Advanced methodology for nanoparticle detection in spICP-ToF-MS time series

Mickaël Tharaud[1], Paul Boniol[2]

1. Université Paris Cité, Institut de physique du globe de Paris, UMR CNRS 7154
2. Inria, ENS, PSL Université, UMR CNRS 8548

1. Introduction

The Anthropocene era is marked by human activities releasing contaminants in the environment[1]. Among these contaminants, nanoparticles (NPs) have increasingly drawn the attention of scientists[2,3]. Due to their very high specific surface area and small size (nm to µm), these NPs are highly reactive particles[4]. The field of environmental nanogeochemistry has recently emerged to study their biogeochemical fate, behavior, and impact within the critical zone[5]. This field particularly focuses on assessing the flow of these NPs, determining their origin (natural vs. anthropogenic), understanding their role (vector of trace metals), and their transformation (homo- vs. hetero-aggregation, dissolution) within environmental ecosystems[6].

To that end, most studies employ inductively coupled plasma mass-spectrometry (ICP-MS) and derivated technologies such as single-particle ICP-MS to detect and characterize NPs within samples. spICP-MS has proven promising in our previous studies on the fate of NPs in aquatic systems, where we demonstrated that land use and hydrological variability affect the biogeochemical properties of rivers, which drive the flows and stability of NPs[7–10]. However, the source of NPs in these ecosystems remains unclear due to the lack of multi-element monitoring. A new generation of time-of-flight mass spectrometer used in single-particle mode (spICP-ToF-MS) has since proved highly effective, offering more complete information on NPs[11–13]. This cutting-edge instrument, **available exclusively at IPGP in France since 2022**, enables real-time detection and elemental characterization of individual NPs, providing unparalleled insights into their composition and size distribution[14,15]. In practice, continuously introducing an environmental sample (usually aqueous) produces element-specific time series consisting of a background noise and peaks corresponding to individual NPs (Figure 1). The main challenge is therefore the accurate detection of NPs events within the time series, distinguishing them from the background noise, followed by their accurate identification thanks to their elemental composition. Current methods use either threshold-based peak detection algorithms, often not robust to variation of peak prominence, shape, and background noise level over time and element[16], or advanced background noise modeling, which require *a priori* knowledge of noise distribution[17]. Coincident peaks, where multiple NPs events occur in close succession, and saturated peaks, where the signal produced by relatively large NPs exceeds the detector's capacity, are not yet adequately addressed (Figure 1). For identification of NPs families, hierarchical clustering is so far preferred by environmental nanogeochemists[13–15], but it requires manual adjustment of several meta-parameters, such as the distance threshold for merging.

Advances in machine and deep learning algorithms (hereafter refereed to as artificial intelligence, or AI) have improved detection capabilities[18], revealed hidden relationships[19], and classified NPs by their properties[20]. However, their application to spICP-ToF-MS time-series remains unexplored. This project aims to fill this gap by building a data processing and analysis pipeline to **offer the environmental nanogeochemistry community with a tailored and standardized methodology** to effectively detect, characterize and subsequently identify NPs in a timely manner. To do so, we must tackle the following challenges: (1) **accurately detect NPs** peaks within the time series despite variations, (2) **identify the NPs families with minimum prior knowledge** based on their elemental composition to understand their origin, role and transformation.

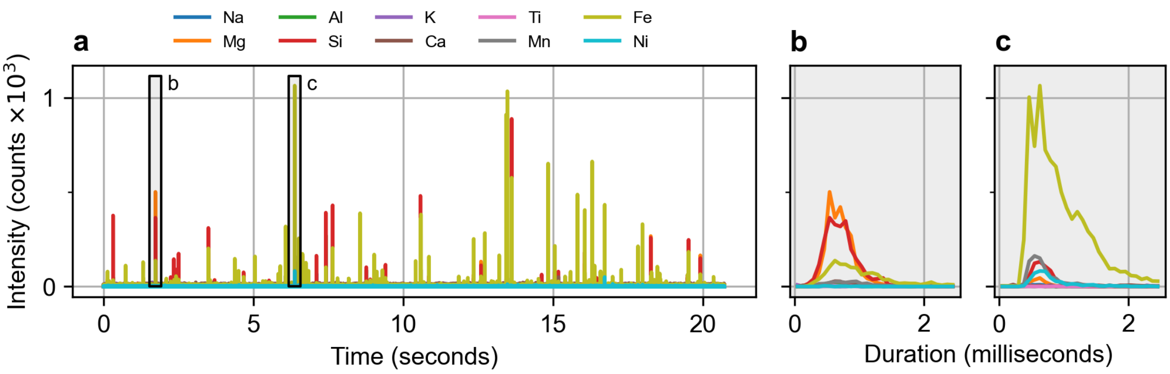

Figure 1: **a.** Part (20 seconds) of an spICP-ToF-MS measurement for a 100-fold diluted surface water sample containing NPs. **b.** Zoom on a single NP-generated peak of **a.** with spikes in Mg, Si and Fe. **c.** Zoom on a single NP-generated peak of **a.** with spikes in Fe, Mn, Si, Ni and Mg.

2. Objectives

The project main objective is to **develop, validate and provide an advanced AI-based methodology** to enhance the accuracy and precision of NPs peak detection and clustering in spICP-ToF-MS time series. We will:

A. Conduct a systematic evaluation and comparison of various AI models to determine the most effective approach(es) for NPs peak detection within the time series,
B. Design and implement advanced methods for the classification and handling of overlapping and saturated peaks, ensuring accurate representation and analysis of complex peak structures,
C. Develop robust clustering algorithms to classify and analyze peaks based on their elemental composition, facilitating the understanding of NPs characteristics and their environmental implications,
D. Create an integrated platform that encompasses tools for spICP-ToF-MS time series visualization, data processing, and comprehensive analysis, providing a standardized and user-friendly interface for researchers.

3. Method and Timetable

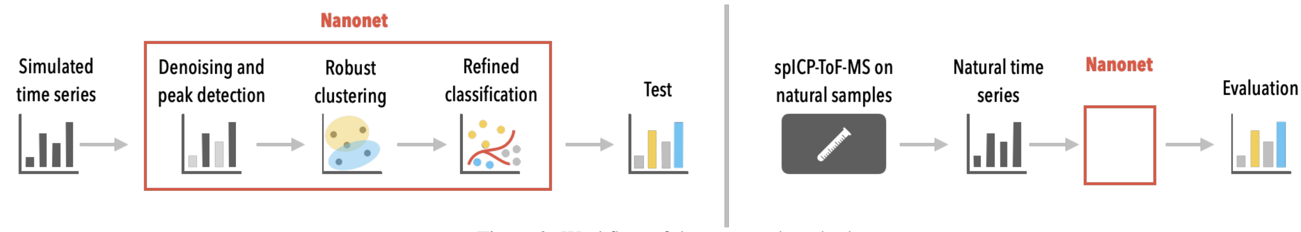

Figure 2: Workflow of the proposed method.

The use of simulated data, with known peak number, position, and characteristic and various noise levels, will provide a valuable resource for training and validating ML/DL models. This will facilitate the development and testing of our *Nanonet* pipeline before applying it to real-world data (Figure 2).

- A1. Generate simulated spICP-ToF-MS time series with varying (known) attributes for training and testing.
- A2. Develop and test a basic threshold-based algorithm for robustly detecting the largest peaks.
- A3. Implement a ML/DL model for peak recognition on A1 data; recognize previously-undetected peaks.
- A4. (if A3 weak) Train a deep-learning model (U-net-like) to denoise A1 data. Combine with A2 model.
- B1. Develop algorithms to classify detected peaks into isolated, overlapping, and saturated categories.
- B2. Implement unsupervised clustering algorithms to identify NP families with minimal prior knowledge.
- C1. Collect and preprocess real spICP-ToF-MS data, obtain successful detection and classification results.
- C2. (if C1 weak) Fine-tune the trained models on real data and/or retrain the models with A1 data.
- D1. Dissiminate the developed tools and methodologies to the community through a paper and presentations.
- D2. Implement an integrated platform for data analysis and visualization, available on a public repository.

References

(1) Naidu, R. et al. Emerging contaminants in the environment: Risk-based analysis for better management. *Chemosphere* **2016**, *154*, 350–357.(2) Sharma, V. K. et al. Natural inorganic nanoparticles–formation, fate, and toxicity in the environment. *Chemical Society Reviews* **2015**, *44*, 8410–8423.(3) Singh, A. K., *Engineered nanoparticles: structure, properties and mechanisms of toxicity*; Academic Press: 2015.(4) Lowry, G. V. et al. Transformations of nanomaterials in the environment, 2012.(5) Montaño, M. D. et al. Opportunities for examining the natural nanogeochemical environment using recent advances in nanoparticle analysis. *Journal of Analytical Atomic Spectrometry* **2019**, *34*, 1768–1772.(6) Hochella Jr, M. F. et al. Natural, incidental, and engineered nanomaterials and their impacts on the Earth system. *Science* **2019**, *363*, eaau8299.(7) Phalyvong, K. et al. Occurrence and origins of cerium dioxide and titanium dioxide nanoparticles in the Loire River (France) by single particle ICP-MS and FEG-SEM imaging. *Frontiers in Environmental Science* **2020**, *8*, 141.(8) Phalyvong, K. et al. Assessing CeO2 and TiO2 nanoparticle concentrations in the Seine River and its tributaries near Paris. *Frontiers in Environmental Science* **2021**, *8*, 549896.(9) Wang, J.-L. et al. Flow and fate of silver nanoparticles in small French catchments under different land-uses: the first one-year study. *Water Research* **2020**, *176*, 115722.(10) Wang, J.-L. et al. Titanium nanoparticles fate in small-sized watersheds under different land-uses. *Journal of Hazardous Materials* **2022**, *422*, 126695.(11) Hendriks, L. et al. Performance of sp-ICP-TOFMS with signal distributions fitted to a compound Poisson model. *Journal of Analytical Atomic Spectrometry* **2019**, *34*, 1900–1909.(12) Azimzada, A. et al. Quantification and characterization of Ti-, Ce-, and Ag-nanoparticles in global surface waters and precipitation. *Environmental Science & Technology* **2021**, *55*, 9836–9844.(13) Baalousha, M. et al. Elemental fingerprints in natural nanomaterials determined using SP-ICP-TOF-MS and clustering analysis. *Science of The Total Environment* **2021**, *792*, 148426.(14) Mehrabi, K. et al. Single-particle ICP-TOFMS with online microdroplet calibration for the simultaneous quantification of diverse nanoparticles in complex matrices. *Environmental Science: Nano* **2019**, *6*, 3349–3358.(15) Tharaud, M. et al. Nanoparticle identification using single particle ICP-ToF-MS acquisition coupled to cluster analysis. From engineered to natural nanoparticles. *Journal of Analytical Atomic Spectrometry* **2022**, *37*, 2042–2052.(16) Shaw, P. et al. Nanoparticle analysis using dwell times between 10 μs and 70 μs with an upper counting limit of greater than 3$\times$ 10 7 cps and a gold nanoparticle detection limit of less than 10 nm diameter. *Journal of Analytical Atomic Spectrometry* **2016**, *31*, 1234–1242.(17) Gundlach-Graham, A. et al. Mass-dependent critical value

expressions for particle finding in single-particle ICP-tofms. *Analytical chemistry* **2023**, *95*, 5618–5626.(18) Seydoux, L. et al. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature communications* **2020**, *11*, 1–12.(19) Zhong, S. et al. Machine learning: new ideas and tools in environmental science and engineering. *Environmental Science & Technology* **2021**, *55*, 12741–12754.(20) Charbuty, B. et al. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends* **2021**, *2*, 20–28.