Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2013.848649

52 (2014) 1171–1176
January

Taylor & Francis
Taylor & Francis Group

Proposal of procedure to determine metals and metalloids background values in contaminated soils. Case study of a national interest site in South Italy

Massimo Blonda, Barbara Valenzano*

Regional Agency for Environmental Prevention and Protection, ARPA Puglia, Corso Triestre 7, 70126 Bari, Italy
Email: b.valenzano@arpa.puglia.it

Received 3 March 2013; Accepted 30 March 2013

ABSTRACT

The present work concerns the development of a procedure to determine background values of metals and metalloids in soils of sites of national interest that are subjected to characterization and remediation. The proposed calculation procedure allows to select in metal-metalloid data-set, the representative parameters of site-specific pollution sources. It allows to identify anthropogenic concentration of substances in the soil resulting both natural pedo-geochemical content and moderate widespread immission in soil. The statistical analysis was performed in samples set of topsoil, soil and underground soil. Chemical tests were carried out on all parameters of metals and metalloids present in samples. This work has analized parameters, as As and Be, which have passed statistical test "Shapiro e Wilk" (W test) with contamination "hot spot" of anthropogenic, respect to a large area of substantially homogeneous values. Specifically, this study focuses on area adjacent to perimeter of site of national interest of Brindisi (South Italy). It is characterized by the presence of many process industries and of large power plants. Due to many human activities in this site and difficulty of access to sampling points, analysis was carried out starting from No. 30 sampling sites located in the province of Brindisi. This study was carried out to define a suitable criterion for determination of background concentrations in industrial area. It was found that cumulative frequency curve deviates greatly from straight trend which represents the trend of other measures. In all cases examined, inflection point was coincident with 90° percentile. On the other hand, value distribution of arsenic concentration in subsoil shows that there are hot spots of contamination probably of anthropogenic origin, different from those found in a other large site that are result substantially homogeneous. However, it is not possible to exclude these points from sample distribution and, therefore, they are estimated. Having to provide a representative background value of concentrations of metals present in area was used a conservative approach based on analysis of discontinuity points in cumulative frequency curve. In this

---

*Corresponding author.

Presented at the Fifth Annual International Conference on "Challenges in Environmental Science & Engineering—CESE 2012" Melbourne, Australia, 9–13 September 2012

way, background value about contamination anthropogenic source corresponds to previous value at point of discontinuity in cumulative frequency curve (90° percentile). Thus, background conditions would be identified by graphic of concentrations distribution, from the zero point to the point of inflection.

Keywords: Heavy metals; Metalloid; Contaminated soil; Remediation; W test

1. Introduction

The identifying of areas representatives has taken account of areas that have the following characteristics:

- geologically comparable with the examined area;
- the distribution of concentrations of metals/metalloids was not attributable to any source of contamination on time and/or specific act.

In the situation under examination, it was not possible to exclude the point sources of contamination such as industrial plants. Finally, it is believed that the identification of representative areas for determination of values "natural" cannot be separated from considering any case of diffuse contamination phenomena that also affect the geochemical characteristics of the outcropping rocks. One of the most significant sources of contamination, both locally and regionally, is represented by agriculture.

2. Location of sampling points and sample taking

Ten representative areas were identified for sampling. The equation used for minimum number determination of samples for verification of distribution mean, against a threshold value [1–3], by of t-test application, was:

$$n = \frac{S^2\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\Delta^2} + 0.5Z_{1-\alpha}^2 \qquad (1)$$

where n is minimum number of samples; s^2 is estimate of true total variance (σ^2); is acceptable probability that test indicating, incorrectly, that concentrations average do not exceed maximum allowable concentration; β is acceptable probability that test indicating, incorrectly, that concentrations average exceed maximum allowable concentration; Δ is defined as minimum detectable difference; Z is value, for a data normal distribution, for which distribution proportion the left of $Z_{1-\alpha}$ is equal to $1-\alpha$.

The assumption of above equation is that data distribution is normal type, data are representative of population, and the data are not correlated in time and space. The acceptable level of error is given by decision-maker and it is expressed by confidence level $(1-\alpha)$ and power $(1-\beta)$.

In each sampling station, three samples were collected: the topsoil, a representative sample of the surface layer to 0–1 m, and a representative sample of the soil depths of 1–6 m; for a total of No. 30 samples [4].

Sample points were identified on geological considerations [5]. Geological study has excluded "natural" presence of As in soil, subsoil, and groundwater investigated in site of Brindisi Province. The Fig. 1 shows No. 30 samples.

The analytical determinations for samples taken for determination of background values were performed with official methods of analysis (methods UNI EN 16173:2012, EN 16174:2012, UNI EN ISO 15587-1:2002 + UNI EN ISO 15587-2:2002) recognized nationally and/or international level. Total concentration of each metal/metalloid was determined analytically with ICP mass (methods UNI EN ISO 17294-1:2007, EN ISO 17294-2:2005) after dissolution in water "regia" [6].

3. Statistical analysis of data—metals/metalloids: arsenic and beryllium

It is been called the data-set corresponding to representative sample. We proceeded to the determination of the background value by the application of statistical tests according to following steps:

- preliminary analysis of data-sets (identification and treatment of outliers and non-detect);
- definition of the distribution of data (statistical tests);
- representation of distribution of data (numerical and graphical descriptors) and definition of concentration values representative of the background value.

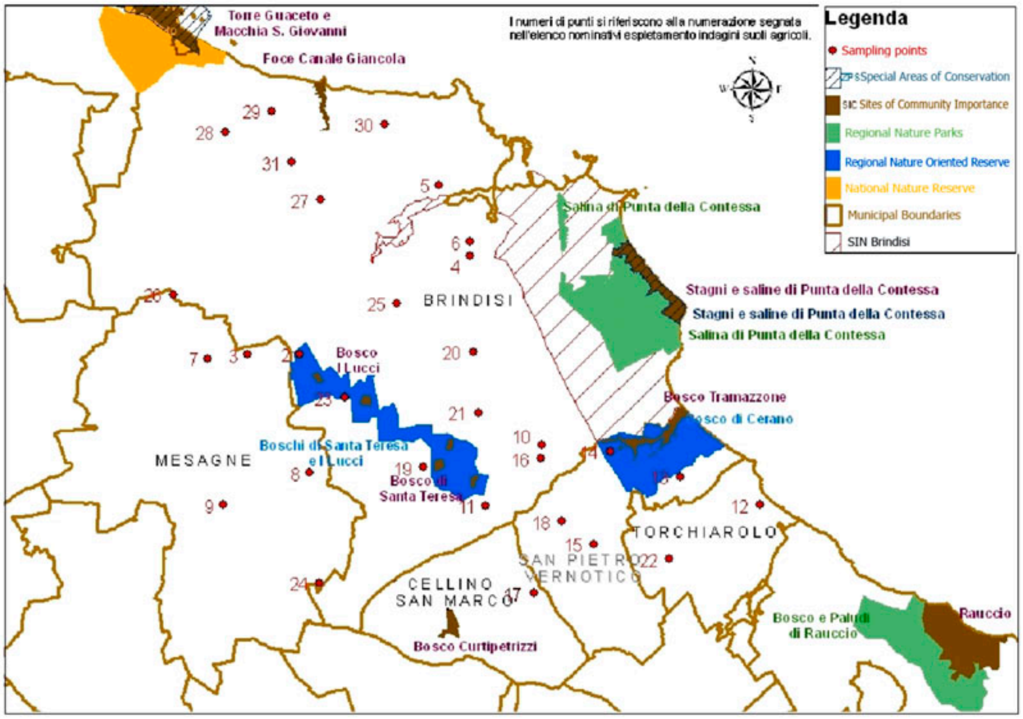

Fig. 1. Location of sampling points.

3.1. Preliminary analysis of data-sets: identification and treatment of outliers and non-detect

The data-set available has been subjected to a preliminary validation procedure. It is continued with a statistical analysis of a second level through the application of a specific statistical test. The identification of outliers was conducted according to the method EPA 2000b, QA/G-9 [7–9].

In this study, Rosner's test was applied both to the data-set "Surface Soil Arsenic" and the averaged data-set "Topsoil—Surface Soil Arsenic." This test has led to exclusion of No. 3 outliers in the first case and No. 1 outlier in the second case.

In the first case, the test has identified outlier values as 49.4, 45.3, and 39.00 mg/kg, (respectively to samples 10–11–22); in the second case, an outlier value was identified as 43.7 relating to sample 10.

3.2. Definition of the data distribution (statistical tests)

All serial numbers (without outliers) are found to have a lognormal distribution by applying the statistical test "Shapiro and Wilk" (*W* test) [10]. As the number of available data was less than No. <50. By way of example, we report the graph with the values of arsenic in subsoil (Fig. 2).

3.3. Representation of the distribution of data (numerical and graphical descriptors)

Data processing is the final point of entire process of determining the value of the background. It is strongly dependent on the number of data that have at their disposal and it is important to determine at design the appropriate number of samples to be taken and analyzed, as described in previous paragraphs. The processing was performed for each data group falling within the homogeneous areas.

In data processing, data of topsoil and topsoil have been assimilated in a single data-set. The statistical parameters are determined as follows: mean value, standard deviation, coefficient of variation (CV%), maximum, median, and minimum (Table 1).

Then, we proceeded to construction of the cumulative frequency curve.

For the construction of the cumulative distribution of frequency, we have adopted the following formula:

$$\mathrm{CF}_i = \sum_{j=1}^{i} \mathrm{AF}_j \tag{2}$$

AF_i, value of the absolute frequency (i.e. in the number of times that value was observed); j, number

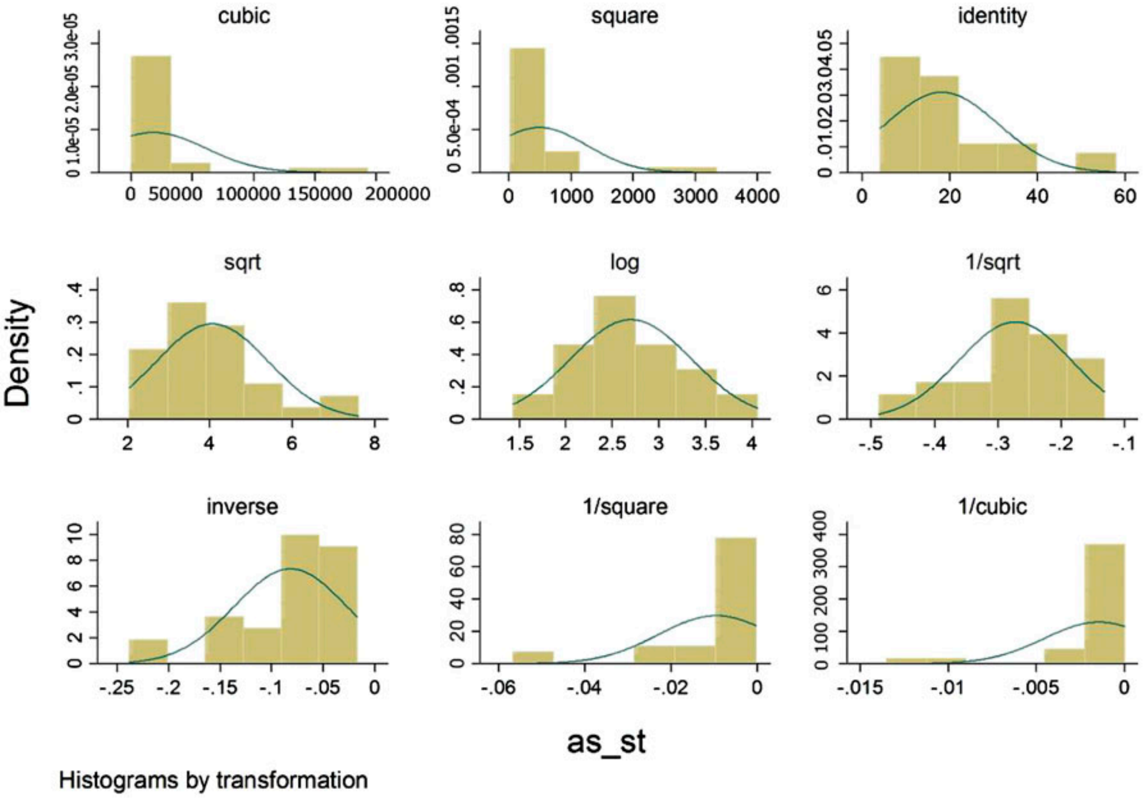

Fig. 2. Arsenic distribution subsurface.

Table 1
Descriptors of statistical data series arsenic and beryllium (observed values)

Data-set median	Minimum maximum	Average	Standard deviation	Median	Coefficient of variation
Topsoil arsenic	5.10–38.00	12.43	7.03	11.20	56.53
Surface soil arsenic	5.70–21.30	11.58	4.17	10.20	35.99
Topsoil arsenic—soil surface	5.50–33.70	12.59	6.75	10.50	53.54
Arsenic underground	4.20–57.80	18.22	12.80	14.85	70.24
Beryllium topsoil	0.60–3.40	1.49	0.61	1.50	41.01
Beryllium soil surface	0.70–2.80	1.52	0.61	1.30	40.00
Beryllium topsoil—soil surface	0.65–2.75	1.51	0.56	1.40	37.09
Beryllium underground	0.20–1.40	0.64	0.39	0.55	60.98

of frequency classes; CF_i, number of observations that are less (or equal) to the value $x(i)$.

The cumulative rates for each value of $\boldsymbol{i}$ are obtained by normalization of CF_i.

From the course of curves obtained, we have information about the distribution of sample. A linear trend indicates a normally distributed sample.

The trends curvilinear were made by the linear logarithmic scale for the values $x(i)$. The distributions are log-normal, as found by applying the W test.

We set out below the cumulative frequency curves calculated with reference to arsenic and beryllium for the three levels of soil investigation, i.e. topsoil (0–0.1 m), surface soil (0–1 m), topsoil—surface (0–0.1 m e 0–1 m), and subsoil (1–6 m). In addition, we report the cumulative curves mediated between the values of arsenic and beryllium detected in the topsoil (0–0.1 m) and soil surface (0–1 m). The distributions of all values were log-normal [11,12] (Figs. 3 and 4).

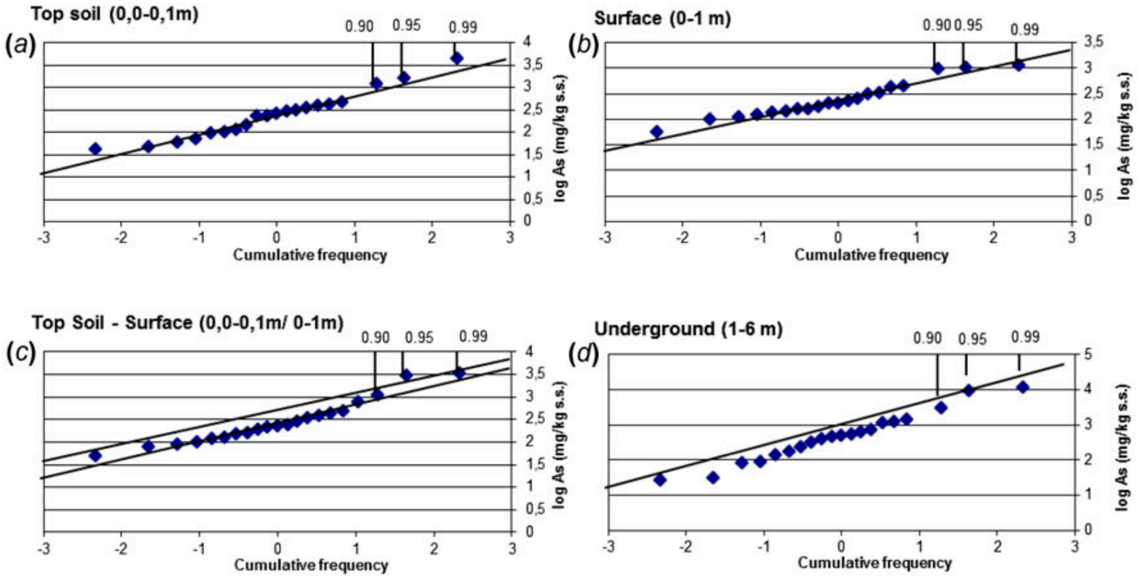

Fig. 3. Cumulative frequency curve of arsenic.

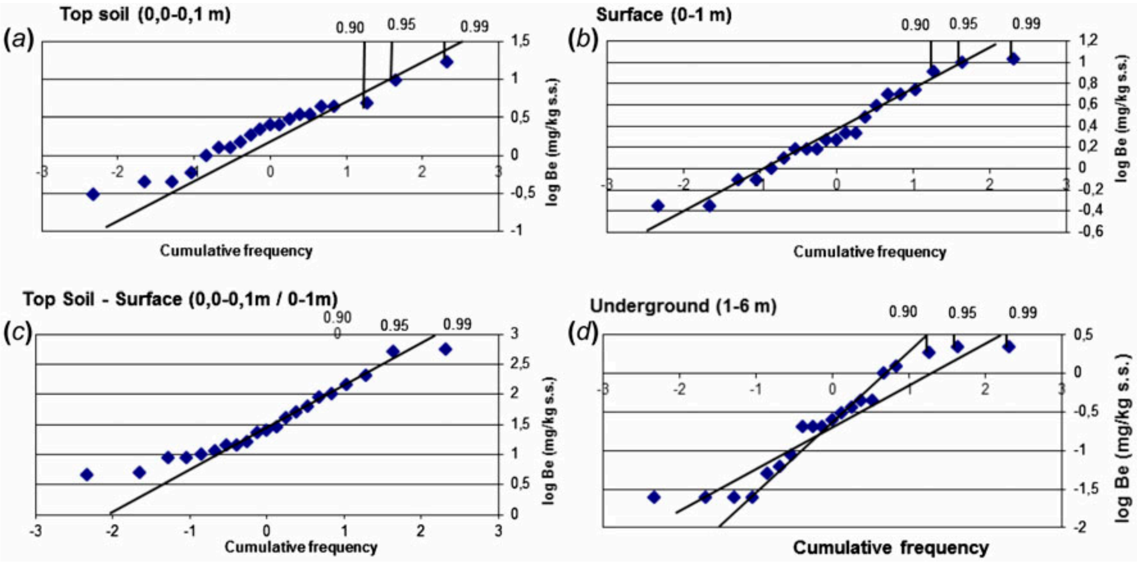

Fig. 4. Cumulative frequency curve of beryllium.

4. Conclusions

The analysis of scientific studies, nationally and internationally, has highlighted the lack of a policy on the shared determination of background concentrations in the area. In the case study, this determination is complicated by the inability to exclude the presence of anthropogenic contamination. The difficulty is to identify areas located near industrial sites with comparable geologic characteristics for determine a "natural background"; as these areas are affected of anthropic activities and of agricultural activities practiced in near areas of industrial site. The proposed procedure has determined "anthropic background" representative and not properly natural.

In fact, in particular as regards the measures of Arsenic in the subsoil, there are three observations which are situated above the 90° percentile, dragging to the right indicators of central tendency, and supporting a high variability of the distribution of the sample values. Looking at the curve of the cumulative distribution of log-transformed values, it is noted that these observations deviate sharply (from the linear trend line) represent the performance of other measures. Moreover, the inflection point, it was found coincident with the 90° percentile. Having to provide a value representative of background concentrations of metals present in the area, it was decided to use a conservative approach based on analysis of discontinuity points in the curve of cumulative frequency. The value of man-made background of this area corresponds to value preceding the discontinuity point (which corresponds to the 90° percentile) in the graph of the cumulative distribution of frequency (quantile of concentration).The conditions of the background value would be identified by the distribution of concentrations of graph, from the origin to inflection point. The main point of inflection was identified through the study of derivatives and the same graph, quantile of concentration.

References

[1] US Environmental Protection Agency, Guidance for Data Quality Assessment. Practical Methods for Data Analysis. EPA QA/G-9, Office of Environmental Information, Washington, DC, 2000.

[2] US Environmental Protection Agency, The Data Quality Evaluation Statistical Toolbox (Data QUEST) Software, EPA QA/G-9D, Office of Research and Development, Washington, DC, 1996.

[3] US Environmental Protection Agency, Guidance on Systematic Planning Using the Data Quality Objectives Process EPA QA/G-4 Final, Office of Environmental Information, US Environmental Protection Agency, Washington DC, 2006.

[4] International Organization for Standardization, Soil Quality—Guidance on the Determination of Background Values. ISO19258—ISO TC 190/SC 7, Austrian Standards Institute, Austrian, 2005.

[5] R. Salminen, T. Tarvainen, A. Demetriades, M. Duris, F.M. Fordyce, V. Gregorauskiene, H. Kahelin, J. Kivisilla, G. Klaver, H. Klein, J.O. Larson, J. Lis, J. Locutura, K. Marsina, H. Mjartanova, C. Mouvet, P. O Connor, L. Odor, G. Ottonello, T. Paukola, J.A. Plant, C. Reimann, O. Schermann, U. Siewers, A. Steenfelt, J. Van der Sluys, B. de Vivo, L. Williams, FOREGS Geochemical Mapping. Field Manual. Geologian Tutkimuskeskus, Opas—Geological Survey of Finland, Guide 47. p. 36, 15 figures, and 1 appendix, 1998.

[6] Italian Environmental Protection Agency—ISPRA (ex APAT), Operational Protocol for the Determination of Background Values of Metals/Metalloids in Soils of National Interest, June 2006.

[7] US Environmental Protection Agency, Data Quality Objectives Process for Hazardous Waste Site Investigations. EPAQA/G-4HW, Final. Office of Environmental Information, US Environmental Protection Agency, Washington, DC, 2000a.

[8] US Environmental Protection Agency, Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA QA/G-9, QA00 Update, Office of Environmental Information, Washington, DC, 2000b.

[9] US Environmental Protection Agency, Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. Office of Emergency and Remedial Response US Environmental Protection Agency, Washington, DC, 2002.

[10] Italian Environmental Protection Agency—ISPRA (ex APAT), Methodology for Applying Absolute Risk Analysis to Contaminated Sites, Rev., ISPRA, Manuali e Linee Guida, Roma, 2008.

[11] US Environmental Protection Agency, RAGS: Volume 3 PART A-Process for Conducting Probabilistic Risk Assessment Appendix A, Environmental Protection Agency, Washington, DC, 2001.

[12] US Environmental Protection Agency, Calculating Upper Confidence Limits for Exposures Point Concentrations at Hazardous Waste Sites. Oswer 9285.6–10, Office of Emergency and Remedial Response, USEnvironmental Protection Agency, Washington, DC, 2002.