A SEM-based method to determine the mineralogical composition and the particle size distribution of suspended sediment

Sylvain Pinet, Bruno Lartiges, Jean-Michel Martinez, Sylvain Ouillon

A SEM-based method to determine the mineralogical composition and the particle size distribution of suspended sediment

Sylvain Pinet[a,*], Bruno Lartiges[a], Jean-Michel Martinez[a,b], Sylvain Ouillon[c]

[a] University of Toulouse (Paul Sabatier), Geosciences Environment Toulouse, CNRS (Centre National de la Recherche Scientifique), IRD (Institut de Recherche pour le Développement), 14 av. E. Belin, 31400 Toulouse, France.

[b] Instituto de Geosciências, Universidade de Brasilia, Campus Universitario Darcy Ribeiro, ICC Centro, 70910-900 Brasilia, Brazil.

[c] University of Toulouse, LEGOS (Laboratoire d'Etudes en Géophysique et Océanographie Spatiales), 14 av. E. Belin, 31400 Toulouse, France.

* Corresponding author. *E-mail address:* sylvain.pinet@yahoo.fr (S. Pinet)

A SEM-based method to determine the mineralogical composition and the particle size distribution of suspended sediment

Keywords: Mineralogy; Scanning Electron Microscopy; Suspended sediment; Machine learning; Particle size distribution

Abstract

A robust method for characterizing the mineralogy of suspended sediment in continental rivers is introduced. It encompasses 3 steps: the filtration of a few milliliters of water, measurements of X-ray energy dispersive spectra using Scanning Electron Microscopy (SEM), and robust machine learning tools of classification. The method is applied to suspended particles collected from various Amazonian rivers. A total of more than 204,000 particles were analyzed by SEM-EDXS (Energy Dispersive X-ray Spectroscopy), i.e. about 15,700 particles per sampling station, which lead to the identification of 15 distinct groups of mineralogical phases. The size distribution of particles collected on the filters was derived from the SEM micrographs taken in the backscattered electron imaging mode and analyzed with ImageJ freeware. The determination of the main mineralogical groups composing the bulk sediment associated with physical parameters such as particle size distribution or aspect ratio allows a precise characterization of the load of the terrigenous particles in rivers or lakes. In the case of the Amazonian rivers investigated, the results show that the identified mineralogies are consistent with previous studies as well as between the different samples collected. The method enabled the evolution of grain size distribution from fine to coarse material to be described in the water column. Implications about hydrodynamic sorting of mineral particles in the water column are also briefly discussed. The proposed method appears well suited for intensive routine monitoring of suspended sediment in river systems.

1. Introduction

Understanding and quantifying the processes controlling the production, transportation, and deposition of sediment in watersheds is a major issue for both river geomorphology and river management. Considering that sediment, and especially clay minerals, are key indicators of various environmental processes (Moriarty, 1977; Naidu & Mowatt, 1983; Petschick et al., 1996; Roddaz et al., 2005; Washner et al., 1999), knowledge of suspended sediment composition as well as its temporal and spatial variability in the river, provides significant insight about the evolution of watersheds. Conventional analytical procedures for water analysis such as spectrophotometry, chromatography, or spectrometry as Inductively Coupled Plasma - Atomic Emission Spectroscopy (ICP-AES) and Inductively Coupled Plasma - Mass Spectroscopy (ICP-MS) involve a heavy protocol of sample preparations and data analysis. The development of simple and efficient tools to characterize the bulk of suspended sediment in rivers is, thus, particularly needed by water agencies and the scientific community to identify any change in the distribution and in the nature of the material transported within the catchment during the hydrological cycle.

Such a need is particularly relevant in large river basins, such as the Amazon River basin, where the amount and the nature of suspended sediment may impact the environment at various scales, from global (Gibbs, 1977; Ludwig & Probst, 1998; Milliman & Meade, 1983; Syvitski et al., 2003) to local (Armijos et al., 2013; Guyot et al., 2005; Meade et al., 1985). Most often, high frequency monitoring is associated with a rough characterization of suspended sediment. The solid discharge of the Amazon River is now systematically monitored every 10 days at various sampling stations by the Observation Service program HYBAM, providing both estimates of Andean erosion rates and of the total amount of sediment transported through the Amazon River basin to the Atlantic Ocean (Filizola et al., 1999; Filizola, 2003; Guyot et al., 1988; Laraque et al., 2004; Martinez et al., 2009; Molinier

et al., 1996; Safran et al., 2005). On the other hand, there is a significant body of literature reporting detailed chemical characterizations of the particulate or dissolved load in the Amazon basin on a one time basis only (Bouchez et al., 2011; Gaillardet et al., 1997; Guyot et al., 1996; Guyot et al., 2007; Irion, 1983; Johnsson & Meade, 1990; Martinelli et al., 1993; Moquet et al., 2011; Sondag et al., 2010; Stallard & Edmond, 1983; Tardy et al., 2005), and it is now envisaged to perform a large analysis of their variations amongst the different sub-basins, using a unique, simple, and robust method.

Scanning Electron Microscopy coupled with automated Energy Dispersive X-ray Spectroscopy (SEM-EDXS) and image analysis, simultaneously provide both the size characteristics and the chemical analysis of thousands of particles. Such a method has previously been used to characterize the occurrence of heavy metal bearing mineral phases in polluted soils and sediments (de Boer & Crosby 1995; El Samrani et al., 2004) or to support the analysis of light-scattering features of particles in water bodies (Peng et al., 2009; Pinet et al., 2017).

The aim of this paper is to introduce a method that determines the mineralogical composition of suspended sediment in continental waters. This method involves a simple sampling protocol combined with the robustness of powerful statistical techniques, making it an ideal tool for processing a significant amount of water samples collected across a watershed. The main purpose is to take advantage of the expertise in mineralogical determination, then to reinforce it by the ability of machine learning tools to identify statistical patterns in large datasets, and finally to supply an efficient tool to characterize suspended sediment in lakes or rivers. In the current paper, the method is fully illustrated for suspended sediment sampled in the Amazon basin, where many one-time analyses are already available in the literature.

2. Materials and methods

2.1. Sampling

Three sampling campaigns were done between March 2013 and December 2014 by the Service National d'Observation (SNO) HYBAM network and its national partners. Water samples were taken from the surface [0-30 cm] and at various depths in rivers belonging to the Amazon basin. Eight liters of water were collected at each sampling depth, using a bucket for the surface samples and a specific depth sampler (Callède, 1994) in the water column. Turbidity was measured with a turbidity meter (Hach 2100Q) giving for each sample a proxy of the sediment load. Depending on the turbidity values, from 1 ml of water (for the samples with the highest turbidity) to 6 ml (for the samples with the lowest turbidity) were collected with a micropipette (Eppendorf), and filtered through 0.4 µm pore sized polycarbonate membrane filters (GE Osmonics). The filters were then dried and kept in PetriSlides (Millipore) until further processing in the laboratory.

2.2. Scanning electron microscopy (SEM)

After preliminary examination with a binocular microscope, a piece of membrane filter was selected, mounted on an aluminum stub using carbon double-sided conductive tape, and coated with carbon. SEM observations were done with a Jeol JSM 6363LV microscope equipped with a silicon drift detector (SDD) PGT operating at 20 kV for chemical analysis. The backscattered electron imaging mode was used to identify the mineral particles from the membrane filter (Fig. 1). A series of SEM images was randomly collected at 400x magnification, taking care not to superimpose successive images. Each identified particle was analyzed by Energy Dispersive X-ray Spectroscopy (EDXS) to determine its elemental composition. To obtain the mineralogical composition of classified groups, the stoichiometric ratios were calculated from the atomic percentages given by the EDXS analysis, and were then compared with known mineralogical compositions.

SEM images were also used to determine the size distribution of the suspended solids using the ImageJ freeware (National Institutes of Health, Bethesda, MD, U.S.) in order to compare

the size distribution of the samples. Projected area, perimeter, and shape descriptors were automatically determined for each particle with ImageJ, while excluding any particle touching the edge of the image. The equivalent diameter, D_{eq}, of each particle was calculated according to Saleh and Guigon (2009) (after Allen, 1975) as:

$$D_{eq} = (S / \pi)^{½} \quad (1)$$

where S is the projected area of the particle in μm^2. The particle size distribution (PSD) was then determined both for each sampling station and for each mineral phase identified, using Junge's power-law distribution (Junge, 1963):

$$N(D) = K \times D^{-J} \quad (2)$$

where $N(D)$ represents the number of particles of diameter D, K is the concentration of particles and the exponent J is the slope of the distribution. The size bins range from 0.14 to 550 µm. However, because of the low magnification used to acquire the SEM images, only the size classes greater than 1 µm were taken into account for the calculation of the exponent J. An aspect ratio (AR) was also obtained from the ratio between the major axis and the minor axis of the best-fit ellipse around each particle. An AR equal to 1 is a circle, whereas an AR greater than 1 indicates elongated particles.

2.3. Machine learning

The particles were partitioned into homogeneous groups based on their elemental composition using the Partitioning Around Medoids (PAM) unsupervised classification (Kaufman & Rousseeuw, 1987). The PAM method is based on a simple to use k-medoids algorithm that does robust classifications (Breitkreutz & Casey, 2008). The elemental content of particles representing 99% of the total of elements detected by the EDXS analysis was considered to derive the mineralogical groups. These elements are sodium (Na, 2.2%), magnesium (Mg, 2.7%), aluminum (Al, 25.7%), silicon (Si, 57.7%), potassium (K, 2.8%), calcium (Ca, 0.7%), titanium (Ti, 1.2%), and iron (Fe, 6.9%).

The randomForest supervised learning classification algorithm (Breiman, 2001) was applied on the same training dataset to test the robustness of the groups and to be compared with the classification resulting from the PAM method. The randomForest classification was then used on all the remaining sampling stations to extend the mineralogical determination. The randomForest classification computes hundreds of classification trees, selecting randomly *m* variables (i.e. elements) that lead to the split of the dataset at each tree node. A 'vote' is assigned to each tree built during the classification process, and then the majority vote over the set of trees built in the process designates the most consistent and recurrent classification as the final result. In the current study, the package used was "randomForest" ported in R by Liaw and Wiener (2002) from the original Fortran program by Breiman and Cutler (available at http://www.stat.berkeley.edu/users/breiman/). The sampling stations used for the training and the validation of the two classifications are described in Table 1 by their dates, locations, depth, and the corresponding numbers of SEM images and particles analyzed.

Building only two mineralogical groups would determine a minimum error rate in the classification, since if one particle does not fit in a group, it necessarily belongs to the other group. However, in that case the consistency within each group is likely to be quite poor, e.g., the intra-group variability of the particle types will be much too high to identify the resultant dominant mineralogy. In contrast, a very high number of groups would correctly classify the particles according to their chemical analysis, but the similarity between those groups would be too close, and the mineralogy would be difficult to assess and misclassification would be frequent. Therefore, the optimal number of groups achieved by the classification was investigated in order to obtain the best consistency (i.e. a single mineralogy) for each group while keeping a reasonable number of groups and minimizing the possible randomForest misclassifications. Such tests required an expertise in earth sciences in support of the statistical analysis of the classification results that was accomplished in collaboration with mineralogists. In order to reach the best agreement between PAM and randomForest

classifications, various parameterizations for the randomForest classification were tested on the training dataset.

Fig. 1. Imaging procedure and data processing from SEM micrographs of suspended particles: (left) image used in grain size analysis and (right) the same image with the particles analyzed by EDXS and manually tagged using green numbers.

2.4. Study area

The Amazon River basin is the most important watershed in the world, both in terms of drainage area (6.1×10^6 km², Goulding et al., 2003) spreading from the Andes to the Atlantic Ocean through a major part of South America, and in terms of water discharge (on average 208×10^3 m^3 s^{-1}, Callède et al., 2010) (Fig. 2). The region is characterized by a humid tropical climate (average rainfall = 2,015 ± 112 mm yr^{-1}, Espinoza Villar et al., 2009) with contrasted seasonal rainfall patterns depending on the latitude. The Amazon River main stream is known as the Solimões River from the border between Peru and Brazil until downstream of the Rio Negro confluence near Manaus. The Madeira River is the third main tributary of the Amazon River, 200 km downstream from the confluence between the Negro River and Solimões River, and drains the southeast part of the Amazon River basin (Fig. 2).

Fig. 2. Location of the sampling stations in the Amazon River basin (source *Google Earth*).

Most sediment transported by the Amazon River originates from the Andean mountains which are composed of highly erodible rocks (Dunne et al., 1998; Filizola & Guyot, 2009). Such hydro-morphological characteristics induce massive erosion processes at a large scale in Ecuador, Peru, and Bolivia (Aalto et al., 2006; Guyot, 1993; Masek et al., 1994); as a result,

the Amazonian basin is one of the most important contributors in the world in terms of sediment input to the ocean (average = 800 x 10^6 t yr^{-1}, Martinez et al., 2009). The Solimões, Madeira, and Amazon rivers are classified as white rivers (Gibbs, 1967; Sioli, 1957) because of their high sediment load, originating mainly from the erosion of the Andes, but also coming from tributaries, river bank erosion, and remobilization of the older sediments trapped in the floodplains (Dunne et al., 1998). These sediments correspond mainly to quartz and assemblages of various clays.

3. Results

3.1. Unsupervised PAM classification

The training of the classifications was done using samples from the Solimões, Madeira and Amazon rivers during the March 2013 sampling campaign (Table 1, Fig. 2), representing various water types and various terrigenous origins of suspended particles.

Table 1. Description of the sampling stations used in this study - N_i stands for the number of SEM images, and N_p for the number of particles analyzed by the machine learning process - Samples used for the PAM classification and the randomForest training are marked with *****.

Date	Station	Depth (m)	N_i	N_p
March 9, 2013	Manacapuru	0	13	1,141*
-	Foz de Madeira	0	10	1,971*
-	Amazonas	0	14	1,525*
-	Manacapuru	10	21	1,729*
April 3, 2014	Belém	0	13	1,481
April 6, 2014	Fonte Boa	0	16	1,154
-	-	10	12	1,478
-	-	20	8	1,603
-	-	30	14	1,236
April 11, 2014	Itapeua	0	14	1,390
-	-	25	12	1,787
-	-	35	8	1,720
April 13, 2014	Manacapuru	0	16	1,874
-	-	30	14	1,902

December 15, 2014	Careiro	0	16	2,291

The unsupervised PAM and the supervised randomForest were both built from the chemical analyses of 6,366 particles sampled during this campaign. The resulting randomForest classification was then used on the stations of the Solimões River sampled during the other field sampling trips (Table 1). The sampling campaigns corresponding to the period of April and December 2014 were not used in the PAM classification procedure. The classification process involved a trial-and-error phase in the choice of the number of groups. The consistency of each group was evaluated through the dispersion of the elementary contents (in percent). The average standard deviation of the total classification was, thus, calculated and compared with classifications built with a number of groups ranging from 2 to 20. A value of less than 3% expressing a great consistency within the different groups built by the PAM classification was first reached for a minimum of 15 groups. Table 2 lists the mean elementary composition of the 15 groups obtained after the PAM classification based on the 6,366 particles sampled in March 2013.

Table 2. Mean values and standard deviations of the elementary contents (in %) of each group representing 99% of the bulk particle composition and built from the unsupervised PAM classification of 6,366 suspended particles.

		Na	Mg	Al	Si	K	Ca	Ti	Fe
Group 1	Mean	0.04	0.01	0.06	99.79	0.00	0.00	0.01	0.07
	St. dev.	0.33	0.14	0.35	0.59	0.03	0.05	0.11	0.30
Group 2	Mean	1.34	11.21	26.35	35.00	1.08	0.59	0.47	21.48
	St. dev.	2.32	7.21	5.75	6.31	1.85	1.23	2.35	7.47
Group 3	Mean	1.60	2.30	36.02	49.23	8.01	0.06	0.09	2.64
	St. dev.	1.80	1.29	2.49	2.67	2.04	0.23	0.47	1.41
Group 4	Mean	1.38	3.34	30.34	54.39	3.22	0.98	0.18	5.47
	St. dev.	2.27	1.91	2.85	2.24	2.11	1.48	0.62	2.09
Group 5	Mean	2.39	0.45	20.90	60.18	14.16	0.09	0.04	0.76
	St. dev.	3.10	1.09	3.42	3.33	4.40	0.40	0.22	1.75
Group 6	Mean	18.94	0.36	21.45	56.35	0.75	1.38	0.03	0.54

	St. dev.	4.02	0.85	3.14	3.33	1.60	2.13	0.26	1.10
Group 7	Mean	0.82	2.26	25.39	63.65	1.73	0.62	0.16	4.75
	St. dev.	1.68	2.55	4.40	4.28	1.96	1.26	1.02	4.10
Group 8	Mean	2.06	5.36	29.71	47.13	2.62	1.39	0.90	9.73
	St. dev.	2.40	3.45	3.66	3.55	2.11	3.90	3.86	3.87
Group 9	Mean	1.31	118	40.90	50.22	0.96	0.17	0.13	2.74
	St. dev.	2.27	1.40	6.04	6.66	1.36	0.47	0.70	1.91
Group 10	Mean	2.08	14.88	10.00	43.35	0.52	11.47	2.03	8.92
	St. dev.	2.63	11.57	7.46	11.24	1.00	13.53	6.16	7.34
Group 11	Mean	1.16	1.40	15.20	76.31	1.61	0.28	0.47	2.96
	St. dev.	2.32	1.54	4.85	4.23	1.65	1.89	3.21	3.77
Group 12	Mean	1.30	2.28	11.34	18.22	0.77	1.27	2.26	57.64
	St. dev.	3.19	4.41	6.59	9.66	2.02	2.55	5.85	15.27
Group 13	Mean	0.64	0.65	8.11	87.61	0.83	0.07	0.17	1.71
	St. dev.	1.25	1.04	2.17	2.72	0.80	0.47	1.07	1.84
Group 14	Mean	0.48	0.97	9.61	13.32	0.33	0.21	69.75	4.28
	St. dev.	1.58	1.79	5.22	7.57	0.65	0.69	16.59	5.96
Group 15	Mean	0.19	0.05	4.02	94.93	0.16	0.00	0.04	0.55
	St. dev.	0.69	0.28	1.20	1.79	0.41	0.07	0.26	0.80

This statistical consistency of the classification was confirmed by the expertise of mineralogists and enabled the association of a unique mineralogy to each group. Two different groups may, however, be attributed to the same mineralogy as natural particles such as clays can occur in a range of elemental compositions. These 15 groups represent the reference from which the remainder of the statistical procedure was applied (i.e. the training of the randomForest classification). The classification reveals an increase in the Al/Si ratio from group 1 (defined as pure quartz with Al/Si close to 0) to the most developed and mature sediments of group 9 (defined as kaolinite with Al/Si close to 1). The classification also detected two poorly represented groups that were very distinct from the others, with major contents in iron or titanium (respectively, iron oxides, and titanium dioxides). Except for these two groups representing less than 3% of the dataset, the classification distinguished 8 groups of clays (represented by chlorite, smectite, illite, and kaolinite) corresponding to 65% of the classified particles, 3 groups of quartz (22% of the dataset) and 2 groups of feldspars (orthoclase and albite, 3% and 7% of the dataset, respectively). The same mineralogies were

established for different groups expressing either a chemical variability within a group of minerals (e.g., the 3 groups of smectite), or the mixing of different particles signatures by the SEM-EDXS measurements (e.g., the 3 groups of silica particles).

3.2. Supervised randomForest classification

By varying the number of trees, N_{tree}, built during the classification and the number of variables, N_{var}, sampled at each split, the minimum error rate was obtained for N_{tree} = 500 and N_{var} = 2. The most representative tree of the forest having the majority 'votes' and corresponding to the best split of the dataset is shown in Fig. 3. This tree highlights the importance of Si, Al, K, Fe, and Na contents in the classification process. On the other hand, the 3 remaining elements (Mg, Ti, and Ca) did not significantly contribute to the repartition of the chemical analyses into distinct groups. These results were confirmed by the internal calculations of variables importance of randomForest called *MeanDecreaseAccuracy* and *MeanDecreaseGini* (Breiman, 2002), which assign the lowest values (i.e. the minimum contribution) to Mg, Ti, and Ca, and the highest values to the Si content.

Fig. 3. The classification tree resulting from the randomForest process and the splitting criterion used to separate the dataset at each node of the tree. Bold numbers on the terminal nodes represent the corresponding classification group.

3.3. Comparison of the classifications

The resulting tree and the corresponding supervised randomForest classification were compared to the PAM partitioning. Table 3 shows a confusion matrix where the misclassification between PAM and randomForest are detailed and quantified via the classification error rate. The entire process was repeated because of the random aspect of the variables used during the classification. All the confusion matrices thus obtained showed

similar results, and the classification error rate between both classification methods was always lower than 4% using the combined (N_{tree} = 500 and N_{var} = 2) parametrization. The randomForest classification was used on the remaining dataset, which included stations sampled during the campaigns of April and December 2014 (i.e., 17,916 particles). The mineralogical content of each Solimões River station were then determined from the number percentage of particles associated with the different groups (Fig. 4).

Table 3. Example of a confusion matrix among the different groups classified by the PAM and the randomForest machine learning.

		Supervised randomForest classification															Class
		G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	error
Unsupervised PAM classification	G1	662	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.00150
	G2	0	399	1	0	0	0	1	12	3	2	0	2	0	0	0	0.05000
	G3	0	0	596	17	0	0	0	8	9	0	0	0	0	0	0	0.05396
	G4	0	0	5	1120	0	5	11	13	7	0	0	0	0	0	0	0.03531
	G5	0	0	0	1	174	1	3	0	0	3	1	0	0	0	0	0.04918
	G6	0	0	0	0	2	436	1	0	0	0	1	0	0	0	0	0.00909
	G7	0	0	0	20	0	0	400	1	0	0	4	0	0	0	0	0.05882
	G8	0	4	4	16	0	0	0	588	7	1	0	0	0	0	0	0.05161
	G9	0	0	6	6	0	0	0	7	467	0	0	0	0	0	0	0.03909
	G10	0	4	0	0	0	1	0	4	0	109	1	0	0	0	0	0.08403
	G11	0	0	0	0	1	3	8	0	0	4	234	0	3	0	0	0.07509
	G12	0	3	0	0	0	0	0	0	0	0	0	163	0	0	0	0.01807
	G13	0	0	0	0	0	0	0	0	0	1	7	0	301	0	4	0.03833
	G14	0	1	0	0	0	0	0	0	0	0	0	1	0	60	0	0.03225
	G15	1	0	0	0	0	0	0	0	0	0	0	0	9	0	415	0.02352

As illustrated in Fig. 4, the Solimões River samples exhibit a very similar pattern in mineralogical composition, with a maximum standard deviation of σ = 2.5%. The comparison between Solimões River and Madeira River stations exhibits contrasting dissimilarities, highlighting the differences in mineralogical composition of the suspended sediments collected during the March 2013 sampling trip and used for the training of the randomForest classification. For all the Solimões River stations, the smectite content (sum of groups 3, 4,

and 10) represents the main component of the mineralogical assemblage (33.01% on average). Group 4 is especially dominant representing on average 19.12% of all particles analyzed by SEM. The remainder of the particles occurs as other clay minerals (illite, chlorite, and kaolinite, on average 19.74, 7.38, and 6.78%, respectively), quartz (on average 19.17%) and feldspars (orthoclase and albite, on average 3.03 and 6.87%, respectively). Finally, the less frequent types of particles are iron oxides and titanium dioxides (on average 2.90 and 1.12%, respectively).

Fig. 4. Distribution of the mineralogical groups from 1 to 15 at 3 different stations on the Solimões River sampled in a) March 2013, c) April 2014, and d) December 2014, and a Madeira River station b) also sampled in March 2013.

The average Pearson's correlation coefficients between Solimões River sampling stations taken two by two reached 0.90, but reached only 0.73 when comparing Madeira River and Solimões River stations. After combining the groups showing analogous mineralogies (e.g., groups 1, 13, and 15 representing quartz), the correlation coefficients are even higher, increasing to 0.97 on average for the Solimões River stations. The mineralogical similarities evidenced in Fig. 4 between Solimões River sampling stations were confirmed by variance analysis, which showed no statistical difference between the randomForest classifications of the Solimões River stations (mean adjusted p-value = 0.96), but established a statistical distinction between the Madeira and the Solimões sampling stations (mean adjusted p-value = 0.48).

3.4. Particle Size Distribution

A total of more than 204,000 particles were scanned by SEM to analyze their size, which represents more than 15,700 particles per station on average. Contrasted types of particles

were examined with projected areas ranging between 0.062 and 25,204 µm² (5.78 µm² on average) and perimeters ranging between 0.70 and 1,874 µm (6.02 µm on average). The resulting equivalent diameters were calculated and ranged between 0.14 and 69.57 µm (0.76 µm on average) (Fig. 5a), which led to the calculation of the grain size distribution of the particles for each station (Fig. 5b), following the least-squares fitting method. The smallest value of the slope exponent was obtained for the suspended particles taken at a 30 m depth at Manacapuru (J = 2.54) and the highest value was for the surface sample of Fonte Boa station (J = 3.36). As expected, a decrease of J was observed with sampling depth in the water column, thus, showing an increase of the coarser fraction of the particles (Fig. 5c). As an example, at the Itapeua station, J is 3.14 at the surface level, 2.96 at a 25 m depth, and is only 2.62 at a 35 m depth. On average, the J value for surface samples is about 3.15, but only reached 2.84 for samples collected at depth. The particle size distribution of each mineralogical group was examined. J values ranging from 2.48 for group 14 to 3.43 for group 4 were obtained. Similar J values (3.31) characterized the smectite groups (including groups 3, 4, and 11), the lowest J values were obtained for iron oxides (2.60, group 12) and with titanium dioxides (2.48, group 14), whereas the other groups (quartz, chlorite, illite, orthoclase, albite, and kaolinite) exhibited a Junge's exponent close to 3 (3.01 on average).

The majority of particles showed aspect ratios in the [1 – 2] range, which reveals that most of the particles were significantly rounded. Such a result was confirmed by the calculation of the circularity coefficient, C (included in the [0 – 1] range) that corresponds to the ratio between the projected area S and the squared value of perimeter P of each particle ($C = 4\pi \times S / P^2$). More than 45% of the particles showed a C value equal to 1, thus, indicating a spherical particle, and it increased to more than 59% for C values higher than 0.9. Those quasi-spherical particles showing an AR value of 1 were characterized by small dimensions with a maximum area of 2 µm² (0.11 µm² on average) and a maximum perimeter of 4.83 µm (0.90 µm on average).

Fig. 5. Grain size distribution of the particles analyzed by SEM: a) histogram of the calculated equivalent diameters; b) particle number/size distribution at each sampling station (for D > 1 μm); c) corresponding Junge's distribution exponent, *J*, values at various depths (black and white triangles represent Manacapuru station sampled in 2013 and 2014, respectively. Black and white circles represent Fonte Boa station and Itapeua station, respectively. Black and white squares represent Belém station and Careiro station, respectively).

4. Discussion

With more than 96% of particles getting a correct assignment (Table 3), the supervised randomForest classification ensures great consistency with the groups built from the unsupervised PAM method. The relative importance of the mineralogical groups, thus, obtained compares favorably with those inferred by X-ray diffraction in Guyot et al. (2007), at similar locations in the Amazon basin (including the Solimões River). These authors used a semi-quantitative method to manually identify the clay minerals present in each sample collected from the banks of the different rivers. In the Brazilian floodplain, Guyot et al. (2007) found approximatively 25% by weight (wt.) for the illite/chlorite assemblage, 10% wt. for kaolinite, and 65% wt. for the smectite/mixed layers assemblage for the Solimões River. In the current case, excluding quartz from the analysis, the composition of the same assemblages are on average 33.55, 8.39, and 58.06% (percent in number), respectively. The clay content in the Solimões River retrieved by the two methods are very close, the smectite content being predominant, and the amount of kaolinite, which indicates the presence of mature sediments, being significant. Despite the analysis of more than 200,000 particles, the initial sampling protocol and the magnification used for the SEM observations preclude the consideration of large particles in the size distribution. Indeed, very few particles larger than

10 µm² and almost none larger than 20 µm² were collected, thus, leading to an overestimation of the *J* exponent of Junge's law which describes the particle size distribution. The absence of the coarse size classes may smooth the differences in the PSD between samples collected at different depths within the water column. Nevertheless, such underestimation of large particles does not conceal the decrease in Junge's parameter, *J*, with increasing depth at a given sampling site, either considering every size class greater than 1 µm, or just taking into account size classes between 1 and 10 µm. In other words, an increasing proportion of coarse sedimentary materials is observed with increasing depth, as illustrated in Fig. 6.

Fig. 6. Two SEM images (x400 magnification) corresponding to the Manacapuru station sampled at the same scale in April 2014, a) at the water surface and b) at 30 m deep.

On the other hand, the analysis of the PSD of the various mineralogical groups showed more contrast. Indeed, iron oxides and titanium dioxides are characterized by coarse material, whereas smectites, and especially those associated with group 4, are mainly composed of small particles. Such results are consistent with Chipera and Bish (2001), who showed that, among various types of clays, smectite represented the finest size fraction. However, there was no significant difference between the other groups in terms of particle size distribution. Slight variations were observed, but the general pattern led to a *J* value of approximately 3. Regarding the aspect ratio analysis, Li et al. (2016) have revealed that particles from deep rivers are generally characterized by a regular shape, with slight roughness. The current results in terms of sediments shape, with more than 59% of particles characterized by a roundness coefficient typical of a quasi-spherical shape, are consistent with that study. It is worth noting that the sphericity index of particles is inferred from projected surfaces and not from direct volume measurements. Thus, taking into account that clays occur as flake-shaped

particles, it is expected that clay minerals present a less spherical shape compared with that of other silicates, such as quartz or feldspar. These differences in mineral shapes are likely to be magnified for coarser materials (Krinsley & Smalley, 1973). It will induce *de facto* an increasing dissociability of the number of particles and associated volume between clays and quartz sediments when considering deeper samples in the water column. It then becomes obvious that the absolute content in Si will increase with depth even if the relative number percentage of quartz particles remains unchanged. On the contrary, the volume percent occupied by clays will decrease with depth, considering that quartz minerals occupy more volume than clay minerals. Thus, using the Al/Si ratio, Bouchez et al. (2011) showed profiles describing a direct relation between depth and mineralogical composition, and suggested that coarser materials are Si enriched. In contrast, the current results indicate that, considering the composition of single particles instead of the bulk composition of suspended sediment, there is no significant pattern between Al/Si ratio and depth (Fig. 7a). In addition, no specific trend allows the description of a split between the different mineralogies based on sampling depth (Figs. 7a and 7b).

Fig. 7. Mineralogical composition of suspended sediments of the Solimões River stations at various depths: a) the Al/Si ratio, illite, chorite, kaolinite and orthoclase contents; b) the smectite, quartz, albite, iron oxide and titanium oxide contents.

The Si content is then truly higher and the Al content is indeed lower as the sampling depth increases, but this should not be related to an increased proportion of quartz or feldspar particles compared with clay minerals. Obviously, the latter result is somehow related to the range of particle size explored in the current study by SEM, and should be confirmed by extending the investigation to both finer and coarser materials. Nevertheless, it should be

pointed out that the levels of turbulence (and, thus, of vertical mixing) were likely different in stations sampled by Bouchez et al. (2011) and in the current study. Such a parameter may be estimated in future sampling to support the analysis of vertical profiles (e.g., Lefebvre et al., 2012).

5. Conclusions

The new PAM-randomForest procedure presented in this paper shows robust results and offers an easy method to characterize suspended sediments. The determination technique of the mineralogical assemblages in rivers is highly representative of the total sediment, even though just a few milliliters of water were sampled. It remains a powerful procedure even in the presence of a large dataset, which is essential in a spirit of maximum representativeness of the measurements. It also allows a more precise characterization of the suspended sediment sampled compared with an estimate of the total mass of the bulk particles. This method is particularly interesting for intensive sampling programs, where a series of samplings are made and where it is easier to keep filters instead of several litters of water or kilograms of sediments. Complementary analyses have to be made to confirm the distribution of the different mineralogies within the water column for larger grain sizes. However, as part of a long term monitoring program, this technique could be routinely used and would provide important information on the sediment content and its variability in large river basins.

Acknowledgements

The authors are grateful to the French Centre National d'Etudes Spatiales (CNES) and the French company NOVELTIS for co-funding the first author's Ph.D. This work was also supported by the Brazilian Agência Nacional de Águas (ANA) for funding the field campaigns. The authors thank also Thierry Aigouy of the French Laboratory Géosciences Environnement Toulouse (UMR 5563), Sophie Gouy from the French Centre de Micro

Caractérisation Raimond Castaing (UMS 3623), and Jacques Yvon from the French laboratory GeoRessources (UMR 7359) for technical and scientific support.

References

Aalto, R., Dunne, T., & Guyot, J. L. (2006). Geomorphic controls on Andean denudation rates. *The Journal of Geology*, *114*(1), 85-99.

Allen, T. (1975). *Particle size measurement*. Bradford: Springer.

Armijos, E., Laraque, A., Barba, S., Bourrel, L., Ceron, C., Lagane, C., …Guyot, J. L. (2013). Yields of suspended sediment and dissolved solids from the Andean basins of Ecuador. *Hydrological Sciences Journal*, *58*(7), 1478-1494.

Bouchez, J., Gaillardet, J., France-Lanord, C., Maurice, L., & Dutra-Maia, P. (2011). Grain size control of river suspended sediment geochemistry: Clues from Amazon River depth profiles. *Geochemistry, Geophysics, Geosystems*, *12*(3).

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Breiman, L. (2002). *Manual on setting up, using, and understanding random forests v3.1*. Statistics Department, University of California at Berkeley, U.S.

Breitkreutz, D., & Casey, K. (2008). *Clusterers: A comparison of partitioning and density-based algorithms and a discussion of optimisations*. Technical Report 11999, James Cook University, Townsville, Australia, 14 p.

Callède, J. (1994). Coletor de amostra d'água. *Publication HiBAm*, Brasília, 16 p. (*In Portuguese*)

Callède, J., Cochonneau, G., Alves, F. V., Guyot, J. L., Guimarães, V. S., & De Oliveira, E. (2010). Les apports en eau de l'Amazone à l'Océan Atlantique. *Revue des Sciences de l'Eau/Journal of Water Science*, *23*(3), 247-273. (*In French*)

Chipera, S. J., & Bish, D. L. (2001). Baseline studies of the clay minerals society source clays: Powder X-ray diffraction analyses. *Clays and Clay Minerals*, *49*(5), 398-409.

de Boer, D.H., & Crosby, G. (1995). Evaluating the potential of SEM/EDS analysis for fingerprinting suspended sediment derived from two contrasting topsoils. *Catena, 24,* 243–258.

Dunne, T., Mertes, L. A., Meade, R. H., Richey, J. E., & Forsberg, B. R. (1998). Exchanges of sediment between the flood plain and channel of the Amazon River in Brazil. *Geological Society of America Bulletin, 110*(4), 450-467.

El Samrani, A. G., Lartiges, B. S., Ghanbaja, J., Yvon, J., & Kohler, A. (2004). Trace element carriers in combined sewer during dry and wet weather: An electron microscope investigation. *Water Research, 38,* 2063-2076.

Espinoza Villar, J. C., Ronchail, J., Guyot, J. L., Cochonneau, G., Filizola, N., Lavado, W., …Vauchel, P. (2009). Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *International Journal of Climatology, 29,* 1574–1594. doi:10.1002/joc.1791

Filizola, N. (2003). *Transfert sédimentaire actuel par les fleuves amazoniens. (Ph.D. dissertation),* Université Paul-Sabatier, Toulouse III, France. (*In French*)

Filizola, N., & Guyot, J. L. (2009). Suspended sediment yields in the Amazon basin: An assessment using the Brazilian national data set. *Hydrological Processes*, *23*(22), 3207-3215.

Filizola, N., Guyot, J. L., & Boaventura, G. (1999). Fluxo de sedimentos em suspensão na Amazônia– a análise a partir da base de dados da ANEEL. *Proceedings of the Hydrological and Geochemical Processes in Large Scale River Basins,* Manaus, 99. (*In Portuguese*)

Gaillardet, J., Dupré, B., Allègre, C. J., & Négrel, P. (1997). Chemical and physical denudation in the Amazon River basin. *Chemical Geology, 142*(3), 141-173.

Gibbs, R. J. (1967). Amazon rivers: Environmental factors that control its dissolved and suspended load. *Science, 156,* 1734-1737.

Gibbs, R. J. (1977). Clay mineral segregation in the marine environment. *Journal of Sedimentary Research*, *47*(1), 237-243.

Goulding, M., Barthem, R., & Ferreira, E. (2003). *The Smithsonian atlas of the Amazon*. Washington, D.C.: Smithsonian Books.

Guyot, J. L. (1993). *Hydrogéochimie des fleuves de l'Amazonie bolivienne. (Ph.D. dissertation),* Editions de l'ORSTOM, Collection Etudes et Thèses, Paris, France.

Guyot, J. L., Bourges, J., Hoorelbecke, R., Roche, M. A., Calle, H., Cortes, J., & Guzman, M. C. B. (1988). Exportation de matières en suspension des Andes vers l'Amazonie par le Rio Béni, Bolivie. In: M. P. Bordas & D. E. Walling (Eds.), *Porto Alegre Symposium, 1988, Sediment Budgets* (pp. 443-451), *IAHS Publication 174*. (*In French*)

Guyot, J. L., Filizola, N., & Laraque, A. (2005). Régime et bilan du flux sédimentaire de l'Amazone à Óbidos (Pará, Brésil) de 1995 à 2003. In: D. E. Walling & A. J. Horowitz (Eds.), *Sediments Budgets 1* (pp. 347-354) *IAHS Publication 291*. (*In French*)

Guyot, J. L., Filizola, N., Quintanilla, J., & Cortez, J. (1996). Dissolved solids and suspended sediment yields in the Rio Madeira basin, from the Bolivian Andes to the Amazon. In: D. E. Walling, & B.W. Webb (Eds.), *Exeter symposium, 1996, Erosion and sediment yield: Global and regional perspectives* (pp. 55-63), *IAHS Publication 236.*

Guyot, J. L., Jouanneau, J. M., Soares, L., Boaventura, G. R., Maillet, N., & Lagane, C. (2007). Clay mineral composition of river sediments in the Amazon Basin. *Catena*, *71*(2), 340-356.

Irion, G. (1983). Clay mineralogy of the suspended load of the Amazon and of rivers in the Papua-New Guinea mainland. *Transport of carbon and minerals in major world rivers*, *2*, 482-504.

Johnsson, M. J., & Meade, R. H. (1990). Chemical weathering of fluvial sediments during alluvial storage: The Macuapanim Island point bar, Solimoes River, Brazil. *Journal of Sedimentary Research, 60*(6), 827-842.

Junge, C. E. (1963). *Air chemistry and radioactivity*. New York: Academic Press.

Kaufman, L., & Rousseeuw, P. (1987). Clustering by means of medoids. In: Y. Dodge (Ed.), *Statistical data analysis based on the L1 norm and related methods* (pp. 405-416) Amsterdam: North-Holland.

Krinsley, D. T., & Smalley, I. J. (1973). Shape and nature of small sedimentary quartz particles. *Science, 180*(4092), 1277-1279.

Laraque, A., Ceron, C., Armijos, E., Pombosa, R., Magat, P., & Guyot, J. L. (2004). Sediment yields and erosion rates in the Napo River basin: An Ecuadorian Andean Amazon tributary. In: V. Golosov, V. Belyaev, & D. E. Walling (Eds.), *Sediment transfer through the fluvial system,* (pp. 220- 225), *IAHS Publication 288.*

Lefebvre, J. P., Ouillon S., Vinh, V. D., Arfi, R., Panche, J. Y., Mari, X., …Torréton, J. P., (2012). Seasonal variability of cohesive sediment aggregation in the Bach Dang-Cam Estuary, Haiphong (Vietnam). *Geo-Marine Letters*, *32*(2), 103-121.

Li, D., Li, Y., Wang, Z., Wang, X., & Li, Y. (2016). Quantitative, SEM-based shape analysis of sediment particles in the Yellow River. *International Journal of Sediment Research*, *31*(4), 341-350.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*, 18–22.

Ludwig, W., & Probst, J. L. (1998). River sediment discharge to the oceans; present-day controls and global budgets. *American Journal of Science*, *298*(4), 265-295.

Martinelli, L. A., Victoria, R. L., Dematte, J. L. I., Richey, J. E., & Devol, A. H. (1993). Chemical and mineralogical composition of Amazon River floodplain sediments, Brazil. *Applied Geochemistry*, *8*(4), 391-402.

Martinez, J. M., Guyot, J. L., Filizola, N., & Sondag, F. (2009). Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data. *Catena*, *79*(3), 257-264.

Masek, J. G., Isacks, B. L., Gubbels, T. L., & Fielding, E. J. (1994). Erosion and tectonics at the margins of continental plateaus. *Journal of Geophysical Research*, *99*, 13941-13956.

Meade, R. H., Dunne, T., Richey, J. E., Santos, U. D. M., & Salati, E. (1985). Storage and remobilization of suspended sediment in the lower Amazon River of Brazil. *Science*, *228*(4698), 488-490.

Milliman, J. D., & Meade, R. H. (1983). World-wide delivery of river sediment to the oceans. *The Journal of Geology*, *9*(1), 1-21.

Molinier, M., Guyot, J. L., De Oliveira, E., & Guimarães, V. (1996). Les régimes hydrologiques de l'Amazone et de ses affluents. In: P. Chevallier, & B. Pouyaud (Eds.), *Tropical hydrology: A geoscience and a tool for sustainability, dedicated to the memory of Jean Rodier, Paris, 1995*, (pp. 209-222), *IAHS Publication 238*. (*In French*)

Moquet, J. S., Crave, A., Viers, J., Seyler, P., Armijos, E., Bourrel, L., …Guyot, J. L. (2011). Chemical weathering and atmospheric/soil CO2 uptake in the Andean and Foreland Amazon basins. *Chemical Geology*, *287*(1), 1-26.

Moriarty, K. C. (1977). Clay minerals in Southeast Indian Ocean sediments, transport mechanisms and depositional environments. *Marine Geology*, *25*(1-3), 149-174.

Naidu, A. S., & Mowatt, T. C. (1983). Sources and dispersal patterns of clay minerals in surface sediments from the continental-shelf areas off Alaska. *Geological Society of America Bulletin*, *94*(7), 841-854.

Peng, F., Effler, S. W., Pierson, D. C., & Smith, D. G. (2009). Light-scattering features of turbidity causing particles in interconnected reservoir basins and a connecting stream. *Water Research, 43*, 2280-292.

Petschick, R., Kuhn, G., & Gingele, F. (1996). Clay mineral distribution in surface sediments of the South Atlantic: Sources, transport, and relation to oceanography. *Marine Geology, 130*(3), 203-229.

Pinet, S., Martinez, J. M., Ouillon, S. Lartiges, B., & Villar, R. E. (2017). Variability of apparent and inherent optical properties of sediment-laden waters in large river basins - Lessons from *in situ* measurements and bio-optical modeling. *Optics Express, 25*(8), A283-A310.

Roddaz, M., Viers, J., Brusset, S., Baby, P., & Hérail, G. (2005). Sediment provenances and drainage evolution of the Neogene Amazonian foreland basin. *Earth and Planetary Science Letters, 239*(1), 57-78.

Safran, E. B., Bierman, P. R., Aalto, R., Dunne, T., Whipple, K. X., & Caffee, M. (2005). Erosion rates driven by channel network incision in the Bolivian Andes. *Earth Surface Processes and Landforms, 30*(8), 1007-1024.

Saleh, K., & Guigon, P. (2009). Caractérisation et analyse des poudres: Propriétés physiques des solides divisés. *Techniques de l'ingénieur. Génie des Procédés,* (J2251). (*In French*)

Sioli, H. (1957). Sedimentation im Amazonasgebiet. *Geologische Rundschau, 45*(3), 608-633.

Sondag, F., Guyot, J. L., Moquet, J. S., Laraque, A., Adele, G., Cochonneau, G., …Vauchel, P. (2010). Suspended sediment and dissolved load budgets of two Amazonian rivers from the Guiana Shield: Maroni River at Langa Tabiki and Oyapock River at Saut Maripa (French Guiana). *Hydrological Processes, 24*(11), 1433-1445.

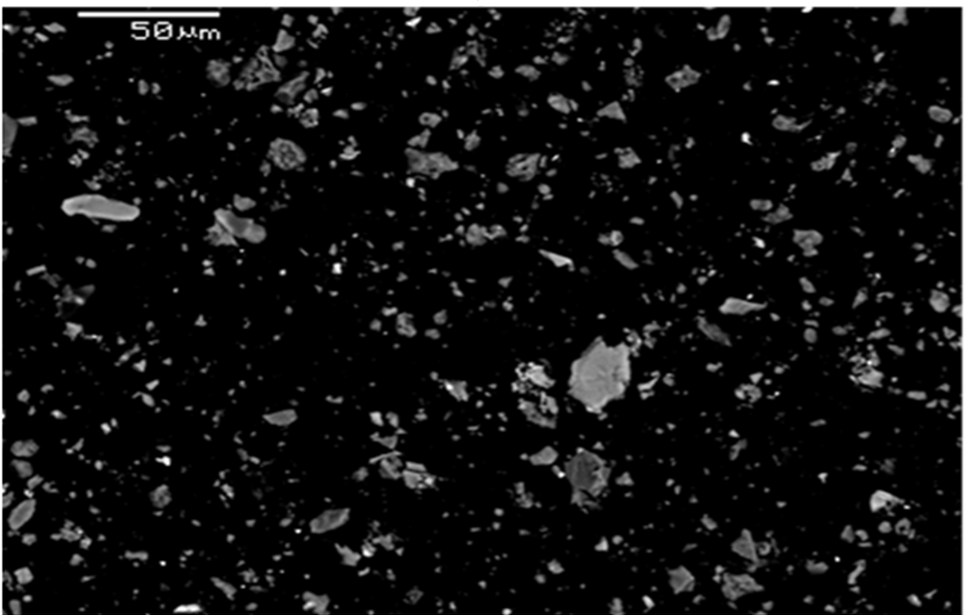
Stallard, R. F., & Edmond, J. M. (1983). Geochemistry of the Amazon: 2. The influence of geology and weathering environment on the dissolved load. *Journal of Geophysical Research: Oceans*, *88*(C14), 9671-9688.

Syvitski, J. P., Peckham, S. D., Hilberman, R., & Mulder, T. (2003). Predicting the terrestrial flux of sediment to the global ocean: A planetary perspective. *Sedimentary Geology*, *162*(1), 5-24.

Tardy, Y., Bustillo, V., Roquin, C., Mortatti, J., & Victoria, R. (2005). The Amazon. Bio-geochemistry applied to river basin management: Part I. Hydro-climatology, hydrograph separation, mass transfer balances, stable isotopes, and modelling. *Applied Geochemistry*, *20*(9), 1746-1829.

Washner, M., Müller, C., Stein, R., Ivanov, G., Levitan, M., Shelekhova, E., & Tarasov, G. (1999). Clay-mineral distribution in surface sediments of the Eurasian Arctic Ocean and continental margin as indicator for source areas and transport pathways - a synthesis. *Boreas*, *28*(1), 215-233.

Scanning Electron Microscopy (SEM)

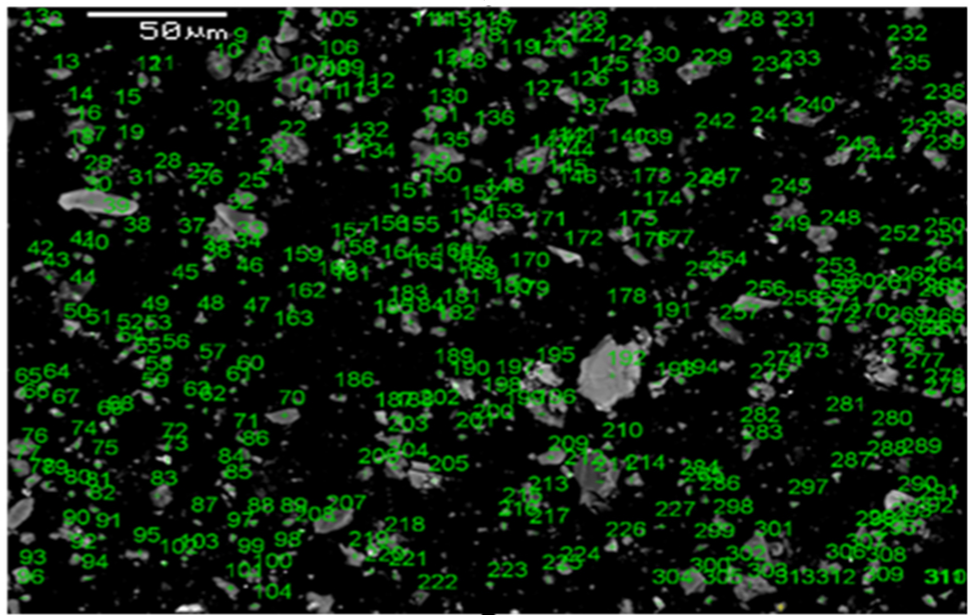


Physical parameters (areas, perimeters…)

Equivalent diameters

Grain size distribution – Junge's coefficient *J*

Energy Dispersive X-ray Spectroscopy (EDXS)

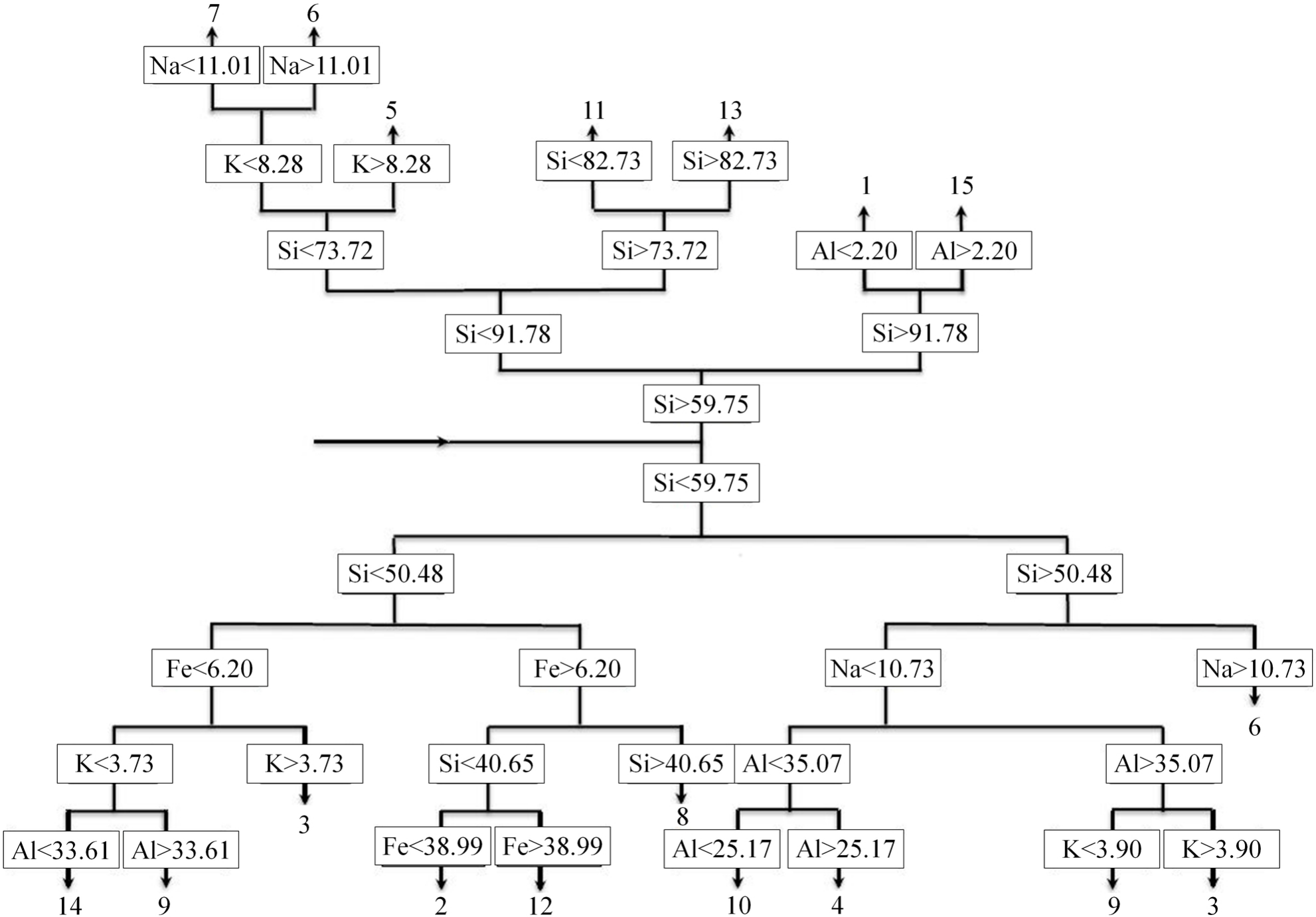


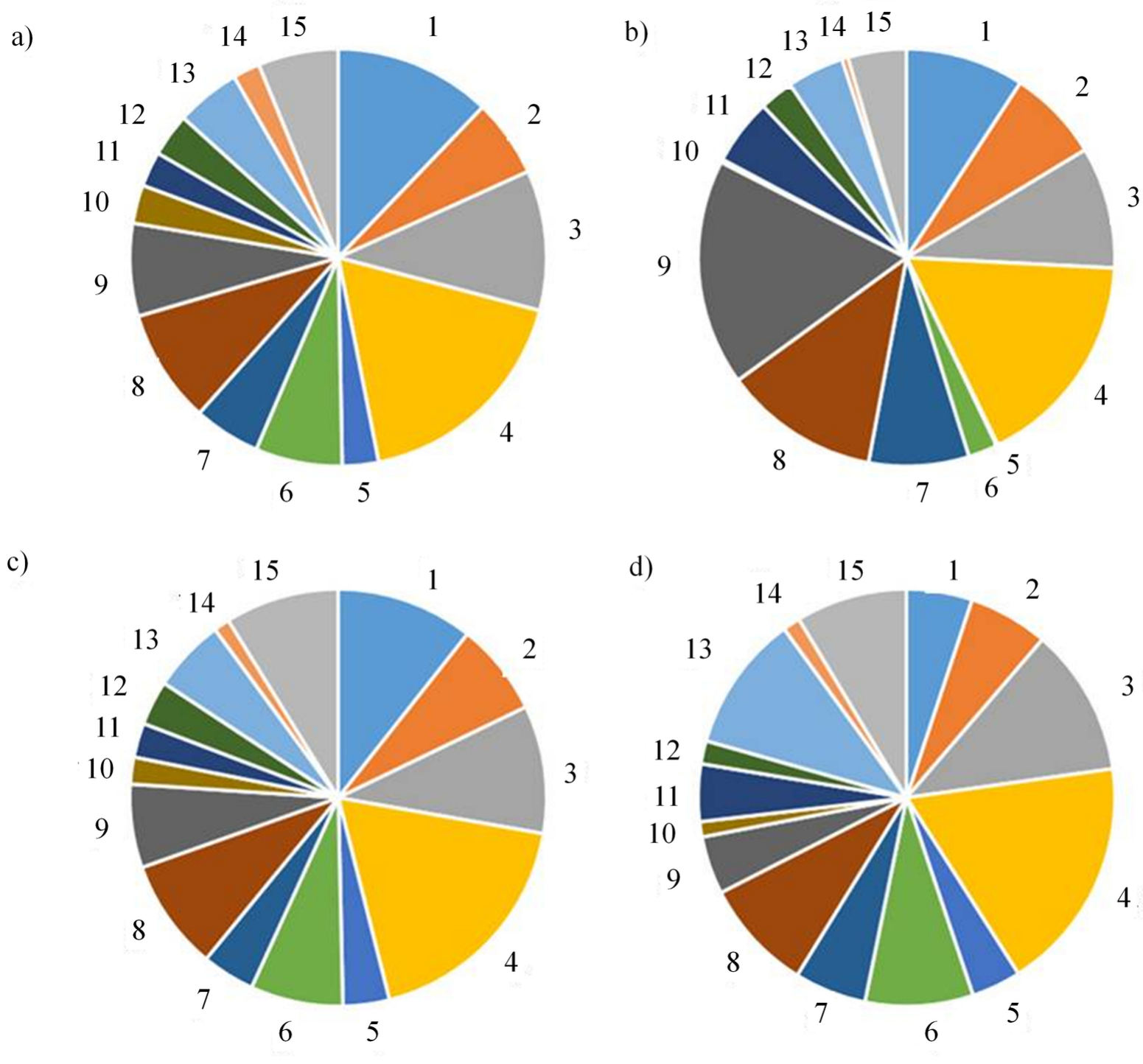
Chemical parameters (Al, Si, Fe…)

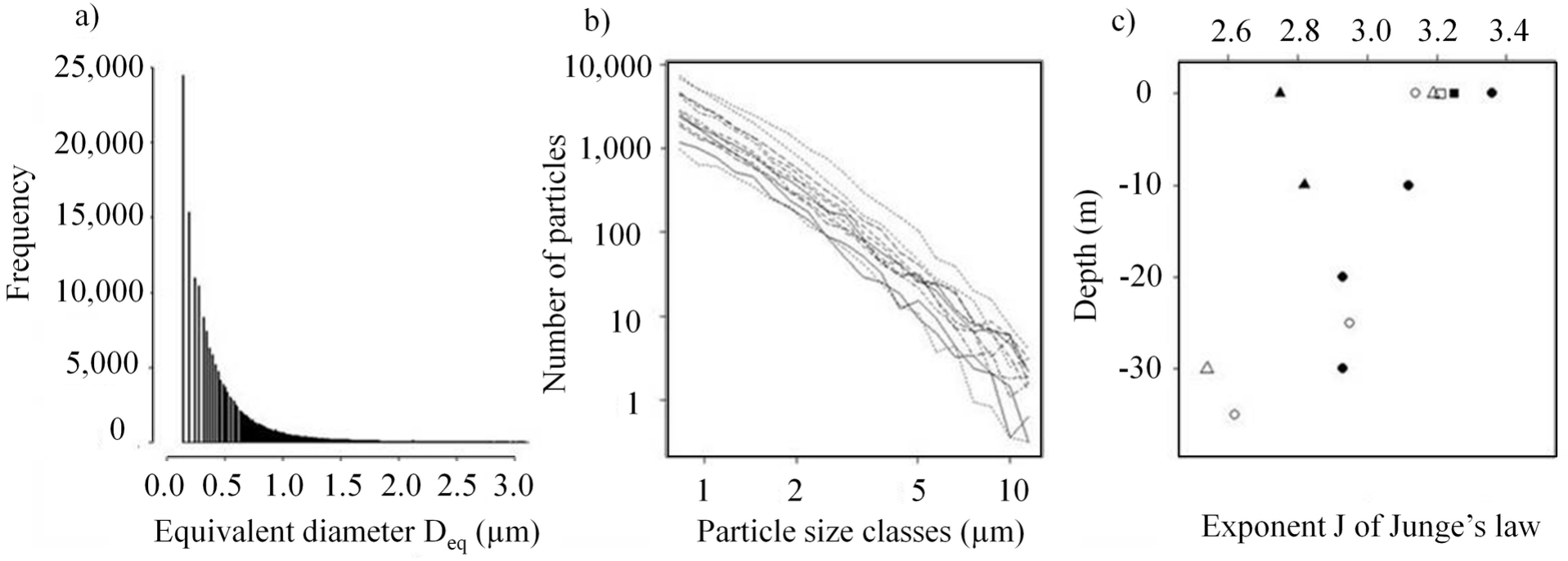
Machine learning (PAM, randomForest)

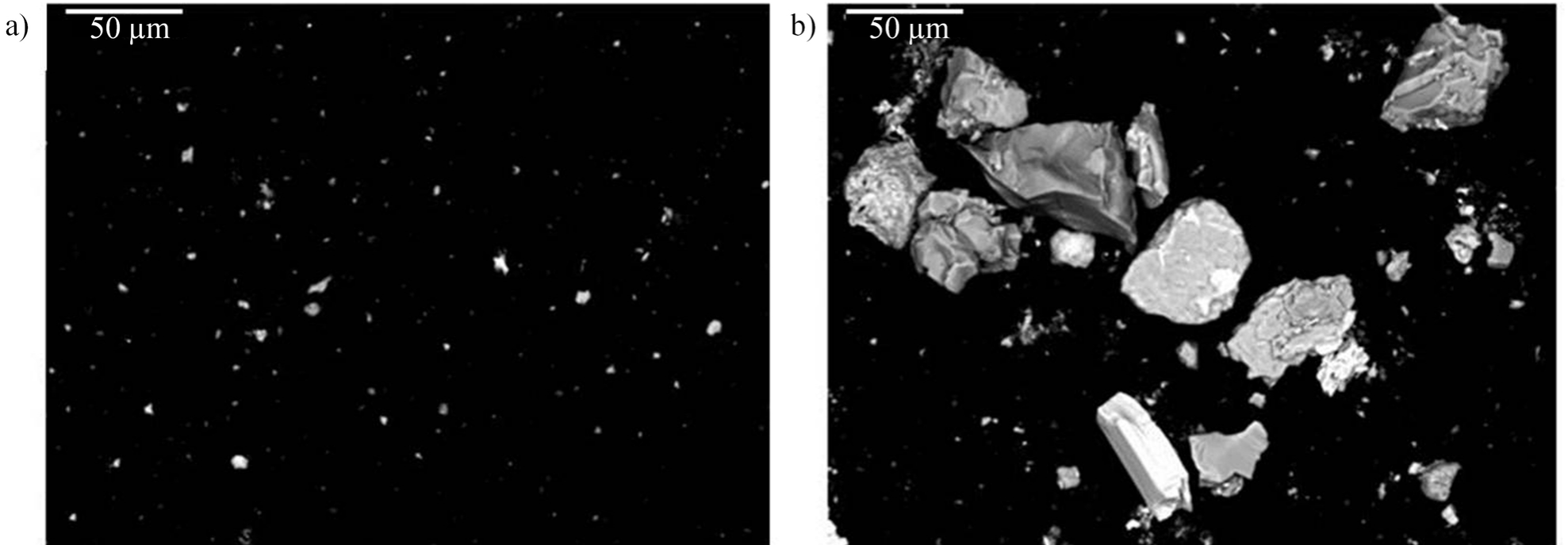
Mineralogical composition

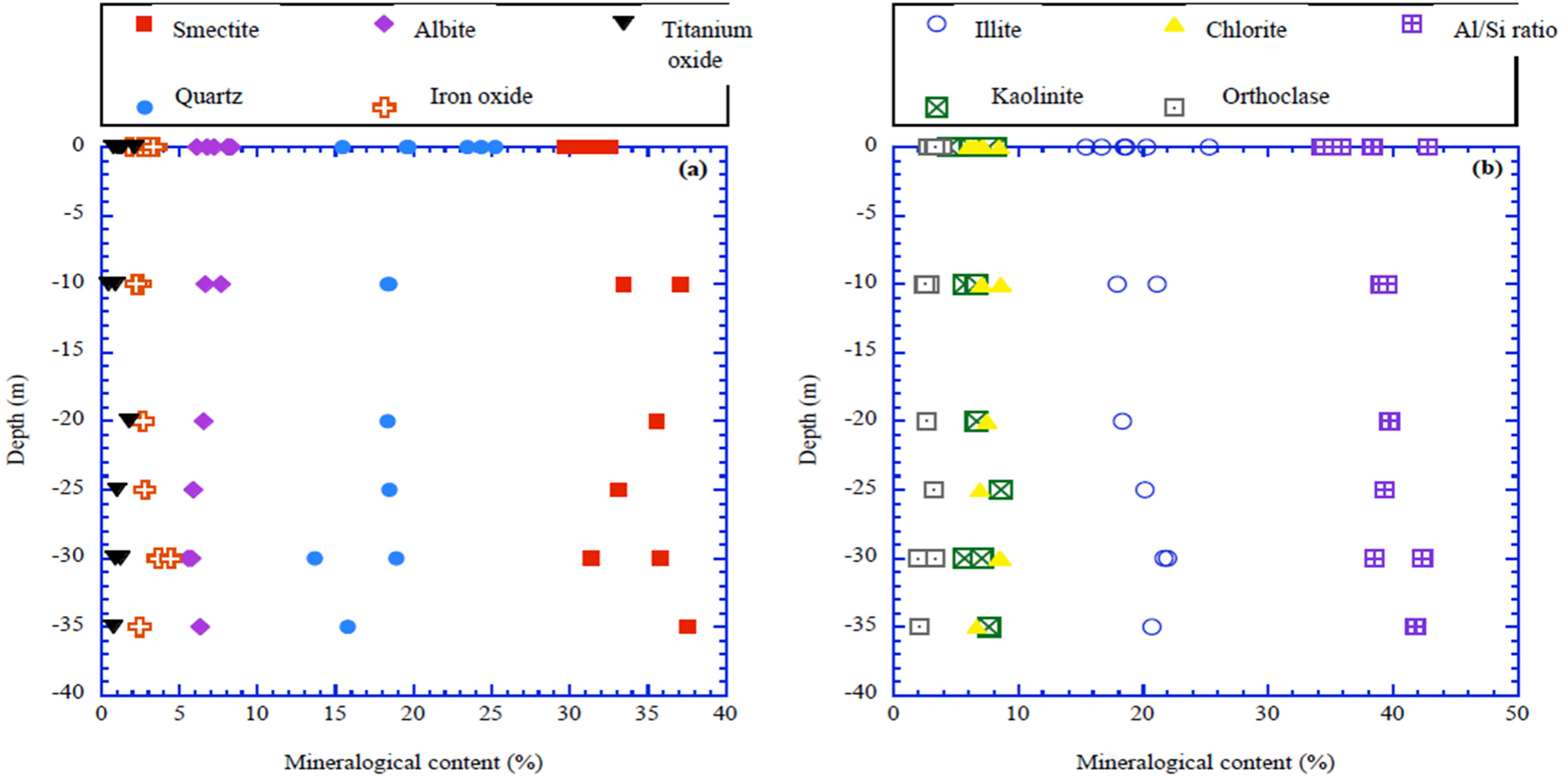

N
Fonte Boa
Negro River
Solimões River
Amazonas
Manacapuru
Careiro
Foz do
Madeira
Belém
Itapeua
Madeira River
Sampling stations
0
50 km


7
6
Na<11.01
Na>11.01
5
11
13
K<8.28
K>8.28
Si<82.73
Si>82.73
1
15
Si<73.72
Si>73.72
Al<2.20
Al>2.20
Si<91.78
Si>91.78
Si>59.75
Si<59.75
Si<50.48
Si>50.48
Fe<6.20
Fe>6.20
Na<10.73
Na>10.73
6
K<3.73
K>3.73
Si<40.65
Si>40.65
Al<35.07
Al>35.07
3
8
Al<33.61
Al>33.61
Fe<38.99
Fe>38.99
Al<25.17
Al>25.17
K<3.90
K>3.90
14
9
2
12
10
4
9
3


a)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
b)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
c)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
d)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15


a)
25,000
20,000
15,000
10,000
5,000
0
Frequency
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Equivalent diameter D_{eq} (µm)
b)
10,000
1,000
100
10
1
Number of particles
1 2 5 10
Particle size classes (µm)
c)
2.6 2.8 3.0 3.2 3.4
0
-10
-20
-30
Depth (m)
Exponent J of Junge's law


a)
50 μm
b)
50 μm


Smectite
Albite
Titanium oxide
Quartz
Iron oxide
(a)
Depth (m)
0
-5
-10
-15
-20
-25
-30
-35
-40
0
5
10
15
20
25
30
35
40
Mineralogical content (%)
Illite
Chlorite
Al/Si ratio
Kaolinite
Orthoclase
(b)
Depth (m)
0
-5
-10
-15
-20
-25
-30
-35
-40
0
10
20
30
40
50
Mineralogical content (%)