https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


# Airborne Microplastics in the $PM_{2.5}$ Fraction: Empirically Calibrated µFTIR-ATR Imaging Detects Particles Down to 2 µm

Yasuhiro Niida[1], Hiroshi Okochi[2*], Norihisa Yoshida[2], Yuto Tani[2], Shunki Sasai[2], Yusuke Fujii[3], Norimichi Takenaka[3]

[1]PerkinElmer Japan G.K., 2F Aquaria Tower, 1-1-32 Shin-urashima-cho, Kanagawa-ku, Yokohama 221-0031, Japan
[2]School of Creative Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-0072, Japan
[3]Graduate School of Sustainable System Sciences, Osaka Metropolitan University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan

Correspondence to: Hiroshi Okochi (hokochi@waseda.jp)

**Abstract.**

Reliable identification of airborne microplastics (AMPs) in the $PM_{2.5}$ fraction remains challenging. Conventional µFTIR imaging studies of atmospheric aerosol have generally targeted larger particles, and the reliability of library-matching scores at this smaller size range has not been systematically examined. We present a µFTIR-ATR imaging workflow that separates spectral processing for chemical identification from image processing for physical characterisation. For chemical identification, particle-level spectra were smoothed and evaluated through library matching combined with diagnostic-band interpretation. For physical characterisation, PCA-based image processing and blank-derived intensity thresholds defined particle boundaries for Feret-dimension measurement and morphological classification. We defined the TopHit ratio as the proportion of top-ranked library assignments agreeing with diagnostic-band interpretation within each maximum-correlation-coefficient ($R_{max}$) interval. Among 674 particles, this ratio rose from 41.1% at $R_{max}$ = 0.4–0.5 to 96.0% at $R_{max}$ = 0.8–0.9 and 100.0% at $R_{max}$ = 0.9–1.0; a Richards growth model best described this relationship across four tested library/bin-width combinations. The workflow also enabled identification and measurement of an airborne polyethylene terephthalate particle collected in the $PM_{2.5}$ fraction, with an apparent minimum Feret dimension of 2.3 µm. To our knowledge, this is the first study to combine µFTIR-based polymer identification of an airborne $PM_{2.5}$ particle approaching 2 µm with explicit empirical calibration of spectral-matching confidence.

## 1 Introduction

AMPs are increasingly recognised as a component of atmospheric particulate matter, with evidence of long-range transport, deposition, and inhalation exposure (Dris et al., 2016; Allen et al., 2019; Zhang et al., 2020; Brahney et al., 2020). AMPs in the $PM_{2.5}$ fraction are of particular interest because this aerodynamic fraction overlaps with respirable particles and can remain suspended for extended periods (Allen et al., 2022). Microplastics have also been reported in remote environments and

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


biological systems, including human lung tissue and wild birds, supporting the environmental relevance of inhalation exposure (Amato-Lourenço et al., 2021; Jenner et al., 2022; Tokunaga et al., 2023).

Despite rapid progress in this field, reliable particle-resolved identification of AMPs with a physical size below 10 µm — particularly those collected in the $PM_{2.5}$ fraction — remains difficult. Conventional forcal plane array (FPA)-µFTIR imaging has identified individual airborne particles down to approximately 11 µm (Vianello et al., 2019), whereas airborne plastic particles below this range have primarily been quantified using thermal analysis, which does not provide particle-resolved polymer identity, size, and morphology (Morioka et al., 2024). Reliability comparisons of automated µFTIR analysis have been reported for 11–500 µm particles in environmental water samples (Moses et al., 2023), but comparable evaluation has not been established for individual airborne $PM_{2.5}$ particles approaching the 2 µm scale. In this size range, chemical identification and physical characterisation impose different preprocessing requirements. Polymer identification requires preservation of weak diagnostic peak positions and shapes; principal component analysis (PCA) reconstruction can modify subtle peak profiles at low signal-to-noise ratios and thereby increase the risk of library-search misclassification when reconstructed spectra are used for identification (Koziol et al., 2018). In contrast, PCA-based noise reduction is useful for delineating particle features in imaging data before size and morphology measurements. These two analytical functions should therefore be separated. In addition, correlation-based matching scores are commonly treated as binary thresholds despite their dependence on particle and measurement characteristics (Primpke et al., 2018; Cowger et al., 2021). Complementary vibrational spectroscopic techniques, most notably Raman microspectroscopy, offer additional capabilities and trade-offs for microplastic identification across different particle-size ranges and sample matrices (Käppler et al., 2016; Araujo et al., 2018). An operational framework is needed that preserves diagnostic spectral information for polymer identification, uses PCA specifically for physical mapping, and replaces binary interpretation of spectral similarity with empirical calibration of spectral-matching confidence.

The validation problem differs fundamentally from that encountered for ideal reference particles. For an unknown, weathered environmental particle, the original polymer composition cannot be independently verified after collection. Commercially available standards below 10 µm are dominated by spherical, unweathered particles and do not reproduce the irregular morphology, heterogeneous weathering, ATR contact, or matrix interactions of airborne fragments. Microplastics should therefore be regarded as a diverse contaminant suite rather than as a single material class (Rochman et al., 2019). Known standards remain valuable for instrument checks, but they do not establish the original composition of the environmental particles analysed here. Under these constraints, agreement between complementary analytical interpretations of the same environmental spectrum provides a practical and empirically testable measure of identification consistency. Systematically defining this agreement creates a common validation strategy for authentic weathered particles that can be reproduced and compared across analytical systems.

In this study, we present an empirically calibrated micro-Fourier transform infrared attenuated total reflection (µFTIR-ATR) imaging workflow (Niida et al., 2026; Wang et al., 2023) for particle-resolved identification and physical characterisation of airborne microplastics below 10 µm, including particles collected in the $PM_{2.5}$ fraction. The workflow comprises two

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


deliberately separated processing pathways. In the chemical-identification pathway, particle-level spectra are processed by Savitzky–Golay (SG) smoothing (Savitzky and Golay, 1964), followed by spectral library matching and diagnostic-band interpretation. The maximum correlation coefficient ($R_{max}$) is treated as a graded spectral-similarity metric and related to the TopHit ratio, defined as the proportion of top-ranked library assignments agreeing with diagnostic-band interpretation. In the physical-characterisation pathway, PCA-based noise reduction is applied to imaging data after chemical identification, and blank-derived 5σ mapping thresholds are used exclusively to delineate particle maps for Feret diameter measurement and morphological classification. The 5σ mapping threshold is not used for polymer identification or TopHit calibration. The objectives were to demonstrate µFTIR-based polymer identification of an airborne $PM_{2.5}$ particle approaching the 2 µm scale, establish the complementary SG-smoothed and PCA-based processing pathways, empirically calibrate the relationship between $R_{max}$ and spectral-matching confidence, test the descriptive robustness of that relationship to library composition and bin width, and provide a basis for reliable quantification of representative airborne microplastic concentrations in each study region. Relative to our related prior work applying this µFTIR-ATR platform to sub-10 µm airborne microplastics (Niida et al., 2026), the specific advance introduced here is the replacement of a fixed spectral-similarity criterion with the graded, empirically calibrated $R_{max}$–TopHit framework developed in the present study.

## 2 Experimental

### 2.1 Sample collection and pretreatment

Atmospheric aerosol samples were collected in 2021 and 2022 at two urban sites in Japan: the Nishi-Waseda Campus of Waseda University in Shinjuku, Tokyo, and the Nakamozu Campus of Osaka Metropolitan University in Sakai, Osaka. Samples were collected in successive 1-week periods and size-fractionated using a multi-nozzle cascade impactor (MCI; Tokyo Dylec Corp., Japan) operated at 20 L $min^{-1}$. The sampler separates particles according to aerodynamic diameter ($d_a$), with nominal 50 % cut-off diameters ($d_{50}$) of 10 and 2.5 µm. Airborne particles were therefore classified into three fractions: $PM_{10+}$ ($d_a$ > 10 µm), $PM_{10–2.5}$ (10 ≥ $d_a$ > 2.5 µm), and PM2.5 (2.5 µm ≥ $d_a$). Particles in each fraction were collected on separate fluoropolymer-binder glass-fibre filters (TX40HI20-WW, Pallflex) and processed independently.

According to the manufacturer's specifications, the MCI $PM_{2.5}$ stage provides the same aerodynamic size-classification performance as an FRM $PM_{2.5}$ sampler (Tokyo Dylec Corp., 2020), supporting comparability with established regulatory $PM_{2.5}$ size selection. Compared with impactors having more stages, the three-stage configuration distinguishes fine, coarse, and giant particles while reducing analytical workload, filter handling, and contamination risk. It therefore balances aerodynamic size resolution with analytical feasibility for routine and interregional monitoring. The representative PET particle described in Sect. 3.2 was recovered from the $PM_{2.5}$ fraction. The sampling site and sampler configuration followed Niida et al. (2026).

Collected particles were extracted with ultrapure water and subjected to filtration through a 0.45 µm hydrophilic polytetrafluoroethylene membrane to remove soluble components. The retained particulate matter was digested using 30 % $H_2O_2$ to remove organic material. Two successive density-separation cycles were then performed using sodium iodide solution

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


(1.5 g $cm^{-3}$) with centrifugation. After the first separation, the supernatant was recovered, and the lower residue was subjected to a second density-separation cycle under the same conditions. The supernatants recovered from the two cycles were combined and filtered onto a 0.2 µm alumina membrane, which was subsequently dried at room temperature prior to spectroscopic analysis.

The pretreatment protocol was designed to retain micron-scale particles while removing soluble, biological, and mineral components. All procedures were conducted within a clean bench. A cotton laboratory coat and cellulose-based gloves were worn throughout the procedure. Laboratory procedural and reagent blanks were analysed to assess contamination; after improvements to reagent filtration and sample handling, no polymer particles were detected in procedural blanks under the present analytical conditions (Wang et al., 2023).

### 2.2 Imaging measurement

Samples were analysed by µFTIR-ATR imaging using a Spotlight 400 FTIR imaging system (PerkinElmer) equipped with a germanium ATR crystal and a 16-element linear mercury cadmium telluride detector. Spectra were acquired from 4000 to 695 $cm^{-1}$ at a spectral resolution of 8 $cm^{-1}$ and a pixel size of 1.56 µm. A germanium ATR crystal was used to ensure high spatial resolution and sufficient signal intensity for micron-scale particles. Five systematically positioned 750 × 750 µm imaging areas were measured on each 4-mm-diameter filtration spot: the centre and four positions midway between the centre and perimeter along orthogonal axes. The five areas represented approximately 22.4 % of the filtration area (Wang et al., 2023).

The measurement strategy generated more than 200,000 spectra per imaging area. Sample acquisition, atmospheric correction, C–H/C=O band-based screening, particle-level spectrum extraction and averaging, SG smoothing, and PCA-based noise reduction were performed using Spectrum Image 1.11. The subsequent analytical sequence deliberately separated Savitzky–Golay-based spectral processing for polymer identification and degradation assessment from PCA-based image processing for particle mapping, Feret-dimension measurement, and morphological classification, as described in Sect. 2.3.

### 2.3 Polymer identification, confidence calibration, and morphological analysis

Following atmospheric correction, the imaging data were first screened using two diagnostic spectral regions: carbon–hydrogen stretching (2700–3000 $cm^{-1}$) and carbonyl stretching (1710–1740 $cm^{-1}$). Absorbance features in either region were used to locate and extract candidate microplastic particles from the imaging data. The complementary use of these two spectral regions for initial candidate extraction is a key methodological innovation of the present workflow because it broadens candidate recovery across polymers with different diagnostic absorption characteristics. This initial band-based screening was used only to locate candidate particles.

For chemical identification, spectra associated with each candidate particle were averaged at the particle level and processed using SG smoothing with a five-point window. SG smoothing was selected to reduce high-frequency spectral noise while retaining the positions and shapes of diagnostic absorption bands (Koziol et al., 2018). The resulting spectra were used for polymer identification and, where applicable, degradation assessment following the spectral-index approach described by

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


Niida et al. (2026). Degradation indices were not used to define the TopHit ratio in the present study. Library searching was performed using Spectrum IR 10.7 (PerkinElmer), and polymer identification was based on the $R_{max}$ between measured and reference spectra together with diagnostic-band interpretation.

An in-house targeted reference library was first constructed by acquiring ATR spectra of approximately 100 compositionally known polymer and non-polymer standards relevant to microplastic analysis, including representatives of the material classes listed in Tables 1 and 2. This targeted library was selected for the primary empirical calibration because the reference materials were known and measured using the same ATR configuration as the environmental particles. The library was subsequently expanded by adding further in-house reference spectra and spectra from the PerkinElmer standard spectral library, resulting in a comprehensive library of approximately 10,000 spectra. Both libraries were used for final particle assignment. Unlike conventional approaches that apply a fixed binary threshold, $R_{max}$ was interpreted as a graded spectral-similarity metric and calibrated against the observed agreement between the top-ranked library assignment and diagnostic-band interpretation. It therefore provides an empirical indicator of spectral-matching confidence rather than an absolute probability of polymer identity.

To evaluate the operational consistency of library-based identification, assignments obtained from the SG-smoothed spectra were compared with diagnostic-band interpretations of the same environmental spectra. Diagnostic-band interpretation was performed by a single analyst with specialised expertise in FTIR spectral interpretation and µFTIR analysis who was blinded to the library-search rank and $R_{max}$ value. Within each $R_{max}$ interval, the TopHit ratio was calculated as the proportion of highest-ranked library assignments that agreed with the diagnostic-band interpretations. The two approaches interrogate the same spectrum but apply different decision rules: the library search ranks whole-spectrum correlations, whereas the diagnostic-band interpretation evaluates polymer-specific positive bands and non-polymer exclusion features. The TopHit ratio therefore quantifies agreement between complementary analytical interpretations under controlled conditions rather than an absolute probability established from independently verified environmental-particle identity. Neither PCA-reconstructed imaging data nor the 5σ mapping threshold was used to determine polymer identity or calculate the TopHit ratio. The instrument, SG smoothing preprocessing procedure, libraries, sample matrix, and interpretation criteria were explicitly specified so that the calibration can be reproduced, evaluated, and compared with corresponding calibrations obtained using other analytical systems.

After polymer identification and, where applicable, degradation assessment had been completed using the SG-smoothed spectra, PCA-based noise reduction was applied separately to the imaging data used for particle mapping and physical characterisation. Particle maps were generated using the same carbon–hydrogen and carbonyl regions employed in the initial candidate screening. For each region, an operational mapping threshold was set at five times the standard deviation (5σ) of the corresponding blank measurements to suppress pixels and define particle boundaries more clearly. The 5σ mapping threshold was used exclusively to determine whether an image feature could be delineated for physical measurement; it was not used to accept or reject a polymer assignment and did not contribute to the TopHit ratio.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


Spatially contiguous pixels exceeding the mapping threshold were grouped into individual particle maps. The resulting maps were analysed using ImageJ 1.54g (National Institutes of Health, Bethesda, MD, USA), from which the minimum and maximum Feret dimensions were measured. The aspect ratio was calculated as the maximum Feret dimension divided by the minimum Feret dimension. Particles were operationally classified as granules (aspect ratio < 1.2), fragments (1.2 ≤ aspect ratio < 3.0), or fibres (aspect ratio ≥ 3.0). Particles for which the Feret dimensions could not be reliably determined were recorded as unmeasurable. Image features below the mapping thresholds were recorded as "below LOD" and were not assigned a morphological class. Here, "below LOD" denotes failure to meet the study-specific physical-mapping threshold rather than failure of polymer identification or a universal instrumental limit of detection. For the representative particle, $PM_{2.5}$ denotes the aerodynamic collection fraction, whereas 2.3 µm is the apparent minimum Feret dimension; the latter demonstrates measurement at the approximately 2 µm scale but is not a validated method-wide sizing limit.

Additional experimental and analytical details are provided in Appendix A.

## 3 Results and discussion

### 3.1 Integrated analytical workflow

Figure 1 summarises the sequential but functionally distinct stages of candidate screening, chemical identification, and physical characterisation. Detailed procedures are provided in Sects. 2.2 and 2.3 and Appendix A. First, following atmospheric correction, the imaging data were screened using the C–H stretching region (2700–3000 $cm^{-1}$) and the C=O stretching region (1710–1740 $cm^{-1}$) to locate and extract candidate microplastic particles. The complementary use of these two spectral regions for candidate extraction is a key methodological innovation of the present workflow. Second, spectra associated with each candidate particle were averaged and processed using SG smoothing to reduce noise while preserving diagnostic peak shapes. The processed spectra were then evaluated by $R_{max}$-based library matching and diagnostic-band interpretation for polymer identification and empirical calibration of spectral-matching confidence. PCA-reconstructed spectra were not used for polymer identification. Third, after chemical identification and, where applicable, degradation assessment, PCA-based noise reduction was applied separately to the imaging data used for physical characterisation. The mapping thresholds in the same C–H and C=O regions were then used to suppress pixels and delineate measurable particle boundaries. The minimum and maximum Feret dimensions were measured, and the particles were operationally classified as fibres, fragments, or granules. Thus, dual-band screening enables the initial extraction of candidate particles. SG smoothing and library matching support polymer identification, while $R_{max}$–TopHit calibration provides an empirical measure of spectral-matching confidence. PCA-based noise reduction and 5σ particle mapping then enable Feret measurement and morphological classification.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


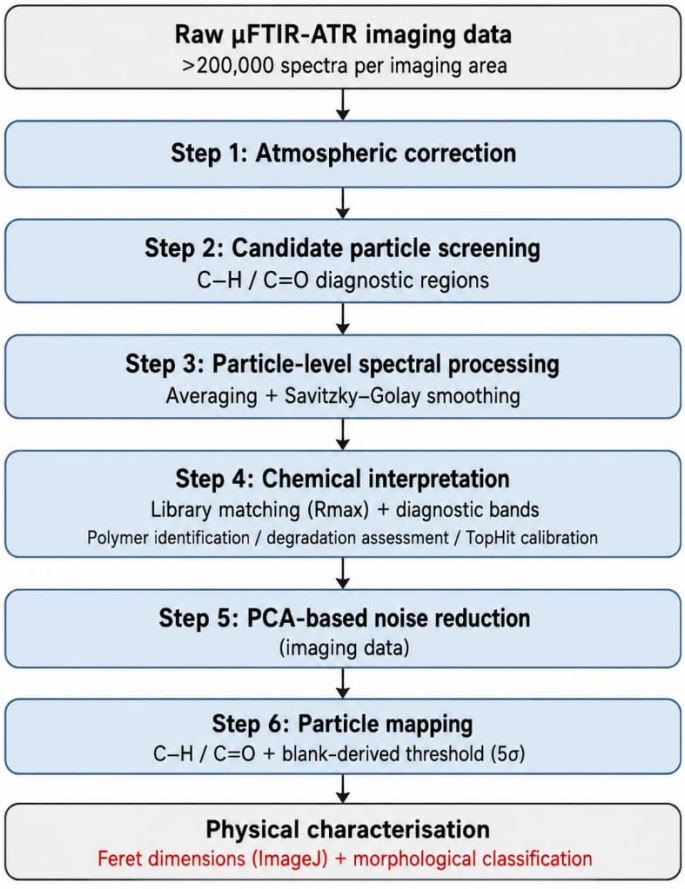


**Figure 1.** Integrated µFTIR-ATR workflow with separate processing pathways for chemical identification and physical characterisation of micron-scale airborne particles. Following atmospheric correction, candidate particles are first screened using the C–H and C=O diagnostic regions. Spectra from each candidate particle are then averaged and processed using SG smoothing, followed by library matching, diagnostic-band interpretation, polymer identification, degradation assessment where applicable, and empirical $R_{max}$–TopHit calibration. PCA-based noise reduction is subsequently applied only to imaging data used for particle mapping. The mapping thresholds are then used to delineate particle boundaries for Feret-dimension measurement and operational morphological classification.

The chemical-identification stage incorporated complementary positive and negative spectral criteria. Table 1 summarises the principal diagnostic bands, weathering-associated features, and major sources of spectral confusion for 12 polymer classes. Table 2 provides complementary exclusion criteria for common non-polymer materials encountered in atmospheric aerosol samples. The combined use of these criteria is important because environmental spectra may contain weathering-related carbonyl and hydroxyl bands, mineral signals, biological material, additives, and mixed-pixel contributions. A polymer

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


assignment was accepted only when the library result was consistent with the overall spectral shape and the diagnostic features in Table 1, and when incompatible non-polymer features listed in Table 2 were absent.

**Table 1.** Polymer-specific diagnostic spectral features used for integrated interpretation of µFTIR-ATR spectra. Key identification bands, weathering-associated features, and major sources of spectral confusion are listed for each polymer class. Band assignments were based on in-house reference spectra and standard infrared spectroscopy references (Colthup et al., 1990).

| Polymer | Key features ($cm^{-1}$) | Weathering features ($cm^{-1}$) | Confusion risk |
|---|---|---|---|
| Polyethylene (PE) | 2915, 2850, 1465 ($CH_2$) | 3200 (O–H), 1725 (C=O), 1173 (C–O) | oils, waxes, oxidized PE |
| Polypropylene (PP) | 2955, 1375 ($CH_3$), 2920 ($CH_2$) | 3420 (O–H), 1730 (C=O), 1130 (C–O) | oils, elastomers |
| Polystyrene (PS) | 3060, 1495 (aromatic ring C–H), 695 (phenyl) | 3230 (O–H), 1715 (C=O), 1205 (C–O) | ABS, styrene copolymers |
| Polyvinyl chloride (PVC) | 2910, 1425 ($CH_2$), 700 (C–Cl) | 3200 (O–H), 1715 (C=O) | plasticizers |
| Polymethyl methacrylate (PMMA) | 2990–2950 ($CH_3$), 1725 (C=O), 1145 (C–O) | 3200 (O–H) | esters, adhesives |
| Polyethylene terephthalate (PET) | 1715 (C=O), 1240, 1095 (C–O), 725 (p-phenylene) | 3420 (O–H), 1715 (broadened C=O), 1215 (C–O) | polyesters, plasticizers |
| Polyamide 6 (PA6) | 3300 (N–H), 1635 (amide I), 1540 (amide II), 1265 (amide III) | 3200 (O–H), 1775, 1710 (C=O), 1090 (C–O) | proteins |
| Polyacrylonitrile (PAN) | 2240 (C≡N), 1450 ($CH_2$), 1360 | 3250 (O–H), 1715 (C=O) 1200 (C–O) | NBR, ABS |
| Polyurethane (PU) | 3290 (N–H), 1720 (urethane C=O), 1530 (amide II), 1090 (C–O) | 3245 (O–H), 1700 (C=O) | polyesters, PA |
| Polycarbonate (PC) | 1770 (carbonate C=O), 1210 (aryl–O-C), 1190 (p-phenylene C-H), 1160 (carbonate C–O) | 3230 (O–H), 1715 (C=O) | PC alloys |
| Silicone (SI) | 2950 ($CH_3$), 1260 (Si–$CH_3$), 1010 (Si–O), 785 (Si–$CH_3$) | 3380 (O–H), 1625 (C=O) | PE overlap, silica |
| Ethylene-vinyl acetate (EVA) | 2915 ($CH_2$), 2850, 1740 (C=O), 1240 (C–O), 1020 | 3200 (O–H), 1715 (shifted/broadened C=O) | degraded PE, stearic acid |

Table 2 provides complementary exclusion criteria for biological materials, minerals, and additives that may produce spectral features overlapping those of polymers.

**Table 2. Spectral features used to distinguish common non-polymer materials from candidate microplastics.**

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


| Material | Key features ($cm^{-1}$) | Confusion risk |
|---|---|---|
| Cellulose | 3300 (O–H), 1160, 1050–1030 (C–O) | silica, oxidized polymers |
| Proteins | 1650 (amide I), 1540 (amide II), 3300 (N–H; broader than PA) | PA |
| Silica/glass | ~1080, 800 (Si–O) | cellulose, sulfates |
| Hydroxides | 3700–3300 (O–H) | water, cellulose |
| Silicates (talc/kaolin) | 3700–3600 (O–H), ~1000 (Si–O) | silica |
| Carbonates | 1400, 870, 710 ($CO_3^{2-}$) | polymers with fillers |
| Sulfates | 1100, 600 ($SO_4^{2-}$) | silica |
| Rocks/sand | 1100–900, 800–700 (Si–O) | mixtures, cellulose |
| Stearic acid | 2915, 1465, 1360–1160 ($CH_2$), 2950 ($CH_3$), 1700 (carboxylic-acid C=O) | polyolefins, waxes, esters |
| Metal stearates (Zn/Ca) | 2915, 1465, 1360–1160 ($CH_2$), 2950 ($CH_3$), 1535, 1395 ($COO^-$, $Zn^{2+}$), 1575, 1540, 1435, 1420 ($COO^-$, $Ca^{2+}$) | polyolefins, carbonates, additives |

The mapping threshold was evaluated using a 66 × 66-pixel region of interest (4356 pixels) extracted from clean alumina-filter images for each diagnostic region (Fig. 2). The corresponding mapping thresholds were $2.945 \times 10^{-3}$ and $1.115 \times 10^{-3}$ for the carbon–hydrogen and carbonyl regions, respectively. These values were applied to PCA-noise-reduced imaging data as conservative operational mapping thresholds for suppressing pixels before Feret-dimension measurement and morphological classification. They were not applied to the SG-smoothed spectra used for polymer identification and did not affect $R_{max}$ or the TopHit ratio. Pixel-level evaluation showed that 70 and 9 pixels exceeded the 3σ and 4σ thresholds, respectively, in the carbon–hydrogen region, whereas no pixel exceeded the 5σ mapping threshold. In the carbonyl region, 13 pixels exceeded 3σ and 2 pixels exceeded 4σ, whereas no pixel exceeded 5σ. These observations indicate that the 5σ mapping threshold suppressed mapped pixels more effectively than the 3σ or 4σ thresholds under the present analytical conditions.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


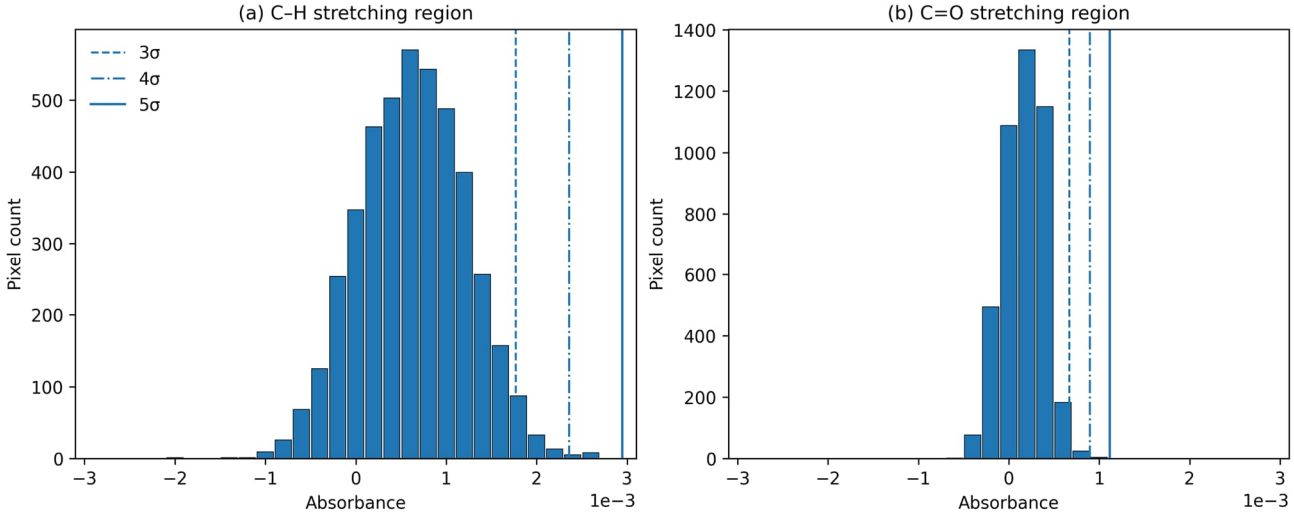


**Figure 2.** Blank absorbance distributions and operational mapping thresholds for µFTIR-ATR imaging. Distributions obtained from 4356-pixel regions of interest on clean alumina-filter images are shown for the C–H stretching region and the C=O stretching region. Lines indicate the 3σ, 4σ, and 5σ mapping thresholds derived from the corresponding blank distributions. These mapping thresholds were evaluated for PCA-based particle mapping and physical characterisation, not for polymer identification.

The 5σ mapping thresholds were subsequently applied to PCA-noise-reduced imaging data from one atmospheric aerosol imaging area (Fig. 3). Mapping of more than 200,000-pixel spectra yielded approximately 50 particle-like image features that could be delineated for subsequent Feret-dimension measurement and morphological classification. Threshold exceedance served only to define mapped image features. Polymer identity was determined independently from the SG-smoothed particle spectrum using library matching and diagnostic-band interpretation.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


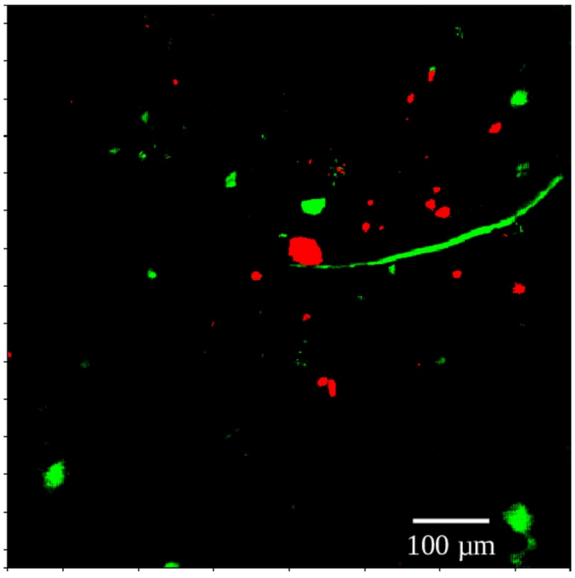


**Figure 3.** Band-based delineation of particle features for physical characterisation in an ambient atmospheric aerosol sample collected in Tokyo in August 2021. After PCA-based noise reduction, pixels exceeding the mapping threshold in the C–H stretching region are shown in red, and those exceeding the corresponding threshold in the C=O stretching region are shown in green. Approximately 50 particle-like image features were delineated. The thresholds were used for particle mapping and not for polymer identification. The scale bar represents 100 µm.

The 5σ mapping threshold was used exclusively as a conservative operational mapping threshold to suppress pixels before particle-size and morphology measurements. It was not a spectral-similarity cutoff and was not used to determine whether a spectrum represented a polymer. Particles with intrinsically weak C–H and C=O absorption, or particles occupying only a small fraction of an imaging pixel, may remain below the mapping threshold, reflecting inherent limitations in the spectral sensitivity and spatial resolution of µFTIR-ATR imaging. Insufficient ATR contact can also reduce image contrast; however, this effect was minimised through standardised crystal-contact procedures performed by an experienced analyst. The 5σ value should therefore be regarded as an operational lower limit for physical particle mapping under the present µFTIR-ATR imaging conditions, rather than as a universal detection or sizing limit applicable to every polymer type and particle morphology. Because the value is calculated from the variability of the corresponding blank measurements, its absolute magnitude may vary among measurement batches. These mapping limitations are distinct from polymer identification, which was performed using SG-smoothed spectra, library matching, and diagnostic-band interpretation.

### 3.2 Representative identification of a $PM_{2.5}$ PET particle approaching 2 µm

Figure 4 presents a representative particle-level assignment obtained from the $PM_{2.5}$ fraction of a Tokyo sample collected in August 2021. The polymer assignment was made from the particle-level spectrum after SG smoothing, whereas the particle

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


map was generated separately using PCA-based noise reduction and the 5σ mapping threshold. Analysis of this map gave an apparent minimum Feret dimension of approximately 2.3 µm and an apparent maximum dimension of approximately 4.0 µm. The minimum dimension corresponds to approximately 1.5 imaging pixels and is therefore reported as an apparent image-derived dimension approaching the 2 µm scale rather than as a validated spatial resolution or method-wide sizing limit. Nevertheless, the independently processed particle spectrum contained multiple PET-specific diagnostic bands.

The measured spectrum was compared with a polyethylene terephthalate (PET) reference spectrum from the spectral library (Fig. 4b). PET was the top-ranked match in the targeted library ($R_{max}$ = 0.92) and the second-ranked match in the comprehensive library. The principal agreement occurred at the ester carbonyl band near 1715–1716 $cm^{-1}$, the C–O stretching region near 1240 and 1095 $cm^{-1}$, and the p-substituted aromatic-ring band near 725 $cm^{-1}$. Differences in relative intensity and baseline shape were present, as expected for a micron-scale environmental particle, but the combined diagnostic-band pattern was consistent with PET. The assignment was therefore supported by agreement between the library result and chemically diagnostic features rather than by the correlation score alone. The lowest interpretable particle size depends on particle morphology, thickness, ATR contact, spectral noise, and interference from adjacent materials.

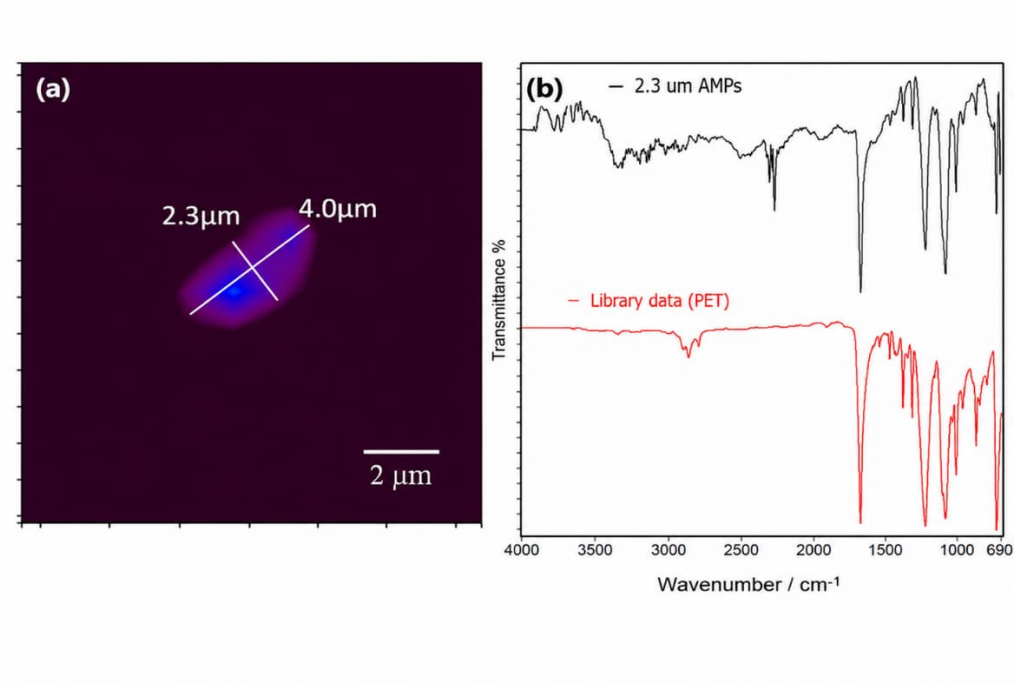


**Figure 4.** Representative assignment and physical measurement of an airborne PET particle collected in the $PM_{2.5}$ fraction in Tokyo in August 2021. (a) PCA-noise-reduced particle map delineated using the 5σ mapping threshold. The particle map gave an apparent minimum Feret dimension of approximately 2.3 µm and an apparent maximum dimension of approximately 4.0 µm. (b) SG-smoothed particle spectrum (black) and PET library spectrum (red). PET was the top-ranked match in the targeted

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


library ($R_{max}$ = 0.92) and the second-ranked match in the comprehensive library. Agreement in the carbonyl, C–O stretching, and aromatic-ring regions supports the PET assignment. The 5σ mapping threshold was not used for polymer identification.

This example demonstrates chemical identification and physical measurement of an airborne $PM_{2.5}$ particle at the approximately 2 µm scale. Conventional FPA-µFTIR imaging studies have generally reported individual airborne particles at approximately 10–11 µm or larger (Vianello et al., 2019; Zhang et al., 2020), whereas the present germanium ATR configuration supported assignment at an apparent minimum Feret dimension of 2.3 µm. $PM_{2.5}$ denotes the aerodynamic collection fraction, whereas 2.3 µm is an image-derived dimension. This result demonstrates detection approaching 2 µm but not a method-wide limit, which varies with particle properties, ATR contact, spectral noise, matrix interference, blank variability, and pixel occupancy.

The representative PET particle illustrates the central analytical design of the workflow. SG smoothing preserved the diagnostic peak pattern used for polymer assignment, while PCA-based noise reduction and the mapping threshold were used separately to delineate the particle and determine its Feret dimensions. Its assignment was supported by the concurrence of an $R_{max}$ value of 0.92 in the targeted library with the ester carbonyl, C–O stretching, and aromatic-ring features characteristic of PET. For $PM_{2.5}$ particles approaching the 2 µm scale, a fixed $R_{max}$ threshold may either reject chemically consistent weathered particles or accept inconsistent matches. The $R_{max}$–TopHit calibration presented below addresses this limitation by replacing binary interpretation with a graded empirical indicator of spectral-matching confidence based on agreement between complementary spectral interpretations.

### 3.3 Empirical calibration of spectral-matching confidence

The internal consistency of the library result was evaluated using the TopHit ratio, defined as the proportion of particles for which the highest-ranked library assignment agreed with diagnostic-band interpretation within each $R_{max}$ interval. The TopHit ratio provides an empirical, agreement-based indicator of spectral-matching confidence for authentic weathered-particle spectra under explicitly defined analytical conditions. It complements, but does not replace, validation with compositionally known reference materials. The calibration dataset comprised 674 particles pooled across all analysed aerodynamic fractions; the $PM_{2.5}$ PET particle presented in Sect. 3.2 is a representative application of the calibrated workflow rather than the sole basis of the calibration. Only candidates with $R_{max} \geq 0.4$ were included because spectra below this operational lower bound could not be interpreted with sufficient confidence at the polymer level, even after diagnostic-band inspection. $R_{max}$ quantifies similarity between environmental-particle spectra and reference spectra; environmental weathering can lower $R_{max}$ by altering spectral shape and relative band intensity even when polymer-specific diagnostic features remain detectable. The TopHit analysis therefore quantifies how frequently the top-ranked library assignment agrees with a complementary interpretation across the observed $R_{max}$ range. Because the analytical configuration and decision criteria were specified, the same procedure can be repeated with different instruments, preprocessing methods, libraries, sample matrices, and analysts to generate directly

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


comparable system-specific calibrations. Wilson 95 % confidence intervals were calculated for each binned TopHit ratio to represent binomial uncertainty. Reference spectra searched against the targeted library yielded a reference-control endpoint at $R_{max}$ = 1.0 and TopHit ratio = 100 %; this endpoint was retained as a boundary anchor in the descriptive curve fitting, and the upper asymptote of the Richards model (Richards, 1959) was fixed at 100 %. Figure 5 shows the relationship obtained using the targeted in-house reference library, which contained approximately 100 spectra acquired from standard polymer and non-polymer materials. The observed TopHit ratio increased non-linearly with $R_{max}$. For $R_{max}$ = 0.4–0.5, 30 of 73 assignments agreed with diagnostic-band interpretation, corresponding to a TopHit ratio of 0.411 and a Wilson 95 % confidence interval of 0.305–0.526. The ratios were 0.656 for $R_{max}$ = 0.5–0.6, 0.679 for 0.6–0.7, 0.842 for 0.7–0.8, and 0.960 for 0.8–0.9. At $R_{max}$ = 0.9–1.0, all 109 evaluated assignments agreed, giving a TopHit ratio of 1.000 with a Wilson 95 % confidence interval of 0.966–1.000. These directly observed proportions and confidence intervals demonstrate that spectral similarity contains useful graded information about empirical spectral-matching confidence, as defined by agreement between complementary spectral interpretations. The pooled counts underlying Fig. 5 are provided in Table A1.

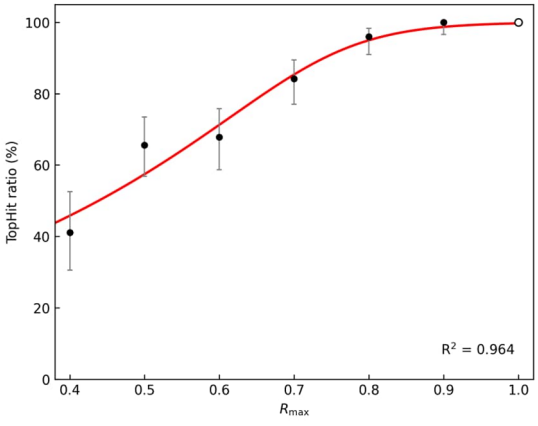


**Figure 5. Relationship between the maximum correlation coefficient ($R_{max}$) and the TopHit ratio for the pooled polymer dataset.** Black circles show environmental-particle observations, grey error bars indicate Wilson 95 % confidence intervals, and the open circle at $R_{max}$ = 1.0 denotes the reference-control endpoint. The red curve is a Richards-type fit in which the reference-control endpoint was retained and the upper asymptote was fixed at 100 %. The coefficient of determination is shown in the panel.

The results support a graded interpretation of $R_{max}$ rather than one universal binary cutoff. Values below approximately 0.6 were associated with low or variable agreement, although some assignments remained chemically consistent. Intermediate $R_{max}$ values required further spectral inspection, whereas values above 0.9 were associated with high agreement in the present dataset. Evidence of mixed spectra, diagnostic-band consistency, and possible non-polymer interference should nevertheless

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


be considered throughout the range. The calibration therefore provides an operational tiered framework: high $R_{max}$ supports higher confidence, whereas intermediate and low $R_{max}$ trigger progressively stronger diagnostic confirmation. This tiered approach offers a transparent reporting structure that can be applied consistently and evaluated across laboratories.

This approach also avoids the systematic underestimation that would result from discarding every candidate below a fixed threshold. Weathering, small particle size, mixed-pixel effects, and imperfect ATR contact can reduce $R_{max}$ even when chemically diagnostic bands remain present. Conversely, a high similarity score can result from a non-polymer material or additive with a partially similar spectral pattern. The TopHit calibration is consequently most useful when combined with the positive and negative interpretation criteria summarized in Tables 1 and 2.

### 3.4 Descriptive calibration curves and robustness assessment

The observed non-linear relationship motivated comparison of alternative descriptive functions. Two-parameter logistic (Berkson, 1944), constrained four-parameter logistic with the upper asymptote fixed at 100 % (De Lean et al., 1978), and Richards growth functions were independently fitted to the same binned TopHit ratios under four reference-library and bin-width conditions: the comprehensive library (approximately 10,000 spectra) and the targeted library (approximately 100 spectra), each evaluated using $\Delta R_{max}$ = 0.02 and 0.1. The purpose of these fits was to visualise the shape and assess the sensitivity of the empirical relationship to library composition, bin width, and functional form; they were not used to estimate a universal particle-level probability or to test a mechanistic model. Fits used unweighted nonlinear least squares so that each $R_{max}$ interval contributed equally to this descriptive shape comparison, while Wilson 95 % confidence intervals were displayed separately to show the binomial uncertainty arising from the number of assignments in each bin. The observed bin proportions and their confidence intervals remain the primary results. The reference-control endpoint at $R_{max}$ = 1.0 and TopHit ratio = 100 % was retained as a boundary anchor in the displayed fits; for the constrained four-parameter logistic and Richards growth models, the upper asymptote was fixed at 100 %. Figure 6 compares the three descriptive functions across the four library and bin-width conditions, and endpoint-exclusion analyses were performed to quantify the influence of the anchor.

Across all four evaluated combinations of library composition and bin width, the Richards model consistently provided the highest $R^2$ and lowest RMSE among the three models examined, indicating the best descriptive fit to the observed relationship. For the comprehensive library, $R^2$ increased from 0.609 for the two-parameter logistic and 0.697 for the constrained four-parameter logistic to 0.723 for the Richards model at $\Delta R_{max}$ = 0.02, while the Richards RMSE decreased to 0.131. At $\Delta R_{max}$ = 0.1, the corresponding $R^2$ values were 0.794, 0.889, and 0.916, and the Richards RMSE was 0.072. For the targeted library, all three models gave $R^2$/RMSE values of approximately 0.887/0.073 at $\Delta R_{max}$ = 0.02, with the Richards model retaining the lowest RMSE; at $\Delta R_{max}$ = 0.1, the corresponding values were 0.959/0.042, 0.960/0.041, and 0.964/0.039. This consistent ranking across both fine and coarse binning and both library sizes supports use of the Richards model as the primary empirical calibration curve. Complexity-penalised criteria provided complementary information. For the comprehensive library at $\Delta R_{max}$ = 0.02, the Akaike information criterion (AIC) marginally favoured the Richards model, whereas the Bayesian information

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


criterion (BIC) favoured the constrained four-parameter logistic model. At $\Delta R_{max}$ = 0.1, the latter two models were effectively indistinguishable (ΔAIC = 0.051; ΔBIC = 0.003). For the targeted library, both AIC and BIC favoured the two-parameter logistic model at both bin widths because these criteria penalise the additional parameters in the more flexible Richards function. Complete statistics are provided in Table A3.

Sensitivity analysis in which the reference-control endpoint was removed showed that the principal targeted-library Richards calibration was not materially dependent on the anchor. For the targeted library at $\Delta R_{max}$ = 0.1, exclusion of the endpoint changed the fitted curve by a maximum of 0.0009 over $R_{max}$ = 0.4–1.0 and yielded $p(R_{max} = 1.0)$ = 0.9968. The targeted-library fits at $\Delta R_{max}$ = 0.02 were similarly stable, with maximum curve differences of 0.0005–0.0013 among the three models. In contrast, endpoint-excluded comprehensive-library two-parameter logistic fits predicted $p(R_{max} = 1.0)$ values of 0.787 and 0.777 for $\Delta R_{max}$ = 0.02 and 0.1, respectively.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


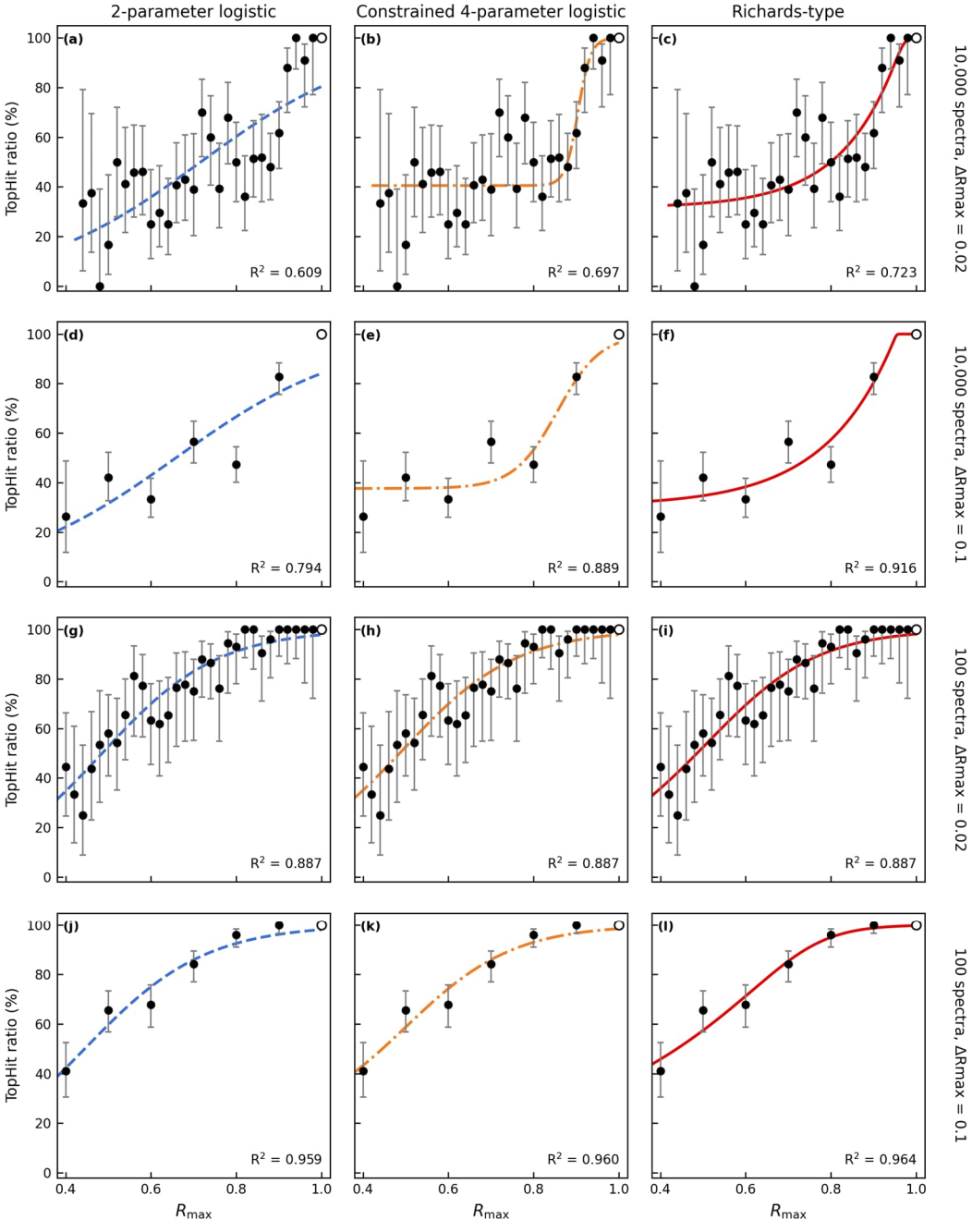


**Figure 6.** Comparison of calibration models under different reference-library and bin-width conditions. Panels (a–c) show the comprehensive library (approximately 10,000 spectra) with $\Delta R_{max}$ = 0.02, (d–f) the comprehensive library with $\Delta R_{max}$ = 0.1, (g–i) the targeted library (approximately 100 spectra) with $\Delta R_{max}$ = 0.02, and (j–l) the targeted library with $\Delta R_{max}$ = 0.1. Within each row, the left, middle, and right panels show the two-parameter logistic, constrained four-parameter logistic, and Richards

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


growth-model fits, respectively. Black circles show environmental-particle observations, grey error bars indicate Wilson 95 % confidence intervals, and open circles at $R_{max}$ = 1.0 denote the reference-control endpoint. The endpoint was retained in all fits; the upper asymptote was fixed at 100 % for the constrained four-parameter logistic and Richards growth models.

Taken together, the model comparisons support selection of the Richards function as the primary empirical calibration curve because it accommodates asymmetry and consistently achieved the highest $R^2$ and lowest RMSE across all four evaluated conditions. The AIC and BIC results show that the magnitude of the improvement must be considered relative to the additional model parameters: for the targeted library, the simpler two-parameter logistic model was preferred after penalising complexity. Thus, the Richards model is identified as the best descriptive representation of the present observations, but not as a unique mechanistic model or a universally optimal function. Its stable performance across library composition, bin width, and endpoint-exclusion analyses provides a practical basis for summarising the graded relationship between $R_{max}$ and the TopHit ratio, while the explicit comparison procedure allows future datasets and laboratories to evaluate the same candidate functions consistently.

### 3.5 Polymer-dependent identification behaviour

The overall analysis combines results for PE, PP, PET, and PS, but polymer-specific exploration revealed distinct relationships between $R_{max}$ and the TopHit ratio (Fig. 7). These four polymer classes were the only classes with sufficient numbers of detected particles for stratified presentation; no particles from the remaining eight classes listed in Table 1 met the $R_{max} \geq 0.4$ inclusion criterion in the present calibration dataset (see Tables A1 and A2). These differences are analytically expected because polymer spectra vary in the number, specificity, and relative intensity of their diagnostic bands, and environmental weathering and spectral interference affect different polymers differently. Richards-type curves were added as descriptive visual guides using the same reference-control endpoint treatment and upper-asymptote constraint as the pooled dataset. Wilson 95 % confidence intervals and the underlying counts, rather than the fitted polymer-specific curves, should be used to judge precision. The polymer-specific counts underlying Fig. 7 are provided in Table A2.

PET showed the clearest graded transition. The TopHit ratio increased from 0.12 at $R_{max}$ = 0.4–0.5 to 0.45, 0.58, 0.79, 0.95, and 1.00 in the successive $R_{max}$ intervals. The presence of the ester carbonyl, C–O, and aromatic-ring bands provides multiple criteria for confirming PET assignments, but small particles, weathering, and matrix effects can reduce the overall correlation score. The PET results therefore illustrate why intermediate $R_{max}$ values cannot be interpreted using a simple pass–fail criterion. PP reached a TopHit ratio of 0.93 at $R_{max}$ = 0.6–0.7 and 1.00 at $R_{max}$ values of 0.7 or greater. Its ratio increased from 0.54 in the lowest interval to 0.84 at 0.5–0.6. In contrast, the PE results were less monotonic at low and intermediate $R_{max}$, with ratios of 0.86, 0.91, 0.67, 0.73, 0.89, and 1.00 across the six intervals. The PE pattern may reflect the comparatively small numbers of particles in several bins and the spectral similarity of PE to oils, waxes, and other aliphatic materials. It should therefore not be interpreted as evidence that low-$R_{max}$ PE assignments are generally reliable.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


PS also showed substantial variability, with TopHit ratios of 0.25, 0.67, 0.67, 0.50, and 1.00 from the lowest to the 0.8–0.9 interval. No PS assignment was available in the 0.9–1.0 interval. Each populated PS interval contained only two to six particles. The apparent irregularity is therefore dominated by limited sample size and associated binomial uncertainty. Larger polymer-specific datasets are required before a stable PS calibration or a polymer-specific threshold can be established.

The polymer-specific patterns differed substantially in transition position and steepness. PP and PET exhibited comparatively smooth increases in TopHit ratio, whereas the PE and PS observations were strongly influenced by bins with small sample sizes and wide confidence intervals. The fitted lines for PE and PS are therefore shown only to visualise the available observations and should not be interpreted as validated polymer-specific calibration functions.

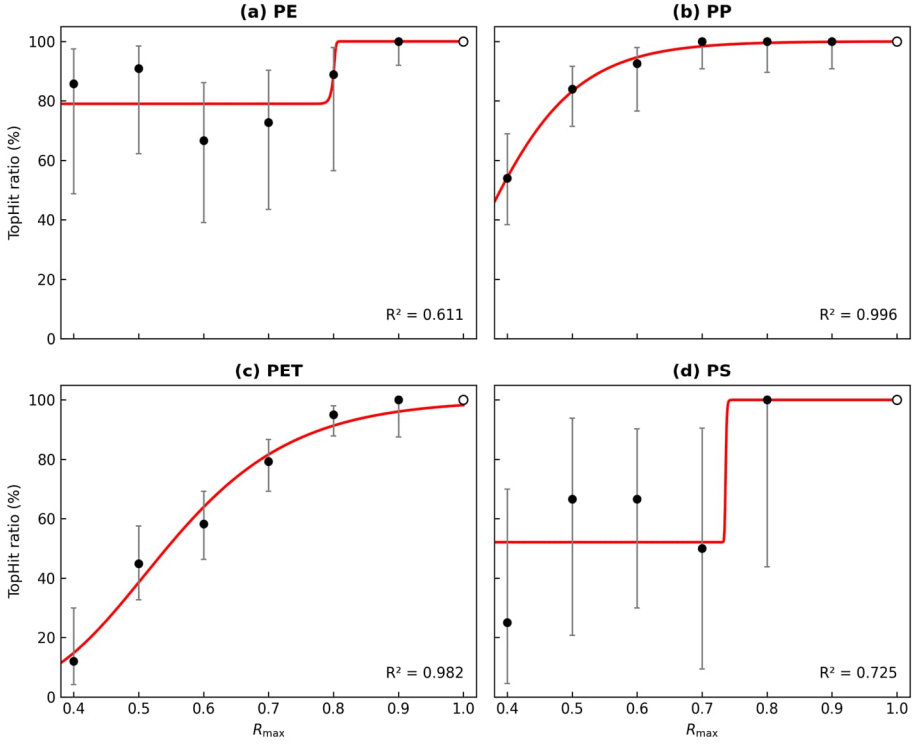


**Figure 7. Polymer-specific relationships between $R_{max}$ and the TopHit ratio for (a) PE, (b) PP, (c) PET, and (d) PS.** Black circles show environmental-particle observations, grey error bars indicate Wilson 95 % confidence intervals, red curves show descriptive Richards-type fits, and open circles at $R_{max}$ = 1.0 denote the reference-control endpoint. No environmental-particle observation is plotted for PS in the 0.9–1.0 interval because no assignments were available in that bin. The upper asymptote was fixed at 100 %. The curves are visual guides; the observed ratios, confidence intervals, and counts in Table A2 provide the basis for interpretation. Because few observations are available per Rmax bin for (a) PE and (d) PS, the Richards-

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


type fits for these two panels are unstable and near-degenerate, and their step-like appearance should not be interpreted as a sharp identification threshold.

The polymer-specific results indicate that a universal $R_{max}$ cutoff is not appropriate. They also show that the overall calibration should be used as an empirical indicator of identification consistency for mixed environmental samples rather than as a substitute for polymer-specific interpretation. Where sufficiently large validation datasets are available, polymer-specific calibration may improve the interpretation of $R_{max}$. With the present dataset, however, the pooled calibration is the more stable operational model, whereas the polymer-specific plots identify classes and $R_{max}$ ranges for which additional diagnostic-band evaluation is particularly important.

### 3.6 Methodological significance, scope, and limitations

The principal contribution of this study is a quantitative and reproducible µFTIR-ATR framework that assigns distinct preprocessing methods to distinct analytical purposes. In the chemical-identification pathway, SG smoothing reduces spectral noise while preserving diagnostic peak shapes, after which library matching, positive-band confirmation, non-polymer exclusion, and the $R_{max}$–TopHit relationship provides graded evidence of polymer identity. In the physical-characterisation pathway, PCA-based noise reduction is applied separately to imaging data, and the 5σ mapping threshold defines image features for Feret measurement and morphological classification. The 5σ mapping threshold is therefore not part of polymer identification. This functional separation is particularly important for airborne particles below 10 µm, including those collected in the $PM_{2.5}$ fraction, because PCA reconstruction can modify subtle spectral features used for identification, whereas PCA remains effective for improving the delineation of particle images (Koziol et al., 2018). The combined framework addresses two previously unresolved problems: reliable polymer assignment without relying on PCA-reconstructed spectra or a binary $R_{max}$ cutoff, and reproducible physical characterisation of particle maps containing substantial pixel-level noise.

The principal uncertainties likewise differ between the two pathways. In chemical identification, residual spectral noise, baseline drift, the SG window, particle thickness, weathering, matrix interference, ATR contact, and library composition can affect diagnostic bands and $R_{max}$. In physical characterisation, PCA settings, blank variability, intrinsic C–H and C=O absorption strength, pixel occupancy, and particle geometry can affect map delineation and the resulting Feret dimensions. Standardised ATR-contact procedures minimise avoidable operator-related variation, but they cannot remove the intrinsic spectral and spatial limitations associated with very small or weakly absorbing particles. Explicitly separating these sources of uncertainty prevents the 5σ mapping threshold from being misinterpreted as a polymer-identification threshold and clarifies which components should be evaluated in future interlaboratory comparisons.

The TopHit ratio is an operational, agreement-based measure of empirical spectral-matching confidence derived exclusively from the chemical-identification pathway. It quantifies consistency between whole-spectrum library matching and diagnostic-band assessment of the SG-smoothed spectra under explicitly defined analytical conditions. PCA-based noise

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


reduction and the 5σ mapping threshold did not enter the TopHit calculation. Because the original polymer composition of an unknown, weathered environmental particle cannot be independently verified after collection, the numerical TopHit calibration should not be interpreted as an absolute probability of polymer identity. This qualification does not limit the broader applicability of the framework: by specifying the instrument configuration, SG smoothing, spectral libraries, sample matrix, analyst blinding, decision criteria, and uncertainty calculation, the study provides a reproducible protocol through which other laboratories can establish and compare their own confidence relationships. Diagnostic-band interpretation in the present study was performed by one blinded analyst; multi-analyst and interlaboratory assessments will provide the next stage of reproducibility testing and support the development of transferable confidence ranges.

The identified $PM_{2.5}$ particle with an apparent minimum Feret dimension of 2.3 µm demonstrates particle-resolved chemical interpretation and physical measurement approaching the 2 µm scale. The aerodynamic fraction and image-derived Feret dimension are distinct, and this example does not define a universal lower limit. The practical lower size range varies with polymer absorption strength, morphology, thickness, ATR contact, pixel occupancy, and matrix complexity. The present analytical framework was developed to enable reliable polymer identification and quantification, ultimately supporting the determination of representative airborne microplastic concentrations in each study region. It therefore provides an analytical benchmark and a defined validation target for subsequent quantitative monitoring, interlaboratory comparison, and inter-study harmonisation.

Within this defined scope, the workflow offers three immediate benefits. First, SG smoothing preserves the diagnostic peak shapes required for polymer identification without relying on PCA-reconstructed spectra. Second, PCA-based noise reduction and the 5σ mapping threshold improve particle-map delineation for size and morphology measurements without being used as polymer-identification criteria. Third, the $R_{max}$–TopHit relationship avoids automatic rejection of chemically consistent weathered particles with intermediate $R_{max}$ values and makes the basis of final classification explicit through complementary positive and negative spectral criteria. Most importantly, to the best of our knowledge, this is the first study to combine µFTIR-based polymer identification of an airborne $PM_{2.5}$ particle approaching 2 µm with explicit empirical calibration of spectral-matching confidence. The framework replaces an arbitrary binary interpretation of $R_{max}$ with a graded confidence indicator derived from agreement between complementary spectral interpretations under defined analytical conditions while retaining spectrum-level chemical review of each environmental particle.

## 4 Conclusions

This study establishes an empirically calibrated µFTIR-ATR imaging framework for particle-resolved identification and physical characterisation of airborne microplastics below 10 µm, including particles collected in the $PM_{2.5}$ fraction. Its central methodological advance is the deliberate separation of preprocessing according to analytical purpose: SG smoothing was used to reduce spectral noise while preserving diagnostic peak shapes for polymer identification and degradation assessment,

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


whereas PCA-based noise reduction was applied only to imaging data used for particle mapping. 5σ mapping thresholds defined which mapped features could be subjected to Feret measurement and morphological classification; they were not used to determine polymer identity or calculate the TopHit ratio. The workflow enabled identification and measurement of an airborne $PM_{2.5}$ PET particle approaching 2 µm, and the TopHit ratio increased from 41.1 % at $R_{max}$ = 0.4–0.5 to 100.0 % at $R_{max}$ = 0.9–1.0. Among the three models examined, the Richards growth model consistently provided the highest $R^2$ and lowest RMSE across both library compositions and bin widths, supporting its use as the primary descriptive calibration curve. By combining peak-preserving chemical identification, separately optimised physical mapping, and empirical $R_{max}$–TopHit calibration, the framework provides an analytical basis for reliable quantification of representative airborne microplastic concentrations, interlaboratory comparison, harmonised reporting, and future method standardisation.

## Appendix A: Additional experimental and analytical details

### A1 TopHit ratio and confidence intervals

The definition and interpretation of the TopHit ratio are provided in Sects. 2.3 and 3.3. The calibration dataset contained 674 particles pooled across all analysed filters. The pooled and polymer-specific analyses used bins of $\Delta R_{max}$ = 0.1. Only candidates with $R_{max} \geq 0.4$ were included because spectra below this operational lower bound could not be interpreted with sufficient confidence at the polymer level, even after diagnostic-band inspection. Two-sided Wilson score 95 % confidence intervals were calculated for each bin using z = 1.96. These observed proportions and confidence intervals constitute the primary statistical evidence; the fitted curves provide descriptive summaries. The pooled counts are listed in Table A1, and the polymer-specific counts are listed in Table A2.

### A2 Calibration-model fitting

Two-parameter logistic, constrained four-parameter logistic, and Richards-type asymmetric sigmoidal functions were fitted independently to binned TopHit ratios using unweighted nonlinear least squares. These fits were used only to compare and visualise the empirical shape across library and bin-width conditions; they were not interpreted as particle-level probability models. For $x = R_{max}$, the fitted functions were as follows.

$$p(x) = \frac{1}{1 + \exp\{-k(x - x_0)\}}$$

$$p(x) = y_{\min} + \frac{1 - y_{\min}}{1 + \exp\{-k(x - x_0)\}}$$

$$p(x) = y_{\min} + \frac{1 - y_{\min}}{[1 + \exp\{-k(x - x_0)\}]^{1/\nu}}$$

In the constrained four-parameter logistic and Richards models, the upper asymptote was fixed at 1; ν controls asymmetry in the Richards function. Robustness was evaluated for the comprehensive library (approximately 10,000 spectra) and targeted

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


library (approximately 100 spectra), each using $\Delta R_{max}$ values of 0.02 and 0.1. Reference spectra searched against the targeted library yielded a reference-control endpoint at $R_{max}$ = 1.0 and TopHit ratio = 100 %, which was retained as a boundary anchor in the displayed fits. Separate endpoint-exclusion analyses quantified the influence of this anchor. Regression input data and fitted coefficients are provided as part of the dataset listed under Data availability.

Model performance was summarised using the following statistics.

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$\mathrm{AIC} = n\ln\left(\frac{\mathrm{SSE}}{n}\right) + 2p$$

$$\mathrm{BIC} = n\ln\left(\frac{\mathrm{SSE}}{n}\right) + p\ln(n)$$

$R^2$ and RMSE were calculated on the 0–1 TopHit-ratio scale as descriptive goodness-of-fit measures: higher $R^2$ and lower RMSE indicate closer agreement between the fitted curve and the observed bin proportions. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) additionally balance goodness of fit against model complexity; lower values indicate stronger relative support within the same dataset. BIC imposes a stronger complexity penalty than AIC as the number of observations increases. Here, both criteria were calculated from a Gaussian least-squares likelihood with common residual variance. The number of fitted binned points, n, included the reference-control endpoint, and p denotes the number of free parameters: 2 for the two-parameter logistic, 3 for the constrained four-parameter logistic, and 4 for the Richards model. $R^2$ and RMSE were used to identify the function that most closely described the observations, whereas AIC and BIC assessed whether improved fit justified the additional parameters. The overall model-comparison statistics corresponding to Fig. 6 are provided in Table A3.

Table A1. Pooled library-ranking counts and Wilson 95 % confidence intervals underlying Fig. 5.

| $R_{max}$ range | TopHit | 2nd–5th | 6th–10th | >10th | Total (n) | TopHit ratio, % (95 % CI) |
|---|---|---|---|---|---|---|
| 0.4–0.5 | 30 | 34 | 3 | 6 | 73 | 41.1 (30.5–52.6) |
| 0.5–0.6 | 80 | 36 | 1 | 5 | 122 | 65.6 (56.8–73.4) |
| 0.6–0.7 | 76 | 31 | 2 | 3 | 112 | 67.9 (58.7–75.8) |
| 0.7–0.8 | 112 | 20 | 0 | 1 | 133 | 84.2 (77.1–89.4) |

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


| $R_{max}$ range | TopHit | 2nd–5th | 6th–10th | >10th | Total (n) | TopHit ratio, % (95 % CI) |
|---|---|---|---|---|---|---|
| 0.8–0.9 | 120 | 5 | 0 | 0 | 125 | 96.0 (91.0–98.3) |
| 0.9–1.0 | 109 | 0 | 0 | 0 | 109 | 100.0 (96.6–100.0) |

Note: The reference-control endpoint at $R_{max}$ = 1.0 and TopHit ratio = 100 % is not included in the environmental-particle bin counts.

Table A2. Polymer-specific library-ranking counts and Wilson 95 % confidence intervals underlying Fig. 7.

| Polymer | $R_{max}$ range | TopHit | 2nd–5th | 6th–10th | >10th | Total (n) | TopHit ratio, % (95 % CI) |
|---|---|---|---|---|---|---|---|
| PE | 0.4–0.5 | 6 | 0 | 1 | 0 | 7 | 85.7 (48.7–97.4) |
| PE | 0.5–0.6 | 10 | 0 | 0 | 1 | 11 | 90.9 (62.3–98.4) |
| PE | 0.6–0.7 | 8 | 1 | 0 | 3 | 12 | 66.7 (39.1–86.2) |
| PE | 0.7–0.8 | 8 | 2 | 0 | 1 | 11 | 72.7 (43.4–90.3) |
| PE | 0.8–0.9 | 8 | 1 | 0 | 0 | 9 | 88.9 (56.5–98.0) |
| PE | 0.9–1.0 | 44 | 0 | 0 | 0 | 44 | 100.0 (92.0–100.0) |
| PP | 0.4–0.5 | 20 | 14 | 0 | 3 | 37 | 54.1 (38.4–69.0) |
| PP | 0.5–0.6 | 42 | 5 | 1 | 2 | 50 | 84.0 (71.5–91.7) |
| PP | 0.6–0.7 | 25 | 2 | 0 | 0 | 27 | 92.6 (76.6–97.9) |
| PP | 0.7–0.8 | 38 | 0 | 0 | 0 | 38 | 100.0 (90.8–100.0) |
| PP | 0.8–0.9 | 33 | 0 | 0 | 0 | 33 | 100.0 (89.6–100.0) |
| PP | 0.9–1.0 | 38 | 0 | 0 | 0 | 38 | 100.0 (90.8–100.0) |
| PET | 0.4–0.5 | 3 | 19 | 2 | 1 | 25 | 12.0 (4.2–30.0) |
| PET | 0.5–0.6 | 26 | 30 | 0 | 2 | 58 | 44.8 (32.7–57.5) |
| PET | 0.6–0.7 | 39 | 26 | 2 | 0 | 67 | 58.2 (46.3–69.3) |
| PET | 0.7–0.8 | 65 | 17 | 0 | 0 | 82 | 79.3 (69.3–86.6) |
| PET | 0.8–0.9 | 76 | 4 | 0 | 0 | 80 | 95.0 (87.8–98.0) |
| PET | 0.9–1.0 | 27 | 0 | 0 | 0 | 27 | 100.0 (87.5–100.0) |
| PS | 0.4–0.5 | 1 | 1 | 0 | 2 | 4 | 25.0 (4.6–69.9) |

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


| Polymer | $R_{max}$ range | TopHit | 2nd–5th | 6th–10th | >10th | Total (n) | TopHit ratio, % (95 % CI) |
|---|---|---|---|---|---|---|---|
| PS | 0.5–0.6 | 2 | 1 | 0 | 0 | 3 | 66.7 (20.8–93.9) |
| PS | 0.6–0.7 | 4 | 2 | 0 | 0 | 6 | 66.7 (30.0–90.3) |
| PS | 0.7–0.8 | 1 | 1 | 0 | 0 | 2 | 50.0 (9.5–90.5) |
| PS | 0.8–0.9 | 3 | 0 | 0 | 0 | 3 | 100.0 (43.8–100.0) |
| PS | 0.9–1.0 | 0 | 0 | 0 | 0 | 0 | – |

Note: The reference-control endpoint at $R_{max}$ = 1.0 and TopHit ratio = 100 % is not included in the environmental-particle bin counts. No PS assignment was available in the 0.9–1.0 bin.

Table A3. Performance of the calibration models shown in Fig. 6.

| Reference library | $\Delta R_{max}$ | Binned points | Model | $R^2$ | RMSE | AIC | BIC |
|---|---|---|---|---|---|---|---|
| Comprehensive (~10,000) | 0.02 | 29 | Two-parameter logistic | 0.609 | 0.156 | −103.935 | −101.200 |
| Comprehensive (~10,000) | 0.02 | 29 | Constrained four-parameter logistic | 0.697 | 0.137 | −109.369 | −105.267 |
| Comprehensive (~10,000) | 0.02 | 29 | Richards | 0.723 | 0.131 | −110.005 | −104.536 |
| Comprehensive (~10,000) | 0.1 | 7 | Two-parameter logistic | 0.794 | 0.113 | −26.559 | −26.667 |
| Comprehensive (~10,000) | 0.1 | 7 | Constrained four-parameter logistic | 0.889 | 0.083 | −28.870 | −29.033 |
| Comprehensive (~10,000) | 0.1 | 7 | Richards | 0.916 | 0.072 | −28.819 | −29.036 |
| Targeted (~100) | 0.02 | 31 | Two-parameter logistic | 0.887 | 0.073 | −158.662 | −155.794 |
| Targeted (~100) | 0.02 | 31 | Constrained four-parameter logistic | 0.887 | 0.073 | −156.671 | −152.369 |
| Targeted (~100) | 0.02 | 31 | Richards | 0.887 | 0.072 | −154.731 | −148.995 |

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


| Reference library | $\Delta R_{max}$ | Binned points | Model | $R^2$ | RMSE | AIC | BIC |
|---|---|---|---|---|---|---|---|
| Targeted (~100) | 0.1 | 7 | Two-parameter logistic | 0.959 | 0.042 | −40.545 | −40.654 |
| Targeted (~100) | 0.1 | 7 | Constrained four-parameter logistic | 0.960 | 0.041 | −38.702 | −38.865 |
| Targeted (~100) | 0.1 | 7 | Richards | 0.964 | 0.039 | −37.559 | −37.775 |

RMSE is reported on the 0–1 TopHit-ratio scale. AIC and BIC were calculated using p = 2, 3, and 4 for the two-parameter logistic, constrained four-parameter logistic, and Richards models, respectively; lower values indicate stronger support within the same library and bin-width condition. The reference-control endpoint at $R_{max}$ = 1.0 and TopHit ratio = 100 % was retained as a boundary anchor in the displayed fits; the upper asymptote was fixed at 100 % for the constrained four-parameter logistic and Richards models. These model statistics are descriptive and do not replace the observed binomial proportions and Wilson confidence intervals.

## Data availability

The data underlying this study — including the particle-level library-matching records, the binned TopHit-ratio calibration data, the fitted calibration-model coefficients, and the accompanying documentation (README, data dictionary, and correction log) — are openly available from the Zenodo repository at https://doi.org/10.5281/zenodo.21848289 (Niida, 2026).

## Author contribution

Yasuhiro Niida conceived the study, conducted the imaging measurements, analysed the data, and wrote the manuscript. Hiroshi Okochi supervised the study and contributed to the study design and interpretation of the results. Norihisa Yoshida, Yuto Tani, and Shunki Sasai performed sample collection, pretreatment, analysis, and data processing. Yusuke Fujii contributed to discussion of the results. Norimichi Takenaka contributed to supervision and critical revision of the manuscript. All authors discussed the results and approved the final manuscript.

## Competing interests

Yasuhiro Niida is an employee of PerkinElmer Japan G.K., the manufacturer of the FTIR imaging instrument and software used in this study. The other authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


**Acknowledgements**

The authors used Claude Opus 4.8 (Anthropic) to assist with English-language editing and the graphical formatting of figures. The tool was not used to generate scientific data, perform statistical analyses, or determine scientific interpretations. All outputs were reviewed and verified by the authors.

**Financial support**

This study was supported by the Environmental Research and Technology Development Fund (JPMEERF20215003 and JPMEERF20245004) of the Environmental Restoration and Conservation Agency of Japan.

**References**

Allen, D., Allen, S., Abbasi, S., Baker, A., Bergmann, M., Brahney, J., Butler, T., Duce, R. A., Eckhardt, S., Evangeliou, N., Jickells, T., Kanakidou, M., Kershaw, P., Laj, P., Levermore, J., Li, D., Liss, P., Liu, K., Mahowald, N., Masque, P., Materić, D., Mayes, A. G., McGinnity, P., Osvath, I., Prather, K. A., Prospero, J. M., Revell, L. E., Sander, S. G., Shim, W. J., Slade, J., Stein, A., Tarasova, O., and Wright, S.: Microplastics and nanoplastics in the marine–atmosphere environment, Nat. Rev. Earth Environ., 3, 393–405, https://doi.org/10.1038/s43017-022-00292-x, 2022.

Allen, S., Allen, D., Phoenix, V. R., Le Roux, G., Durántez Jiménez, P., Simonneau, A., Binet, S., Galop, D., and Ter Halle, A.: Atmospheric transport and deposition of microplastics in a remote mountain catchment, Nat. Geosci., 12, 339–344, https://doi.org/10.1038/s41561-019-0335-5, 2019.

Amato-Lourenço, L. F., Carvalho-Oliveira, R., Ribeiro Junior, G., Dos Santos Galvão, L., Ando, R. A., and Mauad, T.: Presence of airborne microplastics in human lung tissue, J. Hazard. Mater., 416, 126124, https://doi.org/10.1016/j.jhazmat.2021.126124, 2021.

Araujo, C. F., Nolasco, M. M., Ribeiro, A. M., and Ribeiro-Claro, P. J. A.: Identification of microplastics using Raman spectroscopy: Latest developments and future prospects, Water Res., 142, 426–440, https://doi.org/10.1016/j.watres.2018.05.060, 2018.

Berkson, J.: Application of the logistic function to bio-assay, J. Am. Stat. Assoc., 39, 357–365, https://doi.org/10.1080/01621459.1944.10500699, 1944.

Brahney, J., Hallerud, M., Heim, E., Hahnenberger, M., and Sukumaran, S.: Plastic rain in protected areas of the United States, Science, 368, 1257–1260, https://doi.org/10.1126/science.aaz5819, 2020.

Colthup, N. B., Daly, L. H., and Wiberley, S. E.: Introduction to Infrared and Raman Spectroscopy, 3rd edn., Academic Press, Cambridge, MA, USA, 301–314, 1990.

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


Cowger, W., Steinmetz, Z., Gray, A., Munno, K., Lynch, J., Hapich, H., Primpke, S., De Frond, H., and Rochman, C. M.: Microplastic spectral classification needs an open source community: Open Specy to the rescue!, Anal. Chem., 93, 7543–7548, https://doi.org/10.1021/acs.analchem.1c00123, 2021.

De Lean, A., Munson, P. J., and Rodbard, D.: Simultaneous analysis of families of sigmoidal curves: Application to bioassay, radioligand assay, and physiological dose-response curves, Am. J. Physiol., 235, E97–E102, https://doi.org/10.1152/ajpendo.1978.235.2.E97, 1978.

Dris, R., Gasperi, J., Saad, M., Mirande, C., and Tassin, B.: Synthetic fibers in atmospheric fallout: A source of microplastics in the environment?, Mar. Pollut. Bull., 104, 290–293, https://doi.org/10.1016/j.marpolbul.2016.01.006, 2016.

Jenner, L. C., Rotchell, J. M., Bennett, R. T., Cowen, M., Tentzeris, V., and Sadofsky, L. R.: Detection of microplastics in human lung tissue using μFTIR spectroscopy, Sci. Total Environ., 831, 154907, https://doi.org/10.1016/j.scitotenv.2022.154907, 2022.

Käppler, A., Fischer, D., Oberbeckmann, S., Schernewski, G., Labrenz, M., Eichhorn, K. J., and Voit, B.: Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, Raman or both?, Anal. Bioanal. Chem., 408, 8377–8391, https://doi.org/10.1007/s00216-016-9956-3, 2016.

Koziol, P., Raczkowska, M. K., Skibinska, J., Urbaniak-Wasik, S., Paluszkiewicz, C., Kwiatek, W., and Wrobel, T. P.: Comparison of spectral and spatial denoising techniques in the context of High Definition FT-IR imaging hyperspectral data, Sci. Rep., 8, 14351, https://doi.org/10.1038/s41598-018-32713-7, 2018.

Morioka, T., Tanaka, S., Kohama-Inoue, A., and Watanabe, A.: The quantification of the airborne plastic particles of 0.43–11 μm: Procedure development and application to atmospheric environment, Chemosphere, 351, 141131, https://doi.org/10.1016/j.chemosphere.2024.141131, 2024.

Moses, S. R., Roscher, L., Primpke, S., Hufnagl, B., Löder, M. G. J., Gerdts, G., and Laforsch, C.: Comparison of two rapid automated analysis tools for large FTIR microplastic datasets, Anal. Bioanal. Chem., 415, 2975–2987, https://doi.org/10.1007/s00216-023-04630-w, 2023.

Niida, Y.: Data for Airborne Microplastics in the PM2.5 Fraction: Empirically Calibrated μFTIR-ATR Imaging Detects Particles Down to 2 μm [data set], Zenodo, https://doi.org/10.5281/zenodo.21848289, 2026.

Niida, Y., Fujii, Y., Inatsugi, Y., and Takenaka, N.: Quantifying UV-driven aging of sub-10 μm airborne microplastics with high-resolution μFTIR-ATR imaging, Atmosphere, 17, 146, https://doi.org/10.3390/atmos17020146, 2026.

Primpke, S., Wirth, M., Lorenz, C., and Gerdts, G.: Reference database design for the automated analysis of microplastic samples based on Fourier transform infrared spectroscopy, Anal. Bioanal. Chem., 410, 5131–5141, https://doi.org/10.1007/s00216-018-1156-x, 2018.

Richards, F. J.: A flexible growth function for empirical use, J. Exp. Bot., 10, 290–301, https://doi.org/10.1093/jxb/10.2.290, 1959.

Rochman, C. M., Brookson, C., Bikker, J., Djuric, N., Earn, A., Bucci, K., Athey, S., Huntington, A., McIlwraith, H., Munno, K., De Frond, H., Kolomijeca, A., Erdle, L., Grbic, J., Bayoumi, M., Borrelle, S. B., Wu, T., Santoro, S., Werbowski, L. M.,

https://doi.org/10.5194/egusphere-2026-5191
Preprint. Discussion started: 24 September 2026


Zhu, X., Giles, R. K., Hamilton, B. M., Thaysen, C., Kaura, A., Klasios, N., Ead, L., Kim, J., Sherlock, C., Ho, A., and Hung, C.: Rethinking microplastics as a diverse contaminant suite, Environ. Toxicol. Chem., 38, 703–711, https://doi.org/10.1002/etc.4371, 2019.

Savitzky, A. and Golay, M. J. E.: Smoothing and differentiation of data by simplified least squares procedures, Anal. Chem., 36, 1627–1639, https://doi.org/10.1021/ac60214a047, 1964.

Tokyo Dylec Corp.: MCI sampler, English catalogue, https://www.t-dylec.net/service/mci/ (last access: 3 August 2026), 2020.

Tokunaga, Y., Okochi, H., Tani, Y., Niida, Y., Tachibana, T., Saigawa, K., Katayama, K., Moriguchi, S., Kato, T., and Hayama, S.: Airborne microplastics detected in the lungs of wild birds in Japan, Chemosphere, 321, 138032, https://doi.org/10.1016/j.chemosphere.2023.138032, 2023.

Vianello, A., Jensen, R. L., Liu, L., and Vollertsen, J.: Simulating human exposure to indoor airborne microplastics using a Breathing Thermal Manikin, Sci. Rep., 9, 8670, https://doi.org/10.1038/s41598-019-45054-w, 2019.

Wang, Y., Okochi, H., Tani, Y., Hayami, H., Minami, Y., Katsumi, N., Takeuchi, M., Sorimachi, A., Fujii, Y., Kajino, M., Adachi, K., Ishihara, Y., Iwamoto, Y., and Niida, Y.: Airborne hydrophilic microplastics in cloud water at high altitudes and their role in cloud formation, Environ. Chem. Lett., 21, 3055–3062, https://doi.org/10.1007/s10311-023-01626-x, 2023.

Zhang, Y., Kang, S., Allen, S., Allen, D., Gao, T., and Sillanpää, M.: Atmospheric microplastics: A review on current status and perspectives, Earth Sci. Rev., 203, 103118, https://doi.org/10.1016/j.earscirev.2020.103118, 2020.