

Research Article

Case Study of the Relationship between Aqua Regia and Real Total Contents of Harmful Trace Elements in Some European Soils

Ričardas Taraškevičius,¹ Rimantė Zinkutė,¹ Rimutė Stakėnienė,¹ and Marijus Radavičius²

¹ Nature Research Centre, Institute of Geology and Geography, T. Ševčenkos 13, 03223 Vilnius, Lithuania

² Institute of Mathematics and Informatics, Vilnius University, Akademijos 4, 08663 Vilnius, Lithuania

Correspondence should be addressed to Ričardas Taraškevičius; taraskevicius@geo.lt

Received 1 July 2012; Revised 23 August 2012; Accepted 3 September 2012

Academic Editor: Vladislav Chrastny

Copyright © 2013 Ričardas Taraškevičius et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Both *real total* and *aqua regia* contents of trace elements in soils are often denominated by the same word “total” though the results are not identical. The formulas would be helpful for recalculation of *aqua regia* contents to *real totals*. Data for primary formulas were taken from the International Soil-Analytical Exchange Program of Wageningen Evaluating Programs for Analytical Laboratories. The degree of extractability DE of element in a sample was calculated by weighting the ratios of median contents in *aqua regia* to median *real total contents* in different periods with weights proportional to the respective number of determinations. According to descending median values of DE (%) in ISE European soil samples the elements are arranged as follows: Hg(98), Cd(94), Cu(91), Zn(90), Mn(89), Ni(88), Co(86), As(82), Mo(82), Pb(79), V(61), Sn(57), Cr(56), Sb(53), Be(51), B(46), U(35), and Ba(21). These values can be used for approximate recalculation of *aqua regia* contents to *real totals* and adjustment of contamination assessment. An attempt to obtain more explicit prediction by fitting regression models and problems related to high leverage and possibly influential points are discussed and their possible relation to the specificity of soil composition is pointed out.

1. Introduction

Classification and choice of analytical methods for geochemical investigations depend on the field of application. According to the completeness of extraction the analytical methods for determination of trace elements values in soils or sediments can be subdivided into two main groups: (1) for *real total contents (totals)* and (2) for *partial contents* [1].

Aqua regia digestion [2] is widely used for determination of the contents of harmful chemical elements in the soil aiming to solve environmental tasks. Unfortunately, during this time the unsupported tradition to denominate the *aqua regia* contents as “total contents” was established. This denomination is sometimes still used, despite the following statement in description of standard *aqua regia* procedure used for soil improvers [3]: “Elements, extractable in *aqua regia* cannot be described as “totals...” Denomination of “total contents” for *aqua regia* extraction results is especially

widely spread when comparing *aqua regia* contents with the contents obtained by extraction procedures which do not use strong acids [4–6]. There are also cases when the phrase “total contents” is attributed to *aqua regia* contents even if the authors are well aware that actually their research deals with “pseudototal contents” [7–9]. Besides, some researchers point out that in topsoil, where residual silicates do not display high metal concentrations, the values obtained by *aqua regia* digestion are representative of the total metal concentration [10]. According to Sastre et al. [11] *aqua regia* contents of the main hazardous pollutants Hg, Cd, Zn, Cu, and Pb in environmental samples do not much differ from real total contents obtained by digestion with HF.

Despite the wide use of *aqua regia* extraction results, in most cases geochemical exploration or mapping cannot do without determination of *real total* contents [1, 12–17]. According to Reimann et al. [18] “geologist would prefer to use analytical results from “total” dissolution (HF+HClO₄ or others acids compositions) or an analytical technique (e.g.,

XRF or INAA) resulting in total element concentrations." There is an opinion that strong extractants should be taken as a starting point of monitoring, because they give a worst-case estimate of possible long-term changes [19].

The problem arises when the same term "total content" is used for both real totals and *aqua regia* contents in soils and when exact description of the procedure is not given. In this case the possibility to compare quite different results and obtain incorrect conclusions appears. Besides, comparison of screening values of soil in various European countries [20] is problematic, because screening values in some countries are based on *aqua regia* or other weaker extracts; meanwhile in other countries the total content is required.

As both groups of analysis have their own advantages, it might be useful to determine the relationship between their results. Some solutions of this problem can be obtained by presenting approximate formulas for prediction of real total contents according to *aqua regia* contents or vice versa. The idea to investigate comparability of results obtained by different methods was proposed in an overview of ISO and CEN standards used in the European Union for chemical analysis of solid matrix material [21]. The theme of comparison of *aqua regia* contents to *real total* contents (HF-extractable or obtained by XRF) is also topical in international scientific research and geochemical practice [15, 16, 18]. The researchers present correlation coefficients and the percentage of *aqua regia* contents of chemical elements from HF-extractable or obtained by XRF content. Despite these achievements, there is still a lack of formulas for recalculation or prediction of one content to another content.

The aim of this research was to present the principle of preparation of the formulas for the relationship between *aqua regia* contents determined according to ISO 11466 : 1995 [22] and *real total contents* of harmful trace elements on the basis of some soil samples from Europe. The novelty of this research is that it is based on measurements of the same samples by a large number of different laboratories which use various digestion/extraction procedures for *real total contents* and various methods of detection of both *real total* and *aqua regia* contents. International Soil-Analytical Exchange program of Wageningen Evaluating Programs for Analytical Laboratories (WEPAL ISE) gives such possibility.

The tasks were the following: (1) to collect data on medians of *real total contents* and *aqua regia contents* [22] of harmful trace elements Ag, As, B, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Se, Sn, U, V, and Zn from annual or quarterly WEPAL ISE reports; (2) to estimate the values of the degree of extractability of harmful elements from each sample based on measurements during a particular period; (3) to calculate the weighted estimate of the degree of extractability of harmful elements from each samples based on all repeated measurements of samples during different periods; (4) to analyse the strength of the correlation between *aqua regia contents* and *real total contents* and present formulas for prediction of *real totals* according to *aqua regia contents*.

The mentioned elements were chosen because their soil screening values for unacceptable risk are listed in normative documents of various countries [20]; they are often shown as phytotoxic [23]; besides, it is sometimes misjudged that there

is no substantial difference between the values of their real total and *aqua regia* contents.

2. Data and Methods

2.1. WEPAL ISE Program as a Source of Data. The most impartial comparison of results obtained by different extraction procedures is achieved when measurements of the same samples are done by as great as possible number of laboratories which use similar extraction procedures but not necessarily the same equipment. Advantages of WEPAL ISE are the following. (1) The same samples are analysed by many laboratories; (2) *aqua regia* and *real total contents* are processed in each of these groups separately; (3) there is a great number of analysed chemical elements.

In WEPAL ISE program four soil or sediment samples of unknown chemical composition are sent every 3 months to the participating laboratories. The samples (100 g each) are dried and represent a fraction <0.5 mm. Homogeneity and stability of the materials distributed are sufficient for the goal of proficiency testing [24]. The participants analyse them according to their own procedures and for those elements and parameters they are interested in. Before the end of each quarter they must send to WEPAL the results and method indicating codes (MICs). MIC is the basis for assigning analytical results of reported determinand to a certain group. The first character indicates the method of extraction or digestion. The next three characters indicate the method of detection [25]. Till 2005, there were 2 different groups: *real totals* and *so-called totals (acid extractable)* [26]. The results obtained by *aqua regia* extraction were in the group of the *so-called totals*. In 2006, the results obtained by *aqua regia* extraction using the procedure of ISO 11466 [22] were distinguished into a separate group [27]. However, results obtained by this procedure have never been in the group of *real totals*.

Part of the procedures for obtaining *real total contents* is without digestion/extraction: neutron activation, X-ray fluorescence, or some others. Another part requires either acid digestion/extraction with HF and final medium HCl (or H₂SO₄, or HNO₃, or HClO₄) or melting. There are more than 30 possible methods of detection of trace elements in prepared solutions [25]. All these methods of detection can be used also for the measurement of *aqua regia contents* obtained using the procedure described in ISO 11466 [22]. Earlier comparison of proficiency testing results of 7 different methods of detection did not show their significant influence on results [28].

2.2. WEPAL ISE Methodology for Elimination of Outliers. WEPAL ISE participants are informed about their results in quarterly and annual reports which give them valuable information not only about evaluation of their own results but also about other measurements. For each analyte in each group, the reference value for proficiency testing is obtained which is influenced by the algorithm used [29].

One of the possible methodologies is based on estimators for location and scale which give less weight to observations

in the tails [30]. This methodology was used by WEPAL ISE for a long time. If the number of laboratory measurements exceeds 7, it is based on consecutive elimination of outliers using medians and median absolute deviations; the final estimate can be the first, the second, or the third median [27]. In 2009, a new robust and insensitive to outliers methodology was introduced which provides the statistical characteristics (mean and standard deviation) of the highest mode of the data set and is based on normal distribution approximation [31]. So since 2010 only the first median has been presented in WEPAL ISE reports.

2.3. Data and Contribution of the Institute of Geology and Geography. Data on *real totals* and *aqua regia* (ISO 11466:1995 [22]) contents of Ag, As, B, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Se, Sn, U, V, and Zn in 32 samples analysed by WEPAL ISE participants were taken from reports: 4 annual (2006, 2007, 2008, and 2009) [25, 27, 32, 33] and 1 quarterly (2010) [34].

The Institute of Geology and Geography analysed all these samples for determination of *real total contents* of chemical elements by two different methods. Optical atomic emission spectrophotometry OAES was used until 2009 [17]. Vaporization of 1 g of fine-milled soil was direct and done at 5000–6000°C in the electric arc between three graphite nails; measurement of Ag, B, Ba, Co, Cr, Cu, Mn, Mo, Ni, Pb, Sn, V, and Zn was done by optical spectrophotometer DFS 13. In 2007 energy-dispersive X-ray fluorescence EDXRF (Spectro XEPOS) for determination of As, Ba, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Se, Sn, U, V, and Zn in pressed pellets was started [17].

Aiming to analyse only 19 European soil samples, the non-European soil samples and river clay samples were excluded from WEPAL ISE samples. The European soil samples (Table 1) reflect a wide variety of lithological composition (from sandy soil to clay soil or marshland) and some other characteristics (clay particles, loss on ignition, and Ca content).

The intervals of the contents of harmful chemical elements in the study samples are wide and include the median values of the North European agricultural [15], European [16], and Lithuanian soils [13] (Table 2). Median contents of harmful trace elements in the study samples are rather similar to respective medians in European soils. Maximum contents of most trace elements are close to soil screening values which are used in some European Union countries including Lithuania for an intermediate (warning) or unacceptable risk of residential soil use [20, 35].

2.4. Mathematical Treatment of Data. Estimation of *degree of extractability* DE_i of an element in a sample measured in period i was done by dividing medians of *aqua regia contents* [22] by medians of *real total contents*, both given in WEPAL ISE reports. To obtain the percentage of DE_i the following formula was used:

$$DE_i = 100 \cdot \frac{C_{iARM}}{C_{iRTM}} \quad (1)$$

C_{iARM} is a median *aqua regia content* (mg/kg) given in report of period i , C_{iRTM} is a median *real total content* (mg/kg) given in report of period i , and both medians are usually after elimination of outliers done by WEPAL ISE methodology.

The respective numbers of measuring laboratories used for estimations of two medians in this period were N_{iAR} and N_{iRT} . Minimum from these two numbers was used to characterise the number of measuring laboratories N_i related to estimation of DE_i .

As most of the samples were analysed repeatedly, it was necessary to combine information of the same sample analysed during different periods. The following procedure was used for this aim. The percentage of degree of extractability DE of an element in a sample was calculated by weighting according to the following formula:

$$DE = \frac{\sum_{i=1}^k DE_i * N_i}{\sum_{i=1}^k N_i} \quad (2)$$

where DE_i is the percentage of degree of extractability of an element in period i , N_i is the number of measuring laboratories related to estimation of DE_i in this period, and k is the number of periods.

The number of laboratory measurements (N) of both *real total* and *aqua regia contents* of each element in each sample (indicated as a subscript in Table 1) was obtained by summing up the numbers of measuring laboratories in all periods when the sample was analysed.

The primary matrix of degree of extractability in ISE European soil samples was censored in the following way. If $N < 3$, DE would be deleted (too unreliable results). As WEPAL ISE does not perform elimination of outliers when the number of measuring laboratories is below 8 [25], the values of DE when $3 \leq N < 8$ are also not very reliable, but these results were not deleted from the statistical treatment.

Median DE values were calculated for WEPAL ISE soil samples of 2 European countries (The Netherlands—12 samples, Switzerland—6 samples) and denoted by DE_C . Respective DE_C value of soil from Spain was based only one sample.

Analogous DE_C values in agricultural soil from 10 countries around the Baltic Sea were calculated using results given in the tables of “Agricultural soils in Northern Europe: a geochemical atlas” [15]. The contents obtained by HF treatment and XRF analysis were considered as one group of *real totals*; therefore respective medians were averaged. For each of these countries, median *aqua regia* contents were divided by averaged medians of *real totals* and multiplied by 100.

EXCEL software was used for calculation of DE_i , DE, and DE_C values, their statistical characteristics, and preparation of charts and was helpful for calculation of validation parameters. STATISTICA (version 9) software was used for determination of Pearson correlation coefficients R_p and nonparametric Spearman rank correlation coefficients R_s between trace element *aqua regia* and *real total* contents and their significance levels p [36]. This software as well as SPSS (release 8.00) was used for the selection and modification

TABLE 1: Characteristics of samples.

No.	Chemical elements and number of laboratory measurements of both <i>real total</i> and <i>aqua regia</i> contents	Lithology, country	CL	LOI	Ca
986	Ag ₃ , As ₄₉ , B ₇ , Ba ₃₁ , Be ₁₆ , Cd ₃₇ , Co ₄₇ , Cr ₈₅ , Cu ₈₉ , Hg ₃₇ , Mn ₈₉ , Mo ₂₇ , Ni ₇₇ , Pb ₈₅ , Sb ₁₈ , Se ₈ , Sn ₁₃ , U ₈ , V ₅₂ , Zn ₁₀₅	Sandy soil/NL	2.0	3.6	1.2
919	As ₂₆ , B ₂ , Ba ₁₁ , Be ₉ , Cd ₂₄ , Co ₂₁ , Cr ₄₄ , Cu ₄₅ , Hg ₂₂ , Mn ₃₈ , Mo ₁₄ , Ni ₄₀ , Pb ₄₂ , Sb ₉ , Se ₅ , Sn ₆ , U ₄ , V ₁₉ , Zn ₅₁	Sandy soil/NL	3.3	4.8	2.1
997	As ₂₄ , B ₇ , Ba ₁₅ , Be ₁₂ , Cd ₃₀ , Co ₂₃ , Cr ₅₁ , Cu ₅₂ , Hg ₂₉ , Mn ₅₀ , Mo ₁₄ , Ni ₄₉ , Pb ₄₉ , Sb ₁₁ , Se ₆ , Sn ₇ , U ₃ , V ₂₃ , Zn ₅₂	Sandy soil/NL	4.0	4.3	3.7
995	Ag ₆ , As ₅₆ , B ₈ , Ba ₂₉ , Be ₂₀ , Cd ₇₉ , Co ₆₃ , Cr ₁₀₉ , Cu ₁₁₄ , Hg ₅₂ , Mn ₉₀ , Mo ₄₄ , Ni ₁₁₉ , Pb ₁₁₀ , Sb ₃₁ , Se ₉ , Sn ₂₉ , U ₁₁ , V ₅₆ , Zn ₁₁₄	Sandy soil/NL	4.2	5.3	4.3
958	As ₃₃ , B ₃ , Ba ₁₅ , Be ₁₁ , Cd ₂₀ , Co ₃₀ , Cr ₅₀ , Cu ₅₂ , Hg ₂₈ , Mn ₄₂ , Mo ₁₇ , Ni ₅₀ , Pb ₅₂ , Sb ₁₁ , Se ₆ , Sn ₁₁ , U ₆ , V ₃₀ , Zn ₅₉	Sandy soil/NL	4.7	2.8	0.9
867	As ₁₃ , B ₁ , Ba ₆ , Be ₄ , Cd ₁₀ , Co ₁₅ , Cr ₁₈ , Cu ₂₁ , Hg ₁₀ , Mn ₂₀ , Mo ₁₁ , Ni ₂₁ , Pb ₁₆ , Sb ₅ , Se ₁ , Sn ₃ , U ₂ , V ₁₂ , Zn ₂₄	Forest sandy soil/NL	5.1	8.5	0.4
918	As ₂₇ , B ₃ , Ba ₁₀ , Be ₉ , Cd ₁₇ , Co ₁₇ , Cr ₃₉ , Cu ₃₉ , Hg ₁₈ , Mn ₃₉ , Mo ₁₃ , Ni ₃₃ , Pb ₃₈ , Sb ₇ , Se ₄ , Sn ₃ , U ₃ , V ₂₀ , Zn ₄₃	Sandy soil/NL	6.0	6.0	3.3
961	As ₇₄ , B ₇ , Ba ₃₃ , Be ₂₇ , Cd ₅₂ , Co ₇₂ , Cr ₁₀₄ , Cu ₁₀₁ , Hg ₅₆ , Mn ₉₉ , Mo ₂₉ , Ni ₁₀₃ , Pb ₁₀₅ , Sb ₂₆ , Se ₁₁ , Sn ₂₁ , U ₁₀ , V ₆₄ , Zn ₁₁₄	Clay/NL	13.4	3.9	4.0
992	As ₂₂ , B ₆ , Ba ₁₁ , Be ₁₁ , Cd ₁₈ , Co ₂₃ , Cr ₃₃ , Cu ₃₁ , Hg ₁₆ , Mn ₂₇ , Mo ₁₈ , Ni ₃₂ , Pb ₃₁ , Sb ₈ , Se ₇ , Sn ₂₁ , U ₃ , V ₁₉ , Zn ₃₈	Sandy clay soil/NL	14.7	6.8	29.7
885	As ₁₆ , B ₃ , Ba ₇ , Be ₆ , Cd ₁₀ , Co ₁₈ , Cr ₂₄ , Cu ₂₇ , Hg ₁₃ , Mn ₂₄ , Mo ₅ , Ni ₂₃ , Pb ₂₀ , Sb ₆ , Se ₄ , Sn ₅ , U ₂ , V ₁₃ , Zn ₂₇	Braunerde pseudogley/CH	15.0	4.9	3.8
884	As ₂₈ , B ₄ , Ba ₁₂ , Be ₉ , Cd ₂₀ , Co ₃₂ , Cr ₃₉ , Cu ₃₉ , Hg ₂₃ , Mn ₃₉ , Mo ₁₇ , Ni ₄₆ , Pb ₄₀ , Sb ₁₀ , Se ₇ , Sn ₉ , U ₄ , V ₂₀ , Zn ₅₁	Braunerde/CH	19.3	6.7	7.8
881	As ₂₆ , B ₂ , Ba ₁₁ , Be ₇ , Cd ₂₁ , Co ₂₆ , Cr ₄₀ , Cu ₄₀ , Hg ₂₁ , Mn ₃₃ , Mo ₁₂ , Ni ₃₇ , Pb ₄₀ , Sb ₁₁ , Se ₇ , Sn ₈ , U ₄ , V ₁₇ , Zn ₄₉	Clay/NL	20.0	3.9	40.2
900	As ₂₇ , B ₅ , Ba ₁₀ , Be ₁₂ , Cd ₁₉ , Co ₂₆ , Cr ₃₁ , Cu ₃₇ , Hg ₂₀ , Mn ₃₈ , Mo ₁₅ , Ni ₃₆ , Pb ₃₆ , Sb ₃ , Se ₃ , Sn ₇ , U ₃ , V ₂₀ , Zn ₃₇	Calcareous brown soil/CH	21.8	5.4	12.1
865	As ₁₆ , B ₂ , Ba ₇ , Be ₄ , Cd ₇ , Co ₁₃ , Cr ₂₂ , Cu ₁₀ , Hg ₁₁ , Mn ₁₄ , Mo ₁₀ , Ni ₁₇ , Pb ₁₉ , Sb ₃ , Se ₂ , Sn ₆ , U ₄ , V ₁₃ , Zn ₂₁	Loamy soil/CH	22.0	27.1	2.4
889	As ₂₉ , B ₃ , Ba ₁₁ , Be ₁₁ , Cd ₂₂ , Co ₃₂ , Cr ₄₁ , Cu ₄₀ , Hg ₁₆ , Mn ₃₇ , Mo ₁₆ , Ni ₄₀ , Pb ₃₇ , Sb ₁₁ , Se ₂ , Sn ₈ , U ₃ , V ₂₁ , Zn ₄₅	Soil with high carbonate/ES	24.5	5.5	145
962	Ag ₃ , B ₄ , Ba ₁₄ , Be ₁₁ , Cd ₁₉ , Co ₂₉ , Cr ₄₄ , Cu ₃₁ , Hg ₂₂ , Mn ₃₅ , Mo ₁₄ , Ni ₃₉ , Pb ₃₅ , Sb ₆ , Se ₆ , Sn ₁₂ , U ₆ , V ₂₄ , Zn ₄₄	Sandy clay/NL	35.8	5.7	38.9
872	Ag ₄ , B ₃ , Ba ₁₈ , Be ₁₁ , Cd ₂₃ , Co ₃₀ , Cr ₄₇ , Cu ₄₆ , Hg ₂₆ , Mn ₃₈ , Mo ₁₃ , Ni ₄₅ , Pb ₄₁ , Sb ₁₂ , Se ₄ , Sn ₁₄ , U ₇ , V ₂₇ , Zn ₄₆	Braunerde clay/CH	40.8	9.0	53.9
909	As ₂₄ , B ₄ , Ba ₁₀ , Be ₁₃ , Cd ₂₂ , Co ₂₅ , Cr ₄₆ , Cu ₄₂ , Hg ₂₀ , Mn ₃₃ , Mo ₁₂ , Ni ₄₄ , Pb ₃₄ , Sb ₁₀ , Se ₄ , Sn ₉ , U ₂ , V ₂₂ , Zn ₄₉	Marshland/CH	43.5	29.9	21.6
910	As ₂₇ , B ₄ , Ba ₁₁ , Be ₁₂ , Cd ₂₁ , Co ₃₁ , Cr ₃₇ , Cu ₄₁ , Hg ₂₂ , Mn ₄₀ , Mo ₁₁ , Ni ₄₂ , Pb ₃₉ , Sb ₁₀ , Se ₉ , Sn ₈ , U ₂ , V ₂₀ , Zn ₄₉	Clay soil/NL	58.9	11.4	6.9

No.: sample number in WEPAL ISE program; country of sampling is given after slash: NL: the Netherlands, CH: Switzerland, and ES: Spain; CL: median content of clay particles, that is, fraction <2 μm, %; LOI: median content of loss on ignition, %; Ca: median content of Ca, g/kg.

TABLE 2: Statistical characteristics of WEPAL ISE European soil samples compared to other sources.

VAR	MIN	MAX	MEDIAN	NEmed	EUmed	LTmed	SV	LTsv
Ag _{RT}	0.05	2.90	0.23	“_”	0.27	0.07	“_”	2
Ag _{AR}	0.02	2.86	0.10	“_”	“_”	“_”	“_”	“_”
As _{RT}	1.49	29.7	9.69	3.2	7.0	2.5	5–300	10
As _{AR}	1.10	25.0	7.75	1.9	6.0	“_”	“_”	“_”
B _{RT}	13.0	118	47.9	“_”	“_”	26	“_”	50
B _{AR}	3.58	36.8	17.2	“_”	“_”	“_”	“_”	“_”
Ba _{RT}	129	556	252	402	375	345	285–2000	600
Ba _{AR}	7.95	368	55.4	46	65	“_”	“_”	“_”
Be _{RT}	0.25	3.12	1.19	1.17	<2	1.0	2–30	10
Be _{AR}	0.07	2.03	0.60	“_”	“_”	“_”	“_”	“_”
Cd _{RT}	0.08	12.9	0.30	“_”	0.15	0.15	1–20	3
Cd _{AR}	0.07	12.8	0.32	0.13	“_”	“_”	“_”	“_”
Co _{RT}	0.79	18.5	7.62	5.3	7.8	4.3	20–300	30
Co _{AR}	0.58	15.7	6.42	4.1	7.0	“_”	“_”	“_”
Cr _{RT}	11.2	134	77.5	28.5	60	30	50–1000	100
Cr _{AR}	5.58	117	36.8	12.0	22	“_”	“_”	“_”
Cu _{RT}	5.86	65.5	14.8	10.5	13.0	8.1	100–1000	100
Cu _{AR}	4.66	64.7	13.2	7.8	12.0	“_”	“_”	“_”
Hg _{RT}	0.021	0.29	0.057	“_”	0.037	0.075	0.500–56	1.5
Hg _{AR}	0.022	0.27	0.055	“_”	“_”	“_”	“_”	“_”
Mn _{RT}	54.1	1530	404	426	503	427	“_”	1500
Mn _{AR}	32.1	1462	380	255	382	“_”	“_”	“_”
Mo _{RT}	0.26	4.70	1.05	<2	0.62	0.64	5–200	5
Mo _{AR}	0.16	4.42	0.91	<3	“_”	“_”	“_”	“_”
Ni _{RT}	2.10	83.6	23.0	8.5	18	12	30–500	75
Ni _{AR}	1.50	75.0	19.4	8.0	14	“_”	“_”	“_”
Pb _{RT}	10.2	85.5	25.7	16.0	22.6	15.0	40–700	100
Pb _{AR}	6.81	71.0	21.1	8.4	15.0	“_”	“_”	“_”
Sb _{RT}	0.33	4.50	0.70	0.24	0.6	1.0	2–40	10
Sb _{AR}	0.12	2.19	0.35	<10	“_”	“_”	“_”	“_”
Se _{RT}	0.40	13.1	0.88	<5	“_”	0.2	3–100	5
Se _{AR}	0.15	0.9	0.26	0.14	“_”	“_”	“_”	“_”
Sn _{RT}	1.10	17.4	2.56	<2	3.0	2.1	1–900	10
Sn _{AR}	0.53	15.9	1.38	1.1	“_”	“_”	“_”	“_”
U _{RT}	0.44	3.03	2.46	1.4	2.0	2.2	“_”	20
U _{AR}	0.17	1.96	0.80	“_”	“_”	“_”	“_”	“_”
V _{RT}	11.5	153	54.0	38.0	60.4	32.0	90–500	150
V _{AR}	7.66	88.2	34.1	18.0	33.0	“_”	“_”	“_”
Zn _{RT}	10.2	232	68.3	42.5	52.0	26.0	150–3000	300
Zn _{AR}	8.11	214	59.4	33.0	48.0	“_”	“_”	“_”

VAR: variables, that is, chemical element contents (mg/kg). The subscript RT indicates real total content and subscript AR: aqua regia content; MIN: minimum, MAX: maximum, and MEDIAN: median content of WEPAL ISE European soil samples; NEmed: median content in agrarian soil of North European countries [15], EUmed: median content in European soils [16], and “_”: no data; LTmed: median content in Lithuanian soil [13]; SV: interval of soil screening values of intermediate (warning) risk and potentially unacceptable risk (residential soil use) in Austria, Belgium (Brussels, Flanders, Walloon), Finland, The Netherlands, Germany, Denmark, Sweden, Italy, Czech Republic, Slovakia, United Kingdom, and Poland [20]; LTsv: Lithuanian soil screening values of intermediate (warning) risk [35].

of a regression model. Detailed discussion will be given in results.

3. Results and Discussion

3.1. Variability of Degree of Extractability. Degree of extractability DE of As, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Sn, V, and Zn was characterised for all 19 samples and after elimination of a part of the points according to procedure which will be described in 3.2 (Table 3). DE of Se was known only in 17 samples, of B—in 15, of U—in 15, and of Ag only in 4 samples. Besides, results of different elements in different samples are unequally reliable according to the number of laboratory measurements (N) (Table 1): some of them are lower than 8. According to the percentage of samples, for which $N > 7$, the elements can be arranged as follows: Zn, Cr, Ni, Cu, Pb, Mn, Co, As, V, Hg (100%) > Cd, Mo (95%) > Ba (84%) > Be (79%) > Sb (68%) > Sn (58%) > Se (21%) > U (16%) > B (5%) > Ag(0%).

The minimum *real total contents* of all harmful trace elements, except Se, are lower than their soil geochemical background values [13, 35]. Maximum *real total contents* of Ag, As, B, Cd, Cr, Mn, Ni, Se, Sn, and V measured in soil samples exceed Lithuanian soil screening values [35].

The median percentages of DE (Table 3) can be used as the first approximation for recalculation of *aqua regia content* results (ISO 11466:1995) to *real totals* according to the following formula:

$$C_{RT} = 100 * \frac{C_{AR}}{DE}, \quad (3)$$

where C_{RT} is *real total content* of trace element (mg/kg), C_{AR} is *aqua regia content* of harmful trace element (mg/kg), and DE is the median percentage of degree of extractability of this element.

Degree of extractability of 10 harmful chemical elements from soil of different European countries DE_C is variable (Figure 1). Soil samples analysed for “Agricultural soils in Northern Europe: a geochemical atlas” [15] are from 10 countries in the eastern part of Europe; meanwhile, samples analysed by WEPAL ISE program participants are from 3 countries of Western Europe. Two of the latter countries—The Netherlands and Switzerland (Spain was characterised by a single sample)—are distinguished by higher DE_C of As, Cr, Mn, Pb, V, and Zn compared to soil from 10 countries of the eastern part of Europe. This elevated extractability can be determined both by soil characteristics and geochemical peculiarities in various countries and by differences between methodology. Besides, the fact that soils of Switzerland and Spain greatly differ from soils of other countries according to DE_C of Ba is obvious.

Agrarian soils of Lithuania have only slightly higher DE_C of Ba compared to soil from Netherlands. The difference between DE_C of As in soil of The Netherlands and Lithuania is also small; both countries are distinguished from other countries by a higher degree of extractability of this element. Analysis of Spearman rank correlation coefficients R between DE_C values of As, Ba, Co, Cr, Cu, Mn, Ni, Pb, V, and

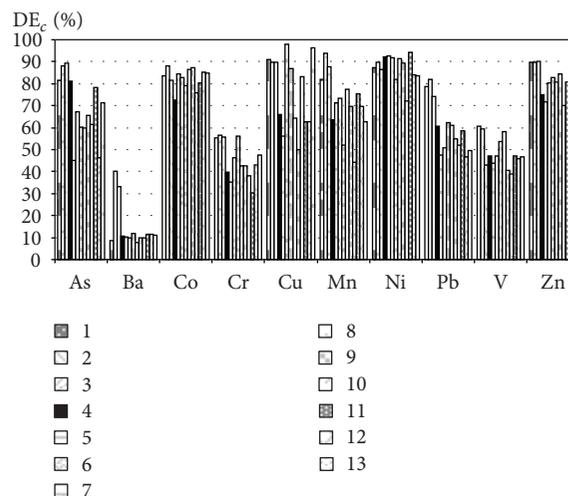


FIGURE 1: Degree of extractability of As, Ba, Co, Cr, Cu, Mn, Ni, Pb, V, and Zn as the percentage of *aqua regia* from real total contents in soils of different European countries (1) The Netherlands, (2) Switzerland, (3) Spain, (4) Lithuania, (5) Belarus, (6) Estonia, (7) Finland, (8) Germany, (9) Latvia, (10) Norway, (11) Poland, (12) Russia, (13) and Sweden. Soil samples from The Netherlands, Switzerland, and Spain were analysed by participants of WEPAL ISE program; while from other 10 countries in the following laboratories, WDXRF analysis was performed at BGR, Hannover; HF extraction and analysis by ICP-MS—at GTK's Espoo Geolaboratory, *aqua regia* extraction and analysis by AAS and ICP-AES—at GTK's Rovaniemi Geolaboratory [15].

Zn characterising different countries has revealed significant positive correlation between soil of The Netherlands and soil of Lithuania: $R = 0.770$, significance level $P = 0.009$. However, soil of Lithuania is even more similar to soil of Poland according to DE_C values of the above-mentioned elements: $R = 0.952$, $P = 0.00002$.

3.2. Equations for Recalculation Aqua Regia Contents to Real Total Contents. According to reliability of results, chemical elements studied can be subdivided into 2 groups: (1) a *less reliably characterised group* which includes Sb, Sn, Se, U, B, and Ag for which there are less than 75% of samples with a sufficient number of laboratory measurements ($N > 7$); (2) a *more reliably characterised group* which includes As, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, V, and Zn for which there are more than 75% of samples with $N > 7$.

Scatterplots of real total versus *aqua regia* contents of elements from the less reliably characterised group (Figure 2) indicate that presently for most of them (Se, Ag, U, and B) it is difficult to find out the equations for recalculation of *aqua regia* contents to real total contents. For Se the task is impossible, because according to all 17 samples both Pearson (R_p) and Spearman (R_s) correlation coefficients are insignificant ($P = 0.572$ and $P = 0.600$, resp.) and even negative due to obvious outlier (sample No. 885, braunerde pseudogley) and a great number of less reliably ($N \leq 7$) characterised samples. For Ag both R_s (1.000) and R_p (0.999, $P = 0.002$) coefficients are very high, but this element is

TABLE 3: Median values of degree of extractability of aqua regia contents to real totals.

El.	MIN	MAX	MEDIAN	MEDIAN*	El.	MIN	MAX	MEDIAN	MEDIAN*
Ag*	37.1	98.6	75.7	53.6	Mn	58.2	95.5	89.4	88.6
As	72.5	89.4	82.4	82.4	Mo	56.5	93.9	82.0	82.0
B	18.3	63.0	39.6	45.9	Ni	69.1	105	88.4	88.4
Ba	5.05	66.1	21.9	20.5	Pb	65.7	90.5	78.8	78.7
Be	25.0	67.2	51.3	51.2	Sb	36.5	69.4	52.1	53.0
Cd	85.0	109	94.1	93.8	Se*	1.53	144	38.7	34.6
Co	74.2	93.3	85.6	85.9	Sn	33.5	93.4	58.2	56.7
Cr	32.2	90.9	55.9	55.9	U	19.0	63.9	35.4	35.4
Cu	79.2	97.9	90.6	90.6	V	43.0	88.8	60.6	61.0
Hg	89.1	107	96.9	97.6	Zn	79.9	98.3	90.0	89.5

El.: harmful chemical elements. Degree of extractability (%): MIN: minimum value, MAX: maximum value, and MEDIAN: median value before elimination of high leverage and possibly influential points, and MEDIAN*: median value after elimination of the high leverage and possibly influential points. The elements are in bold if the median value of degree of extractability after elimination of the above-mentioned points changed not more than 2%. *: indicates that the statistical parameters of their degree of extractability are not reliable due to a low number of laboratory measurements (Table 1), these parameters are given only for information.

represented by only 4 samples, each of them with less than 7 laboratory measurements. High correlation is due to sample No. 995 (sandy soil) with high Ag content. For U and B both R_S (0.815 and 0.812) and R_P (0.811 and 0.929) are positive and significant ($P < 0.00025$) indicating that the relationship between *aqua regia* and real total contents exists. Still these 2 elements have a low number of reliably ($N > 7$) characterised samples: B only one and U only three. Besides, their distribution pattern is complicated, for example, for U the point representing sample No. 889 (soil with high carbonate) and for B sample No. 962 (sandy clay soil) are at great distance from other samples. Therefore equations for these chemical elements will also not be presented. Only for Sb and Sn from this group, the equations will be derived, because their R_S (0.984 and 0.995) and R_P (0.918 and 0.949) are positive and significant at $P < 0.0001$ and more than a half of the samples are reliably characterised.

The scatterplots of 14 chemical elements from the more reliably characterised group (Figure 3) clearly demonstrate that there is a relationship between their *aqua regia* and real total contents; significant ($P < 0.00005$) R_P and R_S values confirm this.

According to R_S all elements from more the reliably characterised group and 2 elements (Sb, Sn) from the less reliably characterised group are arranged as follows: Zn, As(0.998) > Mn(0.995) > Ni(0.993) > Cd, Cu(0.992) > Mo(0.989) > Co(0.988) > Be(0.984) > Hg(0.974) > Pb(0.966) > Sb(0.918) > V(0.907) > Sn (0.891) > Cr(0.867) > Ba(0.810). The equations will be derived for these 16 chemical elements using all 19 samples.

It can be supposed from respective scatterplots of these elements that the relationships are linear. However, the curve estimation procedure from SPSS (release 8.00) was used to check for the possibility to apply several models: linear, 2nd order polynomial, logarithmic, and exponential regression functions. For each model, an ANOVA table was given, the values of F -test were significant ($P < 0.00005$) for all these models. However, the highest values of adjusted R^2

TABLE 4: Adjusted R^2 values of more reliably characterised chemical elements obtained using different regression models without elimination of measurements.

Element	Model			
	Linear	2nd order polynomial	Logarithmic	Exponential
As	0.996	0.996	0.815	0.831
Ba	0.814	0.822	0.707	0.639
Be	0.988	0.989	0.853	0.860
Cd	1.000	1.000	0.707	0.725
Co	0.997	0.997	0.811	0.841
Cr	0.749	0.825	0.769	0.570
Cu	0.998	0.999	0.883	0.871
Hg	0.994	0.997	0.808	0.850
Mn	0.999	0.999	0.763	0.794
Mo	0.997	0.997	0.779	0.802
Ni	0.992	0.992	0.715	0.753
Pb	0.986	0.985	0.854	0.903
Sb	0.967	0.989	0.702	0.895
Sn	0.990	0.993	0.788	0.736
V	0.924	0.923	0.768	0.832
Zn	0.996	0.996	0.801	0.793

The maximum adjusted R^2 values are in bold. The values of F -test were significant ($P < 0.00005$) for all these models.

were obtained either for linear or for 2nd order polynomial regression functions (Table 4). It can also be presumed from scatterplots of some chemical elements, especially Cr, that the 2nd order polynomial functions might be a more precise model. To check whether linear or 2nd order polynomial model is more appropriate for each chemical element, STATISTICA (version 9) software was used. The polynomial regression procedure from GRM (general regression models) module was chosen using a forward stepwise option (critical values to enter a new member and to remove the included member were equal to 0.05). It was applied several times.

TABLE 5: Summary of consecutive stages of stepwise polynomial regression.

El.	ST	Eliminated points	a_0	a_1	a_2	Min	Max	Q^2	Q_{ext}^2
As	S0	None	0.3632	1.1429		1.10	25.0	0.9933	
	S1	910		1.3257	-0.01417	1.10	15.1	0.9959	0.9321
Ba	S0	None	182	1.1243		7.95	368	0.7662	
	S1	909	172	1.3042		7.95	229	0.7078	0.8962
Be	S0	None	0.2167	1.4473		0.072	2.03	0.9864	
	S1	910	0.2118	1.4581		0.072	1.53	0.9779	0.9995
	S2	910, 872	0.1907	1.5144		0.072	1.33	0.9822	0.9643
Cd	S0	None	0.0123	1.0068		0.066	12.8	0.9993	
	S1	995		1.0290		0.066	0.768	0.9781	0.9995
	S2	995, 992		1.0436		0.066	0.768	0.9859	0.7751
Co	S0	None	0.0628	1.1422		0.58	15.7	0.9961	
	S1	910		1.1397		0.58	15.1	0.9962	0.9976
Cr	S0	None	-1.6570	2.3694	-0.01041	5.58	117	0.7456	
	S1	986, 995		1.7540		24.8	78.2	0.6822	-1.2020
Cu	S0	None	0.186	1.1291	-0.00180	4.66	64.7	0.9981	
	S1	995	0.829	1.0475		4.66	45.4	0.9977	0.9956
	S2	995, 889	0.915	1.0376		4.66	45.4	0.9987	0.9895
Hg	S0	None	4.710	0.8848	0.00073	22.1	265	0.8963	
	S1	995		1.0078		22.1	116	0.9806	0.9894
Mn	S0	None	32.86	1.0340		32.1	1462	0.9988	
	S1	872	29.66	1.0438		32.1	1005	0.9985	0.9995
Mo	S0	None	0.1601	1.0277		0.16	4.42	0.9967	
	S1	867	0.1612	1.0264		0.16	2.37	0.9903	1.0000
	S2	867, 910	0.1457	1.0315		0.16	2.37	0.9965	-155.6
Ni	S0	None	-0.0031	1.1213		1.50	75.0	0.9910	
	S1	872		1.1248		1.50	44.5	0.9821	0.9998
	S2	872, 910		1.1076		1.50	39.3	0.9838	0.9805
Pb	S0	None	1.9085	1.1593		6.81	71.0	0.9632	
	S1	995, 910	2.9985	1.1011		6.81	33.4	0.9790	0.9739
	S2	995, 910, 865	3.2856	1.0959		6.81	33.4	0.9912	0.4736
Sb	S0	None	0.3399	0.7841	0.5071	0.12	2.19	0.5253	
	S1	910	0.2321	1.3008		0.12	1.04	0.8837	0.8547
	S2	910, 889	0.1604	1.5092		0.12	0.65	0.8825	0.8354
Sn	S0	None	0.6464	1.3820	-0.0206	0.53	15.9	0.9943	
	S1	995	0.7000	1.3090		0.53	2.86	0.8716	0.9230
	S2	995, 881	0.5606	1.4285		0.53	2.51	0.9055	0.4721
V	S0	None	-3.88	1.7815		7.66	88.2	0.9119	
	S1	910		1.6942		7.66	71.0	0.8715	0.9986
	S2	910, 909		1.6190		7.66	63.7	0.8669	0.8873
Zn	S0	None	3.1815	1.0472		8.11	214	0.9940	
	S1	995, 910		1.1685	-0.0008	8.11	187	0.9954	0.9863
	S2	995, 910, 961, 909		1.1044		8.11	187	0.9985	0.7843

El.: chemical element, ST: stage of stepwise polynomial regression $y = a_0 + a_1 * x + a_2 * x^2$ (S0: preliminary which S1 and S2: main stages), a_0 , a_1 , a_2 : coefficients, significant are given in bold. Min and Max: lower and upper boundary of the interval of aqua regia contents (Hg: $\mu\text{g}/\text{kg}$, for all other elements: mg/kg), where the equation was obtained. Q^2 : cross-validation parameter and Q_{ext}^2 : external validation parameter.

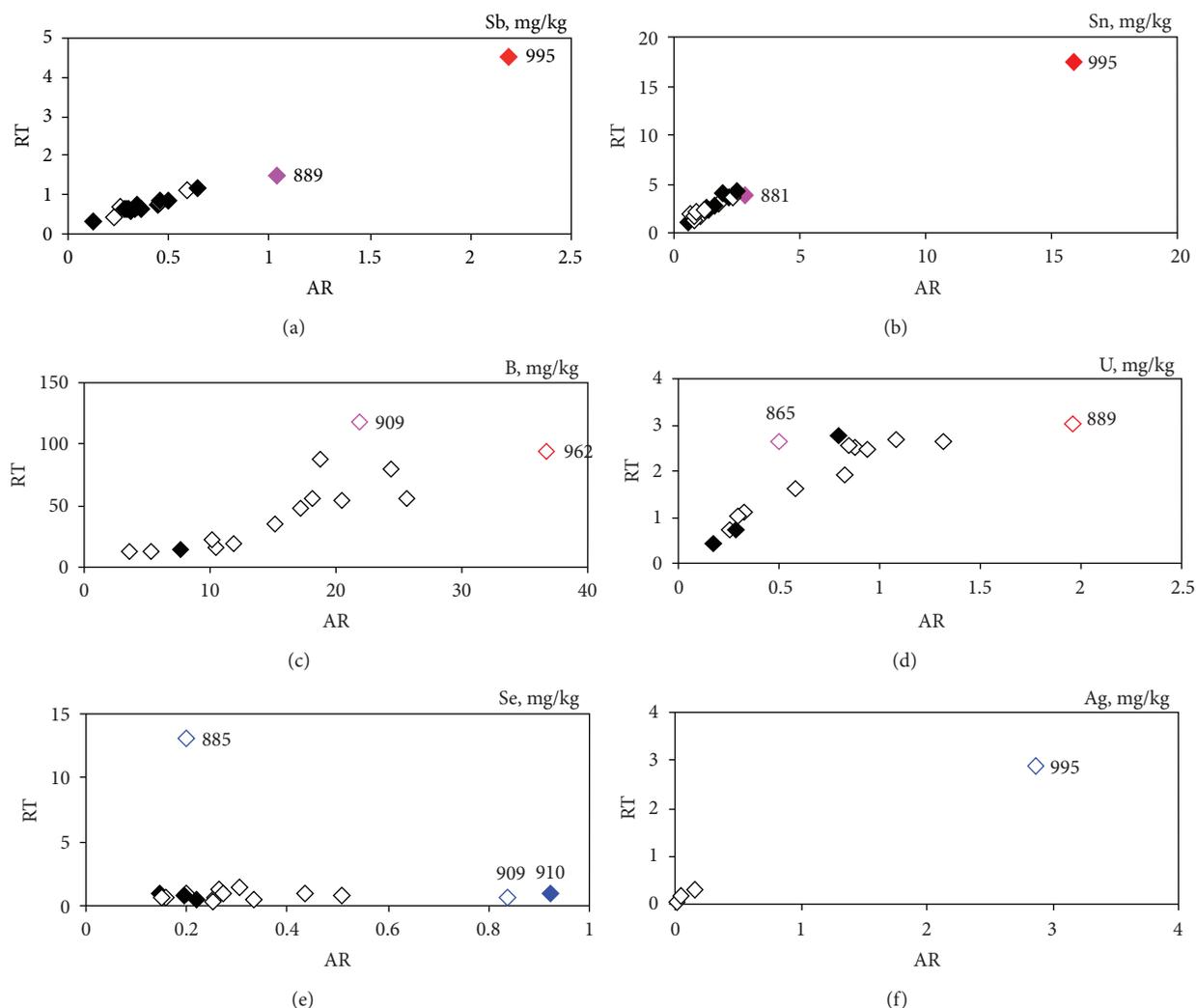


FIGURE 2: Scatterplots of *aqua regia* versus real total contents of less reliably characterised harmful chemical elements. Explanation. Symbol of each trace element is followed by measurement units. Hollow rhombs show samples with less than 8 laboratory measurements of *aqua regia* and real total contents of trace element. High leverage points eliminated from data set after preliminary stage S0 of stepwise polynomial regression are indicated in red; possibly influential points eliminated from the data set in the next stage S1 are indicated in magenta. Outliers for Se and Ag are indicated in blue, since implementation of stepwise polynomial regression for these elements was impossible.

The preliminary stage (S0) was necessary for calculation of leverage values of points for each element [36]. It is obvious from the scatterplots that there are points with high leverage. Since such points might be influential for model, it is expedient to eliminate them from the data set. Different critical values are used for revealing the points with high leverage: $2 * p/n$ [37] and $3 * p/n$ [37, 38], where n is the number of samples and p is the number of parameters in the model. It is obvious that critical values depend on the model. Since our data set is small ($n = 19$), we have chosen the second criteria. So if the linear regression model $y = a_0 + a_1 * x$ (2 parameters) was chosen in the preliminary stage by forward stepwise polynomial regression, the critical value was $6/19 = 0.3158$; meanwhile if the 2nd order polynomial model $y = a_0 + a_1 * x + a_2 * x^2$ (3 parameters) was selected, the critical value was $9/16 = 0.4737$.

During preliminary stage S0 for all 16 elements the linear member was always included during the first step though P values corresponding to F values for inclusion of both linear and quadratic members were lower than 0.05. Such regularity can be explained by higher F values calculated for linear members. Meanwhile during the second step the quadratic member was usually not included due to higher than 0.05 P value of F -test. So forward stepwise polynomial regression resulted in a linear model for most chemical elements (11 from 16), only for Cr, Cu, Hg, Sb, and Sn the 2nd order polynomial model was chosen (Table 5).

To check if coefficients a_0 , a_1 , and a_2 are significant, t -test was used. If P values corresponding to t values are lower than 0.05, the coefficient is significant. For all elements a_1 coefficients were significant both in linear and 2nd order polynomial regression and a_2 coefficients were significant

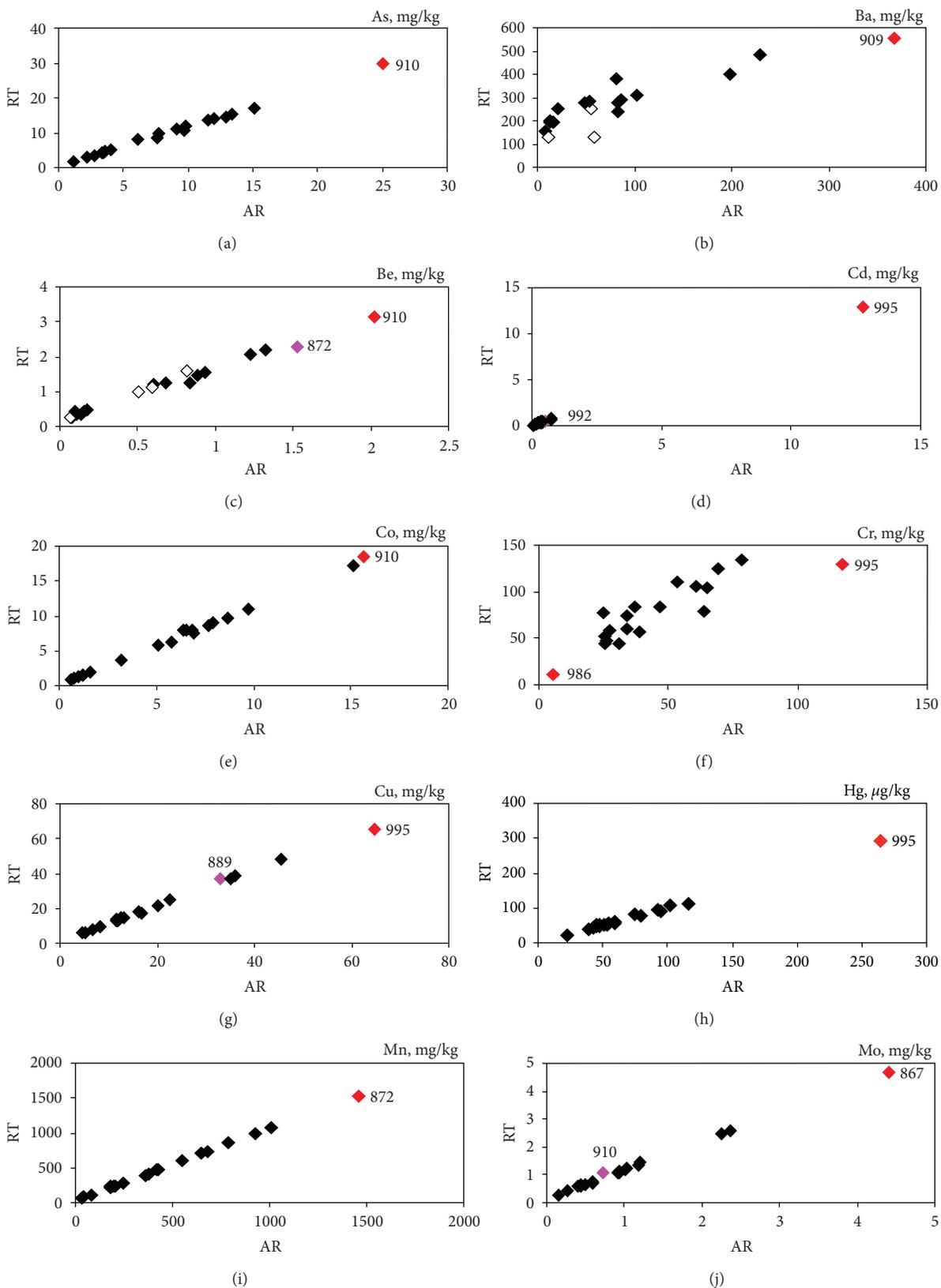


FIGURE 3: Continued.

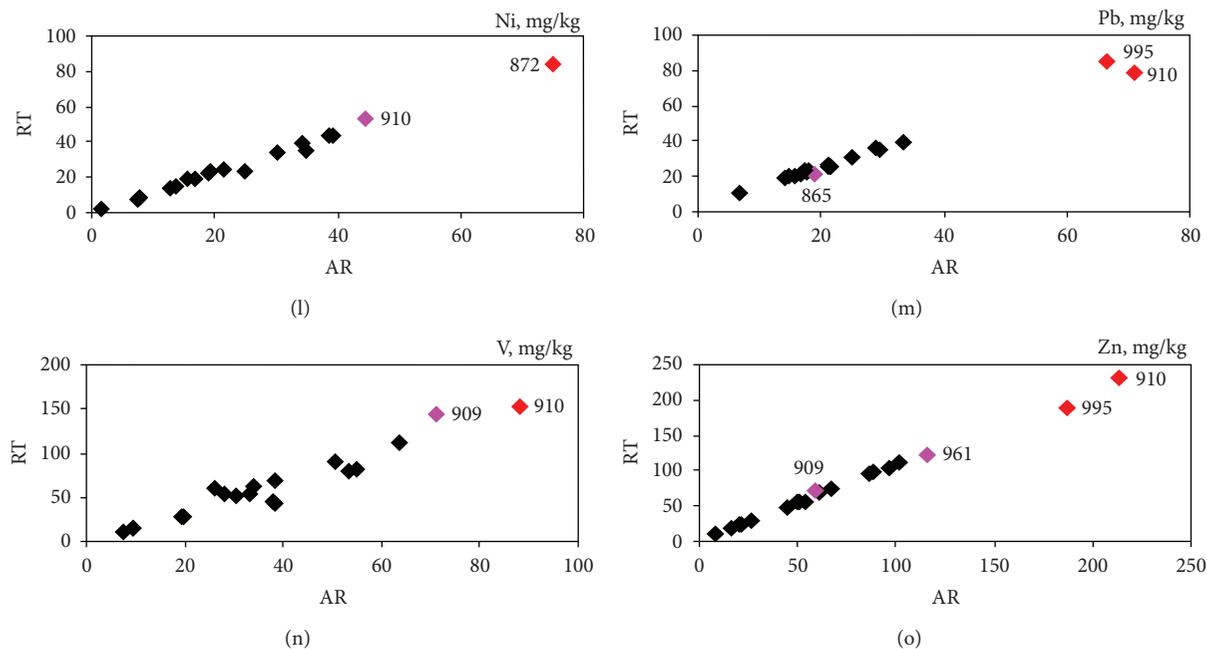


FIGURE 3: Scatterplots of *aqua regia* versus real total contents of more reliably characterised harmful chemical elements. Explanation is given in Figure 2.

for Cr, Cu, Hg, Sb, and Sn. Meanwhile a_0 coefficients were significant only for some chemical elements: Ba, Be, Mn, Mo, and Zn in linear model and Sb and Sn in 2nd order polynomial model.

The preliminary stage S0 resulted in elimination of 1-2 points for each of the 16 elements. The sample No. 995 (sandy soil) was eliminated for 8 elements (Cd, Cr, Cu, Hg, Pb, Zn, Sb, and Sn), No. 910 (clay soil) for 6 elements (As, Be, Co, Pb, V, and Zn), No. 872 (braunerde clay) for 2 elements (Ni and Mn), No. 909 (marshland) for only Ba, No. 867 (forest sandy soil) for Mo. For Cr, there are two samples with high leverage values: one of them with its highest content and the other one with its lowest content. The specificity of soil samples for which high leverage influential measurements were revealed is given in Table 6.

After the preliminary stage, the main part of analysis followed which consisted of two stages. The first stage S1 was applied to a data set without samples eliminated in the preliminary stage. If a_0 coefficient was not significant ($P > 0.05$), the stepwise polynomial regression was repeated without intercept. In this stage the model of Cr, Cu, Hg, Sb, and Sn changed from the 2nd order polynomial to linear indicating that elimination of high leverage points often simplifies the model. On the other hand, the model of As and Zn changed from linear to the 2nd order polynomial regression indicating that reverse process is also possible. The a_0 coefficients in S1 stage were significant for the same elements: Ba, Be, Mn, and Mo, except Zn. However, they became significant also for Cu, Pb, and Sn.

At the end of S1 stage, the data were checked for possibly influential points. They can be either unusual predictor

x -values or unusual y -values (discrepancy values or outliers). STATISTICA software enables to calculate not only leverage values but also characteristics that are helpful for revealing influential data: standardised residuals, studentised residuals, deleted residuals, studentised deleted residuals, Mahalanobis distances, Cook's distances, DFFITS, and standardised DFFITS [39]. We have selected only the last from these characteristics. The reason is that it measures the combination of leverage and discrepancy of observation and that for small to medium sized data sets, cut-off value is simple: if absolute value of standardised DFFITS exceeds 1, the point is possibly influential [40]. Such points were not found for As, Ba, Co, Cr, Hg, and Mn. Meanwhile, for other part of chemical elements possibly influential points were found. Only a part of them was from those in the preliminary stage: sample No. 910 was possibly influential for Ni and Mo, No. 909 for V and Zn, and No. 872 for Be. In addition to them other points were found (Table 5, Figures 2 and 3). For 10 chemical elements the next stage S2 was continued after elimination of possibly influential points.

In S2 stage, the stepwise polynomial regression gave results very similar to S1 stage: the model remained the same for all elements, except Zn; the presence of intercept remained for Be, Cu, Mo, Pb, Sb, and Sn. For Zn, the model changed once more from the 2nd order polynomial without intercept to linear without intercept.

After the main part of the analysis, there were mostly linear equations derived either with intercept (Ba, Be, Cu, Mn, Mo, Pb, Sb, and Sn) or without intercept (Cd, Co, Cr, Hg, Ni, V, and Zn); only for As the 2nd order polynomial regression without intercept was obtained (Table 5).

TABLE 6: Specificity of characteristics of samples for which high leverage and possibly influential measurements were revealed.

No.	Specificity of characteristics
910	Maximum value of CL
909	Maximum value of LOI, high level of CL
872	High levels of CL and Ca
995	Lowest values of CL, LOI and Ca
992	?
986	Minimum value of CL and Ca, low value of LOI
889	Maximum value of Ca
867	Minimum value of Ca and LOI
865	High level of LOI
881	Low value of LOI and high level of Ca
961	?

No.: sample number in WEPAL ISE program; CL, LOI, and Ca are explained in Table 1 where their median values are listed for each sample. ?: the influential component is unknown.

The above-mentioned procedure can formally be applied also for some least reliably characterised chemical elements, that is, B and U aiming to find out high leverage points or possibly influential points and type of model. For B with 15 samples, the procedure resulted in linear regression model without intercept. During preliminary stage, sample No. 962 was eliminated; meanwhile, during S1 stage sample No. 909 was deleted as possibly influential (Figure 2). For U with 15 samples, the procedure resulted in the 2nd order polynomial regression model without intercept. During the preliminary stage, sample No. 889 (soil with high carbonate) was eliminated; meanwhile during S1 stage sample No. 865 (loamy soil) was deleted as possibly influential (Figure 2). For Se and Ag neither linear nor 2nd order model could be applied, because P values to enter were higher than 0.05. However, outliers are clearly seen in the scatterplots and are confirmed by forced application of the linear model despite the fact that both a_0 and a_1 coefficients are not significant ($P > 0.05$).

Cross-validation of results can be done either by leave-one-out or leave-many-out procedures [41, 42]. Due to the small data set the simpler leave-one-out procedure was chosen. The cross-validated correlation coefficient Q^2 was calculated according to the following formula:

$$Q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad (4)$$

where y_i , \hat{y}_i , and \bar{y} are the measured, the predicted and, the averaged values of dependent variable; summing is over all samples. Since the difference $y_i - \hat{y}_i$ comprises deleted residuals the calculation of Q^2 was simplified using deleted residuals, from stepwise polynomial regression. Analogous formula is used for external validation:

$$Q_{\text{ext}}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_{tr})^2}, \quad (5)$$

where y_i , \hat{y}_i , and \bar{y}_{tr} are the measured, the predicted, and the averaged in the training set values and summing is over validation samples which were not used in the training set. We presumed that the samples eliminated either in preliminary stage S0 or in S1 stage can be used for external validation; meanwhile the remaining samples comprise the training set.

Cross-validation was performed after S0, S1, and S2 stages according to measurements in the training set; meanwhile external validation was done only after S1 and S2 stages according to validation set consisting of only several samples which were eliminated in either S1 or S2 stages (Table 5).

Cross-validation in preliminary stage S0 showed that for As, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Sn, V, and Zn (all 16 elements), the initial model was suitable, because Q^2 values were higher than 0.5: for Cd, Mn, Cu, Mo, Co, Zn, As, Ni, and Sn they exceeded 0.99, and for Be, Pb, and V were from 0.90 to 0.99, and for Hg Q^2 was only slightly lower (0.896). The lowest Q^2 values were for Ba (0.766), Cr(0.746), and Sb(0.525).

Cross-validation after stage S1 showed that for 6 chemical elements (Sb, Hg, Pb, As, Zn, and Co) Q^2 value increased in comparison with Q^2 value after S0 stage indicating some improvement of the model; meanwhile for 10 other chemical elements (Sn, Cr, Ba, Cd, V, Ni, Be, Mo, Cu, and Mn) Q^2 value decreased. The improvement is especially obvious for Sb, Hg, and Pb.

The external validation of elements with improvement of the model (As, Co, Hg, Pb, Sb, and Zn) according to samples eliminated due to high leverage values gave high Q_{ext}^2 values ranging from 0.855 for Sb to 0.998 for Co. This fact indicates that elimination of high leverage points was expedient for adjustment of the model; besides, it enables good forecast even in points with high leverage outside the interval of the training set.

The most obvious decrease of Q^2 is for Sn and Cr; they are followed by Ba, Cd, and V, for Ni, Be, Mo, Cu, and Mn the decrease of Q^2 is lower. Since all Q^2 values are higher than 0.5, the model derived in stage S1 is suitable for prediction in the interval without high leverage points.

The external validation of elements with lower Q^2 value in stage S1 compared to S0 gave also high Q_{ext}^2 values ranging from 0.896 for Ba to 1.000 for Cd indicating the possibility to forecast even outside the interval of the training set, the only exception is Cr. For Cr the forecast outside the interval of the training set is not recommended, because Q_{ext}^2 value is negative.

For some chemical elements, S2 stage was implemented; for most of these elements (Sn, Pb, Zn, Cd, Ni, Be, Mo, and Cu) Q^2 value in stage S2 was slightly higher than in stage S1 indicating some improvement of model; only for V and Sb it was slightly lower.

The external validation of elements in stage S2 has shown that Q_{ext}^2 values in S2 stage are always lower than in S1 stage. For Sn and Pb they are lower than 0.47 and for Mo even negative indicating that it is not recommended to predict outside the interval of the training set. For the other 7

chemical elements (Sb, Zn, Cd, V, Ni, Be, and Cu) prediction outside the interval of the training set is possible, because Q_{ext}^2 values range from 0.775 for Cd to 0.990 for Cu.

3.3. Practical Significance of Results. Our results do not confirm the statement of Sastre et al. [11] that the difference between *aqua regia* and *real total contents* of Hg, Cd, Zn, Cu, and Pb in environmental samples is not great. As concerns Pb, our results show that about 20% of its *real total content* is not extracted. Besides, the results of our research show great variability of median degree of extractability for different harmful trace elements (Table 3). Their arrangement according to median percentage of degree of extractability in European soil samples analysed by WEPAL ISE participants from 2006 until the first quarter of 2010 is the following (Table 3): Hg(98), Cd(94), Cu(91), Zn(90), Mn(89), Ni(88), Co(86), As(82), Mo(82), Pb(79), V(61), Sn(57), Cr(56), Ag(54), Sb(53), Be(51), B(46), Se(35), U(35), and Ba(21).

Despite the obvious fact that *aqua regia* extraction differs from digestion with HF in combination with other acids or alkaline melt, the results obtained by *aqua regia* extraction are often denominated as “total contents.”

Controvertial content of some regulatory documents [43] is the subsequence of the tradition to denominate both *real total* and *aqua regia* contents by the same word “total.”

When total contents of a group of chemical elements are used for risk assessment, the term “total content” should be well explained; otherwise due to incomparable results of risk assessment, the problem will be also complicated. For example, in Lithuania soil additive contamination index Z is based on *real total* contents of harmful trace elements used for evaluation of possible health disturbances [35, 44]. Index Z is calculated by summing up the concentration coefficients Kk_i of contaminating elements. The concentration coefficients Kk_i are calculated by dividing the contents of elements in a sample by their background values. The problem of improper determination of index Z arises when the content in numerator is determined in *aqua regia* extract and the content in denominator is *real total content*.

An example is given below to illustrate the problems which are possible during estimation of Z and the category of danger. It is based on Vilnius territory between the works of drills and former works of radio engineering [45]. A new composite sample was taken from one site of this territory in 2008. The site area was about 400 m². It was located 100 m to the east from works of drills in the yard of a dwelling house. *Real total contents* of Ba, Cr, Cu, Mn, Ni, Pb, Se, Sn, U, V, and Zn were determined by EDXRF and of Ag, B, and Co by OAES. They were the following (mg/kg): Ag-0.34, As-3.2, B-24.8, Ba-359, Co-7.18, Cr-53.0, Cu-28.0, Mn-277, Mo-18.7, Ni-12.8, Pb-39.0, Se-1.27, Sn-9.88, U-2.26, V-44.0, and Zn-160. The values of concentration coefficients Kk_i for calculation of Z were obtained using particular background values [35]. The value of Z was 42.8, so site was attributed to dangerous contamination category. Supposing the researchers use *aqua regia* digestion to analyse soil from the same site, the approximate estimates

of *aqua regia contents* in soil of the same site can be obtained according to *real totals* and medians of degree of extractability (Table 3). If these estimates are divided by the same background values [35], other Kk_i values are obtained. Finally, this results in lower value of Z , which is equal to 30.1. So in this case, the site will be attributed to medium dangerous contamination category. Therefore, it is quite possible that the site will not be selected for management. This example shows the necessity to use the identical contents of harmful trace elements both in the numerator and in the denominator for calculating concentration coefficients Kk_i .

The necessity to use identical contents requires to have relevant terminology. In this case, it is useful to remember the definition of total concentration (for inorganics) given in the vocabulary of soil quality, that is, ISO 11074:2005 [46] with the following note: “determination of the total concentration requires use of an instrumental technique such as X-ray fluorescence or a powerful solvent combination such as a mixture of hydrofluoric and perchloric acid, or alkaline melt.” So *real total contents* are total contents.

4. Conclusions

Comparison of *aqua regia* and *real total contents* of Ag, As, B, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Se, Sn, U, V, and Zn (20 elements) in 19 European soil samples analysed by WEPAL ISE revealed the following results.

- (1) *Aqua regia* digestion enables to extract near 98% (median amount) from *real total contents* only of Hg. However, for Cd, Cu, and Zn nearly 10% of *real total content* may be not extracted; for Mn, Ni, Co, As, Mo, and Pb from 11 to 21% of *real total content* may be not extracted; for V, Sn, Cr, Sb, Ag, and Be—from 39% to 49%, for B, Se, U, and Ba—from 54% to 79%. However, presently the equations for prediction can be given for As, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Sn, V, and Zn (16 elements). Data of Ag, B, Se, and U are unsuitable for prediction equations due to low number of samples with a sufficient number of laboratory measurements. Such models can be created in the future when sufficient data are included to the data set.
- (2) Since the data contain high leverage points and possibly influential points, the leave-one-out cross-validation was applied. The cross-validated squared correlation coefficient Q^2 has shown good internal predictivity of *real total contents* of As, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Sn, V, and Zn (16 elements) according to respective *aqua regia* contents by a linear or 2nd order polynomial regression model. Good internal predictivity ($Q^2 > 0.7$) was observed in all stages: preliminary stage S0, stage S1 after elimination of high leverage points, and next S2 stage after elimination of possibly influential points.

- (3) External validation by calculating the values of Q_{ext}^2 according to eliminated samples indicates that extrapolation of prediction for *aqua regia* values higher than upper limit of prediction interval in the final S2 or S1 stages is not recommended for Cr, Sn, Mo, and Pb. On the other hand, such extrapolation is possible for As, Ba, Be, Cd, Co, Cu, Hg, Mn, Ni, Sb, V, and Zn (12 elements).
- (4) The high leverage points and possibly influential points should be explained. Though this is out of scope of the present paper, it has been noticed that the specificity of soil composition (clay content, loss on ignition, and major element contents) should be taken into account.
- (5) For As, Ba, Be, Cd, Co, Cr, Cu, Hg, Mn, Mo, Ni, Pb, Sb, Sn, U, V, and Zn, the difference between the median values of degree of extractability calculated for all data and respective median values calculated according to the data set without high leverage and possibly influential points does not exceed 2%.
- (6) Relevant terminology should always be used in description of analytical methods and results. The contents determined in *aqua regia* should be denominated strictly as *aqua regia* contents; only real total contents can be denominated as total contents.

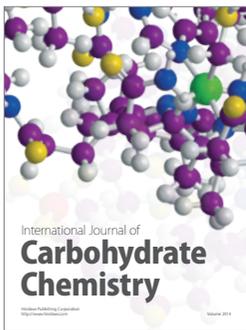
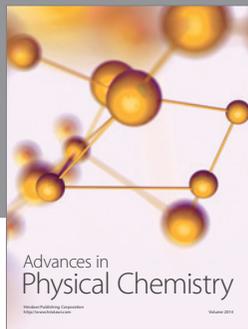
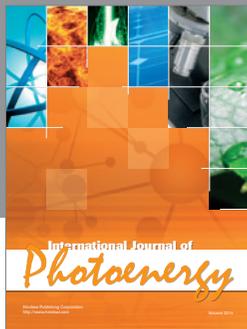
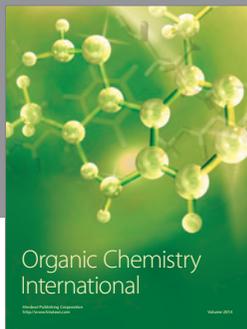
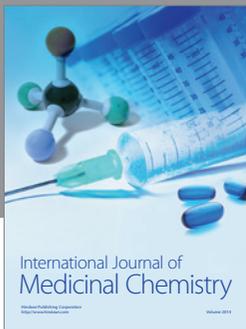
Acknowledgments

This paper was funded by a Grant no. LEK-03/2012 from the Research Council of Lithuania. The authors acknowledge WEPAL working group and especially Bram Eijgenraam for the possibility of participation in the International Soil-Analytical Exchange Program of Wageningen Evaluating Programs for Analytical Laboratories and the permission to use the data for preparing this paper. The authors are also very grateful to Dr. V. Chrastny for the helpful comments and suggestions.

References

- [1] J. Lis, A. Pasieczna, and R. Taraškevičius, "Total and partial extraction of selected elements in soils of the Poland-Lithuania borderland," *Journal of Geochemical Exploration*, vol. 66, no. 1-2, pp. 211-217, 1999.
- [2] Y. K. Soon and S. Abboud, "Cadmium, chromium, nickel," in *Soil Sampling and Methods of Analysis*, M. R. Carter, Ed., pp. 101-108, Lewis, Boca Raton, Fla, USA, 1993.
- [3] EN 13650:2001, International Standard, Soil improvers and growing media, Extraction of aqua regia soluble elements, 2001.
- [4] H. Niskavaara, C. Reimann, V. Chekushin, and G. Kashulina, "Seasonal variability of total and easily leachable element contents in topsoils (0-5 cm) from eight catchments in the European Arctic (Finland, Norway and Russia)," *Environmental Pollution*, vol. 96, no. 2, pp. 261-274, 1997.
- [5] A. L. Horn, R. A. Düring, and S. Gäth, "Comparison of decision support systems for an optimised application of compost and sewage sludge on agricultural land based on heavy metal accumulation in soil," *Science of the Total Environment*, vol. 311, no. 1-3, pp. 35-48, 2003.
- [6] E. Baltrėnaitė, V. Šerevičienė, and D. Paliulis, "Variation of carbon and metal concentration in soil amended with sewage sludge," *Ekologija*, vol. 55, no. 1, pp. 1-8, 2009.
- [7] M. Pueyo, A. Sahuquillo, A. Rigol, J. F. Lopez-Sanchez, and G. Rauret, "A new quality control soil material for monitoring trace metals in accidentally polluted areas," *Analytica Chimica Acta*, vol. 533, no. 1, pp. 41-49, 2005.
- [8] M. García-Delgado, M. S. Rodríguez-Cruz, L. F. Lorenzo, M. Arienzo, and M. J. Sánchez-Martín, "Seasonal and time variability of heavy metal content and of its chemical forms in sewage sludges from different wastewater treatment plants," *Science of the Total Environment*, vol. 382, no. 1, pp. 82-92, 2007.
- [9] E. I. B. Chopin and B. J. Alloway, "Trace element partitioning and soil particle characterisation around mining and smelting areas at Tharsis, Riotinto and Huelva, SW Spain," *Science of the Total Environment*, vol. 373, no. 2-3, pp. 488-500, 2007.
- [10] J. A. Rodríguez Martín, M. L. Arias, and J. M. Grau Corbí, "Heavy metals contents in agricultural topsoils in the Ebro basin (Spain). Application of the multivariate geo-statistical methods to study spatial variations," *Environmental Pollution*, vol. 144, no. 3, pp. 1001-1012, 2006.
- [11] J. Sastre, A. Sahuquillo, M. Vidal, and G. Rauret, "Determination of Cd, Cu, Pb and Zn in environmental samples: microwave-assisted total digestion versus aqua regia and nitric acid extraction," *Analytica Chimica Acta*, vol. 462, no. 1, pp. 59-72, 2002.
- [12] M. Chen, L. Q. Ma, and W. G. Harris, "Baseline concentrations of 15 trace elements in Florida surface soils," *Journal of Environmental Quality*, vol. 28, no. 4, pp. 1173-1181, 1999.
- [13] V. Kadūnas, R. Budavičius, V. Gregorauskienė et al., *Geochemical Atlas of Lithuania*, Geological Survey of Lithuania, Geological Institute, 1999.
- [14] M. Birke and U. Rauch, "Urban geochemistry: investigations in the Berlin metropolitan area," *Environmental Geochemistry and Health*, vol. 22, no. 3, pp. 233-248, 2000.
- [15] C. Reimann, U. Siewers, U. Tarvainen et al., *Agricultural Soils in Northern Europe: A Geochemical Atlas. Geologisches Jahrbuch Sonderhefte*, Reihe D. Heft SD5, Hannover, Germany, 2003.
- [16] R. Salminen, M. J. Batista, M. Bidovec et al., *Geochemical Atlas of Europe. Part 1: Background Information, Methodology and Maps*, vol. 526, Espoo: Geological Survey of Finland, Association of the Geological Surveys of The European Union (EuroGeoSurveys), 2005.
- [17] V. Gregorauskienė, R. Taraškevičius, V. Kadūnas, A. Radzevičius, and R. Zinkutė, "Geochemical characteristics of Lithuanian urban areas," in *Mapping the Chemical Environment of Urban Areas*, C. C. Johnson, A. Demetriades, J. Locutura, and R. T. Ottesen, Eds., chapter 23, pp. 393-409, John Wiley & Sons, 2011.
- [18] C. Reimann, M. Åyräs, V. A. Chekushin et al., *Environmental Geochemical Atlas of the Central Barents Region*, Geological Survey of Norway, 1998.
- [19] W. J. G. M. Peijnenburg, M. Zablotskaja, and M. G. Vijver, "Monitoring metals in terrestrial environments within a bioavailability framework and a focus on soil extraction," *Ecotoxicology and Environmental Safety*, vol. 67, no. 2, pp. 163-179, 2007.
- [20] C. Carlon, Ed., Derivation methods of soil screening values in Europe. A review and evaluation of national procedures towards harmonization. European Commission, Joint Research Centre, Ispra, EUR, 22805-EN, 2007.

- [21] K. J. Andersen and M. I. Kisser, *Digestion of solid matrices*, Desk Study—Horizontal, Eurofins A/S, Denmark, NÖ Umweltschutzanstalt, Austria, 2003.
- [22] ISO 11466:1995, International Standard, Soil quality—extraction of trace elements soluble in agua regia, 1995.
- [23] A. Kabata-Pendias, “Soil-plant transfer of trace elements—an environmental issue,” *Geoderma*, vol. 122, no. 2–4, pp. 143–149, 2004.
- [24] D. Van Dijk and V. J. G. Houba, “Homogeneity and stability of materials distributed within the Wageningen evaluating programmes for analytical laboratories,” *Communications in Soil Science and Plant Analysis*, vol. 31, no. 11–14, pp. 1745–1756, 2000.
- [25] Annual Report 2009, Wageningen Evaluating Programs for Analytical Laboratories, International Soil-analytical Exchange, 2009.
- [26] Annual Report 2005, Wageningen Evaluating Programs for Analytical Laboratories, International Soil-analytical Exchange, 2005.
- [27] Annual Report 2006, Wageningen Evaluating Programs for Analytical Laboratories, International Soil-analytical Exchange, 2006.
- [28] R. Taraškevičius and R. Zinkutė, “Microelement determination methods for mapping purposes,” in *Environmental Geochemical Baseline Mapping in Europe. Abstracts of the 2nd Conference*, pp. 100–103, Geological Survey of Lithuania, Institute of Geology, Vilnius, Lithuania, September 1999.
- [29] M. E. Feinberg, E. Bugner, G. Theiller, V. J. G. Houba, and F. Kadijk, “Expression of the reference value for proficiency tests,” *Journal of Chemometrics*, vol. 9, no. 3, pp. 197–209, 1995.
- [30] M. A. J. Van Montfort, “Statistical remarks on laboratory-evaluating programs for comparing laboratories and methods,” *Communications in Soil Science and Plant Analysis*, vol. 27, no. 3–4, pp. 463–478, 1996.
- [31] W. P. Cofino, I. H. M. Van Stokkum, D. E. Wells, F. Ariese, J. W. M. Wegener, and R. A. L. Peerboom, “A new model for the inference of population characteristics from experimental data using uncertainties. Application to interlaboratory studies,” *Chemometrics and Intelligent Laboratory Systems*, vol. 53, no. 1–2, pp. 37–55, 2000.
- [32] Annual Report 2007, Wageningen Evaluating Programs for Analytical Laboratories, International Soil-analytical Exchange, 2007.
- [33] Annual Report 2008, Wageningen Evaluating Programs for Analytical Laboratories, International Soil-analytical Exchange, 2008.
- [34] Quarterly Report 2010, Wageningen Evaluating Programs for Analytical Laboratories, International Soil-analytical Exchange, 2010.
- [35] HN 60:2004, “Pavojingu cheminiu medžiagu didžiausios leidžiamos koncentracijos dirvožemyje” (Lithuanian), *Valstybės žinios* 41-1357, 2004.
- [36] J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, Irwin, Homewood, Ill, USA, 1985.
- [37] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons, 1980.
- [38] P. F. Velleman and R. E. Welsch, “Efficient computing of regression diagnostics,” *The American Statistician*, vol. 35, no. 4, pp. 234–242, 1981.
- [39] R. R. Hocking, *Methods and Applications of Linear Models. Regression and the Analysis of Variance*, John Wiley & Sons, New York, NY, USA, 1996.
- [40] S. M. A. Khaleelur Rahman, M. Mohamed Sathik, and K. Senthamarai Kannan, “Multiple linear regression models in outlier detection,” *International Journal of Research in Computer Science*, vol. 2, no. 2, pp. 23–28, 2012.
- [41] P. K. Ojha and K. Roy, “Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection,” *Chemometrics and Intelligent Laboratory Systems*, vol. 109, no. 2, pp. 146–161, 2011.
- [42] P. Gramatica, “Evaluation of different statistical approaches to the validation of Quantitative Structure-Activity Relationships,” *Ispra: ECVAM*, 2004.
- [43] “Regulations of ecogeological investigations,” *Valstybės žinios* (State News), 71-2759, 2008 (in Lithuanian).
- [44] J. Kumpiene, E. Brännvall, R. Taraškevičius, Č. Aksamitauskas, and R. Zinkutė, “Spatial variability of topsoil contamination with trace elements in preschools in Vilnius, Lithuania,” *Journal of Geochemical Exploration*, vol. 108, no. 1, pp. 15–20, 2011.
- [45] R. Taraškevičius and R. Zinkutė, “Temporal variability of heavy metal anomalies in topsoil of Naujamiestis district in Vilnius Selected papers,” in *Selected papers, in Proceedings of the 6th International Conference “Environmental Engineering”*, D. Čygas and K. D. Froehner, Eds., vol. 1 of *Environmental Protection, Water Management and Hydraulics*, pp. 292–299, Vilnius Gediminas Technical University Press “Technika”, Vilnius, Lithuania, May 2005.
- [46] ISO 11074:2005, International Standard, Soil quality—Vocabulary, 2005.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

