Systems and Soft Computing 9 (2026) 200537

Contents lists available at ScienceDirect

Systems and Soft Computing

journal homepage: www.journals.elsevier.com/soft-computing-letters

A modular computational framework for integrated microplastic data analysis and environmental monitoring

Edinelson Saldanha [a,*], Vando Gomes [b], João Vitor Araujo [b], Ruan Alcantra Felix [b], Maria Carneiro [b], Ludmila Loureiro [b]

[a] Federal University of Pará (UFPA), Laboratory of Coastal and Oceanic Environmental Geochemistry, Brazil
[b] Federal University of Pará (UFPA), Brazil

A R T I C L E I N F O

Keywords:
Microplastics
Scientific software
Dimensional metrics
Open-source framework
Environmental modelling
Python

A B S T R A C T

Microplastic pollution has emerged as a major environmental concern, increasing the demand for reliable computational tools capable of supporting standardized quantification and interpretation of contamination data. However, current microplastic assessment workflows are often fragmented, separating particle measurement, dimensional aggregation, and statistical analysis into independent procedures, which reduces reproducibility, transparency, and analytical integration. To address these limitations, this study presents IMPa, an open-source computational framework implemented in Python for integrated microplastic data analysis and environmental monitoring. The proposed framework combines count-based concentration (MP), cumulative length metrics (IMPa-L), and projected area metrics (IMPa-A) within a unified architecture that supports automated normalization, morphological stratification, multilevel aggregation, dimensional distribution analysis, matrix-based representation, and embedded multivariate analysis through Principal Component Analysis (PCA). The system was designed using a modular architecture that ensures traceability from particle-level measurements to aggregated environmental indicators while facilitating extensibility and reproducibility. The framework was evaluated using a demonstrative dataset obtained from a tidally influenced river and computational benchmark datasets containing up to 1,000,000 particles. Results showed that dimension-based indices captured variability not detected by count-based concentration alone, highlighting structural differences among aggregation schemes. The embedded PCA explained 91.9% of the total variance in the first two principal components, demonstrating the framework's ability to support exploratory interpretation of multidimensional microplastic datasets. Performance benchmarking confirmed near-linear computational scalability (O(n)) in both execution time and memory consumption, with successful processing of datasets containing up to one million particle records. By integrating measurement, aggregation, visualization, and statistical exploration within a single computational environment, the proposed framework provides a reproducible and transparent solution for dimension-integrated microplastic assessment. The software is distributed under the MIT License, publicly available on GitHub, and archived with a persistent DOI (10.5281/zenodo.18386718), ensuring long-term accessibility and reproducibility.

1. Introduction

Microplastic contamination has emerged as a pervasive environmental issue over recent decades, being reported across freshwater, marine, and terrestrial ecosystems worldwide [1,2]. Due to their persistence, small size, and capacity to interact with both biota and chemical contaminants, microplastics present complex monitoring challenges that require robust and transparent analytical approaches [3].

Environmental assessments of microplastic contamination have predominantly relied on particle counting, commonly expressed as items per unit volume (items/L). This metric has been widely adopted due to its simplicity and comparability across monitoring programs [4,5]. However, growing evidence suggests that particle number alone may

* Corresponding author.
E-mail addresses: edinelsonsaldanha@ufpa.br (E. Saldanha), Vandogomes@ufpa.br (V. Gomes), joao.araujo@salinopolis.ufpa.br (J.V. Araujo), ruanengcosteira@gmail.com (R.A. Felix), mariaeduardacarneiro6789@gmail.com (M. Carneiro), ludmilacruz575@gmail.com (L. Loureiro).

https://doi.org/10.1016/j.sasc.2026.200537
Received 9 March 2026; Received in revised form 23 June 2026; Accepted 29 June 2026
Available online 30 June 2026

not fully capture the dimensional variability inherent to microplastic assemblages, particularly when particle size and morphology differ substantially among samples [6].

Microplastic dimensions influence transport behavior, settling and rising velocities, fragmentation dynamics, bioavailability, and surface-mediated interactions with pollutants and organisms [6,7]. Consequently, length- and area-based descriptors have increasingly been incorporated into analytical workflows as complementary metrics to particle count [8]. Despite these advances, dimensional metrics are often calculated through independent or manual procedures, limiting reproducibility and hindering integration across studies [5].

In parallel, image-based techniques have enabled more precise measurement of particle size and morphology from microscopic imagery [4]. Nevertheless, the absence of integrated computational environments capable of combining particle counting, dimensional aggregation, morphological classification, and structured visualization remains a methodological and computational bottleneck in microplastic research [9]. Many workflows rely on fragmented processing steps that separate measurement, aggregation, and statistical representation, reducing transparency, reproducibility, and scalability.

Recent advances have emphasized the importance of harmonized and reproducible microplastic data workflows, particularly through standardized metadata, open data practices, and computational tools for spectral identification and data integration [10–12]. Similar concerns regarding interoperability, reproducibility, and integration of heterogeneous information have also motivated the development of structured computational frameworks in other data-intensive scientific domains [13]. These studies demonstrate that unified computational architectures can improve information consistency, facilitate analytical integration, and enhance reproducibility across complex datasets. Collectively, these developments highlight the growing need for modular and transparent analytical environments capable of supporting complex environmental datasets.

Existing software solutions, such as Open Specy, have contributed significantly to specific stages of microplastic analysis, especially spectral classification and open reference-library development. However, most available platforms remain focused on individual analytical tasks and do not provide an integrated environment capable of simultaneously combining particle counting, dimensional aggregation, morphological stratification, and multivariate analysis within a single reproducible workflow. This methodological fragmentation motivates the development of integrated computational infrastructures capable of supporting transparent and scalable environmental assessments.

To address this gap, the present study introduces a modular, open-source computational framework for integrated microplastic quantification. The framework combines count-based concentration with length- and area-based aggregation derived from calibrated microscopic images, while preserving morphological classification throughout the analytical pipeline. Rather than prescribing a single optimal metric, the system enables parallel computation and comparison of multiple aggregation schemes within a reproducible and extensible modeling architecture.

The framework is implemented in Python with explicit separation between computational core and graphical interface layers, ensuring methodological invariance and facilitating independent validation. Its performance is evaluated using a real-world dataset collected under variable environmental conditions, demonstrating the framework's ability to integrate measurement, normalization, aggregation, and visualization within an integrated and scalable computational workflow for environmental data analysis.

To demonstrate its practical application, the framework was evaluated using a microplastic dataset collected from a tidally influenced river system. The dataset comprised particle-level measurements obtained from calibrated microscopic images, including length, projected area, morphological classification, and sampling metadata. These data were processed through the complete analytical workflow, enabling demonstration of metric computation, aggregation, stratification, visualization, and multivariate exploration within a real environmental monitoring context.

In addition to dimensional aggregation, the framework incorporates multivariate analysis through Principal Component Analysis (PCA) as an exploratory statistical tool. PCA is included to demonstrate the framework's capacity for integrated multivariate assessment and visualization of relationships among metrics and morphological categories. It is not intended as a predictive or inferential environmental model, and its application in this study should be interpreted as a proof-of-concept analytical component within the broader computational workflow.

The main contributions of this work are:

- Development of an open-source computational framework for integrated microplastic analysis;

Simultaneous computation of count-based (MP), length-based (IMPa-L), and area-based (IMPa-A) metrics from a unified particle-level dataset;

- Automated support for morphological stratification, multilevel aggregation, and dimensional distribution analysis;
- Integration of exploratory multivariate analysis through embedded PCA routines;
- Demonstration of computational scalability and reproducibility through benchmark testing and application to a real-world environmental dataset.

The novelty of the proposed framework lies in the integration of count-based, length-based, and area-based microplastic metrics within a unified computational environment, preserving particle-level traceability while enabling multilevel aggregation, morphological stratification, and direct comparison between aggregation schemes. Unlike existing workflows that treat these metrics independently or require manual integration, the proposed system provides a reproducible and extensible infrastructure for dimension-integrated microplastic assessment.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background and methodological foundations of the framework. Section 3 describes the overall system architecture. Section 4 details the methodological workflow and metric definitions. Section 5 presents the software implementation. Section 6 demonstrates the framework using an environmental dataset. Section 7 evaluates computational performance and scalability. Section 8 discusses the analytical implications, limitations, and future perspectives of the proposed approach. Finally, Section 9 summarizes the main conclusions.

2. Conceptual framework

2.1. Conceptual limitations of count-based microplastic metrics

Particle count metrics have become the dominant indicator in microplastic studies, largely due to their operational simplicity and ease of comparison across studies [5]. However, expressing contamination exclusively as items per unit volume does not explicitly incorporate dimensional attributes of microplastics, including size, shape, and surface area [1].

From a structural standpoint, count-based metrics treat particles as equivalent units regardless of dimensional variability. Two samples with identical microplastic counts may differ substantially in terms of total material extension, interaction surface, and physical characteristics. Conversely, samples with fewer particles may present distinct dimensional profiles if dominated by larger or elongated microplastics, such as fibers [2]. This distinction highlights differences between numerical representation and dimensional characterization, which may influence environmental interpretation.

Moreover, particle counts alone may exhibit limited compatibility with dimension-sensitive environmental modelling approaches, which often rely on physically descriptive variables to represent transport, deposition, and interaction processes [6]. As a result, exclusive reliance on items/L may reduce integration with modelling workflows that incorporate particle-level physical attributes.

2.2. Length-based metrics as a physical descriptor

Length-based metrics represent an alternative dimensional descriptor to particle counts by incorporating information on particle extension. The cumulative microplastic length per unit volume captures differences in size distributions that remain unrepresented under purely count-based approaches, particularly in samples dominated by fibrous particles [7].

Conceptually, cumulative length can serve as a proxy for fragmentation state and dimensional presence, enabling complementary comparison between samples with heterogeneous particle populations [8]. Length-based indicators are computationally tractable and can be directly obtained from image-based measurements, making them suitable for standardized implementation within environmental assessment and modelling frameworks [9].

2.3. Area-based metrics and surface interaction representation

While length-based indicators provide additional dimensional insight, they do not fully characterize non-fibrous microplastics such as fragments, films, and pellets. Projected area-based metrics extend dimensional representation by quantifying total particle surface projection per unit volume [3].

Surface area is associated with key environmental processes, including contaminant adsorption, biofilm development, and organism–particle interactions [3,14]. Consequently, area-based metrics provide additional representation of surface-mediated processes and dimensional variability.

In this framework, projected areas are calculated through polygon-based delineation of particles in calibrated microscopic images. The explicit requirement of geometric closure ensures measurement consistency and analytical reproducibility across analyses [4].

2.4. Role of morphological classification

In addition to dimensional metrics, particle morphology represents an important dimension of microplastic characterization. Categories such as fibers, fragments, films, and pellets are associated with distinct sources, transport behaviors, and interaction profiles [15].

Integrating morphological classification with count-, length-, and area-based metrics enables multidimensional representation of microplastic assemblages. This approach supports structured analysis by linking particle attributes to their dimensional and categorical characteristics within modelling workflows [15,16].

2.5. Integrated framework for multidimensional assessment

The integration of particle count, physical dimensions, and morphology within a single computational framework enables a dimensionally informed characterization of microplastic contamination. Rather than replacing traditional indicators, the proposed approach situates particle counts within a broader dimensional structure, facilitating contextual interpretation across metric types [1,2].

By combining standardized image-based measurements, automated computation, and structured analytical outputs, the framework supports reproducible and transferable assessments across studies and environments. Such integration is particularly relevant for environmental modelling applications, where metric formulation directly influences aggregation behavior, comparison, and interpretation [9].

2.6. Comparison with existing microplastic analysis approaches

Current microplastic quantification workflows typically rely on particle counting as the primary metric, often complemented by independent dimensional measurements obtained through image analysis. However, these approaches are frequently implemented through fragmented procedures, where measurement, aggregation, and statistical analysis are conducted separately. This fragmentation limits reproducibility, complicates comparison across studies, and reduces the ability to integrate dimensional and categorical information.

Recent efforts have emphasized the importance of harmonization and standardization in microplastic analysis, particularly regarding measurement protocols and reporting formats. Nevertheless, most existing approaches do not provide a unified structure that simultaneously integrates count-based, length-based, and area-based metrics with morphological classification and multilevel aggregation.

Several computational tools have been developed to support specific stages of microplastic analysis, including spectral identification, particle classification, image processing, and data harmonization. Recent efforts have increasingly emphasized the importance of standardized, reproducible, and FAIR-compliant computational workflows capable of supporting interoperability among datasets and analytical platforms [11,12]. Among the most widely used platforms are Open Specy and siMPle, which have contributed significantly to spectral classification, reference library development, and workflow standardization. Recent studies have also emphasized the need for harmonized and reproducible analytical workflows capable of improving interoperability across microplastic datasets and analytical platforms [10]. However, these systems primarily focus on specific analytical tasks and do not provide a unified environment for simultaneous computation of count-based, length-based, and area-based metrics integrated with morphological stratification and exploratory multivariate analysis.

In contrast, the proposed framework operates as a unified analytical system in which all metrics are computed from the same particle-level dataset, ensuring internal consistency and enabling direct comparison between aggregation schemes. This integration represents a key methodological advancement, as it allows structural differences between metrics to be explicitly evaluated rather than implicitly assumed. A summary comparison between existing approaches and the proposed framework is presented in Table 1.

The comparison demonstrates that existing tools provide valuable functionality for specific analytical stages, particularly spectral identification, particle management, and workflow standardization. However, they generally require external processing steps to integrate dimensional

Table 1
Comparison between existing microplastic analysis approaches and the proposed framework.

Feature	Traditional Workflows	Open Specy	siMPle	IMPa Framework
Particle counting	✔	✖	✔	✔
Length-based metrics	Partial	✖	Partial	✔
Area-based metrics	Partial	✖	Partial	✔
Morphological stratification	Limited	Partial	Partial	✔
Multilevel aggregation	Manual	Limited	Limited	Automated
Integrated metric computation	✖	✖	✖	✔
Embedded PCA analysis	✖	✖	✖	✔
Unified analytical workflow	✖	✖	✖	✔
Open-source availability	Variable	✔	✔	✔
Reproducible computational pipeline	Limited	Moderate	Moderate	High

aggregation, morphological analysis, and multivariate exploration. In contrast, the proposed framework was designed as an end-to-end computational environment in which particle-level measurements, dimensional metrics, aggregation procedures, and exploratory statistical analyses are generated from a single structured dataset. This integrated architecture reduces workflow fragmentation and enhances methodological reproducibility.

3. System overview

The proposed computational framework follows a modular architecture designed to operationalize the multidimensional assessment described above. The system operates as a structured pipeline that begins with particle-level information obtained from calibrated microscopy images and progresses through data normalization, metric computation, aggregation, and analytical visualization.

Each microplastic particle is treated as an independent computational entity containing length, projected area, morphological class, and sampling metadata. These attributes are propagated through processing modules responsible for computing count-based concentration (MP), cumulative length metrics (IMPa-L), and projected area metrics (IMPa-A). The resulting indicators are then aggregated by sampling unit and morphological category.

The architecture also includes analytical modules for dimensional distribution analysis, categorical decomposition, matrix-based structural representation, and exploratory multivariate analysis. This modular design supports reproducible data processing and allows additional analytical components to be incorporated without modifying the core computational structure.

Fig. 1 presents an overview of the computational architecture and data processing workflow implemented in the framework.

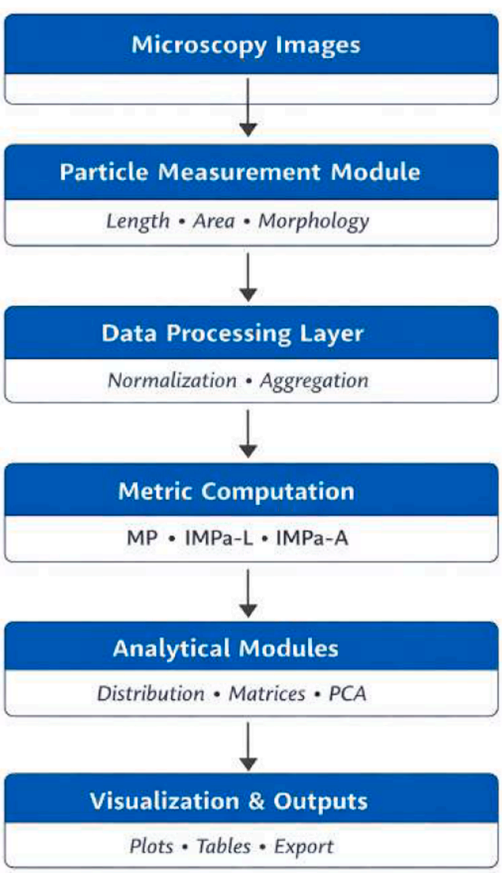

Fig. 1. Computational workflow of the IMPa framework for integrated microplastic data analysis.

4. Methodology

4.1. Methodological overview

The proposed methodology establishes a structured procedure for deriving count-based and dimension-based microplastic indices from particle-level measurements. While numerous studies have addressed particle counting and dimensional descriptors, these elements are often implemented independently or through fragmented analytical workflows [1,2].

In this study, image-based measurement, index calculation, and analytical outputs are defined as a consistent methodological sequence, enabling direct comparison between count-based and dimension-based aggregation schemes. This approach supports dimensionally informed interpretation while preserving the independence of each metric.

The methodological formulation aligns with recent efforts toward harmonized microplastic assessment, particularly in the context of environmental modelling, where metric definition directly influences aggregation behavior and interpretability [5,9].

4.2. Image acquisition and calibration

Microscopic images were acquired under controlled optical conditions to ensure consistency across samples. Prior to particle measurement, each image was spatially calibrated using a reference scale, enabling conversion from pixel-based measurements to metric units (μm). Accurate calibration is essential for dimension-based microplastic analysis and has been emphasized as a key procedural step in recent methodological studies [4,2].

Calibration parameters were stored and consistently applied across images from the same sampling campaign, minimizing systematic bias and ensuring comparability among sampling units.

4.3. Image-based microplastic characterization

4.3.1. Length measurement

Microplastic length was measured by tracing the main particle axis using consecutive points, allowing accurate representation of curved or irregular particles. This procedure is consistent with image-based methodologies commonly applied to elongated microplastics, particularly fibers [15,7].

Cumulative microplastic length was calculated by summing individual particle lengths within each sample. Length measurements were expressed in micrometers (μm) and subsequently normalized by sample volume, enabling standardized comparison across samples.

4.3.2. Area measurement

Projected particle area was determined through polygon-based delineation of the visible particle outline. Area computation was restricted to explicitly closed polygons, ensuring geometric consistency and measurement reproducibility. Area-based metrics have increasingly been incorporated as dimensional descriptors of microplastic assemblages, particularly for non-fibrous particles such as fragments, films, and pellets [3,14].

Projected areas were expressed in square micrometers (μm^2) and aggregated per sampling unit prior to normalization by volume.

4.3.3. Morphological classification

Each microplastic particle was classified according to morphological categories commonly adopted in recent studies, including fibers, fragments, films, pellets, and others [15,16]. Morphological classification provides additional structural information regarding particle geometry and categorical characteristics.

By linking morphology to both length- and area-based metrics, the methodology enables multidimensional aggregation within a unified computational workflow.

4.4. Definition of microplastic indices

4.4.1. Count-based concentration

The traditional microplastic concentration was calculated as:

$$MP = \frac{N}{V} \tag{1}$$

Where Nrepresents the number of detected microplastic particles and $V$is the sample volume (L). Although widely reported, this metric does not explicitly incorporate dimensional attributes, allowing it to be complemented by alternative aggregation schemes [1,2].

4.4.2. Length-based microplastic index (IMPa-L)

The length-based index (IMPa-L) was defined as:

$$IMPa - L = \frac{\sum_{i=1}^{n} L_i}{V} \tag{2}$$

where $L_i$represents the individual length of the i-th particle (µm). Length-based indices function as dimensionally descriptive proxies for cumulative particle extension within a sample [8,7].

4.4.3. Area-based microplastic index (IMPa-A)

The area-based index was defined as:

$$IMPa - A = \frac{\sum_{i=1}^{n} A_i}{V} \tag{3}$$

where A_i denotes the projected area of the i-th microplastic particle (µm^2). Area-based normalization has been recognized as a relevant step for incorporating surface-related dimensional variability into microplastic quantification [3,14]. For clarity and consistency, the principal symbols and notation used throughout the framework are summarized in Table 2.

4.5. Data aggregation and stratification

Indices were calculated at multiple analytical levels, including per sampling unit, per morphological category, and globally across the dataset. Stratified aggregation enables structured comparison between aggregation schemes and supports evaluation of metric behavior across categorical groupings, as recommended in recent harmonization efforts [5].

4.6. Analytical procedures and graphical outputs

Following index computation, a structured set of analytical procedures was applied to explore microplastic datasets across multiple levels of aggregation. The analytical structure was designed to support descriptive, comparative, and structural assessment, ensuring that different aggregation schemes could be evaluated consistently.

Table 2
Summary of symbols and notation used in the proposed framework.

Symbol	Description	Unit
MP	Microplastic concentration based on particle count	items L^{-1}
IMPa-L	Length-based microplastic index	µm L^{-1}
IMPa-A	Area-based microplastic index	µm^2 L^{-1}
n	Total number of particles in a sample	–
Li	Length of particle i	µm
Ai	Projected area of particle i	µm^2
V	Sample volume	L
Σ	Summation operator	–
PC1	First principal component	–
PC2	Second principal component	–

4.6.1. Descriptive analysis of particle dimensions

Descriptive statistical analyses were applied to individual particle measurements to characterize the distribution and variability of microplastic dimensions. Length and projected area data were explored using distribution plots and boxplots, allowing identification of dispersion patterns, asymmetry, and extreme values.

These analyses provide structural context for aggregated indices by highlighting dimensional heterogeneity within samples.

4.6.2. Spatial comparison of aggregated indices

Aggregated indicators—particle count (MP), length-based index (IMPa-L), and area-based index (IMPa-A)—were compared across sampling units using comparative plots. All indices were normalized by sample volume to ensure consistency among samples.

This procedure enables direct comparison between aggregation schemes and illustrates how different metric formulations may produce structurally distinct outputs.

4.6.3. Morphological decomposition of microplastic indicators

To evaluate the contribution of particle types to aggregated outputs, indices were decomposed by morphological category. Both absolute and relative (percentage-based) contributions of fibers, fragments, films, pellets, and other categories were analyzed for MP, IMPa-L, and IMPa-A. Morphological decomposition supports evaluation of how categorical attributes modulate aggregation behavior within each metric type.

4.6.4. Matrix-based structural analysis

Matrix representations (sampling unit × morphology) were constructed for count-, length-, and area-based indices. These matrices provide a structured view of metric behavior across categorical and spatial dimensions, facilitating identification of dominant structural patterns that may not be evident in one-dimensional plots. Matrix-based analysis supports comparative evaluation of aggregation schemes within a unified computational structure.

4.7. Reproducibility and computational implementation

All methodological procedures described in this study were implemented within an open-source computational framework developed in Python. The framework automates particle-level processing, index computation, aggregation procedures, graphical outputs, and exploratory statistical analyses, thereby reducing user-induced variability and improving reproducibility across datasets. Reproducible and transparent computational workflows have become increasingly important in microplastic research, particularly in response to recent efforts aimed at data harmonization, interoperability, and FAIR-compliant analytical practices [4,5,12].

To summarize the methodological workflow, Algorithm 1 presents the sequence of operations implemented within the framework, from calibrated particle measurements to generation of aggregated indicators and analytical outputs.

The algorithm summarizes the sequential integration of measurement, normalization, aggregation, stratification, visualization, and exploratory multivariate analysis within a unified computational workflow. This structure ensures methodological consistency while preserving flexibility for future extensions, including incorporation of additional metrics, analytical modules, and environmental monitoring applications.

5. System architecture and software implementation

5.1. Design principles and objectives

The software implementation was guided by three core principles:

(i) reproducibility of microplastic assessment workflows,

Algorithm 1
Workflow implemented in the IMPa framework.

Input:
Calibrated particle dataset containing:
particle length (Li)
projected area (Ai)
morphological classification
sampling metadata
sample volume (V)
Output:
MP (count-based concentration)
IMPa-L (length-based index)
IMPa-A (area-based index)
aggregation matrices
graphical outputs
PCA results
Step 1: Import particle-level dataset.
Step 2: Validate measurements and associated metadata.
Step 3: Compute count-based concentration (MP).
Step 4: Compute length-based index (IMPa-L).
Step 5: Compute area-based index (IMPa-A).
Step 6: Aggregate indices by sampling unit.
Step 7: Aggregate indices by morphological category.
Step 8: Generate dimensional distribution analyses.
Step 9: Generate comparative plots and matrix representations.
Step 10: Perform exploratory Principal Component Analysis (PCA).
Step 11: Export results, tables, and graphical outputs.

(ii) modularity to support extensibility and adaptation to different datasets, and
(iii) transparency of computational steps to facilitate scientific interpretation.

The system was designed as an operational layer that executes the methodological definitions presented in Sections 2 and 3, ensuring consistent implementation of index calculation, aggregation, and analytical processing. In this context, the software does not redefine the methodology but ensures its standardized and reproducible execution within a structured computational environment.

This design approach is consistent with current practices in scientific software development, which emphasize reproducibility, modularity, and transparency in environmental modelling applications [5,9].

5.2. Software architecture and modular structure

The software was developed in Python, a programming language widely adopted in scientific computing and environmental modelling. Its architecture follows a modular design, separating data processing, index calculation, statistical analysis, and visualization into independent components. The overall project structure and modular organization of the system are illustrated in Fig. 2.

The architecture follows a layered design, separating the data layer (particle attributes), the computational core (metric definitions and aggregation routines), and the presentation layer (visualization and interface components).

Core computational routines responsible for index calculation (count-, length-, and area-based metrics) are implemented as standalone modules, ensuring that methodological definitions remain decoupled from user interaction layers. This separation facilitates code maintenance, testing, and future extensions, while preserving methodological integrity.

Statistical and structural analyses, including data aggregation and distribution-based routines, are implemented as dedicated modules, enabling consistent application across different datasets. Visualization routines are similarly modularized, allowing standardized graphical outputs without interfering with underlying calculations. The IMPa software is openly available via GitHub (https://github.com/edinelsonsaldanha-cloud/Microplastics-Water-Calculator—IMPa-ndice-de-Micropl-stico-) and archived with a persistent DOI through Zenodo (DOI: 10.5281/zenodo.18386718), ensuring transparency, reproducibility, and long-term accessibility of the computational framework.

5.3. Implementation of dimensional and morphological metrics

Length- and area-based indices were implemented directly following the mathematical definitions presented in Section 4. Each microplastic particle is treated as an individual computational entity, with associated attributes including length, projected area, morphological class, and sampling point.

Morphological classification is stored as categorical metadata and propagated throughout the analysis workflow. This design allows stratified aggregation and comparison of metrics by particle type, supporting structured comparison of morphological contributions across metric types.

All metrics are automatically normalized by sample volume, ensuring consistency and preventing manual calculation errors. Intermediate computational steps are preserved internally, enabling traceability from raw measurements to final indicators.

5.4. Workflow automation and error minimization

To reduce operator-induced variability and computational errors, the software automates all post-measurement calculations once particle attributes have been recorded. This automation ensures that index computation, aggregation, and statistical analysis are applied consistently across samples and studies.

Automated checks are implemented to prevent invalid operations, such as area calculation from open polygons or inconsistent calibration parameters. These safeguards enhance methodological rigor and align with best practices for scientific software development in environmental research [4].

All computational steps are logged internally, allowing traceability and facilitating validation across repeated analyses.

5.5. Graphical user interface and user interaction

A graphical user interface (GUI) was developed to facilitate user interaction with the computational framework, particularly during image-based particle measurement and project management. The interface supports microplastic measurement, morphological

```
IMPa_Software/
│
├──── main.py
├──── requirements.txt
├──── README.md
│
├──── core/
│   ├──── calculo_impa.py
│   ├──── calculo_impa_area.py
│   ├──── gerar_base_pca.py
│   ├──── pca_analysis.py
│   ├──── pca_interpretacao.py
│   └──── graficos/
│       ├──── grafico_impa_por_ponto.py
│       ├──── grafico_impa_por_tipo.py
│       ├──── grafico_matriz_impa.py
│       ├──── grafico_boxplot_comprimentos.py
│       ├──── grafico_boxplot_area.py
│       ├──── grafico_impa_area_por_ponto.py
│       ├──── grafico_impa_area_por_tipo.py
│       └──── grafico_percentual_impa_area.py
│
├──── gui/
│   ├──── tela_inicial.py
│   ├──── tela_medicao.py
│   ├──── tela_resultados.py
│   ├──── tela_calibracao.py
│   ├──── tela_criacao_projeto.py
│   ├──── tela_Pontos_Amostrais.py
│   └──── main_window.py
│
├──── resources/
│   └──── logos/
│
└──── dados/
```

Fig. 2. Modular architecture and directory structure of the IMPa software. The system is organized into independent modules for core computation, statistical analysis, graphical outputs, and graphical user interface components, ensuring separation between analytical logic and user interaction layers.

classification, data editing, and visualization of analytical outputs. An overview of the interface structure and main functional modules is presented in Fig. 3.

Importantly, the GUI operates exclusively as an interface layer and does not modify the underlying computational logic. All analytical procedures, including index computation and statistical processing, are executed within dedicated core modules. This architectural separation ensures that methodological definitions remain invariant regardless of user interaction and ensures that analytical outputs remain reproducible independent of graphical interaction.

5.6. Reproducibility, availability, and versioning

The software is distributed as an open-source package under the MIT License, ensuring unrestricted use, modification, and redistribution with appropriate attribution. Source code, documentation, and example workflows are publicly available through a version-controlled repository.

To support reproducibility and long-term accessibility, stable software versions are archived and assigned a persistent Digital Object Identifier (DOI: 10.5281/zenodo.18386718). Versioning follows semantic principles, allowing precise citation of the software version used in each analysis. This practice aligns with emerging standards for software citation and reproducibility in environmental modelling research [9].

5.7. Computational environment and dependencies

The software was developed and tested using Python ($\geq$ 3.10) and relies on widely used scientific libraries for numerical computation, data handling, visualization, and graphical interaction. Dependency management is explicitly documented to facilitate installation and execution across operating systems.

By relying on open and widely supported libraries, the software ensures compatibility with diverse computational environments and supports integration into broader environmental modelling workflows.

6. Results

6.1. Demonstrative dataset and workflow

To demonstrate the operational performance of the proposed computational framework, the system was applied to a real-world dataset obtained from a tidally influenced river in northern Brazil. The dataset consisted of nine water samples collected at a fixed sampling location over short temporal intervals, ensuring natural variability in particle occurrence and size distribution while maintaining spatial

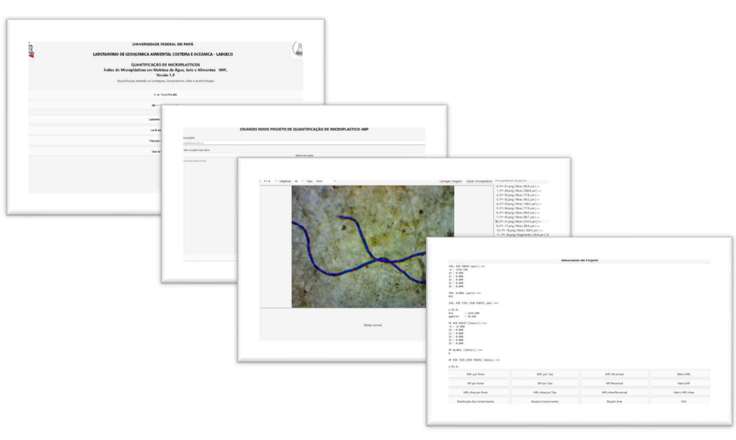

Fig. 3. Graphical user interface of the IMPa system. The interface supports project organization, image-based microplastic measurement, morphological classification, and automated generation of analytical outputs, while maintaining separation from the underlying computational core.

consistency.

The sampling location and its geographical context are presented in Fig. 4. The complete analytical workflow—from environmental sampling and filtration to microplastic measurement, morphological classification, index computation, and graphical output generation—is summarized in Fig. 5.

This dataset was not intended for environmental generalization, but rather to illustrate the framework's ability to process real measurement data under variable conditions. By applying the full computational pipeline, including dimensional quantification and morphological stratification, the system demonstrates its capacity to integrate measurement, classification, normalization, and multilevel aggregation within a unified analytical structure.

All particle attributes (length, projected area, and morphological type) were recorded and automatically processed to compute count-based concentration (MP), length-based index (IMPa-L), and area-based index (IMPa-A), each normalized by sample volume. The results generated through this demonstrative application are presented in the following subsections, emphasizing structural consistency, metric sensitivity, and analytical flexibility of the framework.

6.2. Multimetric computation

The application of the framework to the demonstrative dataset enabled simultaneous computation of count-based concentration (MP), length-based index (IMPa-L), and area-based index (IMPa-A) across all sampling units (Figs. 6–8).

MP exhibited moderate variation across sampling units, whereas dimension-based indices displayed a broader range of values. Notably, sampling units with similar MP values presented distinct IMPa-L and IMPa-A outputs.

Differences between dimensional indices were also observed. IMPa-L showed higher responsiveness to elongated particles, whereas IMPa-A exhibited stronger variation in samples containing particles with larger projected areas.

These patterns indicate that each aggregation scheme captures different aspects of the same particle assemblage, resulting in distinct numerical representations across metrics.

From a computational standpoint, the simultaneous generation of MP, IMPa-L, and IMPa-A confirms the structural coherence of the modular implementation. All indices are derived from the same particle-level dataset, ensuring internal consistency while preserving analytical differentiation.

Together, these results demonstrate that the framework is capable of:

- Computing multiple metrics from a unified measurement base;
- Preserving normalization consistency across aggregation levels;
- Representing distinct dimensional attributes within independent aggregation schemes;
- Supporting direct comparison across metrics without cross-dependence.

This multimetric computation constitutes the analytical core of the proposed system and establishes its operational capacity for dimension-integrated microplastic assessment.

6.3. Morphological stratification

The framework enables automatic stratification of computed indices according to morphological classification, allowing aggregation by particle type across all metric categories (Figs. 9–14). This functionality demonstrates the system's capacity to integrate categorical attributes with dimensional quantification within a unified analytical pipeline.

Across the demonstrative dataset, fibers represented the dominant class in count-based concentration (MP), whereas fragment contribution increased proportionally in dimension-based metrics. This shift illustrates how morphological composition interacts differently with numerical and dimensional aggregation.

When expressed in cumulative length (IMPa-L), elongated particles such as fibers contributed substantially to total values, reflecting their structural geometry. In contrast, area-based aggregation (IMPa-A) amplified the contribution of irregular fragments, whose projected surface areas disproportionately influence cumulative dimensional metrics.

The percentage-based representations further emphasize this metric-dependent modulation. Although fibers accounted for the majority of particle abundance, fragments contributed a larger share of projected area relative to their numerical proportion. This demonstrates that the framework preserves categorical stratification across metric transformations while maintaining internal consistency.

Importantly, the stratification process is fully automated and derived from the same base particle dataset. Morphological classification is stored as an attribute during measurement and propagated through all computational modules, ensuring reproducibility across aggregation levels and eliminating manual recalculation.

From a computational perspective, this capability confirms that the

Fig. 4. Location of the sampling site in a tidally influenced river in Salinópolis, northern Brazil.

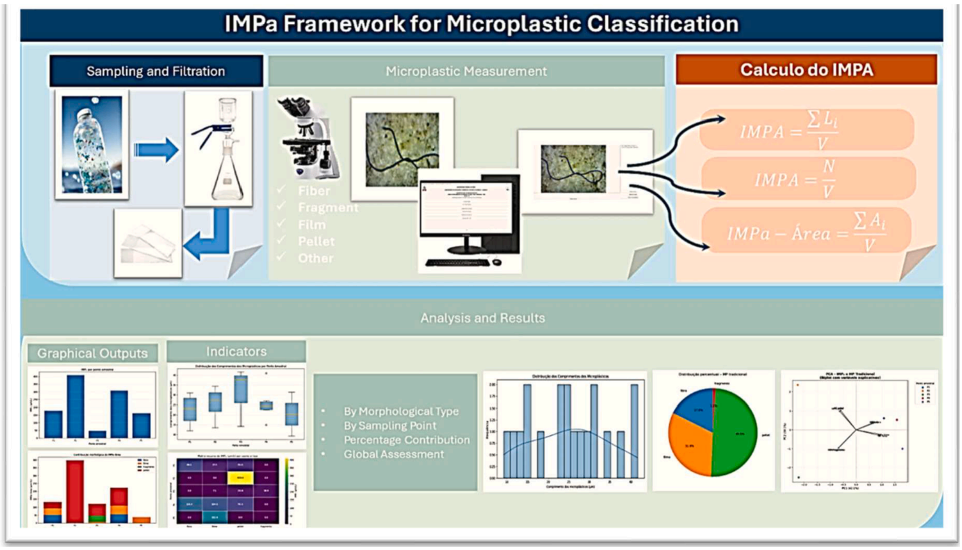

Fig. 5. Conceptual and operational framework of the IMPa system, from environmental sampling and filtration to microplastic measurement, index computation, and analytical outputs.

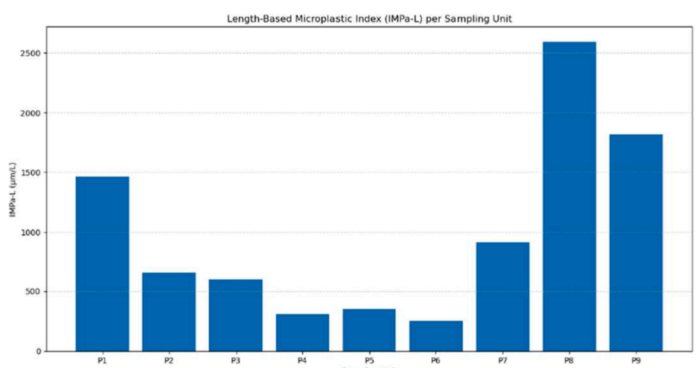

Fig. 6. Length-based microplastic index (IMPa-L, μm L^{-1}) normalized by sample volume for each sampling unit (P1–P9).

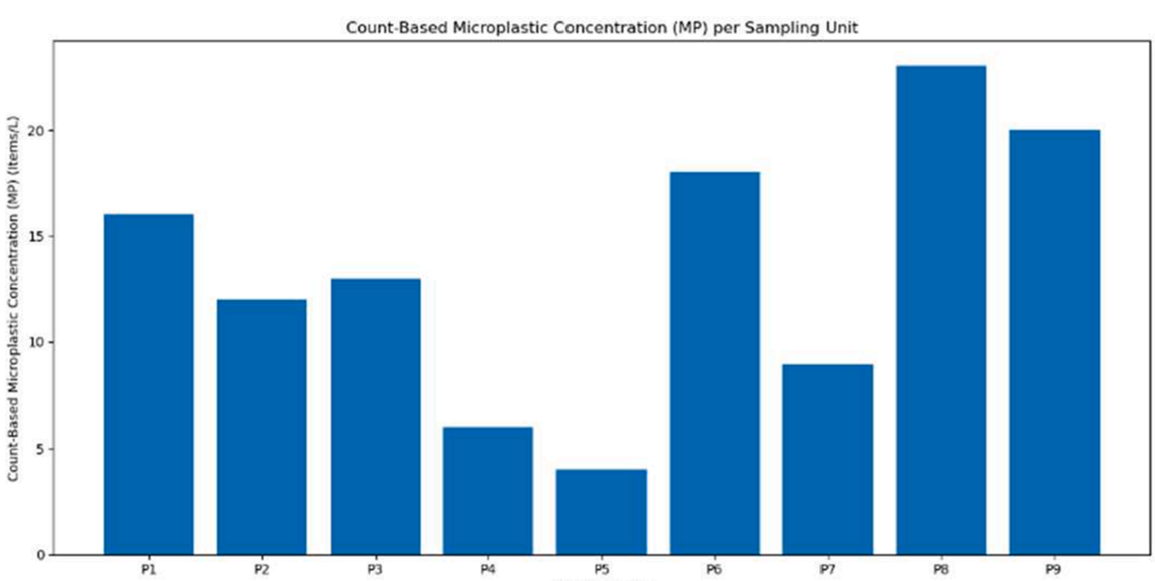

Fig. 7. Count-based microplastic concentration (MP, items L^{-1}) normalized by sample volume for each sampling unit (P1–P9).

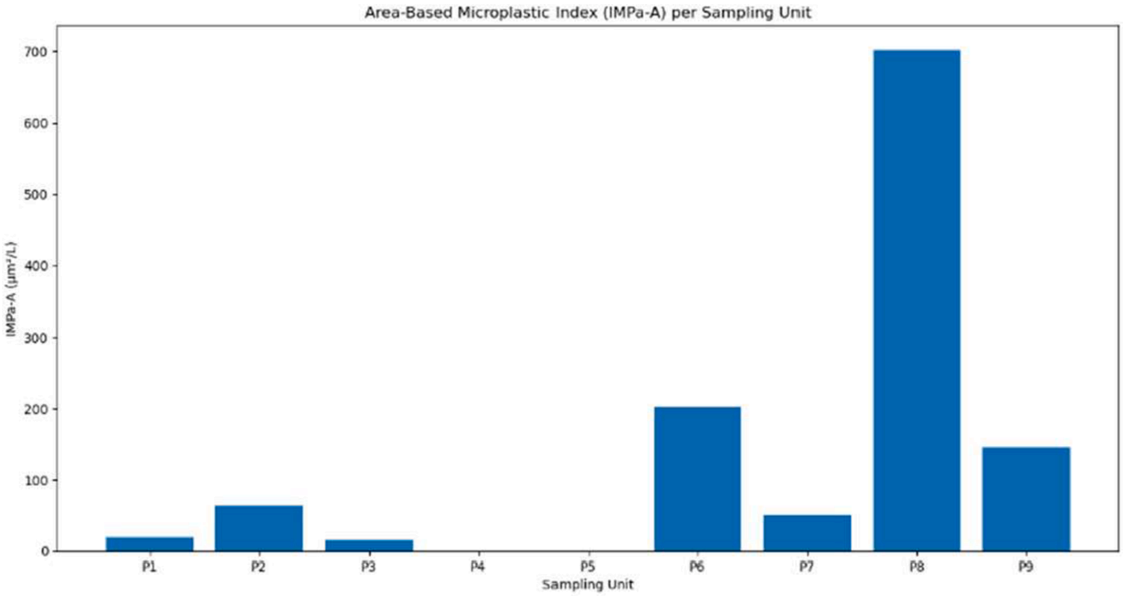

Fig. 8. Area-based microplastic index (IMPa-A, μm^2 L^{-1}) normalized by sample volume for each sampling unit (P1–P9).

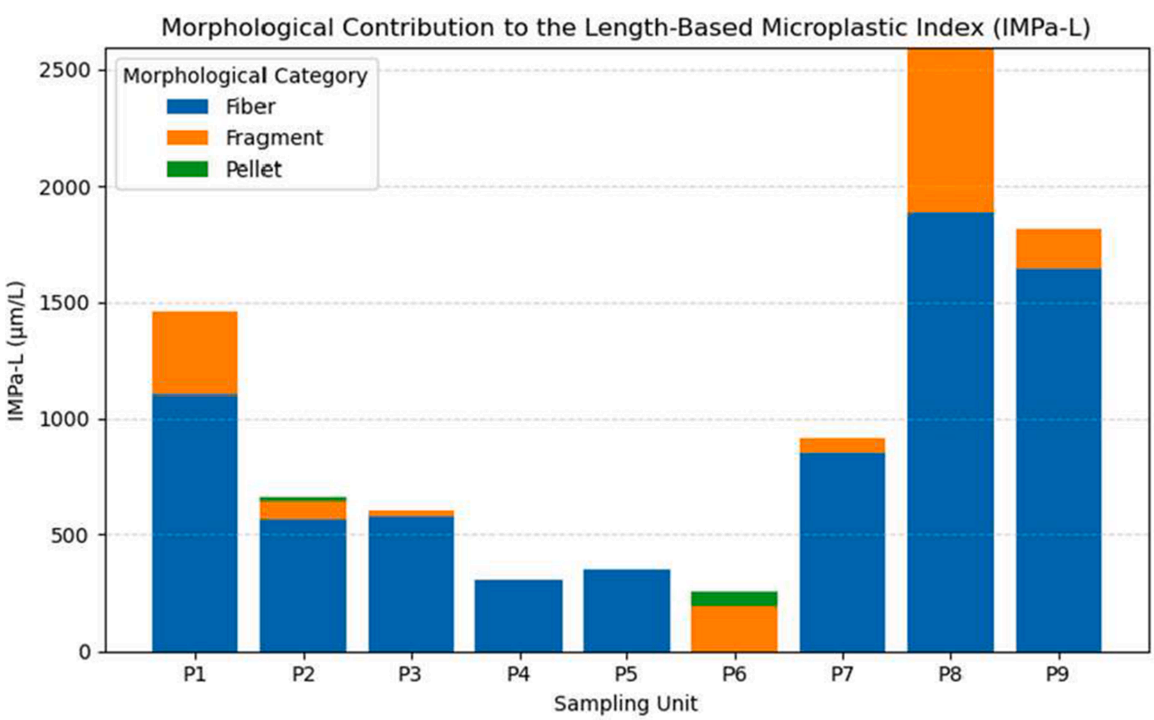

Fig. 9. Decomposition of the length-based microplastic index (IMPa-L) by morphological category across sampling units.

framework:

- Supports simultaneous categorical and dimensional aggregation;
- Maintains metric coherence across transformations;
- Allows cross-comparison between absolute and relative contributions;
- Provides scalable representation for larger datasets.

The morphological stratification module thus reinforces the framework's flexibility and analytical depth, enabling users to explore structural composition without altering the underlying computational logic.

6.4. Dimensional distributions

Beyond aggregated indices, the framework provides distribution-based analysis of particle dimensions, enabling assessment of dispersion, skewness, and extreme-value influence within the dataset (Figs. 15–16).

The boxplot representation of particle length reveals heterogeneous dispersion across sampling units. While some samples exhibit relatively narrow interquartile ranges, others display broader distributions and high-magnitude outliers. This variability illustrates the importance of evaluating dimensional spread rather than relying solely on cumulative metrics.

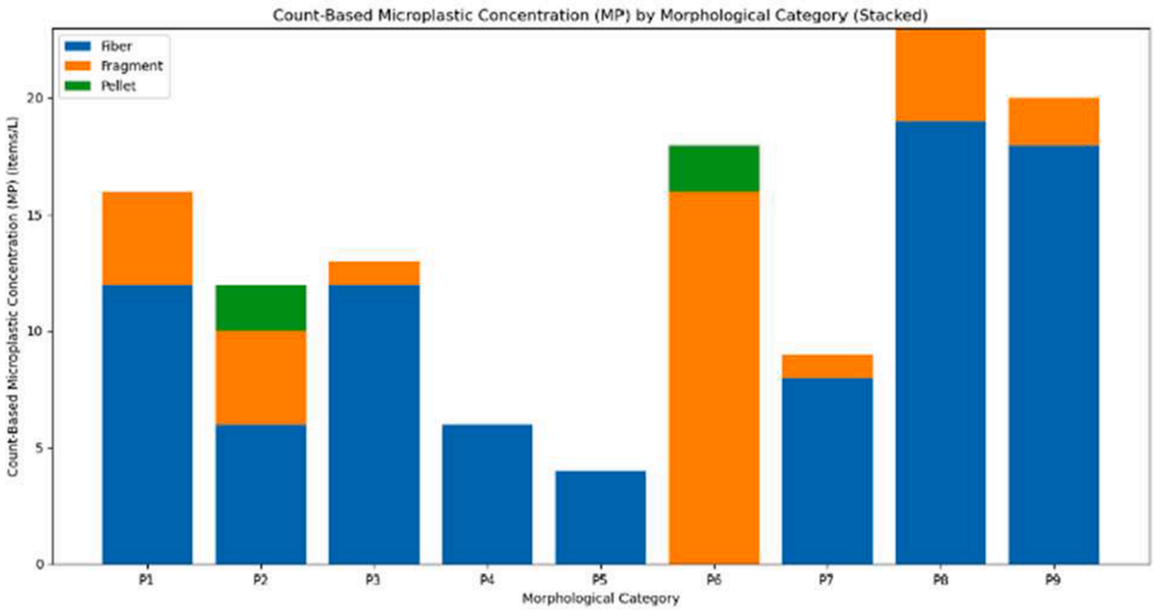

Fig. 10. Morphological decomposition of count-based microplastic concentration (MP, items L^{-1}) across sampling units.

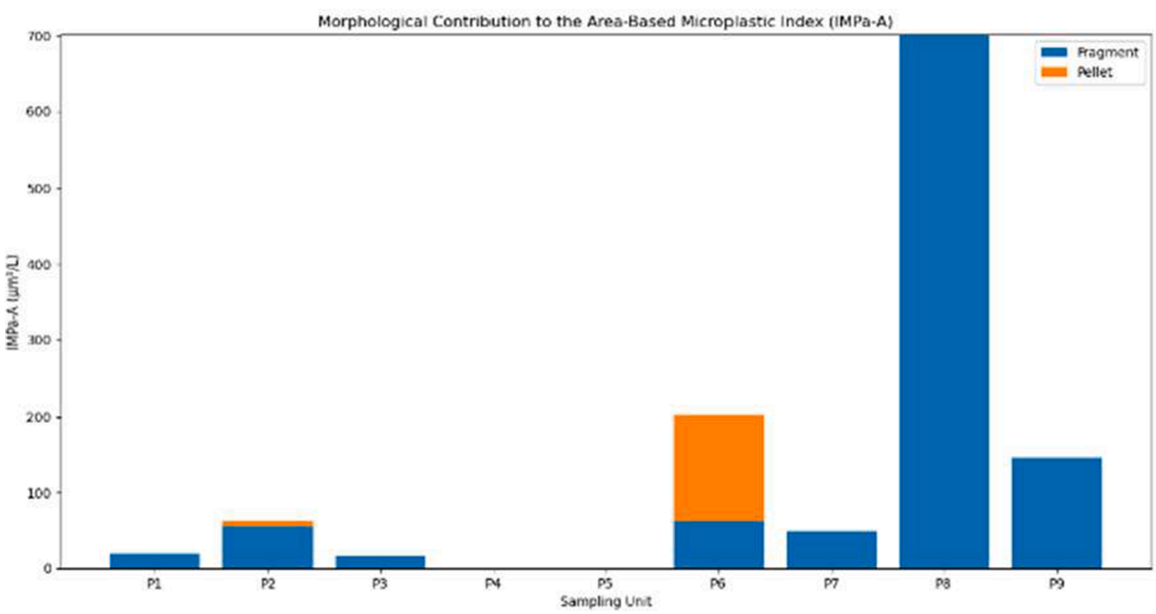

Fig. 11. Morphological decomposition of the area-based microplastic index (IMPa-A, μm^2 L^{-1}) across sampling units.

The projected area distributions show even greater asymmetry. In several sampling units, a small number of large particles contribute disproportionately to total projected area, resulting in extended upper whiskers and isolated extreme values. Such patterns confirm that area-based metrics are particularly sensitive to rare high-magnitude observations.

From an analytical standpoint, these distributional characteristics highlight a key structural property of dimensional quantification: aggregated indices may be strongly influenced by skewed particle-size distributions. The integration of boxplot visualization within the framework allows users to identify such effects and evaluate whether high index values are driven by widespread size increase or by isolated large particles.

Importantly, distribution analysis is computed directly from the same measured particle attributes used for index calculation. No secondary transformation or external statistical processing is required. This integration ensures methodological coherence between descriptive statistics and cumulative metric outputs.

The incorporation of dimensional distribution modules demonstrates that the framework:

- Supports assessment of intra-sample variability;
- Detects asymmetry and outlier-driven amplification;
- Preserves traceability between raw particle measurements and aggregated indices;
- Enables consistent exploratory statistical evaluation.

By combining cumulative metrics with dispersion analysis, the system extends beyond simple quantification and provides structural insight into the dimensional behavior of microplastic datasets.

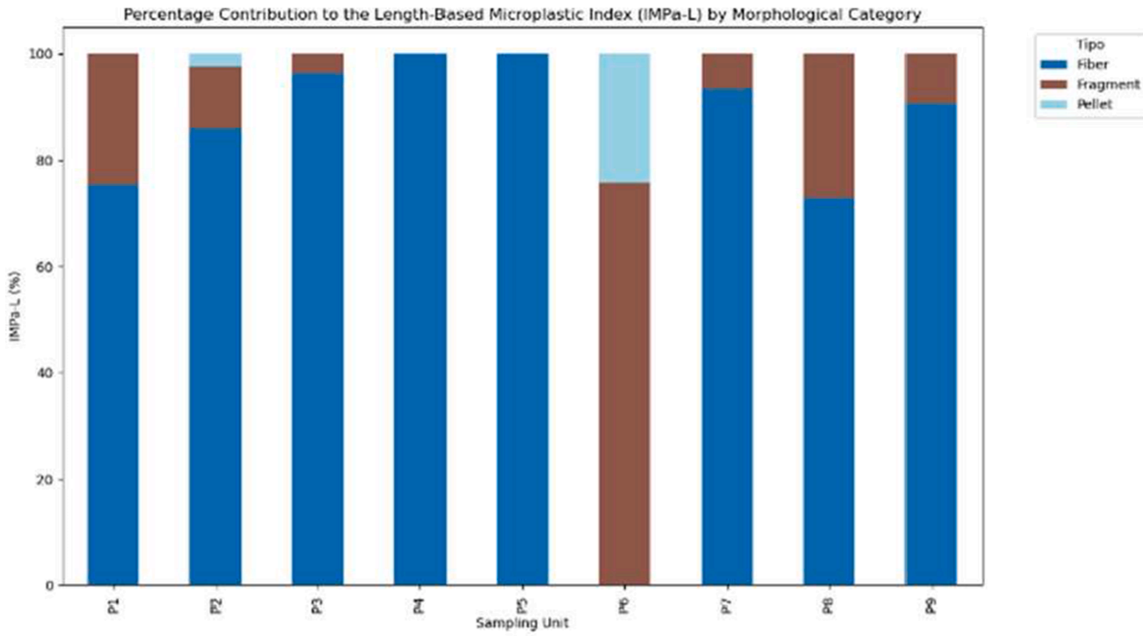

Fig. 12. Relative (percentage) contribution of morphological categories to the length-based microplastic index (IMPa-L) across sampling units.

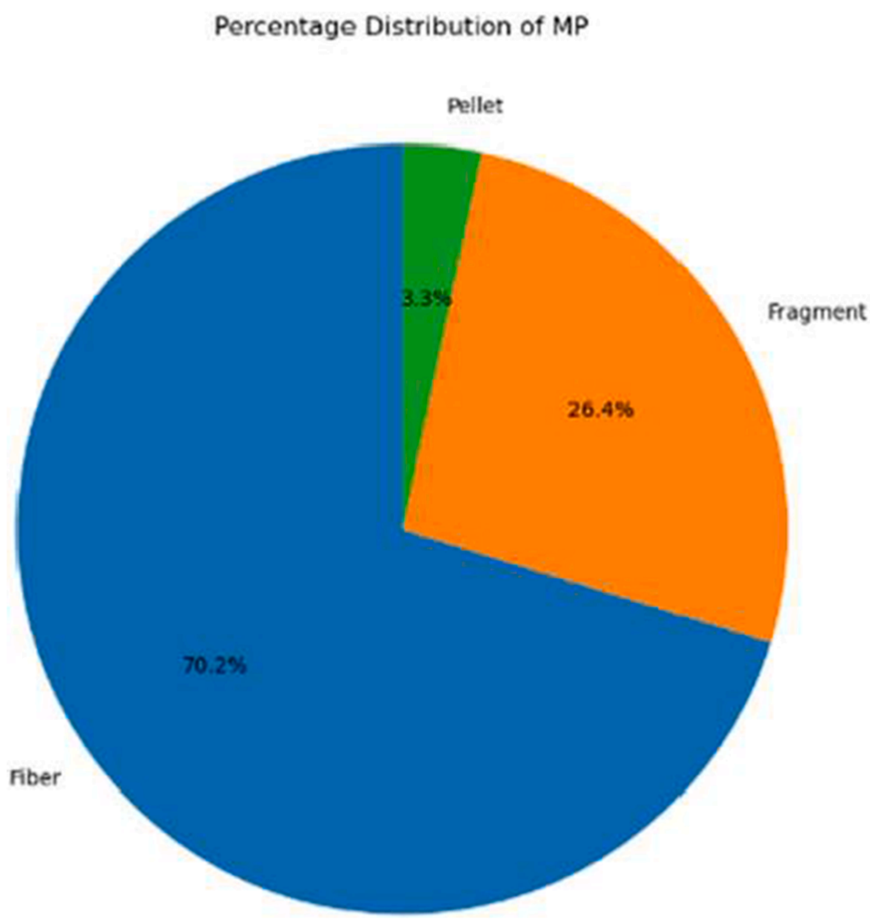

Fig. 13. Global percentage distribution of microplastic particles based on count (MP), highlighting the dominance of fibers in the study area.

6.5. Matrix-based structural representation

To further demonstrate the integrative capacity of the computational framework, matrix-based representations were generated for MP, IMPa-L, and IMPa-A across sampling units and morphological categories (Figs. 17–19).

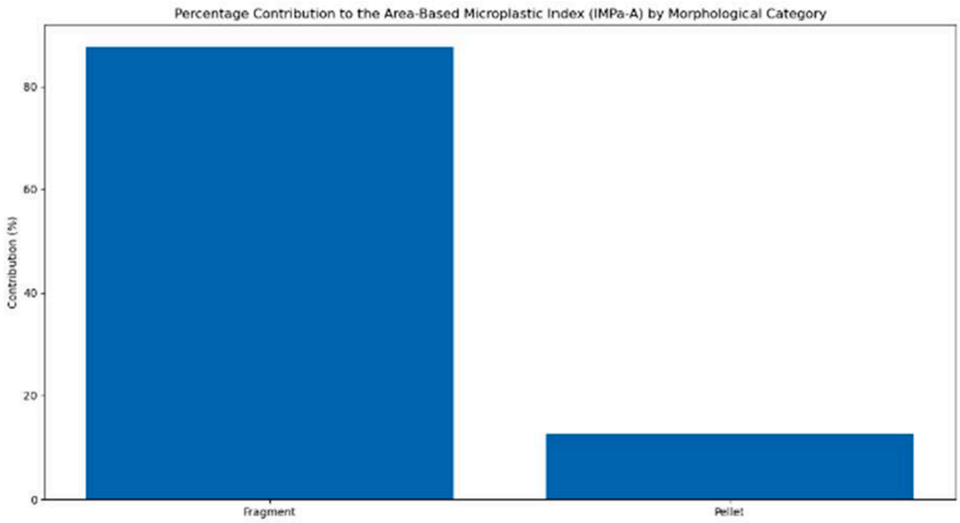

Fig. 14. Percentage contribution of morphological categories to the area-based microplastic index (IMPa-A), aggregated across all sampling units.

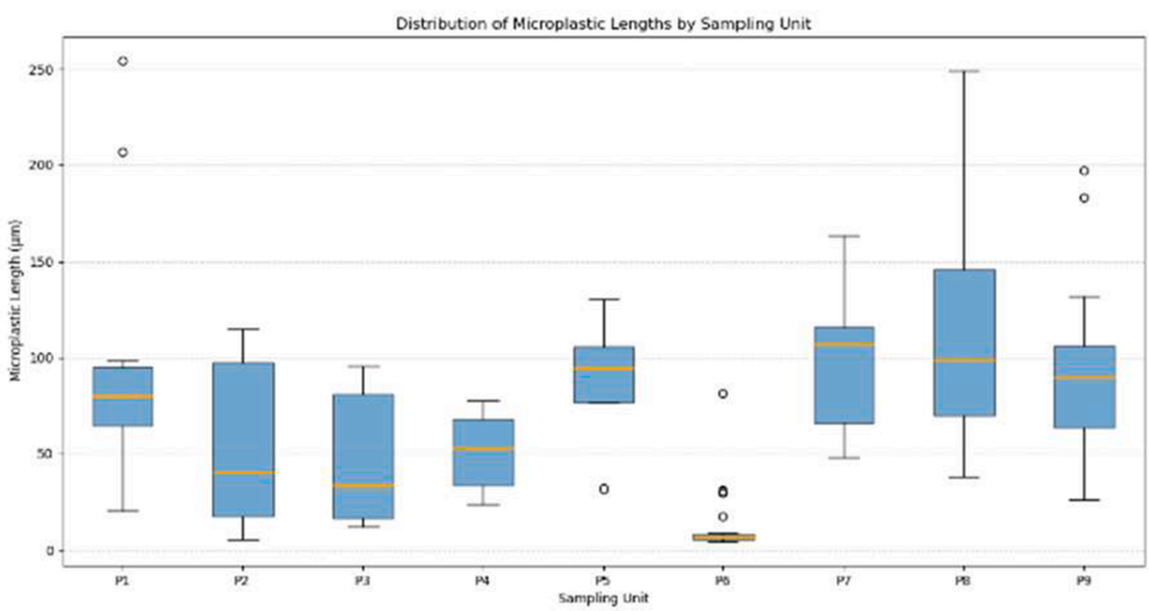

Fig. 15. Distribution of individual microplastic lengths (μm) by sampling unit, represented as boxplots showing median, interquartile range, and extreme values.

These matrices provide a structured view of the dataset by simultaneously organizing information along two primary axes: sampling unit and particle morphology. By combining categorical stratification with metric-specific aggregation, the framework enables multidimensional interpretation within a single analytical representation.

In the count-based matrix (MP), fiber dominance is evident across most sampling units, reflecting the numerical prevalence of elongated particles. However, when examining the length-based (IMPa-L) and area-based (IMPa-A) matrices, shifts in dominance patterns become apparent. Fragment contribution increases proportionally in dimensional matrices, particularly in the projected area representation.

This transition does not represent inconsistency; rather, it demonstrates metric responsiveness to morphological geometry. Because each matrix is computed from the same underlying particle dataset, differences among matrices directly reflect metric transformation rather than data variation.

The matrix module allows users to:

- Identify dominant type–location combinations;
- Compare metric behavior across aggregation levels;
- Detect structural asymmetries;
- Evaluate cross-metric coherence.

Importantly, matrix generation is fully automated within the framework. Once particle attributes are recorded, all cross-tabulated representations are produced without external processing. This ensures reproducibility and scalability for larger datasets.

From a modeling perspective, the multidimensional matrix representation confirms that the framework:

- Maintains structural traceability from raw data to aggregated outputs;
- Preserves independence between metric transformations;
- Supports high-dimensional visualization without compromising computational integrity.

The matrix module therefore consolidates the system's analytical architecture, enabling users to explore structural patterns in a coherent

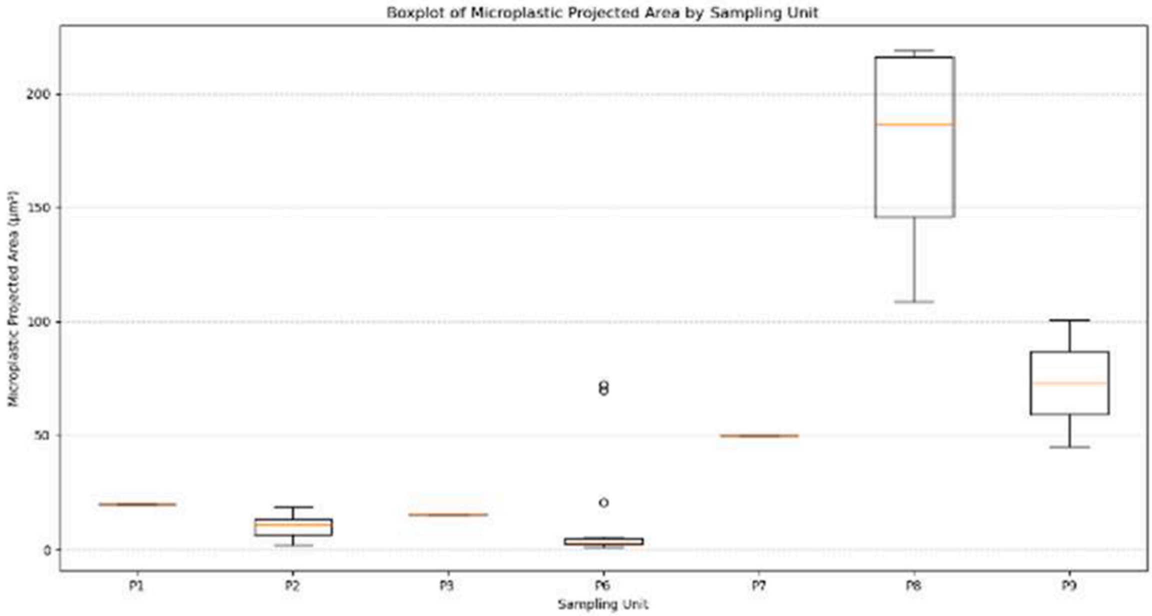

Fig. 16. Boxplot of projected microplastic area (µm^2) by sampling unit, illustrating intra-site variability and the presence of high-area particles influencing the area-based index (IMPa-A).

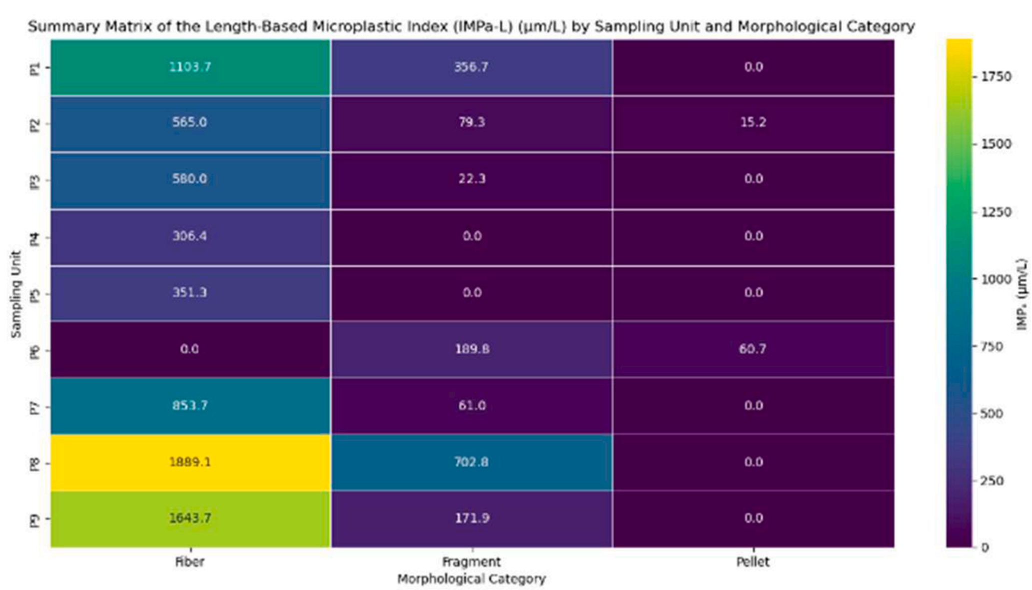

Fig. 17. Summary matrix of the length-based microplastic index (IMPa-L, µm/L) stratified by sampling unit and morphological category. Values represent cumulative length normalized by sample volume.

and reproducible manner.

6.6. *Multivariate Structural Analysis*

To demonstrate the multivariate analytical capability embedded within the IMPa computational framework, a Principal Component Analysis (PCA) was applied to the aggregated dataset combining count-based concentration (MP) and length-based index (IMPa-L), stratified by morphological category.

All variables were standardized prior to analysis to ensure equal weighting and to prevent scale-driven bias. The PCA results are presented in Fig. 20.

6.6.1. *Variance explanation and dimensional structure*

The first principal component (PC1) explained 60.9% of total variance, while the second component (PC2) accounted for 31.0%, resulting in a cumulative variance explanation of 91.9%. This high cumulative variance indicates that the multidimensional structure of the dataset can be effectively represented in a two-dimensional space.

PC1 primarily reflects a gradient associated with overall microplastic load intensity, with strong contributions from IMPa-L and MP variables across morphological categories. PC2 captures contrasts between dominant morphological structures, particularly fragment- versus fiber-associated contributions.

6.6.2. *Morphological gradients and sampling differentiation*

As shown in Fig. 20, sampling units are distributed along PC1

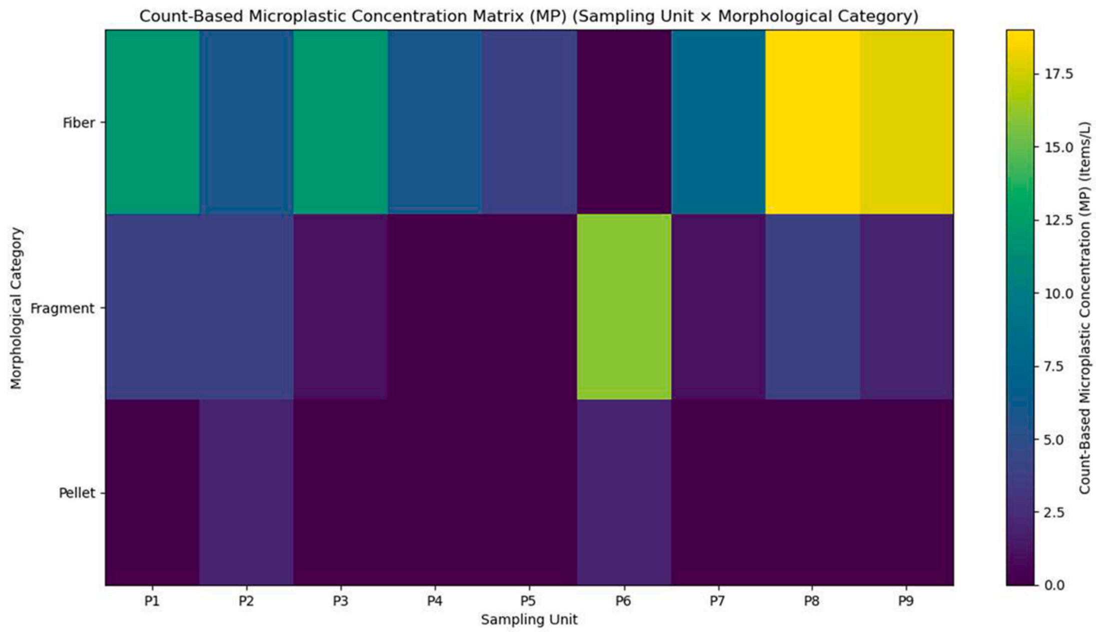

Fig. 18. Matrix representation of count-based microplastic concentration (MP, items/L) across sampling units and morphological categories.

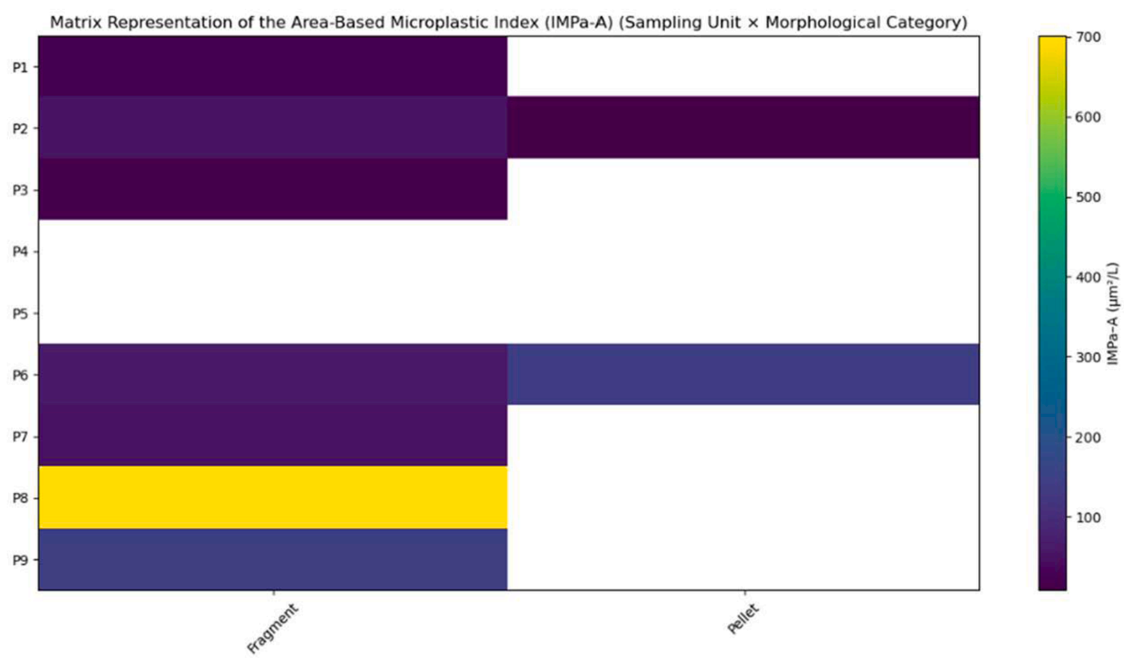

Fig. 19. Structural matrix of the area-based microplastic index (IMPa-A), highlighting the dimensional contribution of morphological categories across sampling units.

according to contamination intensity, while PC2 differentiates samples based on morphological dominance. Vectors associated with fiber-related variables (MP Fiber, IMPa Fiber) cluster along a similar directional axis, indicating strong correlation between count-based and length-based contributions for fibrous particles. In contrast, fragment-associated variables (MP Fragment, IMPa Fragment) display a distinct orientation in the multivariate space, reflecting structural differences in particle composition.

Pellet-related variables show intermediate contributions, consistent with their lower overall representation in the dataset.

This structural separation demonstrates that the framework captures not only contamination magnitude but also morphological composition patterns.

Beyond demonstrating analytical extensibility, PCA contributes to the interpretation of multidimensional microplastic datasets by revealing divergence patterns between count-based and dimension-based metrics. In particular, samples that exhibit similar MP values may occupy distinct positions in the multivariate space due to differences in particle size distribution and morphology. This distinction is relevant for environmental assessment, as it highlights cases where numerical similarity does not correspond to equivalent physical contamination burden.

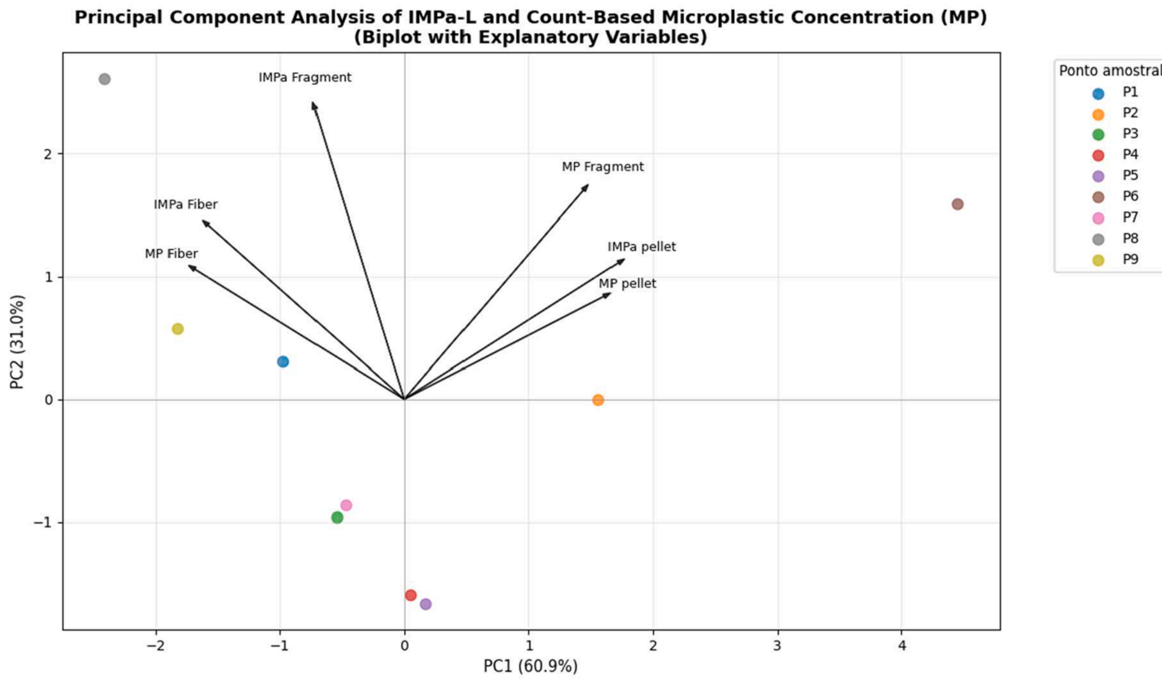

Fig. 20. Principal Component Analysis (PCA) of Length-Based Microplastic Index (IMPa-L) and Count-Based Microplastic Concentration (MP) across sampling units and morphological categories.

6.6.3. *Role of PCA within the framework*

Importantly, PCA is not presented here as an ecological inference tool, but as a demonstration of the analytical extensibility of the IMPa framework. The integration of multivariate procedures within the software architecture enables users to explore structural relationships among indices and morphological categories without external statistical processing.

By embedding PCA directly into the computational workflow, the framework supports:

- dimensionality reduction of complex microplastic datasets
- identification of dominant contamination gradients
- structural comparison among sampling units
- exploratory multivariate interpretation

This integration reinforces the framework's suitability for advanced environmental modelling applications, where multivariate relationships between particle metrics are often relevant.

7. Performance and scalability

7.1. *Framework robustness and analytical flexibility*

The demonstrative application confirms that the proposed framework operates as a coherent and modular analytical system. All outputs—including count-based concentration (MP), length-based index (IMPa-L), area-based index (IMPa-A), morphological stratification, dimensional distributions, and matrix representations—are generated from a single particle-level dataset without requiring external processing or data restructuring.

The framework maintains internal consistency across analytical levels by propagating particle attributes through normalization, aggregation, stratification, and visualization procedures. Consequently, all derived outputs remain directly traceable to the original measurements, ensuring consistency among complementary analytical representations.

Analytical flexibility is supported by the framework's ability to simultaneously compute multiple metric types, perform aggregation at different organizational levels, generate both absolute and relative representations, and integrate descriptive and multivariate analyses within a unified workflow. Because all computational routines operate on shared data structures, additional metrics and analytical modules can be incorporated without modifying existing processing procedures.

The successful application of the framework to the demonstrative dataset indicates that the proposed approach is not restricted to a specific environmental context. Instead, it provides a general computational infrastructure for dimension-integrated microplastic analysis that can be adapted to datasets of varying size and complexity.

Overall, the results demonstrate that the framework provides a reproducible, extensible, and analytically consistent environment capable of integrating measurement, aggregation, stratification, visualization, and exploratory analysis within a single workflow.

7.2. *Computational performance and scalability*

The computational scalability of the IMPa framework was evaluated using synthetic datasets ranging from 100 to 1,000,000 microplastic particles (Table 3). This expanded benchmark was designed to assess framework performance under increasingly demanding workloads and to simulate high-density datasets generated by automated image-based microplastic analysis workflows.

For each dataset size, the complete processing pipeline was executed, including JSON input reading, calculation of MP, IMPa-L, and IMPa-A

Table 3
Temporal sampling design and corresponding sampling units.

Point	Time	Number Image
P1	08:00	14
P2	10:00	11
P3	12:00	13
P4	14:00	5
P5	16:00	4
P6	18:00	11
P7	20:00	9
P8	21:00	23
P9	23:00	17

indices, aggregation by sampling unit and morphological category, normalization procedures, and output generation. Execution time and peak memory consumption were recorded for every benchmark scenario. The results are summarized in Table 4 and illustrated in Fig. 21.

7.2.1. Temporal scalability

Processing time increased consistently with dataset size, demonstrating near-linear empirical scaling throughout the evaluated range (Fig. 21a). Linear regression analysis yielded a coefficient of determination of $R^2 = 0.9991$, indicating strong agreement between execution time and dataset size. The maximum execution time observed was 64.01 s for the 1,000,000-particle dataset.

The observed behavior reflects the underlying algorithmic structure of the framework, which relies primarily on sequential traversal of particle records and aggregation routines without recursive procedures or computationally intensive optimization steps. No evidence of superlinear growth was detected within the evaluated range, supporting the conclusion that the framework exhibits approximately O(n) time complexity under practical operating conditions.

The expanded benchmark substantially strengthens the scalability assessment by extending the evaluation from moderate-sized datasets to workloads representative of high-throughput image-based microplastic analysis. These results demonstrate that the framework remains computationally efficient even when processing datasets containing one million particle records.

7.2.2. Memory scalability

Peak memory consumption also exhibited near-linear growth with increasing dataset size (Fig. 21b), with regression analysis yielding $R^2 = 1.0000$. Memory usage increased from 0.52 MB for 100 particles to 694.46 MB for 1,000,000 particles.

The observed memory profile is consistent with the storage requirements of particle-level attributes and aggregation structures, both of which scale proportionally with the number of input records. Importantly, memory allocation remained predictable across all benchmark scenarios, with no abrupt increases or evidence of inefficient memory management.

The ability to process datasets containing up to one million particles while maintaining linear memory scaling demonstrates the suitability of the framework for large-scale environmental monitoring applications and future integration with automated image-analysis workflows that generate high-density particle datasets.

7.2.3. Computational environment

All scalability experiments were performed on a 64-bit ARM-based system equipped with a Snapdragon® X 10-core X1P64100 processor operating at 3.40 GHz and 16 GB of installed RAM (15.6 GB usable). The framework was executed using Python version 3.14.2 under a standard runtime environment without parallelization or hardware acceleration. Reported execution times and memory usage therefore reflect the intrinsic efficiency of the core algorithm under typical desktop computing conditions.

Table 4
Empirical computational performance of the IMPa framework across increasing dataset sizes.

Number of Particles	Execution Time (s)	Peak Memory (MB)
100	0.116	0.52
1,000	0.128	0.72
5,000	0.373	3.47
10,000	0.610	6.94
20,000	1.308	13.89
50,000	3.230	34.74
100,000	6.511	69.39
500,000	29.983	347.24
1,000,000	64.008	694.46

7.2.4. Implications for environmental applications

The combined temporal and spatial scalability results demonstrate that the IMPa framework is computationally efficient and suitable for high-density microplastic datasets derived from automated image analysis or large-scale monitoring campaigns.

The absence of superlinear growth patterns ensures predictable performance scaling, which is particularly relevant for long-term environmental monitoring programs, multi-site comparative analyses, and automated batch processing workflows.

These findings confirm that the framework maintains methodological rigor while remaining computationally lightweight and deployable on standard desktop environments.

8. Discussion

8.1. Computational implications

The divergence observed between count-based and dimension-based metrics highlights the influence of aggregation structure on environmental interpretation. Rather than reflecting inconsistencies in the dataset, these differences arise from the mathematical properties of each metric, as count-based concentration depends solely on particle frequency, whereas dimension-based indices incorporate particle size and geometry.

The results demonstrate that microplastic quantification can be operationalized within a unified analytical framework that integrates numerical, dimensional, and categorical attributes while preserving traceability to the original particle-level measurements. Rather than replacing count-based metrics, the proposed framework enables the simultaneous computation and comparison of multiple aggregation schemes, allowing users to evaluate how metric formulation influences interpretation.

Recent discussions in environmental modeling emphasize the importance of transparency, reproducibility, and methodological consistency in computational workflows [17,18]. In this context, the proposed framework provides a structured environment in which alternative descriptors can be calculated from the same underlying dataset, ensuring direct comparability among metrics.

The observed differences among MP, IMPa-L, and IMPa-A illustrate the structural consequences of metric selection rather than environmental heterogeneity alone. This distinction is critical because count-based, length-based, and area-based indicators describe different physical dimensions of the same particle assemblage. By making these differences explicit, the framework supports more transparent interpretation of microplastic datasets and facilitates evaluation of how aggregation choices affect analytical outcomes.

8.2. Modeling relevance

Environmental contamination datasets often combine categorical attributes (e.g., morphology), dimensional attributes (e.g., size), and spatial identifiers. However, analytical tools frequently treat these dimensions independently. The present framework integrates these axes within a consistent computational structure, enabling multidimensional aggregation without manual recalculation.

Matrix representations and stratified outputs demonstrate how cross-dimensional relationships can be preserved while maintaining traceability to particle-level data. This modeling strategy aligns with recent calls for harmonized environmental indicators that maintain compatibility across metric transformations [4].

Importantly, the framework does not enforce a single metric hierarchy. Instead, it allows parallel computation of independent aggregation schemes. This flexibility supports scenario-based analysis in which metric choice can be evaluated according to research objectives rather than predetermined methodological constraints.

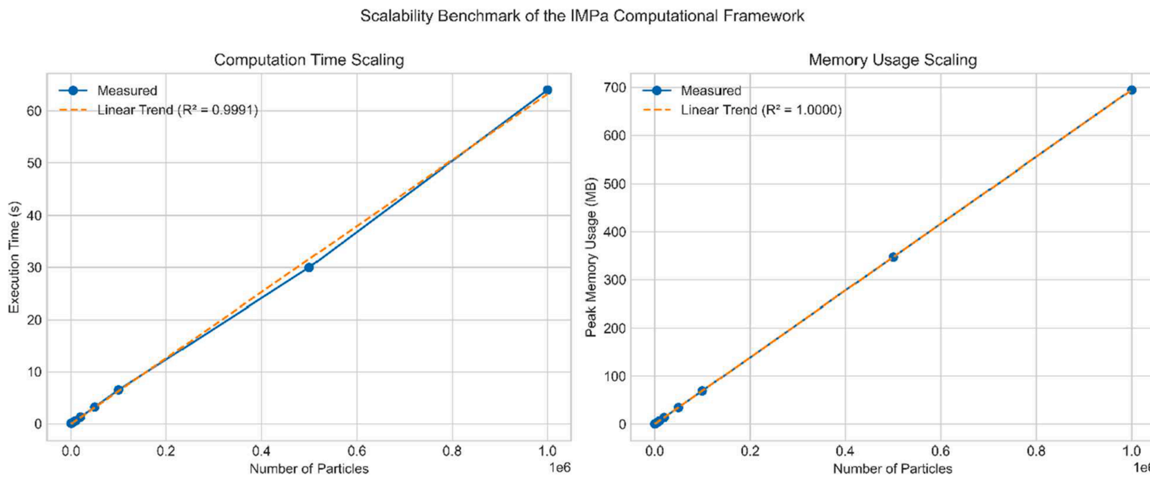

Fig. 21. Computational scalability of the IMPa framework. (a) Processing time scaling as a function of dataset size ($R^2 = 0.9991$). (b) Peak memory usage scaling ($R^2 = 1.0000$). Both results confirm linear empirical complexity (O(n)).

8.3. Analytical sensitivity and structural transparency

The results demonstrate that dimensional aggregation is inherently influenced by particle-size distributions and morphological composition. Unlike count-based concentration, which depends exclusively on particle frequency, length-based and area-based indices incorporate physical attributes that may amplify the contribution of larger or morphologically distinct particles. Consequently, samples exhibiting similar MP values may generate substantially different IMPa-L and IMPa-A outputs, revealing structural characteristics that remain undetected when contamination is assessed solely through particle counts. Recent studies have reinforced that particle size, morphology, and size-distribution structure are essential descriptors for interpreting microplastic contamination patterns beyond abundance alone [19,20].

The integration of distribution-based analyses within a unified computational environment enables direct evaluation of these aggregation effects. By combining cumulative indices with dimensional distributions, boxplots, and matrix-based representations, the framework provides a transparent view of how particle-level properties propagate through different aggregation schemes. This capability reduces the likelihood of misinterpretation arising from highly skewed particle-size distributions or isolated extreme observations, while supporting more robust interpretation of multidimensional datasets. Recent methodological developments in microplastic analysis emphasize the need for reproducible workflows, transparent data processing, and standardized analytical procedures to improve comparability across studies [10,21].

Importantly, the observed differences among MP, IMPa-L, and IMPa-A should not be interpreted as evidence of superiority of one metric over another. Rather, they reflect the distinct mathematical properties associated with each aggregation strategy. Count-based, length-based, and area-based indices capture complementary dimensions of microplastic contamination and therefore provide different, yet equally relevant, perspectives on the same particle assemblage.

From a methodological standpoint, the framework functions not only as a quantification tool but also as an exploratory analytical environment that allows users to evaluate the consequences of metric selection on environmental interpretation. By making aggregation behavior explicit, traceable, and reproducible, the proposed system contributes to greater methodological transparency and supports informed selection of metrics according to specific monitoring objectives, regulatory requirements, and research questions.

8.4. Relationship among count-based and dimension-based metrics

To further investigate whether count-based and dimension-based indicators provide redundant or complementary information, a Spearman correlation analysis was performed using the demonstrative dataset comprising nine sampling units (Fig. 22). Unlike the scalability assessment, which was conducted using synthetic datasets, this analysis employed the real dataset used throughout the framework demonstration.

The correlation matrix revealed a moderate positive association between MP and IMPa-L ($\rho = 0.567$, $p = 0.112$), a strong positive association between MP and IMPa-A ($\rho = 0.845$, $p = 0.004$), and a moderate association between IMPa-L and IMPa-A ($\rho = 0.460$, $p = 0.213$). These results indicate that the three indicators are related, but not interchangeable.

The observed relationships reflect the distinct physical dimensions represented by each metric. MP quantifies particle abundance per filtered volume (items L^{-1}), whereas IMPa-L and IMPa-A quantify cumulative particle length ($\mu m\ L^{-1}$) and projected particle area ($\mu m^2\ L^{-1}$), respectively. Consequently, increases in particle abundance do not necessarily translate into proportional increases in cumulative particle dimensions.

This distinction becomes evident when comparing sampling units

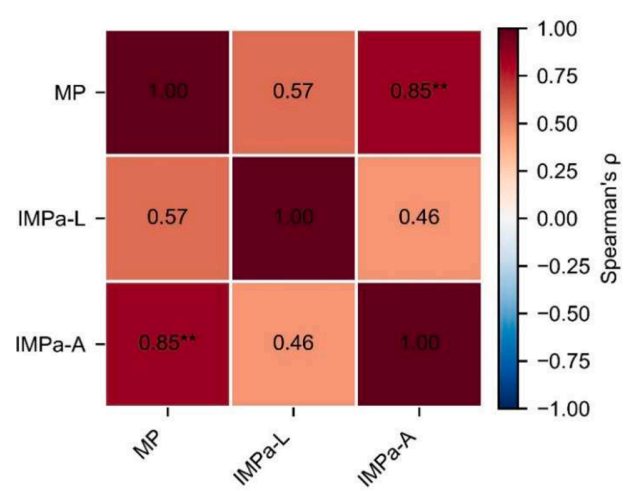

Fig. 22. Spearman correlation matrix among count-based microplastic concentration (MP), length-based microplastic index (IMPa-L), and area-based microplastic index (IMPa-A) calculated from the demonstrative dataset ($n = 9$ sampling units). The results indicate that the metrics are associated but not interchangeable, supporting the simultaneous use of count-based and dimension-based descriptors within the proposed framework.

with similar particle concentrations but contrasting dimensional burdens. For example, some sampling units exhibited comparable MP values while displaying substantially different IMPa-L and IMPa-A values, indicating that particle abundance alone may not adequately represent the dimensional characteristics of microplastic contamination. Conversely, the strong correlation observed between MP and IMPa-A suggests that sites containing larger numbers of particles frequently also exhibit greater cumulative projected area, although the relationship remains imperfect.

Importantly, the moderate and statistically non-significant correlation between IMPa-L and IMPa-A demonstrates that length-based and area-based descriptors capture different aspects of particle geometry. This finding is particularly relevant for heterogeneous assemblages composed of fibers, fragments, and pellets, where particle elongation and projected surface dimensions may vary independently.

Overall, the correlation analysis supports the conceptual rationale of the proposed framework. Rather than functioning as competing indicators, MP, IMPa-L, and IMPa-A provide complementary descriptions of the same particle assemblage. Their simultaneous computation allows the framework to characterize microplastic contamination from numerical, dimensional, and geometric perspectives, thereby improving analytical resolution and supporting more comprehensive environmental assessments.

Although the demonstrative dataset was collected in a tidally influenced river system, the objective of the present study is not to establish environmental drivers of microplastic distribution. Instead, the dataset was selected to demonstrate the analytical capabilities of the proposed framework under real monitoring conditions. Consequently, variations among sampling units may reflect local environmental processes, including hydrodynamic transport, particle retention, and tidal influence, but a detailed environmental interpretation falls outside the scope of this work. Future studies applying the framework to dedicated environmental monitoring datasets may further investigate the relationships between microplastic metrics and environmental forcing factors.

8.5. Positioning within environmental modeling literature

Open-source, modular environmental modeling tools increasingly emphasize reproducibility, extensibility, and transparent computation [18]. The IMPa framework follows this paradigm by:

- Implementing modular core routines;
- Preserving independence between data input and metric computation;
- Allowing scalable integration of additional analytical modules.

By focusing on architecture rather than environmental generalization, the present work contributes to methodological infrastructure in microplastic research. The framework provides a reproducible computational foundation upon which further metric development, sensitivity analysis, and comparative studies may be built.

From an applied perspective, the proposed framework can support environmental monitoring and decision-making processes by enabling simultaneous evaluation of particle abundance, dimensional extension, and morphological composition. In monitoring programs, this allows identification of sampling units that present similar particle counts but differ significantly in dimensional burden, which may be relevant for prioritization and comparative assessment.

Furthermore, the integration of multimetric analysis within a unified computational structure facilitates consistent reporting and cross-study comparison, addressing a common limitation in microplastic research. This capability is particularly relevant for regulatory and long-term monitoring contexts, where methodological consistency and interpretability are essential.

The IMPa framework was intentionally designed to quantify descriptors that can be directly obtained from calibrated two-dimensional images, namely particle abundance, cumulative length, and projected area. Mass-based descriptors were not included because particle mass cannot be directly measured from image data alone and generally requires additional assumptions regarding particle thickness, three-dimensional geometry, and polymer-specific density. Existing approaches that report mass-based microplastic metrics typically rely on indirect estimation procedures rather than direct measurement. Therefore, the framework is deliberately restricted to image-derived metrics that remain fully traceable to observed particle attributes.

It is important to emphasize that the PCA results presented in this study should be interpreted within the context of framework demonstration rather than environmental generalization. The demonstrative dataset was intentionally selected to illustrate the analytical capabilities of the proposed system, including dimensional aggregation, morphological stratification, and multivariate exploration. Consequently, PCA is employed as an exploratory visualization tool to highlight structural relationships among metrics and particle categories, rather than as a basis for ecological inference or predictive environmental assessment. Although the dataset is sufficient to validate software functionality and workflow integration, future applications should evaluate the framework using larger and more diverse environmental datasets to further investigate multivariate patterns under different monitoring scenarios.

9. Limitations and future developments

Despite its modular and reproducible architecture, the proposed framework presents limitations that should be considered when interpreting its outputs and extending its application.

First, the system relies on image-based particle measurement. Although calibration procedures are implemented to ensure dimensional accuracy, measurement precision may still be influenced by image quality, operator consistency, and particle delineation procedures. Future developments may incorporate semi-automated or fully automated image segmentation routines to reduce operator dependency and increase analytical throughput.

Second, the framework currently focuses on descriptors that can be directly derived from calibrated two-dimensional images, namely count-based concentration (MP), cumulative particle length (IMPa-L), and projected particle area (IMPa-A). Mass-based metrics were not included because particle mass cannot be directly measured from image data alone and generally requires additional assumptions regarding particle thickness, three-dimensional geometry, volume, and polymer-specific density. Consequently, the present framework intentionally prioritizes image-derived descriptors that remain fully traceable to measured particle attributes. Future versions may incorporate optional modules for indirect mass estimation when complementary volumetric or polymer characterization data are available.

Third, the demonstrative dataset used in this study was intentionally limited in scope, serving primarily as an operational validation of the computational workflow. Although the framework demonstrated successful scalability for datasets containing up to one million particles, further applications involving multi-site, multi-seasonal, or long-term monitoring datasets would provide additional evidence of analytical robustness under diverse environmental conditions.

Fourth, the current implementation does not include uncertainty propagation or formal error estimation within aggregated indices. Future extensions may integrate statistical resampling approaches, sensitivity analysis, or inter-operator reproducibility assessments to quantify uncertainty associated with dimensional measurements and aggregation procedures.

From a computational perspective, the modular architecture facilitates continued expansion without structural redesign. Future developments may incorporate advanced multivariate analyses, machine learning-assisted classification, automated quality-control routines, and additional environmental indicators while maintaining compatibility with the existing analytical workflow.

Importantly, although data acquisition may currently depend on manual measurement procedures, the computational framework itself is not constrained by this limitation, as demonstrated by the observed linear scalability in both execution time and memory consumption across large particle datasets.

Finally, the framework does not prescribe a single optimal metric for microplastic assessment. Instead, it provides a structured computational environment in which count-based, length-based, and area-based descriptors can be evaluated simultaneously. Future studies may further investigate the performance and applicability of these complementary metrics under specific environmental, monitoring, and ecotoxicological contexts.

10. Conclusion

This study presented IMPa, a modular and open-source computational framework for integrated microplastic analysis based on particle-level measurements derived from calibrated images. The framework combines count-based concentration (MP), cumulative length metrics (IMPa-L), and projected area metrics (IMPa-A) within a unified analytical environment, enabling direct comparison among complementary descriptors of microplastic contamination.

The proposed architecture preserves traceability from individual particle measurements to aggregated indicators while supporting morphological stratification, dimensional distribution analysis, matrix-based representations, and exploratory multivariate assessment through Principal Component Analysis (PCA). By integrating these analytical components within a single workflow, the framework reduces fragmentation commonly observed in microplastic data processing and improves computational transparency and reproducibility.

Application of the framework to a real environmental dataset demonstrated that count-based and dimension-based metrics are associated but not interchangeable, highlighting the importance of evaluating microplastic contamination from multiple analytical perspectives. The simultaneous computation of abundance-, length-, and area-based descriptors provides a more comprehensive characterization of particle assemblages than particle concentration alone.

Computational benchmarking further demonstrated linear temporal and memory scalability across datasets ranging from 10^2 to 10^6 particles, confirming the suitability of the framework for high-density image-based datasets and large-scale monitoring applications. The modular design also facilitates future incorporation of additional analytical routines and metric definitions without altering existing computational procedures.

As an open-source platform distributed under the MIT License and archived with a persistent DOI, IMPa contributes not only a set of analytical tools but also a reproducible computational infrastructure for microplastic research. The framework provides a practical foundation for environmental monitoring, comparative assessment, and future methodological developments aimed at improving the standardization, transparency, and scalability of microplastic data analysis.

Software availability

Software name: IMPa – Integrated Microplastic Quantification Framework

Developers: Edinelson Saldanha et al.

Year first available: 2026

Current version: v1.0.0

Permanent repository: Zenodo

DOI: 10.5281/zenodo.18386718

Code repository: https://github.com/edinelsonsaldanha-cloud/Microplastics-Water-Calculator—IMPa-ndice-de-Micropl-stico-

License: MIT License

Programming language: Python ($\geq$ 3.10)

Operating systems: Platform independent (tested on Windows ARM64 environment)

Dependencies: NumPy, Pandas, Matplotlib, Scikit-learn, and PyQt (see requirements.txt for full list)

Installation instructions: Provided in the GitHub repository README file

Documentation: Included in repository with example datasets and usage workflow

Availability: Open source, unrestricted academic and commercial use under MIT License

CRediT authorship contribution statement

Edinelson Saldanha: Writing – review & editing, Writing – original draft, Software, Methodology. **Vando Gomes:** Writing – original draft, Visualization, Validation, Supervision. **João Vitor Araujo:** Writing – review & editing, Writing – original draft, Software. **Ruan Alcantra Felix:** Writing – review & editing, Writing – original draft, Software, Methodology. **Maria Carneiro:** Writing – review & editing, Writing – original draft, Software, Methodology. **Ludmila Loureiro:** Writing – review & editing, Writing – original draft, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sasc.2026.200537.

Data availability

Demonstrative datasets were used for framework validation. Source code is publicly available on GitHub and archived with a DOI for reproducibility.

References

[1] A.A. Koelmans, P.E. Redondo-Hasselerharm, N.H.M. Nor, M. Kooi, Risk assessment of microplastic particles, Nat. Rev. Mater. 5 (2020) 720–739, https://doi.org/10.1038/s41578-020-00211-y.

[2] M. Simon, N. van Alst, J. Vollertsen, Quantification of microplastics: developing robust methods for particle size, mass and number, Water. Res. 213 (2022) 118109, https://doi.org/10.1016/j.watres.2022.118109.

[3] A.A. Koelmans, M. Kooi, K.L. Law, E. van Sebille, All is not lost: deriving a top-down mass budget of plastic at sea, Environ. Res. Lett. 17 (2022) 013001, https://doi.org/10.1088/1748-9326/ac3d09.

[4] W. Cowger, A.M. Booth, B.M. Hamilton, C. Thaysen, S. Primpke, K. Munno, et al., Microplastic spectral classification needs an open source community: open specy to the rescue!, Anal. Chem. 92 (17) (2020) 11669–11676, https://doi.org/10.1021/acs.analchem.0c00123.

[5] C. Gauchotte-Lindsay, E. Hendrickson, M.E. Miller, A.S. Tagg, Harmonisation of microplastic monitoring methods: a critical review, Environ. Pollut. 288 (2021) 117703, https://doi.org/10.1016/j.envpol.2021.117703.

[6] M. Kooi, E.H. van Nes, M. Scheffer, A.A. Koelmans, Upscaling microplastic removal by wastewater treatment plants: a modelling approach, Environ. Sci. Technol. 55 (15) (2021) 10226–10236, https://doi.org/10.1021/acs.est.1c00476.

[7] K. Waldschläger, H. Schüttrumpf, Effects of particle properties on the settling and rise velocities of microplastics in freshwater under laboratory conditions, Environ. Sci. Technol. 54 (4) (2020) 1958–1966, https://doi.org/10.1021/acs.est.9b05322.

[8] L.D.K. Kanhai, K. Gardfeldt, T. Krumpen, R.C. Thompson, I. O'Connor, Mass-based quantification of microplastics in environmental samples: a critical review, Sci. Total Environ. 753 (2021) 142017, https://doi.org/10.1016/j.scitotenv.2020.142017.

[9] K. Zhang, X. Xiong, H. Hu, C. Wu, Y. Bi, Y. Wu, et al., A framework for standardized microplastic data analysis and environmental interpretation, Environ. Modell. Softw. 164 (2023) 105743, https://doi.org/10.1016/j.envsoft.2023.105743.

[10] W. Cowger, A. Karapetrova, C. Lincoln, A. Chamas, H. Sherrod, N. Leong, K. S. Lasdin, C. Knauss, V. Teofilović, M.M. Arienzo, Z. Steinmetz, S. Primpke, L. Darjany, C. Murphy-Hagan, S. Moore, C. Moore, G. Lattin, A. Gray, R. Kozloski, J. Bryksa, B. Maurer, Open Specy 1.0: automated (Hyper)spectroscopy for

microplastics, Anal. Chem. 97 (32) (2025) 17345–17356, https://doi.org/10.1021/acs.analchem.5c00962.

[11] H. Hapich, W. Cowger, K. Larsen, S. Primpke, H. De Frond, C.M. Rochman, et al., Microplastics and trash cleaning and harmonization (MaTCH): a framework for harmonizing environmental debris datasets, Environ. Sci. Technol. 58 (32) (2024) 14191–14202, https://doi.org/10.1021/acs.est.4c02406.

[12] E.S. Nyadjro, J. Cebrian, T.E. Cox, Z. Wang, Y.H. Lau, A.M. Konefal, G. Turnage, T. Offner, R. Gilpin, T. Boyer, K. Larsen, P. Mickle, E. Sparks, J.A.B. Webster, Gaps and pathways towards standardized, FAIR microplastics data harmonization: a systematic review, Microplastics 5 (1) (2026) 11, https://doi.org/10.3390/microplastics5010011.

[13] M. Akhtar, Q. Zhang, L. Zhang, X. Li, J. Yang, Y. Xu, A. Hussain, Y. Li, Multilingual entity alignment by abductive knowledge reasoning on multiple knowledge graphs, Eng. Appl. Artif. Intell. 139 (2025) 109660, https://doi.org/10.1016/j.engappai.2024.109660.

[14] F. Liu, K.B. Olesen, A.R. Borregaard, J. Vollertsen, Microplastics adsorption of organic contaminants: role of surface area and particle size, J. Hazard. Mater. 402 (2021) 123778, https://doi.org/10.1016/j.jhazmat.2020.123778.

[15] L. Mai, L.-J. Bao, L. Shi, C.S. Wong, E.Y. Zeng, Traceability analysis of microplastic morphology and sources in aquatic environments, J. Clean. Prod. 334 (2022) 130180, https://doi.org/10.1016/j.jclepro.2021.130180.

[16] S. Primpke, R.K. Cross, S.M. Mintenig, M. Simon, A. Vianello, G. Gerdts, J. Vollertsen, Toward the systematic identification of microplastics in the environment: evaluation of spectroscopic methods, Appl. Spectrosc. 74 (9) (2020) 1011–1024, https://doi.org/10.1177/0003702820920266.

[17] Y.-D. Choi, J.L. Goodall, J.M. Sadler, A.M. Castronova, A. Bennett, Z. Li, D. G. Tarboton, Toward open and reproducible environmental modeling by integrating online data repositories, computational environments, and model application programming interfaces, Environ. Modell. Softw. 135 (2021) 104888, https://doi.org/10.1016/j.envsoft.2020.104888.

[18] S.R. Wilkinson, M. Aloqalaa, K. Belhajjame, M.R. Crusoe, B. de Paula Kinoshita, L. Gadelha, D. Garijo, O.J.R. Gustafsson, N. Juty, S. Kanwal, F.Z. Khan, J. Köster, K. Peters-von Gehlen, L. Pouchard, R.K. Rannow, S. Soiland-Reyes, N. Soranzo, S. Sufi, Z. Sun, B. Vilne, M.A. Wouters, D. Yuen, C. Goble, Applying the FAIR Principles to computational workflows, Sci. Data 12 (2025) 328, https://doi.org/10.1038/s41597-025-04451-9.

[19] S. Fraissinet, G.E. De Benedetto, C. Malitesta, R. Holzinger, D. Materić, Microplastics and nanoplastics size distribution in farmed mussel tissues, Commun. Earth. Environ. 5 (2024) 128, https://doi.org/10.1038/s43247-024-01300-2.

[20] A. Thomas, J. Marchand, G.D. Schwoerer, E.C. Minor, M.A. Maurer-Jones, Size distributions of microplastics in the St. Louis Estuary and Western Lake Superior, Environ. Sci. Technol. 58 (23) (2024) 8480–8489, https://doi.org/10.1021/acs.est.3c10776.

[21] E.R. Forgione, C. Ferry, B. Neumann, K.D. Good, Initiating environmental microplastics analysis: a review and planning guide with practical insights from laboratory implementation, Sci. Total Environ. 1014 (2026) 181220, https://doi.org/10.1016/j.scitotenv.2025.181220.