

Article

Vertically Resolved Aerosol Optical-State Identification by Multi-Wavelength Raman–Mie Polarization Lidar in Contrasting Inland and Coastal Environments

Zhen Zhang ¹, Zhigang Li ^{1,*}, Zhichao Bu ² and Yaru Dai ²

¹ College of Physics and Optoelectronic Engineering, Ludong University, Yantai 264025, China

² Meteorological Observation Center, China Meteorological Administration, Beijing 100081, China

* Correspondence: lizg@ldu.edu.cn

Abstract

Multi-wavelength Raman–Mie polarization lidar provides vertically resolved measurements of aerosol scattering, particle shape, and wavelength-dependent response. We analyzed quality-controlled clear-sky observations from June 2025 to March 2026 to identify aerosol optical regimes at inland Beijing Nanjiao and coastal Beihai. The feature space comprised $\log_{10}(\beta_{532})$, $\delta_{p,532}$, and $AE_{\beta,355/532}$, representing scattering intensity, particle nonsphericity, and size-sensitive spectral response. K-means was applied independently at each site. At Beijing Nanjiao, a high-depolarization, strong-scattering, and low- $AE_{\beta,355/532}$ regime was concentrated in the lowest observed layer, whereas lower-loading regimes occurred more frequently aloft. Its dust AOD and dust fraction were descriptively 37.5% and 19.2% above the site means, although the dust-related inter-regime differences were not significant after FDR correction. Beihai was dominated by low-depolarization regimes separated mainly by scattering intensity and wavelength response. Its strongest-scattering regime showed total, sea salt, OC, and sulfate AOD enhancements of 33.0%, 15.5%, 24.5%, and 41.9%, respectively; the inter-regime differences were significant for total and sulfate AODs but not for sea salt AOD. MERRA-2 aerosol diagnostics and trajectory analyses provided auxiliary regional context for interpreting the lidar-defined optical states and were not used as clustering inputs or direct chemical validation. These results demonstrate that combined polarization and multi-wavelength lidar sensing can distinguish vertically varying aerosol optical states that are obscured in surface or column-integrated observations.

Keywords: Raman–Mie polarization lidar; multi-parameter optical sensing; aerosol optical regime; vertical structure; inland–coastal comparison



Received: 30 July 2026

Revised: 25 August 2026

Accepted: 26 August 2026

Published: 27 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Atmospheric aerosols affect climate, air quality, visibility, and human health through processes that depend on particle abundance, size, shape, composition, and mixing state [1–7]. Their impacts also depend strongly on vertical position. Aerosol layers near the surface interact directly with population exposure and boundary-layer dynamics, whereas elevated layers can persist during decoupling, undergo long-range transport, and modify radiative heating at levels that are not represented by surface observations. Consequently, similar column aerosol burdens may arise from substantially different vertical distributions and may have different environmental implications.

Conventional aerosol observations constrain only parts of this three-dimensional variability. Surface particulate measurements describe near-ground conditions, while sun-photometer and satellite aerosol optical depth integrate extinction through the atmospheric column [8]. These observations are indispensable for monitoring aerosol loading, but they cannot alone distinguish near-surface accumulation from residual, advected, or elevated aerosol layers. Satellite retrievals also provide broad spatial coverage, although their sampling and vertical information can be limited by clouds, surface reflectance, and overpass timing. Continuous ground-based profiling is therefore needed to resolve how aerosol optical properties vary with altitude and evolve within the lower troposphere.

Ground-based lidar provides active range-resolved optical sensing with vertical resolutions that are suitable for detecting aerosol layers and transitions [9–11]. Coordinated lidar networks have further demonstrated the value of standardized long-term profiling for regional aerosol characterization and intercomparison [12,13]. Multi-wavelength Raman and polarization measurements extend this capability beyond layer detection by providing complementary information on aerosol amount and particle microphysical response. In the present sensing framework, $\log_{10}(\beta_{532})$ describes scattering intensity over a broad dynamic range, $\delta_{p,532}$ constrains particle nonsphericity, and $AE_{\beta,355/532}$ provides a wavelength-dependent response related to particle-size distributions [14–17]. These parameters are not unique chemical tracers. Their joint variation nevertheless provides a physically interpretable optical space in which layers with similar loading but different shape or spectral characteristics can be separated.

Previous lidar studies have established characteristic multi-wavelength and polarization signatures for dust, pollution, marine aerosol, smoke, and their mixtures [18–23]. Observation-based aerosol-typing schemes have used combinations of depolarization, lidar ratio, and spectral response to assign vertically resolved measurements to predefined classes [24,25]. Such approaches demonstrate the diagnostic value of multivariate optical measurements, but they also expose two limitations. First, the optical ranges of mixed or aged aerosols can overlap, especially when humidity and transport alter particle size and shape. Second, supervised class definitions derived for one instrument or environment may not transfer directly to another site. For long time–height records, an unsupervised organization of the measurements can reveal recurring states without imposing externally defined aerosol types. The resulting groups must, however, be interpreted as optical regimes rather than chemically validated compositions.

This distinction is particularly important when comparing inland and coastal environments. Beijing Nanjiao lies within the North China Plain, where urban emissions, regional pollution, boundary-layer evolution, and episodic northern coarse-particle transport jointly shape aerosol profiles [26–29]. Beihai is located on the South China coast, where sea salt production, high humidity, and transport from terrestrial source regions can produce different optical mixtures and vertical structures [30,31]. A common set of lidar observables may therefore describe both sites, while the distribution and physical meaning of their recurring states remain environment dependent. Directly equating cluster labels between the two locations would obscure this dependence and could imply a level of compositional certainty that is not supported by optical measurements alone.

We therefore identify site-specific aerosol optical regimes from the joint lidar feature space and evaluate their altitude-dependent occurrence at Beijing Nanjiao and Beihai. K-means is used as an organizational tool for quality-controlled multi-parameter observations. Each site is standardized and clustered independently, and cross-site comparison is based on measured optical characteristics and vertical occurrence rather than matched cluster labels. Regional aerosol reanalysis and trajectory-based transport diagnostics provide auxiliary physical context, but they are not clustering inputs or direct chemical validation. This

design addresses whether a common multi-wavelength polarization sensing framework can retain environment-specific optical states and vertical transitions across contrasting inland and coastal settings.

2. Data and Methods

2.1. Observation Sites and Data

This study selected Beijing Nanjiao and Beihai as two contrasting ground-based lidar case sites (Figure 1; Table 1). Beijing Nanjiao is located near the urban area of the North China Plain and is influenced by anthropogenic emissions, regional pollution transport, and northern coarse-particle transport. Beihai is located on the South China coast, where marine influence, humid conditions, and terrestrial regional transport provide a distinct environmental setting. The sites are treated here as an inland urban environment and a marine-influenced coastal environment, respectively.

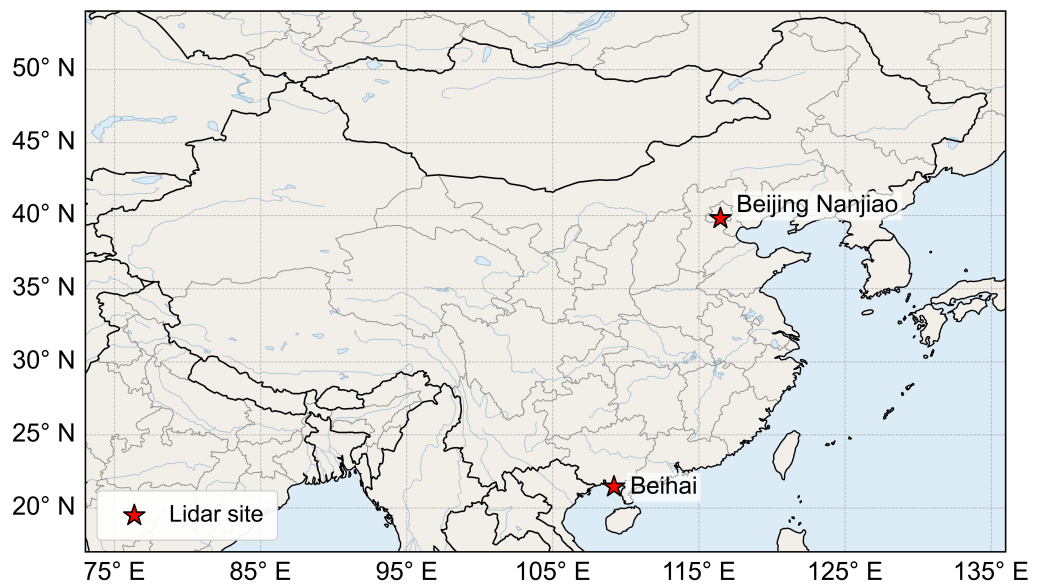


Figure 1. Locations of Beijing Nanjiao and Beihai ground-based lidar observation sites.

Table 1. Observation-site characteristics and quality-controlled lidar data availability.

Site	Environmental Context	Location	Altitude	Valid Days After QC	Valid 15-Minute Profiles	Observation Hours (h)
Beijing Nanjiao	North China urban/regional pollution region	116.4700° E, 39.8067° N	31.5 m	50	3752	938.00
Beihai	South China coastal/marine-influenced region	109.1836° E, 21.4475° N	10.3 m	46	3223	805.75

The observations cover June 2025 to March 2026 and were obtained from CARLNET sites operated by the China Meteorological Administration Meteorological Observation Center. Both sites used the same Nd:YAG-based multi-wavelength Raman–Mie polarization lidar. The system emits at 354.7, 532.1, and 1064.1 nm and receives elastic signals at 355, 532, and 1064 nm, with polarization channels at 355 and 532 nm. Water-vapor Raman detection is provided at 407.6 nm and nitrogen Raman detection at 386.7 and 607.6 nm. The laser operates at 1000 Hz with minimum pulse energies of 0.6/1.5/1.0 mJ and maximum divergence angles of 0.3/0.3/0.5 mrad at 354.7/532.1/1064.1 nm, respectively. A 0.3 m Cassegrain telescope receives the returns within a 1 mrad field of view.

The raw 5 min and 15 m observations were resampled to 15 min and 30 m after common quality control. Each time–height grid point was treated as one sample. Collocated millimeter-wave cloud radar detection and lidar checks removed cloudy, rainy, and abnormal-retrieval periods. Analysis was restricted to 0.3–5.0 km to reduce near-field overlap uncertainty while retaining lower-tropospheric aerosol layers. A dynamic altitude-dependent β_{532} threshold and physically plausible ranges of $\delta_{p,532}$ and $AE_{\beta,355/532}$ removed weak-signal and ratio-amplification artifacts. After screening, 50 and 46 valid clear-sky days remained at Beijing Nanjiao and Beihai, respectively. A total of 3752 and 3223 valid 15-minute lidar profiles were retained, corresponding to 938.00 and 805.75 h of effective observations at Beijing Nanjiao and Beihai, respectively. The data do not represent a continuous annual climatology or seasonal cycle. Because adjacent 15-minute profiles may sample the same aerosol episode and neighboring 30 m range bins may belong to the same vertically coherent aerosol layer, the resulting time–height samples are not statistically independent. K-means is therefore used here as a descriptive tool to organize the observed optical-state space rather than as an inferential method requiring independent samples.

On this basis, a robust z-score method based on the median and median absolute deviation (MAD) was further used to identify statistical outliers before clustering. For each clustering input variable, including $AE_{\beta,355/532}$, $\log_{10}(\beta_{532})$, and $\delta_{p,532}$, the robust z-score was calculated as

$$z_{\text{robust}} = 0.6745 \frac{x_i - \text{median}(x)}{\text{MAD}}, \quad (1)$$

where

$$\text{MAD} = \text{median}(|x_i - \text{median}(x)|). \quad (2)$$

Samples with $|z_{\text{robust}}| > 8$ for any input were removed, while samples with $5 < |z_{\text{robust}}| \leq 8$ were retained only when supported by time–height neighborhood continuity. The threshold of 8 was used as a deliberately conservative operational cutoff rather than as a universal lidar-specific criterion. It was chosen to remove only extreme statistical deviations after the preceding signal-quality and physical-range screening, while samples with $5 < |z_{\text{robust}}| \leq 8$ were not automatically discarded and were retained when supported by neighboring time–height continuity. This conservative treatment was intended to reduce obvious retrieval artifacts without unnecessarily removing infrequent but physically plausible aerosol states. The model was fitted to these quality-controlled core samples and then assigned labels to all the valid samples. High-confidence assignments were those within the cluster-specific 90th percentile of core-sample centroid distances. This internal compactness criterion is not an independent validation of aerosol composition. High-confidence samples were used for feature-space visualization and selected auxiliary analyses. Unless otherwise stated, they were not used to define the denominator of the altitude-dependent occurrence frequencies. These valid observations are intermittent subsets of the June 2025–March 2026 study window rather than uniformly distributed temporal samples. Accordingly, subsequent regime frequencies and vertical occurrence statistics characterize the available quality-controlled clear-sky observations and should not be interpreted as equal-weight seasonal or climatological frequencies.

2.2. Lidar Retrieval Parameters and Clustering Features

The elastic return at range r and wavelength λ is [9–11]

$$P(r, \lambda) = C(\lambda) \beta(r, \lambda) r^{-2} \exp \left[-2 \int_0^r \alpha(r', \lambda) dr' \right], \quad (3)$$

where $C(\lambda)$ is the system constant and β and α are the total backscatter and extinction coefficients. After background subtraction and range-squared correction, β_{532} was retrieved using a Fernald–Klett elastic-backscatter inversion [32,33]. At night, the nitrogen Raman channels provided independent aerosol extinction and backscatter estimates and hence the aerosol lidar ratio [34,35]. Daytime elastic retrievals used lidar ratios informed by nighttime retrievals, contemporaneous weather, and adjacent observations, with permitted ranges of approximately 40–80 sr at 355 nm and 30–70 sr at 532 nm. These choices preceded clustering and did not use cluster labels. The far-end reference region was generally 6–6.5 km, and molecular parameters were obtained from a standard atmospheric model. Daytime backscatter therefore retains uncertainty associated with the assumed lidar ratio. To directly evaluate this sensitivity, selected daytime observations were independently re-inverted using the lower and upper bounds of the prescribed 532 nm lidar-ratio range, $LR_{532} = 30$ and 70 sr. The resulting optical variables and cluster assignments were compared with the corresponding baseline retrievals at identical time–height locations. Because the numerical labels of independently fitted K-means solutions are arbitrary, the alternative-regime labels were matched to the baseline solution by centroid correspondence before calculating ARI and matched-label retention.

The particle depolarization ratio is determined by the volume depolarization ratio, molecular depolarization ratio, and backscatter ratio [16,17,36],

$$\delta_p(r) = \frac{[1 + \delta_m]\delta_v(r)R(r) - [1 + \delta_v(r)]\delta_m}{[1 + \delta_m]R(r) - [1 + \delta_v(r)]}, \quad (4)$$

where $\delta_p(r)$ is the particle depolarization ratio, $\delta_v(r)$ is the volume depolarization ratio, δ_m is the molecular depolarization ratio, and $R(r)$ is the backscatter ratio, defined as the ratio of the total backscatter coefficient to the purely molecular backscatter coefficient. The depolarization ratio was calibrated according to the standard CARLNET processing procedure before calculating the particle depolarization ratio.

The backscatter-related Ångström exponent was calculated using the aerosol backscatter coefficients at 355 and 532 nm:

$$AE_{\beta,355/532}(r) = -\frac{\ln[\beta_{355}(r)/\beta_{532}(r)]}{\ln(355/532)}. \quad (5)$$

We used $\log_{10}(\beta_{532})$ to represent scattering intensity over its broad dynamic range, $\delta_{p,532}$, to constrain particle nonsphericity and $AE_{\beta,355/532}$ to describe the size-sensitive spectral response. The complementary variables form the active optical-sensing space that is used to identify regimes without uniquely retrieving composition [22,23,37,38].

2.3. Identification of Aerosol Optical Regimes

K-means was used to organize continuous lidar observations into optical regimes [39–41]. Candidate values ranged from $K = 2$ to $K = 8$. The cluster number was selected by jointly considering the SSE elbow, silhouette coefficient, resampling stability, and the physical interpretability of the resulting structures in the three-dimensional optical feature space. $K = 4$ was retained because it provided a useful balance between statistical compactness and the resolution of distinct combinations of scattering intensity, particle nonsphericity, and spectral response beyond the broader $K = 2$ partition. The selection was therefore not based on the silhouette coefficient alone [42].

The three inputs were standardized separately at each site,

$$x' = \frac{x - \mu}{\sigma}, \quad (6)$$

where x' is the standardized value and μ and σ are the site-specific mean and standard deviation. Standardization was performed independently for each site using StandardScaler. The mean and standard deviation were estimated from the quality-controlled core samples used for model fitting, and the same site-specific transformation was subsequently applied to all the valid samples before cluster assignment. This prevents variables with larger numerical variance from disproportionately dominating the Euclidean-distance metric. K-means then minimizes the within-cluster sum of squares,

$$J = \sum_{k=1}^K \sum_{x'_j \in C_k} \|x'_j - \mu_k\|^2, \quad (7)$$

where K is the number of clusters, C_k is the k -th cluster, x'_j is a standardized sample, and μ_k is the corresponding centroid. The SSE used for cluster-number evaluation has the same form as Equation (7). The silhouette coefficient of sample j is

$$s(j) = \frac{b(j) - a(j)}{\max[a(j), b(j)]}, \quad (8)$$

where $a(j)$ is the mean distance to samples in the assigned cluster and $b(j)$ is the smallest mean distance to samples in another cluster. Larger values indicate greater within-cluster cohesion and separation. The retained reference solution used K-means++ initialization with a fixed random seed of 42 and 30 internal initializations. Robustness of the $K = 4$ solution was further evaluated using 30 profile-resampling runs. In each run, 80% of the complete 15-minute profiles were randomly sampled without replacement, the features were re-standardized within the subsample, and K-means was refitted using 10 internal initializations. The resulting model was then applied to the full core-sample dataset. Agreement with the reference partition was quantified using the Adjusted Rand Index (ARI), while matched centroids were compared using the maximum centroid shift in the reference-standardized feature space. Thus, this test evaluates sensitivity to both profile sampling and K-means initialization rather than initialization alone. Sensitivity to the selected cluster number was additionally examined by comparing the centroid structures obtained for $K = 3, 4$, and 5. Because cluster indices are arbitrary across independently fitted K-means solutions, the comparison was based on nearest-centroid relationships in the lidar feature space rather than numerical label identity. Physical interpretation of the resulting clusters was based first on the directly retrieved lidar optical coordinates and their altitude-dependent occurrence rather than on the auxiliary regional datasets. Accordingly, an “optical regime” denotes a recurrent combination of scattering intensity, particle nonsphericity, and wavelength-dependent backscatter response and does not imply a unique aerosol chemical type. MERRA-2 aerosol diagnostics and HYSPLIT-based transport analyses were introduced only after clustering to evaluate whether the lidar-defined optical states were consistent with plausible regional aerosol and transport contexts. C1–C4 are site-specific labels without one-to-one cross-site correspondence; cross-site statements therefore use the measured optical variables and vertical patterns. The corresponding cluster centroids in the original lidar feature coordinates are summarized in Table 2 to facilitate cross-site comparison. To quantify within-cluster variability, the standard deviation of each clustering feature was calculated separately from the quality-controlled core samples assigned to each regime. These values are reported together with the corresponding K-means centroids in Table 2.

Classification sensitivity to prescribed perturbations of the retrieved optical variables was additionally evaluated using 100 Monte Carlo realizations while retaining the original site-specific scaler and fixed $K = 4$ centroids. For each sample, independent perturba-

tions were drawn from bounded uniform distributions. The β_{532} values were perturbed multiplicatively, whereas $\delta_{p,532}$ and $AE_{\beta,355/532}$ were perturbed additively. Low, moderate, and high scenarios, respectively, used bounds of $\pm 10\%$, $\pm 20\%$, and $\pm 30\%$ for β_{532} ; ± 0.01 , ± 0.02 , and ± 0.03 for $\delta_{p,532}$; and ± 0.10 , ± 0.20 , and ± 0.30 for $AE_{\beta,355/532}$. No cross-variable error correlation was imposed, and a fixed base random seed of 240824 was used for reproducibility. After perturbation, $\log_{10}(\beta_{532})$ was recalculated and the original scaler and fixed K-means model were applied without refitting. Robustness was quantified using label retention and the Adjusted Rand Index (ARI). These prescribed amplitudes represent sensitivity/stress-test scenarios rather than instrument-specific retrieval-uncertainty estimates.

Table 2. Cluster centroids and within-cluster variability in the original lidar feature coordinates at Beijing Nanjiao and Beihai. Values are reported as the K-means centroid $\pm$ one standard deviation of the core samples used for model fitting. The two sites were standardized and clustered independently; therefore, identical cluster labels do not imply one-to-one correspondence between sites.

Site	Regime	$\log_{10}(\beta_{532})$	$\delta_{p,532}$	$AE_{\beta,355/532}$
Beijing Nanjiao	C1	-3.250 ± 0.216	0.285 ± 0.034	2.097 ± 0.499
	C2	-2.551 ± 0.253	0.302 ± 0.028	0.822 ± 0.311
	C3	-2.803 ± 0.260	0.202 ± 0.032	1.308 ± 0.301
	C4	-3.369 ± 0.164	0.183 ± 0.034	2.527 ± 0.487
Beihai	C1	-2.270 ± 0.274	0.039 ± 0.013	1.029 ± 0.427
	C2	-3.365 ± 0.205	0.081 ± 0.031	2.543 ± 0.426
	C3	-2.937 ± 0.199	0.048 ± 0.011	1.848 ± 0.249
	C4	-2.506 ± 0.238	0.095 ± 0.021	1.386 ± 0.250

Note: Before logarithmic transformation, β_{532} is expressed in $\text{km}^{-1} \text{sr}^{-1}$. Both $\delta_{p,532}$ and $AE_{\beta,355/532}$ are dimensionless. The reported standard deviations were calculated from the quality-controlled core samples used to fit the site-specific K-means models.

Altitude-dependent occurrence fractions were evaluated over 0.3–5.0 km using all the valid and successfully classified samples that satisfied the quality-control criteria described above rather than only the high-confidence subset. For each altitude bin h , the occurrence fraction of regime c was defined as $F_{c,h} = N_{c,h} / N_h$, where $N_{c,h}$ is the number of samples assigned to regime c and N_h is the total number of valid successfully classified samples in that altitude bin. Consequently, the four regime fractions sum to 100% in each non-empty altitude bin. Because the amount of valid lidar sampling varied with time and altitude, no equal-month or equal-season weighting was applied. The resulting occurrence fractions therefore represent sampling-weighted statistics of the available quality-controlled observations rather than uniformly sampled seasonal climatologies. Adjacent 15-minute profiles and neighboring 30 m range bins are not statistically independent. K-means is used here descriptively rather than for inference based on independent samples; profile-level resampling preserves within-profile vertical dependence but does not remove temporal persistence, so the occurrence fractions are interpreted as sample-weighted frequencies rather than frequencies of independent aerosol events.

2.4. MERRA-2 Regional Aerosol Context

Hourly MERRA-2 `tavg1_2d_aer_Nx` diagnostics at a horizontal resolution of $0.5^\circ \times 0.625^\circ$ were used to provide regional aerosol context [43–47]. For each lidar site, the nearest MERRA-2 grid cell was extracted and matched hourly to periods containing high-confidence lidar assignments. The selected fields were total, dust, sea salt, BC, OC, and sulfate AODs, together with dust, pollution, carbon, sulfate, and sea salt fractions. For each optical regime, the mean deviation from the overall site mean was expressed as a percentage. These coarse regional fields were not clustering inputs. Because the lidar

measurements represent a narrow local atmospheric column whereas MERRA-2 represents a substantially larger grid cell, agreement between the two datasets is interpreted as regional-scale contextual consistency rather than point-by-point correspondence or independent validation of aerosol composition. To avoid treating repeated lidar time–height samples matched to the same regional field as independent observations, each site-hour contributed one MERRA-2 record associated with the dominant high-confidence lidar regime. Inferential comparisons were then performed on day-by-regime aggregates using the Kruskal–Wallis test with Benjamini–Hochberg false-discovery-rate correction.

2.5. HYSPLIT Backward Trajectory, WPSCF, and WCWT Analyses

HYSPLIT 72 h backward trajectories driven by the NOAA GDAS1 one-degree archive described potential transport pathways and upwind regions [48–52]. High-confidence samples were aggregated by site, regime, and 3 h window and ranked by

$$\text{Score} = \bar{\beta}_{532} \cdot \sqrt{N}, \quad (9)$$

where $\bar{\beta}_{532}$ and N are the window-mean backscatter and valid-sample count. At most three starting times per day and 120 starting times per site–regime combination were retained. Because trajectories were initialized at three heights for each selected starting time, these corresponded to 159, 123, 114, and 159 trajectories for BN-C1–C4, respectively, and 183, 204, 204, and 147 trajectories for Beihai-C1–C4. The score-based selection preferentially samples stronger and better-sampled lidar occurrences and may underrepresent weak-loading or sparsely sampled conditions. Consequently, the trajectories are interpreted as selected transport cases and differences among regimes as transport tendencies of the selected cases rather than unbiased estimates of source-sector frequency or probability.

Trajectories started at 500, 1500, or 3000 m according to sample altitude and were output hourly. WPSCF and WCWT used a common $0.5^\circ \times 0.5^\circ$ grid over 70° – 135° E and 15° – 55° N; cells with fewer than three endpoints were excluded.

For grid cell (i, j) , the potential source contribution function was [53,54]

$$\text{PSCF}_{ij} = \frac{m_{ij}}{n_{ij}}, \quad (10)$$

where m_{ij} and n_{ij} are target-regime and total endpoint counts. The weighted value was

$$\text{WPSCF}_{ij} = W_{ij} \cdot \text{PSCF}_{ij}, \quad (11)$$

with $W_{ij} = 0.05, 0.42, 0.70$, and 1.00 for endpoint counts $\leq$ the mean and 1 – 1.5 , 1.5 – 3 , and > 3 times the mean, respectively.

WCWT incorporated the starting-time β_{532} as an optical-loading weight [54,55]:

$$\text{WCWT}_{ij} = \frac{\sum_l C_l \tau_{ijl}}{\sum_l \tau_{ijl}}, \quad (12)$$

where C_l is the β_{532} weight and τ_{ijl} is residence time. Sensitivity of the trajectory-based source diagnostics was evaluated by varying the grid resolution, minimum endpoint threshold, and retained trajectory number. The influence of the endpoint-density weighting was additionally examined for WPSCF. Spatial correlation with the baseline field and overlap of high-value grid cells were used to assess pattern stability. Trajectories were also interpolated at 3 h intervals from 0 to -72 h and clustered by their longitude–latitude sequences. Fewer than five trajectories were not subdivided; otherwise, the trajectory-cluster number was selected by the silhouette coefficient.

3. Results

3.1. Multi-Parameter Optical States at Beijing Nanjiao

At Beijing Nanjiao, the silhouette coefficients for $K = 2$ –8 were 0.357, 0.331, 0.354, 0.328, 0.311, 0.310, and 0.324, respectively (Figure 2). Although $K = 2$ yielded a marginally higher silhouette coefficient, increasing to $K = 4$ reduced the SSE from 56,190.6 to 31,905.6. More importantly, the four-cluster solution resolved distinct combinations of scattering intensity, particle nonsphericity, and spectral response that were merged into two broader groups at $K = 2$. We therefore retained $K = 4$ as a compromise between statistical compactness, solution stability, and physically interpretable optical resolution rather than selecting the cluster number from the silhouette coefficient alone.

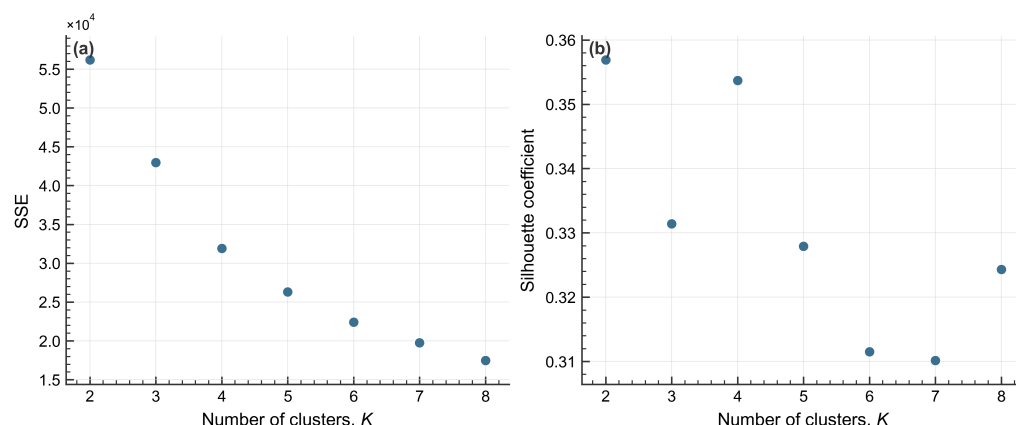


Figure 2. Evaluation of the K-means cluster number at Beijing Nanjiao: (a) within-cluster sum of squares and (b) mean silhouette coefficient for $K = 2$ –8.

The 30-run profile-resampling analysis indicated that the Beijing Nanjiao solution was dominated by a reference-like partition but also admitted an alternative local partition. The median ARI was 0.966, and 21 of the 30 runs (70%) achieved $\text{ARI} \geq 0.95$ relative to the reference solution. The remaining nine runs formed a lower-ARI group, with the minimum ARI reaching 0.728 and correspondingly larger maximum centroid shifts. Across all the runs, the ARI was strongly negatively correlated with maximum centroid shift ($r = -0.998$), confirming that the lower-ARI realizations represented coherent shifts in the partition rather than isolated numerical fluctuations (Figure 3). Importantly, the major optical ordering of the four regimes remained largely preserved across the repeated realizations. The ordering of the regime centroids in $\text{AE}_{\beta_{355/532}}$ and $\log_{10}(\beta_{532})$ was unchanged in all 30 runs. The main variability occurred in the relative particle-depolarization ordering of the two more depolarizing regimes, with C2 having the highest $\delta_{p,532}$ in 21 of 30 runs and C1 in the remaining nine. Thus, the alternative partition mainly affects the separation between neighboring high-depolarization states, whereas the broader contrasts in scattering and spectral response remain stable. The adjacent- K comparison further showed that changing the number of clusters mainly merged or subdivided neighboring optical states. At $K = 3$, the $K = 4$ C1 and C4 states were represented by one broader state, whereas, at $K = 5$, the $K = 4$ C3 state was subdivided. The principal high-depolarization, strong-scattering, and high-/low-spectral-response structures nevertheless remained represented.

The feature-space distributions show four optically distinct regimes (Figure 4). C1 ($n = 8458$ high-confidence samples) combined relatively low β_{532} with moderately high $\delta_{p,532}$ and $\text{AE}_{\beta_{355/532}}$. C2 ($n = 4215$) was the most strongly depolarizing regime, with relatively high β_{532} and low $\text{AE}_{\beta_{355/532}}$. C3 ($n = 11,156$) also showed strong backscatter but lower depolarization than C2. C4 ($n = 9213$) combined high $\text{AE}_{\beta_{355/532}}$ with low β_{532} and low $\delta_{p,532}$.

C2 had high depolarization and a coarse-particle-sensitive optical signature; C3 had strong scattering and lower depolarization; C4 had low loading and a fine-particle-sensitive spectral response; and C1 combined weak loading with greater nonspherical influence than C3 and C4.

The occurrence profiles separated these states vertically (Figure 5). High-depolarization C2 dominated the lowest observed layer and declined rapidly with height, while strong-scattering C3 also contributed mainly in the lower layer. Low-loading C1 increased above the lower layer and dominated much of the middle and upper observed range. C4, distinguished by its high $AE_{\beta,355/532}$ response, also contributed more clearly in the middle and upper profile. The contrasting lower-, middle-, and upper-layer frequencies show that the partition tracks different aerosol-layer optical states rather than only subdividing a large set of near-surface points.

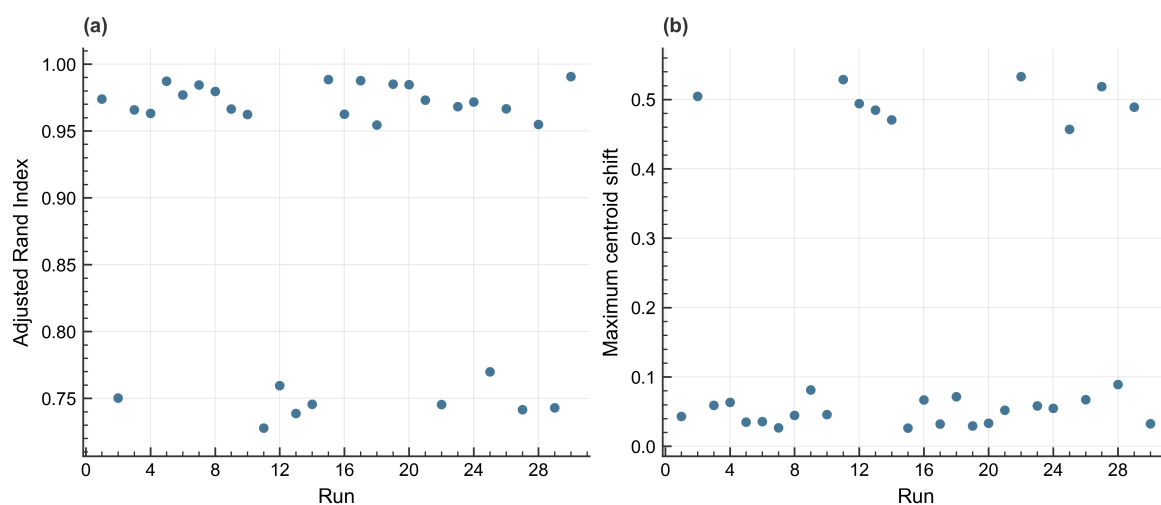


Figure 3. Profile-resampling stability of the four-cluster solution at Beijing Nanjiao across 30 repeated runs based on random selection of 80% of the complete 15-minute profiles: (a) Adjusted Rand Index relative to the reference solution and (b) maximum centroid shift in the reference-standardized feature space. Higher ARI and lower centroid shift indicate greater reproducibility.

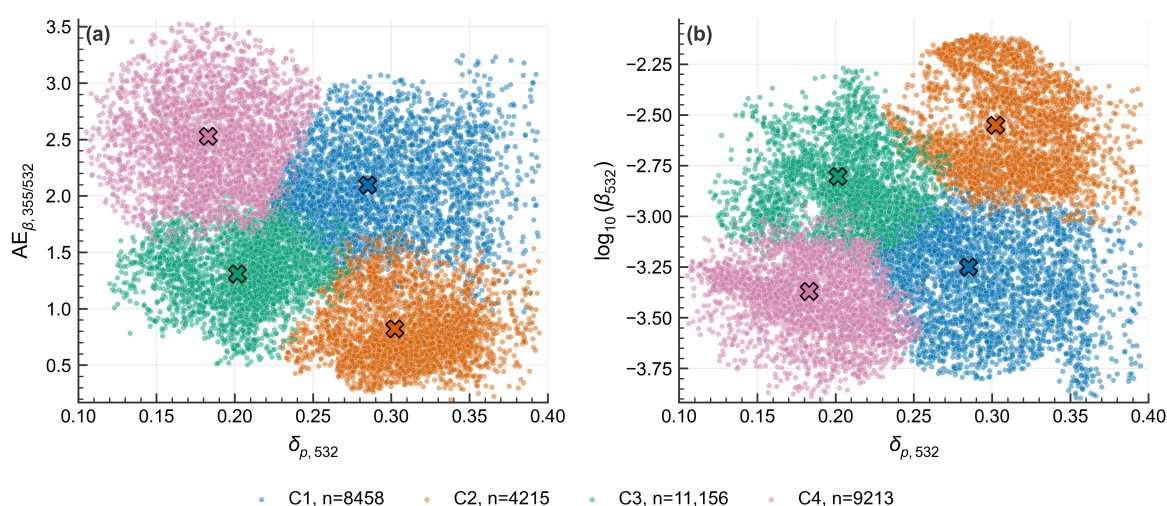


Figure 4. Feature-space distributions of high-confidence samples at Beijing Nanjiao: (a) $AE_{\beta,355/532}$ versus $\delta_{p,532}$ and (b) $\log_{10}(\beta_{532})$ versus $\delta_{p,532}$. Black crosses indicate cluster centroids.

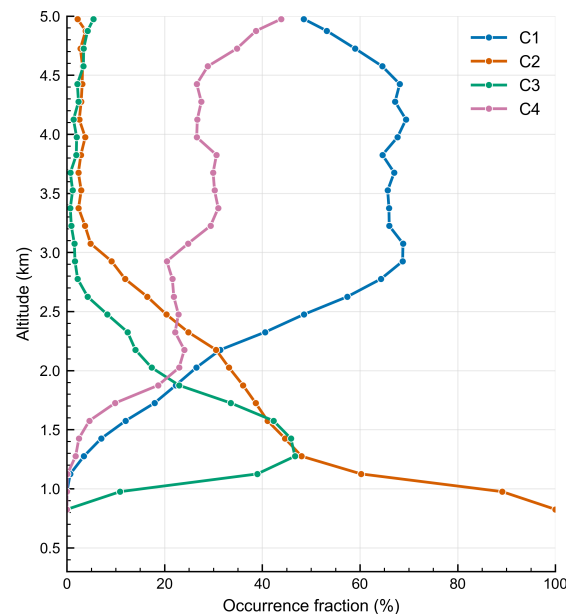


Figure 5. Altitude-dependent occurrence fractions of the four site-specific aerosol optical regimes at Beijing Nanjiao. Fractions were calculated within each altitude bin using all the valid and successfully classified samples included in the occurrence analysis rather than only the high-confidence subset.

3.2. Multi-Parameter Optical States at Beihai

At Beihai, the silhouette coefficients for $K = 2-8$ were 0.363, 0.402, 0.406, 0.346, 0.351, 0.329, and 0.327, respectively, with $K = 4$ yielding the highest value among the evaluated solutions (Figure 6). The SSE simultaneously decreased from 158,538.4 to 87,113.9. Together with the additional separation of physically distinct combinations of backscatter, depolarization, and spectral response, these diagnostics provide stronger statistical support for retaining $K = 4$ at Beihai.

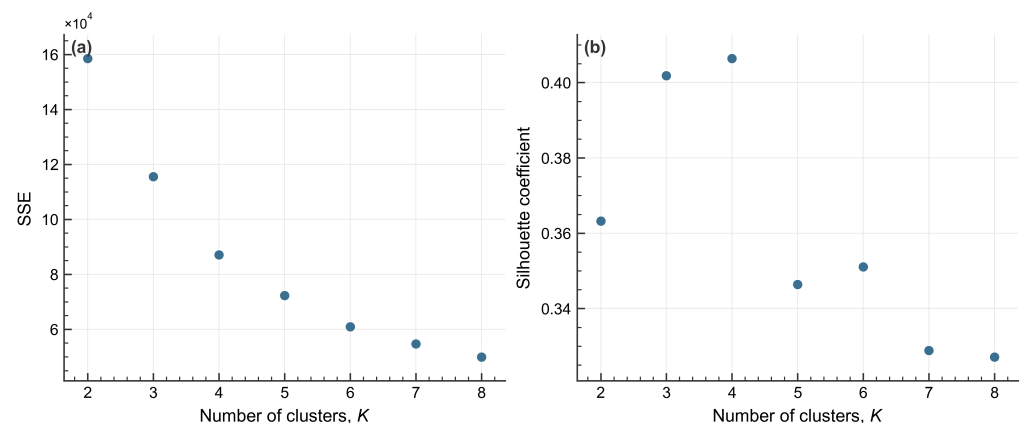


Figure 6. Evaluation of the K-means cluster number at Beihai: (a) within-cluster sum of squares and (b) mean silhouette coefficient for $K = 2-8$.

The Beihai solution showed substantially higher reproducibility across the 30 profile-resampling runs. The median ARI was 0.986, with all 30 runs yielding $\text{ARI} \geq 0.95$ and a minimum ARI of 0.962. The median maximum centroid shift was 0.040 in the reference-standardized feature space. Moreover, the relative centroid ordering of all four regimes in $\text{AE}_{\beta_{355/532}}$, $\log_{10}(\beta_{532})$, and $\delta_{p,532}$ was preserved in all 30 realizations, indicating that both the partition and the major optical characteristics of the Beihai regimes were highly stable (Figure 7). The adjacent- K comparison showed a similar pattern. At $K = 3$, the $K = 4$ C2

and C3 states were represented by one broader optical state, whereas, at $K = 5$, C3 was mainly subdivided. Close counterparts of the other major $K = 4$ optical structures remained identifiable, indicating that changing K primarily merged or subdivided neighboring states rather than producing qualitatively different structures.

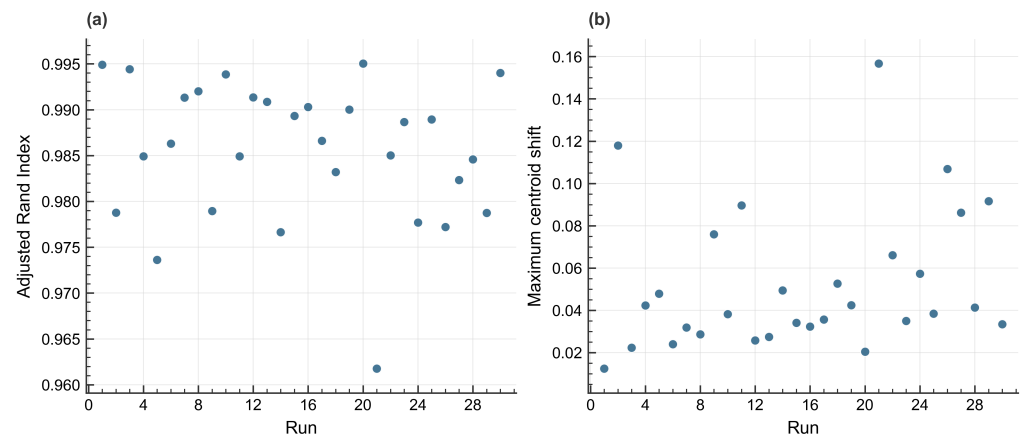


Figure 7. Profile-resampling stability of the four-cluster solution at Beihai across 30 repeated runs based on random selection of 80% of the complete 15-minute profiles: (a) Adjusted Rand Index relative to the reference solution and (b) maximum centroid shift in the reference-standardized feature space. Higher ARI and lower centroid shift indicate greater reproducibility.

The feature-space distributions separated four regimes at Beihai (Figure 8). C1 ($n = 17,520$ high-confidence samples) combined high β_{532} with low $\delta_{p,532}$ and low-to-moderate $AE_{\beta_{355/532}}$. C2 ($n = 18,892$) had the highest $AE_{\beta_{355/532}}$ and relatively low backscatter. C3 ($n = 43,028$) was the largest high-confidence regime and occupied intermediate backscatter and low-depolarization conditions. C4 ($n = 13,864$) showed relatively high backscatter and low $AE_{\beta_{355/532}}$. Its mean depolarization was close to that of C2.

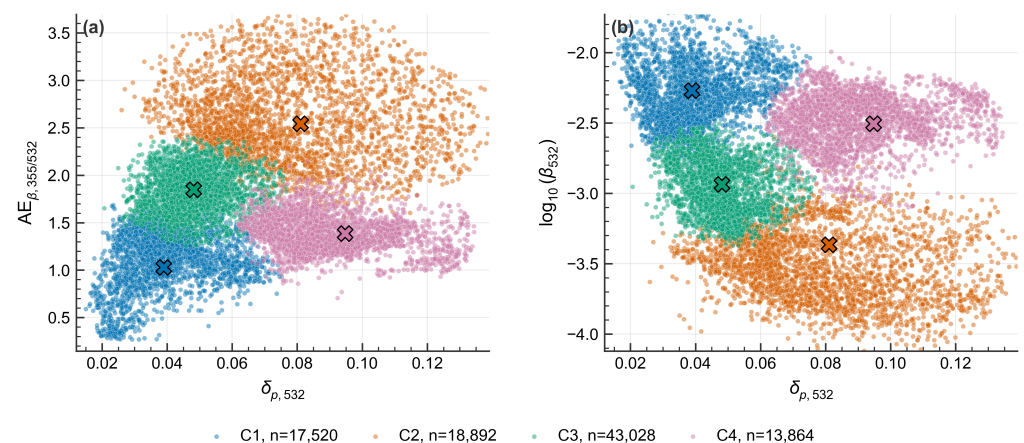


Figure 8. Feature-space distributions of high-confidence samples at Beihai: (a) $AE_{\beta_{355/532}}$ versus $\delta_{p,532}$ and (b) $\log_{10}(\beta_{532})$ versus $\delta_{p,532}$. Black crosses indicate cluster centroids.

The direct optical descriptors distinguish C1 as a low-depolarization strongly scattering regime; C2 as a lower-loading regime with the strongest size-sensitive spectral response; C3 as the site-background low-depolarization regime; and C4 as a relatively strong-scattering low- $AE_{\beta_{355/532}}$ coastal regime. The C2–C4 depolarization difference is small and is not treated as significant without event-level uncertainty analysis.

The Beihai states also changed markedly with altitude (Figure 9). Strong-scattering C1 contributed most in parts of the near-surface and lower layer. C3 dominated much

of the approximately 1.5–3.0 km range and contributed again at higher levels, while the high- $AE_{\beta,355/532}$ C2 state increased through parts of the upper observed layer. C4 remained smaller overall but occurred in portions of the lower and middle layers. Beihai was therefore a predominantly low-depolarization system in which scattering intensity and wavelength response provided the main optical separation. The lower-to-middle-to-upper transitions further show that the clustering retained vertically changing coastal aerosol states.

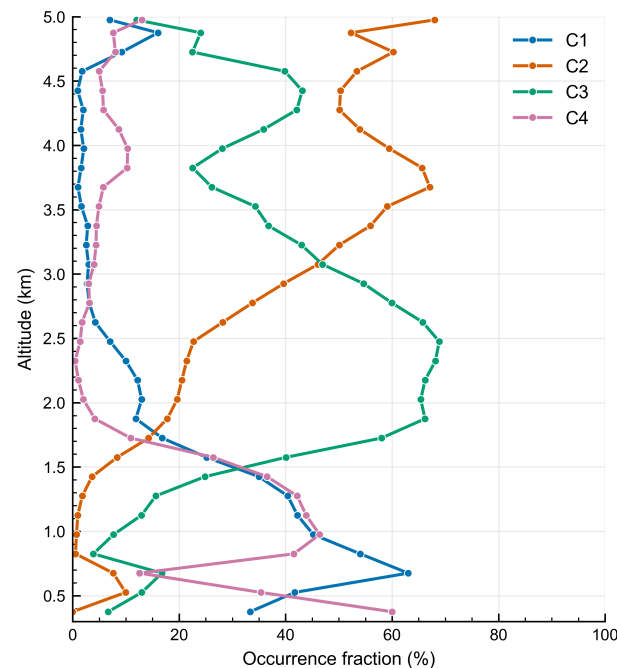


Figure 9. Altitude-dependent occurrence fractions of the four site-specific aerosol optical regimes at Beihai. Fractions were calculated within each altitude bin using all the valid and successfully classified samples included in the occurrence analysis rather than only the high-confidence subset.

3.3. Inland–Coastal Comparison of Optical Regimes

Because clustering was performed independently at the two sites, regimes with the same numerical label are not assumed to be equivalent. As shown by the original-coordinate centroids in Table 2, cross-site comparison is therefore based on the measured optical characteristics and vertical occurrence of each regime rather than on cluster numbering. The site-aware analysis resolved distinctly different aerosol optical-state distributions at the two sites. Beijing Nanjiao contained a pronounced high-depolarization regime (C2, mean $\delta_{p,532} = 0.30$) and a second strong-scattering regime with lower but still appreciable depolarization (C3, mean $\delta_{p,532} = 0.20$). In contrast, most high-confidence samples at Beihai remained below $\delta_{p,532} = 0.15$, and its largest regime (C3) represented the low-depolarization coastal background. These differences demonstrate a stronger nonspherical and coarse-particle influence at the inland site and a greater prevalence of optically spherical or near-spherical aerosol states at the coastal site.

The optical contrast also depended on height. At Beijing Nanjiao, high-depolarization C2 was concentrated in the lowest observed layer, whereas lower-loading C1 and C4 contributed more strongly aloft. At Beihai, low-depolarization C3 dominated much of the middle layer and C2 increased toward parts of the upper layer. Thus, the sites differed in the coupled vertical variation in scattering intensity, particle shape, and size-sensitive spectral response. These range-resolved transitions would be obscured in surface particulate records or column-integrated AOD. The lidar-ratio sensitivity test showed a site-dependent classification response. At Beijing Nanjiao, $LR_{532} = 30$ and 70 sr produced median β_{532} differences of +3.45% and

−8.75%, with ARI values of 0.440 and 0.571 and matched-label retention of 69.3% and 81.3%, respectively. At Beihai, the corresponding differences were +4.21% and −11.00%, with ARI values of 0.859 and 0.675 and retention of 94.2% and 87.0%. The selected cases therefore indicate stronger sensitivity to the lidar-ratio assumption at Beijing Nanjiao. This lidar-ratio test and the Monte Carlo experiment represent different perturbation structures and therefore should not be interpreted as providing the same site-to-site sensitivity ranking. The lidar-ratio change also propagated to the other clustering variables. For $LR_{532} = 30/70$ sr, the median changes in $\delta_{p,532}$ were −0.0062/+0.0176 at Beijing Nanjiao and −0.0009/+0.0028 at Beihai, while the median changes in $AE_{\beta,355/532}$ were +0.479/+0.236 and −0.102/+0.295, respectively. The Monte Carlo sensitivity test further showed that the cluster assignments were generally robust, although Beihai was more sensitive than Beijing Nanjiao. For the low, moderate, and high scenarios, the median label retention was 97.54%, 95.01%, and 92.45% at Beijing Nanjiao and 95.94%, 91.34%, and 85.96% at Beihai, with corresponding ARI values of 0.934, 0.870, and 0.807 and 0.891, 0.779, and 0.659, respectively.

3.4. Regional Aerosol Context from MERRA-2

MERRA-2 placed the lidar-derived states in the regional aerosol context. Figures 10 and 11 show descriptive percentage deviations from each overall site mean.

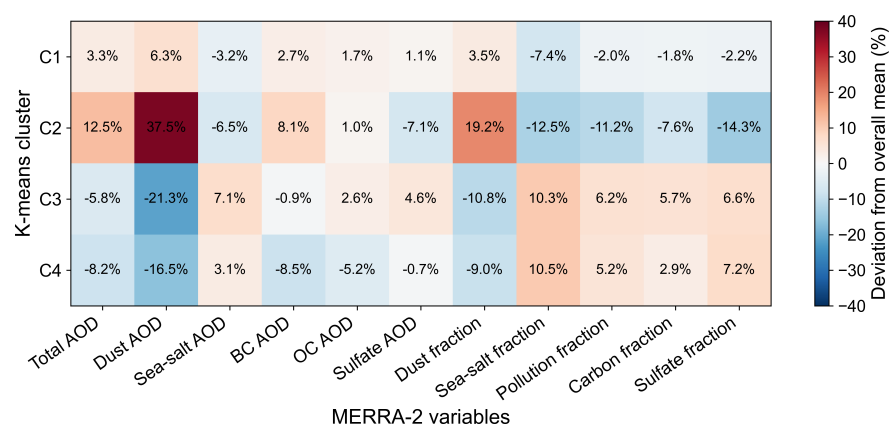


Figure 10. Percentage deviations of MERRA-2 aerosol components and derived fractions from the overall Beijing Nanjiao site means for the four lidar-derived optical regimes.

At Beijing Nanjiao, C2 had the largest dust-related anomalies. The total, dust, sea salt, BC, OC, and sulfate AODs deviated by +12.5%, +37.5%, −6.5%, +8.1%, +1.0%, and −7.1%, respectively, and the dust fraction was 19.2% above the site mean. However, after day-level aggregation, the overall inter-regime differences in total AOD, dust AOD, and dust fraction were not significant after FDR correction ($p_{\text{FDR}} = 0.258, 0.151$, and 0.151 , respectively). These differences are therefore treated as descriptive regional anomalies rather than statistically confirmed compositional separation among the lidar regimes. The larger descriptive dust-related anomalies for C2 are consistent with a possible dust and coarse-particle contribution to its directly observed high-depolarization, low- $AE_{\beta,355/532}$ optical signature.

At Beihai, the total, sea salt, OC, and sulfate AODs for C1 were 33.0%, 15.5%, 24.5%, and 41.9% above the site means, respectively. The sea salt fraction was 16.7% below the site mean despite the higher absolute sea salt AOD, indicating that the total aerosol column was more strongly enhanced rather than being more sea salt-dominated. At the day-by-regime level, the overall inter-regime differences remained significant for the total and sulfate AODs (both $p_{\text{FDR}} = 0.0011$), whereas the sea salt AOD and sea salt fraction were not significant ($p_{\text{FDR}} = 0.278$ and 0.140 , respectively). Together with the strong-scattering low-depolarization

lidar signature, these regional differences are consistent with an enhanced mixed coastal–regional aerosol background for C1. Beihai C4 coincided with positive dust, carbon, and sea salt fraction anomalies and represented a comparatively coarse-particle-sensitive coastal optical state.

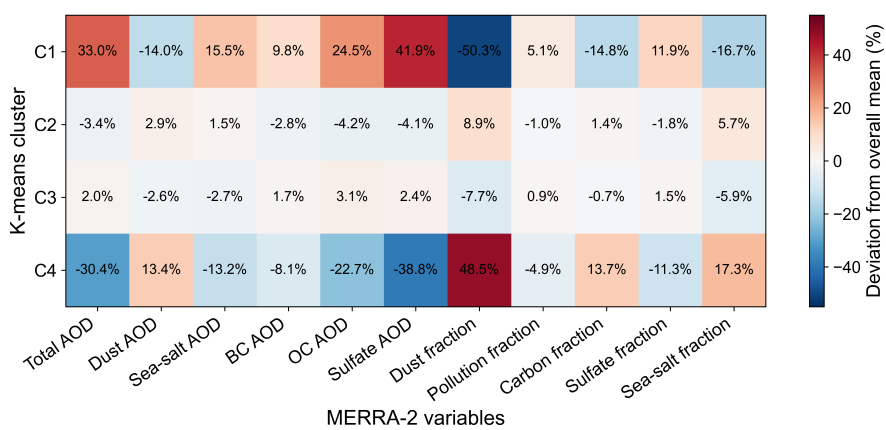


Figure 11. Percentage deviations of MERRA-2 aerosol components and derived fractions from the overall Beihai site means for the four lidar-derived optical regimes.

3.5. Transport Pathways and Potential Upwind Regions

The 72 h trajectories, WPSCF, and β_{532} -weighted WCWT describe transport during selected stronger-signal occurrences (Figures 12 and 13). Beijing Nanjiao was associated mainly with northwestern, northern, and regional North China pathways. Beihai showed pathways from the South China coast, Guangxi–Guangdong, the Beibu Gulf, the Indochina Peninsula, and longer-range western or southwestern sectors.

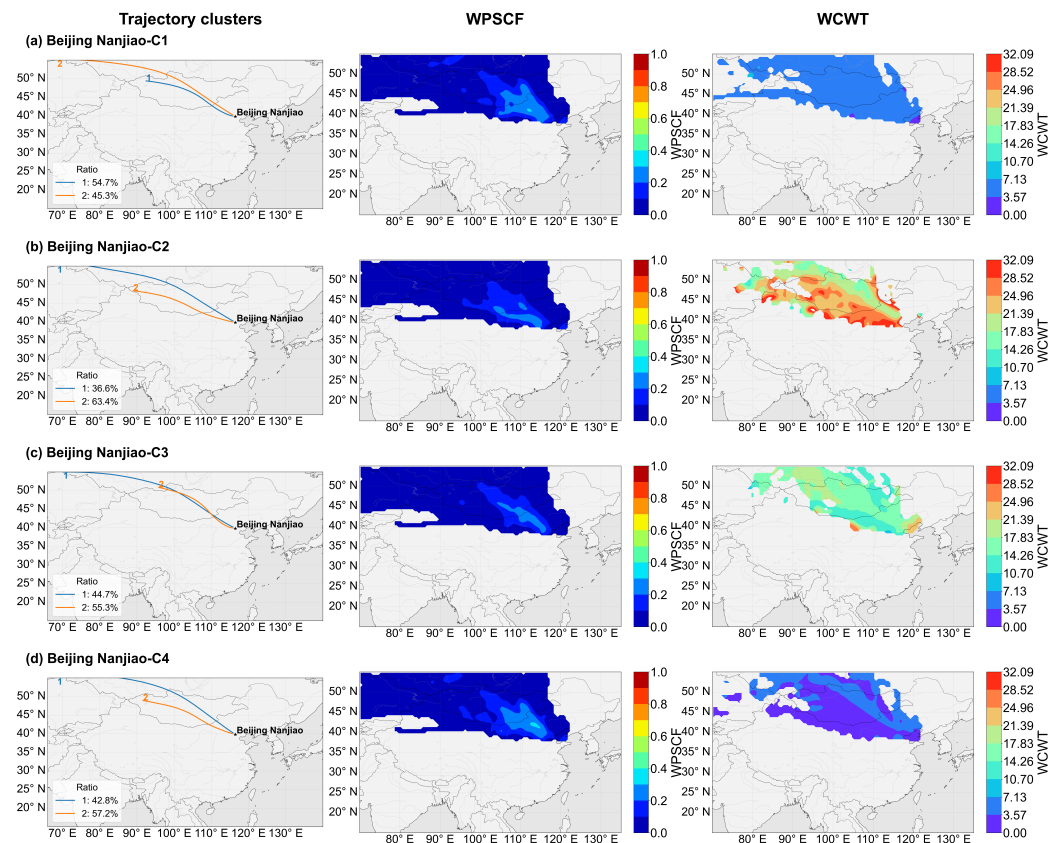


Figure 12. Trajectory clusters, WPSCF, and β_{532} -weighted WCWT for the four site-specific optical regimes at Beijing Nanjiao. Rows (a–d) correspond to C1–C4. The columns show major trajectory clusters, WPSCF, and β_{532} -weighted WCWT fields. WPSCF and WCWT indicate potential transport and upwind regions.

At Beijing Nanjiao, the WPSCF and WCWT high-value regions were concentrated northwest of the site and over northern North China. Their agreement with the high-depolarization, low- $AE_{\beta_{355}/532}$ C2 signature and positive MERRA-2 dust context is consistent with a possible regional dust and coarse-particle transport influence within the mixed urban–regional background. The C3 pathways were compatible with regional transport.

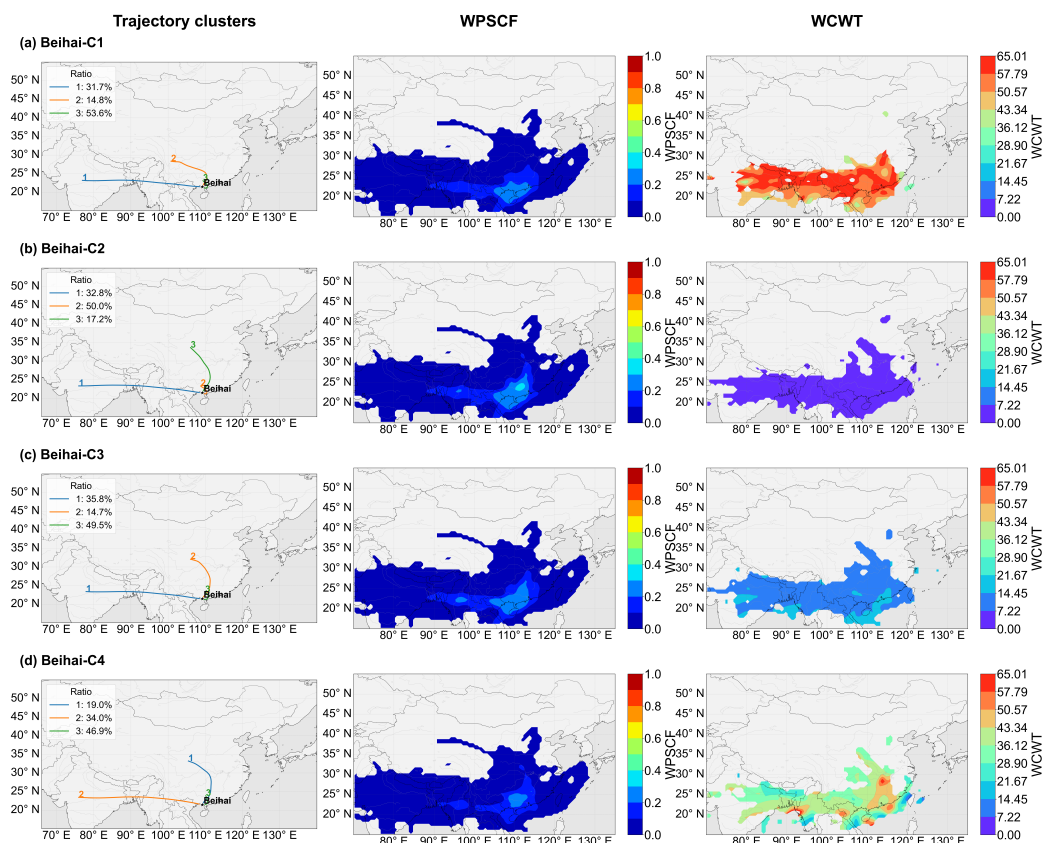


Figure 13. Trajectory clusters, WPSCF, and β_{532} -weighted WCWT for the four site-specific optical regimes at Beihai. Rows (a–d) correspond to C1–C4. The columns show major trajectory clusters, WPSCF, and β_{532} -weighted WCWT fields. WPSCF and WCWT indicate potential transport and upwind regions.

The Beihai potential contributing regions followed the South China coast, Guangxi–Guangdong, the Beibu Gulf, and the Indochina Peninsula. Together with the low-depolarization lidar signature and the regional aerosol context, particularly the enhanced total and sulfate AODs, these pathways are consistent with combined marine and terrestrial regional transport influence on the coastal strong-scattering state.

Because trajectory initialization was conditioned on selected high-confidence and relatively strong-signal episodes, the WPSCF and WCWT patterns are interpreted only as potential source and transport characteristics associated with the sampled regimes. They should not be interpreted as unbiased climatological source frequency distributions. The sensitivity tests showed that the broad WPSCF and WCWT spatial patterns were generally preserved under changes in grid resolution, minimum endpoint threshold, and trajectory number. WPSCF showed greater sensitivity to removal of the endpoint-density weighting, whereas exact WCWT hotspot locations were more sensitive to changes in grid resolution. Accordingly, the transport analysis is interpreted primarily in terms of broad potential upwind regions rather than exact hotspot boundaries.

Table 3 links the direct lidar and vertical characteristics of each regime to regional aerosol anomalies and trajectory diagnostics. The auxiliary evidence provides physical context for the contrasting inland and coastal optical states without acting as a clustering input or chemical validation.

Table 3. Integrated evidence synthesis for the site-specific aerosol optical regimes. Direct lidar characteristics and vertical occurrence are combined with MERRA-2 aerosol anomalies and trajectory diagnostics to support physically traceable environmental interpretations.

Site and Regime	Direct Lidar and Vertical Characteristics	Converging Regional and Transport Evidence	Integrated Interpretation and Scientific Significance
Beijing Nanjiao C1	Relatively high $\delta_{p,532}$, low β_{532} , and moderately high $AE_{\beta,355/532}$; occurrence increased through much of the middle and upper observed layers.	Small MERRA-2 aerosol anomalies; northern and regional transport pathways.	Weak-loading optical regime with nonspherical-particle influence. Its enhanced occurrence aloft indicates that nonspherical aerosol states occurred above the lower layer even under relatively low aerosol loading.
Beijing Nanjiao C2	High $\delta_{p,532}$, high β_{532} , and low $AE_{\beta,355/532}$; dominant in the lowest observed layer and rapidly decreasing with altitude.	Dust AOD +37.5% and dust fraction +19.2%; northwestern and northern transport pathways.	High-depolarization coarse-particle-sensitive optical regime. The combined evidence is consistent with possible regional dust and coarse-particle influence during the sampled C2 occurrences.
Beijing Nanjiao C3	High β_{532} , moderate-to-low $\delta_{p,532}$, and moderate $AE_{\beta,355/532}$; concentrated mainly in the lower observed layer.	Negative dust-related anomalies and regional transport pathways.	Strong-scattering lower-depolarization optical regime occurring under regional-transport conditions. The available regional evidence is consistent with anthropogenic or mixed regional influence but does not uniquely determine aerosol composition.
Beijing Nanjiao C4	High $AE_{\beta,355/532}$, low β_{532} , and low $\delta_{p,532}$; occurrence increased in parts of the middle and upper observed layers.	Low total AOD and negative dust-related anomalies.	Low-loading fine-particle-sensitive elevated optical regime. Its vertical distribution reveals an aerosol state that would be difficult to distinguish using surface measurements or column-integrated AOD alone.
Beihai C1	High β_{532} , low $\delta_{p,532}$, and low-to-moderate $AE_{\beta,355/532}$; important in the near-surface and lower observed layers.	Total AOD +33.0%, sea salt AOD +15.5%, OC AOD +24.5%, and sulfate AOD +41.9%; southern coastal, Beibu Gulf, and regional transport pathways.	Strong-scattering low-depolarization coastal optical regime. The MERRA-2 and trajectory context is consistent with combined marine and terrestrial regional influence but does not permit a unique compositional attribution.
Beihai C2	Highest $AE_{\beta,355/532}$, relatively low β_{532} , and moderate $\delta_{p,532}$; occurrence increased through parts of the upper observed layer.	Most MERRA-2 components remained close to their overall site means.	Lower-loading fine-particle-sensitive upper-layer optical regime. It represents a vertically elevated coastal state distinct from the strongly scattering lower-layer C1 regime.
Beihai C3	Moderate β_{532} , low $\delta_{p,532}$, and moderate $AE_{\beta,355/532}$; dominant through much of the middle observed layer.	Aerosol-component anomalies remained close to the site means; coastal and regional transport pathways.	Site-background low-depolarization optical regime. Its large sample fraction and middle-layer dominance define the prevailing coastal aerosol background during the sampled clear-sky period.
Beihai C4	Relatively strong β_{532} , low $AE_{\beta,355/532}$, and comparatively elevated $\delta_{p,532}$ within the Beihai regime set; present mainly in parts of the lower and middle observed layers.	Dust AOD +13.4%, dust fraction +48.5%, carbon fraction +13.7%, and sea salt fraction +17.3%; mixed coastal and terrestrial transport pathways.	Comparatively coarse-particle-sensitive coastal optical regime. The accompanying regional anomalies and mixed transport pathways provide contextual evidence for possible episodic coarse-particle influence. It reveals an intermittent coarse-particle contrast within the otherwise predominantly low-depolarization coastal environment.

Note: The optical regimes are defined by the lidar-derived optical characteristics and their altitude-dependent occurrence. MERRA-2 aerosol anomalies and trajectory diagnostics are used only as contextual evidence for possible environmental influences and are not treated as chemical validation or unique aerosol-type identification. The specific contribution of humidity-dependent particle growth to the strong scattering of Beihai C1 requires evaluation using collocated relative-humidity profiles.

4. Discussion

The three lidar variables provide complementary optical constraints. $\log_{10}(\beta_{532})$ organizes the wide range of scattering intensity, $\delta_{p,532}$ adds sensitivity to particle nonsphericity,

and $AE_{\beta,355/532}$ captures the wavelength dependence of backscatter and its size-sensitive response. Similar scattering intensity can therefore be separated when polarization or spectral response differs. Conversely, low-depolarization states that appear to be similar in shape space can be distinguished by their loading and wavelength dependence. K-means is useful here because it organizes these jointly measured dimensions, but the sensing capability and physical interpretation arise from the Raman–Mie polarization lidar observations rather than from the algorithm itself. The physical distinctiveness of the regimes is therefore assessed from their different positions in this three-dimensional optical space together with their coherent altitude-dependent occurrence patterns. The clustering stability tests address numerical reproducibility, whereas the separation in directly observed optical characteristics and vertical structure provides the primary physical basis for distinguishing the regimes. Numerical reproducibility and physical distinctiveness were therefore treated as complementary but different properties. The former was evaluated using resampling-based agreement, whereas the latter was assessed from separation in the three directly retrieved lidar coordinates together with coherent altitude-dependent occurrence patterns.

The altitude-dependent occurrence profiles add a critical range-resolved constraint. At Beijing Nanjiao, the high-depolarization coarse-particle-sensitive C2 regime was concentrated in the lowest observed layer, while lower-loading C1 and C4 contributed more strongly in the middle and upper layers. Beihai showed a different vertical sequence among predominantly low-depolarization states: strong-scattering C1 was important in the lower layer, C3 dominated much of the middle range, and the high- $AE_{\beta,355/532}$ C2 state increased in parts of the upper layer. These height-dependent transitions indicate different aerosol-layer optical states rather than a purely mathematical division of pooled time–height points. This is a direct advantage over surface particulate measurements, which emphasize near-ground conditions, and column-integrated AOD, which does not retain the altitude of the contributing layers.

The regional and transport analyses provide auxiliary physical explanations for these optical structures [49,56]. At Beijing Nanjiao, the high-depolarization C2 signature occurred with dust AOD and dust-fraction enhancements and northwestern and northern pathways, providing regional context that is consistent with a possible dust and coarse-particle influence in the inland lower troposphere. At Beihai, enhanced total, sea salt, OC, and sulfate AODs and pathways from the South China coast, the Beibu Gulf, and terrestrial regional sectors were concordant with the strong-scattering low-depolarization C1 state. The lower sea salt fraction despite the higher absolute sea salt AOD indicates a broadly enhanced mixed column rather than increased sea salt dominance alone. MERRA-2 and the trajectory diagnostics are therefore contextual evidence in the chain from lidar state to regional background and potential transport; they are not clustering inputs, point-level composition measurements, or direct chemical validation.

The enhanced backscatter of Beihai C1 should not be attributed exclusively to increased aerosol loading. In the humid coastal environment, hygroscopic growth can increase particle size and optical scattering efficiency even without a proportional increase in dry aerosol mass. Because collocated vertically resolved relative-humidity measurements were not available, the relative contributions of aerosol loading and humidity-driven growth cannot be quantitatively separated here. The enhanced backscatter of C1 is therefore interpreted as reflecting a combination of aerosol abundance, mixing state, and potentially hygroscopic growth.

The high-depolarization Beijing Nanjiao state is consistent with established polarization-lidar observations of nonspherical mineral dust and transported dust layers [16,20,21,38]. The predominantly low-depolarization Beihai states are compatible with reported marine-aerosol and marine-mixture behavior [23,30,31]. The present contribution is the application

of one multi-parameter vertically resolved optical identification framework in two contrasting environments while preserving site-specific regime labels and comparing them through their measured optical coordinates and height dependence.

The framework is suitable for organizing vertically resolved lidar records when direct chemical observations are unavailable provided that the outputs are interpreted as aerosol optical regimes rather than chemically defined aerosol types. The present analysis is limited to quality-controlled clear-sky observations at Beijing Nanjiao and Beihai during the study period. The 50 and 46 valid days correspond to 938.00 and 805.75 h of effective observations, respectively, and represent intermittent subsets of the June 2025–March 2026 period rather than a continuous seasonal climatology. Because no equal-month or equal-season weighting was applied, the reported occurrence fractions describe the available quality-controlled sampling and may depend on the temporal distribution of valid observations. The trajectory analysis has a related sampling limitation because the score-based selection preferentially retains stronger and better-sampled lidar occurrences; the resulting transport patterns therefore represent selected transport contexts rather than unbiased climatological source frequencies. Daytime backscatter also retains uncertainty associated with the assumed lidar ratio, and the present perturbation experiment evaluates classification sensitivity rather than full height-dependent retrieval-uncertainty propagation. In addition, humidity-dependent growth may contribute to strong coastal scattering. Longer continuous observations together with collocated relative-humidity and composition profiles would help to separate humidity, mixing, transport, and seasonal effects and evaluate transferability to additional lidar environments.

5. Conclusions

Ground-based multi-wavelength Raman–Mie polarization lidar resolved aerosol optical states from complementary measurements of scattering intensity, nonsphericity, and size-sensitive spectral response. Their height dependence separated the low-level high-depolarization state at Beijing Nanjiao from lower-loading elevated states, while Beihai states differed mainly in scattering and wavelength response.

The regional aerosol anomalies and transport pathways at the two sites provided contextual evidence that is consistent with the lidar-based physical interpretations without serving as clustering inputs, chemical validation, or unique aerosol-type identification. The results show that multi-parameter active optical sensing retains aerosol-layer information that is unavailable from surface or column-integrated observations. The identified regimes characterize the quality-controlled predominantly clear-sky observations sampled at the two sites during the present campaign and should not be interpreted as a complete seasonal or climatological classification of aerosol conditions. Cloud screening, instrument availability, signal-to-noise requirements, and the finite observational period may influence the sampling distribution. The reported regime frequencies are therefore sampling-weighted by the available quality-controlled observations rather than uniformly sampled seasonal frequencies. Longer continuous records with humidity and composition profiles are needed to evaluate seasonal representativeness and transferability.

Author Contributions: Conceptualization, Z.Z. and Z.L.; methodology, Z.Z. and Z.L.; validation, Z.L. and Z.B.; formal analysis, Z.Z. and Z.L.; investigation, Y.D.; resources, Z.B. and Y.D.; data curation, Y.D.; writing—original draft preparation, Z.Z.; writing—review and editing, Z.L., Z.B. and Y.D.; visualization, Z.Z.; supervision, Z.L.; project administration, Z.L. and Y.D.; funding acquisition, Z.B. and Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2023YFC3007800 and 2023YFC3007802).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The MERRA-2 aerosol diagnostics used in this study were obtained from the MERRA-2 tavg1_2d_aer_Nx product provided by the NASA Global Modeling and Assimilation Office (GMAO) and distributed by the Goddard Earth Sciences Data and Information Services Center (GES DISC), available at <https://doi.org/10.5067/KLICLTZ8EM9D> (accessed on 11 June 2026). HYSPLIT is provided by the NOAA Air Resources Laboratory (ARL), and the GDAS1 one-degree meteorological fields used for backward-trajectory analysis are publicly available from the NOAA ARL archive at <https://www.ready.noaa.gov/gdas1.php> (accessed on 24 June 2026).

Acknowledgments: Special thanks to the Meteorological Observation Center of the China Meteorological Administration for providing the lidar and ground-station data. We also thank the NASA Global Modeling and Assimilation Office (GMAO) for providing the MERRA-2 reanalysis data and the NOAA Air Resources Laboratory (ARL) for providing the HYSPLIT model and the meteorological data used for backward-trajectory analysis.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. IPCC. *Climate Change 2021: The Physical Science Basis*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I.; et al., Eds.; Cambridge University Press: Cambridge, UK, 2021. <https://doi.org/10.1017/9781009157896>.
2. Charlson, R.J.; Schwartz, S.E.; Hales, J.M.; Cess, R.D.; Coakley, J.A., Jr.; Hansen, J.E.; Hofmann, D.J. Climate forcing by anthropogenic aerosols. *Science* **1992**, *255*, 423–430. <https://doi.org/10.1126/science.255.5043.423>.
3. Haywood, J.; Boucher, O. Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review. *Rev. Geophys.* **2000**, *38*, 513–543. <https://doi.org/10.1029/1999RG000078>.
4. Pöschl, U. Atmospheric aerosols: Composition, transformation, climate and health effects. *Angew. Chem. Int. Ed.* **2005**, *44*, 7520–7540. <https://doi.org/10.1002/anie.200501122>.
5. Seinfeld, J.H.; Pandis, S.N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2016. Available online: <https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Change%2C+3rd+Edition-p-9781118947401> (accessed on 25 August 2026).
6. Boucher, O. *Atmospheric Aerosols: Properties and Climate Impacts*; Springer: Dordrecht, The Netherlands, 2015. <https://doi.org/10.1007/978-94-017-9649-1>.
7. Ramanathan, V.; Crutzen, P.J.; Lelieveld, J.; Mitra, A.P.; Althausen, D.; Anderson, J.; Andreae, M.O.; Cantrell, W.; Cass, G.R.; Chung, C.E.; et al. Indian Ocean Experiment: An integrated analysis of the climate forcing and effects of the great Indo-Asian haze. *J. Geophys. Res.* **2001**, *106*, 28371–28398. <https://doi.org/10.1029/2001JD900133>.
8. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanré, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5).
9. Weitkamp, C. (Ed.) *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*; Springer: New York, NY, USA, 2005. <https://doi.org/10.1007/b106786>.
10. Kovalev, V.A.; Eichinger, W.E. *Elastic Lidar: Theory, Practice, and Analysis Methods*; Wiley: Hoboken, NJ, USA, 2004; ISBN 978-0-471-20171-5. <https://doi.org/10.1002/0471643173>.
11. Measures, R.M. *Laser Remote Sensing: Fundamentals and Applications*; Wiley-Interscience: New York, NY, USA, 1984; ISBN 978-0-471-08193-7.

12. Bösenberg, J.; Matthias, V.; Amodeo, A.; Amoiridis, V.; Ansmann, A.; Baldasano, J.M.; Balin, I.; Balis, D.; Böckmann, C.; Boselli, A.; et al. EARLINET: A European Aerosol Research Lidar Network to Establish an Aerosol Climatology. In *MPI Report*; Max-Planck-Institut für Meteorologie: Hamburg, Germany, 2003; No. 348. <https://doi.org/10.17617/2.995285>.
13. Pappalardo, G.; Amodeo, A.; Apituley, A.; Comeron, A.; Freudenthaler, V.; Linné, H.; Ansmann, A.; Bösenberg, J.; D'Amico, G.; Mattis, I.; et al. EARLINET: Towards an advanced sustainable European aerosol lidar network. *Atmos. Meas. Tech.* **2014**, *7*, 2389–2409. <https://doi.org/10.5194/amt-7-2389-2014>.
14. Dubovik, O.; Holben, B.; Eck, T.F.; Smirnov, A.; Kaufman, Y.J.; King, M.D.; Tanré, D.; Slutsker, I. Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci.* **2002**, *59*, 590–608. [https://doi.org/10.1175/1520-0469\(2002\)059%3C0590:VOAAOP%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059%3C0590:VOAAOP%3E2.0.CO;2).
15. Schuster, G.L.; Dubovik, O.; Holben, B.N. Angstrom exponent and bimodal aerosol size distributions. *J. Geophys. Res.* **2006**, *111*, D07207. <https://doi.org/10.1029/2005JD006328>.
16. Freudenthaler, V.; Esselborn, M.; Wiegner, M.; Heese, B.; Tesche, M.; Ansmann, A.; Müller, D.; Althausen, D.; Wirth, M.; Fix, A.; et al. Depolarization ratio profiling at several wavelengths in pure Saharan dust during SAMUM 2006. *Tellus B* **2009**, *61*, 165–179. <https://doi.org/10.1111/j.1600-0889.2008.00396.x>.
17. Sassen, K. The polarization lidar technique for cloud research: A review and current assessment. *Bull. Am. Meteorol. Soc.* **1991**, *72*, 1848–1866. [https://doi.org/10.1175/1520-0477\(1991\)072%3C1848:TPLTFC%3E2.0.CO;2](https://doi.org/10.1175/1520-0477(1991)072%3C1848:TPLTFC%3E2.0.CO;2).
18. Winker, D.M.; Vaughan, M.A.; Omar, A.; Hu, Y.; Powell, K.A.; Liu, Z.; Hunt, W.H.; Young, S.A. Overview of the CALIPSO mission and CALIOP data processing algorithms. *J. Atmos. Ocean. Technol.* **2009**, *26*, 2310–2323. <https://doi.org/10.1175/2009JTECHA1281.1>.
19. Omar, A.H.; Winker, D.M.; Vaughan, M.A.; Liu, Z.; Hu, Y.; Trepte, C.R.; Ferrare, R.A.; Lee, K.-P.; Hostetler, C.A.; Kittaka, C.; et al. The CALIPSO automated aerosol classification and lidar ratio selection algorithm. *J. Atmos. Ocean. Technol.* **2009**, *26*, 1994–2014. <https://doi.org/10.1175/2009JTECHA1231.1>.
20. Liu, Z.; Omar, A.; Vaughan, M.; Hair, J.; Kittaka, C.; Hu, Y.; Powell, K.; Trepte, C.; Winker, D.; Hostetler, C.; et al. CALIPSO lidar observations of the optical properties of Saharan dust: A case study of long-range transport. *J. Geophys. Res.* **2008**, *113*, D07207. <https://doi.org/10.1029/2007JD008878>.
21. Tesche, M.; Ansmann, A.; Müller, D.; Althausen, D.; Engelmann, R.; Freudenthaler, V.; Groß, S. Vertically resolved separation of dust and smoke over Cape Verde using multiwavelength Raman and polarization lidars during Saharan Mineral Dust Experiment 2008. *J. Geophys. Res. Atmos.* **2009**, *114*, D13202. <https://doi.org/10.1029/2009JD011862>.
22. Müller, D.; Ansmann, A.; Mattis, I.; Tesche, M.; Wandinger, U.; Althausen, D.; Pisani, G. Aerosol-type-dependent lidar ratios observed with Raman lidar. *J. Geophys. Res.* **2007**, *112*, D16202. <https://doi.org/10.1029/2006JD008292>.
23. Groß, S.; Tesche, M.; Freudenthaler, V.; Toledano, C.; Wiegner, M.; Ansmann, A.; Althausen, D.; Seefeldner, M. Characterization of Saharan dust, marine aerosols and mixtures of biomass-burning aerosols and dust by means of multi-wavelength depolarization and Raman lidar measurements during SAMUM 2. *Tellus B* **2011**, *63*, 706–724. <https://doi.org/10.1111/j.1600-0889.2011.00556.x>.
24. Burton, S.P.; Ferrare, R.A.; Hostetler, C.A.; Hair, J.W.; Rogers, R.R.; Obland, M.D.; Butler, C.F.; Cook, A.L.; Harper, D.B.; Froyd, K.D. Aerosol classification using airborne High Spectral Resolution Lidar measurements—Methodology and examples. *Atmos. Meas. Tech.* **2012**, *5*, 73–98. <https://doi.org/10.5194/amt-5-73-2012>.
25. Papagiannopoulos, N.; Mona, L.; Amodeo, A.; D'Amico, G.; Gumà Claramunt, P.; Pappalardo, G.; Alados-Arboledas, L.; Guerrero-Rascado, J.L.; Amiridis, V.; Kokkalis, P.; et al. An automatic observation-based aerosol typing method for EARLINET. *Atmos. Chem. Phys.* **2018**, *18*, 15879–15901. <https://doi.org/10.5194/acp-18-15879-2018>.
26. Huang, J.; Minnis, P.; Chen, B.; Huang, Z.; Liu, Z.; Zhao, Q.; Yi, Y.; Ayers, J.K. Long-range transport and vertical structure of Asian dust from CALIPSO and surface measurements during PACDEX. *J. Geophys. Res.* **2008**, *113*, D23212. <https://doi.org/10.1029/2008JD010620>.
27. Guo, S.; Hu, M.; Zamora, M.L.; Peng, J.; Shang, D.; Zheng, J.; Du, Z.; Wu, Z.; Shao, M.; Zeng, L.; et al. Elucidating severe urban haze formation in China. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17373–17378. <https://doi.org/10.1073/pnas.1419604111>.
28. Li, Z.; Guo, J.; Ding, A.; Liao, H.; Liu, J.; Sun, Y.; Wang, T.; Xue, H.; Zhang, H.; Zhu, B. Aerosol and boundary-layer interactions and impact on air quality. *Natl. Sci. Rev.* **2017**, *4*, 810–833. <https://doi.org/10.1093/nsr/nwx117>.
29. Uno, I.; Eguchi, K.; Yumimoto, K.; Takemura, T.; Shimizu, A.; Uematsu, M.; Liu, Z.; Wang, Z.; Hara, Y.; Sugimoto, N. Asian dust transported one full circuit around the globe. *Nat. Geosci.* **2009**, *2*, 557–560. <https://doi.org/10.1038/ngeo583>.
30. O'Dowd, C.D.; Facchini, M.C.; Cavalli, F.; Ceburnis, D.; Mircea, M.; Decesari, S.; Fuzzi, S.; Yoon, Y.J.; Putaud, J.-P. Biogenically driven organic contribution to marine aerosol. *Nature* **2004**, *431*, 676–680. <https://doi.org/10.1038/nature02959>.
31. O'Dowd, C.D.; de Leeuw, G. Marine aerosol production: A review of the current knowledge. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2007**, *365*, 1753–1774. <https://doi.org/10.1098/rsta.2007.2043>.
32. Klett, J.D. Stable analytical inversion solution for processing lidar returns. *Appl. Opt.* **1981**, *20*, 211–220. <https://doi.org/10.1364/AO.20.000211>.

33. Fernald, F.G. Analysis of atmospheric lidar observations: Some comments. *Appl. Opt.* **1984**, *23*, 652–653. <https://doi.org/10.1364/AO.23.000652>.
34. Ansmann, A.; Wandinger, U.; Riebesell, M.; Weitkamp, C.; Michaelis, W. Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar. *Appl. Opt.* **1992**, *31*, 7113–7131. <https://doi.org/10.1364/AO.31.007113>.
35. Whiteman, D.N. Examination of the traditional Raman lidar technique. I. Evaluating the temperature-dependent lidar equations. *Appl. Opt.* **2003**, *42*, 2571–2592. <https://doi.org/10.1364/AO.42.002571>.
36. Bravo-Aranda, J.A.; Belegante, L.; Freudenthaler, V.; Alados-Arboledas, L.; Nicolae, D.; Granados-Muñoz, M.J.; Guerrero-Rascado, J.L.; Amodeo, A.; D’Amico, G.; Engelmann, R.; et al. Assessment of lidar depolarization uncertainty by means of a polarimetric lidar simulator. *Atmos. Meas. Tech.* **2016**, *9*, 4935–4953. <https://doi.org/10.5194/amt-9-4935-2016>.
37. Müller, D.; Mattis, I.; Wandinger, U.; Ansmann, A.; Althausen, D.; Dubovik, O.; Eckhardt, S.; Stohl, A. Saharan dust over a central European EARLINET-AERONET site: Combined observations with Raman lidar and Sun photometer. *J. Geophys. Res.* **2003**, *108*, 4345. <https://doi.org/10.1029/2002JD002918>.
38. Groß, S.; Freudenthaler, V.; Schepanski, K.; Toledano, C.; Schäfler, A.; Ansmann, A.; Weinzierl, B. Optical properties of long-range transported Saharan dust over Barbados as measured by dual-wavelength depolarization Raman lidar measurements. *Atmos. Chem. Phys.* **2015**, *15*, 11067–11080. <https://doi.org/10.5194/acp-15-11067-2015>.
39. MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297. Available online: <https://projecteuclid.org/euclid.bsmsp/1200512992> (accessed on 25 August 2026).
40. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. <https://doi.org/10.1109/TIT.1982.1056489>.
41. Arthur, D.; Vassilvitskii, S. k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2007; pp. 1027–1035. <https://doi.org/10.1145/1283383.1283494>.
42. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
43. Global Modeling and Assimilation Office (GMAO). MERRA-2 *avg1_2d_aer_Nx: 2d, 1-Hourly, Time-Averaged, Single-Level, Assimilation, Aerosol Diagnostics V5.12.4*; Goddard Earth Sciences Data and Information Services Center (GES DISC): Greenbelt, MD, USA, 2015. <https://doi.org/10.5067/KLICLTZ8EM9D>.
44. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>.
45. Randles, C.A.; da Silva, A.M.; Buchard, V.; Colarco, P.R.; Darmenov, A.; Govindaraju, R.; Smirnov, A.; Holben, B.; Ferrare, R.; Hair, J.; et al. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *J. Clim.* **2017**, *30*, 6823–6850. <https://doi.org/10.1175/JCLI-D-16-0609.1>.
46. Buchard, V.; Randles, C.A.; da Silva, A.M.; Darmenov, A.; Colarco, P.R.; Govindaraju, R.; Ferrare, R.; Hair, J.; Beyersdorf, A.J.; Ziemba, L.D.; et al. The MERRA-2 aerosol reanalysis, 1980 onward. Part II: Evaluation and case studies. *J. Clim.* **2017**, *30*, 6851–6872. <https://doi.org/10.1175/JCLI-D-16-0613.1>.
47. Gueymard, C.A.; Yang, D. Worldwide validation of CAMS and MERRA-2 reanalysis aerosol optical depth products using 15 years of AERONET observations. *Atmos. Environ.* **2020**, *225*, 117216. <https://doi.org/10.1016/j.atmosenv.2019.117216>.
48. Draxler, R.R.; Hess, G.D. An overview of the HYSPLIT_4 modelling system for trajectories, dispersion and deposition. *Aust. Meteorol. Mag.* **1998**, *47*, 295–308. <https://doi.org/10.1071/es98032>.
49. Stohl, A. Computation, accuracy and applications of trajectories—A review and bibliography. *Atmos. Environ.* **1998**, *32*, 947–966. [https://doi.org/10.1016/S1352-2310\(97\)00457-3](https://doi.org/10.1016/S1352-2310(97)00457-3).
50. Stein, A.F.; Draxler, R.R.; Rolph, G.D.; Stunder, B.J.B.; Cohen, M.D.; Ngan, F. NOAA’s HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 2059–2077. <https://doi.org/10.1175/BAMS-D-14-00110.1>.
51. Rolph, G.; Stein, A.; Stunder, B. Real-time Environmental Applications and Display sYstem: READY. *Environ. Model. Softw.* **2017**, *95*, 210–228. <https://doi.org/10.1016/j.envsoft.2017.06.025>.
52. NOAA Air Resources Laboratory. *Global Data Assimilation System (GDAS1) Archive Information*; NOAA Air Resources Laboratory: College Park, MD, USA. Available online: <https://www.ready.noaa.gov/gdas1.php> (accessed on 1 July 2026).
53. Ashbaugh, L.L.; Malm, W.C.; Sadeh, W.Z. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmos. Environ.* **1985**, *19*, 1263–1270. [https://doi.org/10.1016/0004-6981\(85\)90256-2](https://doi.org/10.1016/0004-6981(85)90256-2).
54. Polissar, A.V.; Hopke, P.K.; Paatero, P.; Malm, W.C.; Sisler, J.F. Atmospheric aerosol over Alaska: 2. Elemental composition and sources. *J. Geophys. Res.* **1998**, *103*, 19045–19057. <https://doi.org/10.1029/98JD01212>.

55. Hsu, Y.-K.; Holsen, T.M.; Hopke, P.K. Comparison of hybrid receptor models to locate PCB sources in Chicago. *Atmos. Environ.* **2003**, *37*, 545–562. [https://doi.org/10.1016/S1352-2310\(02\)00886-5](https://doi.org/10.1016/S1352-2310(02)00886-5).
56. Fleming, Z.L.; Monks, P.S.; Manning, A.J. Review: Untangling the influence of air-mass history in interpreting observed atmospheric composition. *Atmos. Res.* **2012**, *104–105*, 1–39. <https://doi.org/10.1016/j.atmosres.2011.09.009>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.