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

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]

Published version
DOI:10.1016/j.chemosphere.2015.07.046

Publisher:

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[a]*, *Maria Michela Dell'Anna,*[a,] *Piero Mastrorilli*[a], *Leonardo Damiani*[a], *Nicola Ungaro*[b], *Claudia Belviso*[c], *Saverio Fiore*[c]

[a]DICATECh, Politecnico di Bari, via Orabona, 4 I-70125 Bari, Italy
[b]ARPA PUGLIA, Corso Trieste, 27 I-70126 Bari, Italy
[c]IMAA-CNR, C.da S. Loja - Zona Industriale, I-85050, Tito Scalo (PZ), Italy
*Corresponding author e-mail addresses: matilda.mali@poliba.it (M. Mali); 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.

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

1. Introduction

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

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

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

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

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

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

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

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

In this framework, with the aim to gain insight into the most available techniques useful for determining proper heavy metal RGB concentrations in coast sediments, a specific area (110 Km length) of the South-East Italian coast, called “Coastal Area of Bari” (CAB) was chosen as case study for the application of three different conventional statistical methods: *i)* the linear regression analysis (Covelli & Fontolan, 1997) after geochemical normalization (Redon et al., 2013); *ii)* the data processing with the construction of cumulative distribution of frequency (Reimann et al., 2005); and *iii)* the 2σ technique analysis (Matschullat et. al., 2000). The used data set consisted of 158 sediment samples collected at different depths (up to 3.50 m long) coming from almost uncontaminated zones of 3 major ports, 4 minor ports and 6 marine transect of the near shore area (200 m and 500 m from the coast), thus covering the whole 110 Km CAB coastline.

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

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

2. Materials and methods

2.1 Morphological and geochemical features of the study area

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

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

Among the cities bordering the CAB, Barletta and Bari are the most developed, even if no current heavy industrial activities interested the area. Besides pollution caused by water washout of the intensively cultivated areas alongside costs that end in the sea through local hydrographic network, other two important industrial events interested the area: the bombing campaign during the II WAR and the fuels refinery built at north of Bari in the late thirties of the last century and active until 1974. The area is only recently subjected to reclamation work in order to eliminate hydrocarbons residuals that affected the quality of the ground water.

2.2 *Sampling*

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

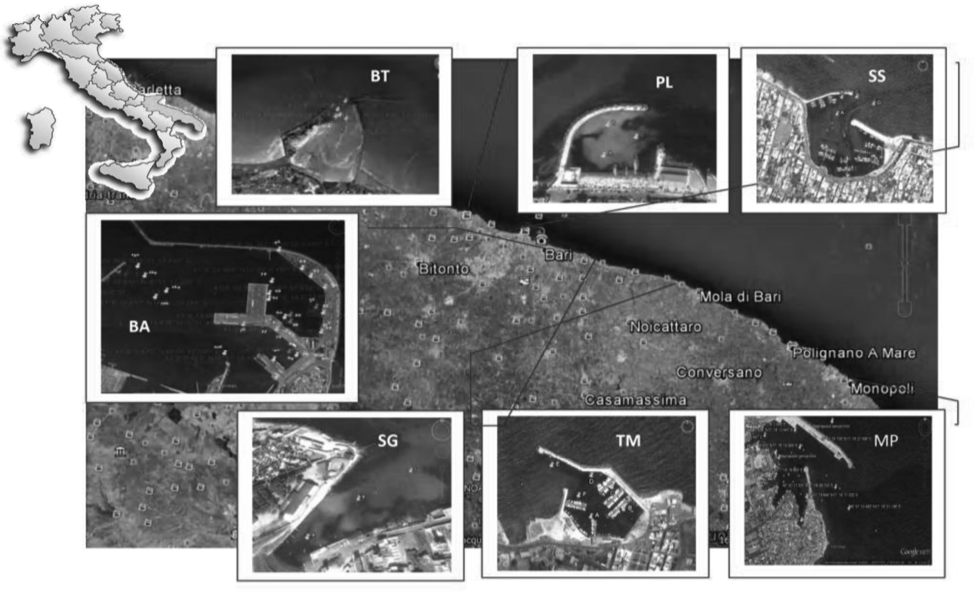


Figure 1: Map of sampling sites and location of sampling stations within ports (insets).

The sampling of sediments from the bottom of sea was carried out by using a vibro-corer PF1, equipped with liner, with the support vessel and provided of the differential system GPS for positioning of sampling cores. The sample cores were up to 3.50 m deep, reaching in any case the rocky hard bottom of the harbor sediment. Each core was divided into several 50 cm length sub-cores. Aliquots were taken by each sub-core and analyzed.

Only in the case of the samples collected from the 6 marine transects, the sediment cores were short (45-60 cm long).

2.3 Analytical methods

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

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

The heavy metal concentrations were determined by inductively coupled plasma mass spectrometry (thermo ICP MS X Series Thermo Fisher Scientific USA) after total acid digestion ($HCl-HNO_3-HF$). Marine Sediment References Material and Standard Reference Materials 2702 were used to control the analysis quality. The Hg content was determined by cold vapor atomic fluorescence spectroscopy following $HNO_3–H_2SO_4$ procedure.

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

2.4 Statistical analyses

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

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

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

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

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

3. Results and discussion

3.1 Sediment characteristics

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

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

BA and BT ports are characterized by a predominance of the mud fraction. Both of them are industrial and commercial ports, therefore designed to minimize hydrodynamic energy within the harbor basin, in order to allow the shipping, loading and unloading activities. In fact, the port structure facilitates the accumulation of mud sediments within port less than other open bay (Hancock G.J., 2001**)**. It is noteworthy that the sandy fraction prevails in MP port and in all the four minor ports (PL – SG – SS – TM). It is assumed that the tidal currents propagating daily through these ports (which can be considered as almost natural bays characterized by a jagged rocky coastline) determine the accumulation of sand within them, explaining the prevalence of this fraction in their sediments.

3.2 Heavy metal distribution

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

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

Investigations of vertical variation of heavy metals in sediment cores permit to reconstruct the history of contamination within an aquatic environment (Swennen and Van der Sluys, 1998). Each sediment core was divided into 50 cm depth intervals. The maximum depth of the sample cores was 2.00 m and 3.50 m, for minor and major ports, respectively, reaching in any case the rocky hard bottom of the harbor sediment. Vertical profiles (amount of metal *vs* depth of the sample core *vs* fine grain size fraction) for Al, Fe, As, Cd, Cr, Cu, Hg, Ni, Pb and Zn for the whole area are illustrated in Figure S4 of the S.I.). The vertical profiles of Al, Fe, Cr, V, Ni and Zn display rather similar trends, while the content of Pb and Hg in some samples is unusually high in almost surficial cores, revealing a probable common source of pollution. Interestingly the As content is unexpectedly high in deep core samples with a low amount of fine grain size fraction (Figure 2).

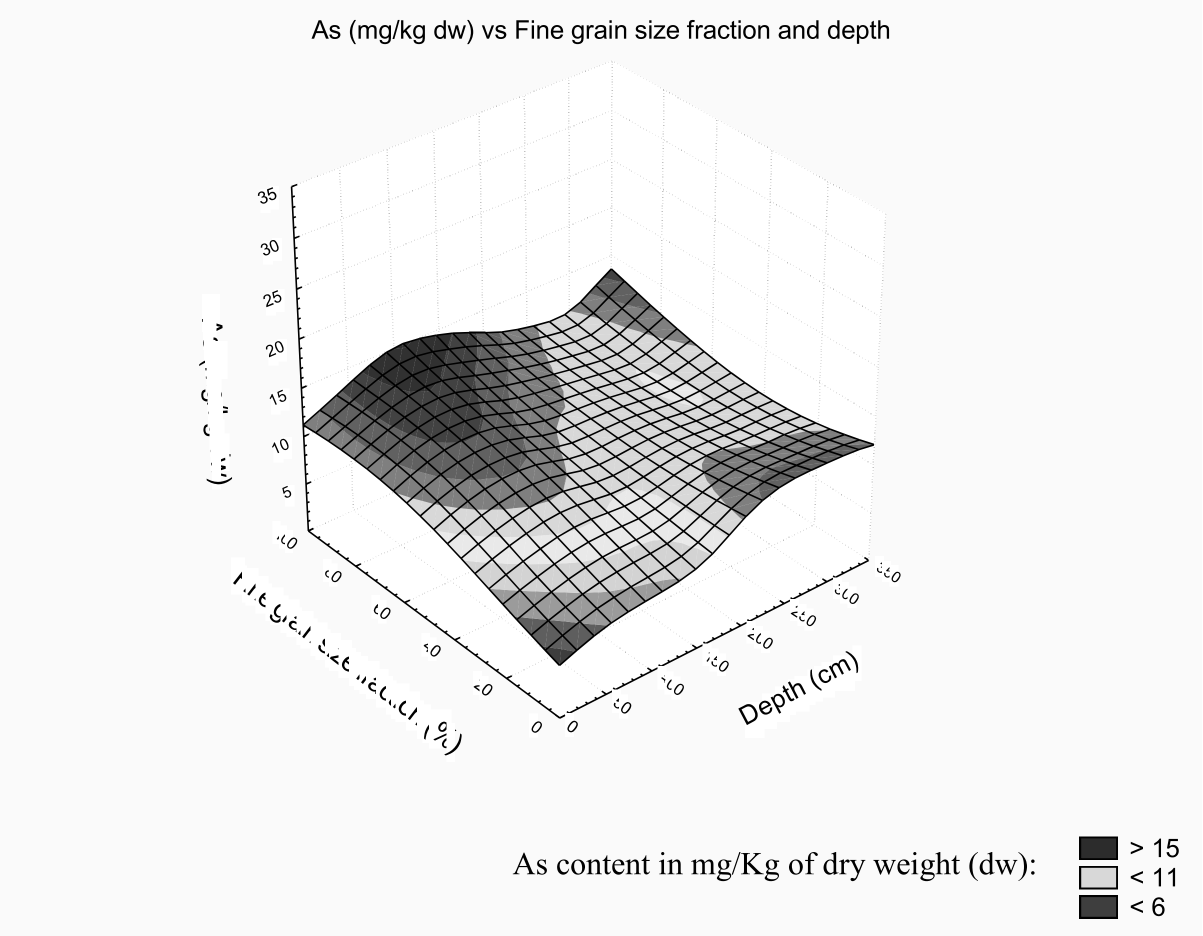


Figure 2: Amount of As *vs* depth of the core sample *vs* fine grain size fraction.

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

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

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

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

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

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

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

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

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

The obtained results allowed us to conclude that, in the presence of a large data set, the use of the cumulative frequency curve or the application of the 2σ technique for the establishing HMe RGB values are more convenient compared to the regression analysis, because the geochemical normalization step is circumvented. On the contrary, when the available data set has a relatively smaller size, the geochemical normalized regression analysis is safer, provided that the studied samples contain fine grain size fraction in the whole percent range (from 3 to 98 % *ca.*) (Covelli and Fontolan, 1997).

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

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

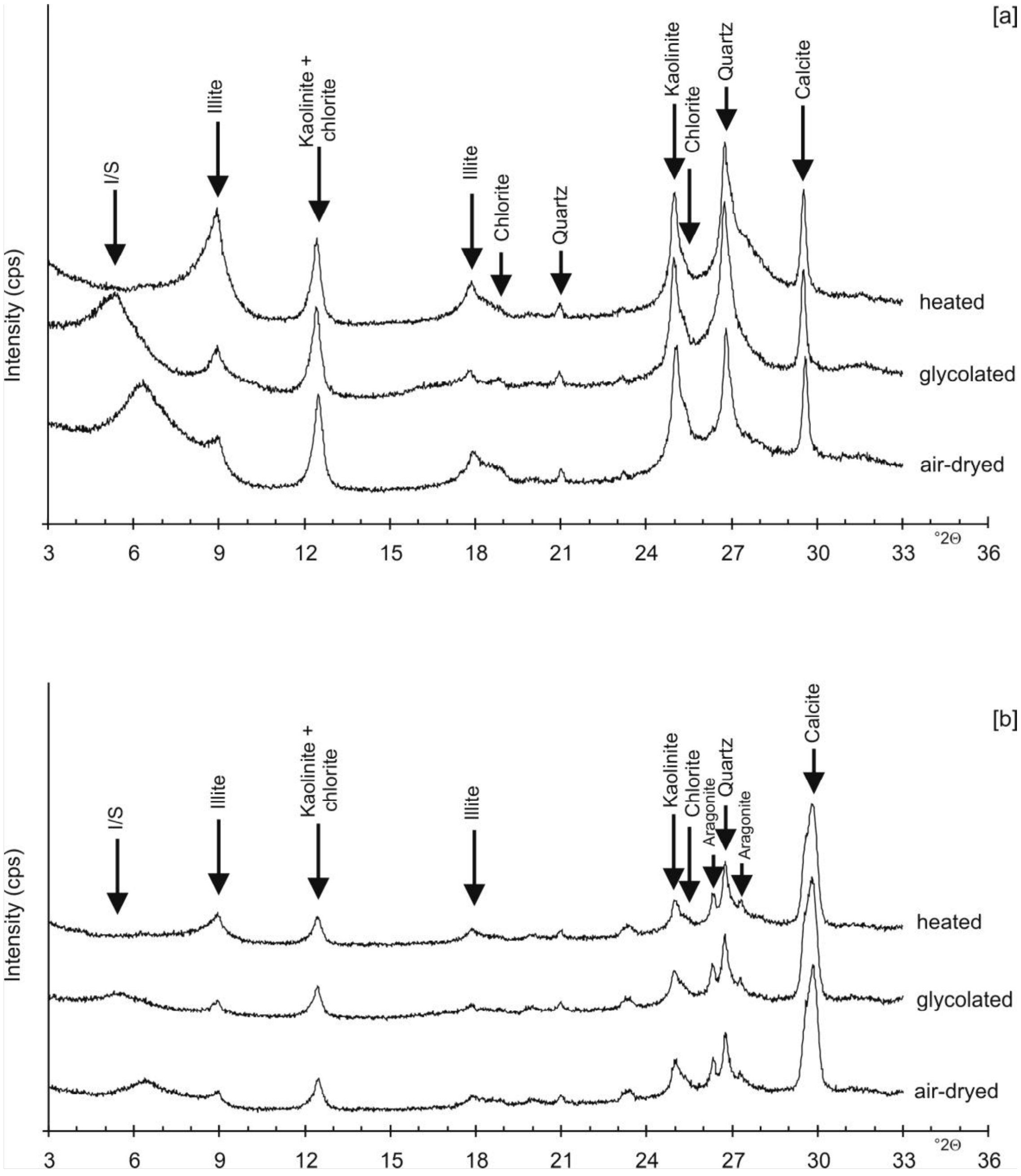


Figure 3. XRD patterns of <2 µm grain-size fraction of a sample representative of [a] BT and [b] BA areas

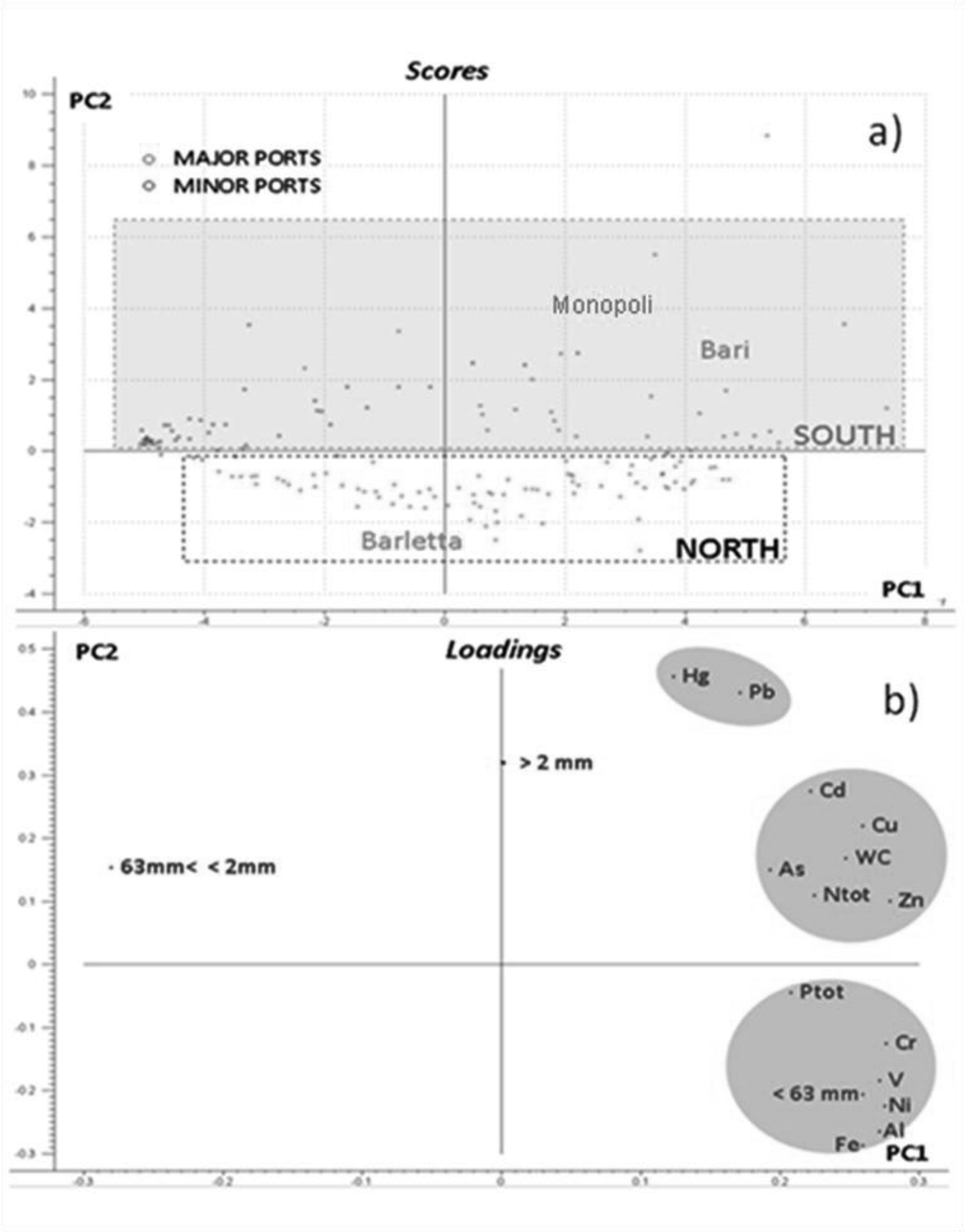


Figure 4. Score (a) and loading (b) plots of PC2 versus PC1 obtained for the data set. In figure 12a, red and blue points indicate samples coming from major (BA, BT and MP) and minor (PL, SS, SG and TM) ports, respectively. In figure 12a, the grey square groups the samples coming from the southern NPU,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

In order to gain insight this first observation, principal component analysis (PCA) as a non-supervised technique was performed on dataset to compare the compositional pattern between the sediment samples and to identify the factors influencing each of them. The entire data set (Tables S1 and S2), excluding three variables (OPR, pH and TOC) which were not available for minor ports, is used for this purpose by means of Unscramble Software. The data set consisted of a matrix with variables in columns (17) and the sediment samples in rows (158) and was mean and centre.

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

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

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

On the base of the PCA results, by using the Al normalization linear correlation method reported above, we calculated for each heavy metal three RGB concentration ranges, belonging to the northern major port (BT port, Table S8 of the S.I.), the southern major ports (BA and MP ports, Table S9 of the S.I.) and all the minor ports (Table S10 of the S.I.), respectively. Differences among the considered zones were obtained. As, Ni, Cr, V and Zn background concentrations were higher in the major port of the northern area (BT port) than in the major ports of the southern one, while background values of Pb, Cd, Hg and Cu were higher in the major ports of the southern part than in the major port of the northern zone. However, for both the considered zones, for some metals the so obtained background concentrations were not in line with the national guidelines. On the other hand, lower concentration of all metals were found in the minor ports and the background values determined for this sub-area resulted similar to the average crustal shale values established by Turekian and Wedepohl (1961) (Figure S7 of the S.I.).

4. Conclusions

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

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

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

The obtained results point out that an ideal method for assessing regional heavy metal background concentrations does not exist, but it must be selected for each specific system. In addition, deeper investigations (such as mineralogical analysis) and multivariate statistical analyses carried out on the whole data set 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.

Acknowledgments.

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

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

References

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

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

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

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

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

Dassenakis, M., Adrianos, H., Depiazi, G., Kostatas, A., Karabela, M., Sakellari, A., 2003. The use of various methods for the stydy of metal pollution in marine sediments, the case of Euvoikos Gulf, Greece. App. Geochem.18, 781-794.

Dell'Anna, L., Laviano, R, 1981.) Mineralogia delle Argille Plioceniche di Potenza. Rendiconti Soc. Ital. Min. Petrol., 37, 179-194. *(in Italian)*.

Desaules, A., 2012. Critical evaluation of soil contamination assessment methods for trace metals. Sci. Total Environ. ,426,120–131.

Dung, T.T.T., 2013. Cappuyns V., Swennen R., Phung N. K., Rev. Environ. Sci. Biotechnology 12, 335-353.

Galuszka, A., 2007. A review of geochemical background concepts and an example using data from Poland. Environ Geol. 52, 861–870

Galuszka, A., Migaszewski, Z.M., 2011. Geochemical Background - an environmental perspective, Miner. 42, 7-17.

Garrett, R.G., 1991. The management, analyses and display of exploration geochemical data. Exploration geochemistry workshop, Ottawa Geological Survey of Canada, Open File 2390.

Hancock, G.J., Olley, J.M. and Wallbrink, P.J. Sediment transport and accumulation in Westernport, *CSIRO Land and Water Technical Report*, 2001 47/01.

Ho, H.H., 2012. Swennen R, Cappuys V, Vassilieva E, Tran TV. Necessity of normalization to aluminum to assess the contamination by heavy metal sand arsenic in sediments near Haiphong Harbor, Vientam. J. Asian Earth. Sci. 56, 229-239.

Hughes A.O., Prosser I.P., Wallbrink P.J. Suspended Sediment and Bedload Budgets for the Western Port Bay Basin CSIRO Land and Water, Canberra Technical Report 4/03

Joksas, K, Stakeniene, R., Galkus, A., Lagunaviciene, L., 2008. Metals in bottom sediments of Sventoji Port Area - Lituania. Geol. 50, 143-155.

Lepeltier, C., 1969. A simplified treatment of geochemical data by graphical representation. Econ. Geol 64, 538-550 .

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

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

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

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

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

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

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

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

Reimann, C., De Caritat, P., 2004. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. Sci. Total Environ. 337, 91 – 107.
Reimann, C., Garrett, R.G., 2005. Geochemical background. Concept and reality. Science of the Total Environment 350, 12–27.
Reimann, C., Filzmoeser, P., Garret, R.G., 2005. Background and threshold: critical comparison of methods of determination. Sci. Total Environ. 346, 1-16.

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

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

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

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

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

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

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

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

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

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

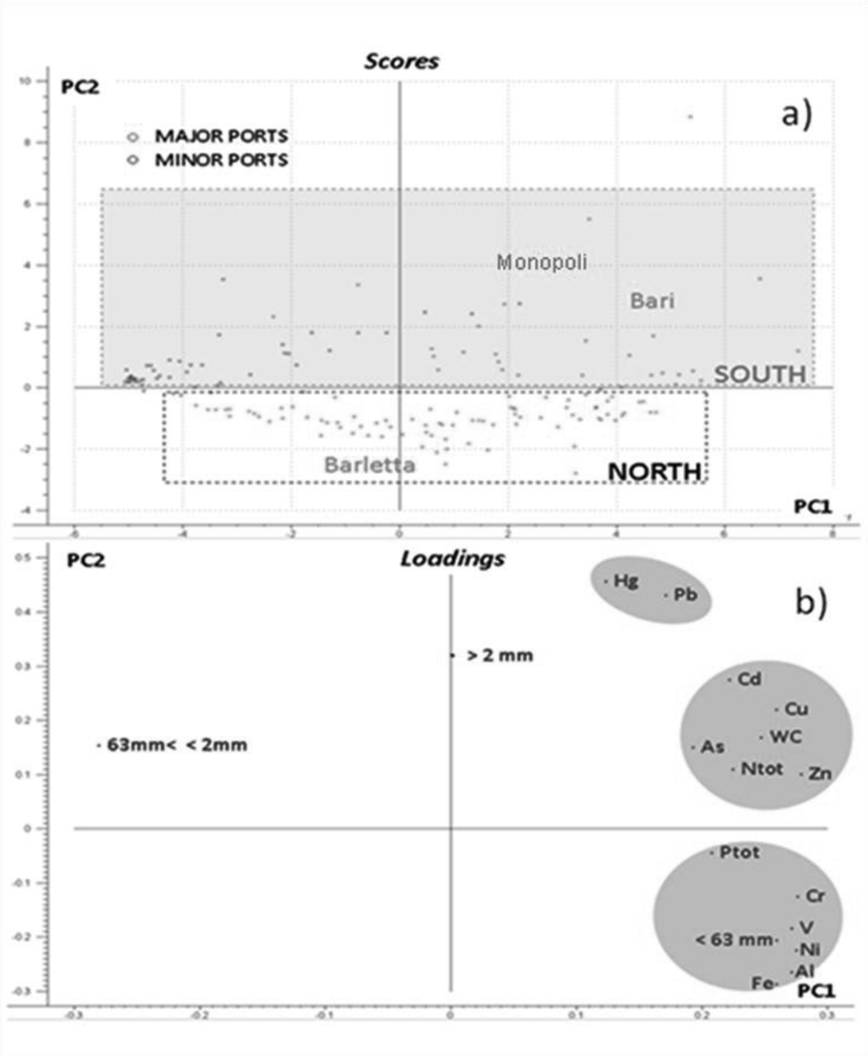
- Heavy metal background concentrations in South-East Italy sediments were calculated
- Three conventional statistical techniques were used and compared to each other
- The conventional techniques led to the same range of heavy metal background values
- Mineralogical and multivariate statistical analyses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

Graphical Abstract (for review)

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[a]*, *Maria Michela Dell'Anna*[a], *Piero Mastrorilli*[a], *Leonardo Damiani*[a], *Nicola Ungaro*[b], *Claudia Belviso*[c], *Saverio Fiore*[c]

[a]DICATECh, Politecnico di Bari, via Orabona, 4 I-70125 Bari, Italy
[b]ARPA PUGLIA, Corso Trieste, 27 I-70126 Bari, Italy
[c]IMAA-CNR, C.da S. Loja - Zona Industriale, I-85050, Tito Scalo (PZ), Italy
*Corresponding author e-mail addresses: matilda.mali@poliba.it (M. Mali); 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 µm 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

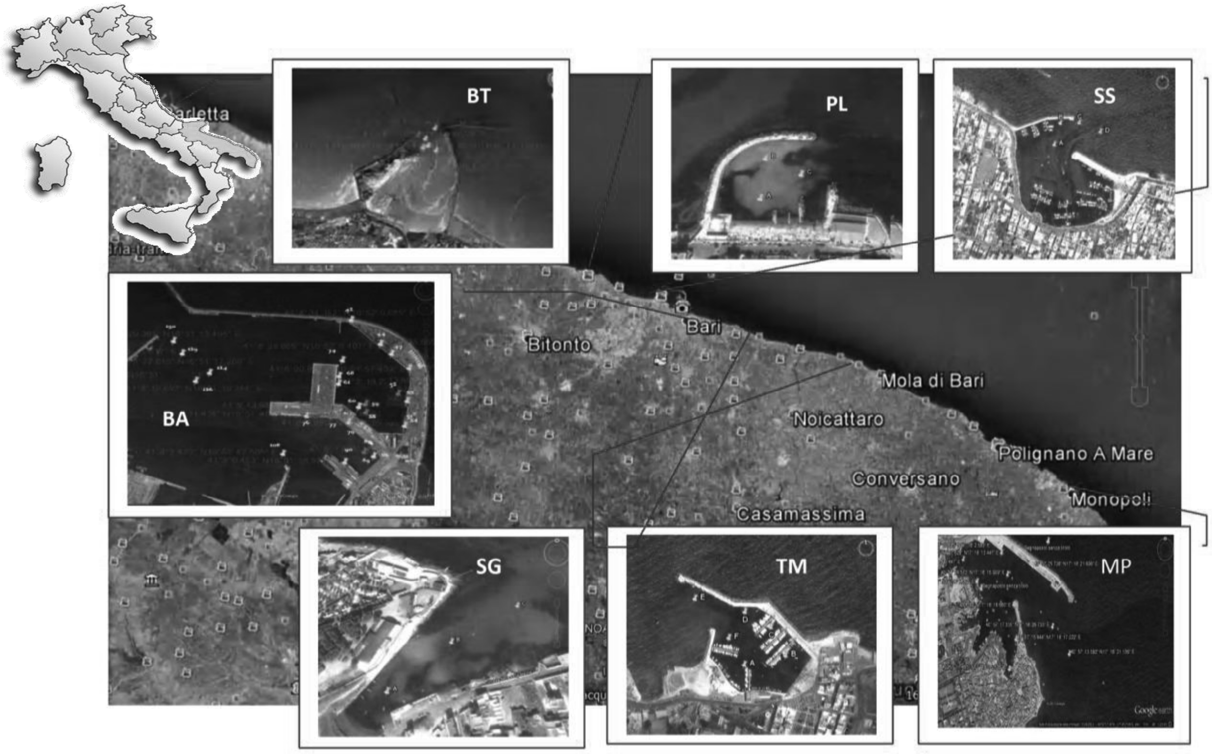

BT
PL
SS
BA
SG
TM
MP
arletta
Bitonto
Bari
Mola di Bari
Noicattaro
Polignano A Mare
Conversano
Monopoli
Casamassima

Figure 2

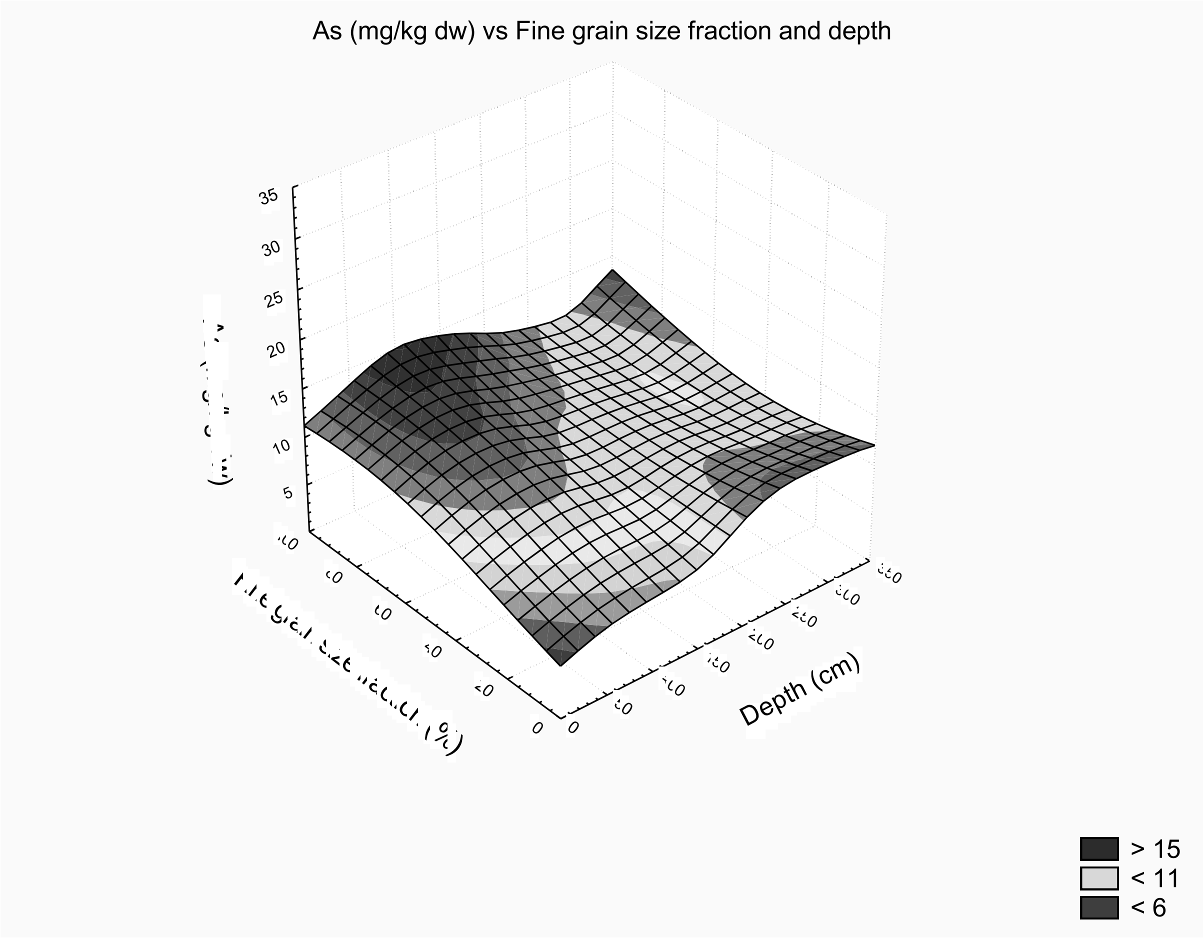

As (mg/kg dw) vs Fine grain size fraction and depth
35
30
25
20
15
10
5
0
0
Depth (cm)
> 15
< 11
< 6

Figure 3

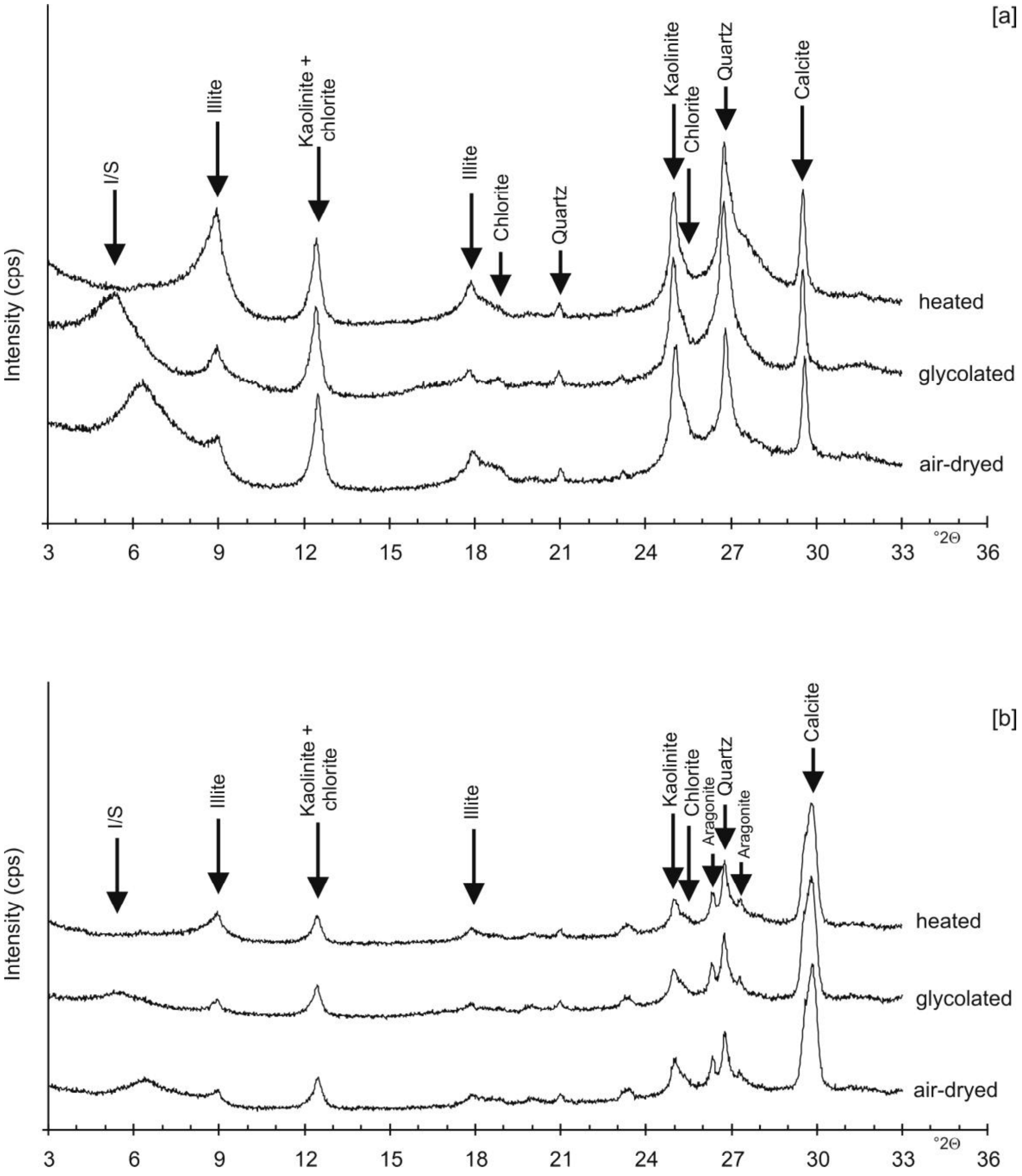
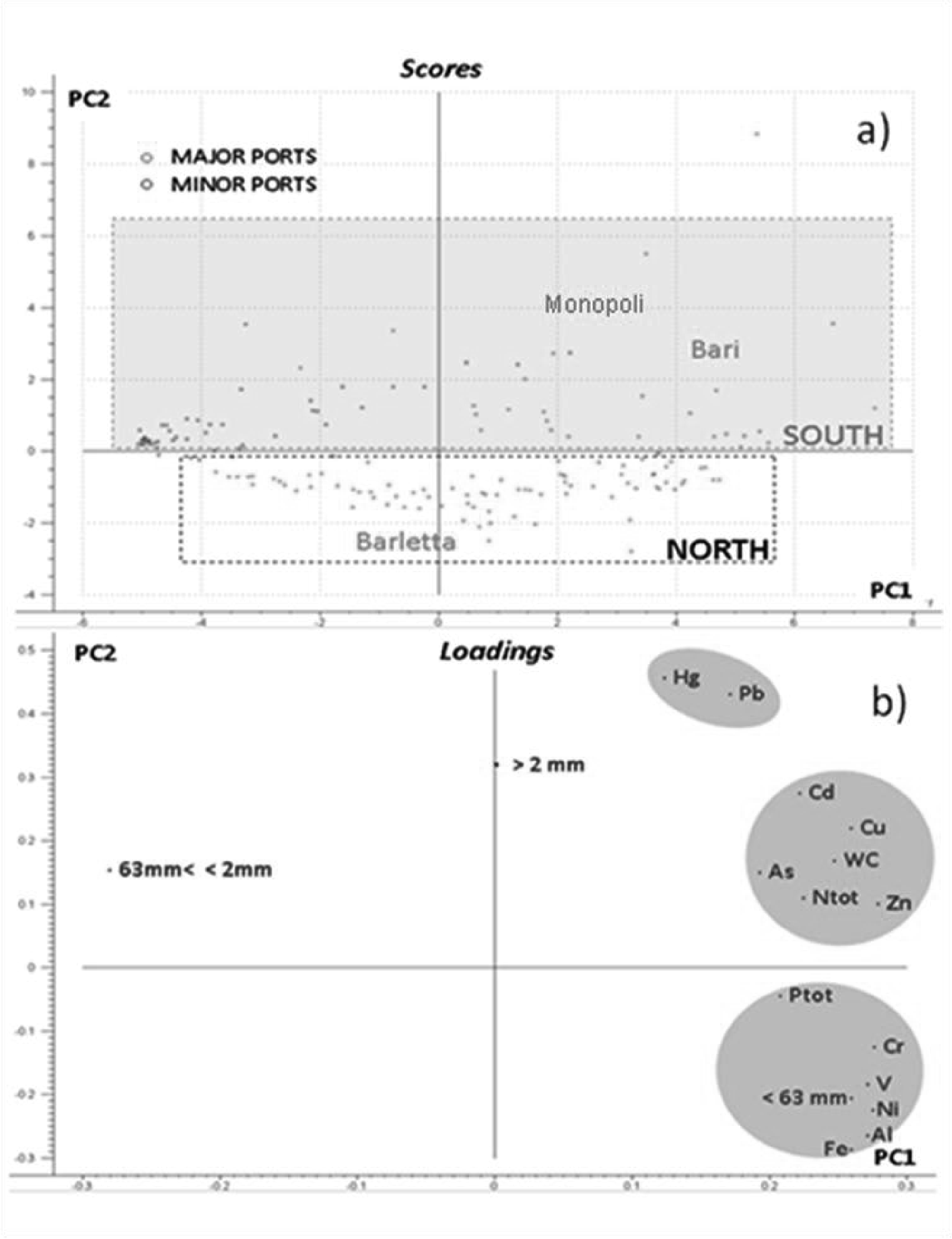

[a]
Intensity (cps)
I/S
Illite
Kaolinite +
chlorite
Illite
Chlorite
Quartz
Kaolinite
Chlorite
Quartz
Calcite
heated
glycolated
air-dryed
3
6
9
12
15
18
21
24
27
30
33
36
°2Θ
[b]
Intensity (cps)
I/S
Illite
Kaolinite +
chlorite
Illite
Kaolinite
Chlorite
Aragonite
Quartz
Aragonite
Calcite
heated
glycolated
air-dryed
3
6
9
12
15
18
21
24
27
30
33
36
°2Θ

Figure 4 a) e b)



Scores
a)
PC2
PC1
MAJOR PORTS
MINOR PORTS
Monopoli
Bari
SOUTH
Barletta
NORTH
10
8
6
4
2
0
-2
-4
-6
-4
-2
0
2
4
6
8
Loadings
b)
PC2
PC1
Hg
Pb
> 2 mm
Cd
Cu
WC
As
Ntot
Zn
63mm< < 2mm
Ptot
Cr
V
< 63 mm
Ni
Al
Fe
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.3
-0.2
-0.1
0
0.1
0.2
0.3

Supplementary Material

Supplementary Material

Supplementary Material

Supplementary Material