Nanoparticle identification using single particle ICP-ToF-MS acquisition coupled to cluster analysis. From engineered to natural nanoparticles

Mickaël Tharaud, Lukas Schlatt, Phil Shaw, Marc Benedetti

Nanoparticle identification using single particle ICP-ToF-MS acquisition coupled to cluster analysis. From engineered to natural nanoparticles

Mickaël Tharaud[§], Lukas Schlatt[£], Phil Shaw[£], Marc F. Benedetti[§]

[§]Université Paris Cité, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

[£]Nu Instruments, 74 Clywedog Road South, Wrexham, UK

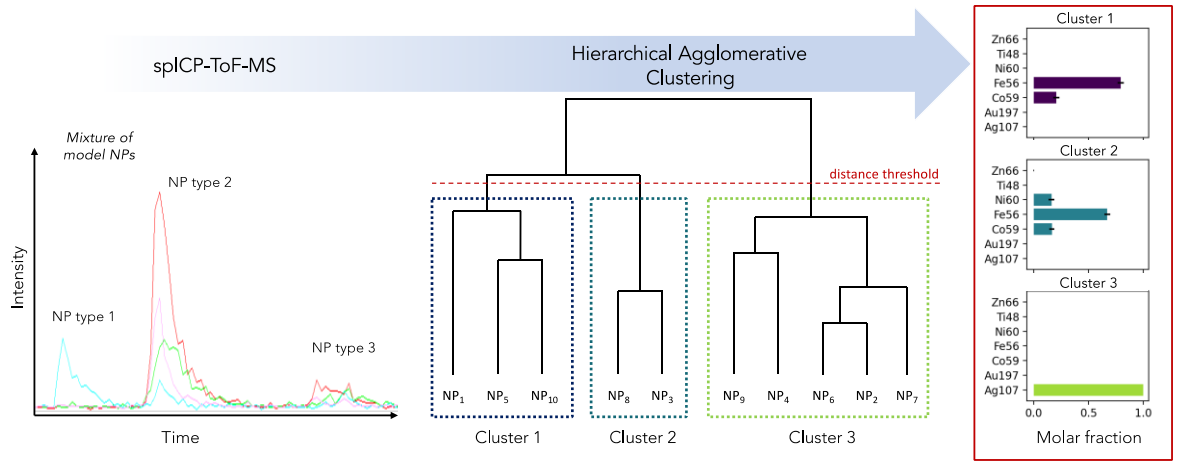
Abstract

In this study, mono-, bi-, and tri-metallic engineered nanoparticles (ENPs) are first analysed using an ICP-ToF-MS in single particle mode. The method is also used to characterize natural NPs (NNPs) (*i.e.* oxide and clay minerals). We provide here the elemental composition and mass distribution of these different NPs. For the first-time using spICP-ToF-MS, the presence of Ti- and Si-NPs as well as that of a second population with a distinct Si-Al ratio are highlighted in kaolinite. Finally, for mixtures of the different original NPs, an automated post-processing approach based on hierarchical agglomerative clustering (HAC) is applied. The elemental signature of the different clusters is in agreement with the results of previous characterizations. This illustrates the applicability of coupling spICP-ToF-MS with HAC to determine elemental fingerprints of NPs and to differentiate one from the other when mixed. The described methodology could be a starting point to simplify the analysis and make reliable interpretations of unknown NP populations and thereby more accurately understand NP populations in natural systems. However, as illustrated, thorough evaluation of the data must be performed in order to avoid misclassification of the NPs in future studies.

Keywords

engineered and natural nanoparticles, elemental signature, spICP-Time of Flight-MS, hierarchical agglomerative clustering

Graphical Abstract



Introduction

A better qualification and quantification of the anthropogenic impact coming from nanoparticles (NPs) release in the environment is linked to a better identification of the NP's origin. Several studies have shown that much of this can be achieved using the NP's elemental composition (*i.e.* fingerprint) [1-3]. However, the composition of engineered nanoparticles (ENPs) (*i.e.* purposely manufactured for specific purposes) can be quite similar to incidental (INPs) (*i.e.* by-products of human activities) and natural nanoparticles (NNPs) (*i.e.* produced by natural processes). Indeed, the production of ENPs requires elements present in soils and/or sediments from where some INPs and most NNPs come. Nevertheless, ENPs are produced for a specific purpose, for which they must have special properties, and therefore their elemental composition could differ from that of NNPs. Nevertheless, an identification of individual or group of NPs can be particularly challenging.

Single-particle inductively coupled plasma mass spectrometry (*i.e.* spICP-MS) is a technique which is able to not only investigate the particle number concentration and size (or mass), but also the elemental composition of individual nanoparticles [4-8]. Briefly, NPs introduced into the plasma are ionized and generated ions are successively extracted through the interface, separated according to their *m/z*, collected at the detector and the signal is recorded by a computer at high frequency. The resulting signal of individual NP is then converted to a mass using an external calibration and the transport efficiency [9]. Typically, the signal of a single NP event usually lasts between 200 and 1200 µs. Thus, a critical point for nanoparticle analysis/identification is the ability to monitor several elements/isotopes in an individual particle in this timeframe. Using a sector field (SF)-MS in single particle mode would allow very low detection limits (i.e. $D_{min} \leq 10$ nm for single element NP) [10]. However, the single detector with milliseconds dwell-times (*i.e.* $t_d \geq 1$ ms) combined with the relatively slow rate of change of the magnetic field (few tens of milliseconds) to switch from one mass to charge ratio (*m/z*) to another prevents the possibility of monitoring multiple elements in an individual NP. Hirata et *al.* (2020) [11], proposed spICP-MS analysis with a multi-collector (MC)-ICP-MS for Re/Os ($^{185}Re/^{188}Os$ and $^{187}Os/^{188}Os$) ratios on NPs produced through a laser-ablation system. Such a mass spectrometer is equipped with several detectors that allow the measurement of multiple

isotopes simultaneously. In the cited study, thanks to four high-time resolution ion counters, isotope signals were recorded at very high acquisition rates (*i.e.* t_d = 10 µs). However, a MC-ICP-MS is not adapted for multiple element analysis as it can only target a narrow range of m/z in one data acquisition (*i.e.* Re/Os in the cited study). Recently, although quadrupole (Q)-MS have only one detector, manufacturers increased the peak-jumping and acquisition rates (*i.e.* $t_d \leq 50$ µs) [12]. These improvements allowed better spICP-Q-MS time resolution for NP events and monitoring multiple elements in each particle became possible. Although settling-times (*i.e.* time separating the measurement of two consecutive masses) are relatively short (*i.e.* ~50 µs), the monitoring is limited to two targeted elements if each NP event is to remain well resolved [13].

Compared to the above, a time of flight (ToF)-MS does not use a magnetic, electrostatic, or RF field to disperse or filter ions for individual *m/z* ion detection but instead uses a time discrimination process to resolve ions of different mass [14]. Moreover, the short spectral-acquisition time presents one of the major advantages of the time-of-flight mass analyzer as it enables the monitoring of all *m/z* ions in the mass spectrum (quasi-)simultaneously (~2µs packets of ions sampled every 25-30 µs with 3-10 spectra summed and stored to disk continuously)[14]. Undeniably, an ICP-ToF-MS continuously collecting the multi-ion beam at this high rate with no interruption allows:

i/ deconvolution of NP signals

ii/ collection of all NP signals without loss

iii/ reduction in the incidence of multiple NP being measured in the same acquisition period

iv/ determination of the (multi)-elemental composition of individual NPs

Thus, the identification of nanoparticles (NPs) in natural complex matrices (*i.e.* environmental, geological, biological…) should be facilitated as the information on each measured NP is much more thorough.

Nevertheless, at least one issue still needs to be addressed since the large amount of data/information makes its interpretation very challenging. Several studies have used single-particle ICP-ToF-MS for the characterization of NPs in complex/environmental matrices [15-19]. Praetorius et *al.* (2017) [15] were the first to use machine learning by means of a supervised gradient decision tree boosting classification after spICP-ToF-MS. Then, Mehrabi et *al.* (2021) [20] and Baalousha et *al.* (2021) [21] combined time-of-flight measurement with another type of unsupervised machine learning algorithm: the hierarchical agglomerative clustering (HAC). Thanks to the ToF-MS, they were able to analyze metal and metalloid-containing NPs in wastewater samples and soil extractions and determine their (multi-)elemental composition. For the first time, the use of the HAC helped them to preliminarily classify the NPs and hypothesize their natural or anthropogenic origin. Although innovative and very promising, the lack of information (*i.e.* composition, transformation) about the NPs potentially present in these artificial and geological reservoirs limited the interpretation of the provided data. To go further and more robustly determine their origin, both studies mentioned the need to have more knowledge (*i.e.* data) about original NPs (*i.e.* engineered and natural).

Thus, to undertake the challenge of the identification of NP origin, this study had two objectives: i/ illustrate the applicability of coupling spICP-ToF-MS at high acquisition rate to an automated post-processing approach based on hierarchical agglomerative clustering for NP population identification

ii/ provide information for future studies (*i.e.* dealing with the question of the origin and the transformation of NPs in the environment).

These two objectives were completed by analyzing/characterizing NPs of known composition either engineered or analogues of natural NPs (*i.e.* alumina and clay minerals).

Material & Methods

Chemicals

Ultrapure water (UPW; resistivity ≈ 18.2 MΩ cm^{-1}) from a Millipore Integral 5 (Millipore, USA) was used throughout the work for dilutions. Mono-elemental standards (Au, Ag, Al, Si, Ti, Fe, Zn, Co, Mg, Ca, La, Ce, Zr) at 1000 mg L^{-1} were obtained from SCP Science (Canada). Normapure nitric acid (VWR, USA) was purified in a sub-boiler DST 1000 (Savillex, USA).

Spherical and citrate capped Au-NPs from British Biocell International (UK) with a certified nominal diameter of 40, 60 and 80 nm were used for the determination of the sample transport rate.

For the proof-of-concept experiments, different types of NPs were used:

- Mono-metallic spherical and citrate capped Au- and Ag-NPs from BBI Solutions (UK) with a certified nominal diameter of 79 ± 6 nm and 80 ± 7 nm, respectively.
- Bi-metallic spherical and citrate capped core-shell Au-Ag-NPs from Nanocomposix (USA) with a certified nominal diameter of 59 ± 6 nm and 79 ± 9 nm with a gold core diameter of 30 ± 3 nm and 51 ± 6 nm, respectively. For both NPs, based on certificates, the respective calculated thickness of the Ag shell is 15 nm and 17 nm. To facilitate the reading, they are referred to as 60 and 80 nm Ag@Au-NPs.
- Tri-metallic Nickel, Cobalt, Iron and Zinc, Cobalt, Iron oxide nanopowders, respectively named FeCoNi and FeCoZn, from US Research Nanomaterials (USA) both with a nominal diameter of 40 nm.

The theoretical molar fractions of all particles are displayed in Table 1.

As analogue of NNPs, two clay minerals, a kaolinite – *KGa-2* – and a Texas montmorillonite – *STx-1b* –, were purchased from the Clay Mineral Society (CMS). Their composition and molar fraction derived from the structure given on the CMS website (http://www.clays.org/sourceclays_data.html) are displayed in the Table 1. For instance, given the structure of STx-1b (*i.e.* $(Ca_{0.27}\ Na_{0.04}\ K_{0.01})[Al_{2.41}\ Fe(III)_{0.09}\ Mn_{tr}\ Mg_{0.71}\ Ti_{0.03}][Si_{8.00}]O_{20}(OH)_4$) and taking the major elements (*i.e.* Al, Si, Ti, Fe, Mg and Ca), the theoretical molar fraction of Al is 2.41/(2.41+8.00+0.03+0.09+0.71+0.27) = 0.21. Note that for model NNPs, the so-called theoretical

molar fraction is actually the molar fraction of the bulk material. However, the term "theoretical" is here used for consistency.

An aluminium oxide (α–Al_2O_3) was obtained from Interchim (pure 99.99%, theoretical size fraction 200 – 500 nm) and purified as described in Janot et *al.* (2012) [22].

Table 1: Theoretical molar fraction of the different NPs used in this study

	Au	***Ag***	***Al***	***Co***	***Zn***	***Ni***	***Si***	***Ti***	***Fe***	***Mg***	***Ca***
80 nm Au-NPs	1.00	/	/	/	/	/	/	/	/	/	/
80 nm Ag-NPs	/	1.00	/	/	/	/	/	/	/	/	/
60 nm Ag@Au-NPs	0.13	0.87	/	/	/	/	/	/	/	/	/
80 nm Ag@Au-NPs	0.25	0.75	/	/	/	/	/	/	/	/	/
FeCoNi	/	/	/	0.17	/	0.17	/	/	0.67	/	/
FeCoZn	/	/	/	0.17	0.17	/	/	/	0.67	/	/
Alum. ox. – Al_2O_3	/	/	1.00	/	/	/	/	/	/	/	/
Kaolinite – KGa-2	/	/	0.46	/	/	/	0.51	0.02	0.01	/	/
Montm. – STx-1b	/	/	0.21	/	/	/	0.70	0.002	0.008	0.06	0.02

Sample preparation and analysis

The clay mineral samples and the aluminium oxide were each dispersed in 50 mL of ultrapure water at 100 mg L^{-1} and vigorously shaken for 15 minutes. In order to remove bigger particles (*i.e.* > 1 µm), suspensions were left to settle overnight, *ca.* 15 hours, and then the top 10 mL from each supernatant were sampled.

All suspensions (*i.e.* model ENPs and NNPs) were first individually analysed and then, two mixtures were prepared and analysed:

i/ a first mixture with the model ENPs (Ag-, Au-, Ag@Au-NPs and FeCoNi/Zn oxides)

ii/ a second mixture with the model NNPs (Al_2O_3, KGa-2 and STx-1b).

The interaction between ENPs and NNPs being beyond the scope of this article, no ENPs-NNPs mixture was prepared.

Prior to spICP-ToF-MS analysis, suspensions were sonicated for 60 seconds and freshly diluted using ultrapure water to avoid NP coincidences as necessary after reviewing initial measurements.

Single particle ICP-ToF-MS

The simultaneous multi-elemental spICP-MS analysis was performed on an ICP-ToF-MS (Vitesse, Nu Instruments). This instrument is able to record mass spectra from ^{23}Na to ^{238}U without

interruption at 27.5µs per spectrum. By reducing the data using a high-power graphics processing unit during collection, uninterrupted acquisitions are possible for multiple minutes at these fast dwell-times. Therefore, 3 raw spectra were accumulated for each fully processed spectrum stored to the solid-state hard drive resulting in a total dwell-time of 83 µs, still allowing the deconvolution of all NP peaks. Spectral processing included the subtraction of electronic noise and spectral baseline followed by an integration of specific mass regions for each isotope within the mass range measured. Table 2 gives representative spICP-ToF-MS operating conditions and parameters. spICP-ToF-MS raw data were processed using the integrated software – *NuQuant* using the methodology described by Shaw & Donard (2016) [23].

Table 2: Instrument operating conditions and spICP-ToF-MS parameters

	spICP-ToF-MS parameters
Cones	standard Ni cones
Uptake rate [$]	0.4 L min^{-1}
RF power [$]	1100 W
Ar Cool gas [$]	13.0 L min^{-1}
Ar auxiliary gas	2.0 L min^{-1}
Ar nebulization gas [$]	1.0 L min^{-1}
Collision cell gas [$]	He at 0.016 L min^{-1} and H_2 at 0.002 L min^{-1}
Acquired mass range	23 amu – 245 amu
Examined isotopes [#]	^{24}Mg, ^{27}Al, ^{28}Si, ^{44}Ca, ^{48}Ti, ^{56}Fe, ^{59}Co, ^{60}Ni, ^{66}Zn, ^{90}Zr, ^{107}Ag ^{139}La, ^{140}Ce and ^{197}Au
Dwell-time*	83 µs *
Oxides [$]	0.2% for BaO/Ba

[$] The instrument was tuned daily to ensure maximum sensitivity and stability, as well as low barium oxide formation. [#] Full spectra from ^{23}Na to ^{238}U were acquired by the ICP-ToF-MS. However, to reduce the data processing time post-acquisition, the isotope set was reduced to key elements. This made processing possible in seconds rather than minutes when calculating the permutations of element coincidence in each particle. * In this case, the dwell-time represents the spectral-acquisition time.

Hierarchical agglomerative clustering

The post-processing strategy for graphical representation, advanced calculations and cluster analysis was programmed in Python 3.8 using calculated moles and masses exported from *NuQuant*. The machine learning library – *Scikit-learn* (*https://scikit-learn.org/stable/*) – and specifically its object – *AgglomerativeClustering* – was used to perform the hierarchical agglomerative clustering (HAC).

The hierarchical clustering consists of grouping clusters recursively either in an agglomerative or divisive way. The hierarchical agglomerative clustering (HAC) is here preferred. Each datapoint is

considered as an individual cluster and is then successively merged in clusters based on similarities [24]. The usual output of the HAC is a 2-D tree-like structure. The so-called dendrogram illustrates the arrangement of the grouped clusters [20, 24]. The number of clusters is determined by cutting this dendrogram at a desired level (*i.e.* distance threshold discussed below). To classify the NPs within each sample, single stage clustering using the Ward minimum variance method calculated with the Euclidean distance is performed here as the recursive merging strategy. It is defined as the smallest increase in the within-cluster sum of squares due to the merging of two clusters. The Euclidean distance used here is the molar fraction normalized as a distance metric.

Results & Discussion

Characterization of individual suspensions of NPs

Gold and/or silver NPs. In this section, 80 nm pure Au- and Ag-NPs as well as 60 nm and 80 nm bi-metallic Ag shelled Au-NPs were individually analyzed using the ICP-ToF-MS in single particle mode.

Figure 1 displays the number of moles of Au *vs* the number of moles of Ag (all in femtomoles and calculated using the sum of moles of the detected elements) for individual solutions.

Pure Au- and Ag-NPs (Figure 1 a and b) respectively give an average size of 76.6 ± 6.5 nm and 76.8 ± 13.0 nm (Figure 1 c and d) with a pure metallic fingerprint (molar ratio = 1). For 60 and 80 nm bi-metallic Ag shelled Au-NPs, the average size per particle is respectively 63.2 ± 10.5 nm with a molar fraction of Ag = 0.90 ± 0.06 and Au = 0.10 ± 0.06 and; 93.7 ± 15.0 nm with a molar fraction of Ag = 0.83 ± 0.08 and Au = 0.17 ± 0.08. Additionally, the calculated size of the Au-core is 29 ± 7 nm and 54 ± 10 nm for 60 and 80 nm for Ag@Au-NPs, respectively. Based on the average diameter of 60 and 80 nm Ag@Au-NPs, the recalculated thickness of the Ag shell is respectively *ca.* 17 nm and *ca.* 19 nm. These results show that calculated sizes as well as mole fraction of pure and bi-metallic NPs are in good agreement with the certified ones.

Table 3: Experimental diameter and molar fractions of the different NPs for pure Au-, pure Ag- and bi-metallic Ag shelled Au-NPs. The theoretical molar fractions of 60 and 80 nm Ag@Au-NPs are 0.13/0.87 and 0.25/0.75, respectively

	Molar fraction		***Nanoparticle***
	Au	***Ag***	***diameter (nm)***
80 nm Au-NPs	1.00	/	76.6 ± 6.5
80 nm Ag-NPs	/	1.00	76.8 ± 13.0
60 Ag@Au-NPs	0.10 ± 0.06	0.90 ± 0.06	63.2 ± 10.5
80 Ag@Au-NPs	0.17 ± 0.08	0.83 ± 0.08	93.7 ± 15.0

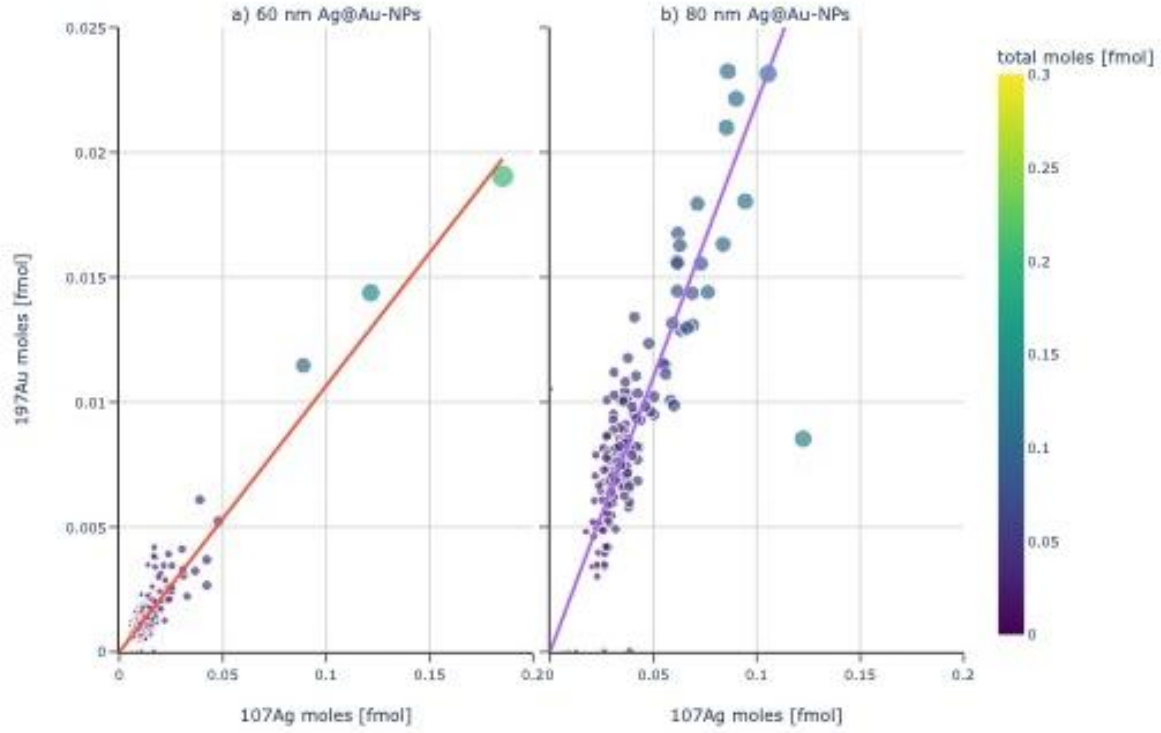


Figure 1: number of moles of Au vs the number of moles of Ag (all in femtomoles). Red and purple lines represent the best fit lines, for a) 60 nm and b) 80 nm Ag@Au-NPs respectively. The theoretical molar fractions of 60 and 80 nm Ag@Au-NPs are 0.13/0.87 and 0.25/0.75, respectively. The colorbar gives the total amount of moles (in femtomoles) in individual datapoints while the size of each point is related to the calculated mass of each NP. Note that the molar fraction, the total amount of moles and the mass are calculated using the sum (mole or mass) of the elements detected within the NP

Nickel/Zinc Cobalt Iron oxide NPs. Individual suspensions of tri-metallic Nickel, Cobalt, Iron and Zinc, Cobalt, Iron oxide NPs were analyzed by spICP-ToF-MS. Figure 2 displays the calculated molar fraction of the three metals in each suspension. In both tri-metallic NP solutions, 3 types of NPs are found with the following composition:

1/ pure Fe NPs,

2/ bi-metallic Fe-Co NPs

3/ tri-metallic Fe-Co-Ni/Zn NPs.

The average particle mass (sum of the mass of the detected elements) displayed in Table 4 shows that bi-metallic NPs have mass lower than tri-metallic NPs.

Considering first only tri-metallic NPs, calculated molar fractions (Table 4) are in good agreement with theoretical ones displayed in Table 1. Calculated sizes (assuming a spherical shape) are much larger compare to the theoretical values (*ca.* 98 ± 25 and 124 ± 30 nm, for Fe-Co-Ni and Fe-Co-Zn respectively, instead of 40 nm). These results suggest that only the larger fraction of the size distribution of the material or aggregates could be detected instead of the average size of individual tri-metallic NPs. It is confirmed in Figure SI-5 where the minimum size detectable of tri-metallic NPs has been computed assuming a spherical shape and a homogeneous theoretical composition such as $Ni_{0.5}Co_{0.5}Fe_2O_4$ and $Zn_{0.5}Co_{0.5}Fe_2O_4$. This leads to respective D_{min} of *ca.* 80 and 87 nm which are

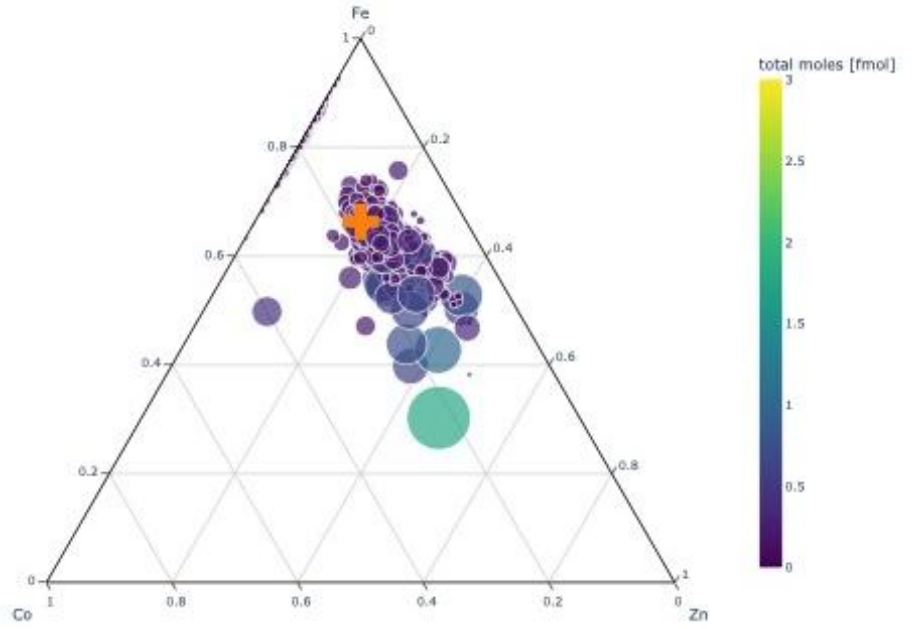
significantly above the theoretical size of 40 nm. Taking into account pure and bi-metallic NPs, two hypotheses can be made. Assuming a homogeneous composition of all NPs, higher M_{min} for Ni and Zn compare to Co displayed in Table SI-2 imply that if the NP mass decreases, the ability of the ICP-ToF-MS to detect Ni and Zn is first affected. This then exhibits NPs with a bi-metallic (Fe-Co) signature. For even smaller NPs, although Table SI-2 shows that $M_{min\text{-}Fe} > M_{min\text{-}Co}$, as Fe is the major constituent of the NP (omitting oxygen) this favors its detection which leads to the measurement of NPs with an apparently pure Fe composition.

A second assumption is that the elemental composition of these NPs is heterogeneous and that pure and bi-metallic NPs are indeed present within the suspension. The heterogeneity observed in Figure SI-1b first tends to confirm the dispersity of the tri-metallic NP composition as well as the presence of mono and bi-metallic NPs (Figure 2). In addition, as the size of the datapoint is related to the mass of the NPs in Figure 2, it clearly appears that there is no direct link between these masses and the presence of one or more elements. Therefore, this confirms that that mono-, bi- and tri-metallic NPs coexist in the solution.

Table 4: Experimental mass and molar fractions of the different types of NPs coming from Fe-Co-Ni/Zn oxides. The theoretical molar fractions of Fe-Co-Ni/Zn oxides are 0.67/0.17/0.17

		Molar fraction			***Average particle***
		Fe	***Co***	***Ni/Zn***	***mass(fg)***
Fe-Co-Ni	***Fe***	1.00	/	/	0.5 ± 0.2
	Fe-Co	0.80 ± 0.03	0.20 ± 0.03	/	1.1 ± 0.6
	Fe-Co-Ni	0.66 ± 0.05	0.16 ± 0.03	0.18 ± 0.05	5.0 ± 5.2
Fe-Co-Zn	***Fe***	1.00	/	/	0.4 ± 0.2
	Fe-Co	0.82 ± 0.05	0.18 ± 0.05	/	1.4 ± 1.3
	Fe-Co-Zn	0.63 ± 0.04	0.14 ± 0.03	0.23 ± 0.05	9.5 ± 8.3

a)



b)

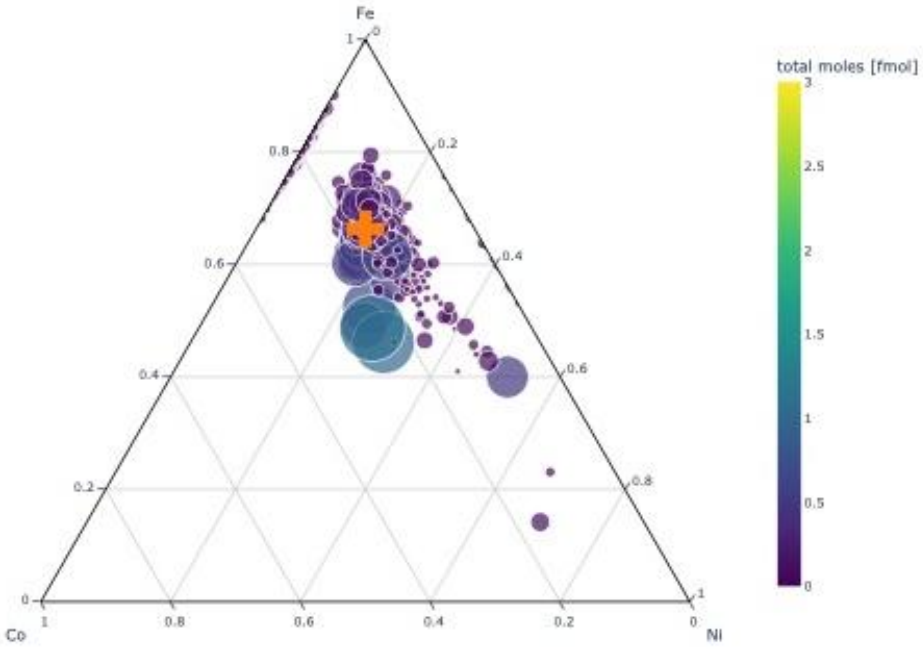


Figure 2: Ternary diagrams of molar fractions of a) Fe-Co-Zn and b) Fe-Co-Ni. Orange cross represents the theoretical molar fraction (i.e. 0.67/0.17/0.17). The color bar gives the total amount of moles (in femtomoles) in individual datapoints while the size of each point is related to the calculated total mass of each NP. Note that the molar fraction, the total amount of moles and the mass are calculated using the sum (mole or mass) of the elements detected within the NP

Analogue of natural NPs. In this section, the three model NNPs (*i.e.* a kaolinite, a montmorillonite as well as an aluminium oxide) are individually characterized using single-particle ICP-ToF-MS. Figure 3 displays the number of moles of Si *vs* the number of moles of Al (*i.e.* the two major elements) for each sample. Fitted lines clearly appear to be different for all samples. For KGa-2 and STx1-b, lines of least squares fit are 1.06 and 2.73, respectively. These fits represent the experimental Si-Al molar ratio. Theoretical ones, calculated from the data in Table 1, are 1.11 and 3.33 for KGa-2 and STx1-b,

respectively. For the kaolinite, the experimental and theoretical molar ratio are in good agreement. For the montmorillonite the experimental ratio slightly diverges from the theoretical one. The presence of pure Al-NPs (datapoints displayed on the x-axis in Figure 3 c) decreases the correlation coefficient and skews the line of fit but this latter is still very distinct from the one of kaolinite.

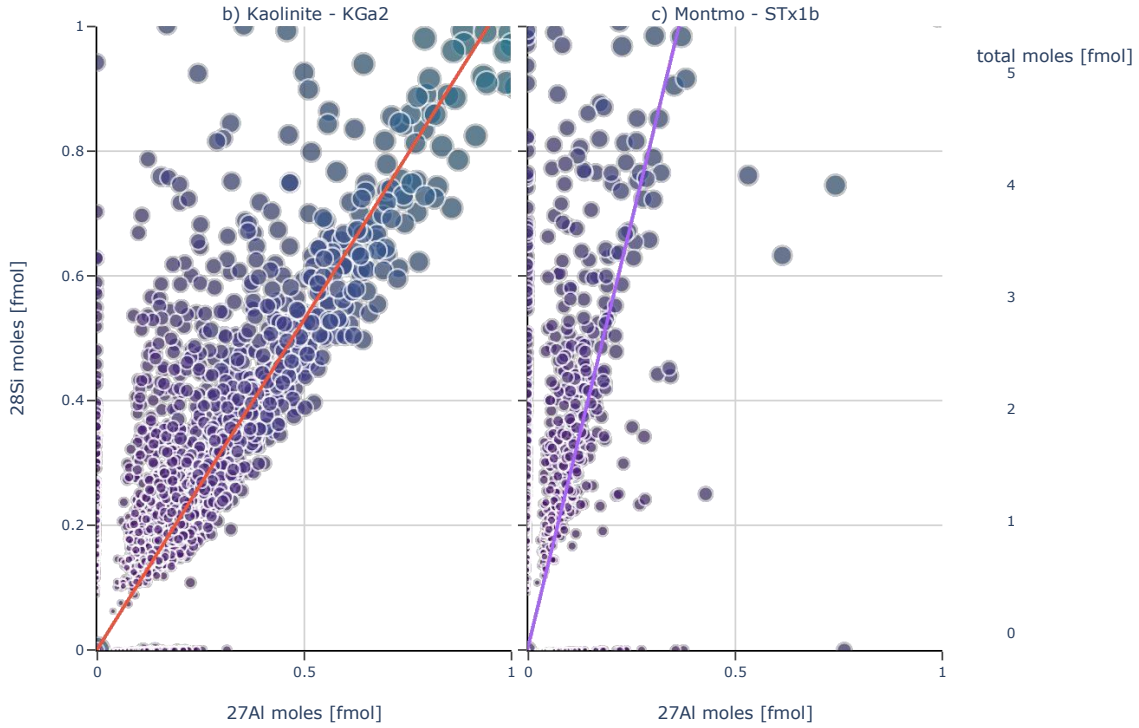


Figure 3: (Zoom of Figure SI-3a for KGa-2 and STx-1b only). Number of moles of Si vs the number of moles of Al (all in femtomoles). Red and purple lines represent the best fit lines, for KGa-2 and STx-1b, respectively. The theoretical Al-Si molar fraction is 1.11 and 3.33 for KGa-2 and STx-1b, respectively. The color bar gives the total amount of moles (in femtomoles) in individual datapoints while the size of each point is related to the calculated total mass of each NP. Note that the molar fraction, the total amount of moles and the mass are calculated using the sum (mole or mass) of the elements detected within the NP

Identification of model ENPs in mixtures

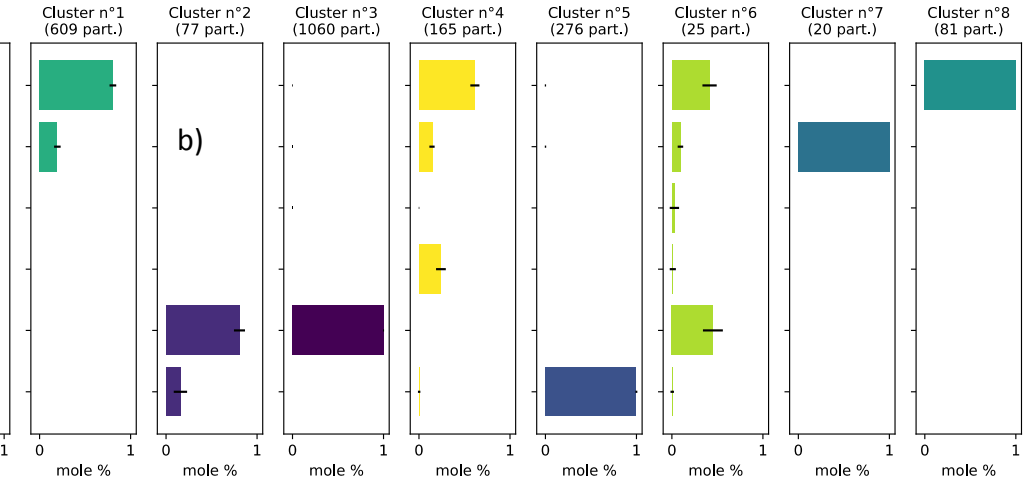
Mixture of model ENPs. The results displayed in the previous sections show that the characterization of multi-metallic NPs is possible using an ICP-ToF-MS in single particle mode. Although sizes provided for the tri-metallic NPs are much larger than expected, sizes of other NPs as well as all molar fractions are individually, precisely determined. To go further, it is important to demonstrate that this kind of analysis is precise and reliable for the identification of NPs in mixtures with a number of monitored elements > 3 as in environmental samples. In order to mimic these kinds of samples, a mixture of the model ENPs characterized above is analyzed by spICP-ToF-MS. Then, the hierarchical agglomerative clustering (HAC) approach described above is used to recursively group similar events (*i.e.* NP) with similar fingerprints (*i.e.* elemental molar fractions). The HAC, being an unsupervised

methodology, performs the clustering "blind" (*i.e.* without influence of a human being). However, as mentioned above, the clustering must be stopped at some level by applying a threshold. Here, this latter was visually determined after careful examination of the dendrogram and in order to obtain minimum dispersion within each cluster. The optimized threshold distance obtain was $d_{threshold} = 3$. Based on the summary displayed in Figure 4, the methodology looks precise and accurate. Indeed, cluster characteristics are in agreement with the model ENPs described above. For instance, considering silver and gold, the pure (cluster 3 and 5) and bi-metallic (cluster 2) populations are displayed with the correct molar fractions. Equivalent spherical diameters also correspond to the ones calculated when analyzed separately (Figure 4 c). For the tri-metallic NPs, pure Fe (cluster 8), bi-metallic Fe-Co (cluster 1) as well as tri-metallic Fe-Co-Ni/Zn (cluster 0 and 4) NPs are still displayed with molar fractions (Figure 4 a) and equivalent spherical diameters (Figure 4 c) in agreement with information provided in the previous section. Two other clusters (cluster 6 and 7) are also displayed in Figure 4 a. The mono-metallic composition (*i.e.* only Co is detected) as well as the very low mass (Figure 4 b) of cluster 7 indicate that these NPs are probably pure Co-NPs dispersed during the sonication and initially categorized as hot-spots in STEM-XEDS mapping (Figure SI-1b). Cluster 6 shows an unexpected multi-metallic composition with molar fraction of *ca.* Ag = 0.50, Fe = 0.45, Co = 0.04 and traces of Ni and Zn. The sample was initially diluted to reduce the probability of coincidences. However, such NP coincidences as well as (hetero-)aggregation within the sample are probable. Time-based graphic records displayed in the supporting information demonstrate that these are easily identified through the continuous acquisition at tens of micro-second dwell-times. In the first scenario (Figure SI-4a), the starting time of the peaks of Fe, Co, Zn and Ag is exactly the same. Based on the NPs initially dispersed in the sample, this indicates a hetero-aggregation between a tri-metallic NP and a pure Ag-NP. The Figure SI-4b shows a different scenario. Indeed, the starting point of the peak of Ag is different than the one of Fe, Co and Ni indicating that an overlap is observed. Thus, the way data are processed in *NuQuant* leads to an estimate that these pure Ag and tri-metallic peaks are actually a single quadri-metallic peak. Although important to note that for future studies and subsequent interpretations, a careful consideration of the raw data should be made to avoid misinterpretation, the proportion of cluster 6 as well as cluster 7 is here negligeable compared to the others (*i.e.* 0.9% and 0.7%) (Figure 4 d).

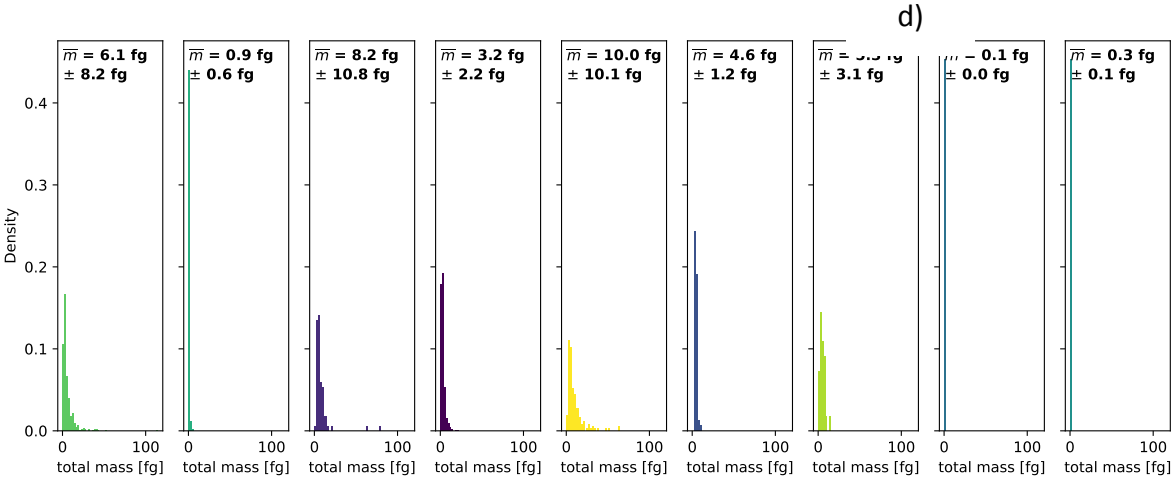
This demonstrates the applicability of spICP-ToF-MS analysis coupled to HAC for model ENPs identification. It especially shows that it is very efficient in differentiating populations even in low proportions and also in highlighting (hetero-)aggregation. As described above, after examining the dendogram, an optimized distance threshold (*i.e.* $d_{threshold} = 3$) is used to end the clustering. The more complex the dataset is, the more complicated the definition of the threshold becomes. For instance, by increasing the distance threshold for this dataset, the total number of clusters decreases as some clusters are merged together. The first populations to be merged are tri-metallic NPs both with Ni and Zn giving one cluster with a quadri-metallic signature. Then by increasing the $d_{threshold}$ further, all

clusters containing either Fe, Co, Ni or Zn are merged together while pure Ag and Ag@Au-NPs are also combined, leaving a third cluster with only Au-NPs. To find the appropriate/optimal distance threshold in HAC (as for environmental samples), Krishnamoorthy et *al.* (2012) [25] propose an Optimized Agglomerative Clustering (OAC) that reduces computational complexity and misclassification errors with good accuracy using multilevel thresholds for finding the optimum number of clusters. However, this optimized methodology is not yet implemented in Python but is a promising line of development for future studies. In the next section, about the clustering of model NNPs, in order to more objectively group the data, the distance threshold was fixed at the previously optimized value found for the clustering of the model ENPs mixture *i.e.* $d_{threshold} = 3$.

a)


Cluster n°1
(609 part.)
Cluster n°2
(77 part.)
Cluster n°3
(1060 part.)
Cluster n°4
(165 part.)
Cluster n°5
(276 part.)
Cluster n°6
(25 part.)
Cluster n°7
(20 part.)
Cluster n°8
(81 part.)
b)
mole %

c)


d)
$\overline{m}$ = 6.1 fg
± 8.2 fg
$\overline{m}$ = 0.9 fg
± 0.6 fg
$\overline{m}$ = 8.2 fg
± 10.8 fg
$\overline{m}$ = 3.2 fg
± 2.2 fg
$\overline{m}$ = 10.0 fg
± 10.1 fg
$\overline{m}$ = 4.6 fg
± 1.2 fg
± 3.1 fg
± 0.0 fg
$\overline{m}$ = 0.3 fg
± 0.1 fg
Density
total mass [fg]

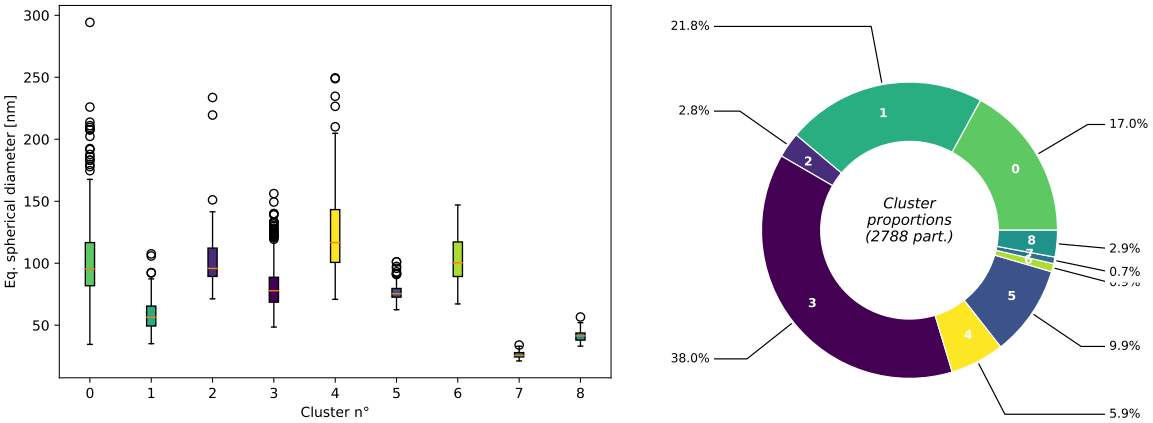


Figure 4: Summary of the hierarchical agglomerative clustering on spICP-ToF-MS data of the model ENP mixture analysis. a) average molar fractions in color and standard deviation in black, b) particle mass distributions calculated with the mass of all detected elements in the cluster, c) equivalent spherical diameters and d) proportions of individual clusters are displayed

Mixture of model NNPs. Previous results for model ENPs have shown that the here described methodology of coupling spICP-ToF-MS analysis with hierarchical agglomerative clustering is reliable for distinguishing them when mixed together as well as determining their individual fingerprints. Furthermore, spICP-ToF-MS has also been proved to be effective for the characterization of individual model NNPs. Thus, in the following section, model NNPs mixed together are analyzed using spICP-ToF-MS. Results are then processed using the HAC approach with the d_t previously optimized with the model ENPs. As for the mixture of model ENPs, our methodology preserves the molar fractions of each individual cluster/NP population (Figure 5). As expected, cluster 0 presents a (nearly-)pure-Al composition (*i.e.* Al molar fraction = 1) while clusters 1 and 4 are respectively composed of Si – Al – Fe – Ti and Si – Al – Mg – Fe – Ca – Ti. Looking at their molar fractions, these compositions are in agreement with the molar fractions of the kaolinite and the montmorillonite (Table 1). Indeed, it appears that in cluster 1, silicon and aluminium respectively represent 0.50 ± 0.04 and 0.49 ± 0.03, followed by Ti and Fe with 0.007 ± 0.025 and 0.003 ± 0.005 clearly showing a KGa-2 signature. Cluster 4 has silicon, aluminium and magnesium with respective molar fraction of 0.74 ± 0.07, 0.23 ± 0.07 and 0.02 ± 0.03 in agreement with the signature of STx-1b. Fe, Ca and Ti are less present in this cluster with molar fractions of 0.003 ± 0.004, 0.002 ± 0.006 and 0.001 ± 0.003, respectively. Two additional (nearly-)pure NP clusters (*i.e.* Si (cluster 3) and Ti (cluster 5) with molar fraction *ca.* 0.99 ± 0.02) are also identified. As mentioned on the CMS website, quartz (SiO_2) was detected in STx-1b using infrared spectroscopy at λ = 697 cm^{-1}. Titanium containing clusters are likely to be TiO_2 NPs. These TiO_2 NPs were previously observed by micro-Raman spectroscopy within clay mineral samples and particularly in kaolinite where Ti is as free-oxide at *ca.* 86% [26]. Surprisingly, a last cluster (*i.e.* cluster 2) is displayed, still with Si – Al – Ti – Fe signature as for KGa-2 but with unexpected molar fractions (0.66 ± 0.06 – 0.33 ± 0.07 – 0.009 ± 0.021 – 0.004 ± 0.005, respectively). Note that this cluster is also present when analyzing a suspension of kaolinite only (Figure SI-3b). We therefore assume that this population is originally in the kaolinite and that this study is the first to

highlight its presence. Going a step further and combining this population with the one displayed in cluster 1, the Si – Al – Fe – Ti molar fraction of the kaolinite appears to be now 0.55 ± 0.08 – 0.44 ± 0.06 – 0.003 ± 0.004 – 0.007 ± 0.016, respectively.

However, this shows that coupling spICP-ToF-MS with HAC can distinguish model NNP populations with close signatures and accurately establishes their elemental fingerprint as well as correctly determines their mass distribution in a mixture (Figure 5).

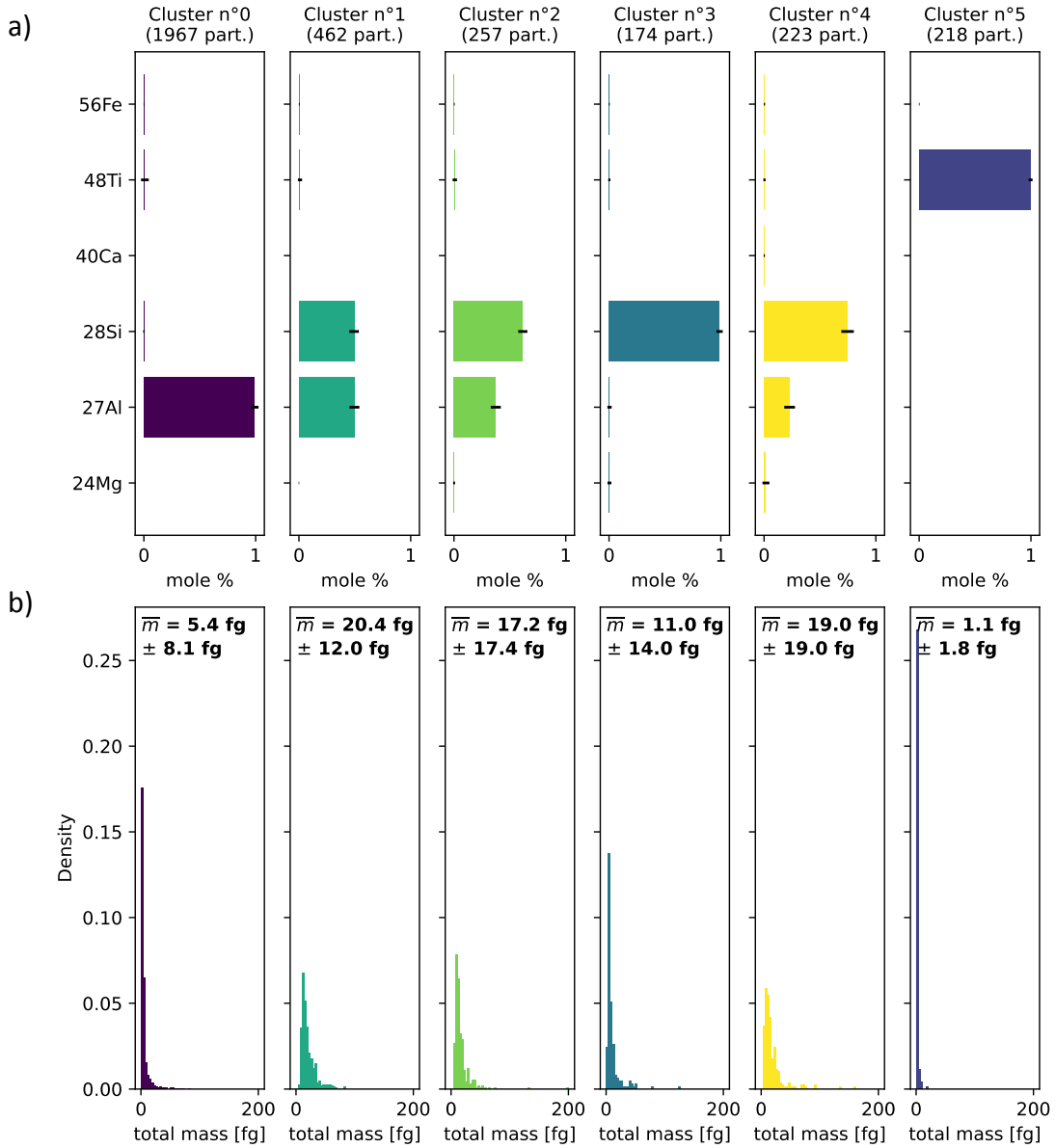


Figure 5: Summary of the hierarchical agglomerative clustering on spICP-ToF-MS data of the model NNP mixture analysis. a) Molar fractions and b) particle mass distributions of individual clusters are displayed

Conclusion

We have successfully analyzed model engineered and natural NPs using an ICP-ToF-MS in single particle mode as well as processed data of mixtures using an HAC approach. Using this

methodology, we first fully determined the elemental fingerprint of model ENPs alone and then mixed together with good precision and accuracy. STEM analysis has confirmed the results qualitatively (*i.e.* bi- or tri-metallic composition of the model NPs). We have subsequently provided elemental composition and mass distribution information about a mixture of three types of model NNPs. The presence of Ti- and Si-NPs also highlighted here is in agreement with infrared spectroscopy data (CMS website) as well as µ-Raman spectroscopy data [26]. For the first time in kaolinite, the presence of a second population with a distinct Si-Al ratio was also demonstrated.

This work has been a starting point to better identify/classify NP populations (*i.e.* anthropogenic *vs* natural NPs). However, in future studies, thorough evaluation of the data must be performed in order to avoid misclassification of the NPs. Other factors such as low particle numbers as well as low masses of NPs (such as cluster 6 and 7 in the mixture of model ENPs) must be meticulously justified. Moreover, we have highlighted that the determined elemental composition is affected by the sensitivity of the mass spectrometer. Indeed, as illustrated with the results of the tri-metallic NPs, M_{min} of each element must be taken into account before assuming the origin of NPs. Experiments on a variety of NP populations, must also be performed. Therefore, along with necessary studies of natural environments, future work must also focus on supplying elemental signatures of model engineered or natural NPs either alone or mixed together. By then comparing such studies with unknown samples, the identification, origin and potential transformations of NPs should be possible.

From an analytical point of view, in order to increase reliability of the hierarchical clustering, the automatic determination of the distance threshold should be improved as this is the starting point of the grouping of NPs based on their similarities. Indeed, HAC miscomputation can obviously lead to misinterpretation of spICP-ToF-MS data. Additionally, as demonstrated here, traces can be more difficult to detect, limiting the understanding of the origin and the transformation of the NPs in the media studied. Indeed, differentiating natural from engineered NPs is made through the detection of elemental impurities within the NPs. Hence, manufacturers should continue to work on the development of more efficient ion extraction and/or detection systems.

References

1. A. Gondikas, F. von der Kammer, R. Kaegi, O. Borovinskaya, E. Neubauer, J. Navratilova, A. Praetorius, G. Cornelis and T. Hofmann, *Environmental Science: Nano*, 2018, **5**, 313-326.
2. K. Phalyvong, Y. Sivry, H. Pauwels, A. Gélabert, M. Tharaud, G. Wille, X. Bourrat and M. F. Benedetti, *Frontiers in Environmental Science*, 2020, **8**.
3. D. L. Slomberg, M. Auffan, N. Guéniche, B. Angeletti, A. Campos, D. Borschneck, O. Aguerre-Chariol and J. Rose, *Frontiers in Environmental Science*, 2020, **8**, 76.
4. C. Degueldre and P. Y. Favarger, *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 2003, **217**, 137-142.
5. C. Degueldre and P. Y. Favarger, *Talanta*, 2004, **62**, 1051-1054.
6. C. Degueldre, P. Y. Favarger and C. Bitea, *Analytica Chimica Acta*, 2004, **518**, 137-142.
7. C. Degueldre, P. Y. Favarger and S. Wold, *Analytica Chimica Acta*, 2006, **555**, 263-268.
8. C. Degueldre, P. Y. Favarger, R. Rossé and S. Wold, *Talanta*, 2006, **68**, 623-628.
9. H. E. Pace, N. J. Rogers, C. Jarolimek, V. A. Coleman, C. P. Higgins and J. F. Ranville, *Analytical chemistry*, 2011, **83**, 9361-9369.
10. M. Tharaud, A. P. Gondikas, M. F. Benedetti, F. von der Kammer, T. Hofmann and G. Cornelis, *Journal of Analytical Atomic Spectrometry*, 2017, **32**, 1400-1411.
11. T. Hirata, S. Yamashita, M. Ishida and T. Suzuki, *Mass Spectrometry*, 2020, **9**, A0085-A0085.
12. M. Montano, H. Badiei, S. Bazargan and J. Ranville, *Environmental Science: Nano*, 2014, **1**, 338-346.
13. M. D. Montaño, J. W. Olesik, A. G. Barber, K. Challis and J. F. Ranville, *Analytical and bioanalytical chemistry*, 2016, **408**, 5053-5074.
14. M. Guilhaus, *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2000, **55**, 1511-1525.
15. A. Praetorius, A. Gundlach-Graham, E. Goldberg, W. Fabienke, J. Navratilova, A. Gondikas, R. Kaegi, D. Günther, T. Hofmann and F. von der Kammer, *Environmental Science: Nano*, 2017, **4**, 307-314.
16. K. Mehrabi, D. Günther and A. Gundlach-Graham, *Environmental Science: Nano*, 2019, **6**, 3349-3358.
17. A. Azimzada, J. M. Farner, I. Jreije, M. Hadioui, C. Liu-Kang, N. Tufenkji, P. Shaw and K. J. Wilkinson, *Frontiers in Environmental Science*, 2020, **8**.
18. F. Loosli, J. Wang, S. Rothenberg, M. Bizimis, C. Winkler, O. Borovinskaya, L. Flamigni and M. Baalousha, *Environmental Science: Nano*, 2019, **6**, 763-777.
19. A. Azimzada, I. Jreije, M. Hadioui, P. Shaw, J. M. Farner and K. J. Wilkinson, *Environmental Science & Technology*, 2021.
20. K. Mehrabi, R. Kaegi, D. Günther and A. Gundlach-Graham, *Environmental Science: Nano*, 2021, **8**, 1211-1225.

21. M. Baalousha, J. Wang, M. Erfani and E. Goharian, *Science of The Total Environment*, 2021, **792**, 148426.
22. N. Janot, P. E. Reiller, X. Zheng, J.-P. Croué and M. F. Benedetti, *Water Research*, 2012, **46**, 731-740.
23. P. Shaw and A. Donard, *Journal of Analytical Atomic Spectrometry*, 2016, **31**, 1234-1242.
24. P. Govender and V. Sivakumar, *Atmospheric Pollution Research*, 2020, **11**, 40-56.
25. R. Krishnamoorthy and S. SreedharKumar, 2012.
26. S. Shoval, G. Panczer and M. Boudeulle, *Optical Materials*, 2008, **30**, 1699-1705.

Acknowlegment

Parts of this work were supported by IPGP multidisciplinary program PARI, by Region île-de-France SESAME Grants no. 12015908 and EX047016 and the IdEx Université de Paris grant, ANR-18-IDEX-0001. Authors also thank Dr. Nicolas Menguy from the Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, UMR 7590 CNRS - Sorbonne Université for performing and interpreting the TEM measurements.

Figure captions

- Figure 1: number of moles of Au vs the number of moles of Ag (all in femtomoles). Red and purple lines represent the best fit lines, for a) 60 nm and b) 80 nm Ag@Au-NPs respectively. The theoretical molar fractions of 60 and 80 nm Ag@Au-NPs are 0.13/0.87 and 0.25/0.75, respectively. The colorbar gives the total amount of moles (in femtomoles) in individual datapoints while the size of each point is related to the calculated mass of each NP. Note that the molar fraction, the total amount of moles and the mass are calculated using the sum (mole or mass) of the elements detected within the NP
- Figure 2: Ternary diagrams of molar fractions of a) Fe-Co-Zn and b) Fe-Co-Ni. Orange cross represents the theoretical molar fraction (i.e. 0.67/0.17/0.17). The color bar gives the total amount of moles (in femtomoles) in individual datapoints while the size of each point is related to the calculated total mass of each NP. Note that the molar fraction, the total amount of moles and the mass are calculated using the sum (mole or mass) of the elements detected within the NP
- Figure 3: (Zoom of Figure SI-3a for KGa-2 and STx-1b only). Number of moles of Si vs the number of moles of Al (all in femtomoles). Red and purple lines represent the best fit lines, for KGa-2 and STx-1b, respectively. The theoretical Al-Si molar fraction is 1.11 and 3.33 for KGa-2 and STx-1b, respectively. The color bar gives the total amount of moles (in femtomoles) in individual datapoints while the size of each point is related to the calculated total mass of each NP. Note that the molar fraction, the total amount of moles and the mass are calculated using the sum (mole or mass) of the elements detected within the NP
- Figure 4: Summary of the hierarchical agglomerative clustering on spICP-ToF-MS data of the model ENP mixture analysis. a) average molar fractions in color and standard deviation in black, b) particle mass distributions calculated with the mass of all detected elements in the cluster, c) equivalent spherical diameters and d) proportions of individual clusters are displayed
- Figure 5: Summary of the hierarchical agglomerative clustering on spICP-ToF-MS data of the model NNP mixture analysis. a) Molar fractions and b) particle mass distributions of individual clusters are displayed

Table captions

- Table 1: Theoretical molar fraction of the different NPs used in this study
- Table 2: Instrument operating conditions and spICP-ToF-MS parameters

- Table 3: Experimental diameter and molar fractions of the different NPs for pure Au-, pure Ag- and bi-metallic Ag shelled Au-NPs. The theoretical molar fractions of 60 and 80 nm Ag@Au-NPs are 0.13/0.87 and 0.25/0.75, respectively
- Table 4: Experimental mass and molar fractions of the different types of NPs coming from Fe-Co-Ni/Zn oxides. The theoretical molar fractions of Fe-Co-Ni/Zn oxides are 0.67/0.17/0.17