Alicja Kicińska, Katarzyna Turek

ESTABLISHING GEOCHEMICAL BACKGROUND OF ELEMENTS PRESENT IN SOIL AND ITS APPLICATION IN THE EVALUATION OF SOIL POLLUTION BASED ON DATA COLLECTED IN THE BESKID SĄDECKI REGION

AGH University of Science and Technology,
Faculty of Geology, Geophysics and Environmental Protection,
Al. Mickiewicza 30, 30-052 Kraków, Poland,
kicinska@geol.agh.edu.pl, katarzynaturek2@gmail.com

Abstract

The present paper describes problems related to establishing geochemical background. Three selected methods: [Av±2σ], [Me±2MAD] and Tukey Box Plot were applied to establish the geochemical background of soils from the Beskid Sądecki geographical region in southern Poland. The ranges of geochemical background calculated for the following elements: Mg, Ca, Mn, Pb, Cd and As, for 2 sets of data with a different degree of human pressure and sampling area showed considerable differences. The results obtained based on a set of samples collected from a stable, regional forest ecosystem seem the most appropriate for establishing geochemical background of large areas. The paper confirms that the [Av±2σ] method should not be used in the case of data with skewed distribution and a large degree of pollution. Analysing numerous sets of data and several methods also allowed for a conclusion that selecting a single universal method for establishing the background range is impossible. When selecting a method (direct, indirect or integrated), the pollution level of a given area, variability of the tested substance concentration, its geochemical characteristics, data distribution etc. should be taken into consideration. The box plot method eliminates fewer outliers and usually results in obtaining a broader background range. This method should be used mostly for areas with a low degree of pollution and a considerable diversity of element concentration.

WYZNACZANIE TŁA GEOCHEMICZNEGO DLA PIERWIASTKÓW WYSTĘPUJĄCYCH W GLEBACH ORAZ JEGO ZASTOSOWANIE W OCENIE ZANIECZYSZCZENIA GLEB, NA PRZYKŁADZIE BESKIDU SĄDECKIEGO

Abstrakt

W pracy scharakteryzowano problematykę związaną z wyznaczaniem tła geochemicznego. Trzy wybrane metody: [Av±2σ], [Me±2MAD] oraz Tukey Box plot zostały zastosowane do wyznaczenia tła geochemicznego gleb pochodzących z obszaru Beskidu Sądeckiego, dla następujących pierwiastków: Mg, Ca, Mn, Pb, Cd and As, dla 2 zbiorów danych, różniących się stopniem antropopresji oraz powierzchnią obszaru opróbowania. Najbardziej odpowiednie do wyznaczania tła geochemicznego dla dużych obszarów wydają się wyniki uzyskane na podstawie zbioru próbek pobranych ze stabilnego ekosystemu leśnego, który miał charakter regionalny. W pracy potwierdzono, że w żadnym wypadku nie należy stosować metody [Av±2σ] do danych charakteryzujących się skośnym rozkładem oraz dużym stopniem zanieczyszczenia. Przeanalizowanie wielu zestawów

danych i wybranych metod, pozwala także wnioskować, że dobór jednej uniwersalnej metody do wyznaczenia zakresu tła jest niemożliwy. Przy doborze metody (pośredniej, bezpośredniej czy zintegrowanej) powinno się uwzględniać np. stopień zanieczyszczenia badanego obszaru, zmienność zawartości badanego składnika i jego charakterystykę geochemiczną oraz rozkład danych. Metoda wykorzystująca wykres pudełkowy (Box plot), prowadzi do eliminacji mniejszej liczby wartości odstających i zazwyczaj przy jej użyciu otrzymywany jest szerszy zakres tła. To właśnie ta metoda powinna być stosowana przede wszystkim do obszarów charakteryzujących się niewielkim zanieczyszczeniem oraz znacznym zróżnicowaniem zawartości pierwiastków.

INTRODUCTION

The term "geochemical background (GB)" is one of the key notions in researching the condition of the environment. Currently, every sample collected for geochemical tests contains an amount of element or a chemical compound, which originates from a natural source (geogenic, biological) or an anthropogenic source. The problem of distinguishing between natural levels of a given substance and those resulting from human activity is particularly important, especially in determining the level of environment pollution. Although numerous articles have addressed this subject, there are a lot of questions which remain unresolved. Also, the methodology for establishing GB seems to remain open for discussion (Matschullat et al. 2000; Reimann, Garrett 2005; Gałuszka, Migaszewski 2011).

Owing to the vast scale of changes that presently occur in the environment as well as the complexity of processes which influence element levels in the soil, establishing the GB is important not only when related to hazardous elements, which pose a threat to life and health, but also those elements whose presence in the soil, and thus in the trophic chain, is necessary for the correct functioning of organisms in a given ecosystem.

This paper presents the analysis of micro and macro-element levels in soils. Four sets of soil samples differentiated by the degree of human pressure and the sampling area were used in the study. The following hypotheses were formulated in the initial phase of the study:

1) selecting a single universal method applicable for establishing the background in different sets of data is impossible;
2) the pollution of the area and variability of the bedrock on which the soil has been formed have a major impact when selecting statistical method;
3) considerable differences of soil's chemical composition, resulting from heterogeneity of parent rock forming the bedrock, may hinder the interpretation of results and suggest anthropogenic nature of the anomalies.

Geochemical background – terminology

The term "geochemical background" was first used in research related to finding mineral deposits as early as in the 1960s. In prospecting geochemistry it was construed as *the normal abundance of an element in barren earth material* (Hawkes, Webb 1962). This definition has been introduced to differentiate the "normal" concentration of a given element and an anomaly, what would indicate a presence of mineral deposits. However, the definition is ambiguous as it contains the word "normal", which may be interpreted in different ways (Reimann, Garrett 2005). In the following years, the term GB was introduced in environmental sciences. Confusion was caused by the introduction of a group of new terms which are unjustifiably considered synonyms of GB, including: *natural background*, *ambient background*, *anthropogenic background*, *naturally occurring background*, *pedogeochemical background*, *baseline*, *threshold value* (Baize, Sterckeman 2001; Appleton et al. 2008). Still, the major purpose of using the term GB is to differentiate the natural quantity of a given substance in the environment from the quantity originating from human activity (Gałuszka 2006). A major difference found in the definitions published is that they either include or exclude the so called scattered pollution sources. Terms such as: *ambient background*, *anthropogenic background* and *area background* may be considered synonymous – they all cover the natural content of a substance as well as anthropogenic pollutants, which are hard to identify (Gałuszka 2007). This approach was followed in the European standard for establishing the background values of soils issued in 2011: *Soil quality – Guidance on the determination of background values* (PN-EN-ISO 19258:2011). The definitions presented therein include the influence of natural geological and soil processes but also scattered pollution sources. The standard mentioned above also contains the notion *pedogeochemical content*, which applies to a natural content of substances in the soil, however excluding any signs of human impact. Some researchers believe that background values should only include the natural

content of substances in the environment (Matschullat et al. 2000). Gałuszka (2006) noticed that in this approach the background has not been presented as a specific value or a range of values.

Methods for establishing geochemical background

As with the definitions, there is no agreement over the best and most widely accepted method for establishing the value of GB. Most often, three group of methods for establishing the background value are distinguished: indirect, direct and integrated methods. An important factor effecting GB is the geological structure, so it is essential to learn about it as well. Moreover, the complexity of geochemical processes occurring in the environment should be taken into consideration during background evaluation (Gałuszka 2007). It is also worth noting that establishing the actual natural content of substances in the environment is practically impossible. We may only talk about a "theoretical" natural content. Reimann and Garrett (2005) emphasise that a single value cannot be regarded as GB. Thus, a given method should allow for establishing a natural content range, including temporal and spatial variability of the background.

Direct methods

When using the direct methods, GB is established by analysing samples collected from unpolluted areas. This approach is implemented in two ways. Test samples are materials dating back to times before the 19th century e.g. archive plants from herbaria, ice cores, deep lacustrine or fluvial sediments. Contemporary samples collected from areas that are considered free from anthropogenic impact can also be subject to the analysis (Baize, Sterckeman 2001; Gałuszka 2007). In direct methods, also known as geochemical methods, a calculated arithmetic mean value (Av_A), geometric mean value (Av_G) or a median (Me) for the data set analysed is adopted as the background value. This method allows for obtaining only a single constant value, which may be interpreted only as a hypothetical natural content of a given substance in the environment, without considering its temporal and spatial variability (Gałuszka 2007). Geochemical methods are criticised for their high degree of subjectivity in evaluating how "unpolluted" a given area is. Also, considerable doubts accompany the assumption that in the pre-industrial era there was no pollution of the environment. Direct methods are also related to high

costs and a large amount of laboratory workload. Additional tests on materials originating from the "pre-industrial" era cause their partial destruction and therefore gradually deplete the resources. Also, such materials are not available in many areas (Gałuszka 2007).

Indirect methods

GB may also be established using statistical (indirect) methods, which to a large degree are based on the assumption that the natural content of elements and chemical compounds in the environment follows the normal or the log-normal distribution (Matschullat et al. 2000). Gałuszka (2006) also lists several other assumptions, which are the basis for this group of methods. First of all, it is assumed that the geochemical analysis of samples is free of critical errors and that there are no losses or no contamination of the analyte during the analysis. The samples analysed should also represent a fairly homogenous environment, characterised by identical lithological or climatic conditions. When using statistical methods, it is also important to have a relatively large number of samples. There is a range of methods, which are strongly based on the assumption of normal distribution of the analyte tested. One of such methods is the *iterative 2-sigma technique* (2-σ). This method is very simple and does not require additional tests, therefore it has numerous supporters.

However, many researchers believe that the indirect methods based on the assumption of normal distribution are faulty (Reimann et al. 2005). Supporters of this approach think that the methods which are not based on statistical assumptions are much more reliable. One of such methods is the [Me±2MAD] technique, which is similar to the [Av_A±2σ] approach. This method uses the median (Me) instead of the arithmetic mean (Av_A) and the MAD (*median absolute deviation*) coefficient, which replaces the standard deviation (σ). These statistics are more resistant to outliers. Outliers can also be eliminated by the use of a box plot based on specific parameters proposed by Tukey (1977).

Also the graphical methods, including the most popular Lepetier's method, relative cumulated frequency curves method and regression analysis method, have gained numerous supporters (Matschullat et al. 2000; Thi Thu Dung et al. 2013). Background evaluation based on advanced statistical methods such as principal component analysis (PCA) or multivariate analysis, is

preformed ever more frequently, too (Bing et al. 2011).

Statistical methods are often considered most reliable (Matschullat et al. 2000). However, Reimann and Filzmoser (1999) are critical of the majority of those methods, stating that the majority of environmental samples do not reveal signs of normal or log-normal distribution. So, using methods based on such assumptions when it is uncertain whether the samples actually conform to them may result in erroneous results. The aforementioned researchers even claim that such distributions are never found in geochemical and environmental research performed on a regional scale. On the other hand, normal or log-normal distribution may characterise samples from small areas, with uniform bedrock. If the data does not follow the normal distribution, using Av_A and σ for establishing the geochemical background is completely unjustified and may provide results that do not reflect the actual conditions. In such cases, Me or Av_G are better measures of data location, and MAD is a better measure of sample propagation.

Integrated methods

An integrated approach to establishing GB combines both, the indirect and direct methods. Samples for tests are collected in areas with low pollution levels such as national parks, nature reserves etc. It is also advisable to use samples of soil taken from fully developed forest ecosystems in this type of research, as they are considered most stable and balanced as regards the supply and release of energy. In this approach, samples collected from areas which are considered relatively unpolluted are the subject of statistical analysis, described in the previous chapter. This allows for considerable and purposeful limitation of the data scope, which yields better statistical results. When using statistical analyses (e.g. *iterative 2-sigma technique* or Me±2MAD method), the result obtained is the GB range, and the knowledge about the area used in the research makes this result even more reliable (Gałuszka 2007).

RESEARCH AREA and MATERIALS

The soil research was conducted in the Beskid Sądecki region. It is fully located within Magurska nappe, consisting of flysch formations, in the southernmost part of Outer Carpathians. This area has a complex geological structure characterised by a high lithological variability of deposits, which were also subject to numerous tectonic disturbances.

Great part of the Beskid Sądecki region is covered with forests, growing mostly in mountainous areas. Human activity can be observed only in the valleys, where the settlements are located. The region has numerous animate and inanimate nature sites.

The present paper uses the results of research conducted by Kicińska (2012) in the years 2006–2010 in
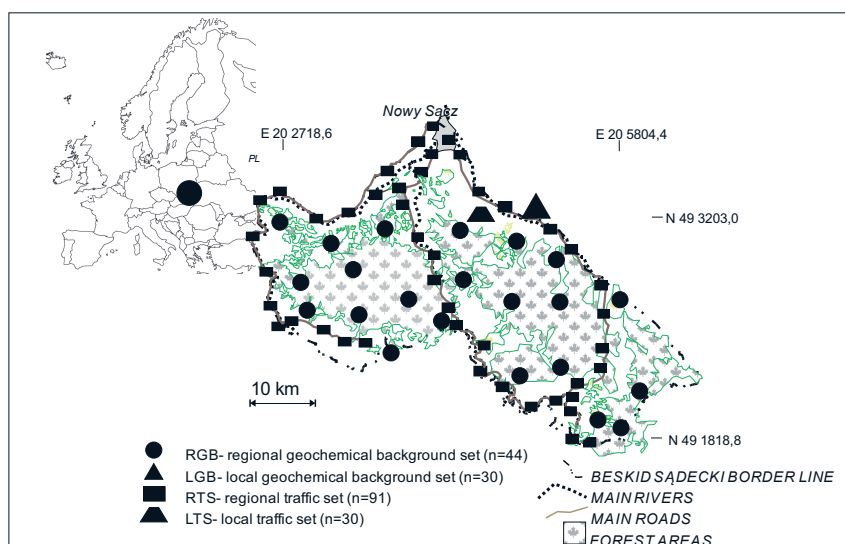

Fig. 1. Localization of sampling sites
Rys. 1. Miejsca poboru próbek

the Beskid Sądecki Mts. The research was planned so as to obtain several sets of data characterised by a different degree of human pressure and varying sampling areas. The first set – Regional Geochemical Background set (symbol: **RGB**, set size: n=44) contained samples collected in a fertile Carpathian beech forest area (*Dentario glandulosae-Fagetum*), located above 600 m ASL (Fig. 1). The entire Beskid Sądecki area was divided using a grid of 6×9 km rectangles. Next, a general sample (consisting of 5 samples from the corners and the centre of a 1×1 m square and a depth of 0–25 cm) was collected from each of them. Such ecosystems are considered very slightly affected by human activity, so they are regarded as the most appropriate for geochemical background research. The second set was the Local Geochemical Background set (symbol: **LGB**, n=30). The samples were obtained from a sod-covered field, from an area of 4×5 m, divided using a regular 1 m square grid. Due to the natural character of the sampling location and a considerable distance from emission sources, it is considered (similarly to the previous set), that the anthropogenic impact on soils is low.

The third, much larger set is the Regional Traffic Set (symbol **RTS**, n=91). It consists of soil samples taken near 3 communication routes with a total length of 181 km, passing through the Beskid Sądecki region. Related to this set is the fourth set, the Local Traffic Set (symbol: **LTS**), consisting of 30 samples collected from a 30m-long section located along one of the roads, taken in a straight line, 1 m apart from one another.

RESEARCH METHODOLOGY

The soil samples were obtained with a furrow method. Then, after the required preparation, they were subjected to extraction in 65% HNO_3 acid. After their break-down in a microwave oven, the levels of selected micro- and macro-elements were determined using the ICP-MS Elan 6100 PerkinElmer method. As many as 195 soil samples collected from an area of about 200 km^2 were tested. The levels of 6 elements were identified in soils from this area: Mg, Ca, Mn and Pb, Cd, As. The laboratory tests were performed at the Accredited Hydrogeochemistry Laboratory of the AGH-UST (certificate of accreditation PCA no. AB1050).

The first step was to establish the distribution and nature of the variables analysed. Basic descriptive statistics were used in this work to thoroughly analyse the data sets, including: Av_A, Av_G, Me, σ and the MAD coefficient. Also, the Shapiro-Wilk test was performed to verify the hypothesis on normal distribution of elements tested in individual data sets.

Based on the two regional data sets with low level of human pressure (RGB set and LGB set), the GB value was established, separately for each of the sets. Taking into account the distribution and nature of data, three methods were selected to establish GB:
1. mean and standard deviation method [$Av_A \pm 2\sigma$];
2. median and MAD coefficient method [$Me \pm 2MAD$];
3. box plot method;

Method 1 was used with full awareness of its assumptions and disadvantages. Nevertheless, owing to its popularity and simplicity, it was decided to compare the results obtained using this technique with the results obtained using techniques whose application is more recommended nowadays. Methods 2 and 3 were used because of their non-parametrical character. Despite the interesting results obtained using other graphical techniques, they have not been presented due to the small size of the data sets analysed.

The techniques applied belong to the group of statistical methods. However, owing to the characteristics of the area (forest ecosystems), we may say that in the case of the RGB and LGB sets the integrated approach was used.

Evaluation of pollution of the soils tested was performed based on selected geochemical indices such as: Geo-accumulation Index I_{geo}, Contamination Factor *CF* and Pollution Load Index *PLI*.

I_{geo} is a geo-accumulation indicator. It was calculated using the following formula:

$$I_{geo} = log_2 \frac{Ce}{1.5 * GB} \tag{1}$$

where: Ce – concentration of a given element in the sample, GB median value of geochemical background established using one of the 3 methods. GB is multiplied by a correction factor of 1.5 to account for variability of the background value. The I_{geo} values established for the samples analysed are classified into 7 pollution classes (Muller 1981). The I_{geo}0 value indicates no pollution, while I_{geo}6 indicates an extreme pollution level.

The *CF* value has been calculated as a quotient of average concentration of a given element in environ-

mental samples and the so called "pre-industrial" concentration of the element tested. In this instance, this is the mean value of GB established using one of the 3 methods. The calculated *CF* values were classified into 4 groups (Hakanson 1980): *CF*<1 – low *CF*; 1≤*CF*<3 – moderate *CF*; 3≤*CF*<6 – considerable *CF*; *CF*≤6 – very high *CF*.

PLI is an indicator used for an overall analysis of an area's pollution level. It is the *n*th root (depending on the number of chemicals taken into consideration) of the product of ConcF, which in turn are the quotients

of a content of a given chemical in the sample and the median value of GB. If *PLI*>1, samples are considered contaminated.

RESULTS AND DISCUSSION

The levels of the elements tested in the soil of the analysed area are largely diversified. The descriptive statistics parameters calculated vary greatly for each of the sets tested (Tab. 1). The results of the Shapiro-Wilk test indicate that with 95% confidence most of element

Table 1. Statistics for the analyzed data sets
Tabela 1. Parametry statystyczne zawartości pierwiastków dla analizowanych zbiorów

Element	Set	Min.	Max.	Av_A	Av_G	Me	σ	MAD
		mg/kg						
Mg	RGB	1278.7	8920.3	4460.2	4038.1	4838.5	1867.8	1673.5
	LGB	3698.7	4415.4	4029.1	4025.2	4022.1	180.1	125.6
	RTS	4645.8	26440.8	8806.7	8340.8	8266.6	3307.6	1438.4
	LTS	9489.5	17038.9	12945.2	12852.2	12936.2	1570.2	807.0
Ca	RGB	248.3	10337.4	2095.6	1183.5	938.0	2702.1	476.9
	LGB	1716.7	3266.3	2322.7	2294.5	2264.4	377.3	263.7
	RTS	2506.6	65867.1	19743.3	15755.5	16348.9	13908.8	6467.3
	LTS	22382.5	38505.0	31210.0	30929.3	31035.3	4201.6	2565.7
Mn	RGB	55.3	921.1	395.1	321.6	382.2	224.9	117.9
	LGB	563.9	742.3	652.1	650.9	653.3	40.4	25.2
	RTS	371.9	2923.6	721.2	636.6	575.3	519.8	93.3
	LTS	490.9	713.1	598.1	595.5	596.3	56.12	44.7
Pb	RGB	10.48	88.72	33.49	30.06	32.07	15.75	9.95
	LGB	24.45	29.59	26.81	26.78	26.54	1.31	0.86
	RTS	13.11	118.58	46.01	40.78	39.21	24.26	13.55
	LTS	37.25	105.70	54.64	52.82	52.40	15.41	10.25
As	RGB	3.53	81.54	8.78	6.14	5.50	14.50	1.14
	LGB	4.85	5.76	5.34	5.34	5.33	0.23	0.13
	RTS	2.86	37.49	5.43	5.03	4.91	3.88	0.62
	LTS	3.24	7.65	4.43	4.36	4.24	0.85	0.30
Cd	RGB	0.09	6.00	0.66	0.38	0.38	1.08	0.19
	LGB	0.38	0.63	0.50	0.50	0.49	0.07	0.05
	RTS	0.14	12.31	1.39	0.86	0.72	2.05	0.32
	LTS	1.05	9.04	1.58	1.39	1.26	1.43	0.19

Sets: RGB- Regional Geochemical Background set, LGB- Local Geochemical Background set, RTS- Regional Traffic Set, LTS- Local Traffic Set, Min. – minimum, Max. – maximum, Av_A – arithmetic average, Av_B – geometric average, Me – median, σ – standard deviation, MAD – median absolute deviation

Table 2. Parameter *p* determined by Shapiro-Wilk test
Tabela 2. Parametr *p* wyliczony za pomocą testu Shapiro-
-Wilka

Set	Mg	Ca	Mn	Pb	As	Cd
RGB	0.12*	0.00	0.01	0.00	0.00	0.00
LGB	0.84*	0.19*	0.94*	0.69*	0.59*	0.57*
RTS	0.00	0.00	0.00	0.00	0.00	0.00
LTS	0.95*	0.78*	0.94*	0.00	0.00	0.00

* statistically significant, at the significance level α= 0.05

levels in individual data sets do not fit the normal distribution (Tab. 2).

In the RGB set, only Mg follows the normal distribution. On the other hand, all elements in LGB set fit the normal distribution. In the case of samples collected in the vicinity of a road (RTS set), the hypothesis on normal distribution was rejected for all elements. Finally, in the LTS set the elements Mg, Ca and Mn follow the normal distribution.

Results obtained using the Shapiro-Wilk test are typical of geochemical research. Frequently, when testing a hypothesis on the normal distribution of variables on a small scale, there is no basis to reject it. However, when regional research is conducted, normal distribution is encountered very rarely (Reimann, Filzmoser 1999). In the case of the regional sets (RGB and RTS), the decisive factor for rejection of the normal distribution hypothesis was a large natural variability of element concentrations in the soil (the RGB set) resulting from a large sampling area and its varied geological structure, as well as different levels of soil pollution resulting from varying vehicle traffic intensity in individual locations (the RTS set).

**Geochemical background of soils
in the Beskid Sądecki region**

GB ranges for individual elements detected in samples, characterised by low human pressure (RGB and LGB), were established using the 3 statistical methods. When comparing the results it was found that (Fig. 2):
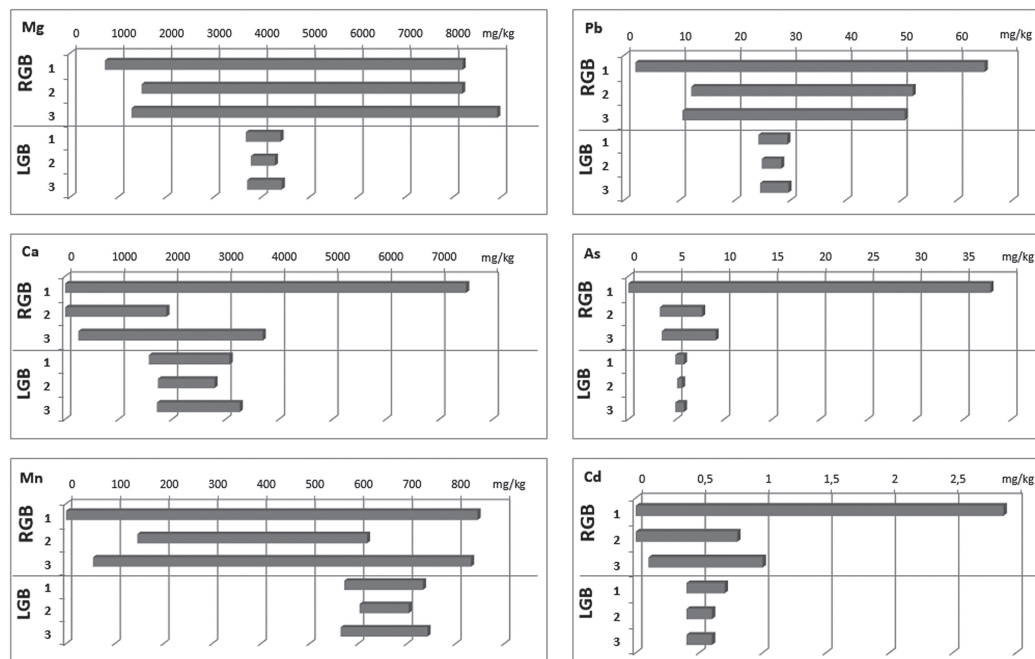

Fig. 2. Geochemical background ranges for Mg, Ca, Mn, Pb, As and Cd for the soils of Beskid Sadecki regions determined using three selected statistical methods: 1. Method [Av±2σ]; 2. Method [Me±2MAD]; 3. Method [Box plot]; RGB – regional set, LGB – local set

Rys. 2. Zakresy tła geochemicznego Mg, Ca, Mn, Pb, As i Cd wyznaczone dla gleb Beskidu Sądeckiego za pomocą: 1. Metody [Av±2σ]; 2. Metody [Me±2MAD]; 3. Metody [Box plot];

- as expected the regional set (RGB) has much wider background ranges (calculated using the 3 methods) than the local set (LGB) for all elements analysed;
- background ranges for the RGB set determined with the use of method [Av±2σ] are several times larger than the ranges established using methods [Me±2MAD] and [Box plot] e.g. for As, Ca and Cd;
- considerably smaller differences in background ranges were calculated for the local LGB set, which was probably the result of lower variability of the variables analysed and their normal distribution (Tab. 1 and 2). Also, in the case of normal distribution of the data, the discrepancies between background ranges established using methods [Av±2σ] and [Me±2MAD] are considerably smaller. If the data does not follow the normal distribution, technique [Av±2σ] is not recommended due to excessive impact of outliers, which may result in errors in establishing the geochemical background.
- the narrowest interval was obtained using method [Me±2MAD]. Background calculated using this method has much narrower ranges compared to the use of method [Av±2σ] based on [Av±2σ]. This demonstrates a very strong impact of outliers, which considerably alter the background ranges in method [Av±2σ]. The Me value and the MAD coefficient are more resistant to outliers. Therefore, when establishing the background range the extreme values are eliminated more effectively;
- similarly to methods [Av±2σ] and [Me±2MAD], background ranges established using method [Box plot] are different for different sets (Fig. 2). However, in this case, elimination of outliers is much more effective compared to methods [Av±2σ] and [Me±2MAD].

Nevertheless, the important question remains: which method and what type of data should commonly be used for establishing GB? The concentration of individual elements in the soil is affected by a great number of complex processes and factors. What is of great importance is the concentration of elements in the parent rock on which the soil was formed, as it largely determines the content of elements in soils. Based on the element concentrations published by Kicińska (2012, 2016), who performed detailed tests of bedrock on which the soils analysed had been formed, GB ranges were compared with the presence of elements in the bedrock. In the case of the RGB set (Fig. 3), method [Av±2σ] was not used due to the fact that only Mg followed the normal distribution (Tab. 2). For the LGB set, the range of elements present in the rock was compared to the background established using all [Box plot] methods (Fig. 4).

GB ranges in soil established for the RGB set using methods [Me±2MAD] and [Box plot] have a much smaller range than the quantities detected in the bedrock (Fig. 3). In the case of the latter, the levels were much higher, especially when it comes to macro-elements e.g. Mg, whose content in the bedrock was 3 times higher than the upper background limit established using both the methods, or Ca, whose content was even more than 10 times higher. In the case of micro-elements (As, Cd and Pb), these differences were less profound. This may stem from the fact that certain amount of elements in the bedrock are not broken down and reach the eluvium layer i.e. the soil. Their considerable quantities are built into the structure of minerals, bound strongly and not subject to transformations.

Also, slight shifts of the lower background limit in relation to minimum concentration of elements in the bedrock can be observed. Lower content of elements found in the soil compared to their concentration in the bedrock might result from the weathering processes e.g. leaching of chemicals caused by infiltrating waters and absorption of micro- and macro-elements by root systems of plants.

In the case of the regional set, the range interval might have been affected by the diversification of geological structure and the range of individual lithological formations. The narrowest GB range for this set was calculated using [Me±2MAD] method. Based on the data analysed, method [Box plot] should be considered the best for establishing the background in regional research Based on the data analysed, [Box plot] method is preferable for establishing the background in regional research as it effectively eliminates outliers and most adequately reflects the element content ranges in rocks.

Considerably larger differences between the GB ranges established using the 3 methods and the concentration of elements in the bedrock are observed in the case of the local set (LGB) (Fig. 4). The bedrock contains much more Ca and Mn and much less As (which, despite the initially assumed low human pressure, indi-
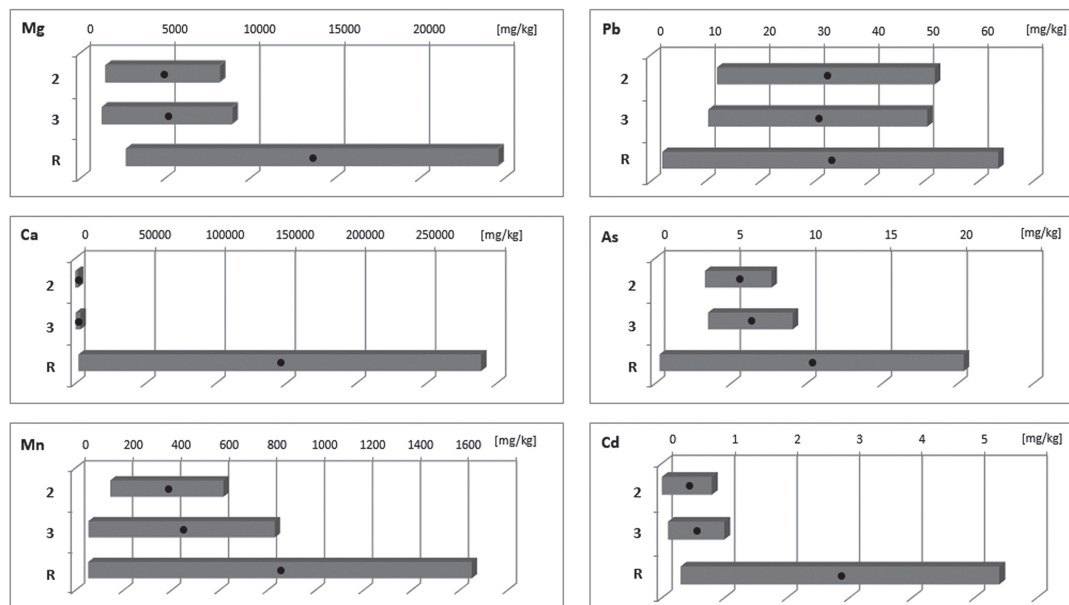
Fig. 3. Geochemical background ranges for Mg, Ca, Mn, Pb, As and Cd in the RGB set and content of elements in rocks of the Beskid Sądecki region determined using the three selected statistical methods: 2. Method Me±2MAD]; 3. Method [Box plot]; R – element content in bedrock; • median

Rys. 3. Zakresy tła geochemicznego Mg, Ca, Mn, Pb, As i Cd w glebach zbioru regionalnego (RGB) oraz ich zawartości w skałach Beskid Sądecki wyznaczone: 2. Metodą [Me±2MAD]; 3. Metodą [Box plot]; R – zawartość pierwiastków w skałach; • mediana
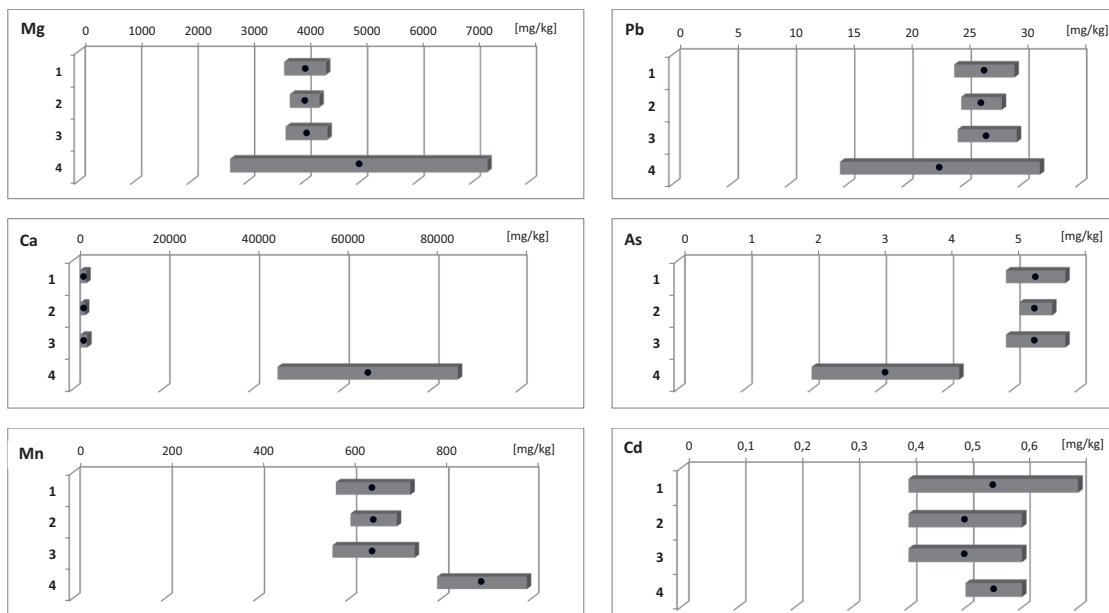

Fig. 4. Geochemical background ranges for Mg, Ca, Mn, Pb, As and Cd for LGB set and content of elements in rocks of *Beloveza Formation*, determined using the three selected statistical methods: 1. Method [Av±2σ]; 2. Method [Me±2MAD]; 3. Method [Box plot]; R – content in the bedrocks; • median

Rys. 4. Zakresy tła geochemicznego Mg, Ca, Mn, Pb, As i Cd w glebach zbioru lokalnego (LGB) oraz ich zawartość w skałach formacji beloweskiej, wyznaczone: 1. Metodą [Av±2σ]; 2. Metodą [Me±2MAD]; 3. Metodą [Box plot]; R – zawartość pierwiastków w skałach; • mediana

cates its anthropogenic origin). The narrowest GB ranges were obtained using [Me±2MAD] method, while the widest were obtained using method [Av±2σ].

The obtained results clearly demonstrate that the use of statistical methods does not yield unequivocal results. In practice, the background range values obtained using the various techniques vary considerably. Method [Av±2σ] is the most susceptible to outliers, therefore it is the least suitable for establishing GB in regional research. Every single outlier can causes multi-fold shift of the limit value. Still, this method may be successfully applied in local research, where the elements analysed follow the normal distribution. When it comes to the other methods ([Me±2MAD] and [Box plot]), method [Me±2MAD] allows for obtaining slightly narrower GB ranges than method [Box plot], both in their upper and lower limits.

Evaluation of soil pollution level

The evaluation of soil pollution level may be conducted in two ways, i.e. by comparing the results obtained with: (*1*) limit values (upper limits) set for different soil types and (*2*) published in legal acts and GB established for a set characterised by low human pressure.

In accordance with the Regulation (2016), soil is considered polluted if the upper limit is exceeded by at least one substance indicated in the document. In the case of the rementioned regulation, when establishing soil pollution, a certain group of soils (determined on the basis of their usage) for which the upper limits were assigned, is investigated. When evaluating the pollution of soils from the 4 sets analysed, it was found that the permissible levels of Pb, As and Cd were not exceeded in the RTS, LTS and LGB sets, and Pb and Cd levels were not exceeded in the RGB set (Tab. 3). The permissible As content (50 mg/kg) was exceeded in 4.5% of the samples from the RGB set.

When evaluating the pollution of soils with regard to their genetic type and using identical permissible values which did not take into account the variability resulting from differences in land usage (Kabata-Pendias et al. 1993), it was found that in the case of the RGB and LGB sets the majority of samples displayed a natural content of Pb (93% of the samples tested in RGB, 100% in LGB) (Tab. 4). For the sets of samples collected along communication routes (RTS and LTS)

Table 3. Classification of soil by quality standards
Tabela 3. Ocena stopnia zanieczyszczenia gleb wg klasyfikacji IUNG

Set	The degree of soil pollution*	Upper limit* in mg/kg (percent of the samples included in each grade soils)	
		Pb	**Cd**
RGB	0	50 (93.2)	0.5 (66.00)
	I	100 (6.8)	1.5 (27.0)
	II	250 (0)	3 (0)
	III	1000 (0)	5 (4.5)
	IV	5000 (0)	10 (2.5)
	V	>5000 (0)	>10 (0)
LGB	0	50 (100)	0.5 (100)
	I	100 (0)	1.5 (0)
	II–V	>250 (0)	3 (0)
RTS	0	50 (64.8)	0.5 (36.3)
	I	100 (30.8)	1.5 (47.2)
	II	250 (4.4)	3 (8.8)
	III	1000 (0)	5 (3.3)
	IV	5000 (0)	10 (3.3)
	V	>5000 (0)	>10 (1.1)
LTS	0	50 (43.3)	0.5 (0)
	I	100 (53.3)	1.5 (73.4)
	II	250 (3.3)	3 (23.3)
	III	1000 (0)	5 (0)
	IV	5000 (0)	10 (3.3)
	V	>5000 (0)	>10 (0)

* According to the recommendations of the IUNG Pulawy (Kabata Pendias et al. 1993)

this condition was met respectively by 65% and 43% of the samples analysed. An increased Pb content was found in 7% of the samples from the RGB set, 31% of the samples from the RTS set and 53% of the samples from the LTS set. Slight pollution with Pb was found in 4% of samples from the RTS set and 3% from the LTS set. In the case of the second analysed element – Cd, the situation was slightly worse. Natural Cd content was found in 66% of the samples from the RGB set, 100% of the samples from the LGB set and 36% of the sam-

Table 4. Classification of soil pollution
Tabela 4. Ocena stopnia zanieczyszczenia gleb wg Rozporządzenia Ministra Środowiska (2016)

Set	A group of land according the type of use*		Element		
			Pb	As	Cd
RGB	III wooded land	permissible content of substances in soils [mg/kg]	500	50	10
		the percentage of samples exceeding the allowable limit [%]	0	4.5	0
LGB	II meadow	permissible content of substances in soils [mg/kg]	250	20	3
		the percentage of samples exceeding the allowable limit [%]	0	0	0
RTS	IV communication areas	permissible content of substances in soils [mg/kg]	600	100	15
		the percentage of samples exceeding the allowable limit [%]	0	0	0
LTS	IV communication areas	permissible content of substances in soils [mg/kg]	600	100	15
		the percentage of samples exceeding the allowable limit [%]	0	0	0

* According to Regulation… (2016)

ples from the RTS set. An increased Cd content was found in 27% of the samples from the RGB set, 47% of the samples from the RTS set and 73% of the samples from the LTS set. Slight pollution with Cd was found in 9% of the samples from the RTS set and 23% of the samples from the LTS set. Moderately pollution with Cd was found in 4.5% of the samples from the RGB set and 3% of the samples from the RTS set. Slightly more than 3% of the samples from the RTS and LTS sets displayed a Cd level of over 10 mg/kg, which would classify them as very strongly pollution with this element. In 1% of samples collected in RTS set, the concentration of Cd exceeded 10 mg/kg, which classifies them as extremely highly polluted with this element.

In order to conduct an evaluation of soil pollution including GB, the following pollution indicators were calculated for the sets analysed: I_{geo}, *CF* and *PLI*. These calculations use the median value, determined with the use of method [Box plot] for the RGB set (see Tab. 5 *GB* values), as the GB value. Because of the non-toxic character of Ca and Mg, we may only talk about pollution when it comes to As, Cd, Pb and Mn. In the case of Ca and Mg such analysis is unnecessary.

The calculated mean I_{geo} values for the elements analysed in the samples from the sets characterised by low human pressure (RGB and LGB) vary in the range between −1.25 and −0.59 (Tab. 5). According to the classification presented by Muller (1981), $I_{geo} \leq 0$ means no pollution (class 0). In the case of samples from the sets subject to greater human pressure (RTS and LTS)

Table 5. *Igeo* values for the data sets
Tabela 5. Wartości *Igeo* dla analizowanych zbiorów

Set	Igeo			
	Mn	Pb	As	Cd
RGB	−1.05	−0.61	−0.62	−1.25
LGB	−0.59	−0.60	−0.60	−0.59
RTS	−0.11	−0.19	−0.88	−0.18
LTS	−0.16	0.20	−1.12	0.63
GB regional value*				
[mg/kg]	443	31	6.3	0.6

* GB- geochemical background – assumed midpoints value calculated by method no. 3 – Box plot

no pollution was found with regard to As and Mn, as well as Cd and Pb in the RTS set. In the LTS set, Cd and Pb values fall into an unpolluted/moderately polluted level. The *CF* coefficient for the elements was analysed and ranged from 0.70 to 2.63 (Tab. 6). Thus, the pollution level is low (*CF*<1) in the case of: Mn in RGB, Pb in LGB and As in almost all sets (with the exception of RGB, where moderate pollution occurs). In the remaining cases, moderate pollution was found (1≤*CF*<3). For each of the sets, the total pollution load indicator *PLI* was calculated taking into consideration the levels of macro-elements (Mg, Ca) and micro-elements (Mn, Pb, As and Cd) and all elements analysed jointly (Tab. 6). The result was *PLI*<1 (which denotes

Table 6. *CF* and *PLI* values for the data sets
Tabela 6. Wartości *CF* i *PLI* dla analizowanych zbiorów

Set	CF*				PLI*		
	Mn	**Pb**	**As**	**Cd**	**Macroelements (Mg, Ca)**	**Microelements (Mn, Pb, As, Cd)**	**All elements**
RGB	0.89	1.09	1.39	1.09	0.96	1.10	1.06
LGB	1.00	0.99	0.99	1.01	0.96	1.00	0.98
RTS	1.56	1.47	0.87	2.14	3.84	1.44	1.99
LTS	1.35	1.79	0.70	2.63	6.34	1.45	2.37

* the average values was calculated for the whole set

no pollution) in the case of macro-elements in sets RGB and LGB, while in sets RTS and LTS *PLI* was the highest (3.84 and 6.34 respectively, denoting increased levels compared to the average concentration in soils of the region). *PLI* values for micro-elements were as follows: RGB 1.10, LGB 1.00, RTS 1.44 and LTS 1.45. Based on the calculated indicators, the analysed area can be considered unpolluted or very lightly polluted (areas neighbouring communication routes), mainly with Cd and Pb.

CONCLUSIONS

Three selected methods: [Av±2σ], [Me±2MAD] and Tukey Box Plot were applied to establish the GB of soils from the Beskid Sądecki geographical region in southern Poland. The ranges of GB calculated for 4 sets of data, with a different degree of human pressure and sampling area, showed considerable differences. Based on the analysis it was confirmed that under no circumstances should the [Av±2σ] method be used in for data with skewed distribution and a high pollution level.

Analysing numerous sets of data and the selected methods also led to a conclusion that selecting a single universal method for establishing the background range is impossible. When selecting a method (direct, indirect or integrated), the pollution level of a given area, variability of tested substance concentration, its geochemical characteristics and data distribution etc. should be taken into consideration. Establishing the background value intervals, based on data characterised by considerable anthropogenic impact, seems difficult when using simple statistical methods. If the data set is affected

by a relatively stable stress factor, the concentration of substances analysed increases, which results in raised values for all statistics upon which the background establishing methods are mostly based.

Based on the research, the following conclusions were formulated:

1. Using different methods, for establishing GB results in markedly different results, so the selection of the appropriate method is the most important phase in researching GB.
2. Due to the complexity of the factors affecting the selection of a method for establishing GB, it is not possible to choose a single universal method. Nevertheless, it is possible to list a set of criteria that should influence this decision.
3. Box plot method used for establishing GB results in eliminating fewer outliers and obtaining a broader background range interval. This method seems appropriate in the case of areas characterised by light pollution and considerable diversity of elements in the soil. This stems from the fact that there is no need to eliminate a large number of samples, which is required when using other methods presented in this work.
4. Using the method based on the MAD coefficient in the case of areas with natural diversity of the analyte levels investigated, may result in establishing a background range that is too narrow, as for Mn in RGB set.
5. Geochemical background values established on the basis of regional research instead of clarke values reflect the actual condition of soils more accurately and are more suitable for evaluation of their pollution.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

Appleton J.D., Rawlings B.G., Thornton I. 2008. National-scale estimation of potentially harmful element ambient background concentrations in topsoil using parent material classified soil: stream–sediment relationships. *Applied Geochemistry* 23: 2596–2611.

Baize D., Sterckeman T. 2001. Of the necessity of knowledge of the natural pedogeochemical background content in the evaluation of the contamination of soils by trace elements. *The Science of the Total Environment* 264: 127–139.

Bing H., Wu Y., Sun Z., Yao S. 2011. Historical trends of heavy metal contamination and their sources in lacustrine sediment from Xijiu Lake, Taihu Lake Catchment, China. *Journal of Environmental Sciences* 23(10);1671–1678.

Gałuszka A. 2006. Geochemical Background of Selected Trace Elements in Mosses Pleurozium schreberi (Brid.) Mitt. And Hylocomium splendens (Hedw.) B.S.G. from Wigierski National Park. *Polish Journal of Environmental Studies*, 15 (2a): 72–77.

Gałuszka A. 2007. A review of geochemical background concepts and an example using data from Poland. *Environmental Geology*, 52: 861–870.

Gałuszka A., Migaszewski Z.M. 2011. Geochemical background – an environmental perspective. *Mineralogia*, 42(1): 7–17.

Hakanson L. 1980. An ecological risk index for aqatic pollution control. A sedimentological approach. *Water Research*, 14, 8: 975–1001.

Hawkes H., Webb J.E. 1962. Geochemistry in mineral exploration. *Harper*, NY, 415.

Kabata-Pendias, Piotrowska M., Witek T. 1993. *Ocena stopnia zanieczyszczenia gleb i roślin metalami ciężkimi i siarką. Ramowe wytyczne dla rolnictwa*. IUNG, Puławy.

Kicińska A. 2012. *Zróżnicowanie geochemiczne obszaru Beskidu Sądeckiego i jego wpływ na obieg wybranych pierwiastków*. Wydawnictwo AGH, Kraków. [In Polish]

Kicińska A. 2016. *Assessment of the road traffic impact on accumulation of selected elements in soils developed on Krynica and Bystrica subunit (Magura Nappe, Polish Outer Carpathians)*, Carpathian J. of Earth and Environmental Science, 11, 1: 245–254.

Matschullat J., Ottenstein R., Reimann C. 2000. Geochemical background – can we calculate it? *Environmental Geology* 39 (9).

Muller G. 1981. Die Schwermetallbelstung der sedimente des Neckars und Seiner Nebenflusse: eine Bestandsanfunahme. *Chemiker-Zeitung*, 105: 156–164.

PN-ENISO19258:2011. *Soil quality-guidance on the determination of background values.*

Regulation of Minister of Environment on 1 September 2016 how to conduct the assessment of contamination of the earth surface (Dz. U. 2016 poz. 1395).

Reimann C., Filzmoser P., Garrett R.G. 2005. Background and threshold: critical comparison of methods of determination. *Science of the Total Environment* 346: 1–16.

Reimann C., Filzmoser P. 1999. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environmental Geology* 39(9).

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

Thi Thu Dung T., Cappuyns V., Swennen R., Ky Phung N. 2013. From geochemical background determination to pollution assessment of heavy metals in sediments and soils. *Reviews in Environmental Science and Biotechnology,* 12: 335–353.

Tukey J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley, Boston.