HEAVY METALS IN TOPSOILS OF THE REPUBLIC OF TATARSTAN, RUSSIA: MULTIVARIATE ANALYSIS AND POLLUTION EVALUATION

S. S. Ryazanov
Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, 28 Daurskaya street, Kazan, 420087, Russia
E-mail: RStanislav.soil@yandex.ru

D. V. Ivanov
Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, 28 Daurskaya street, Kazan, 420087, Russia
E-mail: water-rf@mail.ru

V. I. Kulagina
Research Institute for Problems of Ecology and Mineral Wealth Use of Tatarstan Academy of Sciences, 28 Daurskaya street, Kazan, 420087, Russia
E-mail: viksoil@mail.ru

Abstract. Background. The study is conducted to investigate the content and spatial distribution of Fe and eight heavy metals (Cd, Co, Cr, Cu, Mn, Ni, Pb, and Zn) in soils of the industrially developed region.

Materials and methods. A total of 1,170 soil samples of different land use (natural, agricultural, urban) were collected from topsoils in the territory of the Republic of Tatarstan, Russia. Heavy metals concentrations in soil samples were determined using atomic absorption spectrometry after 5M HNO3 extraction. Multivariate and geostatistical analyses were used to investigate the current state of soil heavy metal contamination and to identify spatial patterns and possible sources of heavy metals on the regional scale.

Results. Zonal soil types of a natural land use were used to assess the regional background values of the heavy metals: Cd – 0.44±0.24 mg/kg, Co – 10.4±3.6 mg/kg, Cr – 23.3±12.7 mg/kg, Cu – 16.4±7.8 mg/kg, Fe – 15275.4±5178.3 mg/kg, Mn – 652.4±228.4 mg/kg, Ni – 29.8±18.8 mg/kg, Pb – 11.5±3.2 mg/kg, Zn – 43.3±12.8 mg/kg. The results of the pollution evaluation showed the absence of regional-scale contamination directly related to agriculture. Urban soils were contaminated by Cu, Pb and Zn. Geostatistical analysis revealed several patterns of regional distribution of heavy metals and suggested an anthropogenic impact to the Cu, Pb and Zn distribution. Principal component analysis allowed distinguishing three regional geochemical groups of heavy metals and showed that at the regional scale the distribution of Cu, Mn and Ni is controlled by the element richness in soil parent material, overlayed by the soil forming factors; the distribution of the Co, Cr and Fe is controlled mainly by lithology; and the distribution of the Pb, Zn and Cd is strongly influenced by the anthropogenic sources. Conclusion. This case study demonstrates that a combination of multivariate statistics and geostatistical analysis together with the pollution assessment allows comprehensive characterizing heavy metals spatial distribution and determining their sources.

Keywords: soil, heavy metals, mapping, multivariate analysis, geostatistics, Republic of Tatarstan, Russia.

Introduction

Identifying sources of heavy metals in the environment is of great importance to our understanding of their pollution patterns and natural cycles in surface reservoirs of Earth [1]. Heavy metals in the topsoil have two sources: a natural input as a result of weathering of soil-forming rocks, and anthropogenic sources such as transport, industrial enterprises and agricultural land use [2, 3]. Over the past decades, there has been a decline in anthropogenic emissions of heavy metals into the environment. Thus, annual emissions of cadmium in European countries decreased from 485 t to 257 t in the period from 1990 to 2003; lead – from 25 kt to 8.6 kt; mercury – from 413 t to 195 t [4]. Despite this, the danger of soil contamination by metals is still acute in industrially developed countries, which is due to their long-term accumulation in upper soil horizons [5].

Concern about increasing levels of heavy metals in soils has led to the development and implementation of numerous programs to determine the baseline levels and contamination rates by heavy metals [6]. The study of soil quality allows us not only to assess the consequences of human activities, but also is an indispensable stage for sustainable development and conservation of depleted soil resources [7]. A combination of multivariate statistics and geostatistical analysis is successfully used for identifying pollution characteristics of heavy metals in soils and distinguishing their natural sources and anthropogenic inputs [6, 8–11].

Previous studies have been dedicated to research influence of soil types and land use types on the total content and mobility of distinct heavy metals in the topsoil, and to investigate geochemical dependence of heavy metals on soil properties (humus content, particle size distribution, pH and content of Fe and Mn) [12].

The present study has been focused on polynomial profilation of heavy metal concentratations in the topsoil in the Republic of Tatarstan, Russia. Multivariate and geostatistical analyses were used to investigate the current state of soil heavy metal contamination in this area and to identify spatial patterns and possible sources of heavy metals on the regional scale.

Materials and methods

Study area. The study was conducted in the territory of the Republic of Tatarstan located in the eastern part of the Russian Federation (55°20'36.1"N; 50°47'31.7"E). The total area of the Republic is 67 847 sq. km. The climate of the region is moderately continental with slight differences between climate-geographical zones. The mean annual temperature is 2–3.1 °C, the mean annual precipitation amount is 460–540 mm. The relief is an elevated stepped plain dissected by a dense network of river valleys [13]. The soil cover, according to the “Classification and Diagnostics of Soils of the USSR” is represented by the following main types: Chernozems (39.7 %), Gray Forest soils (39.5 %), Sod-Podzolic soils (7 %), Sod-
Calcareous soils (3.1 %) [13–14]. Several main types of land use are distinguished within the Republic: agricultural (67 %), settlement (5.2 %), water fund (6.5 %), protected areas (0.5 %) and forest fund (18 %) [15].

The industrial profile of the republic is determined by machine-building enterprises, radio- and electrical instrument making enterprises, production and processing of oil and gas. The main industrial centers of the Republic are the cities of Kazan, Naberezhnye Chelny, Nizhnekamsk, Zelenodolsk, Elabuga and Almetyevsk [16].

**Soil sampling and soil analysis.** A total of 1,170 soil samples were collected from the topsoils (0–20 cm depths) (Fig. 1). Sampling was performed at a distance of at least 200 m away from industry and transport infrastructure. The soil pH was measured in a suspension of 1:2.5 soil: distilled water using a pH meter (pH-150 МИ, LLC “Izmeritel’naya tekhnika”, Russia). The soil humus content was determined by the Tyurin titrimetric method in TsINAO modification [17]. The particle size distribution was determined by the Kachinskiy-Robinson-Kehl pipet method [17]. As an indicator of the particle size distribution, the sum of particles < 0.01 mm was used. According to N. A. Kachinskiy this range of particle sizes is called “physical clay”, and this term is used henceforward [18]. In the soil samples the “total” content of Fe and eight heavy metals (Cd, Co, Cr, Cu, Mn, Ni, Pb, and Zn) were determined according to GD 52.18.191-89 [19]. The soil samples were air-dried, pounded and passed through a 1.0 mm sieve. Then, 2.0 g of the sieved samples were digested for 3 hours in a 10 ml of 5 M HNO₃ on a boiling water bath. After the 3 hours, flasks with the samples were pulled out from the water bath and cooled to room temperature. After cooling, extractions were filtered through ashless filters into 50 ml flasks, and brought to the volume with deionized water. The concentrations in the solution were measured by flame atomic absorption spectrometry (AAnalyst-400 “Perkin Elmer”, USA). It should be considered, that the extraction of heavy metals by 5 M HNO₃ does not in all cases allow estimating absolute values of heavy metals concentrations. The extractable fraction of heavy metals depends on geochemical properties of an element and nature of an organic-mineral matrix, and may be less than 75 % of their actual content. Nevertheless, this extraction allows assessing overall soil pollution and remains relevant from the practical point of view. The term “total concentration” is used henceforward to identify this fraction of metals.
Soil types were determined according to “Classification and Diagnostics of Soils of the USSR” [14]. The type of land use was established according to the site of sampling: agricultural, natural (forest and virgin soils), urban (samples collected on a territory of cities). The distribution of soil samples by soil types and land use types is presented in Table 1. Sod-Podzolic, Sod-Alluvial and Meadow Alluvial soils were mainly represented by the natural land use type. The most samples of Chernozems, Gray Forest and Sod-Calcareous soils were represented by tillable horizons of agricultural soils. All urban soils were united in the type of Urbanozems due to impossibility to determine their original genetic type as a result of anthropogenic transformation and destruction.

### Table 1

<table>
<thead>
<tr>
<th>Soil type</th>
<th>Land use type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>natural</td>
</tr>
<tr>
<td>Sod-Podzolic</td>
<td>38</td>
</tr>
<tr>
<td>Grey Forest</td>
<td>155</td>
</tr>
<tr>
<td>Chernozems</td>
<td>31</td>
</tr>
<tr>
<td>Sod-Calcareous</td>
<td>55</td>
</tr>
<tr>
<td>Sod-Alluvial</td>
<td>61</td>
</tr>
<tr>
<td>Meadow-Alluvial</td>
<td>29</td>
</tr>
<tr>
<td>Urbanozems</td>
<td>0</td>
</tr>
</tbody>
</table>

Background levels and pollution assessment. The Median±2MAD (MAD=Median(|Xi-Me(X)|)) method was used to estimate the regional background levels of heavy metals [2]. It is a robust non-parametric method and does not depend from the data distribution. The estimated background levels are represented by a range, which is more realistic than absolute single-value estimations [20]. Only zonal soil types of natural land use were used to assess the background of heavy metals. Non-zonal soil types, such as calcareous and alluvial soils, were excluded to eliminate the influence of anomalies associated with releases of carbonate rocks and the depositions with alluvial sediments. Agricultural and urban soils were excluded from the background estimation because of a potential contamination due to a high anthropogenic load.

In this study, three indices were applied to qualify the observed degree of heavy metal enrichment and to evaluate anthropogenic influences. The single factor pollution index ($P_i$) was used to assess enrichment ratios of the each heavy metal against background levels [7]:

$$P_i = C_i / S_i ,$$  \hspace{1cm} (1)

where $C_i$ is the measured value of the element $i$ at each sampling point, and $S_i$ is the regional background value of the heavy metal $i$. The soil pollution according to $P_i$ values is classified into three levels: if $P_i < 1$, soil is characterized as unpolluted; if the index is in the range of $1 \leq P_i \leq 3$, soil is moderately polluted; if $3 \leq P_i$, soil is strongly polluted [21].

Two methods to assess overall contamination were used. The first is the Nemerow Comprehensive Index ($P_X$), which is commonly used to assess polyelemental soil pollution [22]:

$$P_X = \sqrt{\left( \frac{P_i}{2} \right)^2 + \left( \max(P_i) \right)^2} ,$$  \hspace{1cm} (2)

where $P_i$ is the single factor pollution index of heavy metal $i$, $\overline{P_i}$ is the mean of single factor index for heavy metal $i$, and $\max(P_i)$ is the maximum of single factor index for heavy metal $i$. Classification of Nemerow Comprehensive Index values is similar to $P_i$ classification.

The second one is the Saet-based Comprehensive Pollution Index ($Z_c$), and it allows assessing ecological threat of contamination since it takes into account geometric mean of single pollution indexes and toxicity of heavy metals [23]:

$$Z_c = n \left( \prod_{i=1}^{n} P_{i} * K_{i} * P_{i} * K_{i} * \cdots * P_{i} * K_{i} \right)^{1/n} - (n-1) ,$$  \hspace{1cm} (3)

where $P_{i1} - P_{in}$ is the single factor pollution indexes of the elements, $n$ is the number of heavy metals, $K_{i}$ is the toxicity coefficient of heavy metal $i$.

Use of the geometric mean in $Z_c$ calculation, instead of the arithmetic mean in the classic Saet Index, decreases estimates of overall pollution in the case of high spread of the pollution indices $P_i$ [23]. The toxicity coefficients of heavy metals depend on their hazard classes. The heavy metals hazard classes and corresponding toxicity coefficients are represented in Table 2. The $Z_c$ is classified into the following pollution classes: $Z_c < 16$ –
pollution is not dangerous, 16<Z_c<32 – moderate danger of pollution, 32<Z_c<128 – dangerous pollution, and Z_c>128 – extremely dangerous pollution [23].

**Table 2**

<table>
<thead>
<tr>
<th>Hazard class</th>
<th>K_i</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>Cd, Pb, Zn, Cr</td>
</tr>
<tr>
<td>2</td>
<td>1.0</td>
<td>Co, Ni, Cu</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>Mn</td>
</tr>
</tbody>
</table>

**Spatial interpolation and accuracy assessment.**
It was shown that ordinary kriging (OK) is the best method to interpolate a single variable in absence of explanatory variables [24–26]. However, OK estimates do not reproduce the sample histogram because of reduced variance as a consequence of a smoothing effect. In the OK estimation process low values are overestimated and high values are underestimated making the estimated histogram narrower than the sample histogram [27]. This can lead to a critical underestimation of regional soil pollution. To get accurate maps with preserved spatial variation of the heavy metals content, the interpolation was done using OK with a smoothing effect correction, as it proposed by Rezaee et al. [27]. The method is based on a normal score transformation of the original variable:

\[ y(x_i) = G^{-1}\left(\frac{r(x_i)}{n+1}\right), \]  

where \( y(x_i) \) is the normal score, \( G(y) \) is the standard normal cumulative density function, and \( r(x_i) \) is the rank for \( i_{th} \) \( z(x_i) \) associated with a set of \( n \) data values.

Further estimate and smoothing correction are done using equation:

\[ Y_{OK}^{**}(x_0) = \left( Y_{OK}(x_0) - E[Y_{OK}(x_0)] \right) \times \left[ \frac{\sqrt{Var[Y(x)]} + E[Y(x)]}{\sqrt{Var[Y_{OK}(x_0)]}} \right], \]  

where \( Y(x) \) is the normal score transformed data, \( Var[Y(x)] \) is the variance of the transformed data, \( E[Y_{OK}(x_0)] \) is the ordinary kriging estimate of the transformed data.

After the estimation and correction were done, \( Y_{OK}^{**}(x_0) \) is back-transformed into an original scale using the inverse operation:

\[ Z_{OK}^{**}(x_0) = F^{-1}(Y_{OK}^{**}(x_0)), \]  

where \( Z_{OK}^{**}(x_0) \) is back-transformed values, \( F^{-1} \) is the function of back-transformation.

More details on the algorithm and examples can be found in the original paper.

Accuracy of the interpolated maps was assessed using leave-one-out cross-validation (LOOCV). In LOOCV each sampling point \( o_i \) is removed sequentially. A spatial interpolation model is fit on \( n-1 \) observation, and a prediction \( p_i \) is made for an excluded observation, using its \( X \) values. Several error measurements were calculated using the difference \( o_i-p_i \):

1) mean error is given by:

\[ ME = \frac{1}{n} \sum_{i=1}^{n} (p_i-o_i); \]  

2) root mean square error is given by:

\[ RMSE = \left[ \frac{1}{n} \sum_{i=1}^{n} (p_i-o_i)^2 \right]^{1/2}; \]  

3) ratio of the observed and the predicted variances is given by:

\[ RVar = \frac{Var[p]}{Var[o]}. \]

Li and Heap made a review of several criteria for using error measurements to judge the performance of spatial interpolation methods [25]. The model is better if ME is closer to zero and RMSE is smaller. The closer RVar is to 1, the better the ability of a spatial interpolation method to preserve an observed variance.

**Statistical analysis.** Principal component analysis was commonly used to distinguish possible sources of heavy metals and pathways in soils [1, 28, 29]. Correlation matrix-based Principal Component Analysis was carried out to determine geochemical associations of heavy metals in a regional scale. The PCA was performed considering the nine metals as variables. Varimax rotation was applied because orthogonal rotation minimizes the number of variables with high loading on each component and facilitates the interpretation of results.
The coefficients of the Principal Vectors, which are the eigenvectors of the correlation matrix, inform which metals contribute the most for each Rotated Principal Component (RC). In this work special attention was given to the metals with loadings higher than 0.5 and smaller than −0.5. This is similar to the Factor Analysis approach, with a difference that variables with loadings in the range −0.5 – 0.5 are not excluded from the analysis [30].

The Kaiser criteria was used to decide the appropriate number of rotated components to be retained (i.e., only factors with eigenvalues >1).

Software. All statistical analyses here presented were performed within the statistical environment R [31]. The spatial interpolation was performed using “gstat” package [32]. Final processing of predicted maps was performed using geoinformatical system QGIS [33].

Results and discussion

Heavy metals content in soils. The descriptive statistics of the heavy metals content in the top soils are summarized in Table 3.

Table 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>CV</th>
<th>Median by LU</th>
<th>Background levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cd (mg/kg)</td>
<td>0.04</td>
<td>1.98</td>
<td>0.40</td>
<td>0.36</td>
<td>0.21</td>
<td>51.4</td>
<td>0.39</td>
<td>0.32</td>
</tr>
<tr>
<td>Co (mg/kg)</td>
<td>0.3</td>
<td>41.9</td>
<td>11.3</td>
<td>10.6</td>
<td>5.0</td>
<td>44.3</td>
<td>10.38</td>
<td>11.03</td>
</tr>
<tr>
<td>Cr (mg/kg)</td>
<td>1.8</td>
<td>93.1</td>
<td>29.8</td>
<td>28.7</td>
<td>11.9</td>
<td>39.9</td>
<td>25.1</td>
<td>30.8</td>
</tr>
<tr>
<td>Cu (mg/kg)</td>
<td>0.3</td>
<td>143.2</td>
<td>19.6</td>
<td>19.0</td>
<td>9.4</td>
<td>47.8</td>
<td>17.43</td>
<td>19.25</td>
</tr>
<tr>
<td>Fe (mg/kg)</td>
<td>775</td>
<td>89830</td>
<td>20312</td>
<td>17980</td>
<td>9120</td>
<td>44.9</td>
<td>16311</td>
<td>18986</td>
</tr>
<tr>
<td>Mn (mg/kg)</td>
<td>49.7</td>
<td>3344</td>
<td>678.5</td>
<td>601.2</td>
<td>347.9</td>
<td>51.3</td>
<td>623.21</td>
<td>599.8</td>
</tr>
<tr>
<td>Ni (mg/kg)</td>
<td>0.8</td>
<td>142.4</td>
<td>37.8</td>
<td>36.0</td>
<td>18.1</td>
<td>47.9</td>
<td>33.82</td>
<td>38.13</td>
</tr>
<tr>
<td>Pb (mg/kg)</td>
<td>0.7</td>
<td>65.6</td>
<td>11.7</td>
<td>11.6</td>
<td>3.9</td>
<td>33.6</td>
<td>11.65</td>
<td>11.46</td>
</tr>
<tr>
<td>Zn (mg/kg)</td>
<td>6.1</td>
<td>244.2</td>
<td>43.7</td>
<td>43.9</td>
<td>10.9</td>
<td>25.0</td>
<td>43.24</td>
<td>43.76</td>
</tr>
<tr>
<td>Humus (%)</td>
<td>0.2</td>
<td>16.2</td>
<td>5.4</td>
<td>5.5</td>
<td>2.75</td>
<td>50.1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Phys. clay (%)</td>
<td>3.4</td>
<td>73.6</td>
<td>41.2</td>
<td>41.3</td>
<td>13.32</td>
<td>30.8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>pH</td>
<td>4.8</td>
<td>8.8</td>
<td>7.0</td>
<td>7.0</td>
<td>0.75</td>
<td>10.2</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: min – minimum; max – maximum; SD – standard deviation; CV – variation coefficient; LU – land use type; Nat. – natural; Agr. – aricultural.

The mean content of Co, Cd, Ni, and Mn in the topsoils of the RT exceeds worldwide estimates according to C. Reimann and P. de Caritat [34] – 10 mg/kg, 0.3 mg/kg, 20 mg/kg and 530 mg/kg for the listed elements respectively. The content of Cu, Zn, Cr, Fe and Pb, on the contrary, is lower than the mean worldwide estimations for soils – 25 mg/kg, 70 mg/kg, 80 mg/kg, 35000 mg/kg and 17 mg/kg respectively [34]. The mean content values of heavy metals in soils of natural land use, normalized by the lower background threshold (Median-2MAD), were decreased as follow: Ni>Cr>Fe>Cu>Pb>Zn>Co>Mn>Cd. Agricultural activity had a certain impact on the accumulation-leaching mode of the elements in topsoil, which slightly affected the decreasing order of normalized heavy metals content: Cr>Ni>Fe>Cu>Co>Zn>Pb>Mn>Cd. The influence of anthropogenic sources is the most powerful in urban soils where the background-normalized content of heavy metals decreased in the following order: Cu>Pb>Ni>Cd>Cr>Zn>Fe>Mn>Co.

Spatial distribution of heavy metals. Parameters of the fitted variogram models are shown in Table 4 and the resulted variogram graphs are shown in Fig. 2, j. The spatial structures of the Cd, Co and Zn content showed the presence of the geometric anisotropy in the NW-SW direction: the autocorrelation range in a minor direction (SE-NW) was 0.6 times less. The nugget/sill ratios showed the strong spatial dependence of the Co and Ni values [35]. The medium strength of autocorrelation was observed for Cu, Fe and Mn (see Table 4). In the case of Cd, Cr, Pb and Zn the nugget/sill ratio is >0.6 showing weak spatial dependence. The high nugget effect represents an uncorrelated variation at the scale of sampling, which in this case could be an indicator of significant local anthropogenic input of Cd, Cr, Pb and Zn [36].
The distribution maps of the heavy metals are shown in Fig. 2.a–i. As it can be seen from the predicted maps, there are several clear patterns of the heavy metals spatial distribution in the topsoil. The first geochemical zone showed elevated values of Cd (0.5–1.4 mg/kg), Co (15–37 mg/kg), Cr (30–68 mg/kg) and Zn (60–104 mg/kg) in the topsoil. It is located in the northern central part of the territory under investigation, in the western part of the so-called Predkam’e climate-geographical region (Fig. 2,b,c,i). The second geochemical zone is located in the southeastern part of the Republic (so-called Bugulmiskaya and Eastern-Zakamskaya climate-geographical regions), and it showed an elevated content of Cr (30–50 mg/kg), Mn (825–2043 mg/kg) and Ni (45–98 mg/kg) (Fig. 2,c,g). Iron showed two zones of elevated values: southwestern part of the Republic (so-called Predvolj’e), where it varied in the range 19500–67600 mg/kg; and the northern part (Predkam’e) with the range of the Fe content 19500–53300 mg/kg (Fig. 2,e). In addition, elements such as Cd, Cu, Pb and Zn showed “bullseye” elevated zones around the cities with developed industry (Fig. 2,a,d,h,i).

Principal component analysis. Only three rotated principal components (RC) with an eigenvalue higher than 1.0 were extracted during the PCA, accounting for over 64 % of the total variance. The results of the PCA showed that nine heavy metals with the strongest loadings on RC1 (0.76, 0.90 and 0.74 for Cu, Ni, and Mn, respectively) represent a regional geochemical group of manganophile elements. According to the geological classification by Goldshmidt, copper belongs to the group of chalcophiles, and nickel to the group of siderophiles [23]. However, it was shown that in the unpolluted soils there is a strong dependence of Cu and Ni content on the content of manganese compounds [23]. High correlation coefficients (Spearman’s rho) also pointed to the strong associations of Cu and Ni to the Mn in the investigated topsoils: \( r_{(Cu-Mn)} = 0.56 \) (adj. p-value<0.01), \( r_{(Ni-Mn)} = 0.66 \) (adj. p-value<0.01). The correlations between RC1 and humus content \( r_{(RC1-Humus)} = -0.42 \), physical clay content \( r_{(RC1-Phys. clay)} = 0.41 \), and pH \( r_{(RC1-Humus)} = 0.36 \) allowed us to conclude that RC1 is controlled by the element richness in soil parent material, overlayed by the soil forming factors (Fig. 4,a,b,c).

The elements Co, Cr and Fe with the highest loadings on RC3 (0.80, 0.54 and 0.80, respectively) represent a geochemical group of siderophiles. Being siderophiles by geological classification, Co and Cr retain a significant dependence on the content of Fe compounds in the topsoil horizons: \( r_{(Co-Fe)} = 0.58 \) (adj. p-value<0.01), \( r_{(Cr-Fe)} = 0.41 \) (adj. p-value<0.01). The coefficient of correlation between RC3 and physical clay content is high \( r_{(RC3-Phys. clay)} = 0.43 \), and confirms the effect of lithological factor. However in contrast to RC1, no significant relationships between RC3 and humus content \( r_{(RC3-Humus)} = -0.09 \), and between RC3 and pH \( r_{(RC3-pH)} = 0.01 \) were found (Fig. 4,g,h,i).

The elements Cd, Pb, and Zn showed the highest loadings on RC2 (0.72, 0.85 and 0.75, respectively). The absence of significant correlations between RC2 and soil properties \( r_{(RC2-Humus)} = 0.09 \), \( r_{(RC2-Phys. clay)} = 0.01 \), \( r_{(RC2-pH)} = 0.79 \), and patterns of the interpolated maps confirmed that the RC2 represents exogenous input of heavy metals in the conditions of increased anthropogenic load (Fig. 4,d,e,f).
Fig. 2. (a-i) Spatial distribution of heavy metals content in the topsoil and j) fitted variogram models. The numbers on the map show the location of cities: 1 – Kazan, 2 – Zelenodolsk, 3 – Naberezhnye Chelny, 4 – Almetyevsk
Fig. 3. Graphical display of the first three RCs of the PCA analysis: a) RC1 vs. RC2; b) RC1 vs. RC3

Fig. 4. Relationships between humus content, particle size distribution and pH and scores of the first three Rotated Components obtained from the PCA. LOESS smooth lines (red color) were added for better interpretability.
Pollution assessment. The summary of the calculated pollution indices is presented in Table 5. Pollution indices calculation was performed on the base of the upper background thresholds of heavy metals in topsoils (Median+2MAD).

According to the single-factor pollution indices, natural and agricultural soils in the territory under investigation were characterized as uncontaminated relatively to the regional background values. Thus, a number of soil samples with a $P_i<2$ did not exceed a 1% of all the natural soils for any of the heavy metals. The $P_i$ values distribution for agricultural soils was similar to that of the natural soils: 68–96% of arable horizon samples had a pollution index $P_i<1$ and were characterized as uncontaminated; 5–30% were characterized as potentially contaminated (1<$P_i<2$) by one or several heavy metals. At the same time, there were several Co and Cu $P_i$ values that indicate a high pollution, which could be an existence evidence of local contamination sources and requires additional confirmation on larger scales (see Table 5).

In urban soils high $P_i$ values were observed for three heavy metals: Cu, Pb, and Zn. A degree of contamination by copper was raised in cities in the following order: Naberezhnye Chelny ($P_i=1.12$, $\max(P_i)=3.35$) < Kazan ($P_i=1.21$, $\max(P_i)=4.35$) < Almetyevsk ($P_i=1.63$, $\max(P_i)=5.92$). Significant exceeding of background zinc values were observed in the cities Almetyevsk ($P_i=1.30$, $\max(P_i)=4.35$) and Naberezhnye Chelny ($P_i=0.92$, $\max(P_i)=3.15$). Lead content significantly exceeded the background level only in the Kazan City ($P_i=1.36$, $\max(P_i)=3.66$); in all other cities lead $P_i$ was less than a 2. The soils of the Zelenodol'sk was characterized as unpolluted with a slight local exceeding of Cd ($P_i=0.75$, $\max(P_i)=1.40$), Pb ($P_i=1.02$, $\max(P_i)=1.55$) and Zn ($P_i=0.70$, $\max(P_i)=1.26$).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Natural</th>
<th>Agricultural</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td>$P_i$</td>
<td>Cd</td>
<td>0.06–2.07</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Co</td>
<td>0.10–2.12</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Cr</td>
<td>0.07–2.10</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Cu</td>
<td>0.02–3.70</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Mn</td>
<td>0.06–3.80</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Ni</td>
<td>0.06–2.87</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Pb</td>
<td>0.07–1.58</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Zn</td>
<td>0.11–1.47</td>
<td>0.77</td>
</tr>
<tr>
<td>$P_N$</td>
<td>0.28–2.85</td>
<td>1.0</td>
<td>0.36–8.99</td>
</tr>
<tr>
<td>$Z_c$</td>
<td>0.0–6.0</td>
<td>0.13</td>
<td>0.0–18.8</td>
</tr>
</tbody>
</table>

The calculated Nemerow polyelemental pollution index for topsoil samples is shown in Fig. 5.a. According to the $P_N$ index, 54.4% of the surveyed natural soils and 60.1% of the agricultural soils were characterized as uncontaminated ($P_N<1$); 43.4% of the natural soils and 37.3% of the agricultural soils were potentially contaminated concerning the regional background levels (1<$P_N<2$). The proportion of soil samples with $P_N>2$ was less than 3% of all surveyed soils of natural and agricultural land use types (see Fig. 5.a). Distribution of the Nemerow index for urban soils reflected higher anthropogenic loads: 60.8% of samples collected on the territory of the largest cities were characterized as potentially contaminated (1<$P_N<2$); 8.9% of the urban samples were contaminated with heavy metals.

According to the calculated Saet-based Comprehensive Pollution Index, the contamination levels revealed by the single factor and Nemerow pollution indices were characterized as not dangerous (Fig. 5.b).

---

**Table 5**

Pollution indices ($P_i$) of heavy metals in topsoil of different land use

![Fig. 5. a) Nemerow Comprehensive Pollution Index, and b) Saet-based Comprehensive Pollution Index](image)
Conclusion

The concentrations, background levels and sources of Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb and Zn in topsoils from the Republic of Tatarstan, Russia were investigated in this study. The results obtained offered relevant information on the distribution of heavy metals on a regional scale.

Pollution assessment using single factor pollution indices and two comprehensive polyelemental pollution indices showed the absence of heavy metals contamination directly related to an agricultural activity: proportions of pollution indices values of agricultural soils were similar to those of the natural soils. However, agricultural activity influences the physical and chemical soil properties, which was reflected on the proportions of heavy metals in the arable horizons compared with the humus horizons of the natural soils. A high anthropogenic load led to an increase of Cu, Pb and Zn content in urban soils, which was also reflected in the high values of the pollution indices.

About 40% of the collected samples were characterized as potentially contaminated, and 3% – as contaminated by heavy metals. Although, the overall soil contamination was characterized as not dangerous according to the Saet-based Comprehensive index, a more detailed and strict contamination monitoring is required at the revealed sites of the elevated heavy metals content.

Geostatistical mapping allowed us to reveal several patterns of regional distribution of heavy metals and confirmed an anthropogenic influence to the Cu, Pb and Zn regional distribution. The PCA allowed distinguishing three regional geochemical groups of heavy metals, confirming previous findings about anthropogenic origin of Cu, Pb and Zn in the topsoils.

This case study demonstrates that a combination of multivariate statistics and geostatistical analysis together with the pollution assessment allows comprehensive characterizing heavy metals spatial distribution and determining their sources. In addition, this study provides a basis for further comparisons and assessments to monitor a long-term heavy metal accumulation.

References


19. ND 52.18.191-89. Metodika vypolneniya izmereniy massovoy doli kislotorastvorimykh form metallov (medi, svintsa, tinika, nikelya, kadmiya) v probakh pochvy atomno-absorbtsionnym analizom [Methods of estimating the mass percentage of acid soluble metal forms (copper, lead, zinc, nickel, cadmium) in soil samples by atomic absorption analysis]. Moscow: Gosudarstvennyy komitet SSSR po gidrometeorologii, 1990. [In Russian]


23. Vodyanitskiy Yu. N. Tyazhelye metally i metalloidy v pochvakh [Heavy metals and metalloids in soils]. Moscow: GNU Pochvennyy institut im. V. V. Dokuchaeva RASKhN, 2008, 86 p. [In Russian]


