Supplement of Atmos. Chem. Phys., 26, 9857–9878, 2026
https://doi.org/10.5194/acp-26-9857-2026-supplement

Supplement of

Characteristics of marine aerosols and cloud condensation nuclei measured during the cruise of R/V *ISABU* in 2024: from the East China Sea to the Indian Ocean

Chanwoo Ahn et al.

Correspondence to: Seong Soo Yum (ssyum@kist.re.kr)

Section S1. Correction procedure of the SMPS data

The aerosol number concentrations measured by the SMPS sometimes exhibited intermittent and discontinuous fluctuations of nearly an order of magnitude. Each set of aerosol size distributions before and after these discontinuity points was similar to others within the set (Fig. S2a). However, the non-refractory PM1 (NR-PM1) was measured by the AMS continuously at the times when there were discontinuities exhibited in the SMPS data (Fig. S3). Furthermore, when compared with CCN number concentrations, the aerosol number concentrations at low levels (corresponding to the red lines in Fig. S2a) were significantly lower than CCN number concentrations. Therefore, these SMPS samples were regarded as anomalous measurements and required correction. Meanwhile, Figure S2b shows that the mean aerosol size distributions for each set of samples before and after the discontinuity points were consistent. These results demonstrate that a constant ratio of change occurred across all size bins, suggesting that a correction based on number concentration can be justified. Unfortunately, because a condensation particle counter directly measuring aerosol number concentrations was not available, the correction relied entirely on data from an Aerodynamic Particle Sizer (APS; model 3321, TSI Inc., USA), which continuously measured aerodynamic diameter-based size distributions (500–20,000 nm) without interruption throughout the cruise. The correction procedure was based on the aerosol size distribution merging procedure proposed by Beddows et al. (2010), modified to suit the objectives of this study. The entire procedure is described as a series of five steps, including preprocessing, sample selection, correction, and validation.

Step 0. Preprocessing and background information

Because the time resolutions of the SMPS and APS differed (3 min and 1 min, respectively) and their measurement start times were not synchronized, the two size distributions could not be compared directly on a one-to-one basis. Consequently, the correction was performed using size distributions averaged over 30 min intervals from 15:00 on March 29 (UTC). Although APS data were unavailable from 04:51 on March 28 (the start of the cruise) to 15:00 on March 29, no anomalous SMPS samples were identified during this period.

To ensure accurate fitting between the SMPS and APS size distributions, Savitzky–Golay filtering was applied to smooth the SMPS size distribution, and the 10.6–14.6 nm range was excluded. The quality of fit (QOF) was calculated according to Equation S1; since it is RMSE-based, a smaller value indicates a better fitting performance.

$$QOF = \frac{RMSE_N}{\sum N} + \frac{RMSE_V}{\sum A} + \frac{RMSE_V}{\sum V}, \quad \text{(S1)}$$

where N, A, and V denote the number, surface area, and volume concentrations, respectively. Because the number concentrations of large particles were low, which can lead to overfitting, surface area and volume concentrations were additionally incorporated.

Step 1. Determination of correction factors for small-size bins of the APS

Due to the scattering detection limit of the 655 nm laser used in the APS, undercounting frequently occurs in the first few size bins. So, concentration corrections were applied to the six bins spanning 542–777 nm. The first bin was excluded because it was especially noisy and the particle size could not be reliably determined.

For comparisons based on the mobility diameter, the aerodynamic diameters measured by the APS were converted to mobility diameters using Equation S2. Since the optimal effective density (ρ_{eff}) had not yet been determined in Step 1, the conversion was performed for all ρ_{eff} values ranging from 1.0 to 2.5 with an increment of 0.01, accounting for realistic aerosol mixing states.

$$D_{mobility} = \frac{D_{aerodynamic}}{\sqrt{\rho_{eff}}} \quad \text{(S2)}$$

APS size bins smaller than the upper-limit diameter of the SMPS (478.3 nm) were excluded, and the number–size distributions of the SMPS and APS were fitted using spline interpolation. Based on the best-fit results, correction factors were derived by comparing the fitted values with the measured values in each APS size bin that had been excluded from the fitting. When fewer than six APS bins were excluded, the correction factors for the remaining bins were assigned to be null. This procedure was applied to all samples at 30 min intervals. The final correction factors for each size bin were determined by averaging the calculated values after excluding the upper 10% as outliers.

Step 2. Determination of the optimal ρ_{eff} and identification of anomalous SMPS samples

After applying the correction factors obtained in Step 1 to the APS data, the conversion was repeated for all ρ_{eff} values from 1.0 to 2.5, with an increment of 0.01. The number, surface area, and volume–size distributions were then each fitted using spline interpolation. Unlike

Step 1, all APS size bins except the first were used in the fitting. The median of the ρ_{eff} values from the best-fit results was determined as the optimal ρ_{eff}.

When the number concentrations in the SMPS size distribution are relatively small, the APS size distribution must be shifted toward smaller diameters to achieve optimal fitting, which requires a larger ρ_{eff}. However, if the number concentrations are anomalously small, a continuous size distribution cannot be obtained even when applying the maximum ρ_{eff} value of 2.5. In such cases, the QOF value increases significantly. This serves as the key principle for identifying anomalous SMPS samples.

Samples showing an abrupt increase in the QOF were selected for correction, and all such selected samples were found to be consistent with anomalous observational results. Nevertheless, some anomalous samples were not identified by this criterion. These data were excluded from the analysis in this study.

Step 3. Determination of SMPS scaling factors and refinement of scaling resolution

Using the optimal ρ_{eff} obtained in Step 2, the APS aerodynamic diameters were converted to the mobility diameter. The scaling factor for the number concentrations of SMPS was then varied from 1 to 15 in increments of 0.1, and the scaled SMPS and converted APS number, surface area, and volume–size distributions were each fitted using spline interpolation. This fitting procedure was applied exclusively to the selected SMPS samples to be corrected.

For each sample, the optimal scaling factor was calculated from the best-fit result using 30-min averaged data. Therefore, an appropriate transformation was required to apply these factors to the original 3-min resolution data. After examining the deviations between the 30-min averaged aerosol number concentration and the corresponding aerosol number concentrations at 3 min intervals, weighted scaling factors were applied to each of the ten 3-min SMPS samples according to the following equation:

$$c_f = f_i \times c_i,$$

$$f_i = 1 + a_k \times w_i,$$

$$w_i = \left(\frac{c_m}{c_i}\right)^{\beta}, \qquad \text{(S3)}$$

where c_i and c_f denote the aerosol number concentrations of the original 3-min data before and after correction, respectively. f_i represents the scaling factor for the 3-min data, and w_i is the weighting. C_m indicates the 30-min averaged number concentration. a_k is the value that satisfies Equation S4 and is expressed as Equation S5. In this study, since β was set to 1, a_k equals $F-1$.

$$\frac{\sum_i c_i f_i}{\sum_i c_i} = F, \tag{S4}$$

where F represents the scaling factor for the 30-min averaged data.

$$a_k = (F-1) \times \frac{\sum_i c_i}{\sum_i c_i w_i} = (F-1), \tag{S5}$$

In this study, the data were averaged over 30-min intervals starting from the hour; for example, from 00:00 to 00:30 and from 00:30 to 01:00. However, the derived scaling factor can vary depending on the specific 30-min window used. Therefore, to reflect the variability of the original data as closely as possible, the 30-min scaling factors were smoothed before calculating the 3-min scaling factors.

Step 4. Validation of SMPS correction through comparison with mass concentrations

To validate the number concentration correction of SMPS, the number concentrations were converted to mass concentrations using Equation S6, assuming spherical particles. These results were compared with $PM_{2.5}$ data based on the beta attenuation method. Mass concentrations were calculated by combining the SMPS and APS data to ensure consistency within the size range.

$$M = \rho \sum_i V_i N_i(d_i) = \rho \sum_i \frac{\pi}{6} d_i^3 N_i(d_i), \tag{S6}$$

where ρ denotes the aerosol density; in this study, we used a value of 1.7 g cm^{-3}, which is the mean aerosol density calculated from the AMS data throughout the entire cruise. V_i and N_i represent the particle volume and number concentrations in the bin with diameter d_i, respectively.

Figure S4 compares mass concentrations, demonstrating that the results after correction were significantly improved compared to those before correction.

Section S2. Preprocessing procedure for a clustering analysis

Step 1. Translation

A parallel translation was performed to align the starting locations. A simple parallel translation based solely on latitude and longitude was not appropriate because clustering analysis relies on distance, and the actual geographical distance corresponding to one degree of longitude varies with latitude. Therefore, all latitude and longitude coordinates were converted into Cartesian (x, y) coordinates using Equation S7. This transformation ensured that the starting locations of all trajectories were aligned at (0, 0) while preserving the linearity and relative structure of each trajectory.

$$\begin{cases} x_i = R \times \cos(\phi) \times (rad(lon_i) - \lambda) \\ y_i = R \times (rad(lat_i) - \phi) \end{cases}, \qquad \text{(S7)}$$

where R is the radius of the Earth, assumed to be 6371 km in this study. The subscript i denotes the time stamp of a back trajectory (from –72 h to 0 h). ϕ and λ represent the latitude and longitude of the starting location in degrees converted to radians, expressed as $rad(lat_0)$ and $rad(lon_0)$, respectively.

However, applying this transformation has the drawback that the actual geographical information of the trajectory is not directly reflected in the clustering analysis. Nevertheless, we did not use a set of back trajectories calculated with a fixed starting location for the clustering analysis because the trajectory originating from the research vessel's actual location at a specific time may differ from that computed from an assigned starting location. Even if the overall pathway appears similar, there might be differences in the regions traversed along the trajectory. This issue became more pronounced because the cruise in this study involved long-distance crossing over the open ocean.

Step 2. Vectorization

The two-dimensional trajectory data (a $\mathbb{R}^{(N+1)\times 2}$ matrix) were converted into one-dimensional data (a $\mathbb{R}^{2(N+1)}$ row vector) to facilitate the calculation of the Euclidean distances between trajectories, as shown below.

$$\begin{bmatrix} x_0 & y_0 \\ x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix} \Rightarrow [x_0, y_0, x_1, y_1, \cdots, x_N, y_N], \quad \text{(S8)}$$

Step 3. Standardization

All one-dimensional trajectory vectors were standardized for each sea area as follows:

$$X^{(j)} = [x_0, y_0, x_1, y_1, \cdots, x_N, y_N] \Rightarrow Z^{(j)} = [z_{x0}, z_{y0}, z_{x1}, z_{y1}, \cdots, z_{xN}, z_{yN}],$$

$$z_{xi}^{(j)} = \frac{x_i^{(j)} - \mu_{xi}}{\sigma_{xi}}, \quad \text{(S9)}$$

where the superscript (j) denotes the numbering of the back trajectory data for each sea area. μ_{xi} and σ_{xi} represent the mean and standard deviation, respectively, of the x (or y) at the same time stamp i across all trajectories (from $X^{(1)}$ to $X^{(j)}$) for each sea area.

This standardization prevented distance-based clustering from being overly sensitive to the variable with an extensive range of values, that is, the variable (latitude or longitude) that exhibits a greater spread.

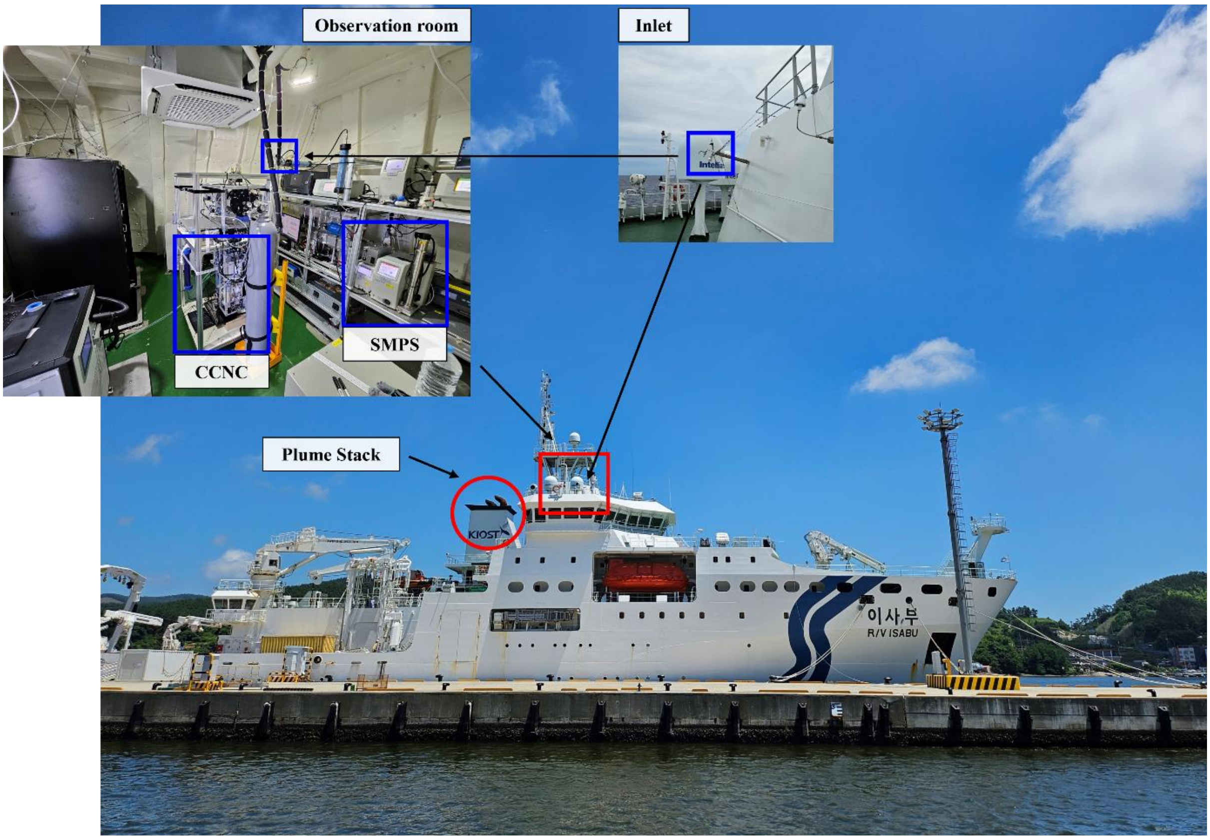

Figure S1. Picture of R/V *ISABU*. The red rectangle denotes the observation room with inlets, and the four blue rectangles in the interior photograph indicate the SMPS, CCNC, and inlets. The red circle represents the vessel's plume stack.

Table S1. Transit voyage schedule of the R/V *ISABU*, including the demarcation by sea area.

	Start	End
Total	2024-03-28 04:51:00 (34.993°N, 128.676°E)	2024-04-17 05:50:00 (20.152°S, 57.494°E)
East China Sea	2024-03-28 19:47:29 (33.000°N, 127.588°E)	2024-03-30 15:29:08 (25.000°N, 123.604°E)
South China Sea	2024-03-31 17:32:59 (20.593°N, 121.083°E)	2024-04-04 07:19:20 (7.068°N, 109.000°E)
Strait of Malacca	2024-04-05 18:14:40 (1.311°N, 104.322°E)	2024-04-07 20:57:46 (6.153°N, 95.088°E)
Indian Ocean	2024-04-07 20:57:47 (6.153°N, 95.088°E)	2024-04-16 23:45:35 (20.000°S, 57.295°E)

Note. All times are in UTC. Because the sea areas near Korea and Mauritius were excluded from the East China Sea and the Indian Ocean, respectively, the start or end times for these regions do not coincide with those of the total.

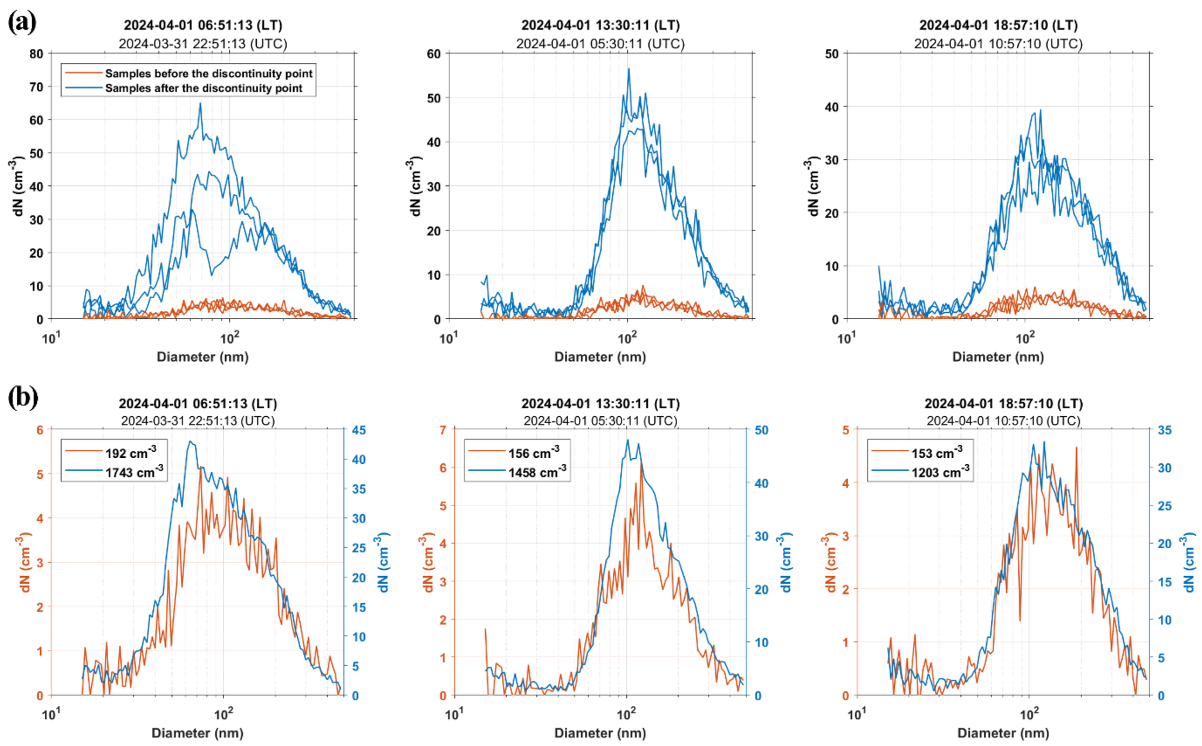

Figure S2. (a) Examples of aerosol size distributions for three samples before (red) and after (blue) the discontinuity point. (b) Mean aerosol size distributions for each set of samples before and after the discontinuity point represented in (a). The values in the legend indicate the mean aerosol number concentrations before and after the discontinuity point, respectively. The times in the titles represent the exact moment the data discontinuity occurred.

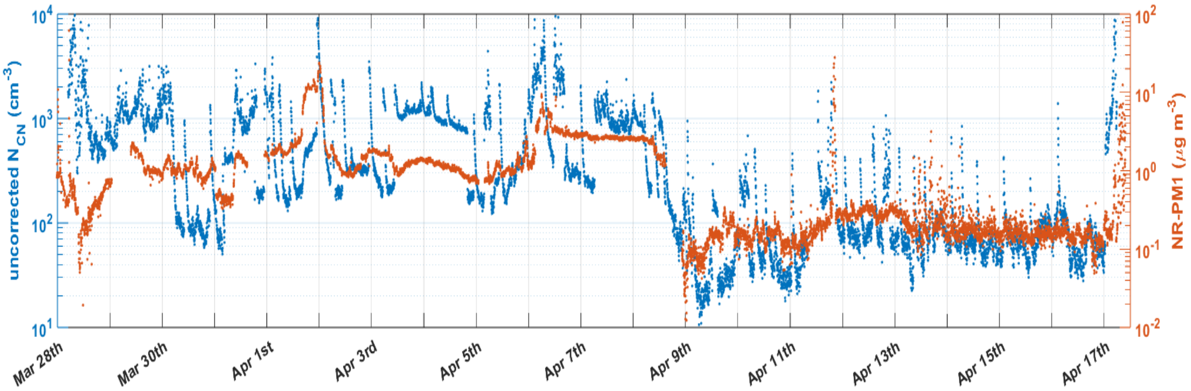

Figure S3. Time series of (left) total aerosol number concentrations before the correction of SMPS and (right) NR-PM1 measured by the AMS.

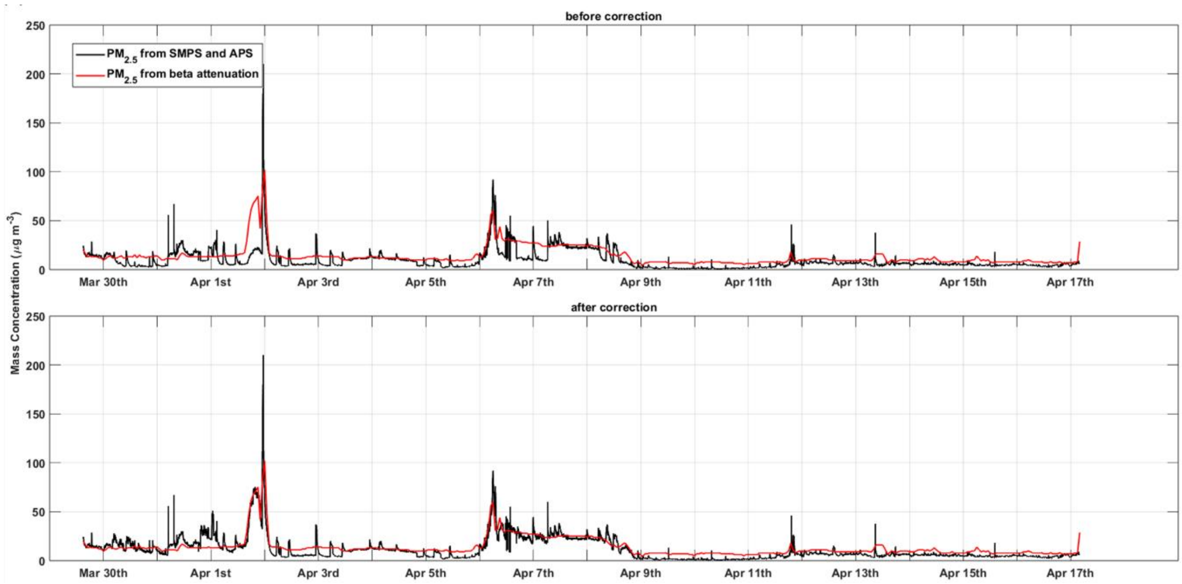

Figure S4. Comparison of mass concentrations of particles smaller than 2.5 μm before (top) and after (bottom) the number concentration correction of SMPS. The black and red lines represent $PM_{2.5}$ calculated by combining SMPS and APS data and $PM_{2.5}$ measured using the beta attenuation method, respectively.

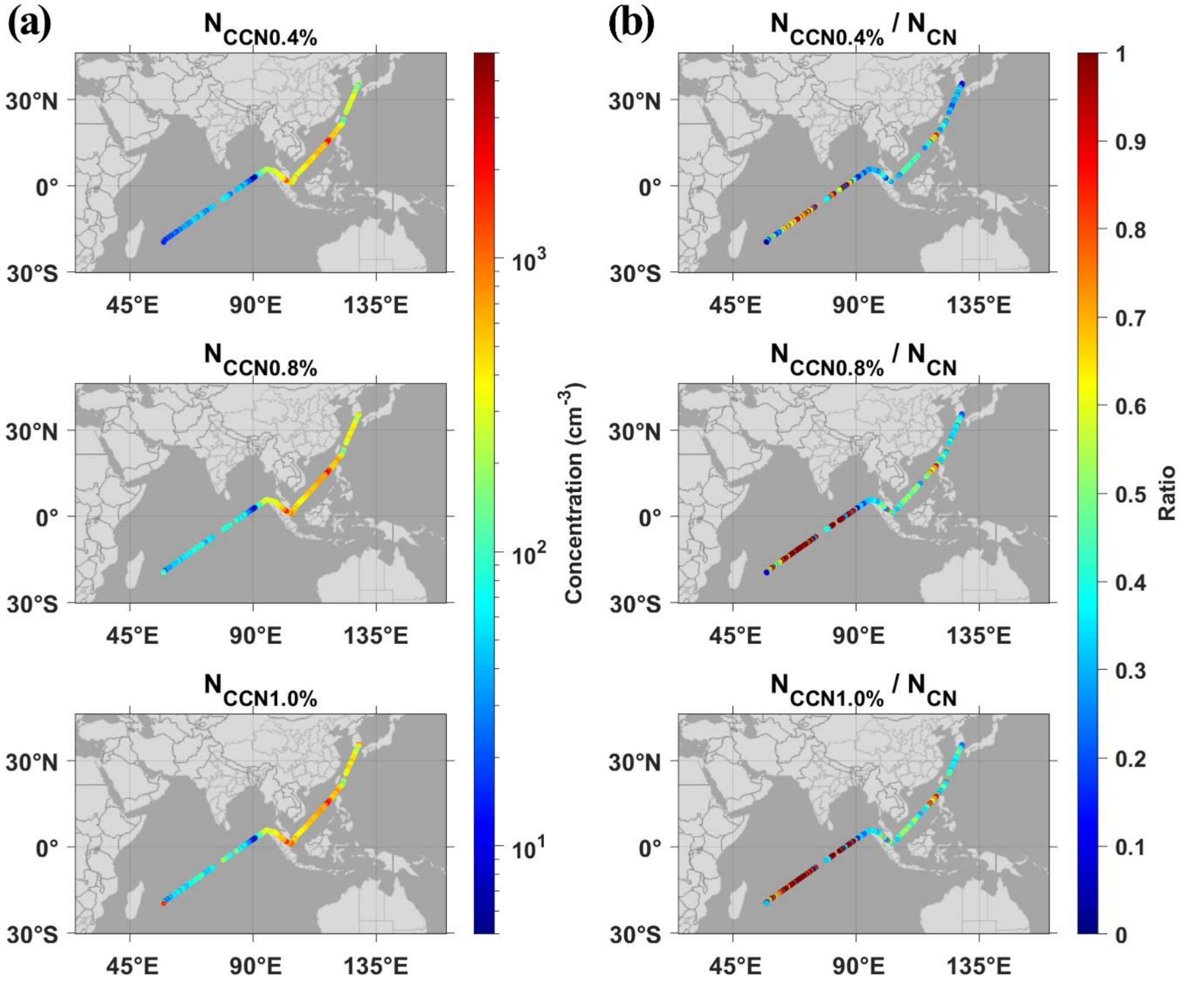

Figure S5. Spatial distributions of (a) CCN number concentrations at 0.4%, 0.8%, and 1.0% supersaturation and (b) their ratios to total aerosol number concentrations during the transit voyage of the R/V *ISABU* (from Natural Earth).

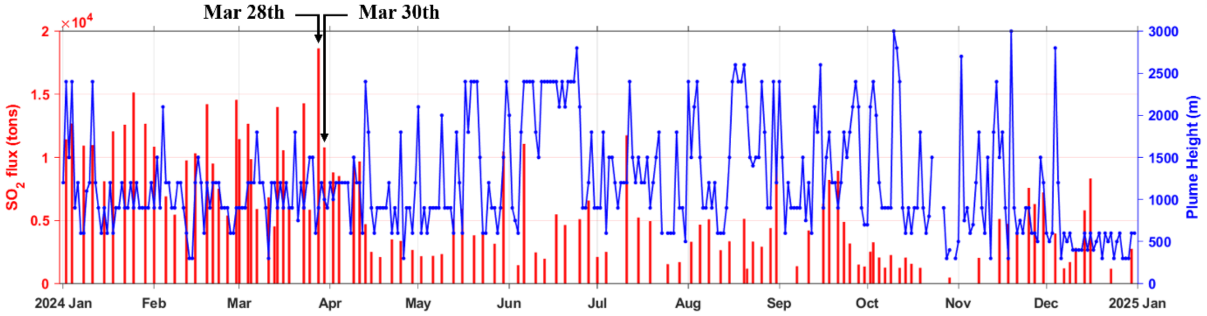

Figure S6. Time series of SO_2 flux (red) and plume height (blue) of the Taal volcano in 2024. It should be noted that the absence of data does not necessarily imply the absence of SO_2 emissions, as SO_2 flux measurements were not conducted daily. The arrows indicate March 28 and 30, respectively.

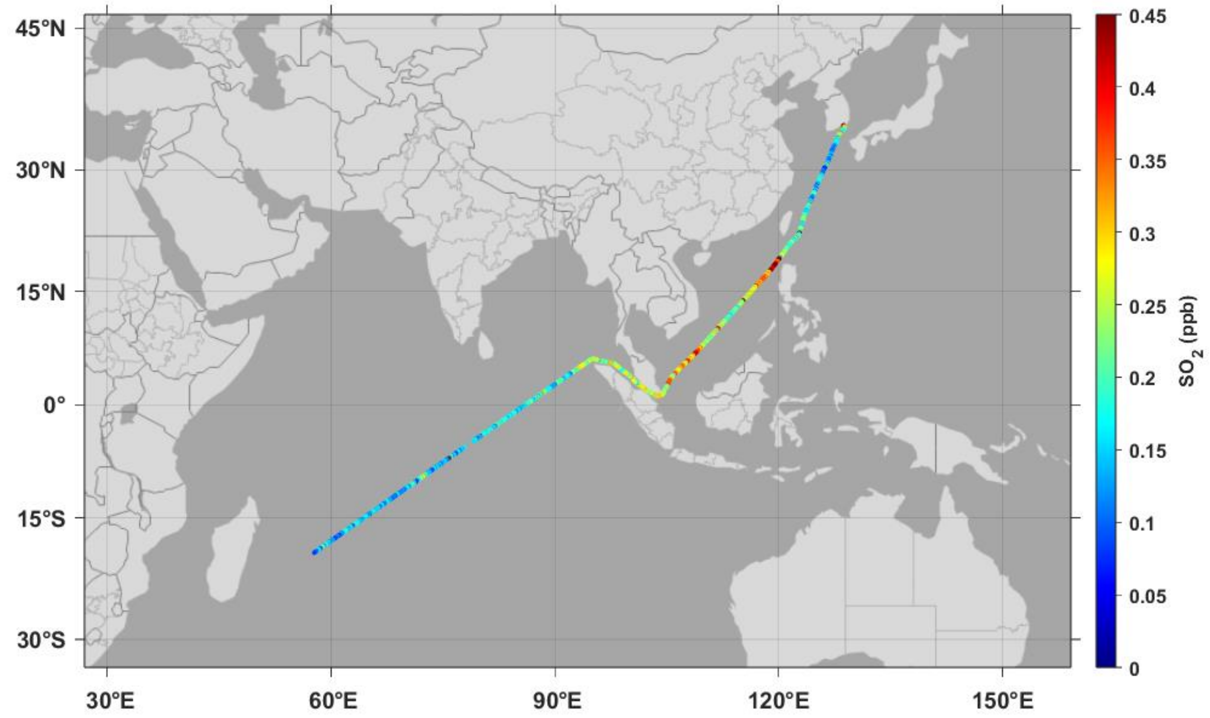

Figure S7. Spatial distribution of SO_2 measured by the Air Quality Monitoring System during the transit voyage of the R/V *ISABU* (from Natural Earth).

Table S2. Ratios of CCN number concentration to total aerosol number concentration (N_{CCN}/N_{CN}) and measurement uncertainties by supersaturation in the South China Sea and the Strait of Malacca.

	South China Sea		Strait of Malacca	
SS (%)	N_{CCN}/N_{CN} ratio	Uncertainty	N_{CCN}/N_{CN} ratio	Uncertainty
0.2	0.23	0.025	0.19	0.021
0.4	0.49	0.053	0.34	0.037
0.6	0.51	0.055	0.39	0.042
0.8	0.52	0.056	0.43	0.046
1.0	0.52	0.056	0.44	0.048

Note. Measurement uncertainties for N_{CN} and N_{CCN} are assumed to be 10% and 4%, respectively, following Wiedensohler et al. (2018) and Uin and Enekwizu (2024). Based on the standard error propagation, the relative uncertainty for N_{CCN}/N_{CN} is 10.8%, and the corresponding absolute uncertainties are presented in the table.

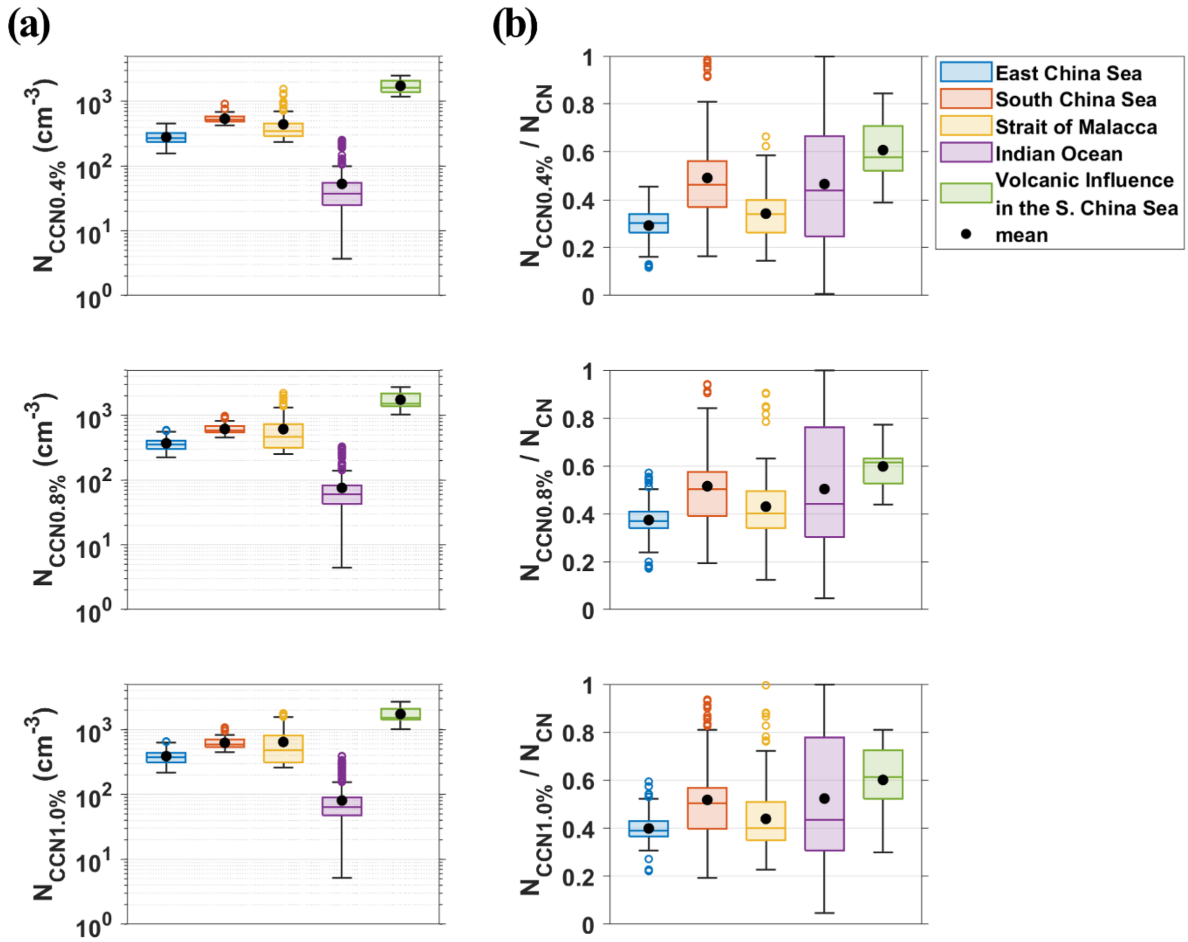

Figure S8. Box plots of (a) CCN number concentrations at 0.4%, 0.8%, and 1.0% supersaturation and (b) their ratios to the total aerosol number concentrations for each sea area. For the South China Sea, the volcanic-influence period was depicted separately. The horizontal line in each box and the black dots represent the median and mean values, respectively. The range of the box is from the 25th percentile to the 75th percentile (IQR). The whiskers extend away from the box to the two extreme values, but if there are data points that exceed $1.5 \times IQR$ from the upper or lower end of the box, they are shown as colored dots as outliers.

Table S3. Summary of fitting parameters and residuals for the lognormal modal fitting in each sea area.

	East China Sea			South China Sea			Strait of Malacca			Indian Ocean			Volcanic Influence		
Fitting parameter															
	N_t	D_{pg}	σ_g	N_t	D_{pg}	σ_g	N_t	D_{pg}	σ_g	N_t	D_{pg}	σ_g	N_t	D_{pg}	σ_g
Nucleation	92.67	24.10	1.50	99.09	18.68	1.50	123.79	15.00	1.32	41.79	16.91	1.50	348.53	16.61	1.50
Aitken	325.39	52.12	1.49	675.62	80.97	1.48	922.87	62.80	1.98	41.38	48.05	1.44	–	–	–
Accumulation	636.96	139.98	1.74	474.98	171.82	1.45	489.72	207.60	1.54	81.74	162.80	1.73	2467.84	165.91	1.43
Coarse	2.71	500.00	1.71	1.63	500	1.88	1.74	500	1.92	1.44	681.24	1.81	1.79	500	1.89
Residuals															
RMSE	4.6953			21.6569			299.9474			8.8404			2.3919		
NRMSE	0.0042			0.0121			0.041			0.0064			0.0166		
MBE	0.5855			–3.3388			–52.4177			0.5868			0.0502		
MARR (15–25)	1.44			14.81			17.31			2.26			1.08		
MARR (25–100)	0.33			4.79			45.36			0.55			0.47		
MARR (100–500)	1.71			9.04			25.25			1.15			3.87		

Note. RMSE and NRMSE are the root-mean-square error and normalized root-mean-square error, respectively. MBE denotes the mean bias error, and MARR denotes the mean absolute relative residual (%).

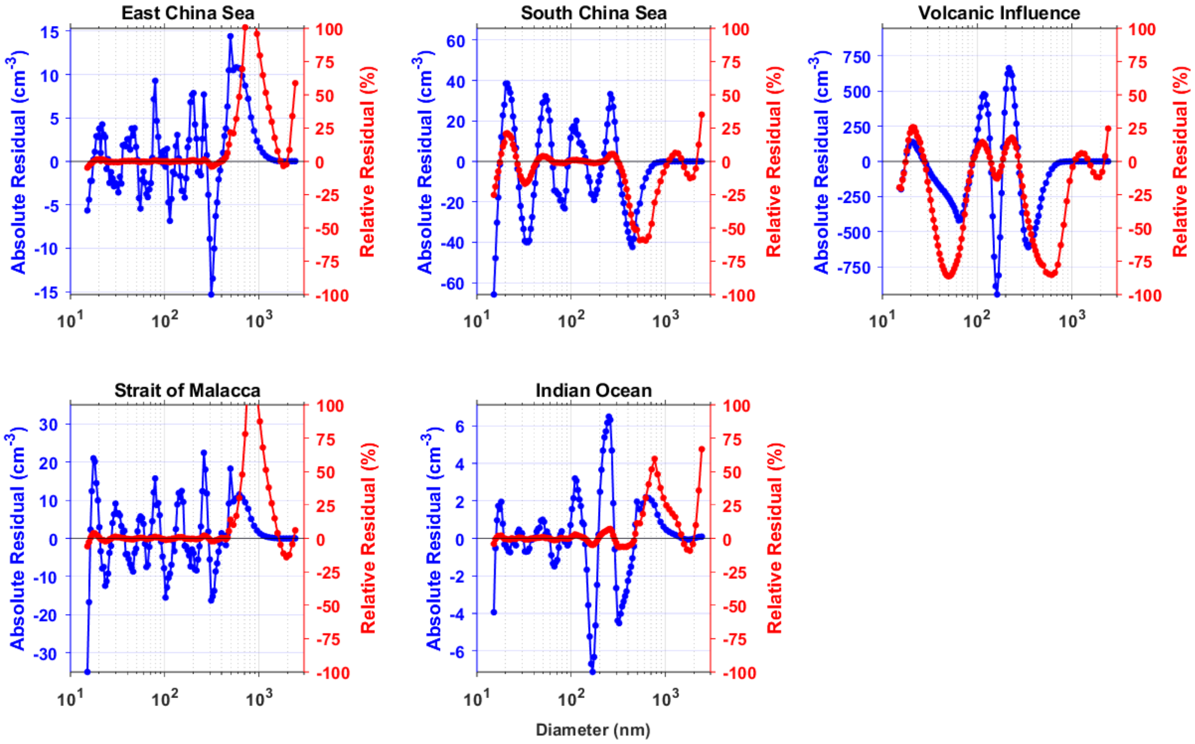

Figure S9. Size-dependent absolute (blue) and relative (red) residuals for the lognormal modal fitting in each sea area.

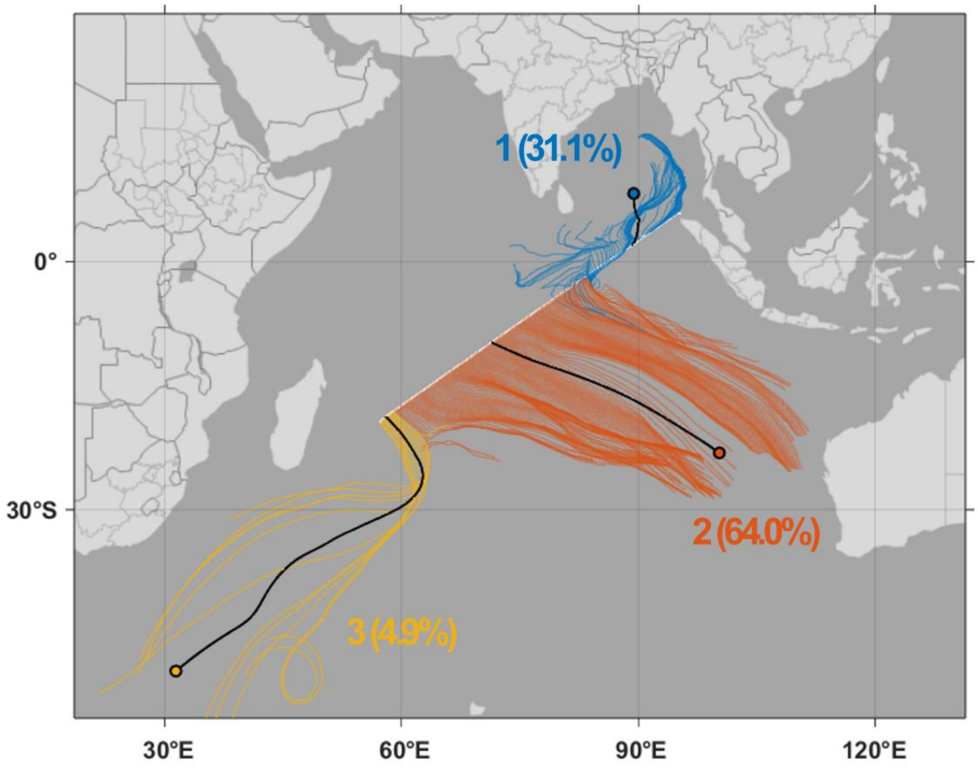

Figure S10. Results of extending each back trajectory to 120 hours in the cluster analysis for the Indian Ocean. The thick black solid lines with dots at the end represent the mean trajectories of each cluster. The percentages in parentheses indicate the proportion of trajectories assigned to each cluster. Map data from Natural Earth.

References

Beddows, D. C. S., Dall'Osto, M., and Harrison, R. M.: An Enhanced Procedure for the Merging of Atmospheric Particle Size Distribution Data Measured Using Electrical Mobility and Time-of-Flight Analysers, Aerosol Sci. Technol., 44, 930–938, https://doi.org/10.1080/02786826.2010.502159, 2010.

Uin, J. and Enekwizu, O. Y. Cloud condensation Nuclei Particle Counter Instrument Handbook, U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington, DOE/SC-ARM-TR-168, https://doi.org/10.2172/1251411, 2024.

Wiedensohler, A., Wiesner, A., Weinhold, K., Birmili, W., Hermann, M., Merkel, M., Müller, T., Pfeifer, S., Schmidt, A., Tuch, T., Velarde, F., Quincey, P., Seeger, S., and Nowak, A.: Mobility Particle Size Spectrometers: Calibration Procedures and Measurement Uncertainties, Aerosol Sci. Technol., 52, 146–164, https://doi.org/10.1080/02786826.2017.1387229, 2017.