Research Article

Rimantė Zinkutė*, Ričardas Taraškevičius, Margarita Jankauskaitė, Vaidotas Kazakauskas, Žilvinas Stankevičius

Influence of site-classification approach on geochemical background values

Abstract: This study of peri-urban minerogenic topsoil on glacigenic or post-glacial deposits shows the influence of the site-classification approach on the differentiated median background (DMB) values of major elements and the potentially harmful elements (PHEs) Ba, Cr, Cu, Mn, Ni, Pb and Zn. Composite samples from forests and meadows were taken in 25 sites, each of which had five sub-sites. A fraction of <2 mm was used to determine the organic matter by loss on ignition (LOI), grain size by laser diffraction and the elemental contents by X-ray fluorescence. The following five site-classification approaches are compared: geochemical (G), using relative median contents of Al, K, Ti; textural (T), according to mean percentages of clay-sized fraction (CLF) and silt fraction (SIF); lithological (L), based on soil parent material texture from the soil database; soil type (S), presented in the soil database; and parent material (P), generalising the underlying Quaternary deposits. Sites were classified into four level groups in which the DMB values were estimated after eliminating anomalies. The average ranks of three scores according to SIF, CLF, LOI, Al, K, Ti, Fe, Mg, Ca and S in the respective groups revealed the highest value for the G approach. It better eliminates the CLF and SIF influences on the median assessment indices of PHEs in sites.

1 Introduction

To estimate the level of soil contamination by potentially harmful chemical elements (PHEs), it is necessary to determine the normal background concentrations [1], often called ambient background concentrations [2,3], or simply background concentrations [4], which are the sum of two components: natural (from geological and pedological processes) and non-natural (from diffuse source input [DSI]). Although background is usually considered as a range [5], the background (B) value from this range is necessary if the on-site contamination level by PHE is expressed as a ratio of the actual on-site concentration to B value. This ratio is always found in formulas of different mono-elemental indices used for the assessment of topsoil pollution level and its associated risk [6–8]. These indices, as well as multi-elemental indices, e.g. those in ref. [6,9], inevitably depend on the proper choice of site-specific B value and the way in which it was estimated. Different mathematical methods are used for the estimation of B values [10]. Two groups can be distinguished among them: (i) methods without prediction, i.e. with constant B value for a group of sites and (ii) methods with prediction (using regression), with variable B (VB) value in sites. Taking into account that geochemical data sets are often heterogeneous, the aim of both method groups is to reduce the influence of natural factors on B values of PHEs.

Methods with prediction always require at least one variable predictor that reflects the variability of PHEs. But methods without prediction also often need discrete or continuous variable classifiers for distinguishing more homogeneous groups in a heterogeneous set of sites, i.e. background differentiation. Additional variables are unnecessary only if the cumulative frequency distribution plot of PHE is analysed. Both groups of methods

* **Corresponding author: Rimantė Zinkutė,** Laboratory of Geoenvironmental Research, Institute of Geology and Geography, Nature Research Centre, Akademijos Str. 2, Vilnius 08412, Lithuania, e-mail: rimante.zinkute@gamtc.lt
Ričardas Taraškevičius, Margarita Jankauskaitė: Laboratory of Geoenvironmental Research, Institute of Geology and Geography, Nature Research Centre, Akademijos Str. 2, Vilnius 08412, Lithuania
Vaidotas Kazakauskas: Laboratory of Quaternary Research, Institute of Geology and Geography, Nature Research Centre, Akademijos Str. 2, Vilnius 08412, Lithuania
Žilvinas Stankevičius: Department of Geology and Mineralogy, Vilnius University, M.K. Čiurlionio 21/27, Vilnius 03101, Lithuania

require elimination of different outliers. Methods without prediction based on various ways to remove the anomalous samples were analysed by Reimann et al. [5], and the Med ± 2MAD method (where Med is the median and MAD is the median absolute deviation) was recommended to reveal a relatively uncontaminated subset (usual concentrations or background range). Some researchers prefer to use the upper threshold background (TB) value to reveal soil contamination [5], while others more appreciate the median background (MB) value [2].

In this study, the phrase "background estimation approach" means the type(s) of variable(s) used either for prediction of VB values in sites or for classification of sites for estimation of MB values in each group, while the word "factor" means one type of variable. The approach is pure if it is based on one factor; otherwise, it is mixed. This meaning of "approach" is taken from the publication by Ander et al. [1], where the approaches for determining background that are based on "sampling and analysis of soil over different parent material groups" are considered as basic, while "association with particle size fractions" [2] and relationship with total Fe or Mn [11] are mentioned as alternatives. They are based on different natural factors: (i) soil parent material, (ii) soil texture and (iii) soil geochemical composition. All three are interrelated and the parent material is decisive: not only is soil texture linked with it [12], but soil geochemical composition is also associated [13]. The relationship between soil geochemistry (Al_2O_3) and soil texture [clay-sized fraction (CLF)] has also been demonstrated [14]. Background estimation approaches can also be based on other factors, e.g. soil type and land use [2]. But land use is often a natural anthropogenic factor, especially in urban areas. Still, natural anthropogenic factors are sometimes used, e.g. for classification of the UK into domains [1].

Valuable experiments with various background estimation approaches were done by Zhao et al. [2]: they used not only mixed approach (which is based on soil texture and organic matter) for determination of differentiated background values of seven PHEs but also a geochemical approach (regression equations between PHEs and one or two major elements). They also determined the non-differentiated background value of Pb using a probability graph. But there is a lack of comparison of MB values estimated using a pure approach for background differentiation, i.e. according to a single selected natural factor. It might be useful for choosing the optimal approach when there is lateral heterogeneity of a geochemical field but the number of soil sampling sites is limited (<150). Experiments with mixed, complex or nested

site-classification approaches using variables attributed to several natural or even natural anthropogenic factors can be achieved only by having a large soil data set, e.g. 5,691 samples in ref. [2].

The aim of this study of minerogenic topsoil from peri-urban sites in Vilnius was to compare the results of five natural factors of site classification into level groups (LGs), possibly reflecting clay-sized and fine fraction increase, to reveal the most suitable for estimation of differentiated median background (DMB) values of major elements and some PHEs and to discuss the usefulness of the values of assessment indices, i.e. contamination index (CI) and enrichment factor (EF). These indices were used in our earlier urban topsoil study [15]. The formula of CI is analogous to contamination factor [16] and pollution index (PI) [7]. In this study, total contamination index [9] will be called the additive index of accumulation (Za).

The tasks were as follows: (i) to classify topsoil sampling sites into the same number of LGs using five pure classification approaches based on qualitative information, i.e. either from Quaternary maps or from the soil database, or on quantitative data, i.e. either the contents of selected major elements or the percentage of clay-sized and silt fractions determined in samples; (ii) to analyse the correspondence between different site-classification approaches; (iii) by using the same Med ± 2MAD method, to estimate the DMB values of PHEs and major elements in LGs obtained using each approach as well as non-differentiated median background (NMB) values in a group of all sampling sites; (iv) to rank approaches according to the criteria that take into account parameters of the soil matrix, i.e. clay-sized and silt fractions and most important major elements; (v) to find the differences in the magnitude of assessment indices in sub-sites and sites when using each approach; and (vi) to reveal the problematic peri-urban soil sites for each PHE where its median assessment index (ASI_M) is elevated and discuss their usefulness and ambiguity of interpretation.

The rationale for the study was to demonstrate the expediency of background differentiation into LGs and that different site-classification approaches followed by the same background estimation method provide slightly different median background values and therefore different magnitudes of assessment indices. Our results support the quite natural hypothesis that a geochemical approach using the contents of selected major elements enables more substantiated background ranges and the DMB values to be revealed in comparison with other approaches, but future research is necessary to reveal the most suitable major elements for this aim.
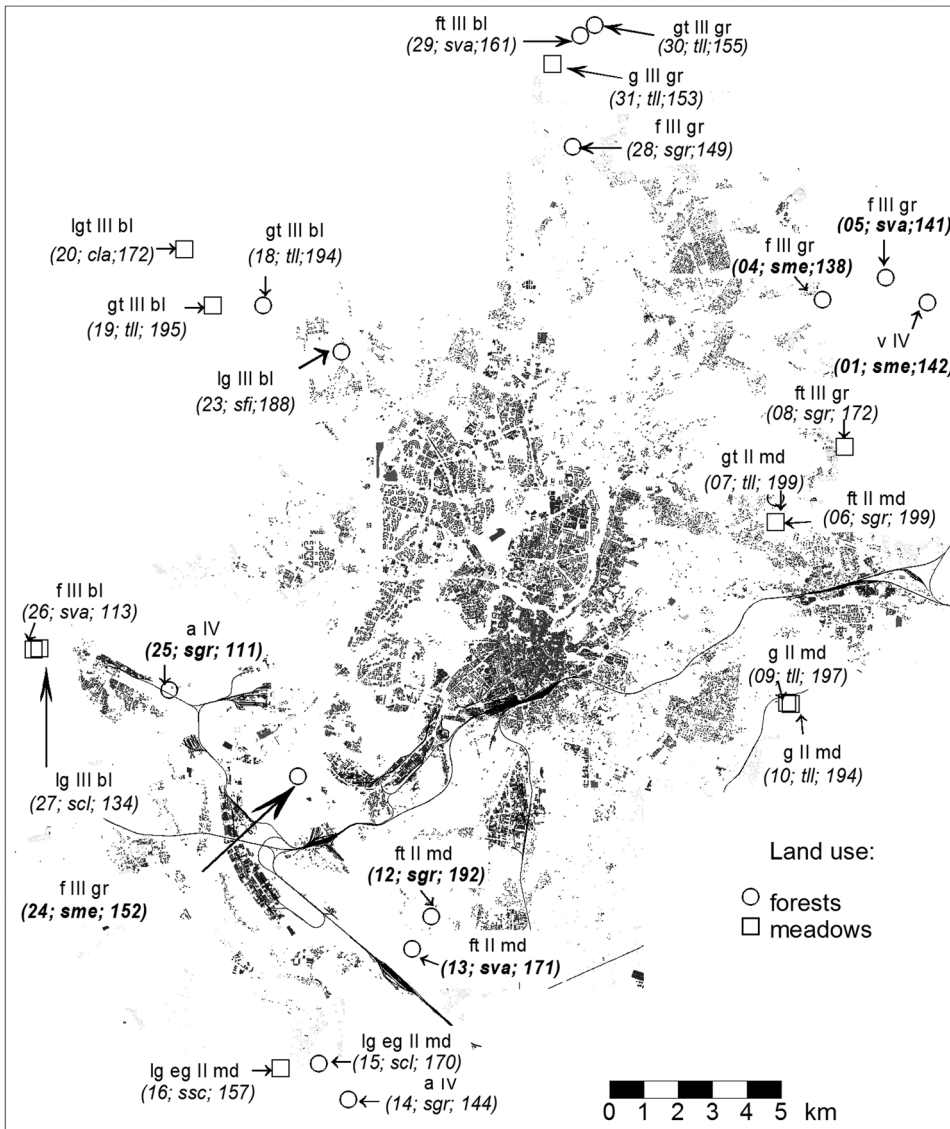
Figure 1: Location of soil sampling sites in peri-urban part of Vilnius and information about the underlying Quaternary deposits, relief elevation and land use. Information on Quaternary deposits was taken from the Lithuanian Geological Survey [17]. The index in the upper row indicates their origin and age. Origin: lg – proglacial glaciolacustrine, lgt – marginal glaciolacustrine, lg eg – englacial glaciolacustrine, g – basal till, gt – marginal till, f – proglacial glaciofluvial, ft – marginal glaciofluvial, a – alluvial, v – aeolian. Age: II md – Middle Pleistocene Medininkai, III gr – Late Pleistocene Grūda, III bl – Late Pleistocene Baltija and IV – Holocene. The lower row contains site ID, lithology of deposits and relief elevation (m). Lithology: cla – clay; tll – till; different types of sand, i.e. ssc – silty-clayey, scl – clayey, sfi – fine, sme – medium, sva – various and sgr – gravelly. The lower row is in boldface when no information about soil type and lithology (texture) is available in the soil database [18].

2 Material and methods

2.1 Sampling and information about sites in the study area

Soil sampling sites were located in the inner and outer peri-urban parts of Vilnius in various directions from the centre (Figure 1).

Although the sites were not in built-up areas, but rather in forests or meadows, diffuse source influence is inevitable as the sites are not far from roads or settlements. The total number of sites was 25, each of which included 5 sub-sites resulting in 125 samples. The minerogenic soil sample in each sub-site was composite, i.e. formed from five sub-samples taken from topsoil to a depth of 25–30 cm excluding the forest litter. The sub-samples were mixed in a plastic bucket in the field and

a portion was taken to the laboratory. Selection of the sampling sites was based on the underlying Quaternary deposits, taking into account their influence on soil texture and elemental composition. Later, using median values of LKS-94 coordinates measured by GPS in five sub-sites, information on these deposits was adjusted based on digital maps from the Geological Survey of Lithuania: (i) for outer peri-urban sites, on the Vilnius Quaternary geological map (scale 1:50,000) [17] and (ii) for inner peri-urban sites, on the Quaternary map of Vilnius (scale 1:16,000) compiled by R. Guobytė.

Information on soil taken from the Geoportal database [18] appeared to be incomplete, i.e. it was absent in the seven sites indicated in boldface type in Figure 1. In other sites, soil types and the respective first-level WRB soil reference groups [19] are presented in Table 1.

In these sites, there is also information on texture (lithological) classes for surface soil and for soil parent material (the latter is in Table 2), which is based on two classifications [20]: (i) the old one (used before 2000) by Kachinskii with 12 classes according to fraction <0.01 mm and (ii) the new one by FAO and modified in Lithuania in 1998 with 15 classes according to the percentages of sand (0.05–2 mm), silt (0.002–0.05 mm) and clay (<0.002 mm) fractions.

2.2 Sample preparation, determination of soil organic matter, texture and total element contents

All samples were sieved through a nylon sieve to <2 mm fractions, which the authors used for analysis: (i) estimation of soil organic matter percentage by the loss-on-ignition (LOI) method with combustion at 550°C [21]; (ii) grain-size measurement by the Fritsch Laser Particle Sizer Analysette 22 MicroTec Plus [22] with calculation of percentage of sand (63–2,000 μm) (SAF), silt (2–63 μm) (SIF) and clay-sized (<2 μm) (CLF) fractions [19]; and (iii) determination of total contents of chemical elements by energy-dispersive X-ray fluorescence Spectro Xepos equipment which uses Turboquant for the pressed pellets calibration method [23].

Additional sample preparation before chemical analysis was the following: (i) part of a <2-mm fraction was milled using an MM 400 mixer mill to obtain two 4 g portions of powdery material, each of which was homogenised with 0.9 g of Licowax binder and (ii) two parallel pressed pellets of 32 mm diameter were prepared from these mixtures using a PP15 press. Total contents of Al,

Table 1: Soil database information on soil types in sites and proposed LGs using S approach

Sites (lithology of underlying Quaternary deposits, see Figure 1)	Lithuanian codes [18] and FAO codes	Explanation	First-level WRB soil reference group [19]	Soil-type LGs
14(sgr), 23(sfl), 28(sgr), 29(sva)	SDp-b, ARh-eu	Eutri-Haplic Arenosols	Arenosols (AR)	S2
08(sgr)	SDp-n, ARh-dy	Dystri-Haplic Arenosols	Arenosols (AR)	S2
06(sgr), 09(tll), 15(scl), 19(tll), 31(tll)	Jlb2, ABe-o	Orthieutric Albeluvisols	Albeluvisols (AB), i.e. Retisols (RT)	S3
18(tll)	Jln2, ABd-o	Orthidystric Albeluvisols	Albeluvisols (AB) i.e. Retisols (RT)	S3
07(tll), 16(ssc), 20(cla), 30(tll)	IDp-t, LVh-or	Orthi-Haplic Luvisols	Luvisols (LV)	S4
10(tll)	IDg8-p, LVg-p-w-ha	Hapli-Epihypogleyic Luvisols	Luvisols (LV)	S4
26(sva), 27(scl)	IDp-t-e2, LVh-or-a2	Orthi-Haplic Endo-Albic Luvisols	Luvisols (LV)	S4

Table 2: Soil database information on soil parent material texture in sites, respective WRB texture classes and proposed LGs using L approach

Sites (lithology of underlying Quaternary deposits, see Figure 1)	Kachinskii classification (code, <0.01 mm, %)	FAO classification modified in Lithuania (code)	WRB texture class [19]	Lithological LGs
14(sgr), 23(sfi), 28(sgr), 29(sva)	Loose sand (s, 0–5%)	Sand (s)	Sand (S)	L2
08(sgr)	Coherent sand (s_1, 5–10%)	Sand (s)	Sand (S)	L2
07(tll)	Sandy loam (ps, 0–20%)	Sandy loam (ps)	Sandy loam (SL)	L3
26(sva), 30(tll)	Light loam (p, 20–30%)	Sandy loam (ps)	Sandy loam (SL)	L3
06(sgr), 15(scl), 16(ssc), 27(scl)	Light loam (p, 20–30%)	Sandy light loam (sp)	Sandy loam (SL)	L3
09(tll), 10(tll), 18(tll), 19(tll), 31(tll)	Medium loam (p_1, 30–40%)	Loam (p_1)	Loam (L)	L4
20(cla)	Heavy loam (p_2, 40–50%)	Clay loam (p_2)	Clay loam (CL)	L4

Ca, Fe, K, Mg, Na, P, S, Si, and Ti (ten major elements) and Ba, Cr, Cu, Mn, Ni, Pb, and Zn (seven PHEs) were used in this study. The joint uncertainty of their contents due to sample preparation and measurement was variable in the different samples, but the arrangement of the major elements according to the median value of the relative standard deviation in the paired pellets (%) divided by the square root of 2 (Med_{RSD}) was as follows: Fe(0.9) < K, Si (1.0) < Ca, Al(1.2) < Ti(1.5) < P(1.7) < Mg(2.1) < Na(2.2) < S (3.6). The sequence of PHEs according to Med_{RSD} was the following: Mn(1.0) < Zn(1.1) < Ba(2.0) < Pb(2.8) < Ni(3.4) < Cr(7.3) < Cu(8.5). The contents of the elements were adjusted using the re-calibration curves compiled by analysing samples of the "International soil-analytical exchange" program organised by Wageningen University.

2.3 Statistical data treatment

MB values were estimated by the Med ± 2MAD method in Microsoft Excel. STATISTICA 9 software was used to calculate the values of Spearman's rank correlation coefficients (r_S) and their significance, to perform hierarchical cluster analysis by Ward's method using $1 − r_S$ distance (for variables) or Euclidean distance (for sites), to test the non-parametric hypotheses using the Mann–Whitney U test and to perform Kruskal–Wallis analysis of variance and multiple comparisons of mean ranks for several groups (their number can be up to ten). Computation of *post hoc* comparisons of mean ranks of all pairs of groups (explained in STATISTICA 9 help) is based on the publication by Siegel and Castellan [24]. First, the z values for each comparison between groups u and v are computed according to the formula:

$$z_{u,v} = |R_u − R_v| / \sqrt{N \cdot (N + 1)/12 \cdot (1/n_u + 1/n_v)}$$

where R with subscripts are the average ranks for groups u and v, n_u and n_v are the numbers of observations in them and N is the total number of observations across all groups. Then p values (two-sided significance levels with a Bonferroni adjustment) associated with each comparison are computed as follows: $p = p(z) \cdot k \cdot (k − 1)$, where $p(z) = 2 \cdot (1 − \Pr(Z < z))$ and k is the total number of groups compared.

Ethical approval: The conducted research is not related to either human or animal use.

3 Results

3.1 Three site-classification approaches using available qualitative information

We selected three different natural factors (approaches) which due to their presumed indirect relationship with the increased clay-sized or fine fractions in topsoil can be used for classification of sites into LGs: P approach, i.e. according to parent material (underlying Quaternary deposits); S approach, i.e. according to soil type; and L approach, i.e. according to lithology (texture) of the soil parent material.

Four LGs for each site classification seemed reasonable, because to classify using S approach, there are three groups of soil types (Table 1) and one unidentified group of sites (bolded in Figure 1). Not only Arenosols (Table 1) but also unidentified soils are on sandy Quaternary deposits. We presumed that the latter group has the lowest level (LG1) and the code S0 was assigned to it (zero emphasises uncertainty), and attributed Arenosols to the group with the lowered level (LG2, code S2). Since Albeluvisols (Retisols) contain more sand and less clay in the upper horizons (A and E) than Luvisols [25], they were classified to the group with the

Table 3: Proposed LGs using P approach based on the underlying Quaternary deposits in sites

Lithology (see Figure 1)	Origin and age	Sites	LGs of P approach
cla, scl, ssc	Glaciolacustrine: marginal (lgt III bl), englacial (lg eg II md), part of proglacial (lg III bl)	20, 15, 16, 27	P4
tll	Till: basal or marginal (g II md, gt II md, g III gr, gt III gr, gt III bl)	09, 10, 07, 31, 30, 18, 19	P3
sfi, sgr, sva	Marginal glaciofluvial (ft II md, ft II gr, ft III bl), part of proglacial glaciolacustrine (lg III bl)	06, **12,**[a] **13,** 08, 29, 23	P2
sme, sva, sgr	Proglacial glaciofluvial (f III gr, f III bl), alluvial (a IV), aeolian (v IV)	**04, 05, 24,** 28, 26, 14, **25, 01**	P1

[a] Site ID is in boldface when no information is available about the soil type and lithology (texture).

elevated level (LG3, code S3), while the Luvisols were classified to the group with the highest level (LG4, code S4).

As the surface soil texture in sites appeared to be very uniform (mainly sandy loam), the texture of the soil parent material was used to classify by L approach (Table 2). Similar to the classification according to the S approach, the unidentified group sites were attributed to LG1 (code L0).

Great variability in Quaternary deposits near Vilnius is obvious in the sampling sites (Figure 1): there are 4 ages, 9 origin types and 15 combinations as well as 8 different lithologies. To classify sites into four LGs according to the P approach, we took into account mainly the generalised origin and lithology (Table 3): (i) glaciolacustrine (P4 or P2 depending on lithology), (ii) glacial (P3), and (iii) other, i.e. glaciofluvial, alluvial, aeolian (the latter are represented by sand, so more detailed information about origin and age was used to subdivide them, presuming that alluvial or aeolian deposits, also relatively younger proglacial glaciofluvial, should be attributed to LG1 with code P1, while relatively older marginal glaciofluvial deposits were attributed to LG2 with code P2).

3.2 Two site-classification approaches using quantitative geochemical and textural data

To distinguish LGs by geochemical approach (G approach), three major elements, Al, K and Ti, were used as variable classifiers based on our previous research [15,26] and presuming that (i) the contents of PHEs should be correlated with them and (ii) they should be correlated with the CLF and possibly fine fraction (<63 μm; [FIF]) percentage. If the closure effect of geochemical data is not taken into account, seven significantly ($p < 0.01$) inter-correlated major elements (ME_7) are observed, i.e. Al, K, Ti, Fe, Mg, Ca and S,

but the association of Al, K, and Ti is the closest (Table 4 and Figure 2). The content of each PHE has a significant ($p < 0.01$) correlation with all ME_7 members, but the association with Al, K, Ti is relatively greater, because (i) the contents of Cr, Ba and Mn are most of all correlated namely with Al, K, Ti; (ii) the contents of Zn, Ni, Cu, Pb have relatively high correlation with at least one of these three major elements (Table 4).

Ignoring the closure effect of geochemical and grain-size data, the contents of all ME_7 members have a significant ($p < 0.01$) correlation with both CLF and silt fraction (SIF) as well as with FIF and LOI percentages (Table 4). Almost all ME_7 members except S have a higher correlation with CLF or SIF than with LOI; further, the correlation with SIF is higher than that with CLF. Our selection of only three variable classifiers was based on their grouping in the dendrogram (Figure 2).

Hence, the site classification into four LGs by G approach was based on the clustering way of Ward's method and the Euclidean distances between the relative median contents of Al, K and Ti (division was by overall median; Figure 3). Ward's method with city block distances, which was ranked higher [27], provided identical classification.

The significant correlation of all PHEs with both CLF and SIF was the basis for finding out by the same clustering way the LGs of the textural approach (T approach), according to the mean values of CLF and SIF (Figure 4).

3.3 Particle size fractions and LOI in LGs distinguished by different site-classification approaches

Four LGs distinguished according to each site-classification approach have analogous similarities according to

Table 4: Multiplied by 100 Spearman's correlation coefficients between variables

	Al	K	Ti	Fe	Mg	Ca	S	LOI[a]	P	Na	Si
Al		96[b]	**93**	**96**	**89**	**75**	**63**	**53**	16	19	**−94**
K	**96**		**95**	**92**	**89**	**75**	**65**	**51**	21	13	**−92**
Ti	**93**	**95**		**91**	**83**	**69**	**64**	**52**	**28**	18	**−87**
Fe	**96**	**92**	**91**		**92**	**80**	**66**	**58**	12	11	**−93**
Mg	**89**	**89**	**83**	**92**		**91**	**63**	**50**	2	11	**−92**
Ca	**75**	**75**	**69**	**80**	**91**		**59**	**41**	11	**25**	**−81**
S	**63**	**65**	**64**	**66**	**63**	**59**		**84**	**30**	13	**−72**
LOI	**53**	**51**	**52**	**58**	**50**	**41**	**84**		11	4	**−61**
P	16	21	**28**	12	2	11	**30**	11		4	**−15**
Na	19	13	18	11	11	**25**	13	4	4		**−9**
Si	**−94**	**−92**	**−87**	**−93**	**−92**	**−81**	**−72**	**−61**	**−15**	**−9**	
Cr	**93**	**94**	**94**	**93**	**88**	**77**	**66**	**52**	22	13	**−89**
Ba	**89**	**92**	**95**	**86**	**79**	**66**	**58**	**49**	21	**31**	**−81**
Mn	**68**	**69**	**76**	**66**	**57**	**50**	**45**	**43**	**35**	**34**	**−63**
Zn	**89**	**86**	**86**	**92**	**85**	**75**	**72**	**63**	27	10	**−88**
Ni	**95**	**92**	**90**	**98**	**93**	**79**	**67**	**57**	8	6	**−92**
Cu	**83**	**83**	**79**	**87**	**87**	**79**	**63**	**50**	16	4	**−82**
Pb	**55**	**51**	**58**	**54**	**41**	**30**	**60**	**71**	18	**36**	**−54**
CLF	**90**	**89**	**87**	**89**	**84**	**66**	**45**	**36**	10	3	**−82**
SIF	**93**	**93**	**96**	**93**	**84**	**73**	**67**	**58**	25	22	**−88**
FIF	**95**	**94**	**97**	**95**	**86**	**74**	**65**	**57**	23	17	**−90**

[a] Abbreviations of variables: LOI – soil organic matter estimated by loss on ignition method (%), CLF – clay-sized fraction (<2 μm) (%), SIF – silt fraction (2–63 μm) (%), FIF – fine fraction (<63 μm) (%). [b] Significant (p < 0.01) coefficients are in boldface.
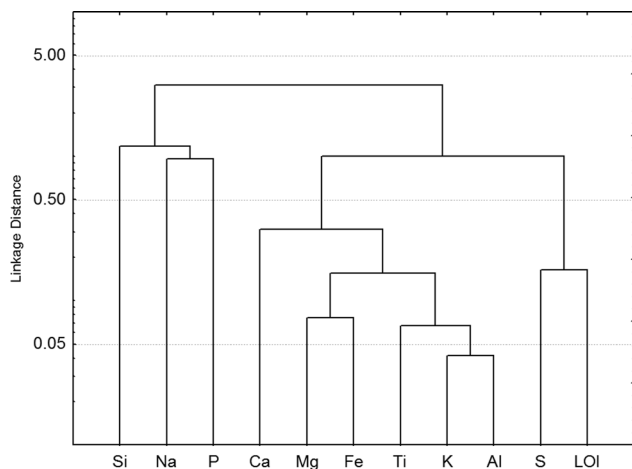
**Figure 2:** Cluster analysis dendrogram indicating association of Al, K and Ti.

the parameters of particle-size and LOI (Supplementary Table S1). The most important of these is the increase in CLF, SIF, FIF and LOI as the index of LGs grows (Figure 5).

All LG2 groups are sufficiently homogeneous, i.e. with relatively low coefficients of variation (CV < 33.3%) of SIF, CLF and LOI, but LGs of other levels have exceptions (Supplementary Table S1). The G and T approaches provide relatively lower CV values in LG1 groups and enable the greatest number of homogeneous LGs to be distinguished: only their LG4 groups are heterogeneous according to CLF. But other approaches also distinguish heterogeneous LG4 groups, because site 20 in them is anomalous. It is the only one with a clay loam texture and clay lithology of underlying deposits (Table 2); and it has much higher CLF in comparison with other sites attributed to LG4 groups. The U test and multiple comparisons test applied to grain-size variables showed different ability of site-classification approaches to separate adjacent LGs (Supplementary Table S2).

3.4 Median background values: non-differentiated and differentiated by each approach

The NMB values of ME7 and PHEs divide the range of DMB values into two intervals (Tables 5 and 6): those of LG1 and LG2 belong to the lower interval and those of LG3 and LG4 belong to the upper one (with only a few exceptions). In contrast, the DMB values of Si are higher in LG1 and LG2 but lower in LG3 and LG4.
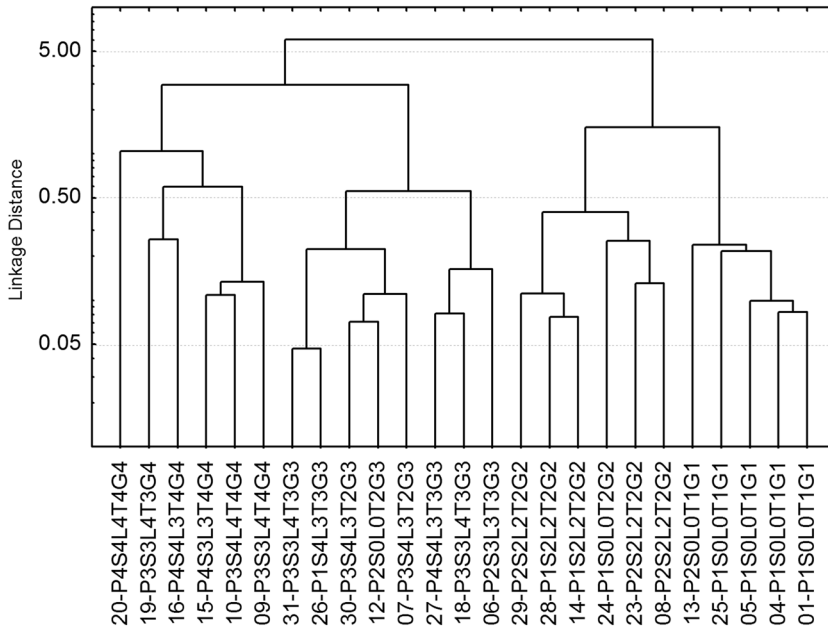
Figure 3: Four LGs of sites distinguished using the geochemical approach (G approach) for site classification and their comparison with those obtained by the other four site-classification approaches: T (textural), L (lithological), S (soil type), P (parent material). Indices after the letters indicate the presumed level of LGs: 1 or 0 – lowest (LG1), 2 – lowered (LG2), 3 – elevated (LG3) and 4 – highest (LG4).
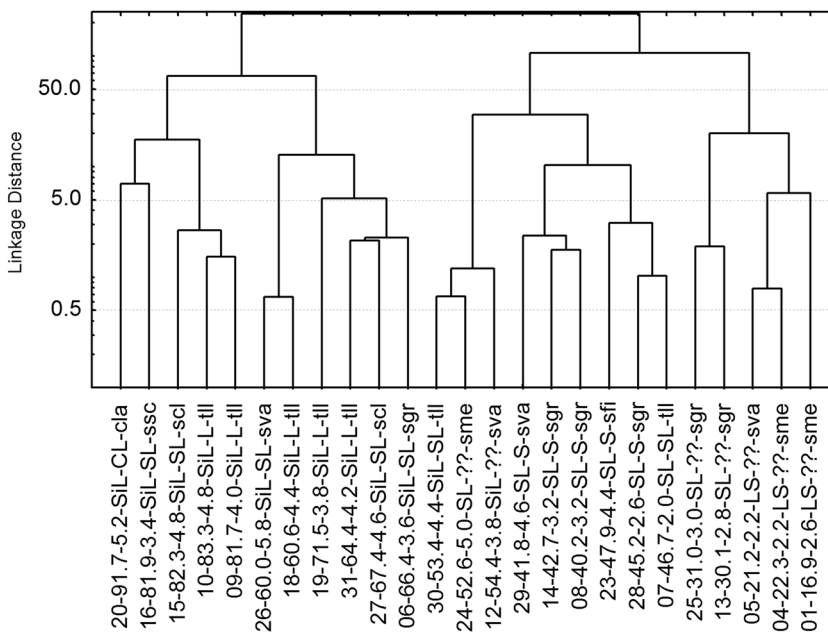
Figure 4: Four LGs of sites distinguished using T approach for site classification. The number after the first dash indicates the average FIF (%); after the second dash it indicates the median LOI (%) in the site. The codes after the third dash indicate the WRB textural class [19] according to average percentages of SAF, SIF and CLF determined by laser diffraction in sites: LS – loamy sand (number of sites $n = 3$), SL – sandy loam ($n = 9$) and SiL – silt loam ($n = 13$). The code after the fourth dash indicates WRB textural class [19] corresponding to that in the soil database for the soil parent material: ?? – no information ($n = 7$, L0), S – sand ($n = 5$, L2), SL – sandy loam ($n = 7$, L3), L – loam ($n = 5$, L4) and CL – clay loam ($n = 1$, L4). The code after the fifth dash indicates the lithology of the Quaternary deposits (Figure 1).
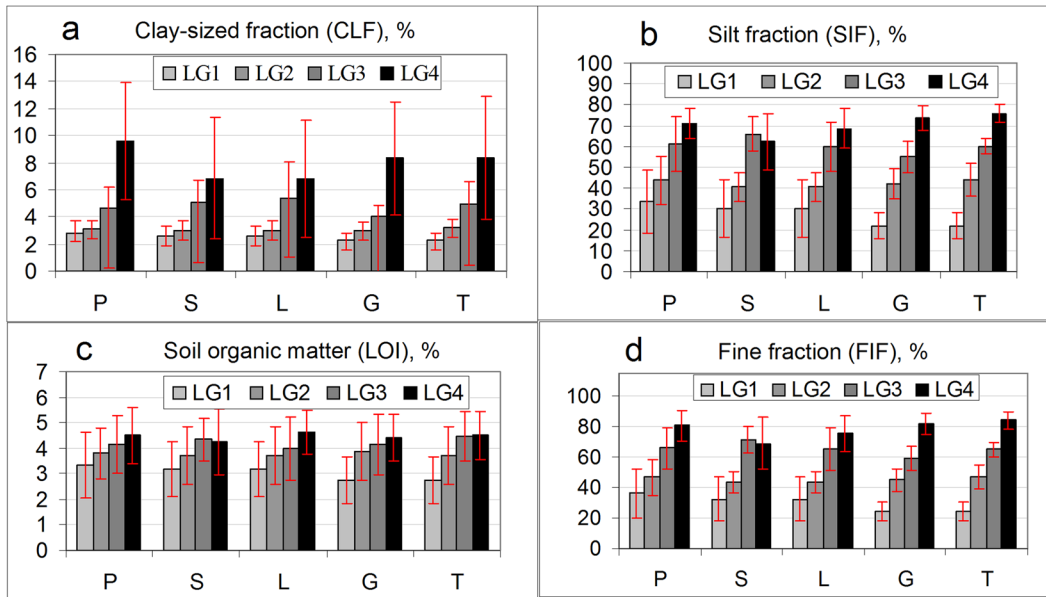
Figure 5: Increase in average percentages of influential grain-size variables (a, b and d) and soil organic matter (c) in four LGs distinguished by five site-classification approaches (see Figure 3). Error bars are based on standard deviation.

The DMB values of ME_7 and PHEs usually increase the higher the index of LGs. But only G approach has clear trends with the growth of the index of LGs: (i) increase of DMB values of all ME_7 and decrease of Si; (ii) increase of DMB values of all PHEs. Other approaches have exceptions, especially the S approach.

The DMB values of PHEs taken from the study of Zhao et al. [2] or from "Geochemical atlas of Lithuania" [29] (Table 6) represent different soil textures, so one may expect them to be somewhat similar to the DMB values obtained by L approach. Although their direct comparison is not substantiated due to different analytical methods and different numbers of LGs, some interesting details can be seen. All the DMD values of PHEs in sandy topsoil taken from "Geochemical atlas of Lithuania" [29], which are based on total contents like in our research, are higher than the DMB values in the L0 and L2 groups, which are also sandy. This is likely due to a different method, i.e. optical emission spectrophotometry, or different qualitative site classification. Although the DMB values in sandy topsoil taken from Zhao et al. [2] are based on partial aqua regia contents and should be lower than the total contents, the results in Table 6 contradict this. For Cu, Zn and Pb, they appear to be higher not only than in the L0 and L2 groups but also than in "Geochemical atlas" [29]; for Ni, the DMB value exceeds the respective values in the L0 and L2 groups; for Cr, it exceeds the value in L0. Possible reasons are differences in the underlying deposits and higher DSI in various types of land use in England and Wales.

3.5 Geochemical differences between adjacent LGs and homogeneity of LGs

Three findings concerning the differences between adjacent LGs in the contents of ME_7 and PHEs (after elimination of their anomalies) and in LOI (Supplementary Table S3) are as follows: (i) all approaches reveal many differences between LG3 and LG2; (ii) most approaches reveal differences between LG2 and LG1; and (iii) G approach is most effective to reveal differences between LG4 and LG3. Further, for Al, K and Ti, the G approach results in a minimum sum of robust coefficients of variation (RCV) across LGs (Supplementary Table S4).

3.6 Scoring and ranking of results of five site-classification approaches

The scoring and ranking were done by taking into account the following ten differentiating variables (Dvar): SIF, CLF, LOI and the contents of ME_7. Three scores were calculated based on the features of LGs obtained by each classification approach (Supplementary Table S5): (i) trend score reflecting the consecutive increase in the central values of Dvar in LGs with a growing index; (ii) separation score indicating the overall degree of significant differences in the values of Dvar between adjacent LGs; and (iii) heterogeneity score estimating the overall

Table 5: Non-differentiated and differentiated median background values (mg kg^{-1}) of major elements compared to non-differentiated values obtained during the Baltic Soil Survey

Gr. (n[a])	Al	K	Ti	Fe	Mg	Ca	S	Na	P	Si
Non-differentiated median background values										
BSS(748)[b]	48,300	15,605	2,601	17,065	3,437	9,934	253	8,977	820	3,31,400
BSS$_{LT}$(26)[c]	37,100	17,930	2,244	13,079	2,985	5,217	207	4,971	574	3,74,900
NMB(125)[d]	29,228	18,348	1,950	8,159	1,902	2,523	96	5,615	438	3,96,384
Differentiated median background values[e]										
G1(25)	18,015	12,999	904	4,401	495	615	7.3	4,080	545	4,23,586
T1(25)	18,015	12,999	904	4,401	495	615	7.3	4,080	545	4,23,586
L0(35)	18,680	13,534	933	4,392	520	1,228	38	4,856	538	4,23,349
S0(35)	18,680	13,534	933	4,392	520	1,228	38	4,856	538	4,23,349
P1(40)	23,267	14,690	948	4,877	677	1,389	32	5,593	408	4,18,416
G2(30)	25,916	15,491	1,524	7,402	1,447	2,072	89	5,996	384	4,00,052
T2(45)	26,092	17,093	1,647	7,546	1,639	2,178	88	5,996	368	3,97,715
L2(25)	25,908	15,312	1,486	7,323	1,447	2,020	80	6,006	381	4,00,696
S2(25)	25,908	15,312	1,486	7,323	1,447	2,020	80	6,006	381	4,00,696
P2(30)	25,225	16,522	1,688	7,455	1,405	2,149	98	5,573	593	3,96,707
G3(40)	30,899	20,387	2,092	9,441	2,564	2,847	117	5,657	400	3,87,656
T3(30)	33,052	21,213	2,191	10,374	2,860	3,627	151	5,081	401	3,78,703
L3(35)	33,854	21,227	2,211	10,405	2,931	3,356	117	5,596	436	3,79,317
S3(30)	34,632	22,017	2,427	10,399	2,742	3,407	142	5,727	511	3,79,603
P3(35)	32,415	21,715	2,142	9,698	2,687	3,491	132	6,152	440	3,80,585
G4(30)	41,063	23,826	3,182	13,611	3,568	3,674	149	5,331	505	3,62,984
T4(25)	40,912	23,798	3,197	12,233	2,756	3,770	152	5,927	527	3,62,984
L4(30)	37,325	23,339	2,565	11,084	2,766	3,695	168	5,870	548	3,77,085
S4(35)	35,470	21,516	2,109	12,132	3,727	3,653	124	5,578	438	3,66,938
P4(20)	43,080	23,642	2,790	16,047	7,649	3,595	151	5,054	449	3,54,178

[a] Number of topsoil samples. [b] In ten Baltic Soil Survey countries [28]. [c] In Lithuania during Baltic Soil Survey [28]. [d] In present research estimated by the Med ± 2MAD method based on the data obtained by the Turboquant calibration method and then adjusted using re-calibration curves. [e] In four LGs obtained in this research using five site-classification approaches (see Figure 3) estimated and adjusted as indicated above.

variability of Dvar within LGs. Three analogous scores were calculated in aforementioned way for seven PHEs.

Trend score is the number of Dvar with consecutive increase in average values (for SIF, CLF and LOI) or DMB values (for ME$_7$) in LGs with the growth of their indices. Its maximum possible value is 10 (number of Dvar). The arrangement of approaches according to decreasing trend scores G&P > T&L > S can be adjusted by adding the respective scores of PHEs; this provides the sequence G > P > T > L > S.

Separation score takes into account not only the number of Dvar (10) but also three pairs of adjacent LGs and three possible values of STR code-variable (statistical test results obtained by comparing adjacent LGs): STR = 2, if the difference is significant ($p < 0.05$) according to the multiple comparisons test; STR = 1, if it is insignificant ($p \geq$ 0.05) according to the multiple comparisons test but significant ($p < 0.05$) according to the U test; STR = 0, if the differences are insignificant ($p \geq 0.05$) according to the U

test. The maximum possible value for each Dvar is $2 \times 3 =$ 6, so the separation score of the approach is <60. The arrangement of approaches according to decreasing separation scores is G > T > P > L > S and is rather similar to that according to PHEs: T > G > L > S > P.

The heterogeneity score is the sum of the CV values of CLF, SIF and LOI in four LGs (Supplementary Table S1) plus the sum of the RCV values of ME$_7$ in all LGs after elimination of anomalies (Supplementary Table S4). The arrangement of approaches according to increasing heterogeneity score, i.e. T < G < L < S < P, as well as respective scores according to PHEs, shows that quantitative approaches provide greater overall homogeneity of LGs than qualitative and that P approach based on Quaternary maps gives the most heterogeneous LGs.

A higher trend rank or separation rank was given to the approach with a higher trend score or separation score, while a higher homogeneity rank was attributed to the approach with a lower heterogeneity score (Table 7).

Table 6: Non-differentiated and differentiated median background values (mg kg^{-1}) of PHEs in topsoil compared to values given in some other references

Gr. (n[a])	Ba	Cr	Cu	Mn	Ni	Pb	Zn
Non-differentiated median background values							
BSS(748)[b]	404	32	10	426	10	15	43
BSS$_{LT}$(26)[c]	344	32	8	349	9	13	33
NMB(125)[d]	320	18.7	4.30	384	5.48	13.8	23.0
Differentiated median background values							
G1(25)[e]	238	5.2	2.32	131	2.36	9.9	15.6
T1(25)[e]	238	5.2	2.32	131	2.36	9.9	15.6
L0(35)[e]	239	5.4	2.77	214	2.36	13.2	16.6
S0(35)[e]	239	5.4	2.77	214	2.36	13.2	16.6
P1(40)[e]	266	7.5	2.92	257	3.04	13.2	18.5
G2(30)[e]	293	14.8	3.68	325	4.80	13.8	21.6
T2(45)[e]	302	17.5	3.75	341	5.18	14.1	21.6
L2(25)[e]	289	14.2	3.70	359	4.78	12.9	20.5
S2(25)[e]	289	14.2	3.70	359	4.78	12.9	20.5
P2(30)[e]	281	17.1	3.70	370	5.03	13.8	21.3
G3(40)[e]	329	22.9	4.79	401	7.60	14.0	28.3
T3(30)[e]	338	27.2	5.93	381	8.51	13.9	30.3
L3(35)[e]	327	27.3	5.14	385	8.51	13.8	30.4
S3(30)[e]	356	30.8	5.31	455	8.21	14.4	30.2
P3(35)[e]	340	27.1	5.16	438	8.11	15.2	29.7
G4(30)[e]	396	35.7	7.21	466	11.8	16.2	31.2
T4(25)[e]	404	35.8	6.59	478	9.12	16.5	32.0
L4(30)[e]	370	30.7	5.80	468	8.36	16.2	30.2
S4(35)[e]	326	29.0	6.12	434	9.12	13.8	30.2
P4(20)[e]	386	39.1	7.65	439	13.8	14.8	36.5
Sandy topsoil (676)[f]	291	24.2	6.50	372	9.1	14.8	20.7
Sandy loamy topsoil (699)[f]	384	35.5	9.40	419	13.3	14.3	28.4
Loamy and clayey topsoil (678)[f]	417	42.5	11.3	423	16.9	14.6	32.9
Sandy topsoil (228)[g]		13	7.4		7.5	22	35
Coarse loamy topsoil (1,141)[g]		27	15		16	36	65
Coarse silty topsoil (182)[g]		39	19		22	42	78
Fine loamy topsoil (2,002)[g]		44	19		25	39	89
Fine silty topsoil (1,061)[g]		48	19		28	39	90
Clayey topsoil (483)[g]		59	23		38	35	106

[a] Number of topsoil samples. [b] In ten Baltic Soil Survey countries [28]. [c] In Lithuania during Baltic Soil Survey [28]. [d] In present research estimated and adjusted as indicated in Table 5. [e] In four LGs obtained in this research using five site-classification approaches (see Figure 3) estimated and adjusted as indicated above. [f] In "Geochemical atlas of Lithuania" [29]. [g] In topsoil samples from England and Wales [2].

The sequence of approaches according to descending average ranks G > T > P > L > S indicates that the S and L approaches can be treated as low-exactness and P approach as medium-exactness, meanwhile the G and T approaches are obviously high-exactness.

3.7 Similarities in site classifications

When different classification approaches are used, soil from the same site is not always attributed to groups of the same level, and the number of sites in them usually differs (Figure 3). But some LGs distinguished using different approaches (S0 and L0, S2 and L2 and G1 and T1) are identical according to the sites; therefore, the respective DMB values are also equal (Tables 5 and 6). Significant ($p < 0.01$) r_S values between the indices of LGs in sites show sufficient similarity in different classifications. According to descending r_S values, the pairs of approaches can be arranged as follows: G&T (0.94) > T&L (0.86) > G&L, L&S (0.85) > G&P (0.81) > G&S (0.80) > T&S (0.78) > T&P, L&P (0.75) > S&P (0.73).

Table 7: Scores determined using ten differentiating variables and seven PHEs, and evaluation of the quality of site-classification approaches according to the respective ranks

Scores and ranks	Site-classification approaches[a]				
	G	T	L	S	P
Trend scores of differentiating variables	10	9	9	5	10
Separation scores of differentiating variables	47	46	32	31	33
Heterogeneity scores of differentiating variables	685	684	874	888	945
(A) Trend ranks of differentiating variables	4.5	2.5	2.5	1	4.5
(B) Separation ranks of differentiating variables	5	4	2	1	3
(C) Homogeneity ranks of differentiating variables	4	5	3	2	1
Average rank of differentiating variables (from A, B, C)	4.50	3.83	2.50	1.33	2.83
Trend scores of PHEs	7	6	4	2	6
Separation scores of PHEs	34	35	27	26	25
Heterogeneity scores of PHEs	330	300	383	402	465
(D) Trend ranks of PHEs	5	3.5	2	1	3.5
(E) Separation ranks of PHEs	4	5	3	2	1
(F) Homogeneity ranks of PHEs	4	5	3	2	1
Average rank of PHEs (from D, E, F)	4.33	4.50	2.67	1.67	1.83

[a] Approaches are given in Figure 3.

3.8 Similarity in background ranges and grain-size variables in LGs of the same level

The similarity in background ranges of ME_7 and PHEs in LGs of the same level obtained using five site-classification approaches was tested by the multiple comparisons test (Supplementary Table S6). In almost all cases where significant ($p < 0.05$) differences were revealed, a small (sometimes even significant) increase was observed in at least one of the influential grain-size variables, i.e. clay, silt or fine fractions. The arrangement of LGs according to the increasing number of significant differences in the contents of each of 14 elements is LG2 < LG1 < LG3 < LG4 (Supplementary Table S7). The analogous arrangement is according to PHEs, but slightly different according to ME_7: LG2 < LG3 < LG1 < LG4. For PHEs, the number of significant differences in all LGs is higher in comparison to that of ME_7.

3.9 Influence of MB values on ASI_M values

When different MB values are used, the number of sub-sites with elevated (>1.3) values of CI or EF calculated using Al as a reference element is rather high and any consistent pattern is hardly visible, most probably due to accidental waste. But the ASI_M values in five sub-sites, i.e. the median CI (CI_M) or the median EF (EF_M), reduce accidental influence and more properly characterise sites (Table 8).

Since the study sites were selected to determine background, those with at least one PHE having $ASI_M > 1.3$ will be called problematic (CI-problematic or EF-problematic), because the reason for the higher values is unknown.

According to the number of CI-problematic sites when using NMB values, these sites occur more often for Ni, Cr, Zn and Cu than for Mn, Ba or Pb.

In comparison with the NMB values, the DMB values of Ni, Cr, Cu and Zn obtained by all approaches reduce the number of CI-problematic sites (except Cu using L approach). But the G approach gives the lowest number of CI-problematic sites, and T approach reveals a lower number of CI-problematic sites of Ni and Zn than the L, S or P approaches. There are fewer EF-problematic sites than CI-problematic sites, even when normalisation by Al is done using NMB values; this is especially obvious for Zn. Analogous reduction is also when the DMB values estimated by various approaches are used (for Ni using G approach the numbers are equal). Each approach for background differentiation of Ni, Cr, Cu and Zn is useful in comparison with NMB values, because it results in a lower number of EF-problematic sites. However, the advantage of G approach is that for Ni and Cr, the number of EF-problematic sites is the lowest.

As concerns the other three PHEs (Mn, Ba and Pb), not all approaches ensure that using respective DMB values, the number of their CI-problematic sites will be lower or at least equal to those obtained using NMB values. But the G approach seems to be the most suitable for their reduction or at least stabilisation.

Table 8: Influence of MB values on the number of sub-sites and sites with elevated assessment indices

MB values (approach[a])	Ni[b]	Cr	Zn	Cu	Mn	Ba	Pb	Ni	Cr	Cu	Zn	Mn	Ba	Pb
	Number of sub-sites with CI > 1.3							Number of sub-sites with EF > 1.3						
NMB	54	47	37	38	19	11	12	39	38	26	7	11	2	14
DMB(G)	23	30	9	26	22	0	15	9	18	17	3	28	0	12
DMB(T)	29	43	11	31	28	0	15	16	25	24	3	32	0	9
DMB(L)	38	40	16	34	21	12	15	28	26	22	3	20	1	9
DMB(S)	37	35	18	30	19	13	14	26	25	13	2	16	2	10
DMB(P)	36	38	17	34	28	3	11	24	28	20	3	15	3	13
MB values (approach)	Number of CI-problematic sites, i.e. where median CI (CI_M) exceeds 1.3							Number of EF-problematic sites, i.e. where median EF (EF_M) exceeds 1.3						
NMB	11	8	8	7	3	2	1	8	7	3	0	1	0	2
DMB(G)	2	4	1	4	3	0	1	2	0	2	0	5	0	0
DMB(T)	4	7	1	5	5	0	1	3	2	3	0	6	0	0
DMB(L)	7	6	3	7	4	3	2	5	2	3	0	3	0	2
DMB(S)	7	5	3	5	4	3	1	5	2	1	0	3	0	2
DMB(P)	7	7	3	6	6	0	0	4	5	1	0	2	0	1

[a] Approaches are given in Figure 3. [b] PHEs are arranged according to decreasing number of CI-problematic sites.

4 Discussion

4.1 Why were such aims and tasks chosen?

The regional geochemical variability in soil metal contents reflecting both natural and anthropogenic sources was demonstrated in e.g. ref. [16]. Due to the inevitable influence of natural factors, including parent material and complexity of Quaternary deposits near Vilnius described by R. Guobytė [30], we presumed that this area also has high local geochemical variability and therefore background differentiation is necessary. According to Wu et al. [31], it is better to use methods with prediction for on-site contamination assessment, because methods that estimate TB values are unsuitable. Since TB values have drawbacks [15,32], we used MB values but excluded methods based on prediction, because variable predictors in the study by Wu et al. [31] are attributed to different factors; further, this or other algorithms, e.g. ref. [33,34], are complicated. The word "approach" in our research means background differentiation factor but not background determination method, which, according to Matschullat et al. [10], can be either empirical or theoretical. Our research can be attributed to the integrated method that was acknowledged later [35] when samples are taken in relatively pristine (i.e. peri-urban) areas, but analytical results are subjected to statistical calculations (i.e. Med ± 2MAD method). Although even in strongly urbanised areas, the influence of geological background has been found [36], and Johnson and Demetriades [37] emphasise the role of parent material and soil types in urban geochemical mapping, the determination of B values in such areas is complicated there due to the extremely variable soil resulting from many anthropogenic factors such as site age, human impact, land use, building rubble and municipal waste [38]. Urban topsoil in Klaipėda [15] also demonstrated that eastern and western parts of the city differ in underlying Quaternary deposits and topsoil Al, K and Ti contents, but due to anthropogenic influence on the sandy part, it was impossible to predict the VB values of some PHEs or correctly assess their DMB values by the Med ± 2MAD method.

Peri-urban soil in Vilnius underlain by Quaternary glacigenic or post-glacial deposits is suitable for experiments with DMB values obtained using different site-classification approaches. It does not matter that some sites are within the city boundaries, because urban areas can also be understood as built-up [37]. The main thing is that the sites are far from the centre and industrial districts with many pollution sources [39,40]. The uppermost layer of the Quaternary cover differs not only according to origin but also according to age, i.e. it is at the boundary of two glaciations: the last one being Nemunas (with older Grūda stage and younger Baltija stage) and the penultimate one being Medininkai [41]. Hence, representation of all ages and origin types needs many more samples.

We excluded the approach based on soil texture and organic matter used in ref. [2,29], because it is based on different factors, i.e. soil with high organic matter content is not a textural class [19]. We studied only minerogenic soil, since peaty or organic soil needs separate research, as it may be greatly influenced by anthropogenic pollution, as shown in the study of peri-urban ombrotrophic bog records [42].

4.2 Insufficiency of NMB values and explanation of approach ranking

In comparison with the median elemental contents of agricultural topsoil from ten European countries [28], the study areas are characterised by 1.2 times higher NMB value of Si and therefore lower NMB values of almost all (except K) other major elements (Table 5) and all PHEs (Table 6).

A comparison with agricultural soil from Lithuania [28] also shows relatively lower NMB values of most elements related to clay or fine fractions. The sequence of elements according to descending ratios of medians S(2.2) > Ca(2.1) > Cu(1.9) > Cr(1.7) > Ni, Fe, Mg (1.6) > Zn(1.4) > P, Al (1.3) > Ti (1.2) > Ba(1.1) indicates the prevalence of light texture in the topsoil. But these NMB values are obviously unable to reveal increased DSI in sites that have low natural background level, as can be seen in the Ni example. Since its NMB value is in the middle of the range of DMB values (Table 6), all CI > 1.3 values of Ni appear to be in sites from LG3 or LG4, while low CI values are found in sites from LG1 or LG2 (Figure 6). This figure clearly demonstrates the advantage of differentiated background values in comparison with non-differentiated.

Although L and S approaches provide greater overall homogeneity of LGs of differentiating variables than P approach, they have the lowest trend and separation scores, especially S approach. Its lowest average rank is in agreement with the finding of Zhao et al. [2] that the major soil
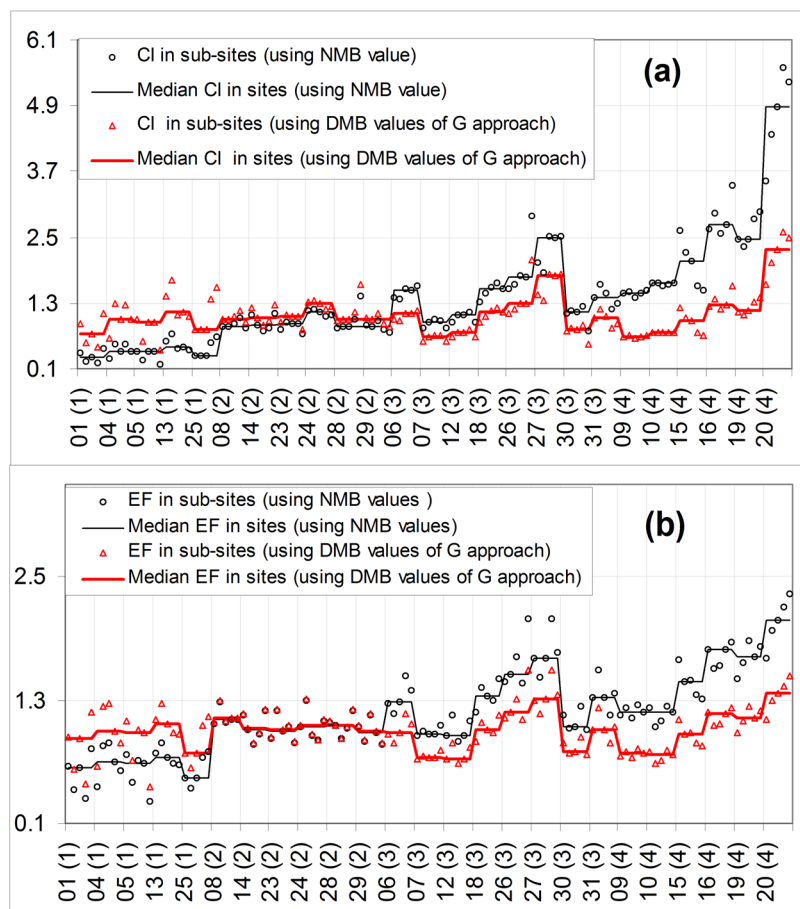
Figure 6: Values of contamination index (a) and enrichment factor (b) of Ni in sub-sites and sites using non-differentiated median background values and differentiated median background values obtained by G approach for classification of sites (LGs are in parentheses).

taxonomical group explains the lower percentage of variation in the contents of PHEs in comparison with soil texture.

The rather low average rank of L approach in comparison with T approach is explained not only by incomplete information from the soil database but also by the horizontal and vertical variability of the soil texture. Generalised information on soil texture in maps can mask its horizontal variability. The vertical variability of texture is obvious from the soil database: in topsoil, it is very uniform; while in soil parent material, it is more variable. In only 22% of study sites, the topsoil texture is similar to that of the parent material; while in about 61% of sites, the parent material is heavier. The vertical variability in soil texture in Luvisols from Northern Poland due to eluviation–illuviation (lessivage) processes was demonstrated by Switoniak et al. [43]. These researchers also found a rather uniform texture of the uppermost layer, and in most cases, a heavier texture in deeper horizons. Leaching of clay particles to illuvial horizons of Luvisols is without chemical degradation [44]. The vertical variability of soil texture is also obvious from discrepancies between the two texture classes, i.e. in topsoil determined by laser diffraction and in soil parent material indicated in the soil database (Figure 4 and Supplementary Table S8). Of course, the discrepancies can also be explained by systematic deviation in laser diffraction data in comparison with traditional pipette method data. The underestimation of clay and sand fractions in favour of silt fraction when using the laser diffraction method was demonstrated by Kun et al. [45]. Therefore, using this method, we obtained silt loam instead of clay or loam as indicated in the soil database and in seven sites where the soil texture class was absent from the soil database, but where sand was expected according to Quaternary deposits, it appeared to be loamy sand, sandy loam or silt loam. Despite this, laser diffraction and the pipette methods provided a similar estimation of sand, silt and clay fractions, since it was confirmed by significant inter-correlation of the respective results [45].

The higher average rank of P approach in comparison with low-exactness site-classification approaches is due to its highest trend rank and higher separation rank (Table 7). But it is close to the average rank of L approach, indicating rather good correspondence between the textural classes given for soil parent material and the lithology of the Quaternary deposits (Figure 4). Presumably, grain-size differences between soil parent material and underlying Quaternary deposits are lower than that between topsoil and soil parent material. Effective control of soil texture by underlying rocks has been demonstrated by Camara et al. [12]. The high heterogeneity score of P approach is most probably due to the existing differences between the topsoil and underlying Quaternary deposits as well as difficulties in classifying them into a few LGs.

An increasing trend seems to be an optional requirement for determination of DMB values, especially in countries with metalliferous mineralisation when classification is done using mixed approaches which include parent material as one of the factors [1]. But in Lithuania, this requirement is natural and useful, since a thick cover of glacigenic Quaternary deposits predetermines the decisive role of soil clay and fine fractions and a regular increase in the contents of most chemical elements. But the increasing trend and the separation of DMB values are not always clearly expressed even using textural approach, e.g. in six LGs of minerogenic soil presented by Zhao et al. [2] (Table 6). The following reasons are possible: (i) too many LGs distinguished during field assessment with small textural differences between some of them and (ii) the use of the aqua regia digestion. The increasing trend of DMB values in LGs distinguished according to minerogenic topsoil texture is better revealed for total contents of elements when the number of LGs is lower. For example, this can be seen in "Geochemical atlas of Lithuania" [29] for Ni, Cr, Cu, Zn, Mn, Ba (Table 6), as well as for Ti; but in comparison with the present research, the separation of adjacent LGs estimated by ratios of respective DMB values is lower. A possible reason is that information on soil texture in atlas was qualitative, i.e. taken from soil maps.

The advantage of T approach compared to L approach is the direct estimation of the percentage of topsoil fractions in samples by laser diffraction. Texture classes in topsoil samples also appeared to be rather uniform and unsuitable for classification into LGs, but clustering of continuous data enabled a higher ranked classification. The average rank of T approach is closer to G approach compared to P approach, because both G and T approaches are based on the topsoil measurements. The higher average rank of G approach compared to T approach is in agreement with the opinion of Dung et al. [46] that normalisation (the aim of which is similar to site classification) by geochemical methods is better than granulometric methods, because similar-in-texture samples can be different according to their mineralogical composition. The drawback of G approach is the difficulty in interpretation of LGs without other information. Qualitative information should never be ignored, even if it is inexact or incomplete. But to interpret the differences between G2 and G1, the grain-size and LOI data were necessary as they show a higher percentage of fine fraction and organic matter in G2 (see Figure 3 and labels in Figure 4).

4.3 Why was it useful to study similarities of site classifications?

The similarities of site classifications by using different approaches indicate that, in general, all of them are acceptable for background differentiation. But the high-exactness approaches (G and T) are preferable as they have higher average ranks according to the essential peculiarities of the soil matrix. Besides, these approaches are most similar. This finding enables us to presume that both will provide the most similar DMB values. The greatest similarity of the high-exactness approaches occurs because both are based on measurements of the same topsoil samples, i.e. determination of Al, K, Ti or clay and silt fractions. This similarity is in accordance with the correlation of Al with the CLF shown by Dreher et al. [14]. It is quite natural taking into account the variables used by researchers for normalisation [46], i.e. fractions <2 and <63 μm for granulometric normalisation and Al, K and Ti for geochemical normalisation.

A rather high similarity of low-exactness classification approaches (L and S) occurs because both are based on the same information source and their LG1 and LG2 contain the same sites. The lowest similarity between classifications based on soil type and Quaternary maps occurs because they use totally different sources of information and because there is a lack of soil-type information in the seven sites. The greater similarity of G approach to other approaches ($r_S \geq 0.80$) in comparison with T approach is in favour of G approach. L approach based on the lithology (texture) of the soil parent material is most similar to both quantitative approaches, but its lower trend rank and especially its separation rank indicate that this approach can be used with caution.

Greater similarity of classifications ensures a lower number of significant geochemical differences between LGs of the same level, especially according to ME_7 (Supplementary Tables S6 and S7). Most similar high-exactness classification approaches have no significant differences according to ME_7 in all LGs. Low-exactness approaches are also rather similar according to ME_7 in three LGs but differ in LG4 according to Ti. However, if different exactness approaches are compared, the number of significant differences increases. For ME_7, all significant differences are observed when comparing LGs distinguished by different exactness approaches. In these cases, many more significant differences are also observed in the DMB values of PHEs. Consequently, different exactness site-classification approaches will provide many more different values of assessment indices.

4.4 Should DMB values be adjusted by additional normalisation?

The finding that most PHEs have a lower number of EF-problematic than CI-problematic sites in peri-urban areas used for background estimation enables us to presume that normalisation by Al can sometimes be a useful tool for adjustment of DMB values. The background in each LG becomes slightly variable and if EF values better reduce the influence of soil grain size as the main natural factor, a comparison of sites according to those values can be more reasonable than according to CI values. But properly choosing between two assessment indices for PHEs, i.e. either EF or CI becomes extremely important, because it enables more reasonable revealing of the influence of other factors (natural or anthropogenic).

We selected Al as the candidate reference element for normalisation, since it was one of the variable classifiers. When choosing between CI and EF, we took into account two restrictions for calculation of EF: (i) PHE should be correlated with the reference element used for normalisation [47] and (ii) PHE should have higher variability than the reference element [8]. In the urban area of Klaipėda, there were only a few PHEs which could be normalised, because they did not satisfy at least one of the restrictions [15]. Meanwhile in peri-urban Vilnius with lower anthropogenic influence, all PHEs appeared to have significant ($p < 0.01$) correlation with Al, but the magnitude of r_S values was different. Four PHEs, i.e. Ni, Cr, Cu and Zn (in comparison with Mn or Pb), show a higher correlation with Al (Table 4); further, they have higher RCV (31.0, 50.1, 32.2, 27.0%) than Al (RCV = 19.5%). Although Mn in comparison with the four aforementioned PHEs has relatively lower r_S with Al, its RCV (21.4%) also slightly exceeds that of Al. Therefore, these five PHEs will be represented by EF values as assessment indices. The expediency of Ni normalisation is obvious, since the difference between EF values obtained by G approach and those obtained without background differentiation is lower than the difference in the respective CI values (Figure 6). But as the RCV values of Ba and Pb (13.3%) are lower than those of Al, they will be represented by CI values.

4.5 What do the values of ASI_M show?

The most important is that in peri-urban areas, these values are an additional tool to check the quality of site-

classification approaches. After site classification by G approach, neither clay nor silt fraction has any influence on ASI_M values of PHEs or values of the respective additive accumulation index Z_a, as is obvious from insignificant Spearman's correlation coefficients r_S (Supplementary Table S9). Other site-classification approaches do not ensure total independence of ASI_M values on fine fraction components. The low-exactness classification approaches retain the greatest number of significant r_S (5), the medium-exactness P approach preserves their lower number (2) and the high-exactness classification T approach has only one. Further, for Ni, Cr, Cu and Zn, the number of EF-problematic sites revealed by G approach (Table 8) is not only lower than or equal to EF-problematic sites revealed using NMB values, but it is also either the lowest or rather low in comparison with other approaches. This fact is an additional indication that G approach reduces the influence of fine fraction best of all. But Mn demonstrates the opposite feature. The number of its EF-problematic sites obtained using G approach is higher in comparison not only with the respective numbers of EF-problematic sites revealed by other approaches or without background differentiation but also with the number of its CI-problematic sites.

We expected that the optimal site-classification approach and therefore the more precise DMB values as well as the proper choice between CI or EF would help us to reveal, according to values of assessment indices, the peri-urban sites with elevated DSI. But it is not strange that we could not fulfil this task, since DSI "usually leads to sites that are relatively uniformly contaminated" [4]. In some sub-sites, high values of assessment index seem to show contamination by local accidental waste (despite that sample was a composite from five sub-samples), but a comparison of sites according to values of ASI_M confirms that DSI is rather uniformly distributed, except for some problematic sites (Supplementary Table S10).

Interpretation of the processes responsible for higher ASI_M values in these sites is complicated. The difficulties in distinguishing between natural and anthropogenic sources with the help of EF were mentioned by Reimann and De Caritat [8] who stated that these indices can be elevated due to many reasons. Background differentiation in peri-urban minerogenic soil followed by optional normalisation reduced the influence of fine fraction, but not of other factors. Although the ASI_M values obtained by G approach are optimal, they do not necessarily reveal anthropogenic influence, since they can be related to some natural processes. For example, two EF-problematic sites are obtained by G approach for Ni (27 and 20) and two for Cu (26 and 20). We can speculate that sites 26 and 27 indicate increased DSI

from the Vilnius–Kaunas highway (Figure 1), but we cannot state that site 20, which is far from pollution sources, is influenced by anthropogenic activity. Elevated values of assessment index most probably reflect the fact that even G approach provides insufficiently precise DMB values for this site. But the absence of EF-problematic sites of Cr and Zn, as well as of CI-problematic sites of Ba, enables us to presume that DSI is low in peri-urban territories. Some doubt arises only concerning one CI-problematic site of Pb (site 25, which is relatively close to the highway).

Another problem is the interpretation of the greatest number (5) of EF-problematic sites of Mn obtained by G approach (Supplementary Table S10). We can only say that although research in the central part of Vilnius [26] has shown that Mn correlates with both natural and anthropogenic factors, it is hardly believable that its EF-problematic sites in peri-urban areas indicate anthropogenic anomalies. First, the number of urban–industrial pollution sources of Mn is relatively low [48]. Second, concentration coefficients of Mn in dust from vents and stacks of some pollution sources in Vilnius are much lower than that of Ni, Zn, Cu and Cr [39]. Third, although Mn is a component of methylcyclopentadienyl manganese tricarbonyl, research along the highways in Canada [49] as well as near the Vilnius–Kaunas highway [50] does not show its significant accumulation, especially taking into account that the distance of the present study sites from roads is >15 m. The fact that Seleznev et al. [34] even use Mn as a conservative element for prediction of VB values of PHEs in urban puddle sediments is in favour of the natural origin of Mn, but revealing the natural reasons for elevated EF values of Mn needs separate research.

Since the G approach applied to sites with minerogenic topsoil reduces the influence not only of grain size but simultaneously of organic matter (Figure 5), very few other natural factors can be mentioned as possibly influencing ASI_M values, e.g. pedogenic processes, natural land use types or differences in relief elevation. Investigation of pedogenic processes is impossible when composite samples are considered. It is also complicated to reveal the influence of relief elevation, since its range is narrow (111–199 m) and most sites are at a great distance.

Hence, we tested only land use influence presuming that in our study this factor can be treated as natural, because we sampled only from forests and meadows. Publications that discuss land use influence present diagrams or tables which contain either the contents of PHEs [51] or the values of their assessment indices, e.g. PI [7]. Sometimes both the contents and EFs are given [52] and their diagrams differ. Our search, with the help of the U

test for significant differences, according to both median contents in sites and ASI_M values also demonstrated that the results are not identical. Significantly ($p < 0.01$) higher median contents of Cr, Ni, Cu, Zn and Ba in meadow sites in comparison with forest sites are obviously predetermined by a significantly higher percentage of clay, silt and fine fractions. Meanwhile, the median values of CI in sites (CI_M) obtained after background differentiation reduce the influence of grain size; therefore, the differences between CI_M values are mainly insignificant ($p \geq 0.5$), especially when DMB values are estimated using G approach for site classification (Supplementary Table S11). Similar results were obtained by Sun and Chen [7] when comparing the PI values of Cu, Ni, Pb, Zn and Cr in several land use types: only a few differences in PI values were significant ($p < 0.05$), but not those between forest and grass land use. The formulas of PI and CI are identical, but the difference is that the background differentiation used by Sun and Chen [7] was according to provinces and it is not clear whether they sufficiently reflect the natural peculiarities of parent rocks, topsoil texture and geochemical composition. But researchers' conclusion – that the impact of land use on heavy metal risks is insignificant – is in accordance with our results. Our results concerning CI_M of Ni, Cr, Cu, Zn (slightly higher in meadows) and Mn (slightly higher in forests) are similar to those based on data from "Geochemical atlas of Lithuania" [29]. Of course, when DMB values obtained by G approach are used and additional normalisation by Al is applied to five PHEs, a significantly higher value for EF of Cu in meadows appears, but the ratio of median EF_M value in meadows to respective value in forests is low (1.13).

4.6 Some problematic aspects for future investigations

Although our results seem logical, further experiments are necessary taking into account the problem that grain-size and geochemical data are compositional (closed). To treat such data, the log–ratio approach was introduced by Aitchison [53] who proposed "additive log–ratio" (arl) and "centered log–ratio" (crl) transformations. Also the "isometric log–ratio" (ilr) [54] was demonstrated as useful in robust principal component analysis [55]. Despite statisticians' huge input to the theoretical development of the compositional data analysis (CoDA) approach, according to Filzmoser et al. [56], this is "only recently being followed by a significant adoption" in geochemistry.

If the CoDA approach is selected for our data treatment, the inter-correlations of chemical elements and their correlation with grain-size fractions will be different. This will lead to new interpretation, changes in the most suitable major elements for site classification, another decision about expediency of normalisation and most suitable reference element and different values of median assessment indices. Possibly, their interpretation, especially as concerns Mn anomalies, will be easier. It may even be possible to distinguish sites with elevated DSI or to better reveal and explain different natural processes, provided that fine fraction influence is more effectively reduced. We presume, however, that when using the CoDA approach for our data treatment, the DMB values obtained after site classification using other major elements demonstrating a strong association with clay-sized or silt fractions will be quite similar to that obtained when classifying using relative median contents Al, K and Ti. This presumption can be done first of all because the latter site classification is similar to that using the relative median contents of all major elements ($r_S = 0.870$). Our present results also cannot be neglected, considering the findings of Fačevicova et al. [57] when studying the Devonian/Carboniferous boundary and comparing the results obtained by the standard approach with those obtained by the compositional approach and revealing that they have "synergic effects:" neither of them has prevalence and both can be further used in geochemical investigations. If so, the results of our research can be useful for future investigations of geochemical approaches based on inter-correlated major elements revealed using transformations that take into account the closure effect of data.

5 Conclusions

1. The present study demonstrates the necessity of site classification for evaluation of DMB values of PHEs (such as Ba, Cr, Cu, Mn, Ni, Pb and Zn) in peri-urban minerogenic topsoil on glacigenic or post-glacial deposits. This need is determined by a significant positive correlation of the contents of PHEs with percentages of clay and silt fractions, as well as with organic matter percentage and the contents of major elements (Al, K, Ti, Fe, Mg, Ca and S). According to the trend, separation and homogeneity ranks of these variables, the average ranks of each approach can be calculated.
2. Quantitative site-classification approaches, i.e. geochemical approach, which reveals LGs according to the relative

median contents of Al, K, Ti and textural approach, which classifies sites according to the mean percentages of clay-sized and silt fractions, have higher ranks in comparison with qualitative approaches that use information from Quaternary maps or the soil database. Due to incomplete information in the soil database, the average rank of the soil-type approach appeared to be the lowest, and the average rank of the soil parent material texture approach was slightly higher. The highest average rank of the afore-mentioned geochemical approach proves its priority for determination of DMB values of PHEs as well as of inter-correlated major elements.

3. The assessment indices of PHEs, e.g. contamination indices or EFs, may carry important information about peri-urban sites, and median assessment indices from several sub-sites are preferable for possible revealing of relative DSI in sites. The geochemical approach based on the relative median contents of Al, K and Ti eliminates the influence of clay and silt fractions on median assessment indices best of all, as can be seen from the absence of their correlation with these grain-size variables.

List of Abbreviations

PHEs	potentially harmful chemical elements
B values	background values
VB values	variable background values
TB values	upper threshold background values
MB values	median background values
LGs	level groups with different geochemical background
DMB values	differentiated median background values
NMB values	non-differentiated median background values
CI	contamination index
EF	enrichment factor
Za	additive accumulation index
LOI	percentage of organic matter determined by the loss on ignition method
SAF	sand fraction (63–2,000 μm) percentage
SIF	silt fraction (2–63 μm) percentage
CLF	clay-sized fraction (<2 μm) percentage
FIF	fine fraction (<63 μm) percentage
ME_7	seven inter-correlated major elements Al, K, Ti, Fe, Mg, Ca, S
Dvar	differentiating variables, i.e. SIF, CLF, LOI and the contents of ME_7
CI_M	median contamination index
EF_M	median enrichment factor
ASI_M	median assessment index (either CI_M or EF_M)
DSI	diffuse source input
PI	pollution index

References

[1] Ander EL, Johnson CC, Cave MR, Palumbo-Roe B, Nathanail CP, Lark RM. Methodology for the determination of normal background concentrations of contaminants in English soil. Sci Total Environ. 2013;454–455:604–18. doi: 10.1016/j.scitotenv.2013.03.005.

[2] Zhao FJ, McGrath SP, Merrington G. Estimates of ambient background concentrations of trace metals in soils for risk assessment. Environ Pollut. 2007;148:221–9. doi: 10.1016/j.envpol.2006.10.041.

[3] Appleton JD, Rawlins BG, Thornton I. National-scale estimation of potentially harmful ambient background concentrations in topsoil using parent material classified soil:stream-sediment relationships. Appl Geochem. 2008;23:2596–611. doi: 10.1016/j.scitotenv.2013.03.005.

[4] ISO 19258:2018(E). International Organisation for Standardisation – 19258 – Soil quality – Guidance on the determination of background values. Geneva: ISO, 2018. Available from: https://www.iso.org/obp/ui/#iso:std:iso:19258:ed-2:v1:en (accessed 15 January 2020).

[5] Reimann C, Filzmoser P, Garrett RG. Background and threshold: critical comparison of methods of determination. Sci Total Environ. 2005;346:1–16. doi: 10.1016/j.scitotenv.2004.11.023.

[6] El-Sherbiny MM, Ismail AI, El-Hefnawy ME. A preliminary assessment of potential ecological risk and soil contamination by heavy metals around a cement factory, western Saudi Arabia. Open Chem. 2019;17:671–84. doi: 10.1515/chem-2019-0059.

[7] Sun R, Chen L. Assessment of heavy metal pollution in topsoil around Beijing metropolis. PLoS One. 2016;11(5):1–13. doi: 10.1371/journal.pone.0155350.

[8] Reimann C, de Caritat P. Distinguishing between natural and anthropogenic sources for elements in the environment: regional geochemical surveys versus enrichment factors. Sci

Total Environ. 2005;337:91–107. doi: 10.1016/j.scitotenv.2004.06.011.

[9] Gregorauskienė V, Taraškevičius R, Kadūnas V, Radzevičius A, Zinkutė R. Geochemical characteristics of Lithuanian urban areas. In: Johnson CC, Demetriades A, Locutura J, Ottesen RT, editors. Mapping the chemical environment of urban areas. Chichester: John Wiley & Sons Ltd; 2011. p. 393–409.

[10] Matchullat J, Ottenstein R, Reimann C. Geochemical background – can we calculate it? Environ Geol. 2000;39(9):990–1000. doi: 10.1007/s002549900084.

[11] Hamon RE, McLaughlin MJ, Gilkes RJ, Rate AW, Zarcinas B, Robertson A, et al. Geochemical indices allow estimation of heavy metal background concentrations in soils. Global Biogeochem Cy. 2004;18(GB1014):1–6. doi: 10.1029/2003GB002063.

[12] Camara J, Gomez-Miguel V, Martin MA. Lithologic control on soil texture heterogeneity. Geoderma. 2017;287:157–63. doi: 10.1016/j.geoderma.2016.09.006.

[13] Rawlins BG, Webster R, Lister TR. The influence of parent material on topsoil geochemistry in Eastern England. Earth Surf Processes. 2003;28:1389–409. doi: 10.1002/esp.507.

[14] Dreher GB, Follmer LR, Zhang Y. Progress report on the chemical composition of soils in Illinois: Cores 1 through 10. Open-File Series 2002-2. Champaign: Illinois State Geological Survey; 2002. Available from: https://core.ac.uk/download/pdf/29158113.pdf, (accessed 19 October 2019).

[15] Zinkutė R, Taraškevičius R, Jankauskaitė M, Stankevičius Ž. Methodological alternatives for calculation of enrichment factors used for assessment of topsoil contamination. J Soils Sediments. 2017;17(2):440–52. doi: 10.1007/s11368-016-1549-4.

[16] Mirzaei R, Esmaili-Sari A, Hemami MR, Rezaei HR, Rodrıguez Martın JA. Levels and variability of metals in soils of the province of Golestan (Iran). Arch Environ Contam Toxicol. 2014;67:617–29. doi: 10.1007/s00244-014-0043-5.

[17] https://www.lgt.lt/zemelap/main.php?sesName=lgt1580379908 (accessed 10 January 2018).

[18] https://www.geoportal.lt/oldmap/index.jsp?mode=zis (accessed 23 August 2015).

[19] World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps. Update 2015. World soil resources reports 106. Rome: FAO, 2015. Available from: http://www.fao.org/3/i3794en/I3794en.pdf (accessed 10 April 2018).

[20] Motuzas AJ, Buivydaitė VV, Vaisvalavičius, Šleinys RA. Dirvotyra (Pedology). Vilnius: Enciklopedija; 2009 in Lithuanian.

[21] Hoogsteen MJJ, Lantinga EA, Bakker EJ, Groot JCJ, Tittonell PA. Estimating soil organic carbon through loss on ignition: effects of ignition conditions and structural water loss. Eur J Soil Sci. 2015;66(2):320–8. doi: 10.1111/ejss.12224.

[22] Šinkovičová M, Igaz D, Kondrlová E, Jarošová M. Soil particle size analysis by laser diffractometry: result comparison with pipette method. IOP Conf Series: Mater Sci Eng. 2017;245(072025):1–14. doi: 10.1088/1757-899X/245/7/072025.

[23] Schramm R, Heckel J. Fast analysis of traces and major elements with ED(P)XRF using polarized Xrays: TURBOQUANT. J Phys IV France. 1998;8:Pr5-335–Pr5-342. doi: 10.1051/jp4:1998542.

[24] Siegel S, Castellan Jr NJ. Nonparametric statistics for the behavior sciences. 2nd ed. New York: McGraw-Hill; 1988.

[25] Mažvila J, Vaičys M, Beniušis R. Causes and consequences of the vertical migration of fine soil fractions. Žemės ūkio mokslai (Agricultural Sciences). 2008;15(3):36–41. Available from: http://mokslozurnalai.lmaleidykla.lt/publ/1392-0200/2008/3/36-41.pdf, (accessed 15 December 2019).

[26] Zinkutė R, Taraškevičius R, Želvys T. Major elements as possible factors of trace element urban pedochemical anomalies. Cent Eur J Chem. 2011;9(2):337–47. doi: 10.2478/s11532-011-0012-z.

[27] Taraškevičius R, Kazakauskas V, Sarcevičius S, Zinkutė R, Suzdalev S. Case study of geochemical clustering as a tool for tracing sources of clays for archaeological and modern bricks. Baltica. 2019;32(2):139–55. doi: 10.5200/baltica.2019.2.2.

[28] Reimann C, Siewers U, Tarvainen T, Bityukova L, Eriksson J, Gilucis A, et al. Agricultural Soils in Northern Europe: A Geochemical Atlas. Geologisches Jahrbuch, Sonderhefte, Reihe D, Heft SD 5. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung; 2003.

[29] Kadūnas V, Budavičius R, Gregorauskienė V, Katinas V, Kliaugienė E, Radzevičius A, et al. Geochemical atlas of Lithuania. Vilnius: Geological Survey of Lithuania, Geological Institute; 1999 (in Lithuanian and English).

[30] https://www.lgt.lt/images/LGT_leidiniai/Vilniaus_Q_geol_ir_geomorfolog_2014.pdf (accessed 15 October 2019, in Lithuanian).

[31] Wu J, Teng Y, Chen H, Li J. Machine-learning models for on-site estimation of background concentrations of arsenic in soils using soil formation factors. J Soils Sediments. 2016;16:1787–97. doi: 10.1007/s11368-016-1374-9.

[32] Zinkutė R, Taraškevičius R, Gulbinskas S, Stankevičius Ž, Jankauskaitė M. Variability of estimated contamination extent depending on calculation methods. Environment. Technology. Resources: Proceedings of the 10th International Scientific and Practical Conference; 2015 June 18–20; Rezekne, Latvia. Vol. II. Rezekne: Rezekne Higher Education Institution; 2015, p. 337–43. doi: 10.17770/etr2015vol2.234.

[33] Cave MR, Wragg J, Harrison H. Measurement modelling and mapping of arsenic bioaccessibility in Northampton. United Kingdom. J Environ Sci Heal A. 2013;48(6):629–40. doi: 10.1080/10934529.2013.731808.

[34] Seleznev AA. Yarmoshenko IV, Sergeev AP. Method for reconstructing the initial baseline relationship between potentially harmful element and conservative element concentrations in urban puddle sediment. Geoderma. 2018;326:1–8. doi: 10.1016/j.geoderma.2018.04.003.

[35] Gałuszka A, Migaszewski ZM. Geochemical background – an environmental perspective. Mineralogia. 2011;42(1):7–17. doi: 10.2478/v10002-011-0002-y.

[36] Rodrigues S, Urquhart G, Hossack I, Pereira MA, Duarte AC, Davidson C, et al. The influence of anthropogenic and natural geochemical factors on urban soil quality variability: a comparison between Glasgow, UK and Aveiro, Portugal. Environ Chem Lett. 2009;7:141–8. doi: 10.1007/s10311-008-0149-y.

[37] Johnson CC, Demetriades A. Urban geochemical mapping: a review of case studies in this volume. In: Johnson CC, Demetriades A, Locutura J, Ottesen RT, editors. Mapping the chemical environment of urban areas. Chichester: John Wiley & Sons Ltd; 2011. p. 7–27.

[38] Greinert A. The heterogeneity of urban soils in the light of their properties. J Soils and Sediments. 2015;15:1725–37. doi: 10.1007/s11368-014-1054-6.

[39] Taraškevičius R, Motiejūnaitė J, Zinkutė R, Eigminienė A, Gedminienė L, Stankevičius Ž. Similarities and differences in geochemical distribution patterns in epiphytic lichens and topsoils from kindergarten grounds in Vilnius. J Geochem Explor. 2017;183(B):152–65. doi: 10.1016/j.gexplo.2017.08.013.

[40] Taraškevičius R, Zinkutė R, Gedminienė L, Stankevičius Ž. Hair geochemical composition of children from Vilnius kindergartens as an indicator of environmental conditions. Environ Geochem Health. 2018;40(5):1817–40. doi: 10.1007/s10653-017-9977-7.

[41] Guobytė R, Satkūnas J. Pleistocene glaciations in Lithuania. In: Ehlers J, Gibbard PL, Hughes PD, editors. Quaternary glaciations – extent and chronology, a closer look. Developments in Quaternary Science 15. The Netherlands: Elsevier; 2011. p. 231–46.

[42] Zinkutė R, Baužienė I, Dilys K, Mažeika J, Taminskas J, Taraškevičius R. Recent Lithuanian peri-urban ombrotrophic bog records: indices derived from the contents of lead, zinc, copper and nickel. Geochem-Explor Env. 2015;15(4):293–318. doi: 10.1144/geochem2013-245.

[43] Świtoniak M, Mroczek P, Bednarek R. Luvisols or Cambisols? Micromorphological study of soil truncation in young morainic landscapes – Case study: Brodnica and Chełmno Lake Districts (North Poland). Catena. 2016;137:583–95. doi: 10.1016/j.catena.2014.09.005.

[44] Volungevicius J, Jukna L, Veteikis D, Vaisvalavicius R, Amaleviciute K, Slepetienė A, et al. The problem of soil interpretation according to the WRB 2014 classification system in the context of anthropogenic transformations. Acta Agr Scand B-S P. 2016;66(5):452–60. doi: 10.1080/09064710/2016.1164231.

[45] Kun A, Katona O, Sipos G, Barta K. Comparison of pipette and laser diffraction methods in determining the granulometric content of fluvial sediment samples. Journal of Environmental Geography. 2013;6(3–4):49–54. doi: 10.2478/jengeo-2013-0006.

[46] Dung TTT, Cappuyns V, Swennen R, Phung NK. From geochemical background determination to pollution assessment of heavy metals in sediments and soils. Rev Environ Sci Bio. 2013;12:335–53. doi: 10.1007/s11157-013-9315-1.

[47] Prohić E, Miko S, Peh Z. Normalisation and trace element contamination of soils in a Karstic Polje – An example from the Sinjsko Polje, Croatia. Geologia Croatica. 1995;48(1):67–86. Available from: http://www.geologia-croatica.hr/index.php/GC/article/view/GC.1995.06 (accessed 15 September 2019) .

[48] Albanese S, Breward A. Sources of anthropogenic contaminants in the urban environment. In: Johnson CC, Demetriades A, Locutura J, Ottesen RT, editors. Mapping the chemical environment of urban areas. Chichester: John Wiley & Sons Ltd; 2011. p. 116–27.

[49] Bhuie AK, Roy DN. Deposition of Mn from automotive combustion of methylcyclopentadienyl manganese tricarbonyl beside major highway of Greater Toronto area. J Air Waste Manag. 2001;51:174–85. doi: 10.1080/10473289.2001.10464350.

[50] Zinkutė R, Brannvall E, Taraškevičius R. Variation of trace element accumulation in topsoil near the Vilnius–Kaunas highway, Lithuania. Geologija. 2007;58:1–9, Available from: http://www.elibrary.lt/resursai/LMA/Geologija/geol72/4.pdf, (accessed 10 April 2020).

[51] Li S, Yang L, Chen L, Zhao F, Sun L. Spatial distribution of heavy metal concentrations in peri-urban soils in eastern China. Environ Sci Pollut Res. 2019;26:1615–27. doi: 10.1007/s11356-018-3691-6.

[52] Li J, Pu L, Liao Q, Zhu M, Dai X, Xu Y, et al. How anthropogenic activities affect soil heavy metal concentration on a broad scale: a geochemistry survey in Yangtze River Delta, Eastern China. Environ Earth Sci. 2015;73:1823–35. doi: 10.1007/s12665-014-3536-7.

[53] Aitchison J. The Statistical Analysis of Compositional Data. London: Chapman and Hall; 1986.

[54] Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueraz G, Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. Math Geol. 2003;35:279–300. doi: 10.1023/A:1023818214614.

[55] Filzmoser P, Hron K, Reimann C. Principal component analysis for compositional data with outliers. Environmetrics. 2009;20:621–32. doi: 10.1002/env.966.

[56] Filzmoser P, Hron K, Tolosana-Delgado R. Statistical analysis of geochemical compositions: Problems, perspectives and solutions. Appl Geochem. 2016;75:169–70. doi: 10.1016/j.apgeochem.2016.11.016.

[57] Fačevicová K, Bábek O, Hron K, Kumpan T. Element chemostratigraphy of the Devonian/Carboniferous boundary – a compositional approach. Appl Geochem. 2016;75:211–21. doi: 10.1016/j.apgeochem.2016.10.002.