npj | clean air

Article

https://doi.org/10.1038/s44407-025-00041-6

Ambient measurements and improved statistical proxies for gaseous sulfuric acid in coastal Hong Kong

Yi Chen[1,2], Penggang Zheng[1], Xin Feng[1], Yan Tan[3,4], Jiali Zhong[1], Tao Wang[3], Shuncheng Lee[5], Jian Zhen Yu[1,2] & Zhe Wang[1] ✉

Gaseous sulfuric acid (SA) is a critical precursor for atmospheric new particle formation and growth. However, its direct measurement remains challenging due to low concentration and fast dynamics. Chemical Ionization Mass Spectroscopy (CIMS) has been developed as an effective technique for SA measurement, but its availability, operational complexity, and complex calibration procedures hinder widespread and long-term observations of SA. To address this gap, proxy models have been developed to estimate SA from routinely measured parameters, but their transferability across regions remains uncertain, underscoring the need for localized improvements. In this study, SA and related parameters were measured at two coastal background sites in Hong Kong, Cape D'Aguilar Supersite in 2018 and the Hong Kong University of Science and Technology Supersite in 2022. Observed SA concentrations showed distinct diurnal patterns, peaking at midday with maxima of $(8.0 \pm 7.5) \times 10^6$ and $(4.2 \pm 2.9) \times 10^6$ molecule/cm³, respectively. We evaluated a series of existing proxies based on SO_2, O_3, alkenes, solar radiation, and condensation sink (CS), and found limited applicability under subtropical conditions. Therefore, new statistical proxies were developed, and locally optimized nonlinear proxies slightly outperformed linear ones, and replacing CS with $PM_{2.5}$ yielded comparable accuracy, greatly enhancing practicality where size-resolved particle data are unavailable. For whole-day estimation, nighttime SA formation was further investigated through box model simulations. Results revealed that regenerated OH radicals ($\sim 4–6 \times 10^5$ molecule/cm³), rather than stabilized Criegee intermediates, played a dominant role in nighttime SA production. Thus, we established new nighttime and piecewise full-day proxies that reproduced observed SA concentrations across both sites. Overall, this work advances understanding of SA dynamics in subtropical coastal environments, provides improved proxies adaptable to data-limited regions, and demonstrates their potential application for long-term monitoring and air quality modeling worldwide.

Sulfuric acid (H_2SO_4, SA) plays a crucial role in atmospheric new particle formation due to its extremely low volatility, strong hygroscopicity, and capacity for hydrogen bonding[1–4]. Current understanding suggests that gaseous SA participates in nucleation through binary or ternary mechanisms, interacting with water, ammonia, amines, and low-volatility organic compounds[5–8]. The nucleation rates are positively correlated with SA concentration, highlighting the importance of accurate SA measurements to quantify particle formation processes[2,9,10]. In addition to initiating nucleation, SA also contributes to aerosol growth through condensation on existing particles[10,11]. Therefore, a comprehensive understanding of its distribution and formation in the atmosphere is essential for a thorough investigation of new particle formation, aerosol growth, and sulfur chemistry.

---

[1]Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong SAR, 999077, China. [2]Department of Chemistry, The Hong Kong University of Science and Technology, Hong Kong SAR, 999077, China. [3]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, 999077, China. [4]School of Environmental and Municipal Engineering, Qingdao University of Technology, Qingdao, 266520, China. [5]Thrust of Sustainable Energy and Environment, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, 511458, China. ✉e-mail: z.wang@ust.hk

While primary emissions contribute to SA in environments such as volcano, tunnels or roadside areas, secondary formation dominates SA in most regions[12]. Precursors of SA originate from both natural and anthropogenic sources, such as ocean emissions, industrial activity, and shipping. SO_2, from either dimethyl sulfide oxidation or primary emission, can be oxidized to SO_3, which reacts rapidly with water to produce SA[13–16]. During the day, OH radicals drive this oxidation, whereas at night, stabilized Criegee Intermediates (SCIs), and in some cases, residual OH radicals play important roles[17,18]. SA is removed from the atmosphere through nucleation, condensation on pre-existing particles, and deposition, with condensation generally recognized as the dominant sink[17,19,20]. Direct measurement of gaseous SA is technically challenging due to its low concentrations and fast formation and loss dynamics. Nitrate-based Chemical Ionization Mass Spectroscopy (CIMS) is the most widely used technique for detecting atmospheric SA, which is ionized and detected as ions of HSO_4^-, $H_2SO_4NO_3^-$, and $H_2SO_4HNO_3NO_3^-$ [21,22]. Calibration is typically performed using in-situ generated SA via reaction of SO_2 with certain OH radicals, but this process remains challenging due to uncertainties in OH radical generation, wall losses, and ion-molecule reaction efficiencies. Operational complexity, instrument sensitivity, and the need for regular calibration limit the deployment of CIMS for long-term or widespread monitoring. Field studies in forest, rural, and urban areas have shown that SA concentrations peak during the day, typically ranging from 10^6 to 10^8 molecules/cm^3 [17,20,23,24]. Nighttime levels, though lower, remain relatively stable in the range of 10^4 to 10^6 molecules/cm^3 [17,20,23,24]. In exceptional cases, such as during volcanic eruptions, SA concentrations can reach up to 10^9 molecules/cm^3 due to elevated emissions of SA and SO_2[25]. Human activities increase SO_2 emissions and, by extension, SA formation, but higher particle concentrations also accelerate its removal, creating a non-linear relationship between SA concentration and pollution level[24].

To overcome the limitations of direct measurements, proxy models have been developed to estimate gaseous SA concentrations based on the formation and loss pathways and routinely monitored atmospheric variables, such as SO_2 concentration, global solar radiation (SR) or ultraviolet radiation B (UVB), relative humidity (RH), and the condensation sink (CS)[17,20,24,26]. Earlier work by Petäjä et al.[20] first proposed such proxies using data measured from the Hyytiälä SMEAR II station in Finland, and exploring different parameterizations with OH radicals, UVB, and SR. Mikkonen et al.[24] expanded this approach by combining data from multiple locations and periods to develop a more generalizable model. However, questions about their applicability in other atmospheric settings are raised due to the environment differential. In particular, the relative importance of key factors, such as SO_2, CS, and RH, differs across environments, highlighting the need for localized optimization. The proxy represented by Eq. 1 has been widely used in subsequent studies to localize parameters for daytime SA vapor, where *a* denotes an apparent coefficient and *k* represents the temperature-dependent reaction rate (cm^3/molecule/s).

$$SA = a \times k \times SR^b \times SO_2{}^c \times CS^d \times RH^e \tag{1}$$

Compared to the daytime proxies, nighttime SA proxies have received relatively little attention. Guo et al.[18] emphasized the important role of SCIs in nighttime SA production in Beijing. Dada et al.[17] incorporated both OH and SCIs (represented by $O_3 \times$ alkenes) as oxidants, alongside CS and nucleation sink, and established a proxy that can estimate SA formation in daytime and nighttime at Beijing and Hyytiälä. Although nighttime SA concentrations are typically lower than daytime levels, understanding and accurate representation of nighttime SA remains crucial for comprehending the fate and atmospheric impacts.

Despite a growing body of research, existing measurements are insufficient to characterize SA's global distribution or temporal variability. SA measurements in coastal regions where anthropogenic emissions interact with ship emissions and oceanic species remain limited. Furthermore, few studies have explicitly examined nighttime formation mechanisms in such regions. These motivated the present study, in which we conducted field measurements of gaseous SA and associated parameters at two coastal sites in Hong Kong, to investigate the SA characteristics in coastal boundary layers. The performance of existing proxies was evaluated, and improved models tailored to coastal conditions were then developed. Additionally, mechanism analysis for SA was performed to identify dominant drivers and sinks. We aim to propose reliable proxies for accurately estimating SA levels in Hong Kong and similar subtropical coastal environments.

Results and discussion

SA concentration and characteristics from Two Field Campaigns

Two field campaigns were, respectively, conducted at the Cape D'Aguilar Supersite Air Quality Monitoring Station (CD supersite: 22.22 °N, 114.25 °E) in 2018 and at the Hong Kong University of Science and Technology (HKUST) supersite (22.33 °N, 114.27 °E) in 2022 during fall and winter. Figure 1 and S1 illustrate the overall observation results of gaseous SA and key parameters relevant to its formation. Due to instrument maintenance and availability issues, SO_2 data at the HKUST supersite were unavailable at the beginning of the campaign, while SR data at the CD supersite were missing on specific days. In both campaigns, SA showed clear diurnal variations, with concentration peaking at midday. At the CD supersite in 2018, the diurnal peak concentration was recorded at $(8.0 \pm 7.5) \times 10^6$ molecules/cm^3, while at the HKUST supersite in 2022, it was $(4.2 \pm 2.9) \times 10^6$ molecules/cm^3. Nighttime SA concentrations decreased to $(6.7 \pm 3.4) \times 10^5$ and $(2.8 \pm 1.3) \times 10^5$ molecules/cm^3 at the respective sites. The median concentrations of SA were 2.4×10^6 molecules/cm^3 in 2018 and 1.3×10^6 molecules/cm^3 in 2022. The lower SA concentration observed at HKUST in 2022 aligned with reduced levels of SO_2, O_3, and isoprene compared to CD supersite. As shown in Fig. S2, no obvious correlations were observed between SA and CO, suggesting the negligible primary emission and the predominant role of secondary formation on SA levels. The secondary formation of SA should be determined by the concentration of SO_2, OH radicals, SCIs, and the values of CS. An extremely strong correlation was observed between SA and SR, due to the decisive role of SR on OH radical formation in the daytime. A positive but moderate correlation was observed between SA and SO_2 (Figs. S2, S3a), suggesting that the elevated SO_2 would lead to an increase in SA, while it did not determine the diurnal variation of SA. In contrast, SA showed poor correlations with CS, $PM_{2.5}$, likely due to the narrow variability of CS and $PM_{2.5}$ during the field campaigns. It also should be noted that CS and SA are dynamically coupled, and SA increase can contribute to particle growth that subsequently enhances CS, leading to a delayed feedback effect rather than an immediate response. The seasonal similarity and comparable SR intensities between the two campaigns suggested analogous daytime OH radical levels, and also similar $PM_{2.5}$ concentrations contributed to comparable CS conditions. However, the average concentration of SO_2, the primary precursor of SA, was three times lower at HKUST in 2022 (0.40 ± 0.29 ppb) compared to the CD supersite in 2018 (1.23 ± 0.61 ppb), which should be the key reason for the reduced SA formation in 2022.

When compared with previous studies (Fig. 2), SA concentrations measured at both sites in this study fell within the typical range reported for other rural sites[13,17,23,24,27]. This similarity can be attributed to comparable levels of SO_2, SR, and CS. Forest sites with lower SO_2 levels generally exhibited lower SA concentrations, whereas in urban areas, despite elevated SO_2 levels due to intensive human emissions, SA concentrations often showed comparable levels to rural and suburban locations[17,18,26,28,29]. This intriguing phenomenon can be explained by the concurrent higher CS values in urban environments, which accelerate SA consumption, offsetting the enhanced formation due to elevated precursor concentrations. Notable exceptions (SC in Fig. 2) include the Kilpilahti site in Finland and Maïdo observatory on Réunion island, which experienced exceptionally high SO_2 and SA concentrations due to unique local influences, such as emissions from the oil refinery at Kilpilahti, and volcanic activity at Maïdo[17,25]. These findings underscore the intricate interplay between SA precursor concentrations, loss rates, and environmental factors, all of which jointly determine SA concentrations across diverse atmospheric settings.

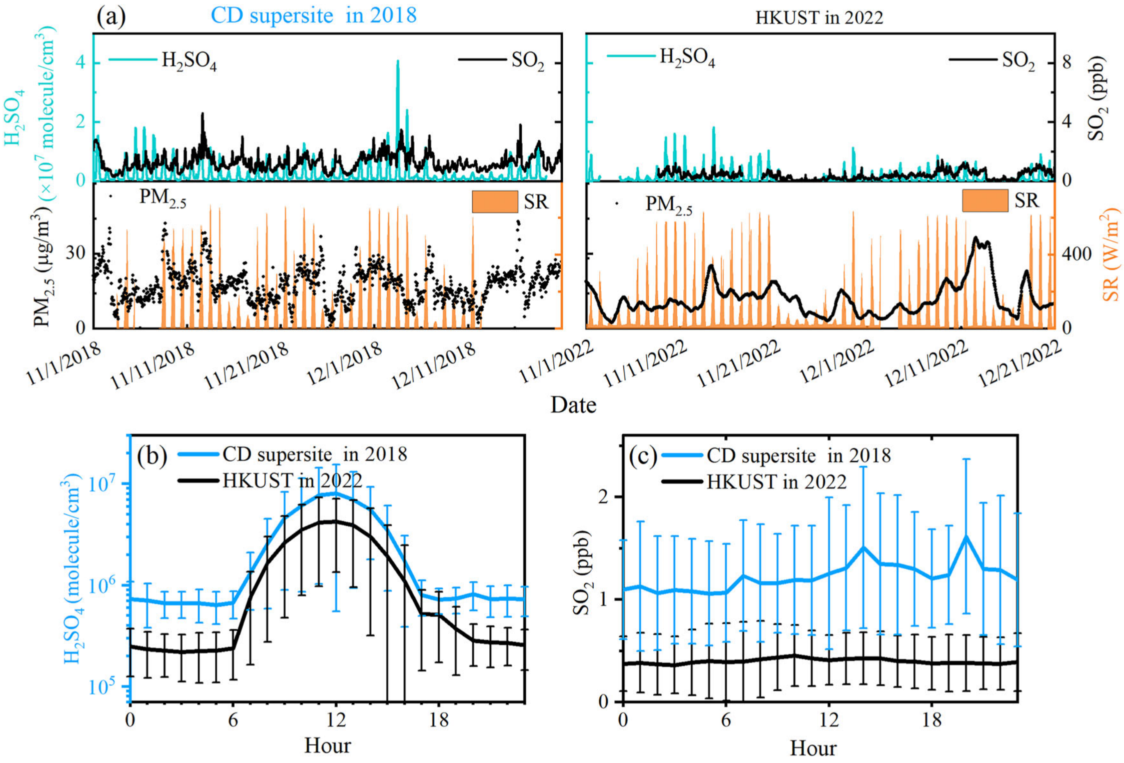

Fig. 1 | Time series and diurnal variations of measured species. **a** Time series of H_2SO_4, SO_2, SR, and $PM_{2.5}$ measured at the CD supersite in 2018 and the HKUST supersite in 2022. **b** Diurnal variation of H_2SO_4 and (**c**) SO_2 measured in both field campaigns.

Evaluating the applicability of previously proposed proxies

Functions of typical proxies used in previous studies were summarized in Table 1, including linear and nonlinear regression-based proxies (D_{L1}-D_{L3}, and D_{NL1}-D_{NL6}) for the daytime SA, and proxy A_1 for the whole day SA estimation, which are described in detail in the Method section. To evaluate the potential of using proxy functions to estimate SA concentrations, we first evaluated the applicability of previously established proxies from other studies (Table S1) in the context of Hong Kong coastal atmospheric conditions. Since SR was the only available indicator for OH radicals during the two field campaigns, proxies based on UVB or directly on OH radical measurements were not included in Table S1. Those previous proxies are all basic functions listed in Table 1, with parameter coefficients empirically derived from different measurement datasets to quantify the sensitivity of SA formation to key atmospheric drivers such as SR, SO_2 oxidation, and condensation sink. Among the selected proxies, P_1-P_6 are specifically designed for daytime SA estimation, while P_7 and P_8 can be used for estimating SA concentrations throughout the entire day. The comparison between the estimated and measured SA concentrations is presented in Fig. 3. Proxies P_1 to P_8 showed strong correlations with observed SA (Pearson $r = 0.69$- 0.87), but variations in the slope, intercept, and the RE values indicate that not all proxies are well-suited to the coastal conditions in Hong Kong.

Proxy P_1 proposed by Petäjä et al.[20] was one of the earliest attempts to estimate SA concentration. It yielded a moderate correlation (0.69) but with a low regression slope of 0.39, indicating its inapplicability to SA estimation in Hong Kong. Proxy P_2 with the function of D_{NL5}, proposed by Mikkonen et al.[24], further considered the influence of RH on CS using multi-site data, and performed better in estimating SA, with a slope of 1.07 and a relative error (RE) of 0.63. Proxy P_3 and P_4, from Kuerten et al.[27] resemble D_{NL4} and D_{NL2}, respectively. Although both proxies produced slopes close to 1 and low RE values, they tended to underestimate SA. Interestingly, the simpler P_4 outperformed P_3, likely because of limited variations in RH and CS during the campaign, and including them added noise rather than explanatory power[27]. This highlights a key trade-off in proxy design that additional variables can improve realism but may reduce robustness when their variability is small or their measurements are uncertain.

Proxy P_5, established by Größ et al.[30] with a linear function of D_{L1}, consistently overestimated SA concentrations in our study, which might be attributed to differences in CS calculation methods. In their work, CS considered particles from 2 nm to 10 μm and adjusted with ambient RH[30], while CS in this work was calculated for dry aerosols (14.1 to 736.5 nm), yielding smaller CS values and thus higher SA. Proxy P_6 further considered nucleation as an additional sink for SA and predicted similar SA concentrations, matching our measurement[17]. It is likely due to similar atmospheric conditions between Agia Marina, Cyprus, where P_6 was developed, and the CD supersite.

P_7 and P_8, with the same function as A_1, were designed to simulate full-day SA concentration under boreal and megacity environments, respectively[17]. However, P_7 with a regression slope of 0.37 and a high RE showed significant inconsistency with SA observation in Hong Kong, suggesting its inapplicability for the coastal measurements. While proxy P_8 performed reasonably well during the daytime, its full-day simulation exhibited a non-linear relationship with the measurements. In summary, previously proposed proxies vary in their transferability due to site-specific dependencies of their empirical coefficients. P_2 and P_6 performed relatively well due to the usage of multi-site datasets covering different conditions or a similar environment to this work, but others did not translate well to Hong Kong's unique coastal atmospheric profile. These results highlight the need for region-specific proxy development, especially for accurate nighttime estimation, where current models remain limited.

Construction of daytime proxies

Based on the proxy functions in Table 1, linear or nonlinear least square regressions were performed with the measurement data at the CD supersite from 8th November to 3rd December 2018 (Fig. 4), and the optimized parameters for each proxy are summarized in Table 2. The proxies were then validated against SA measurements from 4th December to 19th December 2018 at the same site (Fig. S4). Due to the lack of CS data at the

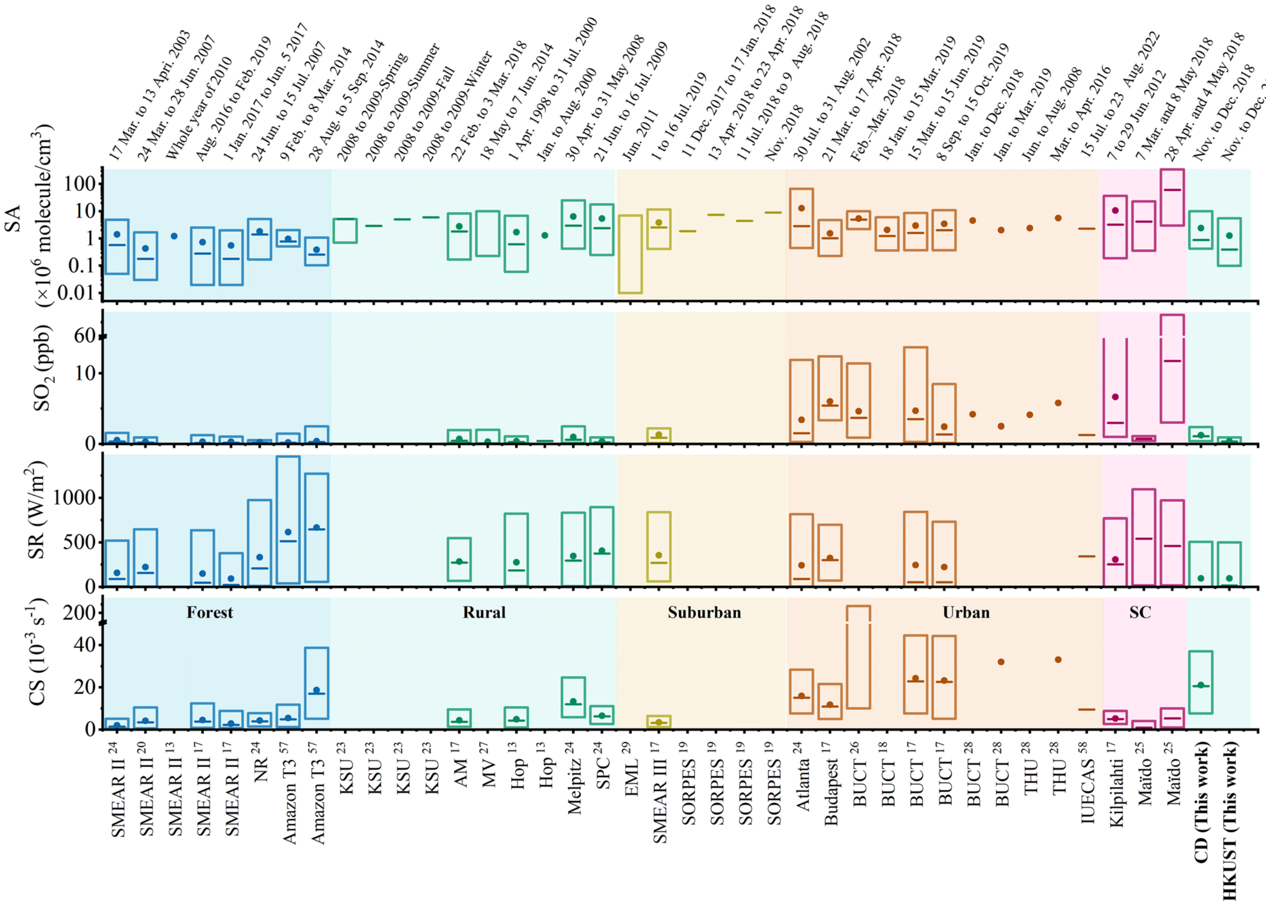

Fig. 2 | Comparison of SA and SO_2 concentration, SR, and CS across multiple studies. Comparison of the results in this work and reported in previous studies (SMEAR II: SMEAR II-station[13,17,20,24], NW: Niwot Ridge[24], Amazon T3: Amazon T3 site[57], KSU: Kent State University[23], AM: Agia Marina[17], MV: Michelstadt–Vielbrunn[27], Hop: Hohenpeissenberg[13], Melpitz[24], SPC : San Pietro Capofium[24], EML: Eagle Mountain Lake[29], SMEAR III: SMEAR III-station[17], SORPES: the Station for Observing Regional Processes of the Earth System in Nanjing[19], Atlanta[24], Budapest[17], BUCT: Beijing University of Chemical Technology[17,18,26,28], THU: Tsinghua University[28], IUECAS: the Institute of Urban Environment, Chinese Academy of Sciences[58], Kilpilahti[17], Maïdo: Maïdo observatory[25]). Bars represent the range of 5 to 95% percentile, lines, and dots indicate medium, and mean values, respectively.

HKUST supersite in 2022, only D_{L2}, $D_{NL2,}$ and D_{NL3} were validated against the measured data (Fig. 5). Across these evaluations, the constructed proxies showed strong correlations with measured SA (R: 0.69–0.91), low relative errors (RE = 0.42–0.97), and slopes close to 1, outperforming earlier proposed proxies.

The fitted parameters for D_{L1} and D_{L2} in our study were consistent with previous reported ranges, while the fitted parameter for D_{L3} was slightly higher than those derived by Mikkonen et al.[24]. Among the linear proxies, simplified D_{L2} excluding CS showed the best performance, reflected by its higher correlation factor and lower RE value. This result can be attributed to the weak correlation between CS and SA during our campaign (Fig. S3b), echoing previous observations by Kuerten et al.[27] and the above discussions in Section 3.2. When CS fluctuates within a limited range, $[SO_2]$ and SR alone may suffice for estimating daytime SA levels.

The non-linear proxies D_{NL1}-D_{NL5} introduce additional parameters, enhancing flexibility in modeling SA behavior. Generally, the exponents for SR (b) and $[SO_2]$ (c) were close to 1, consistent with previous studies by Mikkonen et al.[24] and Kuerten et al.[27]. In our study, the exponent for CS (d), ranged from −0.66 to −0.87, slightly higher than the theoretical value of −1, but still within a plausible range and lower than those reported by Mikkonen et al.[24] and Kuerten et al.[27]. D_{NL1}, which used the same variables as D_{L1} but incorporates nonlinearity, performed similarly to its linear counterpart. D_{NL2}, which excluded CS, performed better than D_{NL1} and also its linear counterparts D_{L2}, and D_{L3}. Although measured SA was negatively correlated with RH (Fig. S3c), likely due to decreasing RH after sunrise as temperatures rise, RH had minimal impact on SA estimation in D_{NL3} and D_{NL4}. Specifically, the values of $RH^{-0.14}$ and $RH^{0.18}$ during the campaign ranged narrowly from 0.53 to 0.57 and 2.1 to 2.3, respectively. This limited variation reduced RH's influence on SA formation in these proxies. In D_{NL5}, the exponents for CS and RH were fixed to the same value (−0.66). As shown in Fig. S5, while RH had little effect on the diurnal variation of 1/CS, combining the two 1/(CS × RH) increased the exponent of CS from −0.86 in D_{NL1} to −0.66 in D_{NL5}. Nonetheless, D_{NL5} still underperformed compared to D_{NL2} and D_{NL3}. D_{NL6} further considered nucleation as an additional sink for SA[17], and the fitted coefficient 'a', representing OH radical as a function of SR, closely matched values in P_6 proposed by Dada et al.[17]. However, the value of K_n obtained in our study was nine orders of magnitude lower than that in proxy P_6, suggesting a negligible contribution of nucleation process to SA consumption under our conditions. While this observation aligns with local atmospheric behavior, further investigation is warranted due to potential uncertainties in proxy construction. Overall, all constructed proxies (D_{L1}-D_{L3}, D_{NL1}-D_{NL6}) effectively estimated daytime SA. Evaluated from the value of R, RE, and the slope of the reconstructed SA versus measured SA, non-linear functions slightly outperformed linear ones, and among them, D_{NL2} demonstrated the highest accuracy.

Given the strong correlation between CS and $PM_{2.5}$ (Fig. S6), we explored the use of $PM_{2.5}$ as a potential surrogate for CS in conditions where direct size-resolved data is unavailable. Although CS is strongly influenced by the size distribution, our measurements showed a clear positive correlation between CS with $PM_{2.5}$ concentration, and $PM_{2.5}$ with the mean aerosol diameter (Fig. S6), suggesting that $PM_{2.5}$ can serve as a reasonable indicator of CS variability. However, caution is warranted because $PM_{2.5}$ assumes a relatively homogeneous particle composition and hygroscopicity, which may not always hold true across regions and seasons, and thus validation is recommended before broader application. Using this approach, we constructed $D_{NL1\text{-}PM}$, $D_{NL4\text{-}PM}$, and $D_{NL5\text{-}PM}$, as shown in Table 2. In $D_{NL1\text{-}PM}$, the exponent *'d'* for $PM_{2.5}$ was only −0.007, resulting in PM^d values close to 1. Consequently, the parameters in $D_{NL1\text{-}PM}$ were nearly identical to those in D_{NL2}. Again, RH exponent had a minimal effect in $D_{NL4\text{-}PM}$, while its addition increased the parameter *'d'* to −0.67. Figures 4, 5, and S4 show that $D_{NL1\text{-}PM}$, $D_{NL4\text{-}PM}$, and $D_{NL5\text{-}PM}$ performed comparably to D_{NL1}-D_{NL6}, suggesting that replacing CS with $PM_{2.5}$ can be a practical alternative for daytime SA proxy, especially in scenarios where direct CS measurements are not available.

Table 1 | Proxy functions used to construct proxies in this work

Proxies	Functions/parameters			
	$SA = a \times k \times SR^b \times SO_2{}^c \times CS^d \times RH^e$			
Power of the Parameters	*SR*	SO_2	*CS* or PM	RH
D_{L1}	1	1	−1	0
D_{L2}	1	1	0	0
D_{L3}	1	1	−1	−1
D_{NL1} / $D_{NL1\text{-}PM}$[a]	b	c	d	0
D_{NL2}	b	c	0	0
D_{NL3}	b	c	0	e
D_{NL4} / $D_{NL4\text{-}PM}$[a]	b	c	d	e
D_{NL5} / $D_{NL5\text{-}PM}$[a]	b	c	d	d
D_{NL6}	$[SA] = -\frac{CS}{2 \times k_n} + \left[\left(\frac{CS}{2 \times k_n}\right)^2 + \frac{[SO_2]}{k_n}(a \times k \times SR)\right]^{1/2}$			
N/ $N_{\text{-}PM}$[a]	$[SA] = (k \times (a' \times [In.]^{b'} + g) + f \times [O_3] \times [Alkene]) \times [SO_2] \times CS^d$			
A_1	$[SA] = -\frac{CS}{2k_n} + \left[\left(\frac{CS}{2k_n}\right)^2 + \frac{[SO_2]}{k_n}(a \times k \times SR + f \times [O_3] \times [Alkene])\right]^{1/2}$			
A_2/ $A_{2\text{-}PM}$[a]	Daytime: D_{NL1} with c = 1 Nighttime: Proxy N			

[a]Note: For $D_{NL1\text{-}PM}$, $D_{NL4\text{-}PM}$, and $D_{NL5\text{-}PM}$, $N_{\text{-}PM}$ and $A_{2\text{-}PM}$, $PM_{2.5}$ concentration was used as alternative to CS.

Construction for the whole day SA proxies

To develop a proxy capable of estimating gaseous SA concentrations over both day and night, A_1 was initially established by considering OH radicals as the daytime oxidant and SCIs for the whole day. As shown in Table 3, the fitted parameter *'a'* was 7.7×10^3, yielding an average product *a*×*k* of 7.3×10^{-9}, which is similar to the value of 8.6×10^{-9} (P_7), previously reported by Dada et al.[17]. However, the estimated apparent reaction rate of alkenes with O_3 (3.7×10^{-30} cm^3/$molecule^1$/s^1) was 4–16 times lower than previously reported values (see Table S1). Consistent with the above proxy D_{NL6}, the 'k_n' value representing the nucleation consumption of SA revealed only a minor role for the SA sink. Overall, A_1 still underperformed, particularly during nighttime, where SA formation was significantly underestimated (Fig. 6a). This suggests that using SCIs or $[O_3] \times [Alkenes]$ may not sufficiently account for nighttime SA formation in coastal Hong Kong.

To investigate the reaction mechanisms of SA, an observation-based photochemical box model (PBM) built on the Master Chemical Mechanism (MCM v3.3.1) was employed to simulate SA concentrations (Figs. 6b and S7). A good linear relationship was observed between the simulated and observed SA levels; however, the model slightly underestimated the observed SA concentrations, particularly during nighttime. Since SA is generated from the reaction of SO_3 with H_2O, the budget analysis of SO_3 can provide insights into the formation pathways of SA (Fig. S8). Notably, OH radicals were identified not only as the main contributor to daytime SA formation but also accounted for 96% of nighttime SA formation in the default model M_0. This result contrasts with previous findings suggesting that SCIs should dominate the nighttime SA formation[17]. The reduced significance of SCIs role on SA formation might be attributed to the incomplete mechanisms of SCIs reacting with SO_2 in M_0. Therefore, modifications were made to improve the model performance and further investigate the formation mechanism of SA. The default reaction rates of SCIs with SO_2 are 7×10^{-14} cm^3/molecule/s, which are lower than the recommended values by the International Union of Pure and Applied Chemistry (IUPAC) for most SCIs (https://iupac.aeris-data.fr/). Given the abundance of different SCIs in our model, the reaction rates for CH_2OO, CH_3CH_2OO, MACROO, and MVKOO reacting with SO_2 were adjusted to the recommended values (M_1 in Table S2). These modifications in M_1 resulted in an increase in the average formation rate of SA from 4.5×10^3 to 3.9×10^4 molecule/cm^3/s during nighttime, and from 7.6×10^4 to 1.1×10^5

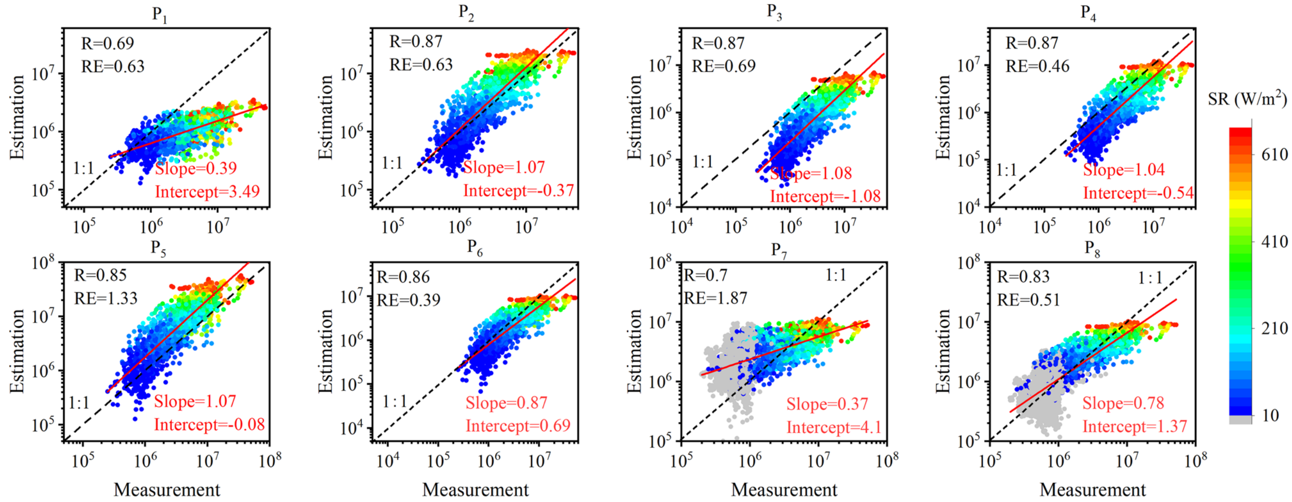

Fig. 3 | Comparison of estimated SA concentrations from previous proxies with measurements. Comparison between SA estimated using proxies P_1-P_8 proposed in previous studies (Table S2) and measured SA concentrations from 8th November to 3rd December 2018 at the CD supersite.

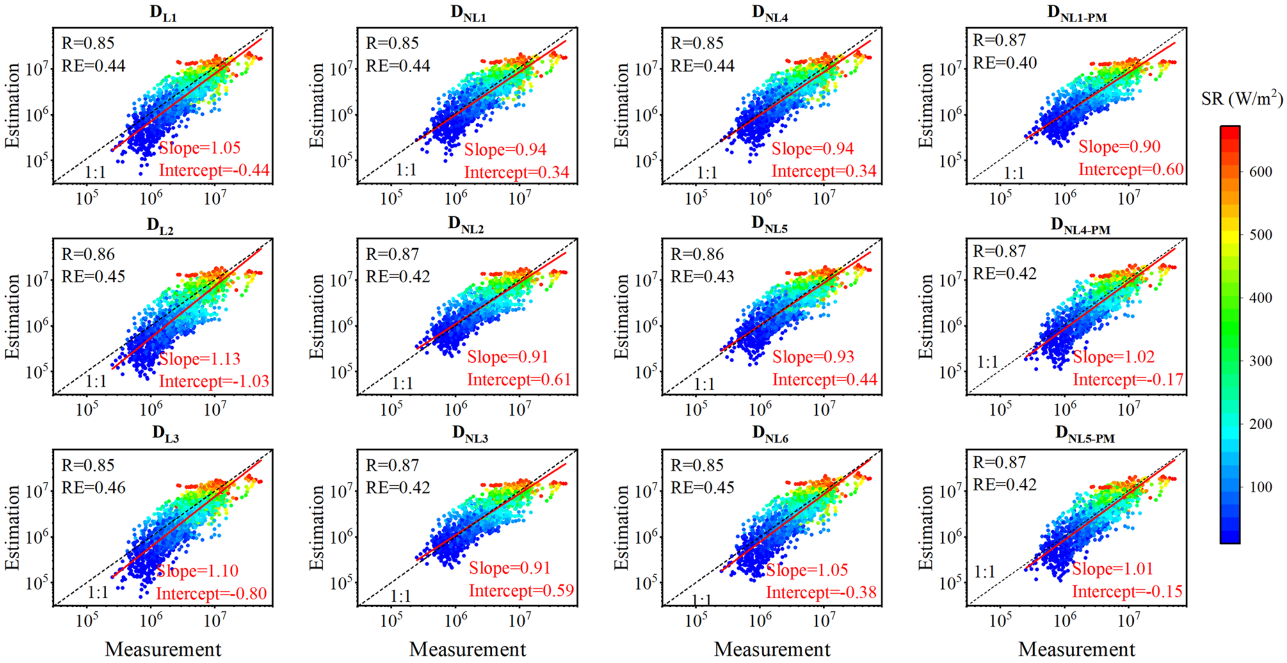

Fig. 4 | Reconstruction of SA concentrations at the CD supersite using different proxies. Reconstructed SA concentrations from linear (D_{L1}-D_{L3}) and non-linear (D_{NL1}-D_{NL6}, D_{NL1-PM}, D_{NL4-PM}, D_{NL5-PM}) proxies versus measured SA from 8th November to 3rd December 2018 at the CD supersite.

Table 2 | Obtained parameters for daytime proxies in Table 1

Num.	a	b	c	d	e	k_n	R^a	RE^a	R^b	RE^b	R^c	RE^c
D_{L1}	6.97e3						0.85	0.44	0.69	0.92		
D_{L2}	4.58e5						0.86	0.45	0.79	0.83	0.86	0.64
D_{L3}	4.83e5						0.85	0.46	0.69	0.97		
D_{NL1}	5.84e3	0.87	1.07	−0.86			0.85	0.44	0.70	0.92		
D_{NL2}	1.60e9	0.82	0.72				0.87	0.42	0.79	0.87	0.91	0.54
D_{NL3}	2.67e9	0.81	0.72		−0.14		0.87	0.42	0.80	0.88	0.90	0.55
D_{NL4}	2.92e3	0.88	1.06	−0.87	0.18		0.85	0.44	0.70	0.91		
D_{NL5}	1.00e6	0.81	1.02	−0.66	−0.66		0.86	0.43	0.73	0.91		
D_{NL6}	7.4e−9					1.03e-18	0.85	0.45	0.70	0.96		
D_{NL1-PM}	1.95e9	0.82	0.71	−0.007			0.87	0.40	0.71	0.83	0.91	0.52
D_{NL4-PM}	1.84e5	0.93	1.17	−0.67	−0.12		0.87	0.42	0.79	0.91	0.88	0.61
D_{NL5-PM}	1.28e6	0.89	1.18	−0.63	−0.63		0.87	0.42	0.70	0.99	0.87	0.67

[a] parameterization with data collected from 8th November to 3rd December 2018 at the CD supersite.

[b] validation of data collected from 4th December to 19th December 2018 at the CD supersite.

[c] validation of data collected from 14th November to 19th December 2022 at the HKUST supersite.

molecule/cm^3/s during daytime. While the simulated daytime SA from M_1 still aligned well with observations, the nighttime simulation was overestimated (Fig. S7b).

As reported in previous studies, H_2O or $(H_2O)_2$ could be dominant sinks for some SCIs. However, the default mechanisms only included reactions between H_2O and SCIs with rates of 1×10^{-17} cm^3/molecule/s, which do not accurately reflect the real atmospheric conditions[31–33]. Additionally, the reaction rates between SCIs with NO_2 were set to 1×10^{-15} cm^3/molecule/s, while the recommended reaction rates of $CH_2OO + NO_2$ and $CH_3CHOO + NO_2$ are 3×10^{-12} and 2×10^{-12} cm^3/molecule/s, respectively[34,35]. Therefore, mechanisms related to SCIs with H_2O and NO_2 were added or modified in the model as M_2. Due to the competition of H_2O and NO_2 with SO_2 for SCIs, the formation rate of SO_3 decreased in M_2. Similar to models M_0 and M_1, the daytime SA was well reproduced in M_2, and the simulated SA in the early evening showed better consistency with observations compared to M_0 and M_1 (Fig. S7c). However, the nighttime SA simulated by M_2 was still lower than the observations, although the discrepancy was smaller than that observed in M_0. These results suggest that, in addition to SCIs, there may be another source contributing to nighttime SA formation.

It is noteworthy that the average simulated nighttime OH concentration in M_0, M_1, and M_2 was approximately 1.4×10^5 molecule/cm^3 (Fig. 6c), which was three times lower than the levels observed in 2020 at the same site and season[36]. Nighttime OH radicals are mainly generated from the oxidation of VOCs by NO_3 radicals and O_3 via the formed RO_2 and HO_2 radicals reacting with NO or Criegee intermediates. Although nighttime O_3 concentrations are typically low in many urban areas, but high nighttime O_3 concentration (annual average of 20–30 ppbv) was observed in this region,

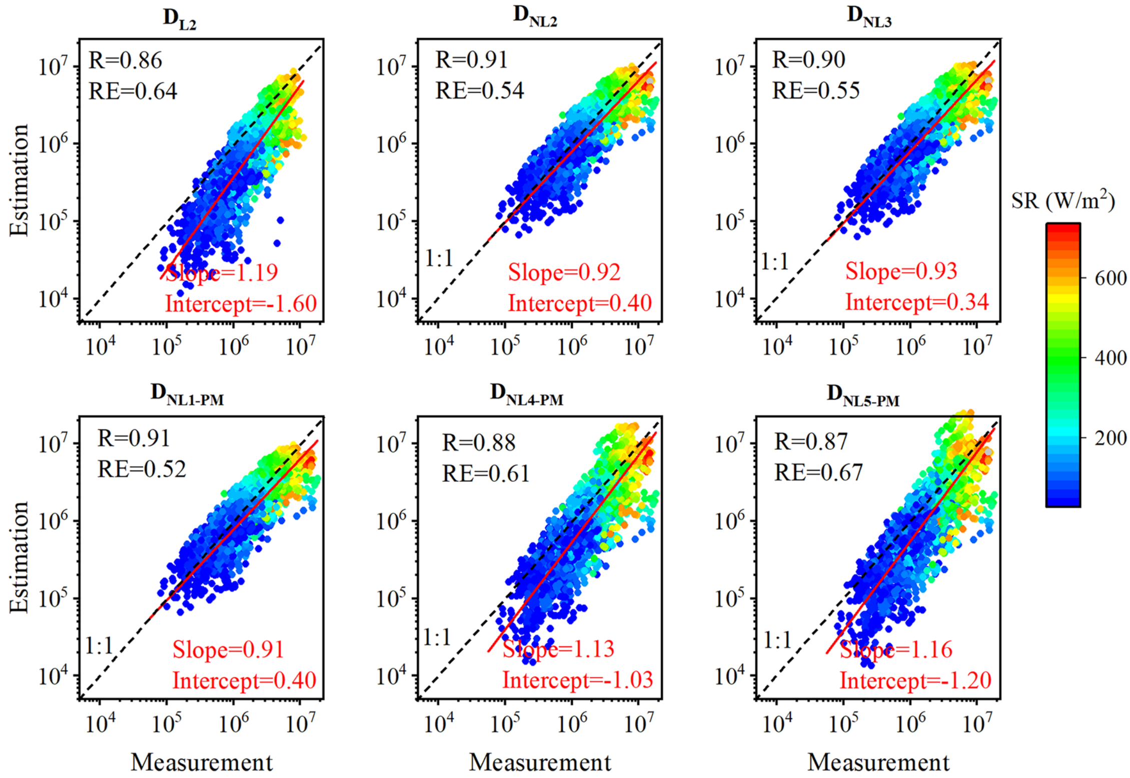

Fig. 5 | Comparison of estimated SA concentrations from proxies with measurement. Estimated SA concentrations from proxies D_{L2}, D_{NL2}, D_{NL3}, D_{NL1-PM}, D_{NL4-PM}, and D_{NL5-PM} versus measured values from 14th November to 20th December 2022 at the HKUST supersite.

Table 3 | Obtained parameters for proxies designed for nighttime and full day

Num.	a	b	d	f	g	k_n	R^a	RE^a	R^b	RE^b	R^c	RE^c
N			−0.79	1.7e−30	5.9e5							
A_1	7.7e3			3.7e−30		9.6e−19	0.87	0.51				
A_2	4.0e4	0.87	−0.80	1.7e−30	5.7e5		0.89	0.39	0.82	0.65		
N_{-PM}			−0.72	3.9e−28	1.4e8							
A_{2-PM}	5.4e6	0.91	−0.58	1.6e−28	1.0e8		0.90	0.39	0.86	0.47	0.88	0.56

[a]parameterization with data collected from 8th November to 3rd December 2018 at the CD supersite.
[b]validation of data collected from 4th December to 19th December 2018 at the CD supersite.
[c]validation of data collected from 14th November to 19th December 2022 at the HKUST supersite.
Note: a' and b' were not shown here due to their negligible role in the proxies.

favoring the formation of NO_3 radicals and subsequent OH production. These processes are complex and are not yet fully represented in the MCM mechanism[37]. Moreover, a number of VOCs remain undermeasured or are not fully represented in the model. The combination of incomplete mechanisms and missing VOC reactivity potentially led to the underestimation of OH radicals. To determine whether the underestimation of nighttime SA was due to potentially missing sources of OH radicals, a direct unknown source with generation rates of 3×10^6, 4×10^6, and 5×10^6 molecules/cm^3/s was added to the model (M_3, including M_{3-1}, M_{3-2}, and M_{3-3}). As shown in Fig. 6c, the additional source of OH radicals resulted in average nighttime OH radical concentrations of 3.9×10^5, 4.7×10^5, and 5.4×10^5 molecules/cm^3 for M_{3-1}, M_{3-2}, M_{3-3}, respectively. The simulated SA from M_3 demonstrated better consistency with observations (Figs. 6d and S9). Based on the budget analysis of M_{3-2}, OH radicals and SCIs contributed 69 and 31% to the nighttime SA formation, respectively, while daytime SA remained predominantly influenced by OH radicals (Fig. 6e). The role of SCIs in SA formation was more important during the evening, which aligns with the findings of Guo et al.[18]. These results emphasize that not only SCIs but also OH radicals played important roles in nighttime SA formation in Hong Kong. Due to the challenges associated with directly measuring OH radicals, particularly at low levels during nighttime, identifying reliable surrogate indicators remains essential for practical applications.

However, according to box model simulations, nighttime OH concentrations showed limited variation (Figs. 6c and S10). Observation data from Zou et al.[36] also present no significant fluctuations during the night. This stability suggests that a representative value may be sufficient to represent nighttime OH radicals. To explore this, we evaluated several potential indicators, including benzene, toluene, and NO_x, to estimate nighttime OH radicals in proxies N (Table 1) and N_{-PM} (see SI). Regardless of the indicator used, the fitted parameter *g*, representing the initial OH concentration in the evening, remained consistent at 5.9×10^5 molecule/cm^3 (Table S3), while the extremely low values of $a' \times [In.]^{b'}$ from different indicators had minimal impact on OH radicals estimation. The 25th, 50th, and 75th percentiles of observed SA during the same period were 5.6×10^5, 7.1×10^5, and 8.6×10^5 molecule/cm^3, respectively, reflecting low variation in nighttime SA. These results suggest a relatively stable nighttime chemical environment in the coastal background environment[36]. The fitted OH

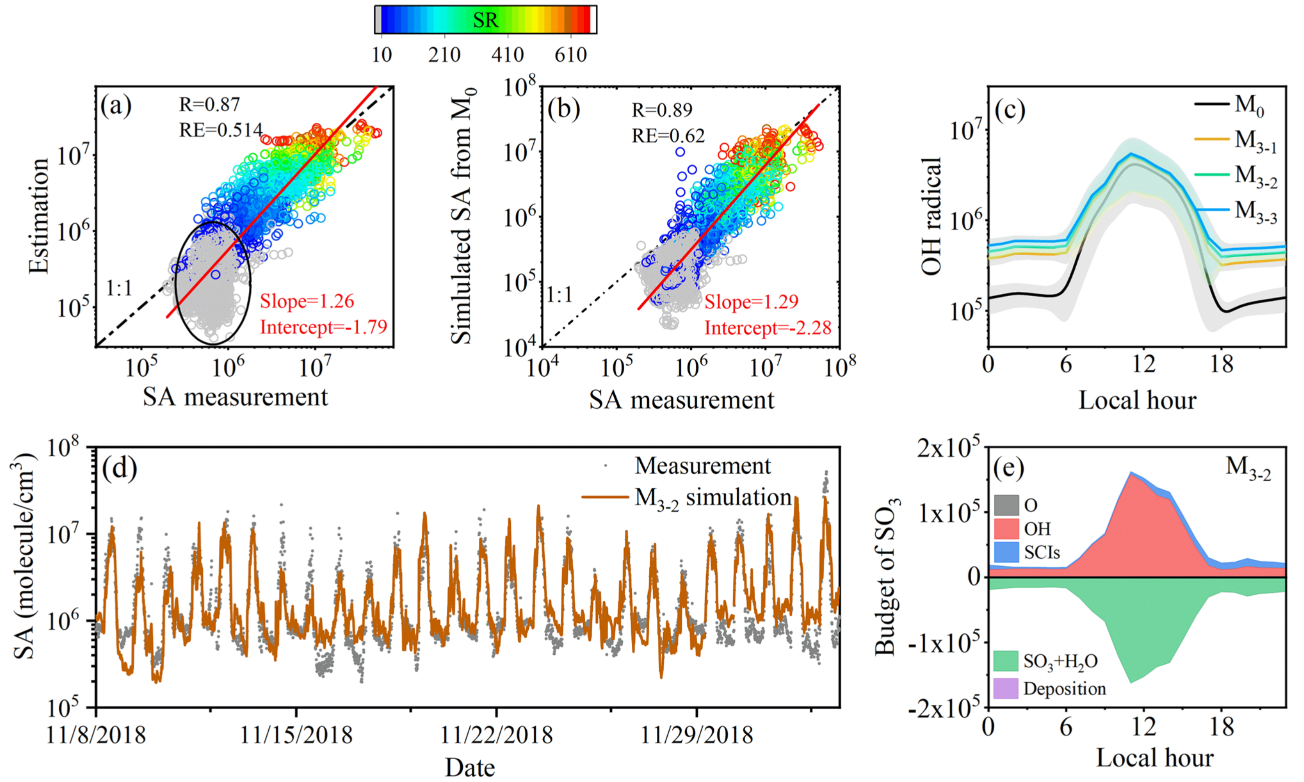

Fig. 6 | SA estimation and box model simulation. **a** The estimated SA from proxy A_1 and (**b**) the simulated SA from box model M_0 versus measured SA from 8th November to 3rd December 2018 at the CD supersite (unit: molecule/cm^3). **c** Simulated OH radicals (molecule/cm^3) in the model M_0 and M_3. **d** Time series of estimated SA from $M_{3\text{-}2}$ and observed SA. **e** Budget analysis of SO_3 (molecule/cm^3/s) in model $M_{3\text{-}2}$.

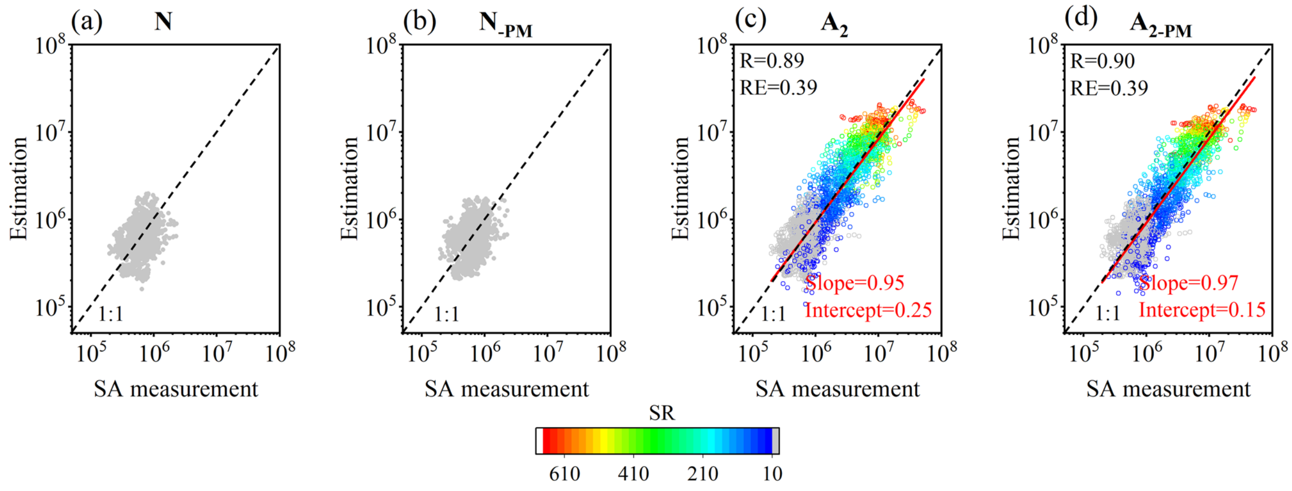

Fig. 7 | Comparison between estimation and measurement. The estimated SA from the proxy of N (**a**), $N_{\text{-PM}}$ (**b**), A_2 (**c**), and $A_{2\text{-PM}}$ (**d**) versus measured SA from 8th November to 3rd December 2018 at the CD supersite.

concentrations were a little higher than the values obtained in model $M_{3\text{-}2}$, which could be partly explained by the higher exponent value of CS than −1 fitted in proxy N. Then, the final expression for Proxy N and Proxy $N_{\text{-PM}}$ can be obtained. Figure 7a, b compare the estimated nighttime SA from proxies N and $N_{\text{-PM}}$ and observed values, both showing strong agreement, supporting the effectiveness of using a constant nighttime OH value in proxy development.

Based on these insights, we proposed a piecewise full-day proxy, A_2, which uses proxy N for nighttime SA and the daytime proxy D_{NL1} (Table 1). To maintain consistency, the exponents for CS and SO_2 were held constant across both day and night components. Regress results showed an OH radical concentration of 5.7×10^5 molecule/cm^3, with parameter values for CS and the alkene oxidation term similar to those in proxy N. Parameters for daytime function ($a = 4.0 \times 10^4$ and $b = 0.87$) also aligned well with those used in earlier non-linear functions in Table 2. Besides, proxy $A_{2\text{-PM}}$ constructed with $PM_{2.5}$ concentration was also obtained (Table 3). Figure 7c and Fig. 7d show strong linear correlations between the simulated SA from both proxies and the observed SA, validating their accuracy.

The constructed proxies N, $N_{\text{-PM}}$, A_2, and $A_{2\text{-PM}}$ were further validated using data collected from 4th to 19th December 2018 at the CD supersite and from 14th November to 19th December 2022 at the HKUST supersite. As shown in Fig. S11, all proxies maintained good correlations with observed

SA, confirming their robustness and applicability across different locations and timeframes. The results underscore the significance of generated OH radicals in nighttime SA formation, with concentrations averaging between 4 to 6×10^5 molecules/cm^3. Despite the challenges in developing reliable nighttime SA proxies, N and $N_{\text{-PM}}$ effectively reproduced nighttime SA dynamics. The piecewise proxies A_2 and $A_{2\text{-PM}}$ offer practical solutions for estimating SA throughout the entire day and show promise for broader application in Hong Kong and similar coastal environments. Further research is needed to better understand nighttime SA formation mechanisms and to refine indicators for OH radicals, which remain critical to improving proxy accuracy and usability.

Atmospheric implication

This study provides a comprehensive analysis of SA formation in Hong Kong, combining field measurements, box model simulations, and statistical proxy development. A series of localized linear and nonlinear daytime proxies tailored for subtropical coastal environments were constructed, with D_{NL2} (incorporating SO_2 and SR) demonstrating the best performance. In addition, for cases where CS data is unavailable or uncertain, $PM_{2.5}$ was found to be a practical substitute, enhancing the usability of the proxies for long-term or resource-limited monitoring efforts. Based on the important role of nighttime OH radicals on SA formation, new nighttime proxies (N and N_{PM}) and full-day piecewise proxies (A_2 and $A_{2\text{-PM}}$) were developed.

$$A_2 = \begin{cases} [SA] & = \quad 4.0 \times 10^4 \times k \times SR^{0.87} \times [SO_2] \times CS^{-0.8}\ (D) \\ [SA] & = \quad (k \times 5.7 \times 10^5 + 1.7 \times 10^{-30} \times [O_3] \times [Alkene]) \times [SO_2] \times CS^{-0.8}\ (N) \end{cases}$$

$$A_{2-\text{PM}} = \begin{cases} [SA] & = \quad 5.4 \times 10^6 \times k \times SR^{0.91} \times [SO_2] \times PM^{-0.58}\ (D) \\ [SA] & = \quad (k \times 1.0 \times 10^8 + 1.6 \times 10^{-28} \times [O_3] \times [Alkene]) \times [SO_2] \times PM^{-0.58}\ (N) \end{cases}$$

Although the current work focuses on the autumn–winter transition period, the underlying physical and chemical relationships embedded in the proxies remain valid across seasons. The observed air mass types and meteorological conditions were representative of typical regional patterns, but the influence of precursor levels should be evaluated across seasons to confirm robustness. While the proxies are optimized for Hong Kong's coastal subtropical conditions, their structure and variable dependence suggest transferability to similar maritime or coastal atmospheres with similar chemical and meteorological regimes. The ranges of key parameters in Fig. 2 can serve as diagnostic indicators to assess whether these proxies are suitable for use in other locations. Overall, this study offers several robust and practical proxy models for estimating SA concentrations and provides a framework for extending SA analysis to regions and periods lacking direct measurements. Further research is still needed to better understand the role of nucleation in the SA sink and the underlying mechanisms of nighttime OH radicals and SA productions, therefore, developing a complete understanding of the fate and impacts of SA in the atmosphere.

Methods

Field measurements

Two field campaigns were conducted to measure atmospheric SA concentration in Hong Kong during fall and winter. The first campaign took place at the CD supersite from 1st November to 19th December 2018, and the second campaign was held at the HKUST supersite over the same periods in 2022. The CD supersite, located at the south tip of Hong Kong Island and faced the South China Sea, serves as a regional background station[38]. It is surrounded by vegetation and situated about 10 km from the nearest urban center. The site is influenced by local biogenic emissions, marine air masses from the South China Sea, and long-range transport from the eastern coastal area of south China and the Pearl River Delta (PRD) region. Previous back-trajectory analyses have shown the site alternates between clean marine conditions and polluted continental outflow[39,40]. The HKUST supersite is located on the university campus, positioned on a coastal cliff facing Port Shelter and Silver Strand Bay. This site represents a typical suburban environment with minimal nearby residential or commercial activity. It experiences similar regional air mass influences as the CD supersite. Unlike the CD supersite, the HKUST site is also affected by on-campus human activities and constructions[41].

Gaseous SA concentrations were measured by the nitrate-based Time-of-Flight Chemical Ionization Mass Spectrometer (ToF-CIMS, Aerodyne Inc., USA) in both campaigns. A HR version and a long version of ToF were employed at the CD supersite and HKUST supersite, respectively[39], with higher mass resolution (9000 for m/z > 200) for the latter than the former (5200 for m/z > 200). Both resolutions were sufficient for the precise identification of targeted ions. Both instruments share the same ionization and detection principles. Primary ions, $(HNO_3)_nNO_3^-$ ($n = 0,1,2$), were generated by exposing a mixture of HNO_3 stream (5–8 mL/min) and sheath gas (20–30 L/min) to soft X-ray radiation. Ambient air was pumped through a sample inlet at a flow of 10 L/min into the center of a coaxial laminar flow reactor, where SA was ionized by $(HNO_3)_nNO_3^-$ to form HSO_4^-, $H_2SO_4NO_3^-$, and $H_2SO_4HNO_3NO_3^-$, and then detected by ToF mass spectrometer. Instrument calibration was performed using a custom-built SA calibrator onsite before and at the end of each filed campaign[42,43]. The calibration methodology is described in detail in our previous studies[39,40]. Briefly, known SA was produced by SO_2 reacting with certain OH radicals, which were generated via the photolysis of H_2O under the irradiation of UV lamp (185 nm). Then, the calibration factor of SA (γ) can be obtained from Eq. 2, which was 9.88×10^9 and 6.14×10^9 molecule cm^{-3} in CD and HKUST campaigns, respectively. The limit of detection, calculated as three times the standard deviation of the zero signal, was 1.2×10^5 and 3.1×10^4 molecules cm^{-3} in CD and HKUST campaigns, respectively. Zero air was injected periodically, and the background signal was subtracted before the calculation.

$$[H_2SO_4] = \gamma \times \text{In}\left(1 + \frac{(HSO_4^- + H_2SO_4NO_3^- + H_2SO_4HNO_3NO_3^-)}{\sum_{j=0}^{2} NO_3^- \cdot (HNO_3)_j}\right) \tag{2}$$

Volatile organic compounds (VOCs) were measured by a Proton Transfer Reaction ToF-MS (PTR-QiToF-MS, Ionicon Inc., Austria)[44,45] and online gas chromatography at the CD supersite (Syntech Spectras GC955 Series 600/800, The Netherlands), and a VOCUS-PTR-ToF-MS (VOCUS 2 R, Aerodyne, Inc., USA) at the HKUST supersite. Both PTR-ToF-MS were periodically calibrated using multi-component VOC gas standards, which included 27 and 15 types of VOCs (covering key VOCs such as benzene, toluene, and isoprene) for the campaigns in 2018 (RESTEK canister, IONICON Analytik, Austria) and in 2022 (Apel-Riemer Environmental, Inc., USA), respectively[45]. Concentrations were retrieved using the sensitivities of calibrated species and theoretical estimation through proton-transfer reaction rate coefficients (k_{PTR}) for uncalibrated species[41]. Trace gases, including NO, NO_2 O_3, SO_2, and CO (42i, 49i, 43i, and 48i, Thermo Fisher Scientific Inc., USA) and meteorological parameters (wind, temperature, RH, and SR) were also concurrently measured at both sites. $PM_{2.5}$ mass concentration was detected at the CD supersite by a SHARP monitor (model 5030, Thermo Scientific Inc., USA), and the particle number and size distribution ranging from 14.1 to 736.5 nm was measured using a Scanning Mobility Particle Sizer spectrometer (SPMS 3938, TSI, USA). Unfortunately, particle size and number concentration were not available at the HKUST supersite during the campaign. Instead, $PM_{2.5}$ concentrations were obtained from the nearby Tap Mun station, where previous studies have shown to correlate closely with HKUST measurement under similar environmental conditions[46].

Box model simulation

An observation-based photochemical box model (PBM) built on the Master Chemical Mechanism (MCM v3.3.1) (http://mcm.york.ac.uk)[47–50] was employed to simulate the formation of gaseous SA and OH radicals at the CD supersite from November 8 to December 3, 2018.

The PBM simulates atmospheric chemical processes of inorganic and organic species in a zero-dimensional framework, and concentrations of trace gases, VOCs, and meteorological parameters are used as time-resolved constraints, while the model integrates coupled gas-phase reactions to reproduce the evolution of target SA species. This approach has been successfully used in our previous studies[39,51,52]. In addition to the default reaction mechanisms for SA formation in the MCM, a condensation sink (CS) term of SA was included as baseline case (M_0, Table S2). Further modifications (M_1, M_2, and M_3 in Table S2) were also made to improve the model performance and investigate the gaseous SA formation mechanisms, including adjusted reaction rates for SCIs with SO_2 based on experimental results (M_1) and additional reactions with H_2O and NO_2 (M_2). An additional nighttime OH source (with generation rate of 3×10^6 to 5×10^6 molecules/cm^3/s) was also added to explore the nighttime SA formation in M_3. Measured SO_2, VOC species, other trace gases, CS, and meteorological parameters were input into the model every 10 min to constrain the simulations. The CS values ranged from 0.005 s^{-1} to 0.07 s^{-1}, with an average value of 0.02 ± 0.01 s^{-1}. Details on the CS calculation are provided in the next section. Further information on the model framework and input parameters can be found in our previous study, which used the same model setup to investigate the formation and sink of nitro-phenolic compounds during the same field campaign[39].

Proxy construction

As mentioned above, atmospheric SA is primarily formed through the oxidation of SO_2 by OH radical and SCIs, while its removal is controlled by the CS and nucleation processes. The changes in gaseous SA concentration can be expressed as Eq. 3:

$$-\frac{d[SA]}{dt} = k\times[OH]\times\left[SO_2\right] + k_{SCIs}\times[SCIs]\times\left[SO_2\right] - CS\times[SA] - k_n\times[SA]^2 \tag{3}$$

Here, [SA], [OH], [SO_2], [SCIs] are the concentrations of SA, OH radicals, SO_2, and SCIs, respectively, with unit of molecules/cm^3. k and k_{SCIs} denote the reaction rates of SO_2 with OH radical and SCIs, and k_n represents the apparent consumption rate of SA via nucleation. The temperature-dependent reaction rate k (cm^3/molecule/s) can be calculated from Eq. 4, Eq. 5[24]. The constants are defined as: $k_1 = 4\times10^{-31}$, $k_2 = 3.3$, $k_3 = 2\times10^{-12}$, $k_4 = -0.8$, and $[M] = 0.101\times(1.381\times10^{-23}\times T)^{-1}$. The average reaction rate k_{SCIs} reflects the contribution of multiple types of SCIs present in the atmosphere:

$$k = \frac{A\times k_3}{(A+k_3)}\times \exp\left(k_4\times\left[1+log_{10}\left(\frac{A}{k_3}\right)^2\right]^{-1}\right) \tag{4}$$

$$A = k_1\cdot[M]\cdot(300/T)^{k_2} \tag{5}$$

CS was calculated from Eq. 6, where D is the diffusion coefficient of gaseous SA, r_i is the geometric mean particle radius in size bin i, and N_i is the particle number concentration[53]. The transitional correction factor, β_{M_i} is expressed as Eq. 7. Knudsen number (K_n) is defined as $\frac{\lambda_v}{r_i}$, with λ_v being the mean free path. The mass accommodation coefficient (α) is typically assumed to be 1[53]:

$$CS = 4\pi D\sum_i \beta_{M_i} r_i N_i \tag{6}$$

$$\beta_M = \frac{K_n+1}{0.377K_n + 1 + \frac{4}{3}\alpha^{-1}{K_n}^2 + \frac{4}{3}\alpha^{-1}K_n} \tag{7}$$

To simplify the system, the formation and consumption of gaseous SA were assumed to be in a steady state[20]. Under this assumption, Eq. 3 can be solved as Eq. 8. If nucleation is neglected, the simplified form becomes Eq. 9. Given that OH oxidation is the dominate SA source during the daytime, SCIs can be neglected, and the equation further simplifies to Eq. 10. Neglecting nucleation, which is a valid assumption on most non-NPF days, and the relatively minor contribution of SCIs during daytime conditions yields the simplest and widely used daytime SA proxy[20,24,27,30], shown in Eq. 11:

$$[SA] = -\frac{CS}{2\times k_n} + \left[\left(\frac{CS}{2\times k_n}\right)^2 + \frac{\left[SO_2\right]}{k_n}\times\left(k\times[OH] + k_{SCIs}\times[SCIs]\right)\right]^{1/2} \tag{8}$$

$$[SA] = \left(k\times[OH] + k_{SCIs}\times[SCIs]\right)\times\left[SO_2\right]\times CS^{-1} \tag{9}$$

$$[SA] = -\frac{CS}{2\times k_n} + \left[\left(\frac{CS}{2\times k_n}\right)^2 + \frac{\left[SO_2\right]}{k_n}\times(k\times[OH])\right]^{1/2} \tag{10}$$

$$[SA] = k\times[OH]\times\left[SO_2\right]\times CS^{-1} \tag{11}$$

Direct measurements of OH radicals and SCIs remain technically challenging and less practical than SA measurements. However, previous studies indicate that daytime OH radical levels correlate strongly with SR, particularly UVB, due to its formation via O_3 and HONO photolysis, and the reaction between HO_2 with NO[26,30,54,55]. Since only SR was measured during the two field campaigns, SR was used as an indicator for daytime OH concentrations in the present study. SCIs, mainly generated from the ozonolysis of alkenes[56], were represented using the product of O_3 and alkene concentrations ([O_3] × [Alkene])[17,20,24,26,27,30]. At both CD and HKUST sites, isoprene was the dominant alkene species observed during the campaigns, with average concentrations of 0.81 ± 0.48 ppb and 0.28 ± 0.21 ppb, much higher than monoterpenes. Moreover, isoprene reacts rapidly with O_3 and produces SCIs with high yield, making it a reasonable proxy variable for alkene-derived SCIs under subtropical coastal conditions[41,45].

Then, a series of gaseous SA proxies can be developed based on Eq. 1, Eq. 8 to Eq. 11, and the indicators for OH radical and SCIs. For the daytime, a general proxy expressed as Eq. 1 can be used. Three linear regression-based proxies (D_{L1}-D_{L3}) can be obtained based on Eq. 1 and Eq. 11, as summarized in Table 1. In these equations, k is derived from Eq. 4. Proxy D_{L1} is based on the fundamental theory and assumes a linear relationship between SR and OH. D_{L2} represents a simplified proxy excluding CS, based on findings from Kuerten et al.[27] that this omission still performed adequately when CS variability was small. D_{L3} incorporates RH, suggested by Mikkonen et al.[24], who noted its influence on aerosol distribution and CS. To account for the non-equilibrium conditions and the existence of other influencing and competing pollutants in the SO_2 + OH reactions[24], nonlinear proxies with individual powers for proxy variables were also derived and tested, including proxies D_{NL1}-D_{NL5}. In addition, considering nucleation as a sink of SA, D_{NL6} could be derived from Eq. 10[17].

The CS calculation requires continuous measurement of SMPS; however, such instrumentation is not available at many air quality monitoring stations, and long-term data is often limited. This limitation reduces the practical applicability of CS-based proxies. Despite greatly affected by the size distribution, CS has been shown to correlate positively with $PM_{2.5}$ concentration in our measurement (Fig. S6). Therefore, alternative proxies ($D_{NL1\text{-}PM}$, $D_{NL4\text{-}PM}$, and $D_{NL5\text{-}PM}$) were constructed by substituting CS with $PM_{2.5}$ in the corresponding nonlinear proxies, i.e., D_{NL1}, D_{NL4}, and D_{NL5}. $PM_{2.5}$ data were collected with a time resolution of 2 min and subsequently averaged to 10 min to match the proxy construction interval.

In addition, to simulate the SA concentrations through the entire day, we also applied a function derived from Eq. 8, designated as proxy A_1[17]. However, regression results indicated that A_1 was insufficient for capturing SA variability over a full day, particularly during nighttime periods. This limitation will be further explored in the results section. The inadequacy of A_1 stems from its exclusive consideration of SCIs as nighttime contributors to SA formation. Several studies have shown that OH radical, typically associated with daytime photochemistry, may also be generated during the night, through the oxidation of VOCs by NO_3 radical and O_3, resulting in non-negligible nighttime OH values. For example, Zou et al.[36] reported an average nighttime OH concentration of $5.1 \pm 1.8 \times 10^5$ molecule cm^{-3} during the 2020 autumn field campaign at the CD supersite. Similar findings from Myers et al.[57] highlight the significance of nighttime OH in SA formation in the Amazon forest.

Given these insights, we attempted to identify suitable indicators of nighttime OH levels that could be used for constructing a nighttime proxy for SA, and more detailed discussions are provided in the SI. By replacing [OH] in Eq. 9 with an empirical expression of $a' \times [In.]^{b'} + g$, proxy N can be derived for nighttime SA estimation, in which [In.] is the concentration of selected chemical indicators, including benzene, toluene, and NO_x (molecule/cm^3). To provide a complete daily representation of SA formation, we developed a piecewise function, designated as A_2, which combines the daytime proxy D_{NL1} with $c = 1$ and the nighttime proxy N. The reason for assigning c to 1 is to keep the exponents of variables coexisted in the daytime and nighttime part of A_2 with the same values. Additionally, PM was also tested as an alternative to CS in constructing parallel versions of the nighttime and whole-day proxies, referred to as $N_{\text{-PM}}$ and $A_{2\text{-PM}}$, respectively.

To assess the general applicability of these proxies, we divided the collected dataset into two parts: one for the proxies fitting and derivation, and the other for validation. The proxy parameters were derived using data collected at the CD supersite from 8th November to 3rd December 2018, and validation was performed using data collected at the same location from 4th to 19th December 2018. To further test transferability across environments, further verification was conducted using data collected at the HKUST supersite from 14th November to 19th December 2022. Proxy performance was assessed by comparing reconstructed and measured SA concentration, $[SA]_s$ and $[SA]_m$. The relative error (RE) was calculated by Eq. 12, where Num. is the total number of the SA samples[19].

$$RE = \frac{1}{Num.} \sum \frac{|[SA]_s - [SA]_m|}{[SA]_m} \tag{12}$$

Data availability

All experimental data can be obtained at CHEN, Yi, 2025, "Ambient Measurements and Improved Statistical Proxies for Gaseous Sulfuric Acid in Coastal Hong Kong", https://doi.org/10.14711/dataset/KECGV2, DataSpace@HKUST, V1, UNF:6:o3ANOzamePIdleOAnr4+FQ = = [fileUNF].

Received: 30 April 2025; Accepted: 13 November 2025;

References

1. Kuang, C., McMurry, P. H., McCormick, A. V. & Eisele, F. L. Dependence of nucleation rates on sulfuric acid vapor concentration in diverse atmospheric locations. *J. Geophys. Res*. **113** https://doi.org/10.1029/2007jd009253 (2008).
2. Sipilä, M. et al. The role of sulfuric acid in atmospheric nucleation. *Science* **327**, 1243–1246 (2010).
3. Berndt, T., Böge, O., Stratmann, F., Heintzenberg, J. & Kulmala, M. Rapid formation of sulfuric acid particles at near-atmospheric conditions. *Science* **307**, 698–700 (2005).
4. Lee, S. H. et al. New particle formation in the atmosphere: from molecular clusters to global climate. *J. Geophys. Res.: Atmos.* **124**, 7098–7146 (2019).
5. Yao, L. et al. Atmospheric new particle formation from sulfuric acid and amines in a Chinese megacity. *Science* **361**, 278–281 (2018).
6. Cai, R. et al. Sulfuric acid–amine nucleation in urban Beijing. *Atmos. Chem. Phys.* **21**, 2457–2468 (2021).
7. Lehtipalo, K. et al. Multicomponent new particle formation from sulfuric acid, ammonia, and biogenic vapors. *Sci. Adv.* **4**, eaau5363 (2018).
8. Vuollekoski, H. et al. Atmospheric nucleation and initial steps of particle growth: Numerical comparison of different theories and hypotheses. *Atmos. Res.* **98**, 229–236 (2010).
9. Wang, Z. B. et al. The simulations of sulfuric acid concentration and new particle formation in an urban atmosphere in China. *Atmos. Chem. Phys.* **13**, 11157–11167 (2013).
10. Kyrö, E. M. et al. Trends in new particle formation in eastern Lapland, Finland: effect of decreasing sulfur emissions from Kola Peninsula. *Atmos. Chem. Phys.* **14**, 4383–4396 (2014).
11. Zheng, P. et al. Molecular characterization of oxygenated organic molecules and their dominating roles in particle growth in Hong Kong. *Environ. Sci. Technol.* **57**, 7764–7776 (2023).
12. Olin, M. et al. Traffic-originated nanocluster emission exceeds H_2SO_4-driven photochemical new particle formation in an urban area. *Atmos. Chem. Phys.* **20**, 1–13 (2020).
13. Boy, M. et al. Oxidation of SO_2 by stabilized Criegee intermediate (sCI) radicals as a crucial source for atmospheric sulfuric acid concentrations. *Atmos. Chem. Phys.* **13**, 3865–3879 (2013).
14. Huang, H. L., Chao, W. & Lin, J. J. Kinetics of a Criegee intermediate that would survive high humidity and may oxidize atmospheric SO_2. *Proc. Natl. Acad. Sci. USA* **112**, 10857–10862 (2015).
15. Aaltonen, E. T. & Francisco, J. S. $HOSO_2$-H_2O radical complex and its possible effects on the production of sulfuric acid in the atmosphere. *J. Phys. Chem. A* **107**, 1216–1221 (2003).
16. Jayne, J. T. et al. Pressure and temperature dependence of the gas-phase reaction of SO_3 with H_2O and the heterogeneous reaction of SO_3 with H_2O/H_2SO_4 surfaces. *J. Phys. Chem. A* **101**, 10000–10011 (1997).
17. Dada, L. et al. Sources and sinks driving sulfuric acid concentrations in contrasting environments: implications on proxy calculations. *Atmos. Chem. Phys.* **20**, 11747–11766 (2020).
18. Guo, Y. et al. Formation of nighttime sulfuric acid from the ozonolysis of alkenes in Beijing. *Atmos. Chem. Phys.* **21**, 5499–5511 (2021).
19. Yang, L. et al. Toward building a physical proxy for gas-phase sulfuric acid concentration based on its budget analysis in polluted Yangtze River Delta, East China. *Environ. Sci. Technol.* **55**, 6665–6676 (2021).
20. Petäjä, T. et al. Sulfuric acid and OH concentrations in a boreal forest site. *Atmos. Chem. Phys.* **9**, 7435–7448 (2009).
21. Eisele, F. L. & Tanner, D. J. Measurement of the gas phase concentration of H_2SO_4 and methane sulfonic acid and estimates of H_2SO_4 production and loss in the atmosphere. *J. Geophys. Res.* **98**, 9001–9010 (1993).
22. Berresheim, H., Elste, T., Plass-Dülmer, C., Eiseleb, F. L. & Tannerb, D. J. Chemical ionization mass spectrometer for long-term measurements of atmospheric OH and H_2SO_4. *Int. J. Mass spectrom.* **202**, 91–109 (2000).
23. Erupe, M. E. et al. Correlation of aerosol nucleation rate with sulfuric acid and ammonia in Kent, Ohio: An atmospheric observation. *J. Geophys. Res*. **115** https://doi.org/10.1029/2010jd013942 (2010).
24. Mikkonen, S. et al. A statistical proxy for sulphuric acid concentration. *Atmos. Chem. Phys.* **11**, 11319–11334 (2011).
25. Rose, C. et al. Investigation of several proxies to estimate sulfuric acid concentration under volcanic plume conditions. *Atmos. Chem. Phys.* **21**, 4541–4560 (2021).
26. Lu, Y. et al. A proxy for atmospheric daytime gaseous sulfuric acid concentration in urban Beijing. *Atmos. Chem. Phys.* **19**, 1971–1983 (2019).
27. Kuerten, A. et al. Observation of new particle formation and measurement of sulfuric acid, ammonia, amines and highly oxidized organic molecules at a rural site in central Germany. *Atmos. Chem. Phys.* **16**, 12793–12813 (2016).

28. Li, X. et al. Responses of gaseous sulfuric acid and particulate sulfate to reduced SO(2) concentration: a perspective from long-term measurements in Beijing. *Sci. Total Environ.* **721**, 137700 (2020).
29. Kim, S. et al. Potential Role of Stabilized Criegee Radicals in Sulfuric Acid Production in a High Biogenic VOC Environment. *Environ. Sci. Technol.* **49**, 3383–3391 (2015).
30. Größ, J. et al. Atmospheric new particle formation at the research station Melpitz, Germany: connection with gaseous precursors and meteorological parameters. *Atmos. Chem. Phys.* **18**, 1835–1861 (2018).
31. Chao, W., Hsieh, J.-T., Chang, C.-H. & Lin, J. J.-M. Direct kinetic measurement of the reaction of the simplest Criegee intermediate with water vapor. *Science* **347**, 751–754 (2015).
32. Lewis, T. R., Blitz, M. A., Heard, D. E. & Seakins, P. W. Direct evidence for a substantive reaction between the Criegee intermediate, CH2OO, and the water vapour dimer. *Phys. Chem. Chem. Phys.* **17**, 4859–4863 (2015).
33. Lin, L. C. et al. Competition between H2O and (H2O)(2) reactions with CH2OO/CH3CHOO. *Phys. Chem. Chem. Phys.* **18**, 4557–4568 (2016).
34. Welz, O. et al. Direct kinetic measurements of Criegee intermediate (CH_2OO) formed by reaction of CH_2I with O_2. *Science* **335**, 204–207 (2012).
35. *IUPAC Task Group on Atmospheric Chemical Kinetic Data Evaluation* https://iupac-aeris.ipsl.fr/datasheets/pdf/CGI_2.pdf (2025).
36. Zou, Z. et al. OH measurements in the coastal atmosphere of South China: possible missing OH sinks in aged air masses. *Atmos. Chem. Phys.* **23**, 7057–7074 (2023).
37. Novelli, A. et al. Importance of isomerization reactions for OH radical regeneration from the photo-oxidation of isoprene investigated in the atmospheric simulation chamber SAPHIR. *Atmos. Chem. Phys.* **20**, 3333–3355 (2020).
38. Ding, A., Wang, T. & Fu, C. Transport characteristics and origins of carbon monoxide and ozone in Hong Kong, South China. *J. Geophys. Res.: Atmos.* **118**, 9475–9488 (2013).
39. Chen, Y. et al. Secondary formation and impacts of gaseous nitro-phenolic compounds in the continental outflow observed at a background site in south China. *Environ. Sci. Technol.* **56**, 6933–6943 (2022).
40. Nie, W. et al. Secondary organic aerosol formed by condensing anthropogenic vapours over China's megacities. *Nat. Geosci.* **15**, 255–261 (2022).
41. Hui, L. et al. Abundant oxygenated volatile organic compounds and their contribution to photochemical pollution in subtropical Hong Kong. *Environ. Pollut.* **335**, 122287 (2023).
42. Kuerten, A., Rondo, L., Ehrhart, S. & Curtius, J. Calibration of a chemical ionization mass spectrometer for the measurement of gaseous sulfuric acid. *J. Phys. Chem. A* **116**, 6375–6386 (2012).
43. Pu, W. et al. Development of a chemical ionization mass spectrometry system for measurement of atmospheric OH radical. *Atmos. Meas. Tech. Discuss*. https://doi.org/10.5194/amt-2020-252 (2020).
44. Yuan, Q. et al. Origin and transformation of volatile organic compounds at a regional background site in Hong Kong: Varied photochemical processes from different source regions. *Sci. Total Environ.* **908**, 168316 (2024).
45. Tan, Y. et al. Characteristics and source apportionment of volatile organic compounds (VOCs) at a coastal site in Hong Kong. *Sci. Total Environ.* **777**, 146241 (2021).
46. Wang, Q. et al. Chemical evolution of secondary organic aerosol tracers during high-PM2.5 episodes at a suburban site in Hong Kong over 4 months of continuous measurement. *Atmos. Chem. Phys.* **22**, 11239–11253 (2022).
47. Jenkin, M. E., Young, J. C. & Rickard, A. R. The MCM v3.3.1 degradation scheme for isoprene. *Atmos. Chem. Phys.* **15**, 11433–11459 (2015).
48. Saunders, S. M., Jenkin, M. E., Derwent, R. G. & Pilling, M. J. Protocol for the development of the Master ChemicalMechanism, MCMv3 (Part A): Tropospheric degradation of non-aromatic volatile organic compounds. *Atmos. Chem. Phys.* **3**, 161–180 (2003).
49. Jenkin, M. E., Saunders, S. M., Wagner, V. & Pilling, M. J. Protocol for the development of the Master ChemicalMechanism, MCMv3 (Part B): Tropospheric degradation of aromatic volatile organic compounds. *Atmos. Chem. Phys.* **3**, 181–193 (2003).
50. Bloss, C. et al. Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons. *Atmos. Chem. Phys.* **5**, 641–664 (2005).
51. Ling, Z. et al. Formation and sink of glyoxal and methylglyoxal in a polluted subtropical environment: observation-based photochemical analysis and impact evaluation. *Atmos. Chem. Phys.* **20**, 11451–11467 (2020).
52. He, Z. et al. Contributions of different anthropogenic volatile organic compound sources to ozone formation at a receptor site in the Pearl River Delta region and its policy implications. *Atmos. Chem. Phys.* **19**, 8801–8816 (2019).
53. Kulmala, M. et al. On the formation, growth and composition of nucleation mode particles. *Tellus B* **53**, 479–490 (2001).
54. Yang, X. et al. Observations and modeling of OH and HO(2) radicals in Chengdu, China in summer 2019. *Sci. Total Environ.* **772**, 144829 (2021).
55. Rohrer, F. & Berresheim, H. Strong correlation between levels of tropospheric hydroxyl radicals and solar ultraviolet radiation. *Nature* **442**, 184–187 (2006).
56. Taatjes, C. A. Criegee intermediates: What direct production and detection can teach us about reactions of carbonyl oxides. *Annu. Rev. Phys. Chem.* **68**, 183–207 (2017).
57. Myers, D. C. et al. Sulfuric acid in the Amazon basin: measurements and evaluation of existing sulfuric acid proxies. *Atmos. Chem. Phys.* **22**, 10061–10076 (2022).
58. Yang, C. et al. Machine Learning Reveals the Parameters Affecting the Gaseous Sulfuric Acid Distribution in a Coastal City: Model Construction and Interpretation. *Environ. Sci. Tech. Lett.* **10**, 1045–1051 (2023).

Acknowledgements

This study is supported by the National Natural Science Foundation of China (42122062), the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China (16209022, 16201623, 16211824), the Hong Kong Environmental Protection Department (EPD), and the Guangdong Natural Science Foundation (GDST23SC13). The authors would like to acknowledge the Hong Kong EPD and the Environmental Central Facility of HKUST for providing the air quality supersite and equipment support on ambient measurement.

Author contributions

Y.C., Z.W., P.Z., and J.Z. conducted the field campaign; X.F., Y.T., T.W., and S.L. provided the trace gases, meteorological parameters and VOCs data; Y.C. did the data analysis and performed the box model analysis; J.Z.Y. contributed valuable comments to the writing of this paper; Z.W. and T.W. provided the funding sources; Y.C. wrote the original draft, and all the authors participated in the review and editing. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44407-025-00041-6.

Correspondence and requests for materials should be addressed to Zhe Wang.