACR ; % 1.05D-14*EXP(-2000/TEMP)*0.17 : O3 + C5H8 = CH2OOE + MVK ; % 1.05D-14*EXP(-2000/TEMP)*0.19 : O3 + C5H8 = HCHO + MACROOA ;

% 1.03D-14*EXP(-1995/TEMP)*0.2 : O3 + C5H8 = HCHO + MVKOOA ; % KDEC*0.095 : MACROOA = C3H6 ; % KDEC*0.095 : MACROOA = CH3CO3 + HCHO + HO2 ; % KDEC*0.56 : MACROOA = MACROO ; % KDEC*0.25 : MACROOA = OH + CO + CH3CO3 + HCHO ; % KDEC*0.095 : MVKOOA = C3H6 ; % KDEC*0.095 : MVKOOA = CH3O2 + HCHO + CO + HO2 ; % KDEC*0.56 : MVKOOA = MVKOO ; % KDEC*0.25 : MVKOOA = OH + MVKO2 ; % KDEC*0.56 : CH2OOE = CH2OO ; % KDEC*0.19 : CH2OOE = CO ; % KDEC*0.25 : CH2OOE = HO2 + CO + OH ;	% 1.05D-14*EXP(-2000/TEMP)*0.23 : O3 + C5H8 = HCHO + MVKOOA ; % KDEC*0.10 : MACROOA = C3H6 ; % KDEC*0.11 : MACROOA = CH3CO3 + HCHO + HO2 ; % KDEC*0.16 : MACROOA = MACROO ; % KDEC*0.63 : MACROOA = OH + CO + CH3CO3 + HCHO ; % KDEC*0.08 : MVKOOA = C3H6 ; % KDEC*0.09 : MVKOOA = CH3O2 + HCHO + CO + HO2 ; % KDEC*0.33 : MVKOOA = MVKOO ; % KDEC*0.50 : MVKOOA = OH + MVKO2 ; % KDEC*0.95 : CH2OOE = CH2OO ; % KDEC*0.05 : CH2OOE = HO2 + CO + OH ;
Mechanism Update Notes for isoprene ozonolysis [c]: To distinguish the fate of $[CH_2OO]^*$ in different alkene ozonolysis systems, CH2OOE is used here to represent the chemically activated Criegee intermediate $[CH_2OO]^*$. MACROOA and MVKOOA are used to represent the chemically activated Criegee intermediate $[(C(CH_3)=CH_2)CHOO]^*$ and $[(CH=CH_2)(CH_3)COO]^*$, respectively. The reaction of isoprene with O_3 initially forms a pair of energy-rich primary ozonides, POZ-1 and POZ-2. POZ-1 rapidly decomposes via two primary channels to produce either methyl vinyl ketone + $[CH_2OO]^*$ or HCHO + $[(CH=CH_2)(CH_3)COO]^*$. Based on reported primary yields of methyl vinyl ketone and HCHO, the branching ratios for these two channels are assigned as 0.17 and 0.23, respectively. POZ-2 rapidly decomposes via two primary channels to produce either methacrolein + $[CH_2OO]^*$ or HCHO + $[(C(CH_3)=CH_2)CHOO]^*$. Based on reported primary yields of methacrolein and HCHO, the branching ratios for these two channels are assigned as 0.41 and 0.19, respectively. The preferred total yield of stabilized Criegee intermediates, including CH_2OO, E-$(CH=CH_2)(CH_3)COO$, Z-$(CH=CH_2)(CH_3)COO$, E-$(C(CH_3)=CH_2)CHOO$ and Z-$(C(CH_3)=CH_2)CHOO$, is 0.65 ± 0.10. The calculated yield of CH_2OO is 0.55. Accordingly, the branching ratio for the CH_2OOE = CH_2OO reaction is estimated as 0.55/(0.41+0.17) = 0.95. The calculated yields of E-$(CH=CH_2)(CH_3)COO$, Z-$(CH=CH_2)(CH_3)COO$, E-$(C(CH_3)=CH_2)CHOO$ and Z-$(C(CH_3)=CH_2)CHOO$ are 0.075, 0.032, 0.031, and 0.014, respectively. Under tropospheric conditions, Z-$(CH=CH_2)(CH_3)COO$ and Z-$(C(CH_3)=CH_2)CHOO$ are expected to undergo mainly unimolecular decomposition. Therefore, the $[(CH=CH_2)(CH_3)COO]^*$ and $[(C(CH_3)=CH_2)CHOO]^*$ stabilization channels are used mainly to represent the formation of E-$(CH=CH_2)(CH_3)COO$ and E-$(C(CH_3)=CH_2)CHOO$, respectively. Accordingly, the branching ratio for the MVKOOA = MVKOO reaction is assigned as 0.075/0.23=0.33, and that for the MACROOA = MACROO reaction is assigned as 0.031/0.19=0.16. The decomposition pathways of E-$(CH=CH_2)(CH_3)COO$ and E-$(C(CH_3)=CH_2)CHOO$ are represented by incorporating the reaction list in the subsequent table, including 5.13D+01 : MVKOO = and 3.02D+01 : MACROO = MVKOO = . The preferred total OH yield is 0.26 ± 0.04, arising mainly from decomposition of the C4 Criegee intermediates. The OH yield from the $[CH_2OO]^*/CH_2OO$ pathway is estimated as (0.41+0.17) ×0.05=0.03. The remaining OH yield, mainly attributed to decomposition of $[Z\text{-}(C(CH_3)=CH_2)CHOO]^*$, Z-$(C(CH_3)=CH_2)CHOO$, $[Z\text{-}(CH=CH_2)(CH_3)COO]^*$ and Z-$(CH=CH_2)(CH_3)COO$, is estimated as 0.26	

- 0.03 = 0.23. Accordingly, the branching ratios for the MACROOA = OH + CO + CH3CO3 + HCHO and MVKOOA = OH + MVKO2 reactions are assigned as 0.63 and 0.50, respectively.	
Thermal reactions of stabilized Criegee intermediates	
CH_2OO [d]	
% 1.20D-15 : CH2OO + CO = HCHO ; % 1.00D-14 : CH2OO + NO = HCHO + NO2 ; % 1.00D-15 : CH2OO + NO2 = HCHO + NO3 ; % 7.00D-14 : CH2OO + SO2 = HCHO + SO3 ; % 6.00D-18*H2O : CH2OO = HCHO + H2O2 ; % 1.00D-17*H2O : CH2OO = HCOOH ;	% 1.20D-15 : CH2OO + CO = HCHO ; % 1.00D-14 : CH2OO + NO = HCHO + NO2 ; % 3.00D-12 : CH2OO + NO2 = HCHO + NO3 ; % KC2CISO2 : CH2OO + SO2 = HCHO + SO3 ; % 2.80D-16*0.5*H2O : CH2OO = HCHO + H2O2 ; % 2.80D-16*0.5*H2O : CH2OO = HCOOH ; % 8.45D-41*EXP(5625/TEMP)*0.5*H2O*H2O : CH2OO = HCHO + H2O2 ; % 8.45D-41*EXP(5625/TEMP)*0.5*H2O*H2O : CH2OO = HCOOH ; Complex reactions: KC2CISO2 = 3.72D-11*(TEMP/298)@(-2.05) ;
CH_3CHOO [d]	
% 1.20D-15 : CH3CHOO + CO = CH3CHO ; % 1.00D-14 : CH3CHOO + NO = CH3CHO + NO2 ; % 1.00D-15 : CH3CHOO + NO2 = CH3CHO + NO3 ; % 7.00D-14 : CH3CHOO + SO2 = CH3CHO + SO3 ; % 6.00D-18*H2O : CH3CHOO = CH3CHO + H2O2 ; % 1.00D-17*H2O : CH3CHOO = CH3CO2H ;	% 1.20D-15 : CH3CHOO + CO = CH3CHO ; % 1.00D-14 : CH3CHOO + NO = CH3CHO + NO2 ; % 2.00D-12 : CH3CHOO + NO2 = CH3CHO + NO3 ; % 1.40D-10 : CH3CHOO + SO2 = CH3CHO + SO3 ; % 1.30D-14*0.5*H2O : CH3CHOO = CH3CHO + H2O2 ; %1.30D-14*0.5*H2O : CH3CHOO = CH3CO2H ; % 5.06D-34*EXP(1549/TEMP)*0.5*H2O*H2O : CH3CHOO = CH3CHO + H2O2 ; % 5.06D-34*EXP(1549/TEMP)*0.5*H2O*H2O : CH3CHOO = CH3CO2H ;
C_2H_5CHOO [d]	
% 1.20D-15 : C2H5CHOO + CO = C2H5CHO ; % 1.00D-14 : C2H5CHOO + NO = C2H5CHO + NO2 ; % 1.00D-15 : C2H5CHOO + NO2 = C2H5CHO + NO3 ; % 7.00D-14 : C2H5CHOO + SO2 = C2H5CHO + SO3 ; % 6.00D-18*H2O : C2H5CHOO = C2H5CHO + H2O2 ;	% 1.20D-15 : C2H5CHOO + CO = C2H5CHO ; % 1.00D-14 : C2H5CHOO + NO = C2H5CHO + NO2 ; % 2.00D-12 : C2H5CHOO + NO2 = C2H5CHO + NO3 ; % 1.40D-10 : C2H5CHOO + SO2 = C2H5CHO + SO3 ; % 1.30D-14*0.5*H2O : C2H5CHOO = C2H5CHO + H2O2 ;

% 1.00D-17*H2O : C2H5CHOO = PROPACID ;	% 1.30D-14*0.5*H2O : C2H5CHOO = PROPACID ; % 5.06D-34*EXP(1549/TEMP)*0.5*H2O*H2O : C2H5CHOO = C2H5CHO + H2O2 ; % 5.06D-34*EXP(1549/TEMP)*0.5*H2O*H2O : C2H5CHOO = PROPACID ;
$(CH_3)_2COO$	
% 1.20D-15: CH3CCH3OO + CO = CH3COCH3 ; % 1.00D-14 : CH3CCH3OO + NO = CH3COCH3 + NO2 ; %1.00D-15 : CH3CCH3OO + NO2 = CH3COCH3 + NO3 ; % 7.00D-14 : CH3CCH3OO + SO2 = CH3COCH3 + SO3 ; % 6.00D-18*H2O : CH3CCH3OO = CH3COCH3 + H2O2 ;	% 1.20D-15 : CH3CCH3OO + CO = CH3COCH3 ; % 1.00D-14 : CH3CCH3OO + NO = CH3COCH3 + NO2 ; % 2.10D-12 : CH3CCH3OO + NO2 = CH3COCH3 + NO3 ; % 1.55D-10 : CH3CCH3OO + SO2 = CH3COCH3 + SO3 ; % 6.00D-17*H2O : CH3CCH3OO = CH3COCH3 + H2O2 ; % 6.90D-37*EXP(1549/TEMP)*H2O*H2O : CH3CCH3OO = CH3COCH3 + H2O2 ; % 4.00D+02 : CH3CCH3OO = ;
$(C(CH_3)=CH_2)CHOO$ [e]	
% 1.2D-15 : MACROO + CO = MACR ; % 1.0D-14 : MACROO + NO = MACR + NO2 ; % 1.0D-15 : MACROO + NO2 = MACR + NO3 ; % 7.0D-14 : MACROO + SO2 = MACR + SO3 ; % 1.0D-17*H2O : MACROO = MACO2H ; % 6.0D-18*H2O : MACROO = MACR + H2O2 ;	% 1.2D-15 : MACROO + CO = MACR ; % 1.0D-14 : MACROO + NO = MACR + NO2 ; % 1.0D-15 : MACROO + NO2 = MACR + NO3 ; % 1.4D-10 : MACROO + SO2 = MACR + SO3 ; % KC4EH2O2*0.5*H2O : MACROO = MACO2H ; % KC4EH2O2*0.5*H2O : MACROO = MACR + H2O2 ; % KC4EH2O22*0.5*H2O*H2O : MACROO = MACO2H ; % KC4EH2O22*0.5*H2O*H2O : MACROO = MACR + H2O2 ; % 3.02D+01 : MACROO = ; Complex reactions: KC4EH2O2 = 2.93D-19*TEMP@1.66*EXP(-973/TEMP) ; KC4EH2O22 = 3.72D-42*TEMP@1.65*EXP(2820/TEMP) ;
$(CH=CH_2)(CH_3)COO$ [e]	
% 1.2D-15 : MVKOO + CO = MVK ; % 1.0D-14 : MVKOO + NO = MVK + NO2 ; % 1.0D-15 : MVKOO + NO2 = MVK + NO3 ; % 7.0D-14 : MVKOO + SO2 = MVK + SO3 ;	% 1.2D-15 : MVKOO + CO = MVK ; % 1.0D-14 : MVKOO + NO = MVK + NO2 ; % 1.0D-15 : MVKOO + NO2 = MVK + NO3 ; % 4.2D-11 : MVKOO + SO2 = MVK + SO3 ;

% 6.0D-18*H2O : MVKOO = MVK + H2O2 ;	% KC4EH2O1*H2O : MVKOO = MVK + H2O2 ; % KC4EH2O21*H2O*H2O : MVKOO = MVK + H2O2 ; % 5.13D+01 : MVKOO = ; Complex reactions: KC4EH2O1 = 7.07D-19*TEMP@1.46*EXP(-3132/TEMP) ; KC4EH2O21 = 8.77D-42*TEMP@1.45*EXP(874/TEMP) ;

[a] The MCM v3.3.1 chemical mechanism is provided in FACSIMILE format and can be directly implemented in AtChem2. Chemical reactions in FACSIMILE format are written as: % k : A + B = C + D; where k denotes the rate coefficient, A and B are the reactants, C and D are the products. The rate coefficient k is either a constant or calculated as a function of temperature (TEMP) and water vapor concentration (H_2O). [b] The MCM v3.3.1g chemical mechanism is provided in FACSIMILE format and can be directly implemented in AtChem2. [c] The relative contributions of the different reaction pathways for each alkene ozonolysis system were based on the information reported in Supplement A of Cox et al. (2020). [d] The unimolecular decomposition pathway of CH_2OO is considered together with that of [CH_2OO]*, with their contribution determined according to the OH yield; therefore, the unimolecular decomposition reaction of CH_2OO is not explicitly included. CH_3CHOO represents only E-CH_3CHOO. Its unimolecular decomposition pathway is considered together with that of [E-CH_3CHOO]*, with their contribution determined according to the yields of their decomposition products. [Z-CH_3CHOO]* and Z-CH_3CHOO are assumed to predominantly undergo unimolecular decomposition, with their contribution determined from the OH yield. Therefore, the unimolecular decomposition reaction of E-CH_3CHOO and the bimolecular reactions of Z-CH_3CHOO are not explicitly included. [e] E-($C(CH_3)=CH_2$)CHOO and E-($CH=CH_2$)(CH_3)COO both refer to the E-stereoisomers. The corresponding Z-stereoisomeric conformers are assumed to predominantly undergo unimolecular decomposition, and this contribution is included in the Criegee intermediates unimolecular decomposition section based on the OH yield. Therefore, the bimolecular reactions of Z-($C(CH_3)=CH_2$)CHOO and Z-($CH=CH_2$)(CH_3)COO are not explicitly included.

Table S4. Summary statistics of all variables in the long-term observational dataset from 9 October 2019 to 30 June 2022.

	Average	Std. dev.	Minimum	5th perc.	25th perc.	Median	75th perc.	95th perc.	Maximum
$PM_{2.5}$	43.2	54.1	1.0	8.0	17.0	29.0	51.0	116.0	985.0
O_3	62.9	42.2	1.0	4.0	25.0	64.0	92.0	133.1	254.0
SO_2	26.9	39.2	1.0	2.0	5.0	12.0	33.0	101.0	502.0
NO_3^-	3.2	5.2	0.0	0.0	0.4	1.1	3.6	14.0	58.3
Fe	985.8	3330.6	0.0	91.5	274.6	512.0	973.3	2710.5	222439.7
VOCs	90.9	393.9	2.0	20.8	38.0	61.4	105.6	214.2	40500.9
alkene%	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.2	0.7
NO_x	53.7	64.7	1.0	5.0	12.0	29.0	73.0	179.0	768.0
NO_2%	0.8	0.2	0.0	0.3	0.7	0.9	0.9	1.0	1.0
SO_4^{2-}	4.7	8.0	0.0	0.0	1.3	2.4	5.1	16.1	184.0
WS	1.9	1.0	0.0	0.5	1.1	1.7	2.5	3.7	8.7
RH	37.6	20.0	0.0	11.0	22.0	34.0	50.0	76.0	100.0

Figure S1. Correspondence between normalized or encoded input values and their original values.

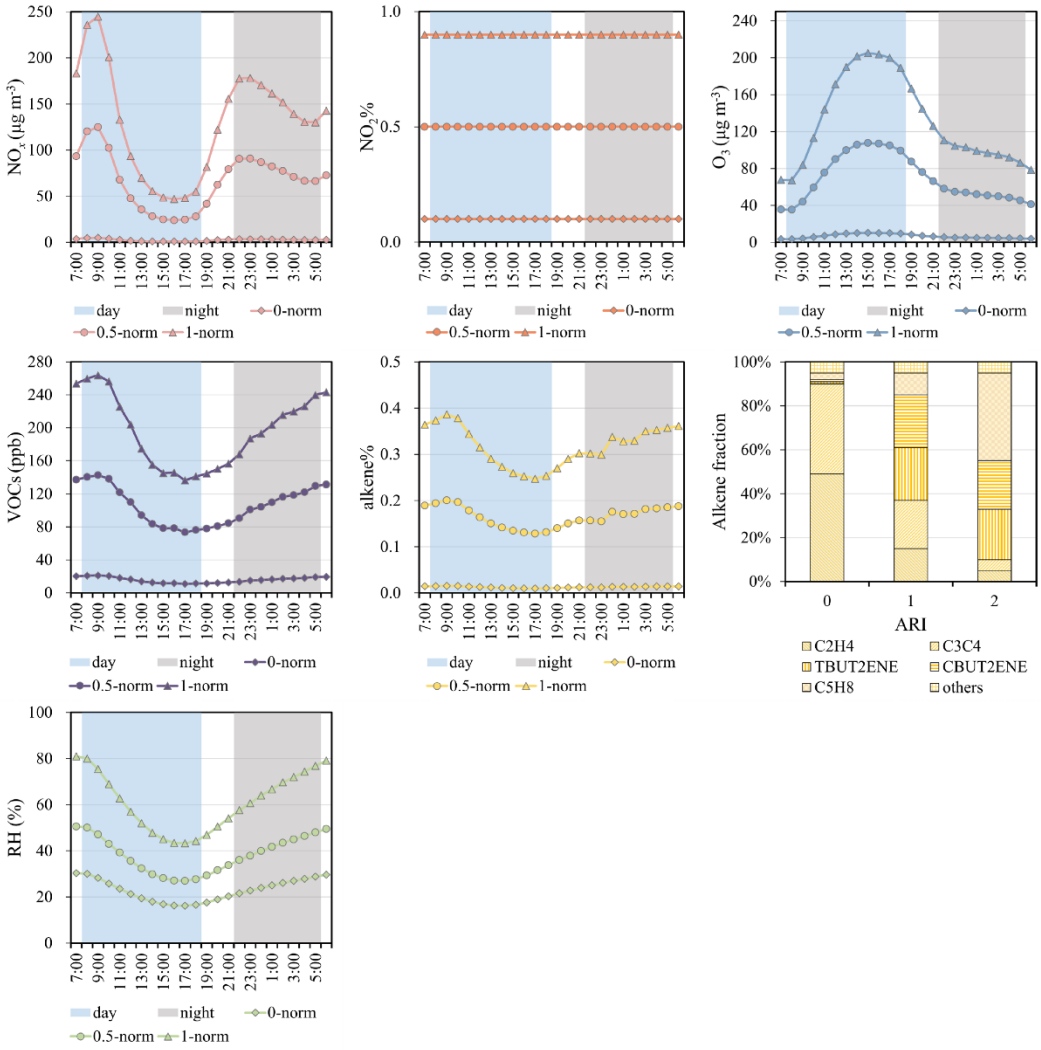


Figure S2. Baseline settings of non-feature variables

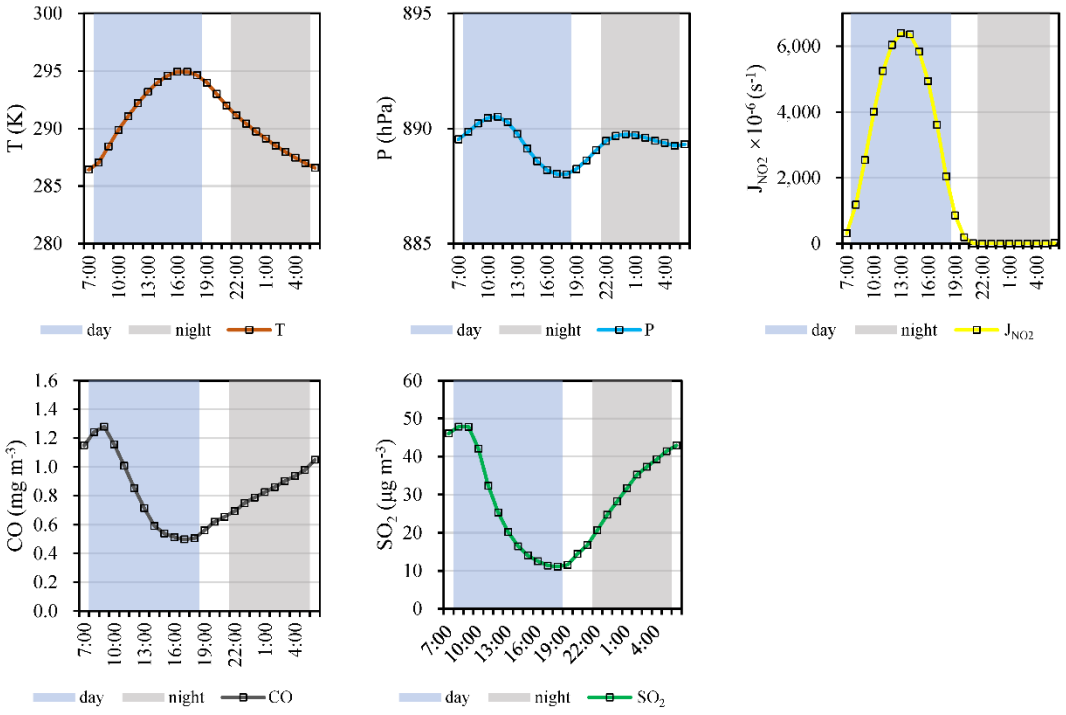


Figure S3. Time series of pollutant mixing ratios and meteorological parameters measured at the Wuhai City Atmospheric Environment Super Monitoring Station from 1 June to 15 July 2021.