



# Politecnico di Bari

Repository Istituzionale dei Prodotti della Ricerca del Politecnico di Bari

Are conventional statistical techniques exhaustive for defining metal background concentrations in harbour sediments? A case study: The Coastal Area of Bari (Southeast

This is a pre-print of the following article

*Original Citation:*

Are conventional statistical techniques exhaustive for defining metal background concentrations in harbour sediments? A case study: The Coastal Area of Bari (Southeast Italy) / Mali, M.; Dell'Anna, Maria Michela; Mastrorilli, Pietro; Damiani, Leonardo; Ungaro, N.; Belviso, C.; Fiore, S.. - In: CHEMOSPHERE. - ISSN 0045-6535. - STAMPA. - 138:(2015), pp. 138.708-138.717. [10.1016/j.chemosphere.2015.07.046]

*Availability:*

This version is available at <http://hdl.handle.net/11589/3284> since: 2021-04-06

*Published version*

DOI:10.1016/j.chemosphere.2015.07.046

Publisher:

*Terms of use:*

(Article begins on next page)

1 **Importance of using multivariate statistical techniques in assessing the heavy**  
2 **metal background concentrations in harbour sediments. A case study: the**  
3 **Coastal Area of Bari (South-East Italy)**

4 *Matilda Mali<sup>a\*</sup>, Maria Michela Dell'Anna,<sup>a</sup> Piero Mastrorilli<sup>a</sup>, Leonardo Damiani<sup>a</sup>, Nicola*  
5 *Ungaro<sup>b</sup>, Claudia Belviso<sup>c</sup>, Saverio Fiore<sup>c</sup>*

6  
7 <sup>a</sup>DICATECh, Politecnico di Bari, via Orabona, 4 I-70125 Bari, Italy

8 <sup>b</sup>ARPA PUGLIA, Corso Trieste, 27 I-70126 Bari, Italy

9 <sup>c</sup>IMAA-CNR, C.da S. Loja - Zona Industriale, I-85050, Tito Scalo (PZ), Italy

10 \*Corresponding author e-mail addresses: [matilda.mali@poliba.it](mailto:matilda.mali@poliba.it) (M. Mali);  
11 [mariamichela.dellanna@poliba.it](mailto:mariamichela.dellanna@poliba.it) (M. M. Dell'Anna); tel:+39 080 5963666; fax: +39 080 5963414

12  
13 **Abstract**

14 Sediment contamination by heavy metals poses risks to coastal ecosystems and is considered to  
15 be problematic to dredging operations. Determination of background values for heavy metal  
16 distribution based on site-specific variability is fundamental for assessing the pollution level in  
17 harbor sediments. With the aim to gain insight into the most available techniques useful for  
18 assessing the background heavy metal concentrations in port sediments, specific Regional  
19 Geochemical Background (RGB) values were calculated through three different conventional  
20 statistical methods, taking the southern area of the Italian Adriatic Coast, called “Coastal Area of  
21 Bari” (CAB) area as an example. The used data set consisted of 158 sediment samples collected  
22 at different depths coming from almost uncontaminated zones of 3 major ports, 4 minor ports  
23 and 6 marine transect of the near shore area, thus covering the whole 110 Km CAB coastline. All  
24 of them led to approximately the same RGB range of values for each of the nine considered  
25 heavy metals, seemingly effectual for the whole study area.

26 However, mineralogical analyses carried out on selected uncontaminated samples and  
27 multivariate statistical analyses performed on the whole data set revealed the heterogeneous  
28 variability of the study area and suggested to divide the CAB into three different zones, before  
29 calculating the heavy metal RGB values for each of them.

30 The obtained results pointed out that deeper investigations are necessary when the homogeneity  
31 of the study area is uncertain, since they reveal important peculiar aspects, which can be masked  
32 by using conventional statistical techniques.

33

34 **Keywords:** heavy metal background concentration, multivariate statistical techniques, site-  
35 specific sediment features, mineralogical analyses.

36

## 37 1. Introduction

38 The port sediments are the largest depository for potential source of metallic contamination in  
39 the marine environment, especially considering that ports are semi-enclosed systems where the  
40 dispersion of contaminants to the open sea is often difficult, since they are designed to minimize  
41 hydrodynamic energy inside (Dassenakis et al., 2003; M.C. Casado-Martinez et al., 2007).  
42 Nevertheless, the distribution of heavy metals within harbor sediments are controlled both by  
43 anthropogenic activities within port areas and by natural factors (*e.g.* mineralogical and chemical  
44 compositions of bottom sediments, textural and geochemical changes that may affect sediments  
45 during their transport and deposition, diagenetic processes). For this reason, a correct assessment  
46 of the quality of port sediments needs to consider both natural and anthropogenic contributions  
47 and requires a correct evaluation of each one. Methods to establish a "threshold value" or a  
48 "geochemical background concentration range" are widely discussed in the scientific community  
49 (Garrett, 1991, Covelli et al., 1997; Matschullat et. al., 2000; Reimann et al. 2005; Reimann C,  
50 Filzmoser P. & Garrett R., 2005; Galuszka et al., 2011; Dung et al, 2013). However, the  
51 development of worldwide accepted guidelines useful for distinguishing the polluted sediments  
52 from the unpolluted ones is a goal not yet achieved.

53 For the whole Italian coast, generic national baseline levels introduced by a national law (D.M.  
54 364/2002) are currently used for assessing the heavy metal pollution of the Italy port sediments.

55 However, these heavy metal threshold values cannot reflect the complex situation of the harbor  
56 sediment matrix of each single Italian area, because they do not take into account neither the  
57 grain-size effect nor the great heterogeneity of Italian coasts, especially from a geochemical  
58 point of view (Loring, 1991). In fact, it is often difficult to apply the national Sediment Quality  
59 Guidelines (SQGs) for metals, because some of them may have high geochemical background  
60 concentrations, such as chromium and nickel in the Ligurian Sea (Mugnai et al., 2010).

61 Empirical (geochemical), theoretical (statistical) and integrated methods (combining both  
62 empirical and theoretical methods) are the main approaches described in literature for  
63 determination of Regional Geochemical Background (RGB) concentrations (Reimann and  
64 Garrett, 2005; Dung et al., 2013).

65 In geochemical methods, also called direct methods, a deep core sample (30 m long) or samples  
66 collected at a certain distance from anthropogenic pollution sources are used to establish the  
67 background levels of heavy metals in the sediments or soils of a target area (Matschullat et. al.,  
68 2000, Desaules, 2012). Although the direct methods provide the heavy metal background  
69 concentration as a single site- specific value, the disadvantages of these techniques are their high  
70 cost and the need of a deep knowledge of the study area.

71 The statistical methods determine and eliminate the outliers which are related to anthropogenic  
72 sources-within the data set consisting of element concentrations in soils or sediments (Gałuszka,  
73 2007) and are associated with several advantages such as a wide selection of different statistical  
74 tests, graphical methods and easily available computer programs for data processing (Ho et al.,  
75 2012; Varol et al., 2011).

76 The combination of statistical and geochemical methods is referred to as “integrated method”  
77 (Gałuszka and Migaszewski, 2011). In this method, the analytical results coming from several  
78 deep core samples or from samples collected in pristine and pre-polluted areas are subjected to  
79 statistical calculations (Qi et al., 2010; Bini et al., 2011). Although this method has the  
80 advantages of both empirical and theoretical methods, it is quite expensive and it cannot always  
81 be applied owing to lacking of deep information on the study area.

82 Nevertheless, when other data are missing, several authors (Zahra et al., 2014) still use as  
83 background values for heavy metals the world average concentrations reported for the shale by  
84 Turekian & Wedepohl (1961).

85 In this framework, with the aim to gain insight into the most available techniques useful for  
86 determining proper heavy metal RGB concentrations in coast sediments, a specific area (110 Km  
87 length) of the South-East Italian coast, called “Coastal Area of Bari” (CAB) was chosen as case  
88 study for the application of three different conventional statistical methods: *i*) the linear  
89 regression analysis (Covelli & Fontolan, 1997) after geochemical normalization (Redon et al.,  
90 2013); *ii*) the data processing with the construction of cumulative distribution of frequency  
91 (Reimann et al., 2005); and *iii*) the  $2\sigma$  technique analysis (Matschullat et. al., 2000). The used  
92 data set consisted of 158 sediment samples collected at different depths (up to 3.50 m long)  
93 coming from almost uncontaminated zones of 3 major ports, 4 minor ports and 6 marine transect

94 of the near shore area (200 m and 500 m from the coast), thus covering the whole 110 Km CAB  
95 coastline.

96 Nevertheless, mineralogical analyses carried out on selected uncontaminated samples and  
97 multivariate statistical analyses performed on the whole data set, revealed that the CAB was not  
98 homogeneously constituted, suggesting to divide the coastline into three different areas and to  
99 calculate for each of them the proper heavy metal RGB ranges of values. Two of the three zones  
100 belonged to the major ports of the northern and the southern part of the coastline, respectively,  
101 and the third one was constituted by the minor ports of the southern part of CAB.

102 In the present study it is pointed out the importance of carrying out deeper investigations, such as  
103 mineralogical and multivariate statistical analyses, when assessing the heavy metal background  
104 concentrations in a specific area whose homogeneity is uncertain, since these studies can reveal  
105 important peculiar aspects, which do not emerge by using other conventional statistical  
106 techniques.

107

## 108 **2. Materials and methods**

### 109 *2.1 Morphological and geochemical features of the study area*

110 The area object of this study is a part of the regional basin, called “Coastal Area of Bari” (CAB)  
111 as defined by the Regional Coast Plan of Apulia Region. The costal line is 110 km in length,  
112 from the south of the Ofanto river estuary up till the Monopoli port, having in the middle the  
113 Bari port. The coast is composed mainly of micritic and calcarenitic limestones and by sands as  
114 in the case of most of the Adriatic coasts. The coastal tract of the metropolitan city of Bari is  
115 currently heavily modified by massive work of artificial burying. This tract presents  
116 morphological features of low-profile characterized mainly by sandy and silty sediments and  
117 development with residual beaches only partially preserved from human intervention.

118 The only major water course of the area is the Ofanto river, 170 km long; it is the longest river  
119 among those flowing into the South Adriatic sea. All of its branches play a fundamental role in  
120 providing a constant fluid and solid flow during the whole year. In addition numerous ephemeral  
121 water courses, with torrential regime, rhythmically affect the coast, especially the coast of Bari  
122 (central part of the whole study area).

123 Among the cities bordering the CAB, Barletta and Bari are the most developed, even if no  
124 current heavy industrial activities interested the area. Besides pollution caused by water washout  
125 of the intensively cultivated areas alongside costs that end in the sea through local hydrographic

126 network, other two important industrial events interested the area: the bombing campaign during  
127 the II WAR and the fuels refinery built at north of Bari in the late thirties of the last century and  
128 active until 1974. The area is only recently subjected to reclamation work in order to eliminate  
129 hydrocarbons residuals that affected the quality of the ground water.

130

## 131 2.2 Sampling

132 The sediment samples used in the present study were collected in years 2010-2011 from selected  
133 almost uncontaminated zones of 3 major ports (Barletta (BT), Bari (BA) and Monopoli (MP) and  
134 4 minor ports, considered as natural bays (Santo Spirito (SS), Palese (PL), Torre a Mare (TM)  
135 and San Giorgio (SG)) (Figure. 1). The Bari port is the biggest of them. It is one of the most  
136 important industrial port and of the most ancient of the southern Adriatic coast. Only 8 stations  
137 situated on the approach channel of Bari port were analyzed, because considered as  
138 representatives of natural sediments of the port area. It is well documented that the sediment  
139 transport caused by natural processes (wave propagation, litoral and tidal currents) determine an  
140 increase of sedimentation in the access port channel, that are comparable with coastal sediments  
141 (Hughes et al., 2003). To the whole data set, data analyses of 6 coastal sample sediments were  
142 added. The sites of these samples, taken by marine transects, situated at 200 m and 500 m near  
143 shore coast, are respectively in front of the Ofanto river, in front of Bari coastal area, and in front  
144 of Monopoli port. The number of all obtained samples was 158.

145

146

147

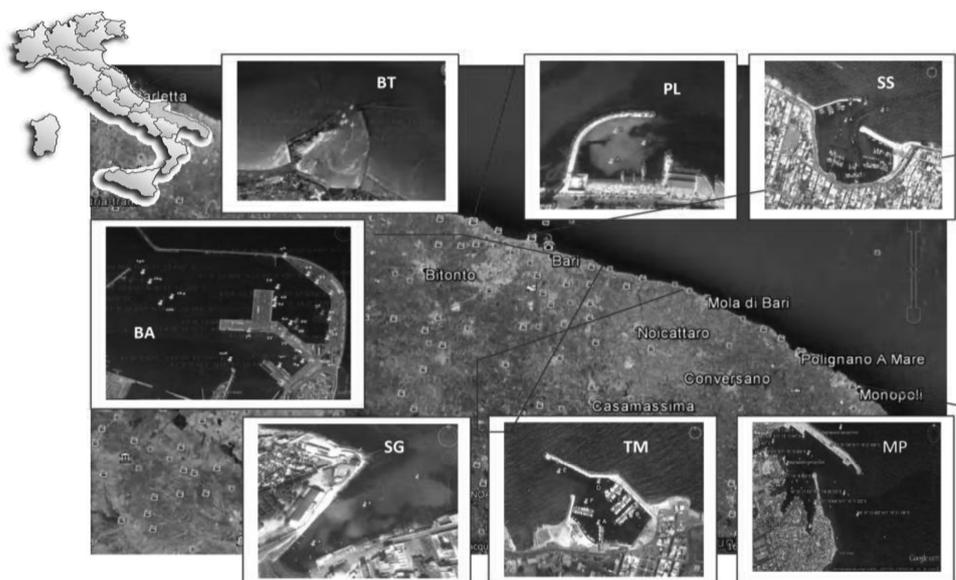
148

149

150

151

152



153 **Figure 1:** Map of sampling sites and location of sampling stations within ports (insets).

154 The sampling of sediments from the bottom of sea was carried out by using a vibro-corer PF1,  
155 equipped with liner, with the support vessel and provided of the differential system GPS for  
156 positioning of sampling cores. The sample cores were up to 3.50 m deep, reaching in any case  
157 the rocky hard bottom of the harbor sediment. Each core was divided into several 50 cm length  
158 sub-cores. Aliquots were taken by each sub-core and analyzed.  
159 Only in the case of the samples collected from the 6 marine transects, the sediment cores were  
160 short (45-60 cm long).

161

### 162 2.3 *Analytical methods*

163 Just after sampling, pH and redox potential (ORP) were measured using a multiparametric probe  
164 (Milwaukee, model SM 802), then the sediments were stored at 4°C during transportation to  
165 laboratory. Afterwards, they were frozen at -20°C and stored until analysis. Samples were  
166 quartered and portions of the samples were used for determining granulometric distribution,  
167 water content, total organic carbon (TOC), total nitrogen ( $N_{tot}$ ) total and phosphorous ( $P_{tot}$ ),  
168 according ICRAM procedures.

169 The sediments were classified according to Shepard (1954) into four sections: gravel: > 2 mm;  
170 sand: 2 mm - 0.063 mm; silt: 0.063 mm - 0.002 mm; clay < 0.002 mm. The fine fraction (<0.063  
171 mm) here reported is the sum of silt and clay fraction.

172 The heavy metal concentrations were determined by inductively coupled plasma mass  
173 spectrometry (thermo ICP MS X Series Thermo Fisher Scientific USA) after total acid digestion  
174 (HCl-HNO<sub>3</sub>-HF). Marine Sediment References Material and Standard Reference Materials 2702  
175 were used to control the analysis quality. The Hg content was determined by cold vapor atomic  
176 fluorescence spectroscopy following HNO<sub>3</sub>-H<sub>2</sub>SO<sub>4</sub> procedure.

177 Mineralogical analyses were performed on selected samples assumed as representative of the  
178 geologic background of BA and BT ports. The mineralogy of the whole rock and fine fraction (<  
179 2 μm) of the samples was determined (Srodon et al., 2001; Lezzerini et al., 1995; Moore and  
180 Reynolds, 1997) by X-ray diffraction (XRD) using a Rigaku Miniflex powder diffractometer  
181 with Cu-K $\alpha$  radiation (V: 30kV; I: 15 mA).

182

### 183 2.4 *Statistical analyses*

184 Conventional statistic approaches were normally used to analyze and assess interrelationship  
185 among elements in the sediments.

186 The cumulative distribution of frequency was calculated according the following formula:

$$CF_i = \sum_{j=1}^i AF_j$$

187  
188 Where  $AF_i$  is the value of the absolute frequency (*i.e.* in the number of times that the value was  
189 observed),  $j$  is the number of frequency classes and  $CF_i$  is the number of observations that are  
190 less (or equal) to the value  $x(i)$ . The cumulative rates for each value of  $i$  were obtained by  
191 normalization of  $CF_i$ .

192 Multivariate statistical analysis including Pearson correlation analysis (CA) and principal  
193 component analysis (PCA) have been carried out to check for geochemical associations and to  
194 ascertain the factors that control the geochemical origin and behavior in the coast sediments.  
195 Data treatment was performed by means of two different chemometric tool: the CA was  
196 conducted using Statistica Software (v.7), while the PCA analyses were carried out with the  
197 Unscrambler Program (v. 9.2 Camo, Oslo, Norway).

198

### 199 **3. Results and discussion**

#### 200 *3.1 Sediment characteristics*

201 The general characteristics of the sediments varied among the sampling points. The area is  
202 constituted by the sandy fraction in the range of 8÷74%<sub>w</sub>, the coarse fraction in the range of  
203 3÷9%<sub>w</sub>, and the fine-size fraction (silt and clay) in the range of 22÷90%<sub>w</sub> (Table S1 of the  
204 Supporting Information, S.I). The fine-size fraction (<0.063 mm) represents the main component  
205 with an average of 63%<sub>w</sub> for the northern part of the coastline (from BT to BA), while the sandy  
206 fraction (0.063 mm ÷ 2 mm) represents the main fraction with an average of 60 %<sub>w</sub> for the  
207 southern part of the coastline (from BA to MN).

208 The granulometric distribution diagrams elaborated for three representative major ports, BA, BT  
209 and MP. A very homogeneous distribution of clays (<0.002 mm) is revealed in the whole area.  
210 The Shepard diagram are elaborated for all of them (Figure S1 of S.I).

211 BA and BT ports are characterized by a predominance of the mud fraction. Both of them are  
212 industrial and commercial ports, therefore designed to minimize hydrodynamic energy within the  
213 harbor basin, in order to allow the shipping, loading and unloading activities. In fact, the port

214 structure facilitates the accumulation of mud sediments within port less than other open bay  
215 (Hancock G.J., 2001). It is noteworthy that the sandy fraction prevails in MP port and in all the  
216 four minor ports (PL – SG – SS – TM). It is assumed that the tidal currents propagating daily  
217 through these ports (which can be considered as almost natural bays characterized by a jagged  
218 rocky coastline) determine the accumulation of sand within them, explaining the prevalence of  
219 this fraction in their sediments.

220

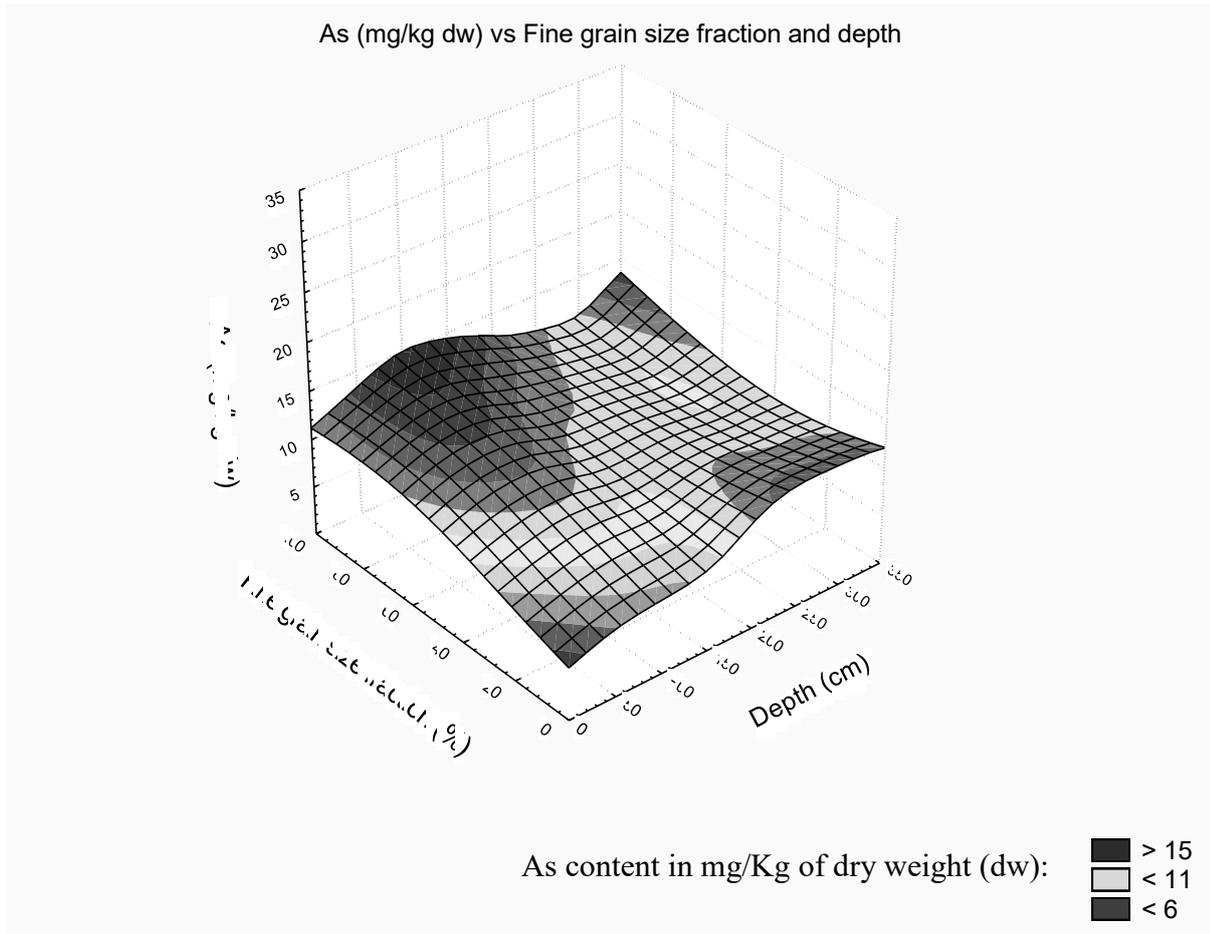
### 221 3.2 *Heavy metal distribution*

222 The metal distribution pattern (Table S2 of the S.I.) was in the following order of concentrations:  
223 Al>Fe>Zn>V>Cr>Ni>Pb>Cu>As>Cd>Hg, which is similar to the crustal average shale  
224 abundances (Wedepohl, 1995) with the exception of Ni being higher than Pb and Cu, and the  
225 concentration of Zn being higher than V. The lowest HMe concentrations of the whole area were  
226 observed in stations of the minor ports. Furthermore, the Al concentration determined in the area  
227 under investigation is in accordance with the background level expected for marine sediments  
228 indicated by Wedepohl (1995). Lower values of Al were found in the minor ports characterized  
229 by low amount of clay sediments. A positive gradient in Al amount is observed ongoing from the  
230 south ports toward the northern part of the considered coastline. No enrichment of this element  
231 in the sediments was found, as confirmed by the normal distribution of Al in the whole coastal  
232 area and confirmed by the positive test of Kolmogorov-Smirnov (Figure S2 of S.I.) (Reimann C.,  
233 De Caritat P., 2004).

234 The metal concentration showed a normal distribution (Figure S3 of the SI) in the case of Ni,  
235 Cr, V, Zn and Fe, while a non-normal distribution was revealed for As, Hg, Cu, Pb and Cd,  
236 confirmed by several outliers that indicate a possible anthropogenic input of these elements,  
237 although the studied sediment samples were collected from potentially uncontaminated zones.  
238 Apart for As, Pb and Hg, all the other heavy metals showed the highest concentrations in the  
239 fine-grained sediments, and the lowest concentrations in the sandy sediments, being metals  
240 sequestered by finest minerals or forming their own mineral phases.

241 Investigations of vertical variation of heavy metals in sediment cores permit to reconstruct the  
242 history of contamination within an aquatic environment (Swennen and Van der Sluys, 1998).  
243 Each sediment core was divided into 50 cm depth intervals. The maximum depth of the sample  
244 cores was 2.00 m and 3.50 m, for minor and major ports, respectively, reaching in any case the  
245 rocky hard bottom of the harbor sediment. Vertical profiles (amount of metal vs depth of the

246 sample core vs fine grain size fraction) for Al, Fe, As, Cd, Cr, Cu, Hg, Ni, Pb and Zn for the  
 247 whole area are illustrated in Figure S4 of the S.I.). The vertical profiles of Al, Fe, Cr, V, Ni and  
 248 Zn display rather similar trends, while the content of Pb and Hg in some samples is unusually  
 249 high in almost surficial cores, revealing a probable common source of pollution. Interestingly the  
 250 As content is unexpectedly high in deep core samples with a low amount of fine grain size  
 251 fraction (Figure 2).



252

253 **Figure 2:** Amount of As vs depth of the core sample vs fine grain size fraction.

254

### 255 3.3 Determination of heavy metal Regional Geochemical Background (RGB) concentrations

256 Geochemical background concentrations must be taken into account when establishing quality  
257 criteria for sediments. Once the geochemical background is determined, the pollution status can  
258 be more accurately assessed. In the present work, three different statistical methods were used: *i)*  
259 the linear regression analysis after geochemical normalization to Al, *ii)* the data processing with  
260 the construction of cumulative distribution of frequency, and *iii)* the  $2\sigma$  technique analysis.

261 The first statistical method employed was the regression analysis. In this method, a linear  
262 regression is performed between the concentrations of an element and one or several  
263 conservative factors (fine fraction, Al, Fe, Li...), which are considered as “inert” or not  
264 influenced by anthropogenic activities. Data that falls beyond the confidence interval (95 %) are  
265 considered as anthropogenic loads, while data that are well fitted with the linear regression  
266 conditions represent the background values (Covelli and Fontolan 1997; Matschullat et al. 2000;  
267 Aloupi and Angelidis, 2001; Galuska 2007).

268 We used this statistical method employing aluminum as normalizing element because it is the  
269 main constituent of the studied sediments, characterized by phyllosilicates and quartz associated  
270 to abundant biogenic carbonates (*vide infra*). Furthermore, Al has a low water solubility and  
271 therefore it can be assumed that its content in sediments has better comparability with that in the  
272 rock (Prohic et al., 1995; Joksas et al., 2008; Ho et al., 2010, 2012). In addition, Al is an  
273 immobile element during weathering and its concentration decreases linearly with increasing  
274 grain size in the sediment. According to the procedure proposed by Covelli and Fontolan (1997),  
275 we first checked for a high correlation in our data set between Al and fine grain-fraction (<0.063  
276 mm) amounts. As expected, a strong linear correlation at 95% level of confidence, characterized  
277 by high values of Pearson’s correlation coefficient ( $R=0.93$  at  $p < 0.00001$ ) was found (Table S3  
278 of the S.I.).

279 Statistically significant linear correlations were also observed between Al concentration and the  
280 content of the majority of heavy metals analyzed (Table S3 of the S.I.). The linear regression  
281 functions between Al amount and each heavy metal concentration (known as Geochemical  
282 Background Functions, GBF) were used to define proper RGB values (Lin et al., 2008; 2013). In  
283 order to increase the correlation coefficient values, box-Plot and Whiskers graphs were  
284 performed, aiming at determining and removing the outliers considered as anthropogenic inputs.  
285 The R values for each regression functions increased significantly, up to 0.79 for As, 0.73 for  
286 Cd, 0.96 for Cr, 0.59 for Hg, 0.87 for Cu, and 0.81 for Pb (Table S4 of the SI).

287 For each metal the threshold level defined as [mean + 2 standard deviation ( $\sigma$ )] was defined and  
288 the values obtained were compared with maximum levels admitted by the Italian national  
289 guidelines and the average baseline crust levels reported by Turekian and Wedepohl (1961)  
290 (Table S5 of the S.I.). It is observed that As, Cd, Hg and Ni background values were lower than  
291 the maximum levels reported in the Italian guidelines, while the geochemical distribution  
292 estimated in the CAB area for Pb, Zn, and Cu resulted higher than the limits imposed by the  
293 national law. For all metals, the calculated RGB range of concentrations were higher than the  
294 corresponding average crust baseline values reported by Turekian and Wedepohl (1961).

295 In order to test the reliability of the linear regression method, we compared the obtained data  
296 with those deriving by application of another statistical technique, such as the one enclosing the  
297 construction of cumulative frequency curve for each metal. These curves, as function of  
298 elemental concentrations, were used by several authors (Lepeltier, 1969; Bauer and Bor, 1995) to  
299 identify the turning points that separate the “background values” from the contaminated samples.  
300 Figure S5 of the S.I. reports the cumulative frequency curves for As and Cu, respectively. For  
301 each analyzed HMe, the threshold limit deriving from the  $CF_i$  calculation was coincident with  
302 the highest background value obtained with the linear regression method.

303 The advantage of using this technique lies in the fact that the geochemical normalization is  
304 circumvented and the data can be processed without preliminary analyses. Anyway, the data set  
305 should be big enough in order to provide statistically significant results.

306 In addition, according to Reimann et al. (2005) geochemical background concentrations for each  
307 element correspond to the values within the range of [mean  $\pm$  2 standard deviation ( $\sigma$ )] or, better,  
308 [median  $\pm$  2 maximum absolute standard deviation (MAD)] in a big size data set. Consequently,  
309 the exact value of [mean + 2 $\sigma$ ] or [median + 2 MAD] should be referred to the upper limit of  
310 geochemical background variation and this was suggested as “threshold level” useful for  
311 environmental legislation. Also in this case, at least for As and Cu (Tables S6 and S7 of the S.I.),  
312 the “threshold level” obtained by applying the [median + 2 MAD] formula is consistent with the  
313 highest background value obtained with the linear regression method. Also by applying the 2 $\sigma$   
314 technique, the geochemical normalization could be avoided. Nevertheless, this method can be  
315 used only when the size of the data set is really big, but the requirement of a large number of  
316 data needed to perform significant statistical tests is not always met in real case situations.

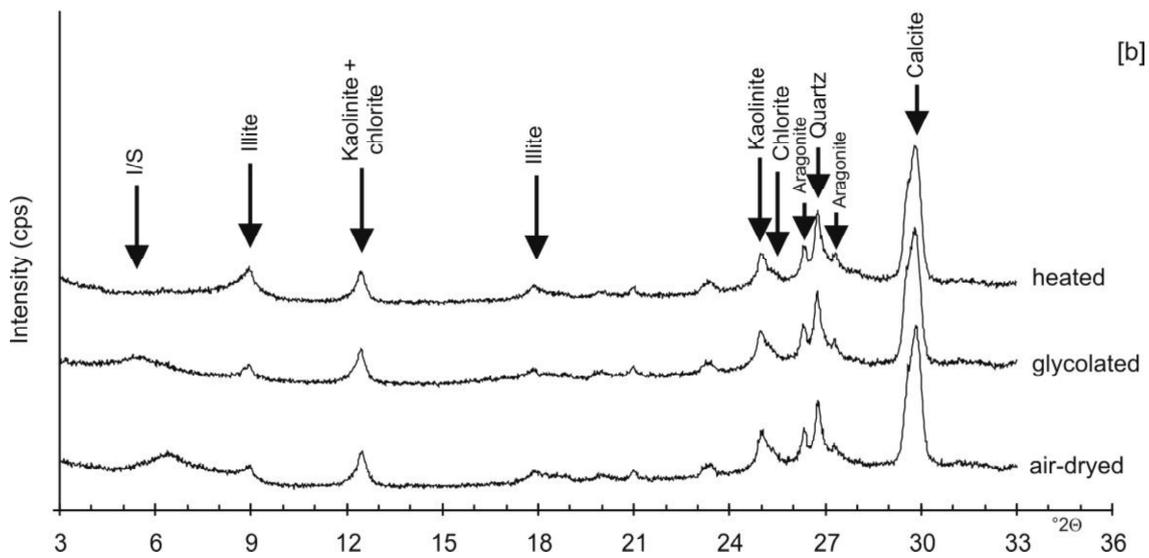
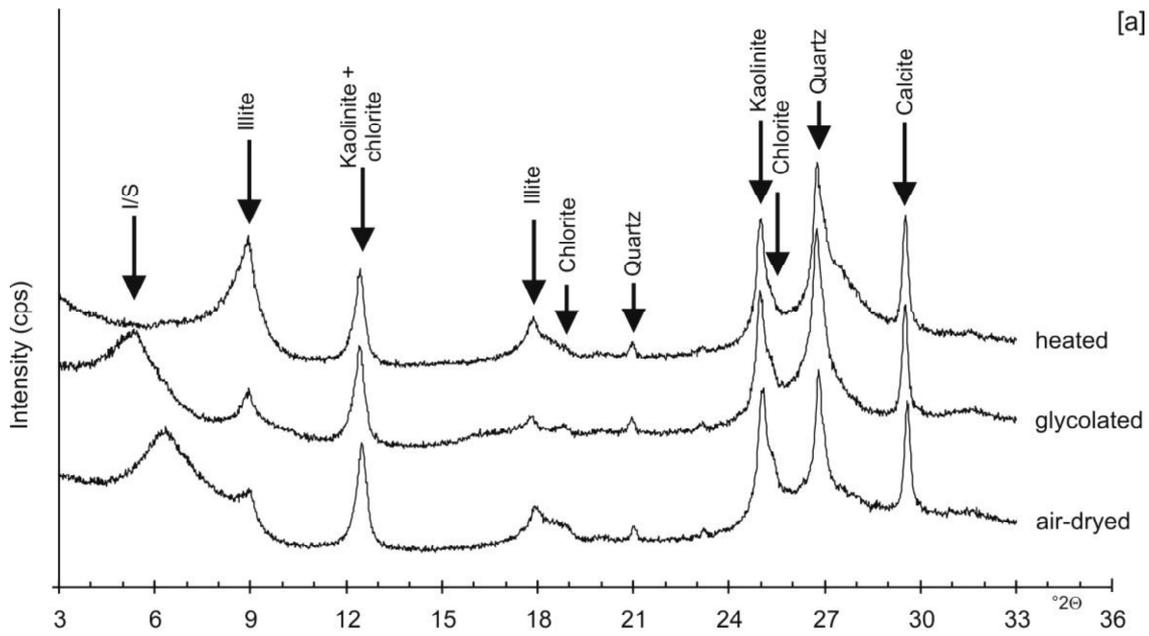
317 The obtained results allowed us to conclude that, in the presence of a large data set, the use of the  
318 cumulative frequency curve or the application of the 2 $\sigma$  technique for the establishing HMe RGB  
319 values are more convenient compared to the regression analysis, because the geochemical

320 normalization step is circumvented. On the contrary, when the available data set has a relatively  
321 smaller size, the geochemical normalized regression analysis is safer, provided that the studied  
322 samples contain fine grain size fraction in the whole percent range (from 3 to 98 % *ca.*) (Covelli  
323 and Fontolan, 1997).

324 Although statistical methods are associated with several advantages, such as the wide selection  
325 of different statistical tests, graphical methods and easily available computer programs for data  
326 processing (Gałuszka et al., 2011), these methods have been criticized because the particular  
327 characteristics of geochemical data. Generally, geochemical data from a particular site are  
328 strongly influenced by spatial, temporal and mineralogical variability (Rencz et al. 2006,  
329 Paikaray 2012).

330 In order to check for a possible mineralogical variability, mineralogical analyses were performed  
331 on selected samples assumed as representative of the lithological background of BA and BT  
332 ports. The mineralogical assemblages of bulk samples show the presence of phyllosilicates and  
333 quartz associated with abundant biogenic carbonate. The amount of phyllosilicates is higher in  
334 the samples collected in the area of the BT port than those collected in the BA port area, the  
335 latter showing the presence of higher content of carbonates (calcite, Mg-rich calcite and  
336 aragonite) (Figure S6 of the S.I.). However, differences in mineralogical compositions of the  
337 sediments collected in the two areas are revealed by the analysis of the clay fraction. The  
338 samples collected in the BT area show larger amount of illite/smectite whereas higher percentage  
339 of illite characterizes the samples coming from the BA area (Figure 3). This is in agreement with  
340 literature data reporting differences between Pliocene and Plio-Pleistocene Apennine clay facies  
341 (Dell'Anna and Laviano, 1981) from which the Adriatic Sea sediments derived (Tomadin, 2000).  
342 These mineralogical differences strongly suggest that heavy metal RGB concentration ranges  
343 could not be the same for the BA and BT areas.

344

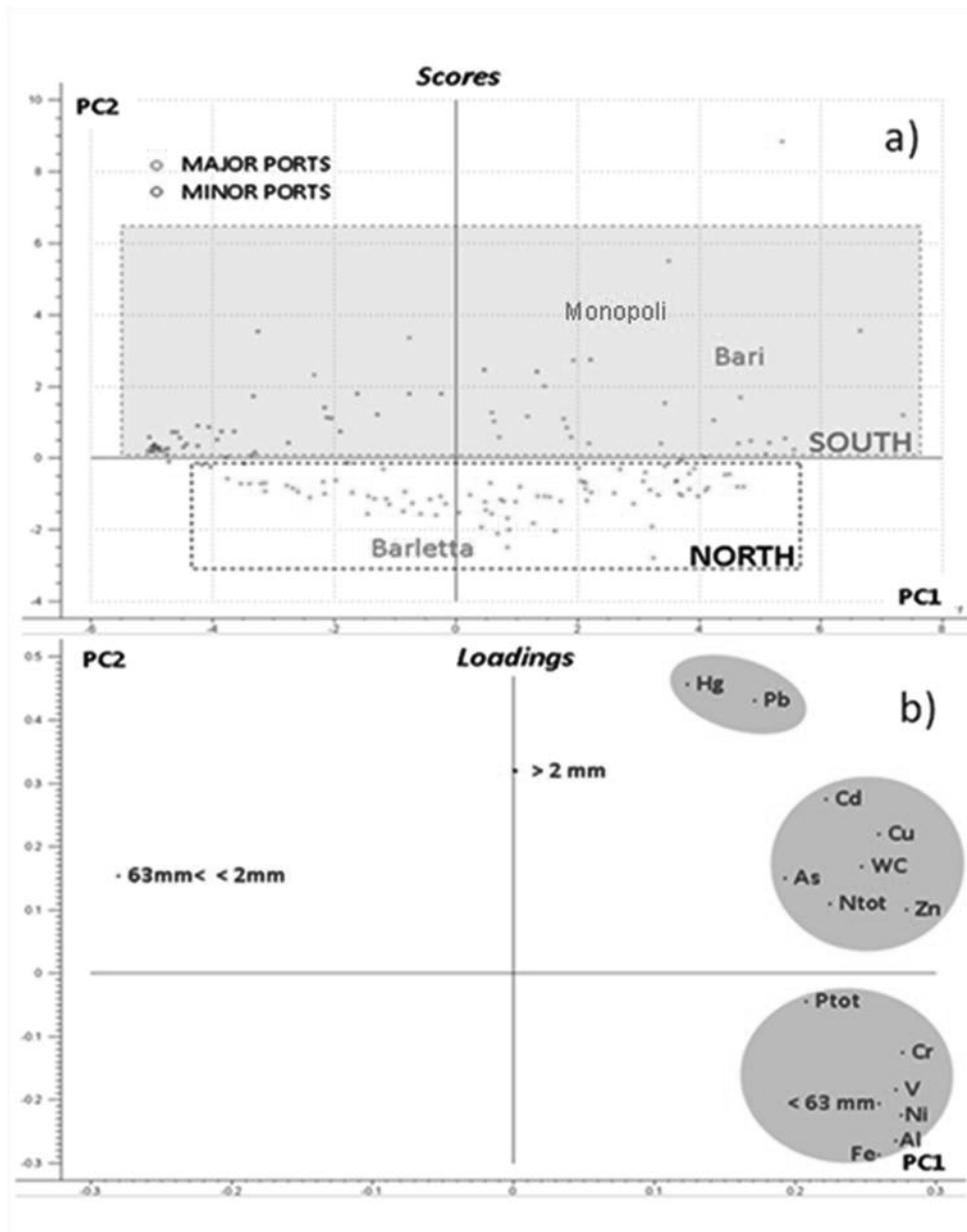


345

346

347 **Figure 3.** XRD patterns of  $<2 \mu\text{m}$  grain-size fraction of a sample representative of [a] BT and [b]

348 BA areas



349  
 350 **Figure 4.** Score (a) and loading (b) plots of PC2 versus PC1 obtained for the data set. In figure  
 351 12a, red and blue points indicate samples coming from major (BA, BT and MP) and minor (PL,  
 352 SS, SG and TM) ports, respectively. In figure 12a, the grey square groups the samples coming  
 353 from the southern NPU, while the uncolored square included the samples collected in the  
 354 northern NPU.

355  
 356 In order to gain insight this first observation, principal component analysis (PCA) as a non-  
 357 supervised technique was performed on dataset to compare the compositional pattern between  
 358 the sediment samples and to identify the factors influencing each of them. The entire data set  
 359 (Tables S1 and S2), excluding three variables (OPR, pH and TOC) which were not available for

360 minor ports, is used for this purpose by means of Unscramble Software. The data set consisted of  
361 a matrix with variables in columns (17) and the sediment samples in rows (158) and was mean  
362 and centre.

363 A model with two PCs (62% and 15% of explained variance, respectively) was selected (Figure  
364 4), which explains 77% of the total variance within the data set, since the results did not  
365 significantly change when the third principal component (PC3 = 6% of explained variance) was  
366 considered using Leverage correction as validation method.

367 Three clusters of variables grouped together were obtained in the loading plot (Figure 4b). In the  
368 first group there were Hg and Pb concentrations, with strong positive loadings on PC2 (Hg: 0.46;  
369 Pb: 0.453); both of them are considered as metals with high level of toxicity in environmental  
370 studies. The second group in the PCA consisted of water content (wc),  $N_{tot}$ , Cd, Cu, As and Zn  
371 concentrations, characterized by lower positive loadings on PC2. The third group includes  $P_{tot}$ , Cr,  
372 V, Ni, Al and Fe amounts, having positive loadings on PC1 and negative ones on PC2. The  
373 formation of these three groups suggested that the metals in the same cluster may have a  
374 common source, and that their content is affected also by  $N_{tot}$ ,  $P_{tot}$  and water content and not only  
375 by the fine-size fraction (<0.063 mm) amount.

376 In addition, the score plot in the space formed by the two PCs (Figure 4a) reveals the existence  
377 of three different groups of sediment samples with specific pollution profiles. The first one, with  
378 negative score projection values on the PC1 and positive ones on the PC2, includes all the  
379 sediments collected from minor ports, such as SS, PL, SG and TM. In the superposed  
380 representation of the scores and loadings this group was not affected by any metal concentration  
381 but only by the sand fraction amount. The second group (positive score projection values on both  
382 PC1 and PC2) was formed by sediment samples coming from major ports of the southern part of  
383 the study area (BA and MP), and was influenced by Hg, Pb, Cd, As, Cu and Zn concentrations,  
384 correlated with  $N_{tot}$  and water content. Finally the third group with negative score projection  
385 values on the PC2 includes all the sediments collected from the northern part of CAB (Barletta)  
386 and characterized by negative loading values of Cr, Ni, V on the PC2 and positive ones on the  
387 PC1.

388 On the base of the PCA results, by using the AI normalization linear correlation method reported  
389 above, we calculated for each heavy metal three RGB concentration ranges, belonging to the  
390 northern major port (BT port, Table S8 of the S.I.), the southern major ports (BA and MP ports,  
391 Table S9 of the S.I.) and all the minor ports (Table S10 of the S.I.), respectively. Differences  
392 among the considered zones were obtained. As, Ni, Cr, V and Zn background concentrations

393 were higher in the major port of the northern area (BT port) than in the major ports of the  
394 southern one, while background values of Pb, Cd, Hg and Cu were higher in the major ports of  
395 the southern part than in the major port of the northern zone. However, for both the considered  
396 zones, for some metals the so obtained background concentrations were not in line with the  
397 national guidelines. On the other hand, lower concentration of all metals were found in the minor  
398 ports and the background values determined for this sub-area resulted similar to the average  
399 crustal shale values established by Turekian and Wedepohl (1961) (Figure S7 of the S.I.).

400

#### 401 **4. Conclusions**

402 In conclusion, a specific area (110 Km length) of the South-East Italian coast has been chosen as  
403 case study for applying three different conventional statistical methods (the cumulative  
404 frequency curve technique, the  $2\sigma$  method and the regression analysis) in order to determine  
405 proper heavy metal RGB concentrations in coast sediments. All three methods led to  
406 approximately the same RGB concentration range for each studied heavy metal, suggesting that  
407 in the presence of a large data set, the use of the cumulative frequency curve technique or the  
408 application of the  $2\sigma$  method for the calculation of HMe RGB values is more convenient  
409 compared to the regression analysis, because the geochemical normalization step is  
410 circumvented. On the contrary, when the available data set has a smaller size, the geochemical  
411 normalized regression analysis is the best compromise, provided that the considered samples  
412 contain fine grain size fraction in the whole percent range (from 3 to 98 % *ca.*), because they  
413 should be representative of the whole system.

414 However, all these methods can be applied only when the study area is homogeneously  
415 constituted. In fact, mineralogical studies carried out on selected uncontaminated samples and  
416 multivariate statistical analysis performed on the whole data set revealed the heterogeneity of the  
417 albeit small study area, suggesting to divide it in three zones before calculating for each of them  
418 the proper HMe RGB ranges of value. These RGB values were then assessed by the Al  
419 normalized regression analysis, obtaining for the zones belonging to the major ports that the As,  
420 Cr, Cu, Ni, V and Zn background concentrations were higher in the northern area than in the  
421 southern one, while background values of Pb, Cd and Hg were higher in the southern part than in  
422 the northern zone. However, for both the considered zones, for some metals the so obtained  
423 background concentrations were not in line with the national guidelines.

424 Finally, the third RGB ranges of metal concentrations, obtained by using data set of samples  
425 collected from the minor ports, resulted similar to the average crustal shale values.

426 The obtained results point out that an ideal method for assessing regional heavy metal  
427 background concentrations does not exist, but it must be selected for each specific system. In  
428 addition, deeper investigations (such as mineralogical analysis) and multivariate statistical  
429 analyses carried out on the whole data set are necessary when the homogeneity of the study area  
430 is uncertain, since they reveal important peculiar aspects, which can be masked by using other  
431 conventional statistical techniques.

432

### 433 **Acknowledgments.**

434 This study was supported by ARPA Puglia, the Port Authority of Bari and fund of TEN  
435 ECOPORT project within SEE programme (SEE/A/0189/2.2/X).

436 The authors thank Prof. Alberto de Diego and Prof. Silvia Fernández-Ortiz de Vallejuelo of  
437 University of Basques Country, Department of Analytical Chemistry for useful discussion.

438

### 439 **References**

440 Aloupi, M., Angelidis, M.O., 2001. Geochemistry of natural and anthropogenic metals in the  
441 coastal sediments cores for the Aegean Sea. Greece Mar. Environ. Pollut. 113, 211-219.

442 Bini, C., Sartori, G., Wahsha, M., Fontana., 2011. Background levels of trace elements and soil  
443 geochemistry at regional level in NE Italy. J. Geochem. Explor. 109, 125–133.

444 Bouer, I., Bor, J., 1995. Lithogene, Geogene und anthropogene Schwermetallgehalte von  
445 Losboden and den Beispielen von Cu, Zn, Ni, Pb, Hg und Cd. Mainzer. Geovz. Mitt.24, 47-70.

446 Casado-Martinez, M.C., Fernandez, N., Forja, J.M., DelValls, T.A., 2007. Liquid versus solid  
447 phase bioassays for dredged material toxicity assessment. Environ. Int.33, 456-462..

448 Covelli, S., Fontonlan, G., 1997. Application of a normalization procedure in determining  
449 regional geochemical baselines. Environ. Geol 30, 34-45.

- 450 Dassenakis, M., Adrianos, H., Depiazi, G., Kostatas, A., Karabela, M., Sakellari, A., 2003. The  
451 use of various methods for the study of metal pollution in marine sediments, the case of  
452 Euvoikos Gulf, Greece. *App. Geochem.* 18, 781-794.
- 453 Dell'Anna, L., Laviano, R., 1981. ) Mineralogia delle Argille Plioceniche di Potenza. *Rendiconti*  
454 *Soc. Ital. Min. Petrol.*, 37, 179-194. (*in Italian*).
- 455 Desaules, A., 2012. Critical evaluation of soil contamination assessment methods for trace  
456 metals. *Sci. Total Environ.* ,426,120–131.
- 457 Dung, T.T.T., 2013. Cappuyns V., Swennen R., Phung N. K., *Rev. Environ. Sci. Biotechnology*  
458 12, 335-353.
- 459 Galuszka, A., 2007. A review of geochemical background concepts and an example using data  
460 from Poland. *Environ Geol.* 52, 861–870
- 461 Galuszka, A., Migaszewski, Z.M., 2011. Geochemical Background - an environmental  
462 perspective, *Miner.* 42, 7-17.
- 463 Garrett, R.G., 1991. The management, analyses and display of exploration geochemical data.  
464 *Exploration geochemistry workshop, Ottawa Geological Survey of Canada, Open File 2390.*
- 465 Hancock, G.J., Olley, J.M. and Wallbrink, P.J. Sediment transport and accumulation in  
466 Westernport, *CSIRO Land and Water Technical Report*, 2001 47/01.
- 467 Ho, H.H., 2012. Swennen R, Cappuys V, Vassilieva E, Tran TV. Necessity of normalization to  
468 aluminum to assess the contamination by heavy metal and arsenic in sediments near Haiphong  
469 Harbor, Vietnam. *J. Asian Earth. Sci.* 56, 229-239.
- 470 Hughes A.O., Prosser I.P., Wallbrink P.J. Suspended Sediment and Bedload Budgets for the  
471 Western Port Bay Basin *CSIRO Land and Water, Canberra Technical Report 4/03*
- 472 Jokšas, K., Stakeniene, R., Galkus, A., Lagunaviciene, L., 2008. Metals in bottom sediments of  
473 Sventoji Port Area - Lithuania. *Geol.* 50, 143-155.
- 474 Lepeltier, C., 1969. A simplified treatment of geochemical data by graphical representation.  
475 *Econ. Geol.* 64, 538-550 .

476 Lezzerini, M., Sartori, F., Tamponi M., 1995. Effect of amount of material used on  
477 sedimentation slides in the control of illite "crystallinity" measurements. *Eur. J. Mineral.* 7, 819-  
478 823.

479 Loring, D.H., 1991. Normalization of heavy metal data from marine and coastal sediments. *ICES*  
480 *J. Mar. Sci.* 48, 101-115.

481 Matschullat, J., Ottenstein R, Reimann C., 2000. Geochemical background - can we calculate it?  
482 *Environ. Geol.* 39, 990-1000.

483 Moore, D.M., Reynolds, R.C.JR., 1997. X-ray Diffraction and the identification and analysis of  
484 clay minerals. Second edition. Oxford University Press, Oxford and New York 378 pp.

485 Mugnai, C., Bertolotto, R.M., Gaino, F., Tiberiade, C., Bellucci, L.G., Giuliani, S., Romano, S.,  
486 Frignani, M., Albertazzi, S., Galazzo, D., 2010. History and trends of sediment contamination by  
487 heavy metals within and close to a marine area of national interest: the Ligurian Sea off  
488 Cogoleto-Stoppani (Genoa, Italy). *Water Air Soil Poll.* 211, 69–77.

489 Paikaray, S., 2012. Environmental hazards of arsenic associated with black shales: a review on  
490 geochemistry, enrichment and leaching mechanism. *Rev Environ. Sci. Biotechnol.* 11, 289–303.  
491 Regione Puglia. Piano Regionale delle Coste. BURP n. 108 del 06-08-2013.

492 Prohic, E., Miko, S., Peh, Z., 1995. Normalisation and trace element contamination of soils in a  
493 Karstic Poje. An expalme from the Sinjsko Poje, Croatia. *Geol Croat.* 48, 67-86.

494 Qi, S., Leipe, T., Rueckert, P., Di, Z., Harff, J., 2010. Geochemical sources, deposition and  
495 enrichment of heavy metals in short sediment cores from the Pearl River Estuary, Southern  
496 China. *J. Mar. Syst.* 82, S28–S42.

497 Reimann, C., De Caritat, P., 2004. Distinguishing between natural and anthropogenic sources for  
498 elements in the environment: regional geochemical surveys versus enrichment factors. *Sci. Total*  
499 *Environ.* 337, 91 – 107.

500 Reimann, C., Garrett, R.G., 2005. Geochemical background. Concept and reality. *Science of the*  
501 *Total Environment* 350, 12–27.

502 Reimann, C., Filzmoeser, P., Garret, R.G., 2005. Background and threshold: critical comparison  
503 of methods of determination. *Sci. Total Environ.* 346, 1-16.

504 Rencz, A.N., Garrett, R.G., Adcock, S.W., Spirito, W.A., 2006. Bonham-Carter GF.  
505 Geochemical background in soil and till. Geolog Surv Canada Open File 6645, 1 CD-ROM.

506 Shepard F.P., (1954) Nomenclature based on sand–silt–clay ratios. *J. Sed. Petr.*, 24, 151-158

507 Srodon, J., Drits V.A., Douglas, K., McCarty, K., Hsieh, C.C., Eberl, D., 2001. Quantitative X-  
508 Ray diffraction analysis of clay-bearing rocks from random preparations. *Clay. Clay Miner.* 49,  
509 514-528.

510 Swennen, R., Van Der Sluys, J., 1998. Zn, Pb, Cu, Cd and As distribution patterns in overbank  
511 and medium order stream sediment samples: their use in environmental geochemistry and in  
512 exploration of base metal ore bodies in heavily polluted areas. *J. Geochem. Explor.* 65, 27–45.

513 Tomadin, L., 2000. Sedimentary fluxes and different dispersion mechanisms of the clay  
514 sediments in the Adriatic Basin. *Rendiconti Lincei*, 11, 161-174.

515 Turekian, K.K., Wedepohl, K.H., 1961. Distribution of the elements in some major units of the  
516 earth's crust. *Bull. Geol. Soc. Amer.* 72, 175–192.

517 Varol, M., 2011. Assessment of heavy metal contamination in sediments of the Tigris River  
518 (Turkey) using pollution indices and multivariate statistical techniques. *J. Hazard. Mater.* 195,  
519 355– 364.

520 Wedehpol, K.H., 1995. The composition of the continental crust. *Geochim. Cosmochim. Ac* 59,  
521 1217-1232.

522 Zahra, A., Hashmi, M.Z., Malik, R.N., 2014. Ahmed Z. Enrichment and geo-accumulation of  
523 heavy metals and risk assessment of sediments of the Kurang Nallah—Feeding tributary of the  
524 Rawal Lake Reservoir, Pakistan. *Sci. Total Environ* 470-471, 925–933.

525

526

Dear Editor,

we have proceeded with the online submission of the paper titled “Importance of using multivariate statistical techniques in assessing the heavy metal background concentrations in harbour sediments. A case study: the Coastal Area of Bari (South-East Italy)” by Matilda Mali, Maria Michela Dell’Anna, Piero Mastrorilli, Leonardo Damiani, Nicola Ungaro, Claudia Belviso, Saverio Fiore.

In this paper, with the aim to gain insight into the most available techniques useful for assessing the background heavy metal concentrations in port sediments, specific Regional Geochemical Background (RGB) values were calculated through three different conventional statistical methods, taking the southern area of the Italian Adriatic Coast, as an example. All the three methods lead to approximately the same general RGB range of values for each of the nine considered heavy metals (As, Hg, Cd, Ni, Pb, Zn, V, Cu and Cr). However, mineralogical and multivariate statistical analyses revealed the heterogeneous variability of the albeit small study area and suggested to divide it into three different zones, before calculating the heavy metal RGB values for each of them. The obtained results pointed out that deeper investigations (such as mineralogical analysis) and multivariate statistical analyses are necessary when the homogeneity of the study area is uncertain, since they reveal important peculiar aspects, which can be masked by using other conventional statistical techniques.

The manuscript slightly exceeds the 6000 word length suggested by the journal guidelines. We were not able to reduce it anymore without rendering it unclear.

We hope you can find this manuscript suitable for publication in Chemosphere in the Environmental Chemistry section.

Best regards,

Matilda Mali

## Highlights

- Heavy metal background concentrations in South-East Italy sediments were calculated
- Three conventional statistical techniques were used and compared to each other
- The conventional techniques led to the same range of heavy metal background values
- Mineralogical and multivariate statistical analyses revealed new specific aspects
- Heterogeneously constituted areas need deeper investigations

## Importance of using multivariate statistical techniques in assessing the heavy metal background concentrations in harbour sediments. A case study: the Coastal Area of Bari (South-East Italy)

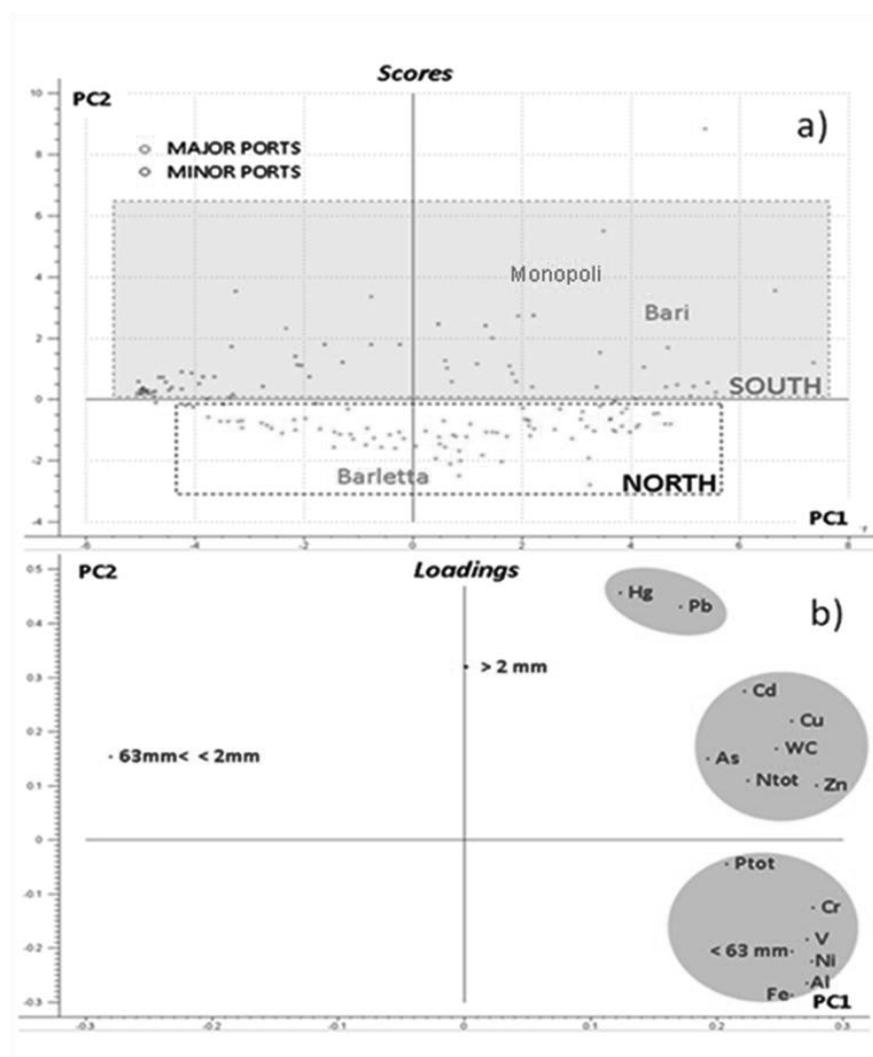
Matilda Mali<sup>a\*</sup>, Maria Michela Dell'Anna<sup>a</sup>, Piero Mastrorilli<sup>a</sup>, Leonardo Damiani<sup>a</sup>, Nicola Ungaro<sup>b</sup>, Claudia Belviso<sup>c</sup>, Saverio Fiore<sup>c</sup>

<sup>a</sup>DICATECh, Politecnico di Bari, via Orabona, 4 I-70125 Bari, Italy

<sup>b</sup>ARPA PUGLIA, Corso Trieste, 27 I-70126 Bari, Italy

<sup>c</sup>IMAA-CNR, C.da S. Loja - Zona Industriale, I-85050, Tito Scalo (PZ), Italy

\*Corresponding author e-mail addresses: [matilda.mali@poliba.it](mailto:matilda.mali@poliba.it) (M. Mali); [mariamichela.dellanna@poliba.it](mailto:mariamichela.dellanna@poliba.it) (M. M. Dell'Anna); tel:+39 080 5963666; fax: +39 080 5963414



## **Abstract**

Sediment contamination by heavy metals poses risks to coastal ecosystems and is considered to be problematic to dredging operations. Determination of background values for heavy metal distribution based on site-specific variability is fundamental for assessing the pollution level in harbor sediments. With the aim to gain insight into the most available techniques useful for assessing the background heavy metal concentrations in port sediments, specific Regional Geochemical Background (RGB) values were calculated through three different conventional statistical methods, taking the southern area of the Italian Adriatic Coast, called “Coastal Area of Bari” (CAB) area as an example. The used data set consisted of 158 sediment samples collected at different depths coming from almost uncontaminated zones of 3 major ports, 4 minor ports and 6 marine transect of the near shore area, thus covering the whole 110 Km CAB coastline. All of them led to approximately the same RGB range of values for each of the nine considered heavy metals, seemingly effectual for the whole study area.

However, mineralogical analyses carried out on selected uncontaminated samples and multivariate statistical analyses performed on the whole data set revealed the heterogeneous variability of the study area and suggested to divide the CAB into three different zones, before calculating the heavy metal RGB values for each of them.

The obtained results pointed out that deeper investigations are necessary when the homogeneity of the study area is uncertain, since they reveal important peculiar aspects, which can be masked by using conventional statistical techniques.

## Captions to figures

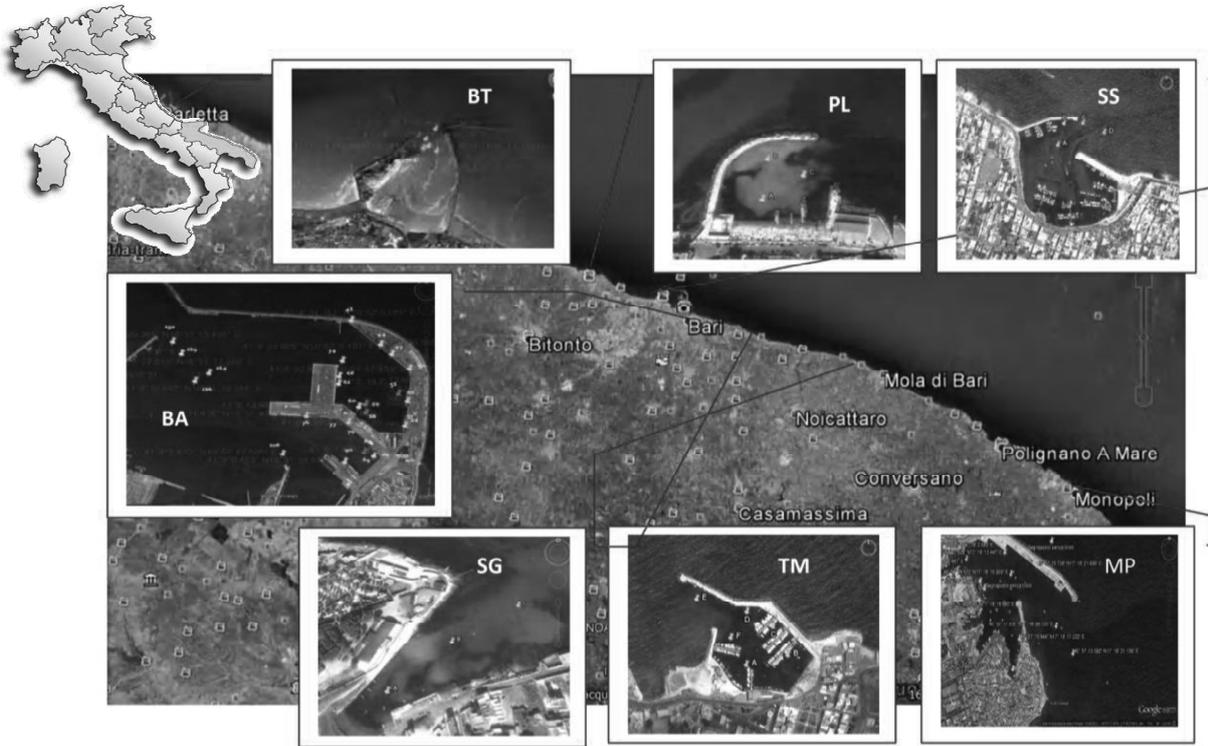
**Figure 1:** Map of sampling sites and location of sampling stations within ports (insets).

**Figure 2:** Amount of As vs depth of the core sample vs fine grain size fraction

**Figure 3:** XRD patterns of  $<2\ \mu\text{m}$  grain-size fraction of a sample representative of [a] Barletta and [b] Bari areas

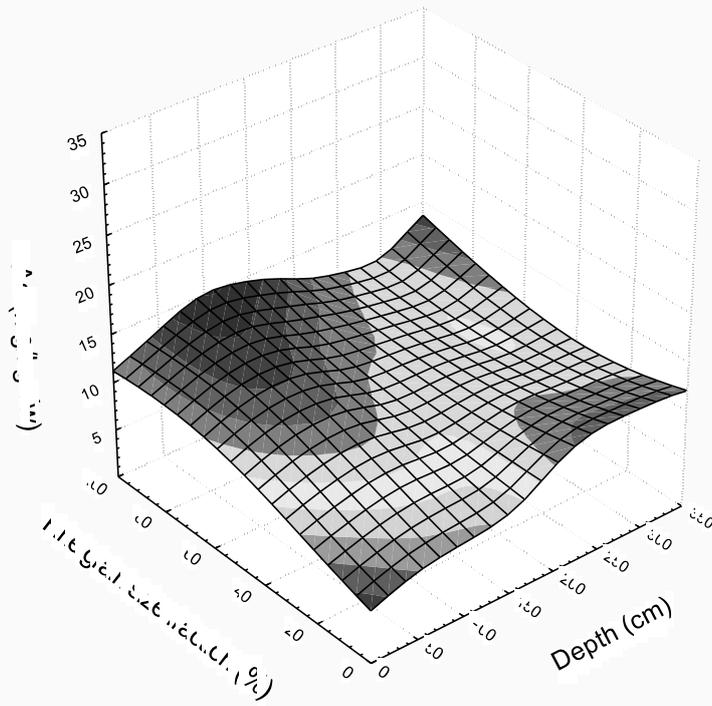
**Figure 4.** Score (a) and loading (b) plots of PC2 versus PC1 obtained for the data set. In figure 11a, red and blue points indicate samples coming from major (Bari, Barletta and Monopoli) and minor (PL, SS, SG and TM) ports, respectively. In figure 11a, the grey square groups the samples coming from the southern NPU, while the uncolored square included the samples collected in the northern NPU.

Figure 1



**Figure 2**

As (mg/kg dw) vs Fine grain size fraction and depth



**Figure 3**

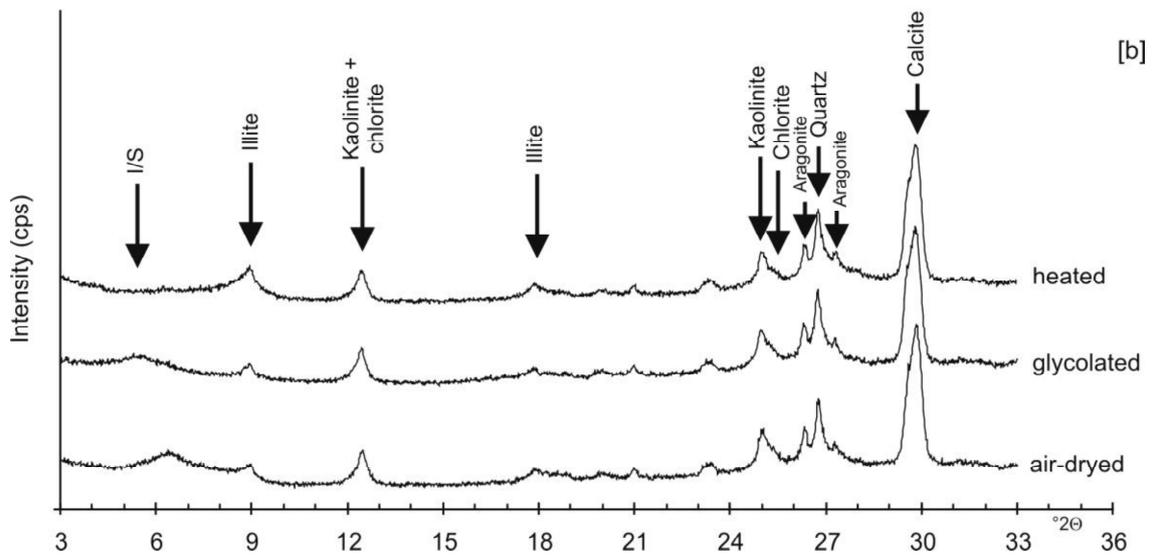
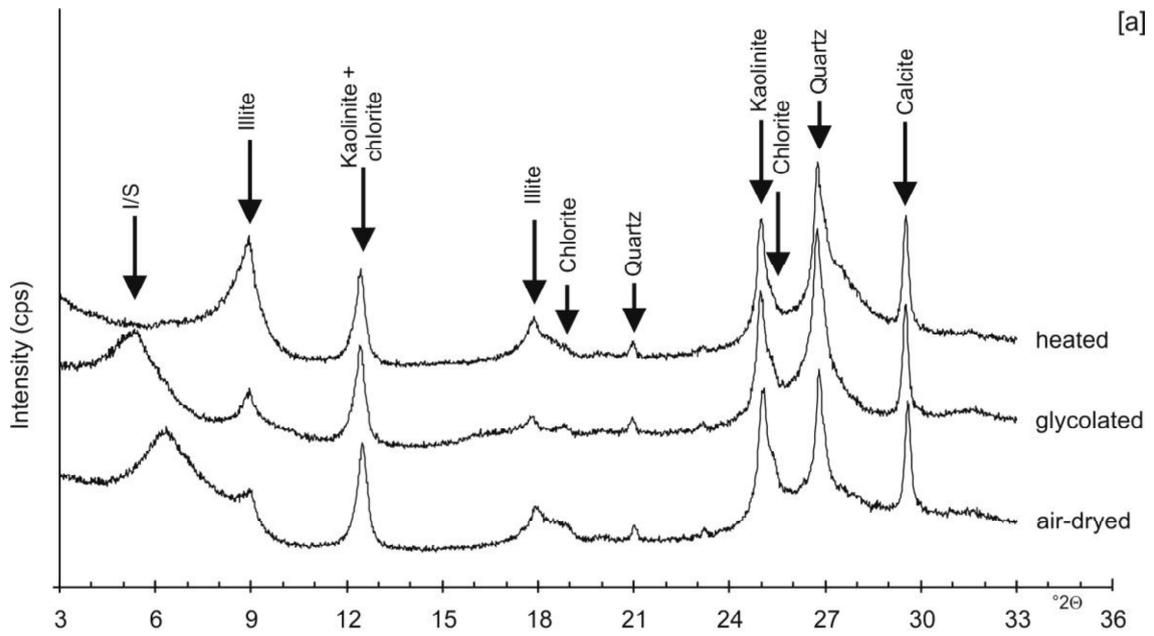
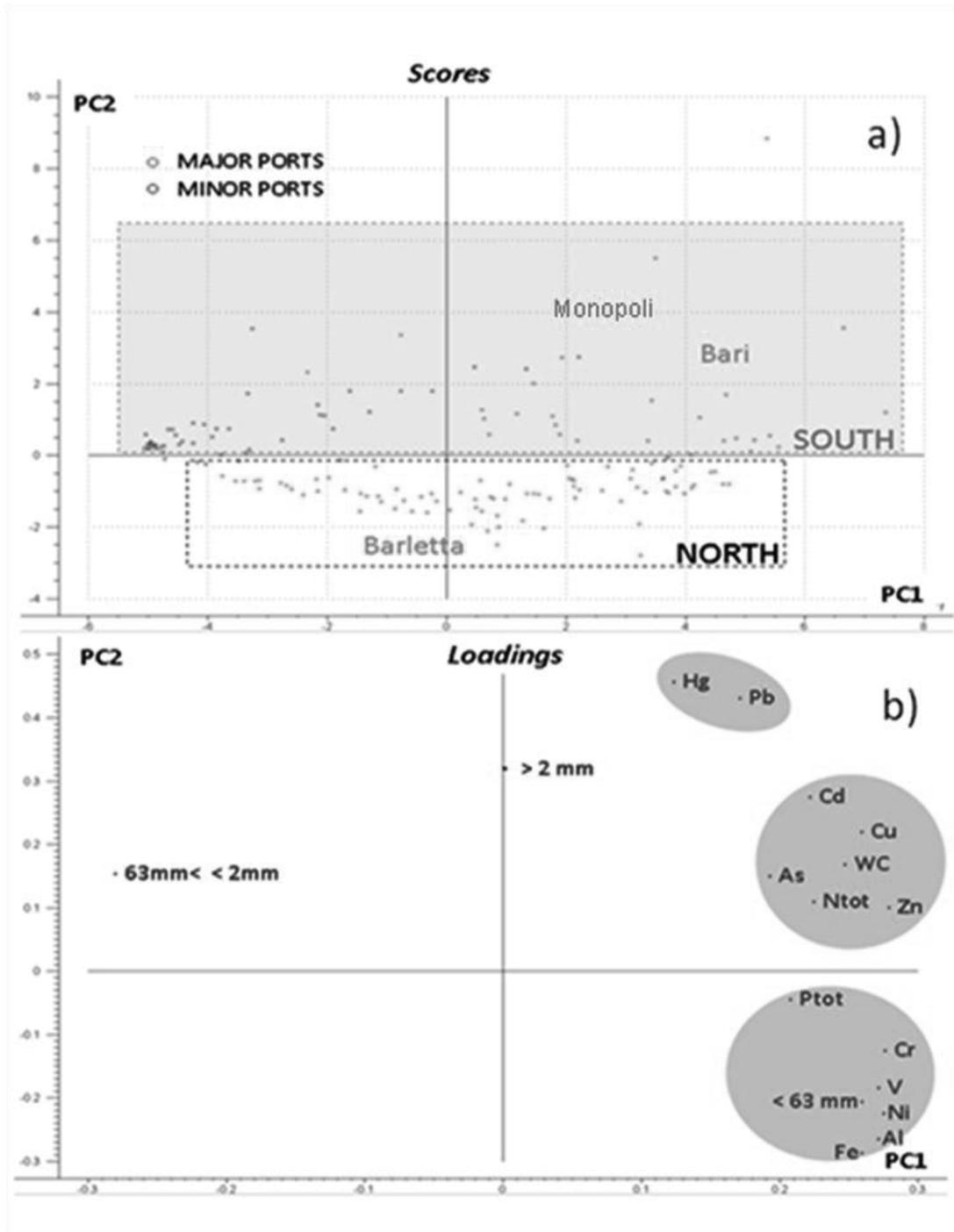


Figure 4 a) e b)



**Supplementary Material**

[Click here to download Supplementary Material: Figures S1\\_Granulometric Distribution and Shepard Diagrams.doc](#)

**Supplementary Material**

[Click here to download Supplementary Material: Figures S2\\_Test of Kolmogorov for AI distribution in coast.doc](#)

**Supplementary Material**

[Click here to download Supplementary Material: Figures S3\\_Box Plot of Heavy Metal Distribution RBL.doc](#)

**Supplementary Material**

[Click here to download Supplementary Material: Figures S4\\_Vertical Profile Heavy Metals vs depth vs grain size.rar](#)

**Supplementary Material**

[Click here to download Supplementary Material: Figures S5\\_Frequency Curves for As and Cu.doc](#)

**Supplementary Material**

[Click here to download Supplementary Material: Figures S6\\_XRD pattern for two areas BT and BA.doc](#)

**Supplementary Material**

[Click here to download Supplementary Material: Figures S7\\_Comparision RGBs site specific - ICRAM - Wedephol.doc](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S1.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S2.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S3.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S4.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S5.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S6.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: table S7.docx](#)

**Supplementary Material**

[Click here to download Supplementary Material: Table S8\\_Sub\\_RGB range North.xls](#)

**Supplementary Material**

[Click here to download Supplementary Material: Table S9\\_Sub\\_RGB range SOUTH.xls](#)

**Supplementary Material**

[Click here to download Supplementary Material: Table S10\\_Sub\\_RGB range MINORI.xls](#)