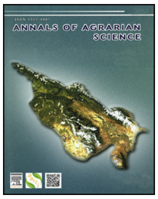
Annals of Agrarian Science xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Annals of Agrarian Science

journal homepage: www.elsevier.com/locate/aasci



Distribution law and evaluation of chemical elements contents in soils below the detection limit

Yu N. Vodyanitskii[a,*], A.T. Savichev[b,c]

[a] *Lomonosov Moscow State University, Faculty of Soil Science, 1, Leninskie Gori, Moscow, 119991, Russia*
[b] *Geological Institute, Russian Academy of Sciences, 7, Pizevski Per, Moscow, 119017, Russia*
[c] *Dokuchaev Soil Science Institute, 5, Pizevski Per., Moscow, 119017, Russia*

ARTICLE INFO

Keywords:
Lanthanides
Halogens
X-ray fluorescence method
Regression equations
Complete metrological errors
Relative regression errors

ABSTRACT

There are samples in soils, where the content of any element falls below the limit of detection (for example, X-ray fluorescence analysis). To reveal a very low element content, we can achieve by regression analysis to obtain the dependence of its content from a common chemical analogue in profile. The most effective approach is the application of power regression equation, connecting low-Clarke lanthanides contents with high-Clarke lanthanides contents and low-Clarke bromine content with the high-Clarke halogen chlorine in soils.

There is a gradual transition from the normal distribution of high-Clarke lanthanides to lognormal distribution in low-Clarke lanthanides. The power regression approach allowed us to separate the regression relative error from the total metrological error. Estimation of the content of dispersed elements in the soil below the detection limit is limited by the magnitude of the regression relative error, and an estimation of the regression relative error is important for the construction of geochemical relationships of the elements.

Introduction

The chemical composition of soils is the basis of any serious soil analysis. Many agrochemical tasks are solved on the basis of these data: related with plant nutrition and soil contamination. However, when using this total chemical composition of soils, some questions have accumulated, which require resolution. One of the problems is connected with the study of micro- and nanoelements.

Occasionally there are samples in the soil profile, where the content of an element is reduced to below the limit detection of this analysis method. It is known, that by using the rapid and low-cost X-ray fluorescence (XRF) method, it is impossible to identify the element content below 3–4 mg/kg [1–3]. In recent years, this method was modified, using the so-called X-ray radiometric analysis (XRR) [4]. It is possible to expand the range of available trace elements, but detection limit is the same. There are similar limitations in other methods, for example, mass spectrometry with inductively coupled plasma. Arsenic is impossible to determine when the content < 1 mg/kg, Nb < 0.1 mg/kg, molybdenum < 0.05 mg/kg [5]. As a result, there are gaps in analytical tables, which impede statistical processing of results. This concerns not only the total chemical composition, but also the fractional composition.

Both geochemists and agricultural chemists face constantly with the problem of gaps in the database, especially analyzing the effect of lanthanides as micronutrients [6]. Probably, the spaces in the table hide different concentrations of rare elements.

Researchers have proposed a variety of statistical methods for the definition of element concentrations below detection limits. These methods are used in the inductively coupled plasma mass spectrometry (ICP-MS). The detection limits for this method are by several orders of magnitude lower than for XRF and its variant XRR. Elements, inaccessible for ICP-MS, are more scattered, and they are distributed in the soil and rocks not normally, but lognormally. According to this fact, the researchers offer a variety of computer programs, based on the logarithmic transformation [7,8].

In XRF and XRR the detection limit is much higher, and their problems are different, they are associated with the identification of the more common elements, the content of which is below 4.5 mg/kg. Among them, some of the lanthanides (Pr, Sm) and a halogen like Br. Their distribution in soils corresponds to the normal rather than the lognormal.

Therefore, to fill the gaps, obtained by using XRF-analysis, we can use simple statistical methods. One of them is correlation analysis, when the content of a rare element is below the detection limit in only a

---

Peer review under responsibility of Journal Annals of Agrarian Science.

[*] Corresponding author.
E-mail address: yu.vodyan@mail.ru (Y.N. Vodyanitskii).

https://doi.org/10.1016/j.aasci.2017.12.011
Received 11 October 2017; Accepted 12 December 2017
1512-1887/

Please cite this article as: Vodyanitskii, Yu.N., Annals of Agrarian Science (2018), https://doi.org/10.1016/j.aasci.2017.12.011

portion of the data array. Then we can get the regression equation for connection of this rare element with more prevalent its chemical analogue. For example, we can fill up a database containing low Clarke lanthanides: Pr, Sm (their soil Clarkes are 7 and 4.6 mg/kg, respectively [9]), if it is known the contents of the more common of chemical analogues: Y, La, Ce, Nd (their soil Clarkes are 23, 27, 56.7 and 26 mg/kg, respectively [9]).

Similarly, when it is known contents in the portion of bromine array (its soil Clarke is 10 mg/kg), we can supplement the information about it, based on distribution data of another halogen - chlorine (its soil Clarke is 300 mg/kg).

Objective: to apply regression analysis in order to calculate contents of chemical elements below the detection limit for methods XRF and XRR.

Method of calculation

Calculation of missing data is based on the use of regression equations for depending the content of analyzed elements on element content with greater Clarke or even less Clarke, but which causes fewer problems with its definition in the soil. For example, samarium (Sm) at low concentrations in the soil can not be detected by the XRR. But other lanthanide, lanthanum (La) has soil Clarke five times higher. Significant statistical association Sm = f (La) in the soil profile can be detected due to the chemical properties similarity of the lanthanides. Using the regression equation for the initial concentrations, we can extrapolate the approximate content of samarium for the sample with a known content of lanthanum. In order to improve the accuracy of calculation is necessary to repeat the same operation with the regression equation between samarium content and other chemical elements, and to get a few more calculated values of the samarium content, then to calculate its average content and to determinate the error.

The proposed regression calculation method is suitable only for those elements, for which are defined chemical and geochemical analogues in the soil; this fact ensures high correlation coefficients between the elements. The regression method can be applied to calculate contents of low-Clarke lanthanides (Sm, Pr), if there is data on the distribution in profile of high-Clarke (Y, La, Ce). It can also be used to calculate the content of low-Clarke bromine, if there is data about the content of high-Clarke chlorine. The strong affinity of elements within group of lanthanides and group of halogens allows such calculations. In some soils, such correlations are sufficiently reliable, making it possible to fulfill our calculations.

Thus, there is a certain group of chemical elements (lanthanides, halogens) in soils, the contents of which can be calculated using the regression method.

The success of this procedure depends on the adequacy of the regression equation. We consider two types of regression equations: linear and power. We suppose, that there are two arrays of measured concentrations of elements « *x*» and «*y*»:

$x_{meas\ i}$ and $y_{meas\ i}$,

where the subscript i takes values $i = 1, 2, 3, \ldots, N$. We find the functional relationship between the arrays. To do this, we introduce an array of calculated concentrations for the element « *y*»: $y_{calc\ i} = f\ (x_{meas\ i})$.

The linear regression equation. The relationship between arrays $x_{meas\ i}$ and $y_{calc\ i}$ is as follows:

$$y_{calc\ i} = \mathrm{A}\ x_{meas\ i}\ +\ \mathrm{B}; \tag{1}$$

free term B should be minimal. Otherwise, when extrapolating the equation to lower concentrations, it is possible an error, that is easy to identify. In fact, the detection of low-Clarke lanthanide content, for example Sm, may give an excessively high value, higher than the sensitivity of the device, which is absurd.

Since extrapolation is conducted in the direction of very low concentrations (in the limit - zero), then the model should reflect this. For example, in the polynomial (in the particular case, linear) model, the free term should be close to zero, and in the power model the condition of extrapolation to zero is obtained automatically.

The power regression equation. In this case, the relationship between arrays $x_{meas\ i}$ and $y_{calc\ i}$ is as follows:

$$y_{calc\ i} = \alpha\ x_{meas\ i}{}^{\beta}, \tag{2}$$

where the degree index β is slightly different from unity. The obvious advantage of this approach is the fact, that the calculated array $y_{calc\ i}$ vanish automatically to zero with decreasing values $x_{meas\ i}$, which is not necessarily using linear regression. The parameters α and β are determined by minimizing the expression:

$$\sum_{i=1}^{N} [\ (\ y_{calc\ i} - y_{meas\ i})^2 / y_{meas\ i}] = \sum_{i=1}^{N} [(\alpha\ \ x_{meas\ i}\beta - y_{meas\ i})^2 / y_{meas\ i}] = \min. \tag{3}$$

Minimizing the expression (3) indicates that the unknown parameters are searched by minimizing the *K*-factor ($\Delta\ y/\sqrt{y}$), instead of minimizing absolute ($\Delta\ y$) or relative deviation ($\Delta\ y/y$), as is customary in standard mathematical software.

The *K*-factor for estimating the error in measuring the concentration of elements is used in the methods of XRF and XRR [4]. The use of the *K*-factor as an error measure has a very strong advantage over the relative error. Indeed, in the case of a wide range of changes in the concentration of the element, it is impossible to estimate the accuracy of the studies performed with a constant value of the relative error: the relative error increases with decreasing concentration. The *K*-factor remains constant in these conditions, which gives an indisputable advantage for the optimal construction of calibration curves.

The physical basis for the *K*-factor application is that in X-ray fluorescence analysis line intensities are measured by summation of the pulses, which is a Poisson process, where the dispersion of the quantity is linearly related to the square root of the mean. In our opinion, the use of *K*-factor is desirable in all methods of so-called current spectrometry (in atomic absorption, emission with inductively-coupled plasma, etc.).

Thus, an accurate error value is needed to evaluate the possibility of applying a correlation analysis for this purpose. For example, under the expected content of an estimated element, using the correlation analysis, about 3 mg/kg, *K*-factor should not exceed 0.7–0.8; while the relative error will not exceed 30–40%, otherwise there will be great uncertainties in the construction of the geochemical relationships of the elements.

With the known degree index β, coefficient α is calculated as:

$$\alpha = \left(\sum_{i=1}^{N} x_{meas\ i}\beta \right) \Big/ \left[\sum_{i=1}^{N} (\ x_{meas\ i} 2\beta / y_{meas\ i}) \right]. \tag{4}$$

The procedure for calculating the parameters α and β is carried out by the following algorithm. The degree index β is changed in small step (for example, 0.01) from unity upwards and downwards and at each step coefficient α is calculated by formula (4), then elements of array $y_{calc\ i}$ are calculated using formula (2), and the *K*-factor dispersion is calculated Δ_K:

$$\Delta K\ = \left\{ \frac{1}{N-1}\ \sum_{i=1}^{N} [\ (\ y_{calc\ i} - y_{meas\ i}) 2 / y_{meas\ i}] \right\}^{1/2}. \tag{5}$$

The procedure for changing the degree index β is continued until a minimum value of *K*-factor dispersion by formula (5). Then the determination coefficient R^2 between arrays $y_{calc\ i}$ and $y_{meas\ i}$ is calculated for optimally selected degree index β.

The mathematical software « Excel » holds a power regression, but it carries out the minimization of absolute deviation, so the correlation coefficients obtained are lower than in our proposed procedure.

Table 1

Contents of the lanthanides in peat-podzolic and peat soils and peat podburs of Khibiny-Lovozero province.

Horizon	Depth, cm	Y	La	Ce	Pr	Nd	Sm
Peat-podzolic soil, profile 10							
E	12–18	8	16	24	4	8	–
BT	18–28	10	17	27	2	7	–
C	28–47	15	30	46	4	15	–
Peat-podzolic soil, profile 11							
T	0–8	23	57	97	13	30	8
E	8–10	8	20	30	3	7	–
BT	10–29	12	21	33	3	6	–
C	29–49	12	25	44	3	11	–
Peat podbur, profile 12							
T2	6–18	16	26	38	5	12	5
BHF	18–25	13	25	38	5	10	3
Peat soil, profile 1							
T2	32–40	107	171	307	22	87	19
Sod podbur, profile 2							
A1	5–12	35	61	120	14	44	11
BHF	12–40	61	108	198	13	57	10
BC	40–50	43	97	185	12	59	10

As examples we use data about contents of certain elements, contents of which in the soil profiles suffers incomplete data. In one or more samples their contents were below the detection limit. Let us turn to specific examples.

Evaluation of samarium content depending on high-Clarke lanthanides in soils of Khibiny-Lovozero Province.

The original contents of lanthanides in the peat-podzolic soils in the Khibiny-Lovozero province according to data [4] are given in Table 1. We were able to determine the Sm-contents in seven samples; these contents ranged from 3 to 19 mg/kg. The average content of samarium is 9.4 mg/kg. In another seven soil samples Sm-contents could not be detected. But in all 14 samples were found La-contents, which ranged from 16 to 171 mg/kg.

Linear regression. For $N = 7$, the equation depending of the Sm-content on the contents of high-Clarke lanthanides (Ln) in peat-podzolic and peat soils and peat podburs are presented in Table 2. In calculating the regression equations we used the original data on the content of samarium and other lanthanides. Then these equations, by extrapolation, were used to calculate the Sm-contents in the horizons, where due to the smallness, they could not be detected instrumentally.

Table 2 also shows the calculated values of the Sm-contents, for cases where the contents of high-Clarke lanthanides in horizons reaches a maximum. The calculated values of the Sm-content raises questions in the horizons, where they were based on the maximum values of high-Clarke lanthanides. For example, in the horizon C with a peat-podzolic soil (profile 10), where the Sm-contents are not detected because of the smallness, its calculated content, based on Y-content (15 mg/kg), reaches $Sm_{max} = 5.4$ mg/kg; it seems very strange. Indeed, the used device allows to determine the content of a chemical element more than 3–4 mg/kg and hence, in this horizon the expected content of 5.4 mg/kg could have been recorded. Such a strange situation was found, when the content of samarium calculated using functions Sm = f (La, Nd), when the content of calculated samarium was awkward high: 5.0 and 4.6 mg/kg. Obviously, the extrapolation of regression equations gives an error.

Table 2

Linear equations for depending Sm-content, calculated from the measured contents of high-Clarke lanthanides (Ln) in peat-podzolic and peat soils and peat podburs of Khibiny-Lovozero Province.

Ln	Equation	R^2	Ln_{max}	Sm_{max}
Y	$Sm_{calc} = 0.146\ Y_{meas} + 3.21$	0.881	15	5.4*
Ce	$Sm_{calc} = 0.0504\ Ce_{meas} + 1.224$	0.879	46	3.5
La	$Sm_{calc} = 0.0924\ La_{meas} + 2.235$	0.874	30	5.0*
Nd	$Sm_{calc} = 0.1753\ Nd_{meas} + 1.939$	0.898	15	4.6*

Note: * - too high content of samarium calculated.

Table 3

The power regression equation of the relationship between high-Clarke lanthanides in the soils of the Khibiny-Lovozero Province.

Horizon	Y_{meas}	$Y_{calc} = 0.3747 \cdot (La_{meas})^{1.08}$	$Y_{calc} = 0.2699 \cdot (Ce_{meas})^{1.00}$
Peat-podzolic soil, profile 10			
E	8	7.5	7.1
BT	10	8.0	8.0
C	15	14.8	14.2
Peat-podzolic soil, profile 11			
T	23	29.5	28.8
E	8	9.5	8.9
BT	12	10.0	9.8
C	12	12.1	13.1
Peat podbur, profile 12			
T2	16	12.6	11.3
BHF	13	12.1	11.3
Peat soil, profile 1			
T2	107	96.7	91.5
Sod podbur, profile 2			
A1	35	31.8	35.6
BHF	61	58.9	58.8
BC	43	52.4	54.9
Dispersion of the relative regression K_{reg}-factor		0.739	0.875
Determination coefficient, R^2		0.974	0.957

Let's see what the error of linear regression equations. The reason for the error is obvious in the high values of the constant term in the equation of direct dependence. Indeed, extrapolating content of high-Clarke lanthanide to a hypothetical value Ln ~0 kg/mg, we have quite significant quantities of samarium content of 3.2, 2.2 and 1.9 mg/kg using the equations due to the Y, Ce and Nd, respectively. This is contrary to common sense, because taking into consideration the geochemical community of lanthanides, in a material with the absence of high-Clarke lanthanides, we cannot count on significant content of low-Clarke samarium. Thus, the linear regression equation Sm = *f* (Y, Ce, La, Nd, Pr) cannot be used to extrapolate data.

Power regression. Table 3 shows the application of power regression for high-Clarke lanthanides Y, La, Ce. It is seen that the determination coefficient R^2 reaches high values (~0.95–0.97). This shows the high accuracy of lanthanides determination by X-ray fluorescence and X-ray radiometric methods; this fact can reliably identify geochemical correlation between these elements. For low-Clarke lanthanides (Tables 4 and 5) Nd and Pr determination coefficients is slightly below: ~0.91–0.98 (Nd) and ~0.85–0.88 (Pr). The lower the content of the element, so the lower the determination coefficient, that indicates an increase in the relative measurement errors. It should be noted, that the degree index β in the regression equation for high-Clarke lanthanides exceeds unity slightly, while at the transition to low-Clarke lanthanides it becomes less than unity.

The lower the lanthanide Clarke, the greater the deviation becomes (Tables 5 and 6). One of the reasons for this behavior of the degree index is in the gradual transition from the normal distribution for high-Clarke lanthanides, to the lognormal distribution for low-Clarke ones. The deviation of the degree index β from unity in the regression equation between two elements can serve as a measure of the difference between the distributions of these elements. In the case of the regression equation between the high-Clarke and low-Clarke lanthanides, the deviation of the degree index β from unity is a measure of the distribution deviation of the low-Clarke lanthanide from the normal one. Obviously, for logarithms of concentrations, the regression relationship will always be linear.

The application of power regression for Sm (Table 6) to the measured part of the samples shows a fairly high correlation coefficient ~ 0.86–0.89, this fact made possible to evaluate the content of the

Table 4
The power regression equation of the relationship between the neodymium and high-Clarke lanthanides in the soils of the Khibiny-Lovozero Province.

Horizon	Nd_{mean}	$Nd_{calc} = 0.3707 \cdot (La_{meas})^{1.09}$	$Nd_{calc} = 0.2678 \cdot (Ce_{meas})^{1.02}$	$Nd_{calc} = 0.9672 \cdot (Y_{meas})^{0.99}$
Peat-podzolic soil, profile 10				
E	8	7.2	6.8	7.6
BT	7	7.7	7.7	9.4
C	15	14.3	13.9	14.1
Peat-podzolic soil, profile 11				
T	30	28.8	28.5	21.6
E	7	9.2	8.6	7.6
BT	8	9.7	9.5	11.3
C	11	11.7	12.7	11.3
Peat podbur, profile 12				
T2	12	12.2	10.9	15.0
BHF	10	11.7	10.9	12.2
Peat soil, profile 1				
T2	87	95.2	92.2	98.8
Sod podbur, profile 2				
A1	44	31.0	35.4	32.7
BHF	57	57.7	58.9	56.6
BC	59	51.3	55.0	40.1
Dispersion of the relative regression K_{reg}-factor		0.744	0.543	1.131
Determination coefficient, R^2		0.963	0.984	0.915

Table 5
The power regression equation of the relationship between praseodymium and high-Clarke lanthanides in the soils of the Khibiny-Lovozero Province.

Horizon	Pr_{meas}	$Pr_{calc} = 0.240 \cdot (La_{meas})^{0.88}$	$Pr_{calc} = 0.1947 \cdot (Ce_{meas})^{0.82}$	$Pr_{calc} = 0.5080 \cdot (Y_{meas})^{0.82}$
Peat-podzolic soil, profile 10				
E	4	2.8	2.6	2.8
BT	2	2.9	2.9	3.3
C	4	4.8	4.6	4.7
Peat-podzolic soil, profile 11				
T	13	8.4	8.3	6.6
E	3	3.4	3.2	2.8
BT	3	3.5	3.4	3.9
C	3	4.1	4.3	3.9
Peat podbur, profile 12				
T2	5	4.2	3.8	4.9
BHF	5	4.1	3.8	4.2
Peat soil, profile 1				
T2	22	22.1	21.3	23.4
Sod podbur, profile 2				
A1	14	8.9	9.9	9.4
BHF	13	14.8	14.9	14.8
BC	13	13.4	14.1	11.1
Dispersion of the relative regression K_{reg}-factor		0.662	0.665	0.741
Determination coefficient, R^2		0.880	0.885	0.850

element in the samples, where Sm has not analytically been identified.

A separate issue is the estimation of the accuracy of extrapolated values. First, let us pay attention to the following fact. Earlier, the determination of the error in the X-ray radiometric measurement of the lanthanides concentrations showed that the value of the *K*-factor is of the order of unity [4]. However, for these soils, as can be seen from Tables 3–6, the *K*-factors of the relative regression dependences of the lanthanides are less than unity and sometimes very significantly (the only exception is the Nd-Y dependence in Table 4). But the magnitude of these *K*-factors includes both an error in the measurement of concentrations and an "error" in the geochemical connection of these elements, due to the lack of an exact functional relationship between a pair of compared elements that we do not know.

Table 6
The power regression equation of the relationship between samarium and high-Clarke lanthanides in the soils of the Khibiny-Lovozero Province.

Horizon	Sm_{meas}	$Sm_{calc} = 0.3299 \cdot (La_{meas})^{0.77}$	$Sm_{calc} = 0.3000 \cdot (Ce_{meas})^{0.70}$	$Sm_{calc} = 0.6945 \cdot (Y_{meas})^{0.70}$
Peat-podzolic soil, profile 10				
E		2.8	2.8	3.0
BT		2.9	3.0	3.5
C		4.5	4.5	4.6
Peat-podzolic soil, profile 11				
T	8	7.4	7.4	6.2
E		3.3	3.2	3.0
BT		3.4	3.5	4.0
C		3.9	4.2	4.0
Peat podbur, profile 12				
T2	5	4.1	3.8	4.8
BHF	3	3.9	3.8	4.2
Peat soil, profile 1				
T2	19	17.3	16.5	18.3
Sod podbur, profile 2				
A1	11	7.8	8.6	8.4
BHF	10	12.1	12.1	12.3
BC	10	11.2	11.6	9.7
Dispersion of the relative regression K_{reg}-factor		0.558	0.552	0.545
Determination coefficient, R^2		0.879	0.867	0.893

There is no contradiction between these figures. A *K*-factor of the order of unity is a complete metrological error, let's call it the metrological K_{metr}-factor. It is known, the total metrological error (K_{metr}-factor) consists of two factors: systematic error (K_{sys}) and reproducibility errors (K_{repr}). The relative regression error (K_{reg}-factor of the mutual regression in Tables 3–6) reflects the relative error of measuring the concentrations of lanthanides (relative to each other), which consists of a reproducibility error (K_{repr}) and an "error" due to the incomplete functional geochemical connection of these elements (geochemical K_{geo}-factor). As can be seen, K_{reg}-factor is much less than unity within the framework of the method under consideration. Indeed, the nature of the systematic error is the same for all elements of the method in question and it will not affect the relative regression error, as well as the determination of the content of the element below the detection limit and the construction of geochemical relationships and correlations between the elements, regardless of whether this element is defined directly by this method or by a regression method.

On the other hand, it is necessary to reveal geochemical relations between the elements determined by various analytical methods with caution, since the nature of the systematic error for different methods may be different. Within the framework of the XRF- and XRR-methods, it is hardly possible to reveal the true geochemical "error": for lanthanides, it appears to be much less than the error of reproducibility, since K_{reg} is obtained in the order of K_{repr}. In order to identify the true geochemical "error" in the analysis of lanthanides, a method with a much smaller K_{repr} (for example, ICP MS-method) is required.

Thus, K_{metr}-factor of the total metrological error of the extrapolated data is in the region of unity, while K_{reg}-factor, which is extremely important in determining the content of the element below the detection limit, and for constructing the geochemical relations of the lanthanides, is 0.7–0.8. As noted above, such a requirement for *K*-factor is just necessary to evaluate the content of elements at concentrations of about 3 mg/kg. Note, that the separation of the relative regression error from the total metrological error became possible only with the use of the power regression approach.

Evaluation of bromine content based on the chlorine content in brown soils of Sakhalin Island.

Table 7
Power regression relationship between contents of chlorine and bromine in the brown soils of Sakhalin Island.

Horizon, depth, cm	Cl_{meas}	Br_{meas}	$Br_{calc} = 0.1395 \cdot (Cl_{\ meas})^{0.77}$
Typical brown soil, profile 3			
O, 0-6	2060	46	49.7
AY1, 6-15	230	16	9.2
AY2, 15-26	150	10	6.6
BM, 26-46	140	8	6.3
B, 46-93	100	5	4.8
BC, 63-90	240		9.5
Gleyic brown soil, profile 1			
O, 0-7	3770	94	79.1
AY1, 7-14	1900	34	46.7
AY2, 14-25	1680	64	42.4
BM, 25-40	250	41	9.8
B1g, 40-63	300	14	11.3
B2g, 63-85	170	5	7.3
Cg, 85-105	230	8	9.2
Podzolized brown soil, profile 4			
O, 0-7	2300	49	54.1
AYe, 7-16	300	5	11.3
Be, 14-24	130	11	5.9
BM1, 24-45	100		4.8
B2, 45-67	160		6.9
C, 67-85	130	5	5.9
Dispersion of the relative regression K_{reg}-factor			1.862
Determination coefficient, R^2			0.827

In the southern part of Sakhalin Island on the east coast, thanks to the proximity of the ocean, brown soils are highly enriched by halogens [2]. This applies both chlorine and bromine. In the brown gleyic soils bromine was be identified in all horizons. But in a typical brown soils in the bottom horizon BC the Br-content was lower than detection limit. In podzolized brown soils the bromine content was too low in horizons BM1 and BM2. Let us try to evaluate the bromine content in the three brown earth.

Linear regression. Initially we made linear regression equation f = Br (Cl), using the data about halogens in all three brown soils. There were obtained the following regression equation:

$$Br_{calc} = 0.0218 \cdot Cl_{meas} + 6.844,\ R^2 = 0.838.$$

As can be seen, the free term in the equation is too large in order to a linear equation can be used.

Power regression. The power approach allows to establish a functional relationship between the chlorine and bromine contents (Table 7) and interpolate unmeasured bromine content. It should be noted, that this approach reveals the disagreement in the profile distribution of bromine and chlorine. This is clearly reflected in the horizons AY2 and BM in brown soils, where calculated values of bromine is 1.5–4 times lower than the actual. Such deviations were caused by the abnormal accumulation of bromine in these horizons, violating the "typical" ratio Cl/Br.

We could not evaluate the bromine content in horizon BC of typical brown soil and in horizons BM1 and B2 of podzolized brown soil. It turned out to be overpriced design content: 9.5 mg/kg, it cannot be considered credible, as XRF-method fixes such concentrations. The situation with podzolized brown soils is slightly better: a bromine content (4.8 mg/kg) in the horizon BM1 can be considered as plausible. Thus, the two halogens (bromine and chlorine) have not entirely consistent behavior in the profile of brown soil, which makes it difficult to use a regression to estimate the dependence of bromine content below the detection limit.

Conclusion

1. There are samples in soil profile, where the element content is reduced to below the detection limit of applied analysis method. When using the rapid and low cost X-ray fluorescence analysis, we cannot reveal the contents of the elements below 3–4 mg/kg. But in some soils we can reveal low levels of the elements by regression analysis, due to their chemical connection with analogous elements with higher contents in a common profile.

   So, using power regression equations can fill in the gaps in the contents of low Clarke lanthanides, depending on contents of high-Clarke lanthanides in soil profile. Similarly, we can calculate the content of low-Clarke Br, receiving the equation, depending on the content of its high-Clarke analogue - Cl.
2. The use of the power regression equations for the example of lanthanides made it possible to estimate the gradual transition of the normal distribution characteristic of high-Clarke lanthanides to the lognormal distribution for low-Clarke ones. An estimation of the distribution deviation from the normal one can be the deviation of the degree index from unity in the regression equation. With a decrease in the Clarke, the degree of deviation from the normal distribution increases.
3. The difference between the total metrological error in determining the concentration of an element and the error of the regression coupling between the elements is established using the example of lanthanides. As a measure of error, two *K*-factors are taken: the metrological K_{metr}-factor and the K_{reg}-factor of the relative regression coupling between the elements. The use of power regression equations made it possible to separate the error of the relative regression coupling between the elements and showed that the latter in the framework of this method is substantially lower than the total metrological error. Estimation of the contents of dispersed elements in the soil below the detection limit is possible only at a low value of the relative regression error, and for the construction of geochemical relationships of elements, it is important to estimate the relative regression error.
4. Using the power equation regressions yielded plausible values for the samarium content in the soils of the Khibiny-Lovozero Province, where its content was below the detection limit by XRF and XRR-methods. Success was ensured by the fact, that lanthanides in these soils exhibit not only chemical but geochemical affinity.

   The situation with the expectation of bromine content in soils is more difficult. Not in all brown soils of Sakhalin Island we can obtain adequate equation between bromine and chlorine. Possible cause is a various geochemical behavior of halogens in these soils.

References

[1] Yu.N. Vodyanitskii, N.D. Kiseleva, O.G. Lopatovskaya, A.T. Savichev, Geochemistry of carbonatic/sulphatic soils in the Angara region, Russia, J. Ann. Agrar. Schriftenr. 14 (2) (2016) 140–151.

[2] Yu.N. Vodyanitskii, D.V. Manakhov, A.T. Savichev, Macro- and microelements, including rare earths, in some soils of Sakhalin Island, Eurasian Soil Sci. 48 (10) (2015) 1210–1221.

[3] Yu.N. Vodyanitskii, A.T. Savichev, S.Ya Trofimov, E.A. Shishkonakova, Accumulation of heavy metals in oil-contaminated peat soils, Eurasian Soil Sci. 45 (10) (2012) 977–982.

[4] A.T. Savichev, Yu.N. Vodyanitskii, Improvement of the X-ray radiometric method of lanthanides' diagnostics in soils, Eurasian Soil Sci. 43 (7) (2012) 665–674.

[5] D.M. Levitan, M.E. Schreiber, R.R. Seal, R.J. Bodnar, J.G. Aylor, Developing protocols for geochemical baseline studies: an example from the Coles Hills uranium deposit, Virginia, USA, Appl. Geochem. 43 (2014) 88–100.

[6] Yu.N. Vodyanitskii, Soil lanthanides and their influence on plant, Agrochemistry 4 (2012) 84–96 (in Russian).

[7] J. Palarea-Albaladejo, J.A. Martin-Fernandez, A modified EM alr-algorithm for replacing rounded zeros in compositional data sets, Comput. Geosci. 34 (2008) 902–917.

[8] J. Palarea-Albaladejo, J.A. Martin-Fernandez, A. Bucciani, Compositional methods for estimating elemental concentrations below the limit of detection in practice using R, J. Geochem. Explor. 141 (2014) 71–77.

[9] A. Kabata-Pendias, Trace Elements in Soils and Plants, fourth ed., CRC Press, 2011.