Text S1. IntelliSEM-EPAS analysis details

Developed by the RJ Lee Group in the United States, the IntelliSEM-EPASTM is an automated particle analysis system designed specifically for the characterization of environmental particulates, leveraging scanning electron microscopy-energy dispersive X-ray spectroscopy (SEM-EDS) technology. This system facilitates the rapid, precise, and comprehensive examination of environmental particles, including their morphological features, size distributions, and elemental compositions. Composed of two core software modules—Prospector and Workbench—the system not only controls the hardware to conduct automatic sample analysis but also interfaces with databases for the secure storage and efficient retrieval of analytical data.

The IntelliSEM-EPAS system utilized in this research is housed in the Joint Laboratory for Electron Microscopy Analysis of Atmospheric Particles, located in Beijing, China. The scanning electron microscope integrated into the system is a TESCAN field emission model (TESCAN CLARA), boasting a maximum resolution of 0.9 nm. It is paired with two Bruker XFlash 6|60 energy dispersive spectrometers, which exhibit a resolution of 129 eV (full width at half maximum of the Mn Kα peak). Equipped with a silicon drift detector (SDD) featuring a 60 mm^2 crystal area, the system ensures rapid data acquisition, high sensitivity, and exceptional accuracy for quantitative elemental analysis.

The IntelliSEM-EPAS was operated in backscattered electron (BSE) detection mode, with a working distance of 12 mm, an accelerating voltage of 20 kV, and a beam current of 1 nA. An internal reference standard containing carbon and aluminum was used to establish a consistent threshold for particle detection. During the computer-controlled scanning electron microscopy (CCSEM) analysis, the electron beam scanned the sample surface within a 120 μm field of view until the BSE signal surpassed the predefined detection threshold. Particles with average diameters ranging from 0.2 to 10 μm were subjected to further analysis by increasing the magnification, ensuring that each particle occupied approximately 50% of the field of view to improve the precision of size measurements. After performing 90 Rotated Feret measurements on each particle, the electron beam was positioned at the particle's centroid to collect an EDS

spectrum for 1 second. Subsequently, a digital image of the particle was captured and stored in the database.

The physicochemical data stored in the database for each particle includes: (1) a high-resolution microscopic image; (2) three shape parameters derived from the image: projected area (Aera), perimeter (Perim), and grayscale intensity (Intensity); (3) five size parameters: average diameter (Davg), equivalent circle diameter (Dcirc), maximum diameter (Dmax), minimum diameter (Dmin), and perpendicular diameter (Dperp, defined as the diameter perpendicular to Dmax); and (4) 24 chemical variables, namely carbon (C), aluminum (Al), silicon (Si), calcium (Ca), iron (Fe), sulfur (S), tin (Sn), titanium (Ti), manganese (Mn), nickel (Ni), copper (Cu), lead (Pb), sodium (Na), magnesium (Mg), potassium (K), chlorine (Cl), zinc (Zn), barium (Ba), vanadium (V), chromium (Cr), cobalt (Co), phosphorus (P), oxygen (O), and selenium (Se).

Text S2. Technical details of the k-means method for particulate matter classification

(1) Variable selection and preprocessing: After careful consideration, 24 chemical element variables (C, Al, Si, Ca, Fe, S, Sn, Ti, Mn, Ni, Cu, Pb, Na, Mg, K, Cl, Zn, Ba, V, Cr, Co, P, O, Se) were selected as input variables for cluster analysis. This selection was motivated by the fact that shape parameters (e.g., aspect ratio, roundness) and element concentration variables possess fundamentally different physical meanings and dimensional characteristics. Mixing all variables in a single clustering analysis could lead to unbalanced influences due to differing magnitudes, making interpretation challenging. Therefore, a stepwise strategy was adopted: clustering was first performed based on chemical composition to ensure that particles within the same cluster shared similar elemental signatures, and shape parameters were subsequently employed for in-depth discussion and auxiliary verification of each cluster's morphological characteristics. This approach guarantees the chemical rationality of the clustering results while fully leveraging morphological information for source identification.

(2) Noise reduction: During data preprocessing, denoising was applied to the raw X-ray energy-dispersive spectroscopy (EDS) data. Specifically, if the X-ray count of a given element for a particular particle was less than twice the square root of the total X-ray count of that particle, the count for that element was set to zero. This noise reduction threshold was established based on statistical signal-to-noise ratio considerations, effectively eliminating background noise and random interference from trace elements while preserving element signals with genuine physical significance, thereby enhancing the accuracy and stability of subsequent cluster analysis.

(3) Data transformation and standardization: Two critical preprocessing steps were performed prior to cluster analysis. First, logarithmic transformation (log-transformation) was applied to the elemental concentration data. This step is essential because elemental concentrations span multiple orders of magnitude (from trace to major elements), and logarithmic transformation effectively compresses the distribution range, preventing high-concentration elements from disproportionately dominating the clustering results. Second, the X-ray spectrum of each particle was normalized to unit length (vector normalization), ensuring that all 24 chemical element

variables carried equal weight in the clustering classification and avoiding biases introduced by dimensional differences. This standardization strategy balances the contribution of each element in the clustering process, facilitating the identification of particle categories based on genuine chemical similarities.

(4) Cluster analysis: Data from all seasons and sampling sites were combined, and cluster analysis was performed using the K-means algorithm. For determining the optimal number of clusters, the elbow method was primarily employed, following the approach of Ault et al. (2012). The elbow method plots the within-cluster sum of squares (WCSS) against the number of clusters k, and the "elbow point"-where the rate of decrease in WCSS sharply slows-indicates the optimal k. Additionally, multiple cluster validity indices, including the silhouette score and the Davies-Bouldin index, were used for comprehensive evaluation. As shown in Fig. S1, when k = 46, the WCSS curve exhibits a markedly decelerated decline, while the silhouette score reaches a satisfactory level. Considering both cluster compactness and interpretability, 46 clusters were ultimately selected for analyzing all particles in the IntelliSEM-EPAS dataset. This number strikes a favorable balance between minimizing intra-cluster error and avoiding excessive fragmentation of categories.

(5) Manual classification and merging: Following clustering, 46 chemical clusters were obtained. These were subsequently merged into 10 source categories with clear physical meanings based on the spectral similarity of EDS fingerprints, morphological characteristics, and established source emission profiles: biomass particles, carbonaceous particles, coal-fired ash, construction dust, metal particles, organic particles, salt particles, soil dust, sulfur-calcium particles, and sulfur-containing particles. This merging process relied primarily on signature elements and characteristic morphologies. For example, irregular flake-like particles enriched in Si and Al were classified as soil dust, whereas chain-aggregated particles dominated by pure C were identified as carbonaceous particles. For minor particle types with insufficiently distinct source characteristics or similar origins across different shapes, further subdivision was not performed to maintain a concise and practical classification scheme; these were collectively grouped into the "Others" category. The detailed

correspondence between the 46 chemical clusters and the 10 source categories, along with the judgment criteria for each category, is systematically presented in Table S1. The characteristics of the 46 particulate matter types are illustrated in Fig. S2, and the spectra (elemental spectra + shape spectra) of the 11 particle types are shown in Fig. S3.

(6) Software environment: The cluster analysis was conducted in Python 3.9. Core dependencies included scikit-learn (version 1.0.2), specifically the sklearn.cluster.KMeans module for executing the K-means algorithm and sklearn.preprocessing.StandardScaler for data standardization. NumPy (version 1.21.0) and Pandas (version 1.3.0) were used for numerical computation and data management, respectively. It should be noted that the K-means algorithm was applied exclusively to the 24 chemical element variables; the three shape parameters and five size parameters were not included as clustering inputs but were instead used for characteristic description and source identification verification after clustering was completed.

Text S3. Passive sampler (UNC-PAS)

UNC-PAS (University of North Carolina Passive Aerosol Sampler) is a passive sampler developed by the University of North Carolina. It collects particles through gravitational sedimentation and diffusion, and then analyzes them in a laboratory environment. This sampler does not require complex calibration procedures (Wagner and Leith, 2001; Castillo et al., 2019), and its structural diagram is shown in Figure S4 (Castillo et al, 2019).

Considering that particles of different sizes exhibit different settling velocities, the settling velocity of each individual particle must be calculated when using passively collected single particles for ambient concentration estimation. The relevant methodology had been provided in our previous study (Zhao et al., 2023). The specific calculation steps are outlined below.

To estimate an average airborne mass concentration over the sampling period, a deposition velocity model was developed by Wagner and Leith (Wagner and Leith, 2001). The ambient particle concentration (C) is calculated as:

$$C = \sum_{i=1}^{n} C_i = \sum_{i=1}^{n} \frac{F_i}{v_{dep,i}} = \sum_{i=1}^{n} \frac{\pi d_{ev,i}^3 \rho_p}{6 A_i t} \frac{1}{v_{dep,i}} \quad (1)$$

Where n is the total number of particles examined in a specific sample, C_i is the contribution of particle i to total particle concentration, F_i is the flux of the particle i to a specific filter, $d_{ev,i}$ is the equivalent volume diameter of particle i (= $d_{pa,i}$ /S_v), $d_{pa,i}$ is the projected-area-equivalent diameter of particle i, S_v is the assumed volume shape factor (= 1.6), ρ_p is the assumed particle density (2 g/m^3, taken from Sparks and Wagner (2021)), A_i is the analysis area of the substrate of the particle size range including particle i, t is the sampling duration and $v_{dep,i}$ is modeled particle deposition rate defined by:

$$v_{dep,i} = v_{amb,i} \gamma_{m,i} \quad (2)$$

where $v_{amb,i}$ is the ambient deposition velocity, $\gamma_{m,i}$ is an empirical correction designed to account for the combined effect of sampler and mesh.

However, recent work has shown that submicron-sized particles are

underestimated due to the lack of mesh cap porosity dependence on diffuse particles and other reasons (Shirdel et al., 2018). To improve this deficiency, a new deposition velocity model was developed to include explicit expressions for convective diffusion and cap interception (Castillo et al., 2019). Equation (1) is also applicable to the new model. Different from equation (2), the $v_{dep,i}$ of the new model is calculated as follows:

$$v_{dep,i} = P_c \times \left(1 - \frac{d_{pa,i}}{d_{pore}}\right)^2 \times \left\{ v_{amb,i} + \left[\frac{D}{X_d} \times \frac{1}{P_c + \left(X_c / X_d\right)} \right] \right\} \tag{3}$$

where P_c is the porosity of the UNC-PAS cap screen, d_{pore} is the diameter of cap screen pores, D is the Brownian diffusion coefficient, X_d is the distance between the cap screen and collection substrate and X_c is cap screen thickness. v_{amb} can be calculated as in Wagner and Leith (2001):

$$v_{amb} = \frac{-v_t}{[(1 - 0.67\tau^{0.49} u_*^{-0.02} \nu^{-0.49} v_t) e^{-v_t I}] - 1} \tag{4}$$

$$v_t = \tau g \tag{5}$$

$$\tau = \frac{\rho_0 d_a^2 C_{c,d_a}}{18\mu} \tag{6}$$

$$d_a = d_{pa} \left(\frac{C_{c,d_{ev}}}{C_{c,d_a}} \right)^{\frac{1}{2}} \left(\frac{\rho_p}{\rho_0} \frac{1}{S_d} \right)^{\frac{1}{2}} \frac{1}{S_v} \tag{7}$$

$$I = \frac{I^+}{u_*} \tag{8}$$

$$I^+ = \frac{1}{\frac{3\sqrt{3}}{29\pi} Sc^{-\frac{2}{3}} + 6.2 \times 10^{-4} (\tau^+)^2} \tag{9}$$

$$Sc = \frac{\nu}{D} \tag{10}$$

$$d_{es} = d_{pa} \left(\frac{f}{\pi} \right)^{\frac{1}{2}} \tag{11}\backslash$$

$$d_a = d_{pa}\left(\frac{C_{c,d_{ev}}}{C_{c,d_a}}\right)^{\frac{1}{2}}\left(\frac{\rho_p}{\rho_0}\frac{1}{S_d}\right)^{\frac{1}{2}}\frac{1}{S_v} \tag{12}$$

$$\tau^+ = \frac{\tau u_*^2}{v} \tag{13}$$

$$\begin{cases} u_* = 10^{-25} & u = 0 \\ u_* = \dfrac{0.4u}{\ln(\dfrac{z}{z_0})} & u \neq 0 \end{cases} \tag{14}$$

where v_t is the terminal settling velocity, Cc is the Cunningham correction factor, μ is the dynamic viscosity, ν is the kinematic viscosity, ρ_0 is the unit particle density, g is the gravitation acceleration (9.8 m/s^2), D is the Brownian diffusion coefficient, k is the Boltzmann constant, T is the ambient temperature, d_{es} is the equivalent surface diameter, d_a is the aerodynamic diameter, f is the surface shape factor, and S_d is the dynamic shape factor, Sc is the Schmidt number, u is the wind speed at height z above the ground and z_0 is the surface roughness, u_* is the turbulent friction velocity.

Table S1 Source Type Identification Characteristic Table

Source Type	Core Elements	Micro-morphology Characteristics
Biomass particles	• Predominantly C and O, containing minor amounts of K, P, and other elements	• Irregular morphology, rough surface
Carbonaceous particles	• Predominantly C	• Amorphous structure, typical chain-like aggregates
Coal-fired ash	• High contents of Al and Si, containing metal and non-metal elements such as Fe, Ca, Mg, K, and Na, exhibiting multi-element mixed characteristics	• Three-dimensional spherical morphology
Construction dust	• Predominantly Ca	• Irregular morphology, appearing as massive, flaky or fibrous shapes
Metal particles	• Rich in metal elements such as Fe, Al, Ti, Zn, Ba, and Mg	• Small spherical aggregates with bright surfaces
Organic particles	• Predominantly C and O	• Mainly spherical or dome-shaped, easily decomposed under electron beam
Salt particles	• Predominantly Na and Cl	• Regular cubic or sub-cubic structure, smooth crystal surface
Soil dust	• Predominantly Si and Al, containing minor amounts of crustal elements such as Fe, Ca, and Mg	• Irregular flaky shape
Sulfur-calcium particles	• Predominantly S and Ca	• Platy, prismatic or irregular shapes
Sulfur-containing particles	• Rich in S	• Diverse morphology, typically spherical or amorphous, with relatively smooth surface

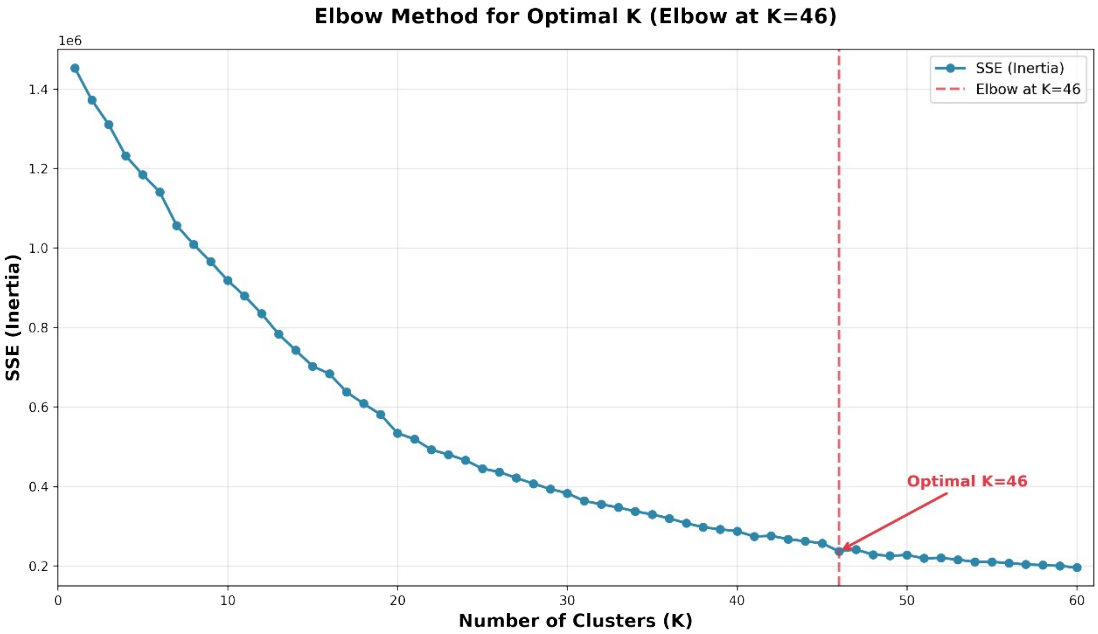
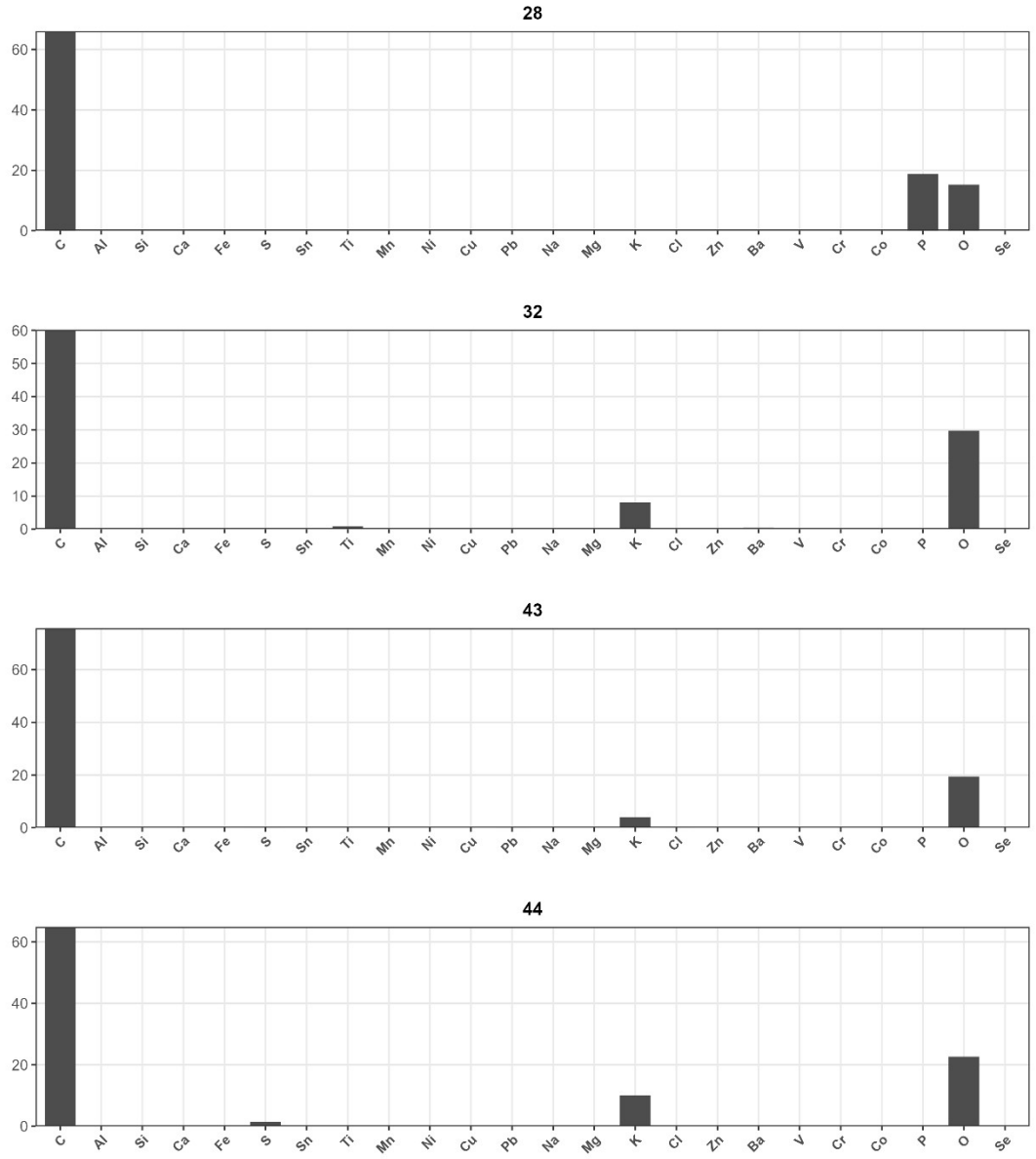
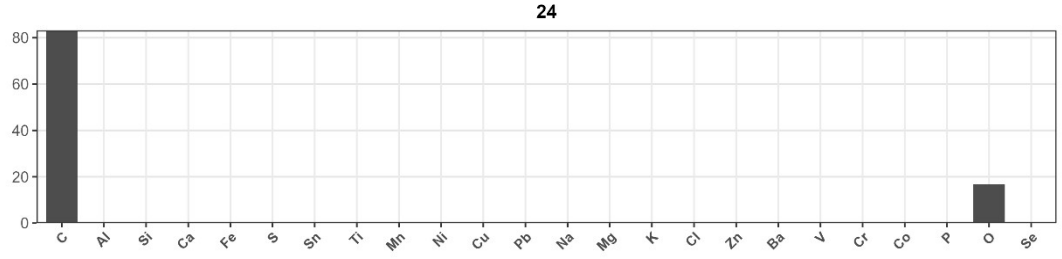


Fig S1 K-means clustering elbow plot

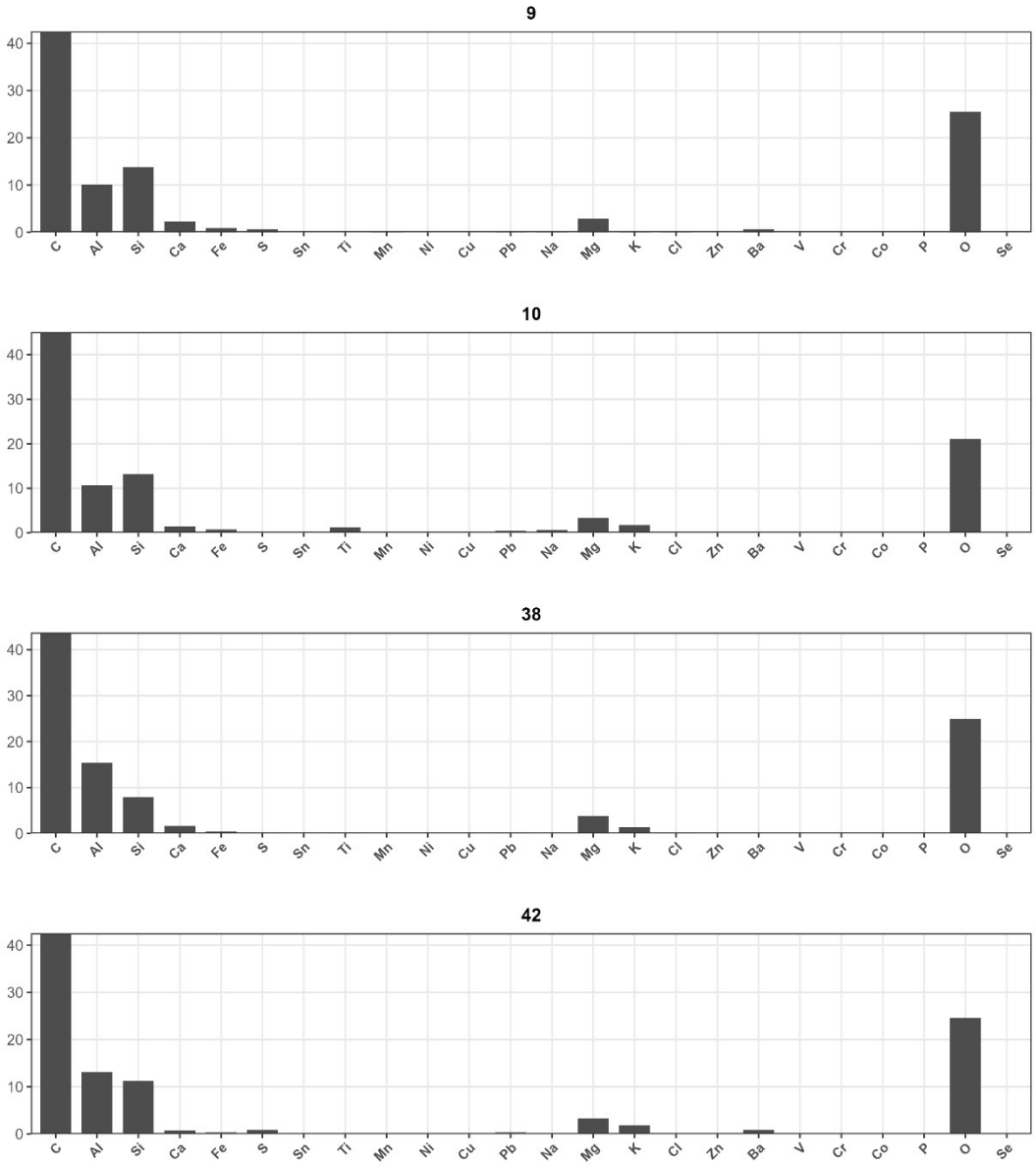
Biomass particles



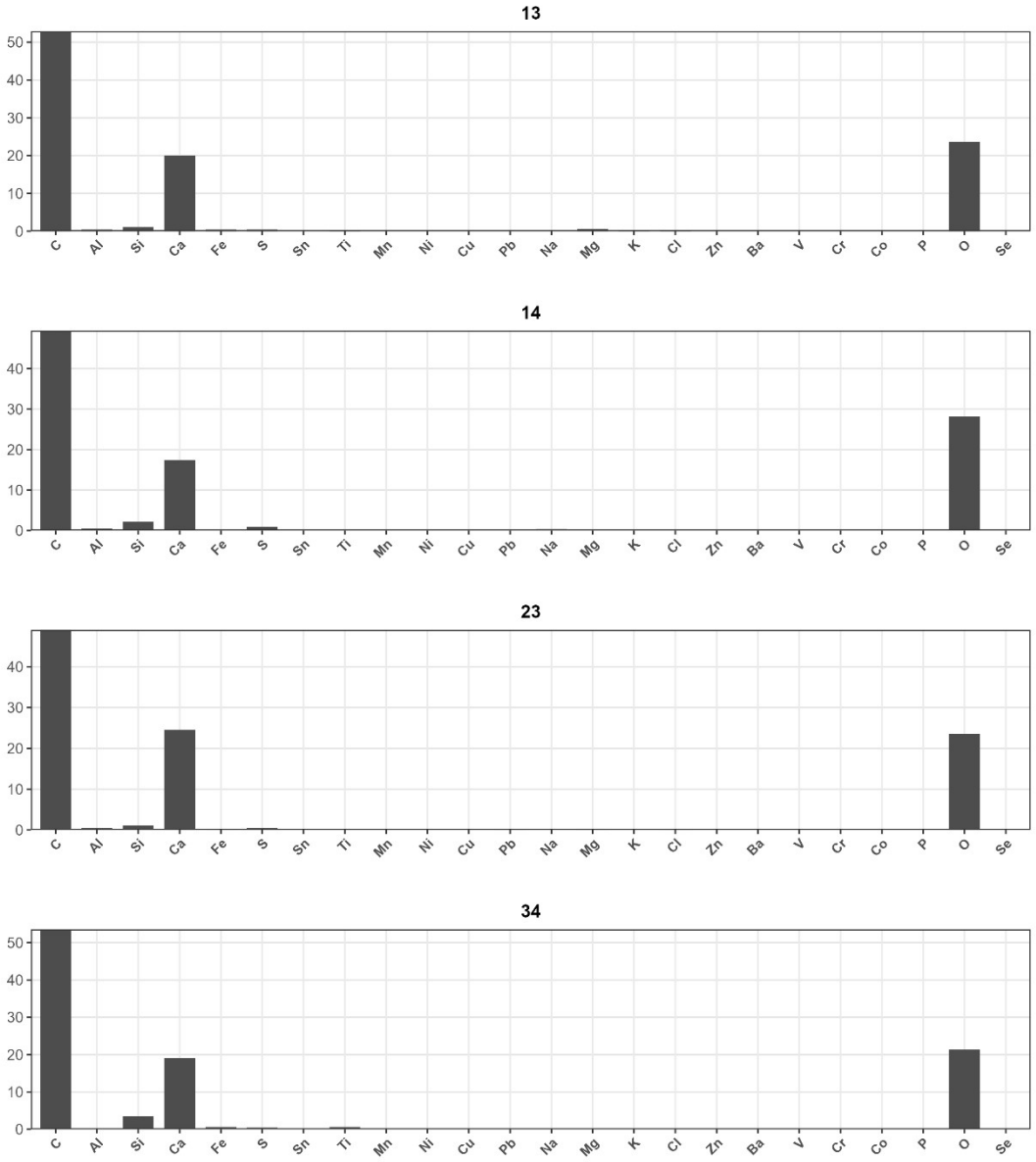
Carbonaceous particles



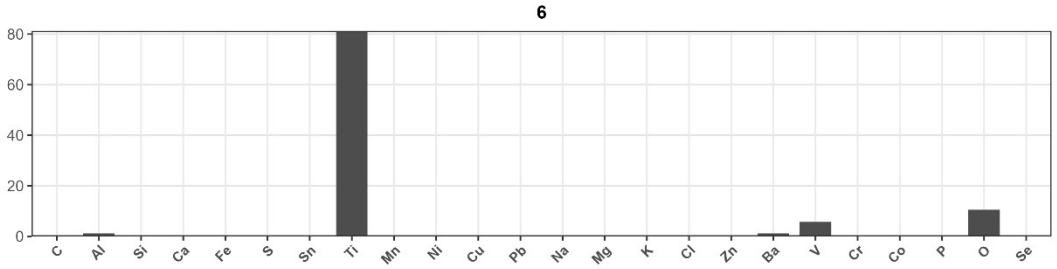
Coal-fired ash

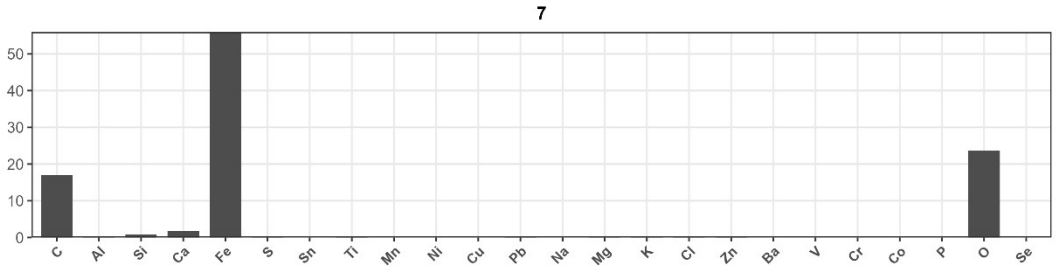


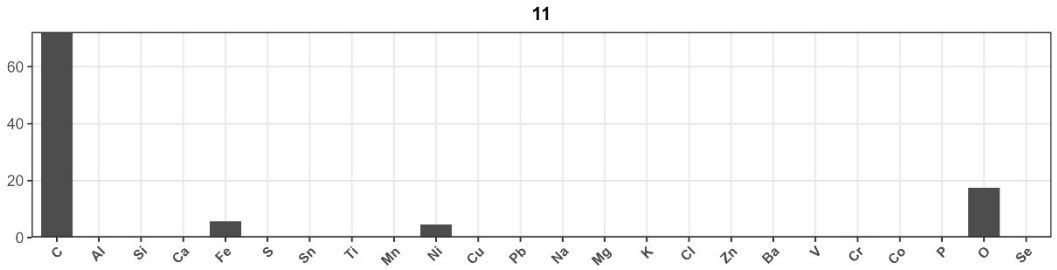
Construction dust

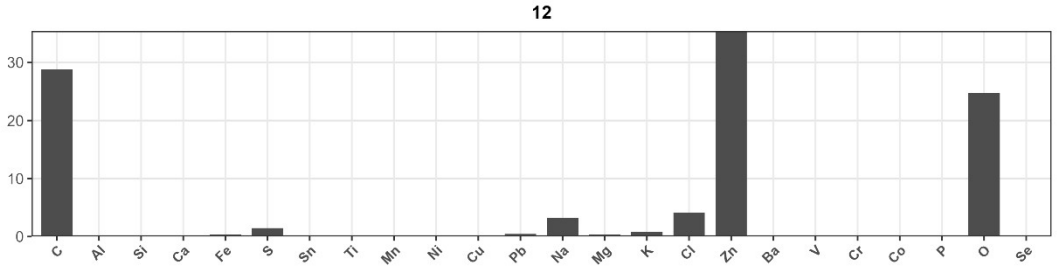


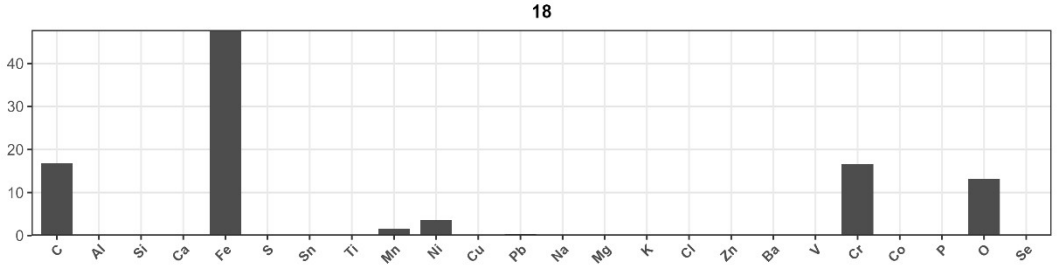
Metal particles

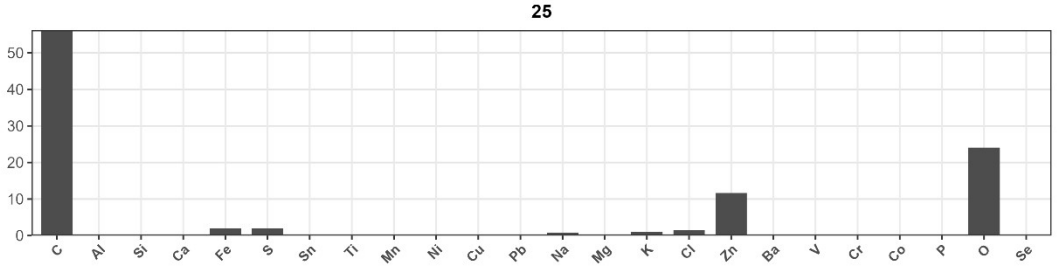

6
80
60
40
20
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

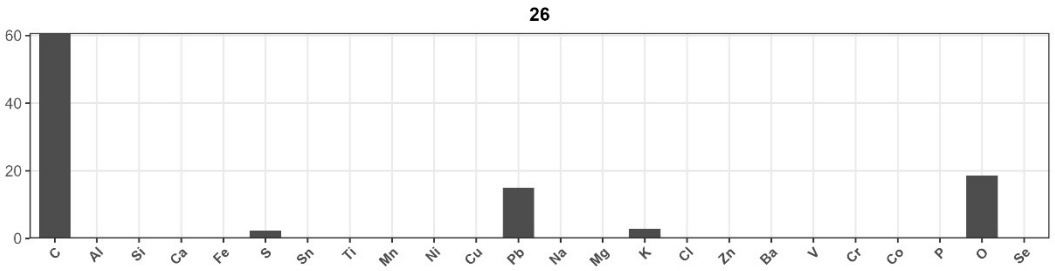

7
50
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

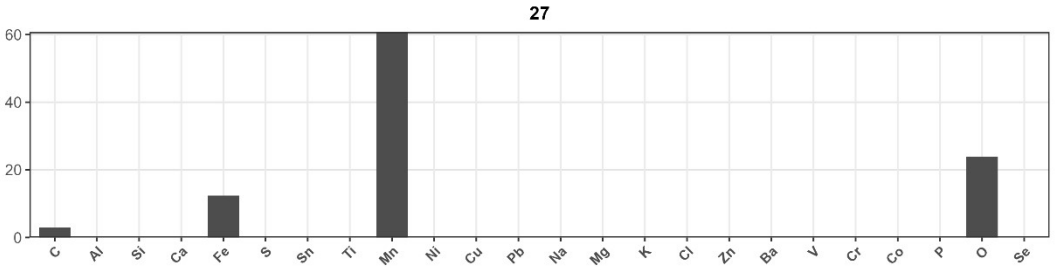

11
60
40
20
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

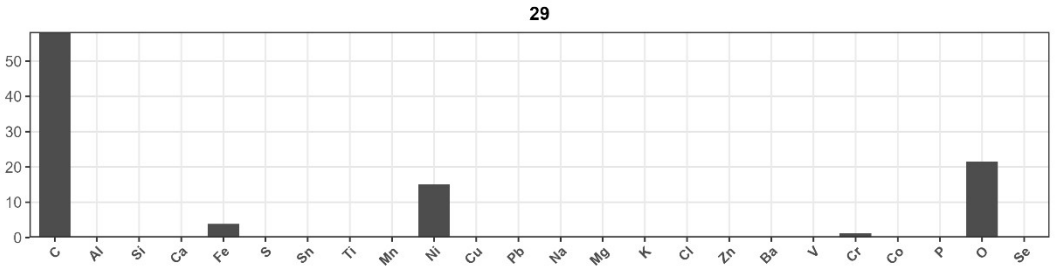

12
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

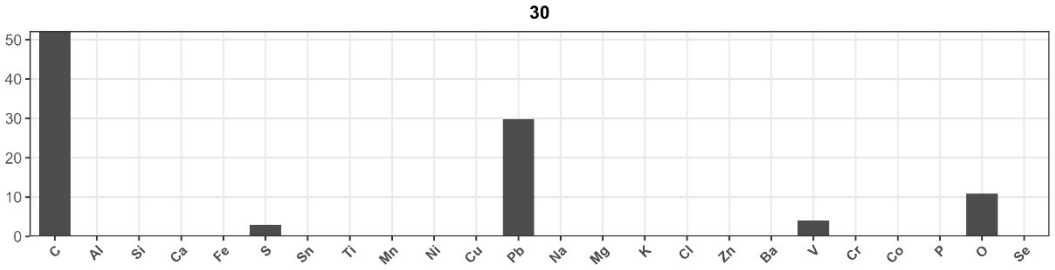

18
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

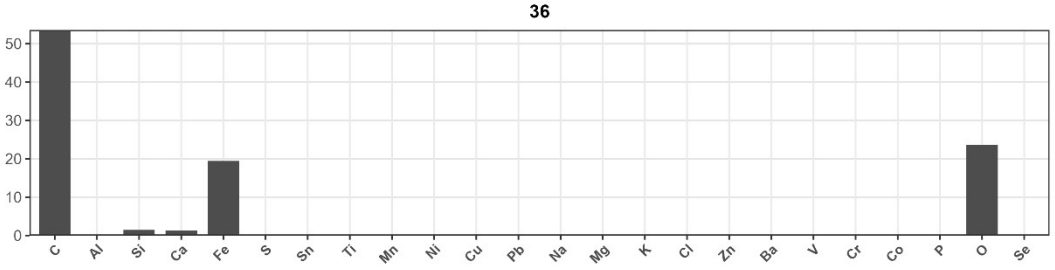

25
50
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

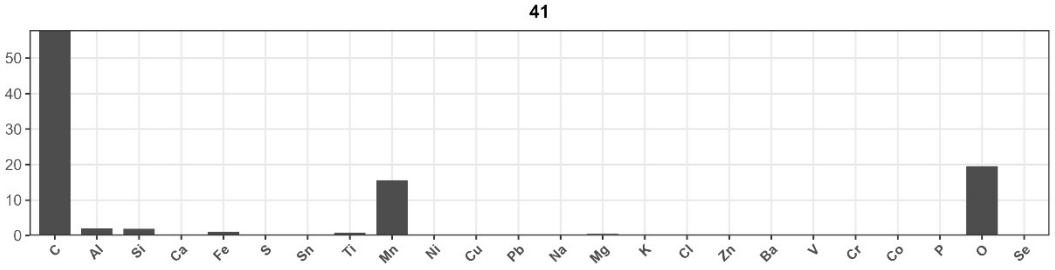

26
60
40
20
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se


27
60
40
20
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

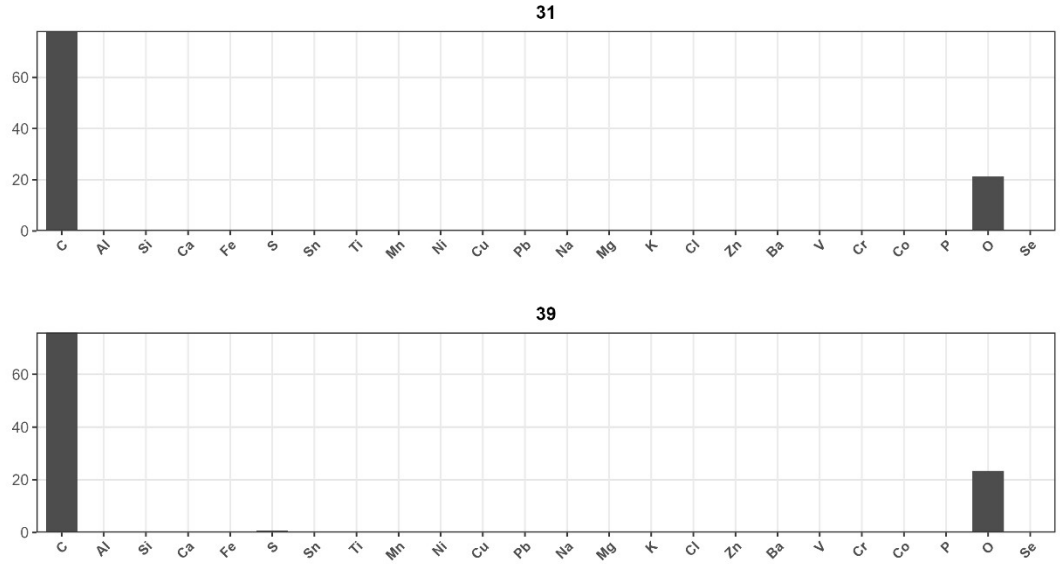

29
50
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se


30
50
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

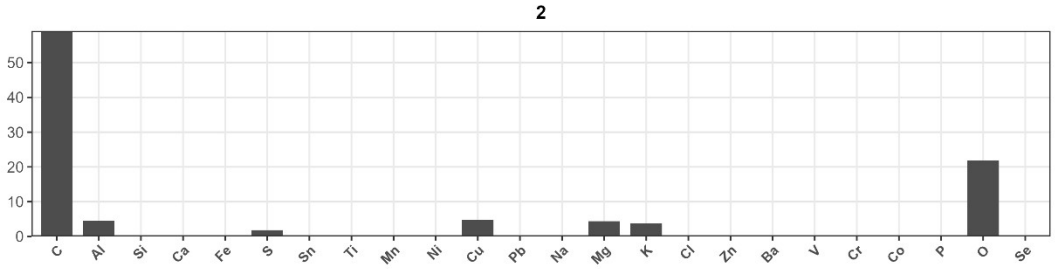

36
50
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

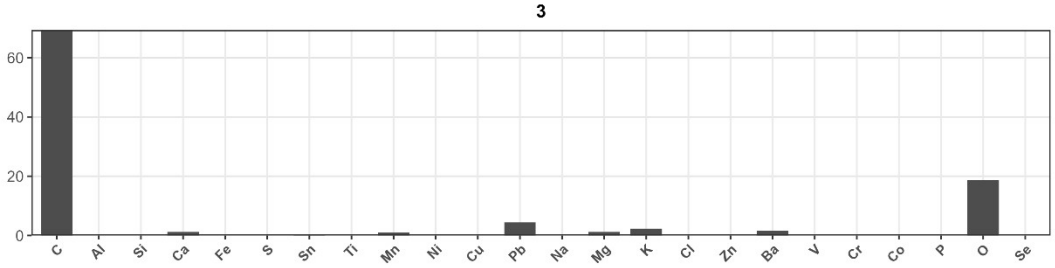

41
50
40
30
20
10
0
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se

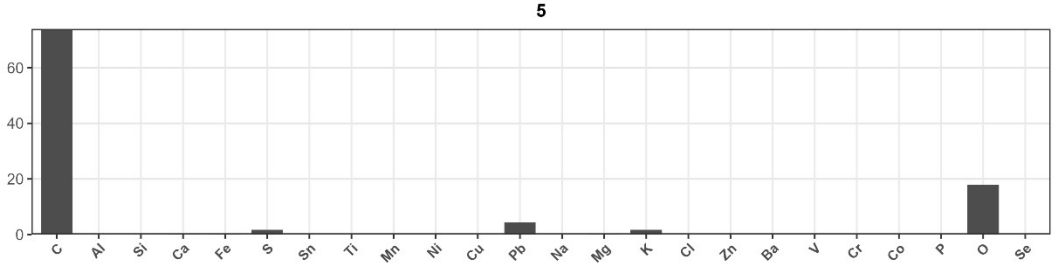
Organic particles

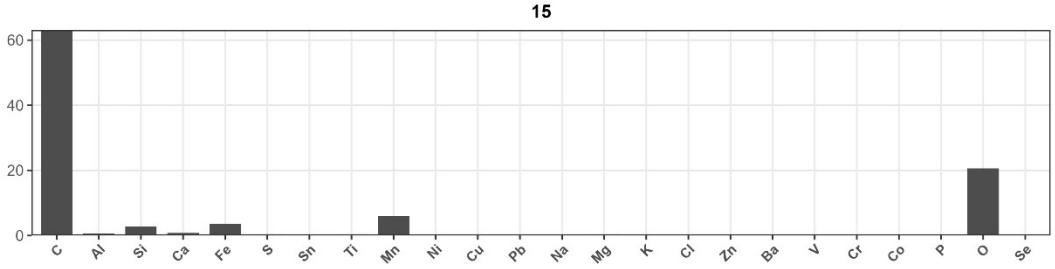


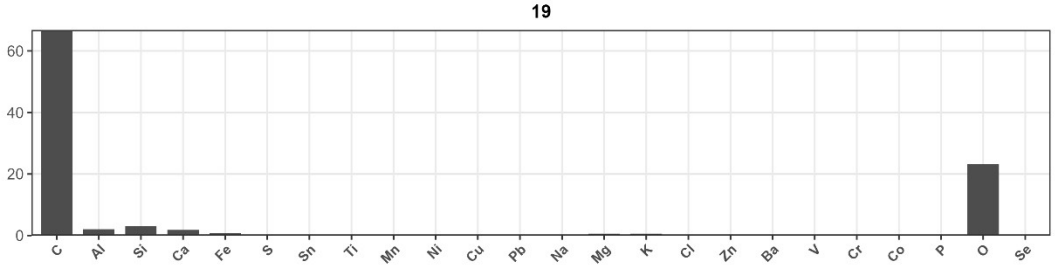
Other

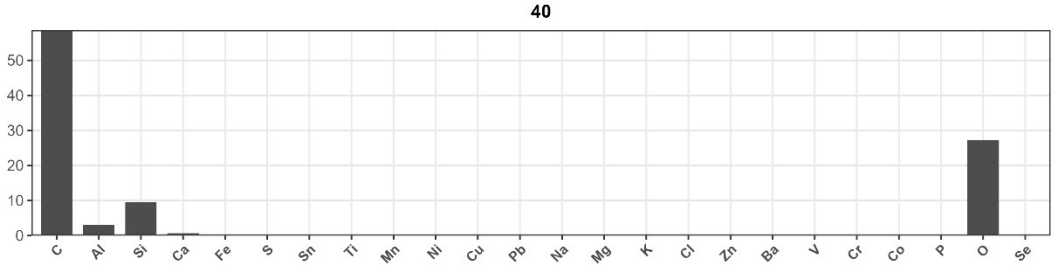

2
0
10
20
30
40
50
C
Al
Si
Ca
Fe
S
Sn
Ti
Mn
Ni
Cu
Pb
Na
Mg
K
Cl
Zn
Ba
V
Cr
Co
P
O
Se


3
0
20
40
60
C
Al
Si
Ca
Fe
S
Sn
Ti
Mn
Ni
Cu
Pb
Na
Mg
K
Cl
Zn
Ba
V
Cr
Co
P
O
Se

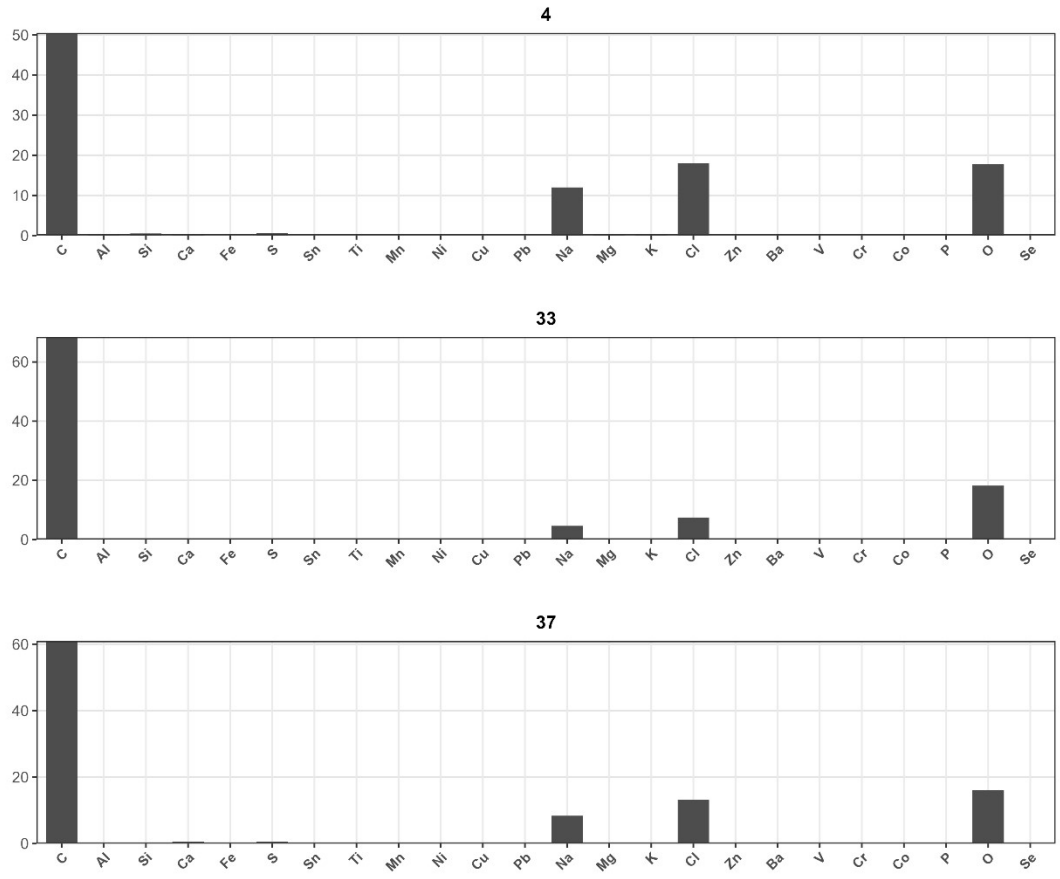

5
0
20
40
60
C
Al
Si
Ca
Fe
S
Sn
Ti
Mn
Ni
Cu
Pb
Na
Mg
K
Cl
Zn
Ba
V
Cr
Co
P
O
Se


15
0
20
40
60
C
Al
Si
Ca
Fe
S
Sn
Ti
Mn
Ni
Cu
Pb
Na
Mg
K
Cl
Zn
Ba
V
Cr
Co
P
O
Se

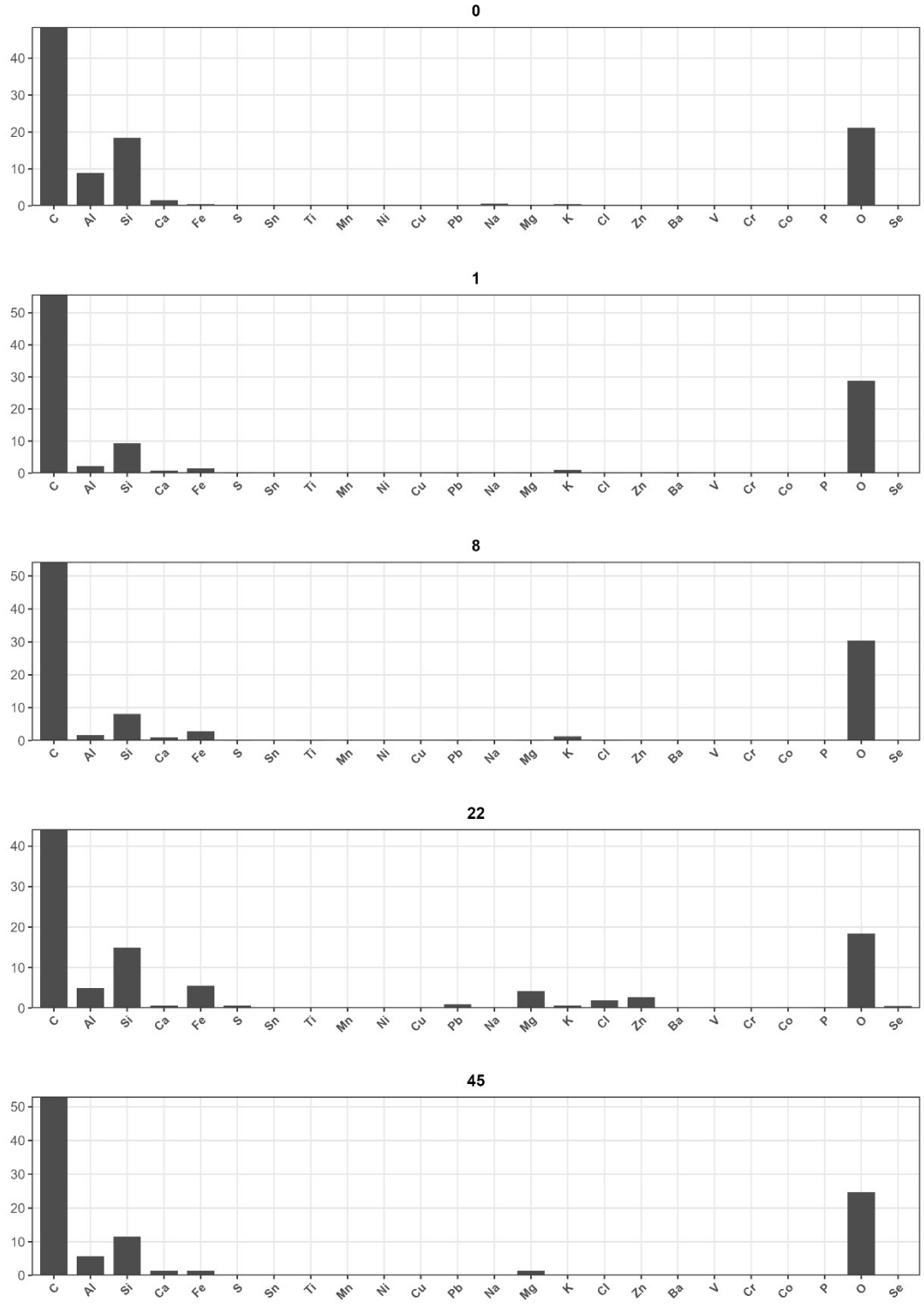

19
0
20
40
60
C
Al
Si
Ca
Fe
S
Sn
Ti
Mn
Ni
Cu
Pb
Na
Mg
K
Cl
Zn
Ba
V
Cr
Co
P
O
Se


40
0
10
20
30
40
50
C
Al
Si
Ca
Fe
S
Sn
Ti
Mn
Ni
Cu
Pb
Na
Mg
K
Cl
Zn
Ba
V
Cr
Co
P
O
Se

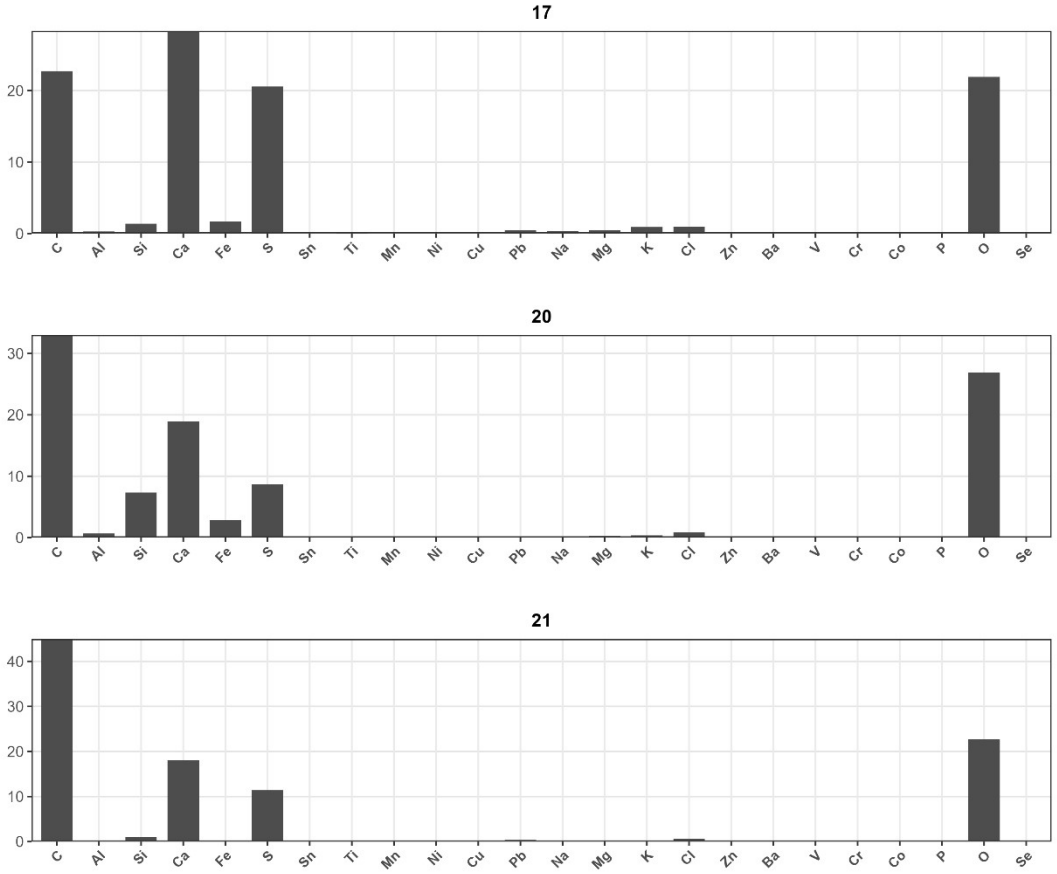
Salt particles



Soil dust



Sulfur-calcium particles



Sulfur-containing particles

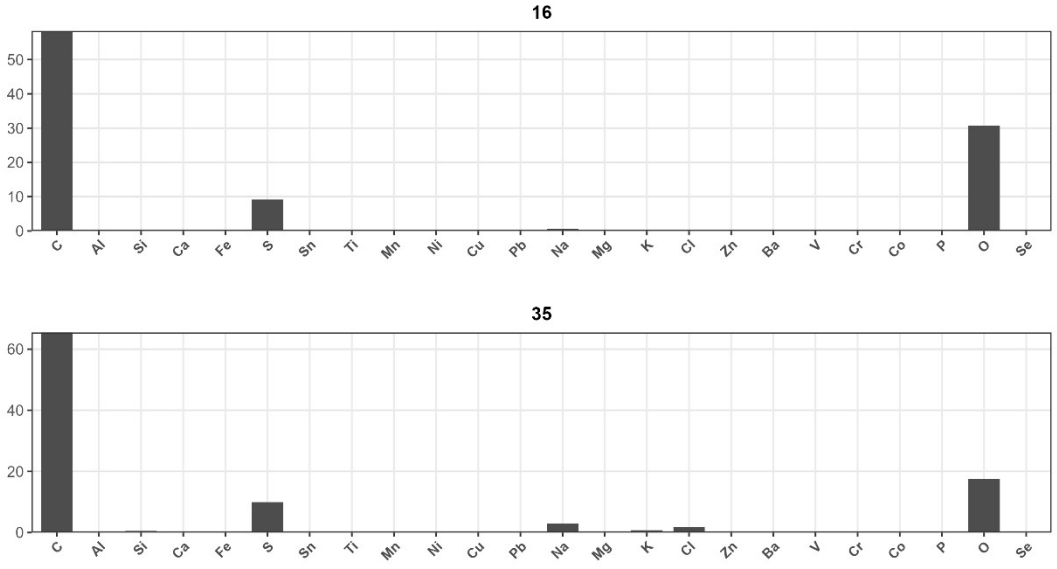


Fig S2 The correspondence between 11 types of particles and 46 categories

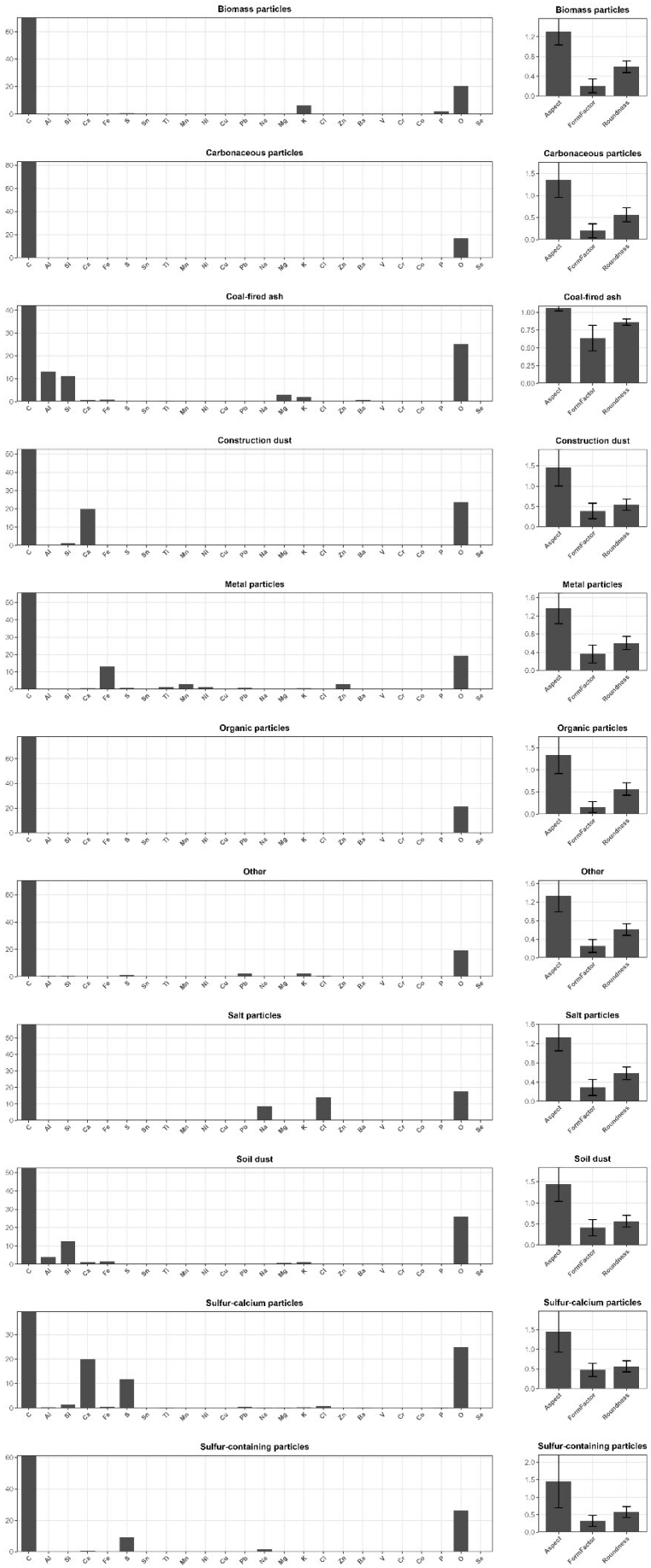

Biomass particles
Carbonaceous particles
Coal-fired ash
Construction dust
Metal particles
Organic particles
Other
Salt particles
Soil dust
Sulfur-calcium particles
Sulfur-containing particles
C Al Si Ca Fe S Sn Ti Mn Ni Cu Pb Na Mg K Cl Zn Ba V Cr Co P O Se
Aspect FormFactor Roundness

Fig S3 The spectrum (elemental spectrum + shape spectrum) of 11 types of particles

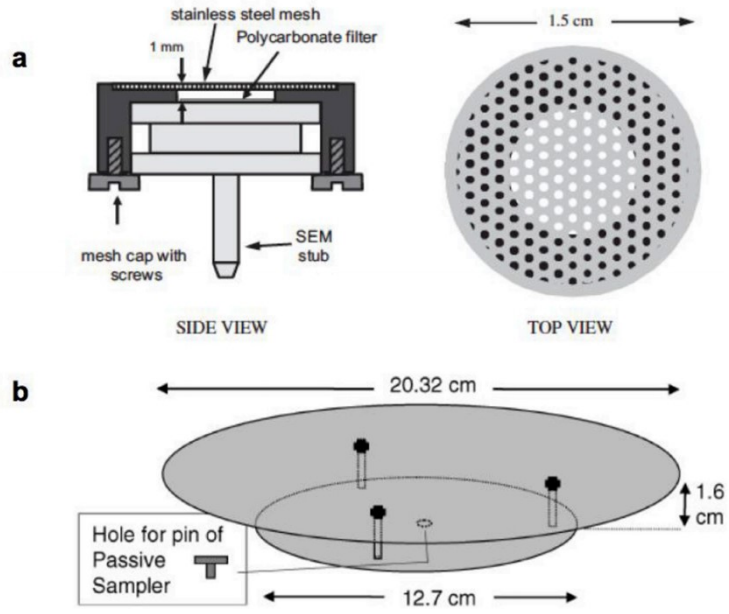


Fig. S4. UNC-PAS filter side and top views (a), and the weather shelter consisting of two circular aluminum plates (b).

REFERENCES

M. D. Castillo, J. Wagner, G. S. Casuccio, R. R. West, F. R. Freedman, H. M. Eisl, Z. M. Wang, J. P. Yip, P. L. Kinney, Field testing a low-cost passive aerosol sampler for long-term measurement of ambient PM2.5 concentrations and particle composition. Atmos. Environ., 2019, 216, 116905.

P. Zhao, P. S. Zhao, J. Tang, G. S. Casuccio, J. Gao, J. Li, Y. Y. He, M. Y. Li, Y. C. Feng, Source identification and apportionment of ambient particulate matter in Beijing using an advanced computer-controlled scanning electron microscopy (CCSEM) system. Sci. Total Environ., 2023, 861, 160608.

J. Wagner, D. Leith, Passive aerosol sampler. Part I: principle of operation. Aerosol Sci. Technol., 2001, 34, 186–192.

M. Shirdel, B. M. Andersson, I. A. Bergdahl, J. N. Sommar, H. Wingfors, I. E. Liljelind, Improving the UNC passive aerosol sampler model based on comparison with commonly used aerosol sampling methods. Ann. Work. Expo. Health, 2018, 62, 328–338.