A CHEMOMETRICAL LOOK AT
HEAVY METALS IN SOILS

Helvi HÖDREJÄRV

Tallinna Tehnikaülikooli keemiainstituut (Department of Chemistry, Tallinn Technical University), Ehitajate tee 5, EE-0026 Tallinn, Eesti (Estonia)

Abstract. Multivariate chemometrical methods – common factor analysis and principal component analysis – were applied to describe the behaviour of Pb, Cu, Mn, and Zn in soils. Some Cambisol and Luvisol profiles were analysed by AAS spectrometry and the sequential extraction method to obtain the total concentrations of heavy metals and their soluble and bioavailable fractions. The pH values and content of organic carbon were used as soil characteristics. Estimated communalities, loadings of elements for the two first factors, and biplots for the first two principal components are presented.

Key words: chemometry, microelements, soil.

Heavy metals found in soils are important from the viewpoint of normal life of plants from two contradictory aspects (i) as bioelements and (ii) as potential phytotoxicants. Over the years several studies have pointed out the importance of heavy metals in the biogeochemical cycling of elements in soil [1–3]. The heavy metals atmospheric input and potential pollution of agricultural land are essential in the oil shale region in Virumaa, Northeast Estonia [4]. However, the conclusions in these and other studies concerning the relationship of heavy metals to soil are of empirical character. Contemporary analytical methods provide us with numerous data sets of elemental analyses, but the resolution of these data into understandable form has remained difficult. Several researchers have utilized statistical techniques to reduce their data to meaningful terms. Krieg and Einax [5] made an attempt to give a multivariate statistical characterization of the metal status in soil to distinguish polluted soil layers or geogenic enrichment from background values. Cluster analysis as well as discriminant and factor analyses have been applied to the description of the metal distribution in soil profiles. Hopke et al. [2] showed that the data set of 18 elements in urban aerosols can be interpreted by the methods of common factor analysis and hierarchical aggregative cluster analysis.

This study was planned to identify factors that influence the distribution and accumulation of HM in soil horizons using a chemometrical solution. Four elements (Pb, Cu, Mn, Zn) were selected for this study. The content of organic carbon was used as a tracer of soil. The value of pH of soil in North Estonia is to a large extent determined by the influence of the calcareous geological base and is used as a tracer of deeper horizons. The estimated data set as well as the correlation matrices for the soils are rather large and can be successfully treated only by multivariate computing methods. We have used factor analysis and the principal component method to look for patterns of correlation among the heavy metals in soils.

MATERIALS AND METHODS

As the behaviour and distribution of heavy metals depend on the character of pedogenesis as well as on the genetic origin of soils formed [6] soil sections of clay-accumulative and clay-translocative differentiation on yellowish-grey calcareous till were obtained for the heavy metal analytical techniques from the Institute of Soil Science and Agrochemistry, Estonian Agricultural University. Soils had been sampled by genetic horizons already in 1966 and generally analysed in the late sixties [7]: Cambisol (typical brown soil) under an oak stand of *Hepatica* site type at Mihkli, West Estonia, and Luvisol (brown lessive soil) under a spruce stand of the same site type at Kaarma, Central Estonia. Later on another Cambisol section sampled by horizons in 1980 under the spruce stand at Mihkli and an arable Luvisol section sampled in 1989 at Inju (close to Kaarma), Central Estonia, were received for complementary analysis. Unpublished analytical data on organic carbon and pH of soil suspension were kindly given us for the use in this investigation.

On the basis of four sections (two forest Cambisols from Mihkli, forest Luvisol from Kaarma, and arable Luvisol from Inju) also four data matrices were compiled with soil horizons representing rows in the matrices (Table 1).

The concentrations of lead, copper, manganese, and zinc, and their soluble and bioavailable parts were determined by sequential extraction speciation and the atomic absorption spectrometry (AAS) method in the Laboratory of Environmental Analysis of Tallinn Technical University [8]. The columns of the matrices included the following characteristics: the total concentrations of metals (t), the soluble parts (s), the bioavailable parts (b), soil pH_{KCl}, percentage of organic C, and the average depth of the soil horizons (D, cm). The soluble part of an element is the part that is not included into silicate lattice and is soluble in sequentially used solutions of EDTA (pH 10), ammonium acetate (pH 4.8), hydroxylamin-chloride (pH 2.4), oxalate buffer (pH 3.0), and hydrogen peroxide-nitric acid (pH 1.3). The bioavailable part of an element is soluble in the two first solutions of that system (EDTA and ammonium acetate). The computing was carried out with the program STATGRAPHICS [9].

Table 1

Total concentrations of Pb, Cu, Mn, and Zn in soil horizons, mg/kg

D, cm	C, %*	pH*	Pb_t	Cu_t	Mn_t	Zn_t
			Cambisol under an oak stand in 1966			
0–5	2.35	6.5	43.5	18.2	771	51.7
5–15	1.57	5.9	42.6	13.6	694	45.7
15–25	0.82	5.7	37.0	20.0	485	53.0
25–35	0.50	6.0	33.6	25.0	862	52.4
35–45	0.49	6.8	42.7	14.0	467	41.9
45–55	<0.1	7.3	26.5	17.2	545	42.7
60–70	<0.1	7.9	29.5	12.0	353	19.6
			Cambisol under a spruce stand in 1980			
0–5	3.13	6.3	38.1	9.7	785	45.6
5–15	2.32	5.8	28.4	11.2	727	41.0
15–25	1.07	5.2	28.2	11.1	686	28.3
25–35	0.52	5.4	18.1	11.0	548	36.1
35–45	0.58	5.6	18.7	16.3	693	46.3
45–55	0.58	6.6	24.8	26.6	621	53.2
60–70	<0.1	7.4	26.3	9.6	383	16.3
			Luvisol under a spruce stand in 1966			
0–4	3.07	4.92	22.4	20.5	320	40.9
10–18	1.79	5.40	21.1	11.2	403	32.3
25–35	0.48	4.90	24.3	11.4	332	38.9
45–50	0.12	4.88	26.2	7.1	262	27.2
60–70	0.25	5.48	32.6	6.8	521	59.0
90–100	<0.1	7.20	55.1	6.5	350	30.3
			Arable Luvisol in 1989			
0–20	1.69	6.75	36.6	80.9	705	43.9
25–31	1.69	6.85	31.5	37.3	610	31.5
32–42	0.30	6.65	29.5	52.6	673	36.1
50–65	<0.1	6.80	30.3	32.8	739	22.7
70–85	<0.1	7.00	47.0	41.0	818	47.0
95–100	<0.1	7.25	41.4	39.2	630	38.7

* Determined at the Estonian Agricultural University [7].

RESULTS AND DISCUSSION

Common factor analysis was applied to identify the dimensions of the data set that are not directly observable. A correlation matrix of variables was used.

The estimated communalities, which tell what proportion of the variability of each variable is shared with the other variables in the data set, were also calculated (Table 2). The organic carbon, pH of soil, and the total concentrations of the elements were applied as the variables in the calculation.

Estimated communalities for the organic C, pH of soils, and total concentrations of heavy metals

Soil	Organic C	pH	Pb_t	Cu_t	Mn_t	Zn_t
Cambisol						
oak stand	0.89872	0.66587	0.85338	0.74247	0.73810	0.92161
spruce stand	0.96010	0.21791	0.74499	0.69627	0.92978	0.89415
Luvisol						
spruce stand	0.94779	0.96350	0.95063	0.94096	0.04553	0.02238
arable	0.89895	0.88617	0.91313	0.90017	0.84155	0.97289

With a few exceptions (pH of Cambisol, Mn and Zn of Luvisol under spruce stand) the estimated communalities were high for all variables. The estimated communalities were used to replace the diagonal elements of the correlation matrix. The factor matrix transformation based on an orthogonal Varimax rotation was used to organize the data into an easily understandable framework. The table for two first factors (F1, F2), which explain on average 90% of the variance, is presented (Table 3). If the variance for a given variable is distributed equally throughout the factors, the loading for that variable will be 0.5; consequently, a factor loading higher than 0.5 indicates an association of the variable with that factor. In Table 3 the loadings smaller than 0.6 are omitted for clarity's sake.

The loadings of variables for the first factors (F1, F2) produced by varimax rotation

Variable	Cambisol				Luvisol			
	oak stand		spruce stand		spruce stand		arable	
	F1	F2	F1	F2	F1	F2	F1	F2
Pb		0.92	0.86			0.91		0.99
Cu	0.87				0.94		0.97	
Mn	0.73			0.79				0.78
Zn	0.88			0.94				0.80
pH	−0.70				0.96			0.66
C		0.93	0.91		0.95		0.91	

100

The loadings of variables in Table 3 indicate that only lead in Cambisols and copper in Luvisols are strongly connected with the organic carbon. Manganese and zinc have nearly equal loadings in Cambisols and arable Luvisol. In forest Luvisol their loadings are not significant. The relationship of different metals to pH of soil is different or missing.

Principal component analysis was carried out to complement factor analysis and detect elemental patterns of the soil horizons. The reduced solution of the principal component analysis consists of two components, which explain 79–98% of the variance of the data sets of the investigated soils. These two components were used to draw biplots (Fig. 1), which
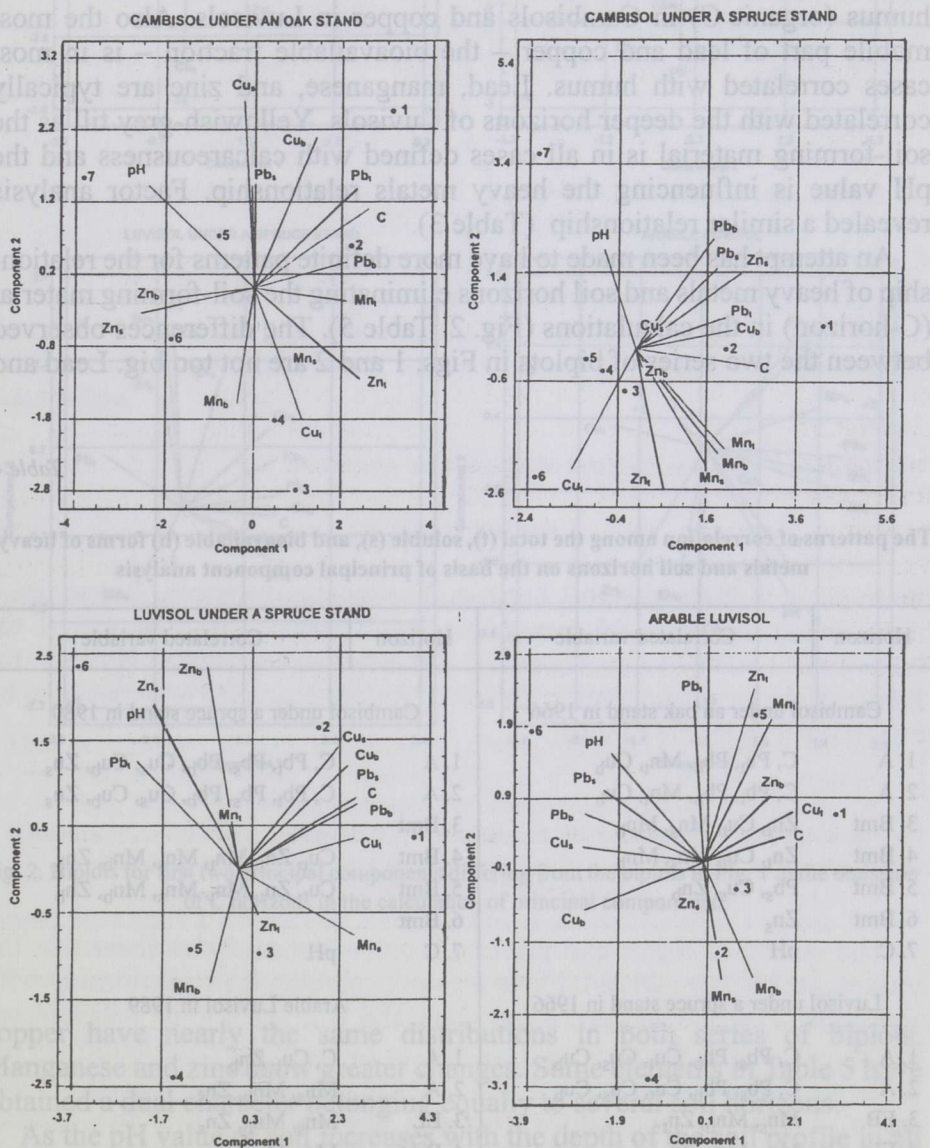
Fig. 1. Biplots for first two principal components. Vectors reflect relationships between organic carbon, the value of pH_{KCl} of the soil, and the total contents and soluble and bioavailable parts of lead, copper, manganese, and zinc in soil. Dots represent the soil horizons in the order of growing depth.

reflect the relationship between the characteristics as well as between the characteristics and soil horizons. In the biplot the component weights (loadings) for the two first components are plotted with one data point determining the length of each vector for each element (variables or characteristics). The vectors are intersecting at the (0.0) origin. The length of each vector is proportional to its contribution to the principal components. The angle between any two vectors is inversely proportional to the correlation between them. The distributions of the scores for soil horizons for the first two principal components have a graphical perspective as a scatterplot.

The results of the principal component analysis (Table 4) are in good agreement with factor analysis data and indicate that lead is related to humus (organic C) in Cambisols and copper in Luvisols. Also the most mobile part of lead and copper – the bioavailable fraction – is in most cases correlated with humus. Lead, manganese, and zinc are typically correlated with the deeper horizons of Luvisols. Yellowish-grey till as the soil-forming material is in all cases defined with calcareousness and the pH value is influencing the heavy metals relationship. Factor analysis revealed a similar relationship (Table 3).

An attempt has been made to have more definite patterns for the relationship of heavy metals and soil horizons eliminating the soil-forming material (C-horizon) in the calculations (Fig. 2, Table 5). The differences observed between the two series of biplots in Figs. 1 and 2 are not too big. Lead and

Table 4

The patterns of correlation among the total (t), soluble (s), and bioavailable (b) forms of heavy metals and soil horizons on the basis of principal component analysis

| Horizon	Correlated variable	Horizon	Correlated variable
Cambisol under an oak stand in 1966		Cambisol under a spruce stand in 1980	
1. A	C, Pb_t, Pb_b, Mn_t, Cu_b	1. A	C, Pb_t, Pb_s, Pb_b, Cu_s, Cu_b, Zn_s
2. A	C, Pb_t, Pb_b, Mn_t, Cu_b	2. A	C, Pb_t, Pb_s, Pb_b, Cu_s, Cu_b, Zn_s
3. Bmt	Zn_t, Cu_t, Mn_s, Mn_b	3. Bmt	
4. Bmt	Zn_t, Cu_t, Mn_s, Mn_b	4. Bmt	Cu_t, Zn_t, Mn_t, Mn_s, Mn_b, Zn_b
5. Bmt	Pb_s, Cu_s, Zn_b	5. Bmt	Cu_t, Zn_t, Mn_t, Mn_s, Mn_b, Zn_b
6. Bmt	Zn_s	6. Bmt	
7. C	pH	7. C	pH
Luvisol under a spruce stand in 1966		Arable Luvisol in 1989	
1. A	C, Pb_s, Pb_b, Cu_t, Cu_s, Cu_b	1. A	C, Cu_t, Zn_b
2. A	C, Pb_s, Pb_b, Cu_t, Cu_s, Cu_b	2. A	Mn_s, Mn_b, Zn_s
3. EB	Mn_s, Mn_b, Zn_t	3. EL	Mn_s, Mn_b, Zn_s
4. EL	Mn_s, Mn_b, Zn_t	4. Bmt	Mn_s, Mn_b, Zn_s
5. Bt	Mn_s, Mn_b, Zn_t	5. Bmt	Pb_t, Mn_t, Zn_t
6. C	pH, Pb_t, Mn_t, Zn_s, Zn_b	6. C	pH, Pb_s, Pb_b, Cu_s, Cu_b
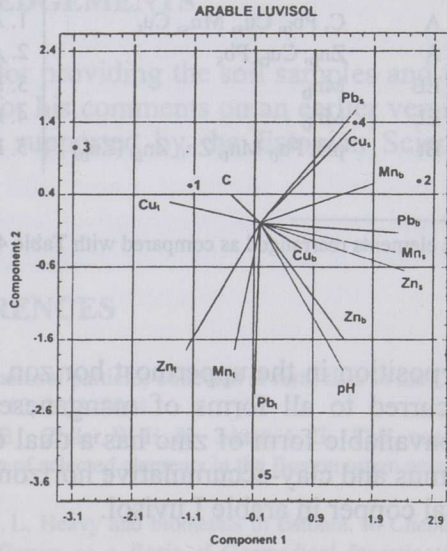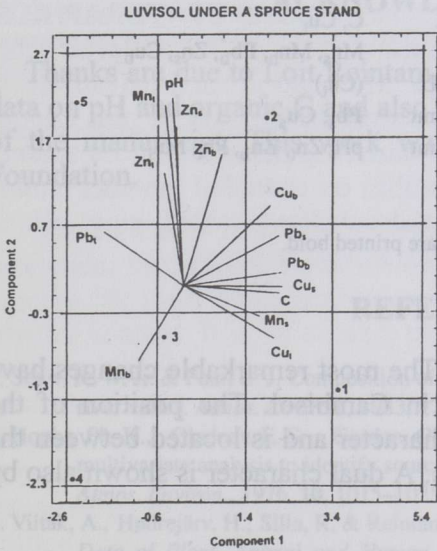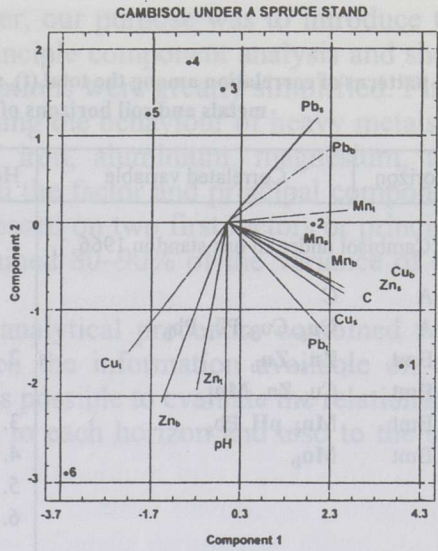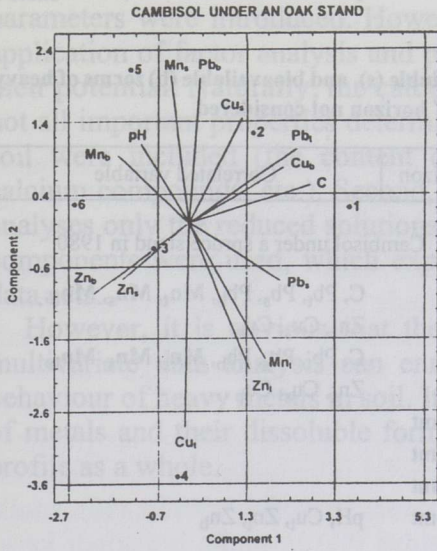
Fig. 2. Biplots for first two principal components differing from the biplots in Fig. 1 in the omission of C horizons in the calculation of principal components.

copper have nearly the same distributions in both series of biplots. Manganese and zinc show greater changes. Some elements in Table 5 have obtained a dual character belonging equally to several soil horizons.

As the pH value of soil increases with the depth of the soil profile in all soils (Table 1) the elimination of the parent till did not change the position of pH in the pattern. The same is valid for organic carbon, which is keeping

Table 5

The patterns of correlation among the total (t), soluble (s), and bioavailable (b) forms of heavy metals and soil horizons of C horizon not considered

Horizon	Correlated variable	Horizon	Correlated variable
Cambisol under an oak stand in 1966		Cambisol under a spruce stand in 1980	
1. A	C	1. A	C, Pb_t, Pb_s, Pb_b, Mn_t, Mn_s, Mn_b,
2. A	Cu_s, Cu_b, Pb_t, Pb_b		Zn_s, Cu_s, Cu_b
3. Bmt	Zn_s, Zn_b	2. A	C, Pb_t, Pb_s, Pb_b, Mn_t, Mn_s, Mn_b,
4. Bmt	Cu_t, Zn_t, Mn_t		Zn_s, Cu_s, Cu_b
5. Bmt	Mn_s, pH, Pb_s	3. Bmt	
6. Bmt	Mn_b	4. Bmt	
		5. Bmt	
		6. Bmt	pH, Cu_t, Zn_t, Zn_b
Luvisol under a spruce stand in 1966		Arable Luvisol in 1989	
1. A	C, Pb_b, Cu_s, Mn_s, Cu_t	1. A	C, Cu_t
2. A	Zn_b, Cu_b, Pb_s	2. A	Mn_s, Mn_b, Pb_b, Zn_s, Cu_b
3. EB	Mn_b	3. EL	(Cu_t)
4. EL	Mn_b	4. Bmt	Pb_s, Cu_s
5. Bt	pH, Pb_t, Mn_t, Zn_t, Zn_s, (Zn_b)	5. Bmt	pH, Zn_t, Zn_b, Pb_t, Mn_t

The elements rearranged as compared with Table 4 are printed bold.

its position in the uppermost horizon. The most remarkable changes have occurred to all forms of manganese in Cambisol. The position of the bioavailable form of zinc has a dual character and is located between the humus and clay-accumulative horizons. A dual character is shown also by total copper in arable Luvisol.

CONCLUSIONS

The results of factor and principal component analyses show that the investigated heavy metals fall into groups according their relationship with soil organic carbon and pH. Manganese and zinc seem to be less affected by organic carbon than lead and copper. In Luvisols the contribution of pH in the principal component weights is more essential than that of organic carbon. In Cambisols organic carbon has an essential contribution in most cases. There are considerable differences in the relationships of heavy metals and soil horizons between Luvisols and Cambisols.

Probably more relationships would be revealed if some additional soil parameters were introduced. However, our purpose was to introduce the application of factor analysis and principle component analysis and show their potential. Naturally, the calculations were greatly simplified. First, not all important properties determining the behaviour of heavy metals in soil were included (the content of iron, aluminium, magnesium, and calcium compounds, etc.). Second, in the factor and principal component analyses only the reduced solutions based on two first factors or principal components were used, which explained 80–90% of the variance of the data sets.

However, it is obvious that the analytical procedure combined with multivariate data analysis can enrich the information available on the behaviour of heavy metals in soil. It is possible to evaluate the relationship of metals and their dissoluble forms to each horizon and also to the soil profile as a whole.

ACKNOWLEDGEMENTS

REFERENCES

1. Shaw, R. W. Jr. & Paur, R. J. Composition of aerosol particles collected at rural sites in the Ohio river valley. *Atmos. Environ.*, 1983, **17**, 10, 2031–2044.

2. Hopke, Ph. K., Gladnev, E. S., Gordon, G. E., Zoller, W. H. & Jones, A. G. The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol. *Atmos. Environ.*, 1976, **10**, 1015–1025.

3. Viitak, A., Hødrejärv, H., Silla, R. & Reintam, L. Heavy and biometals in Estonia. In *Chemical Data of Plant, Animal and Human Tissues as a Basis of Geomedical Investigations* (Låg, J., ed.). The Norwegian Academy of Science and Letters, Oslo, 1996, 115–124.

4. Hödrejärv, H., Viitak, A., Lepane, V., Vaarmann, A. & Treumann, M. Background and bioaccumulation of heavy metals in North-Estonia. In *Problems of Contemporary Ecology*. Estonian Council of Ecology, Tartu, 1994, 120–124.

5. Krieg, M. & Einax, J. Chemometric characterization of soil depth profiles. *Fresenius J. Anal. Chem.*, 1994, **348**, 490–495.

6. Kabata-Pendias, A. Heavy metals in soils – issues in Central and Eastern Europe. In *Heavy Metals in the Environment* (Wilken, R.-D., Förstner, U., Knöchel, A., eds.), vol. 1. Hamburg, 1995, 20–27.

7. Reintam, L. About the characteristics of the brown forest soil type. In *Trans. of Estonian Agricultural Academy, 65. Soil Regimes and Processes* (Reintam, L., ed.). Tartu, 1970, 9–37.

8. Hödrejärv, H., Viitak, A. & Vaarmann, A. Chemical speciation analysis of micrometals in soils. *Proc. Estonian Acad. Sci. Chem.*, 1995, **44**, 2/3, 171–178.

9. *Andmeanalüüs personaalarvutil programmipaketi STATGRAPHICS abil.* Eesti Majandusjuhtide Instituut, Tallinn, 1991.

KEMOMEETRILINE HINNANG RASKMETALLIDE SISALDUSELE MULDADES

Helvi HÖDREJÄRV

On esitatud multivariantsete kemomeetriliste meetodite rakendus-võimalusi raskmetallide seisundi uurimiseks muldades. On uuritud Pb, Cu, Mn ja Zn sisaldust ning nende lahustuvust kollakashallil karbonaatsel moreenil kujunenud leostunud ja leetjate metsa- ja põllumuldade genee-tilistes horisontides. Mullaproovid saadi Eesti Põllumajandusülikooli (EPMÜ) mullateaduse ja agrokeemia instituudist. Mikroelementide sisal-duse määras Tallinna Tehnikaülikooli keskkonnaanalüütika töörühm, andmed orgaanilise süsiniku ja mulla pH kohta saadi EPMÜ-st, kusjuures osa neist oli varem avaldatud. Arvutustes lähtuti andmemaatriksitest, milles objektideks olid muldade horisondid ja omadusteks raskmetallide üldsisaldus, lahustuvad vormid (%), biomastatavad vormid (%), mulla pH_{KCl}, mulla orgaanilise süsiniku sisaldus (%) ja mulla horisondi kesk-mine sügavus (cm). Andmete töötlemiseks arvutil rakendati peakompo-nentide ja klassikalise faktoranalüüsi programmipaketti STATGRAPHICS. Andmetöötluse tulemusel rühmitusid uuritud raskmetallid vastavalt seos-tele mulla orgaanilise süsiniku ja pH-ga. Mangaan ja tsink näivad olevat vähem mõjustatud mulla orgaanilisest süsinikust kui vask ja plii. Leetjas mullas on pH osakaal raskmetallide jaotumises mulla horisontides suurem kui orgaanilisel süsinikul. Leostunud mullas on enamikel juhtudel oluline osakaal orgaanilisel süsinikul. Elementidevahelised suhted ning suhted leostunud ja leetja mulla horisontidega on üsnagi erinevad.