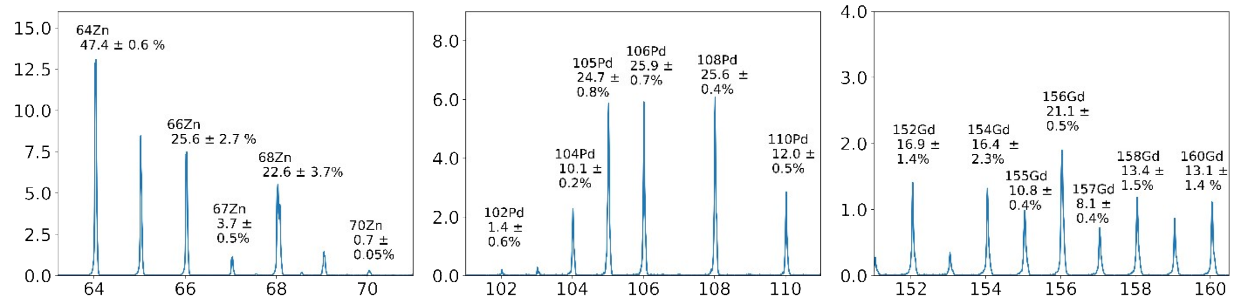
Electronic Supplementary Material (ESI) for Journal of Analytical Atomic Spectrometry.

Supplementary information

1 Extraction of colloids from environmental samples (cont.)

The detailed method is as follows, the preparation of extractions solutions, Solution A 10% v/v of Triton-x 114 (Sigma Aldrich, USA) was made by the addition of 1 mL of Triton-x into a 10ml conical polypropylene centrifuge tube containing 9 m: of MQ-water. This vial was then vortexed for 5 min at 2850 rpm. Solution B, a saturated EDTA (Titriplex III, Merck, Germany) solution was prepared by adding 4 mg of Triplex III and 10 mL of MQ water into a conical tube and was then vortexed for dissolution for 5 min at 2850 rpm. Solution C, a 1.25 M Na-Acetate solution (Merck, Germany) was prepared by adding 1 g of Sodium acetate anhydrous 99.99% Suprapur® (Merck, Germany) and 10 mL water into a conical tube and vortexing for 5 min at 2850 rpm. Solution D, 1M Acetic acid 96% was created by adding 600 ul and 9.4 mL of MQ-water in a conical tube and gently mixed by hand. The reported stability of solutions A, B, C, and D is 1 month.

A mixture of 25 µl of Solution A, 25 µl Solution B, 2.5 µl Solution C, and 10µl Solution D and 2 ml MQ-water in a conical tube was vortexed for 5 minutes at 2850 rpm or until completely dispersed. 100mg of each of the soil samples were added to vials containing the mixture of additives and vortexed for 5 min at 2850 rpm. The extraction mixture was then placed in a temperature-controlled water bath at 40±0.2°C for 2 h until the clouding of the samples was visible, indicating the colloids were encapsulated in micelles. The samples were then centrifuged for 12 min at 2500 rpm (equal to 3238 g). After the supernatant was removed with a pipette. The remaining pellet consisting of matrix and colloids is suspended in 2 mL of MQ-water and vortexed 5 min at 2850 rpm, 5 min ultrasound and5 min at 2850 rpm. The resuspended suspension was then centrifuged for 2 min at 3000 rpm (equal to 1459 g). At this g value, the particle size cut-off is ~300nm.



SI Figure 1. Isotopic patterns for Zn, Pd, and Gd, the spectra are the aggregated mean of all detected particle events and averaged for S1-3.

2 Particle quantification

After the measurement of the samples, the raw data is processed using the following steps and follow the methods outlined in the Tofwerk solution and particle processing guide provided at soft.tofwerk.com for their ToFPilot interface:

1. Calculate average signals and deviations for all analytes for standards where A is the average signal, SD is the standard deviation, a_i is the intensity of the analyte, n is the number of measurement points. $\bar{a}$ is the mean of all a_i:

$$A = \frac{\sum_{i=1}^{n} a_i}{n}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (a_i - \bar{a})^2}{n-1}}$$

2. Perform blank corrections for calibration analytes

$$A_{corr} = A_{raw} - A_{blank}$$

3. Calculate calibration equations using linear fit where y_i is the signal intensity in counts, a_i is the slope of the curve in counts/ppb, x is the analyte concentration in ppb and b_i is the intercept in counts.

$$y_i = a_i x + b_i$$

4. Calculate transport efficiency (T_{eff}) for standard particle where y_{cts_0} is the sum of fit intercept and slope for the standard particle, y_{cts_1} is the integration time corrected y_{cts_0}, x_o is the median value of all cts per peak of the standard particle after the iterative signal, and background separation and split event correction. t_{int_1} the integration time from particle standard, t_{int_2} the integration time from analyte from liquid standard, m_{sp} the particle mass of the standard particle calculated from size and density. f_{mass} the mass flow, l_{sol} the concentration of liquid standards, m_a the analyte mass.

$$y_{cts\_1} = \frac{y_{cts\_0} * t_{int\_1}}{t_{int\_2}}$$

$$m_a = \frac{m_{sp} * y_{cts\_1}}{x_0}$$

$$T_{eff} = \frac{m_a * 10^9}{f_{mass} * t_{int\_1}}$$

5. Calculate the mass entering the plasma m_{ML} for all standard concentrations (l_{conc}), correct the values for integration time resulting in $y_{cts\_int\_1}$, then perform another linear fit resulting in mass calibration curves.

$$m_{ML} = l_{conc} * m_a$$

$$y_{cts\_int\_1} = \frac{y_{cts}}{t_{int\_1} / t_{int\_2}}$$

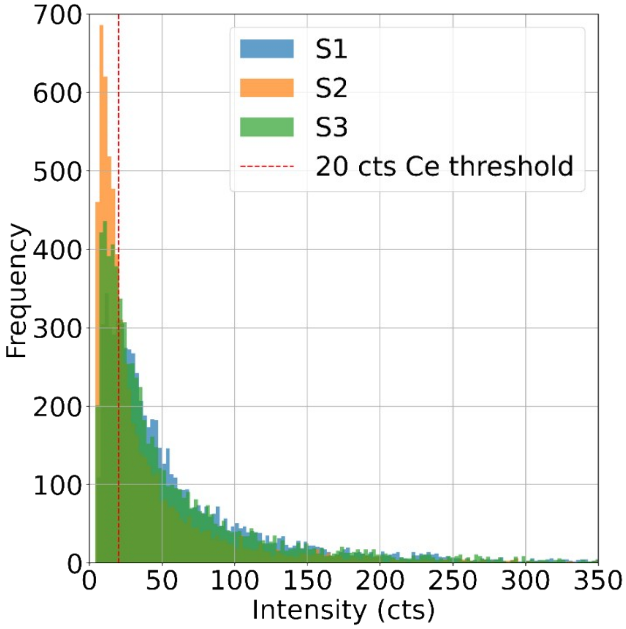
6. Calculate mass LODs using the IUPAC method using the SD of the blank and the slope of mass calibration curves a_g.

$$LOD = \frac{3.29 * SD + 2.72}{a_g}$$

7. Perform iterative particle separation for particle samples using a 1000-point averaging window and three cycles of thresholding based on the mean and standard deviation of the sample. Following this, split-peak processing is done by adding up neighboring signals that are higher than zero maintaining their positions. Using the found peaks, the intensity is converted to mass using the mass calibration curves.

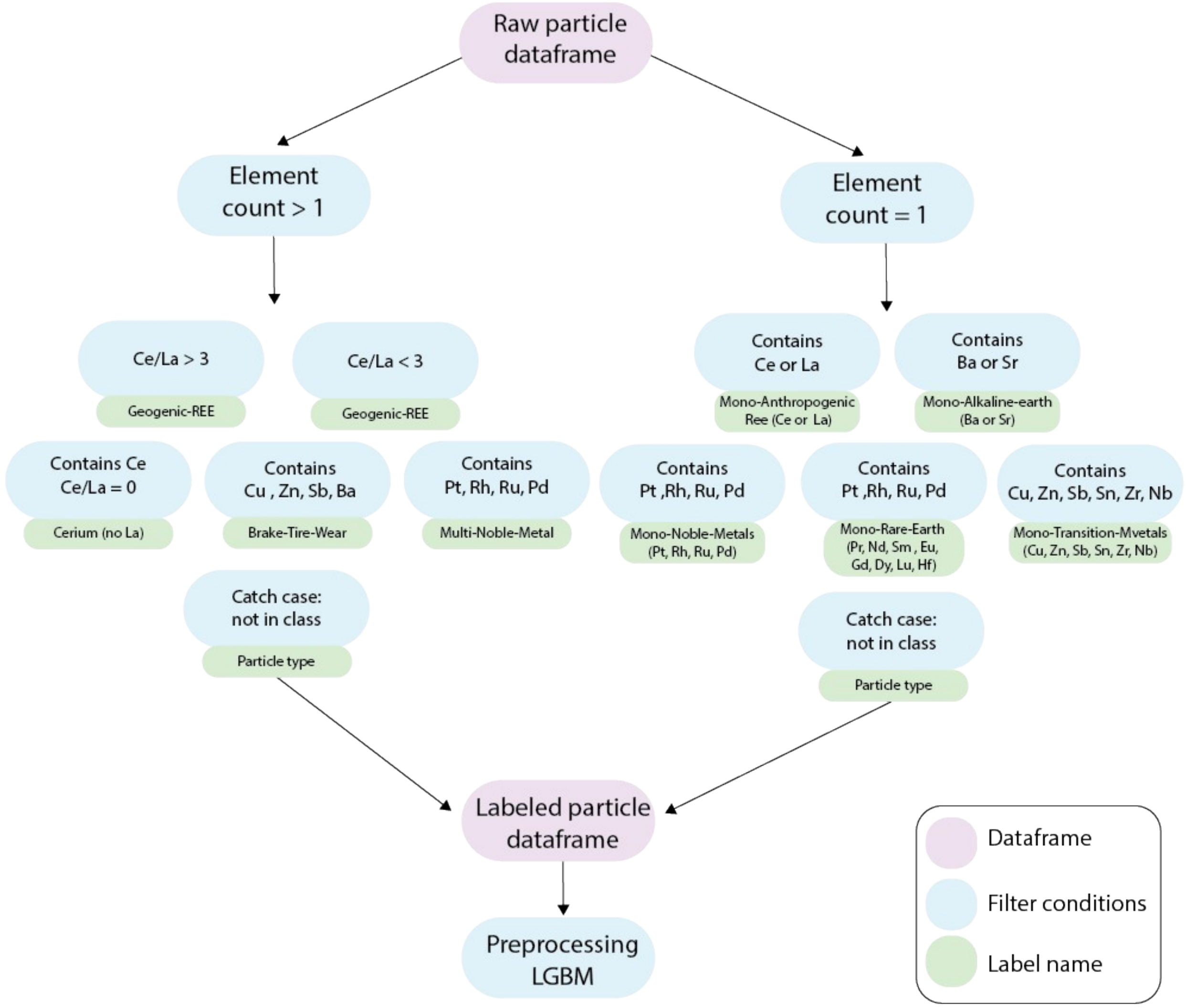
$$N_{thresh} = \mu_{sample} + (3.29 * \sigma_{sample}) + 2.72$$

8. The particle datasets are then further filtered to remove particle-containing fewer than 20 cts of Ce to minimize the error from sampling noise in the Ce/La ratio which may cause false classifications during training of the classifier. The threshold and Ce count distribution are depicted for S1-3 in SI Figure 2.



SI Figure 2. Cts histogram of Ce particles indicating a 20 cts cutoff threshold (Red-dotted line). S1: Blue, S2: Orange, S3: Green

9. The new filtered data frames then have 4 new columns appended to them that included the number of elements per particle (int32), particle type (str), and encoded particle type (int32), and CeLa ratio (float32).
10. The labeled and encoded datasets are then subjected to homo and hetero coincident corrections. The datasets are split into single and multi-elemental particles (SEP, MEP). Homo double event corrections are done by calculating the probability of an elemental event occurring during a dwell time over the measurement time for SEPs. These probabilities are then used to predict the number of homo coincident events throughout an entire measurement run (N_{homo}). The mean of the particle distribution is then calculated for each element that was detected in a SEP form. The value N_{homo} is used to subtract N particles closest to 2 times the distribution mean. The values are then re-appended to the MEP dataset for use in the hetero-particle correction step.
11. Hetero particle coincident correction follows the same approach as the one recently published in an emerging investigated series.[1] The process starts by finding all combinations of two-particle (f_1,f_2) types that appear in the dataset. From these combinations, we determine which calculated combinations appear in the dataset and append it to a new dataset for calculations. The frequency of occurrence for f_1,f_2 is found and appended to the dataset. The multiplication of the f_1,f_2 divided by the total number of measurement points gives use the predicted number of events for each hetero coincident particle ($C_{f1,f2}$). These values are converted to an integer to make sure predicted particle counts are discrete. The scoring method described in the cited publication is then applied to the particle where the element's intensity is normalized to the largest intensity in the particle. The similarity score is then calculated to determine how similar each particle is to the population and the value $C_{f1,f2}$ is then used to subtract the particles with the largest difference from the population.
12. The datasets are then further labeled into respective classes used for training the LGBM multi-class classifier. This process is depicted in SI Figure 3



SI Figure 3. Flow chart of the labeling scheme used for the annotation of the particle classes described in the main text. The labeled particle data frame was used for the training of the LGBM classifier.

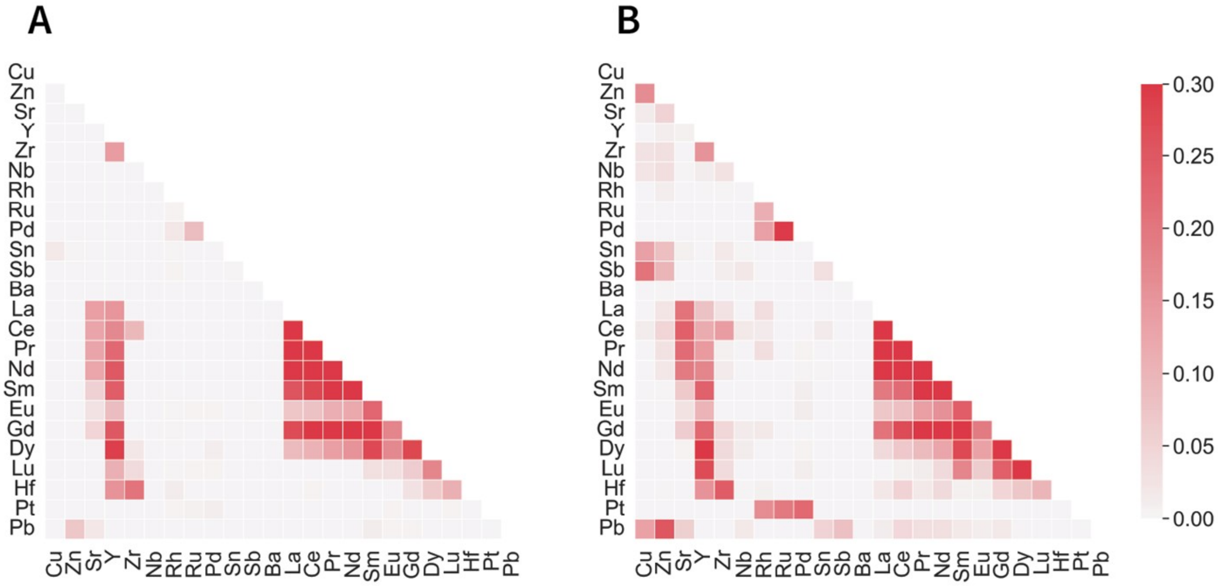
3 Results/data

SI Table 1. Size characterization (DLS & SEM) of the cloud point extracted sample suspension

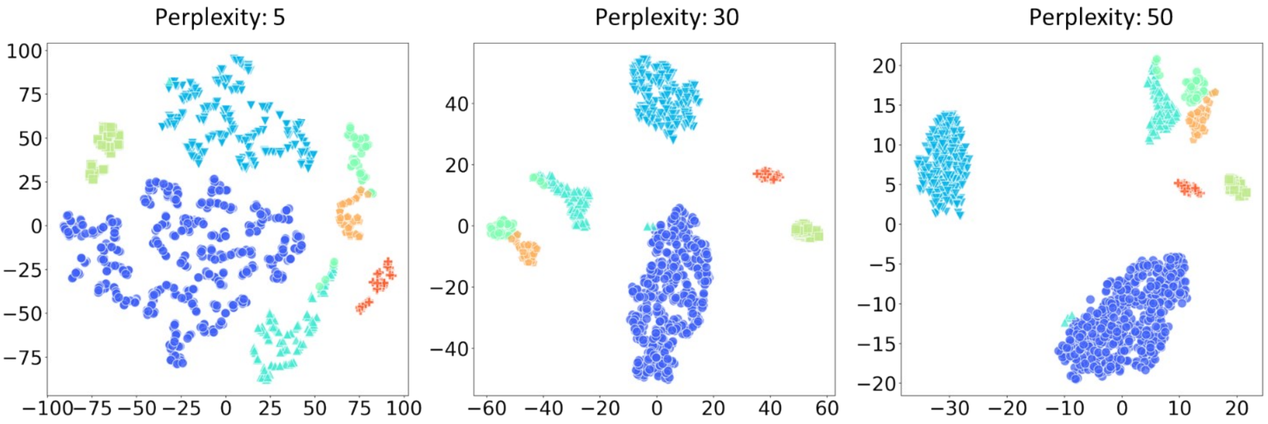
	Characterization technique	
Sample	**Dynamic light scattering (DLS) Hydrodynamic diameter (nm)**	**Scanning electron microscopy Feret diameter (nm)**
S1	358 ± 27	308 ± 35
S2	359 ± 19	310 ± 4
S3	327 ± 12	354 ± 24

SI Table 2. Elemental limits of detection for liquid blank LA-spICP-ToF-MS method applied on the three road sediment samples.

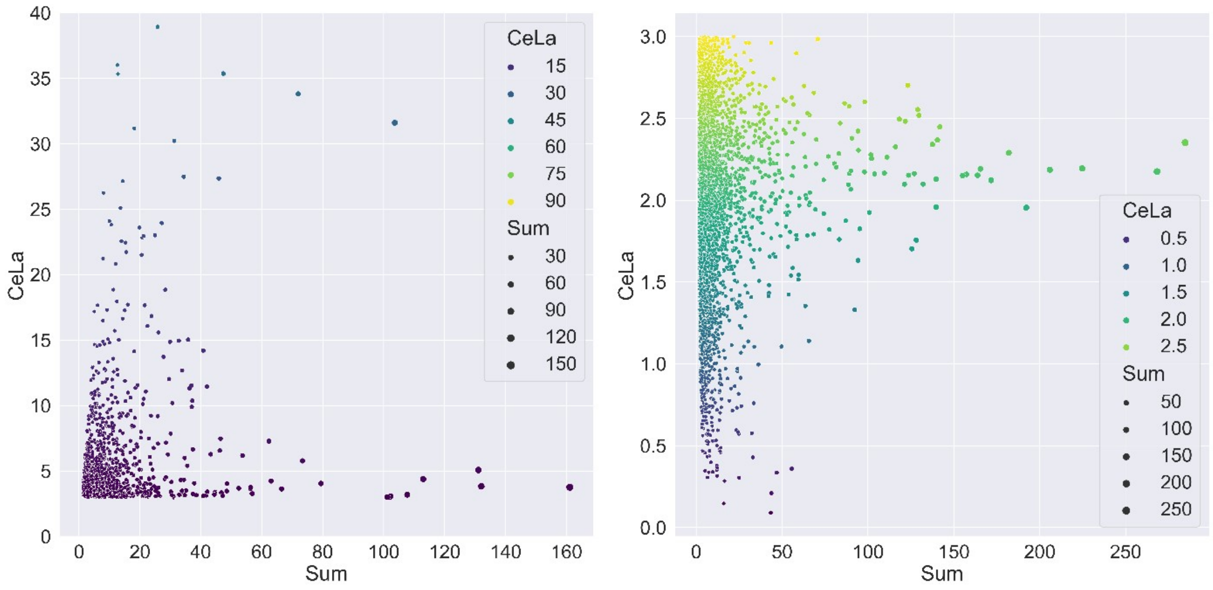
element	LOD (fg)	element	LOD (fg)
[63Cu]+	0.41	[139La]+	0.03
[64Zn]+	0.68	[140Ce]+	0.04
[88Sr]+	0.12	[141Pr]+	0.02
[89Y]+	0.07	[144Nd]+	0.08
[90Zr]+	0.13	[152Sm]+	0.07
[93Nb]+	0.08	[153Eu]+	0.08
[103Rh]+	0.05	[157Gd]+	0.09
[104Ru]+	0.17	[163Dy]+	0.07
[105Pd]+	0.29	[175Lu]+	0.02
[118Sn]+	0.32	[178Hf]+	0.07
[121Sb]+	0.13	[195Pt]+	0.10
[138Ba]+	0.06	[208Pb]+	0.05



SI Figure 4. Pearson correlation heatmaps for S2 (A) and S3 (B).



SI Figure 5. Parametrization comparison for the perplexity argument for the sedimentation basin using a t-SNE dimensionality reduction.



SI Figure 6. Scatter plot comparison of CeLa ratio for particle events for class 1 and 2 particles. CeLa >3 (Left), CeLa <3 (Right).