JOURNAL
OF THE BULGARIAN GEOGRAPHICAL SOCIETY

Application of geographic data for spatial modeling of lead in contaminated fluvial soils

Tsvetan Kotsev, Velimira Stoyanova

National Institute of Geophysics, Geodesy, and Geography – Bulgarian Academy of Sciences, Sofia, Bulgaria
E-mails: tsvetankotsev@mail.bg, stoyanovavelimira@gmail.com

Key words:
maximum entropy, MaxEnt, mine tailings, river floodplain, soil pollution

ABSTRACT

The present study aims to determine the spatial distribution of soils with lead (Pb) content above the quality thresholds in a section of the Ogosta River valley (NW Bulgaria). The study area was contaminated with mine waste from the extraction and flotation of iron, lead-silver, and gold-bearing ores in the second half of the XX century. Predictive modeling was performed with the software Maximum Entropy Species Distribution Modeling (MaxEnt), Version 3.4.4, which uses machine learning algorithms and applies the maximum entropy method. The choice of predictors of contaminated soil distribution is consistent with the main factor for Pb dispersal within the valley floor - flooding from the Ogosta River. The following six parameters explained the environmental settings related to the accumulation of contaminated floodplain sediment: vertical distance to the river channel, lateral distance to the Ogosta River, terrain slope, land cover (CORINE Land Cover, 2019), morphographic units of topography, and elevation. The results represent the average values of 10 replicates of the model. We evaluated the individual models by the value of the area under the relative operating characteristic curve (AUC) and the geographic logic of the obtained results. The AUC score for the test samples was 0.666 for the soil group 1 with Pb ≤120 mg/kg, 0.782 for group 2 with Pb (120-500] mg/kg, and 0.934 for group 3 with Pb>500 mg/kg. The most significant predictors for the models are the vertical and lateral distance to the river and the slope of the terrain. Lead concentrations tend to decrease with the distance from the main river and by increasing the elevation above the river channel due to lower inundation frequency and deposition rate of polluted river sediments. The soils with a Pb concentration below the permissible threshold of 120 mg/kg cover more than 58.42% of the valley floor of the studied section, and lands with Pb content above the intervention value of 500 mg/kg occupy nearly 10.82% of the investigated territory. The selected predictors describe the distribution of highly contaminated soils well and define the range of soils with lower Pb content worse. Combining clean and contaminated soil samples into one group is considered the main reason for the poor performance of MaxEnt for soils with Pb ≤120 mg/kg. However, the results prove the model's ability to predict the spatial distribution of not only biological species but also the dispersal of hazardous substances in soil.

Article processing
Submitted: 2 October 2022
Accepted: 22 November 2022
Published: 25 November 2022
Academic editor: Bilyana Borisova

1. Introduction

Contamination of soil with the potentially toxic element of lead (Pb) is a global issue due to its worldwide occurrence (Kisku et al., 2000; Markus and McBratney, 2001; Zeng et al., 2007; Collin et al., 2022). Kabata-Pendias and Pendias (1984) consider the average natural content of Pb in soils worldwide to be near 25 mg/kg. Human activity has contributed to considerable increases in the concentration of Pb in soils, especially along roadsides due to the use of leaded gasoline by vehicles in the past, near battery recycling sites and buildings painted with lead dyes, and mainly in the vicinity of lead-zinc ore mines (Mapanda et al., 2005; Maneva and Vatchev, 2013; Balkhair and Ashraf, 2016; Kan et al., 2016; Brown, 2016; Laidlaw et al., 2017; Pan et al., 2018; Milke et al., 2022). Lead tends to accumulate in the soil, and once deposited in the past, it can still be a problem today for people and the environment (Markus and McBratney, 2001). The

World Health Organization has listed Pb as one of the ten chemicals of highest public health concern (Collin et al., 2022).

According to the US Environmental Protection Agency, lead exposure is a health concern, especially for young children and pregnant women (EPA, 2020). Lead can affect almost every organ and system in the human body. The health effects include damage to the nervous system of children, which can cause reduced IQ and attention span, hyperactivity, impaired growth, and learning disabilities (Markus and McBratney, 2001; Peng et al., 2019). This necessitates investigation and quantification of the potential risk of lead contamination. Numerous studies have collected data on Pb content in natural and contaminated soils at various scales. However, only some of these studies have made serious attempts to explain the spatial distribution of contaminants by using geographical data (Cattle et al., 2002; McGrath et al., 2004; Liu et al., 2006; Qi et al., 2016). Creating maps of Pb content in soil can help identify the patterns of the spatial distribution of the trace metal. The maps can outline pollution hotspots with dangerously high element concentrations and are essential for site assessment and subsequent risk assessment (Markus and McBratney, 2001). The high cost of chemical analyzes and the time required to obtain chemical data at sufficient locations to enable mapping are often prohibitive. An alternative to extensive sampling is the application of regression and machine-learning models to explain the spatial distribution of contaminants in soil. Very often, such models are complex to apply and require specific knowledge and experience, which limits their widespread use. The MaxEnt predictive model features a user-friendly interface and can work with a relatively small number of sampling sites. MaxEnt uses the maximum entropy method and a machine learning algorithm and has been mostly applied to determine the distribution range of biological species within a territory (Merow et al., 2013; Glover-Kapfer, 2015). We do not find in the available literature applications of MaxEnt for determining the spatial distribution of soils contaminated with hazardous substances, although the model allows other applications than species distribution modeling.

The present study aims to determine the range of soils with different levels of lead contamination in the Ogosta River valley in NW Bulgaria by applying the MaxEnt predictive model and geographical variables as predictors.

2. Study area and research methodology

2.1. Study area

The study area is located in the upper stretch of the Ogosta River valley between the villages of Belimel and Gavril Genovo (Figure 1). The site covers 571.98 ha with an elevation range between 214.72 and 294.66 m. Extraction and dressing of iron, lead-silver and gold-bearing ores took place in the upper reach of the river basin near the town of Chiprovtsi from 1951 to 1999. Due to a tailings dam failure in 1964 and mine waste discharge into the Ogosta River afterward until 1979, the soils in the floodplain of the Ogosta Valley received significant amounts of arsenic, lead, and other potentially toxic elements (Jordanova et al., 2013).

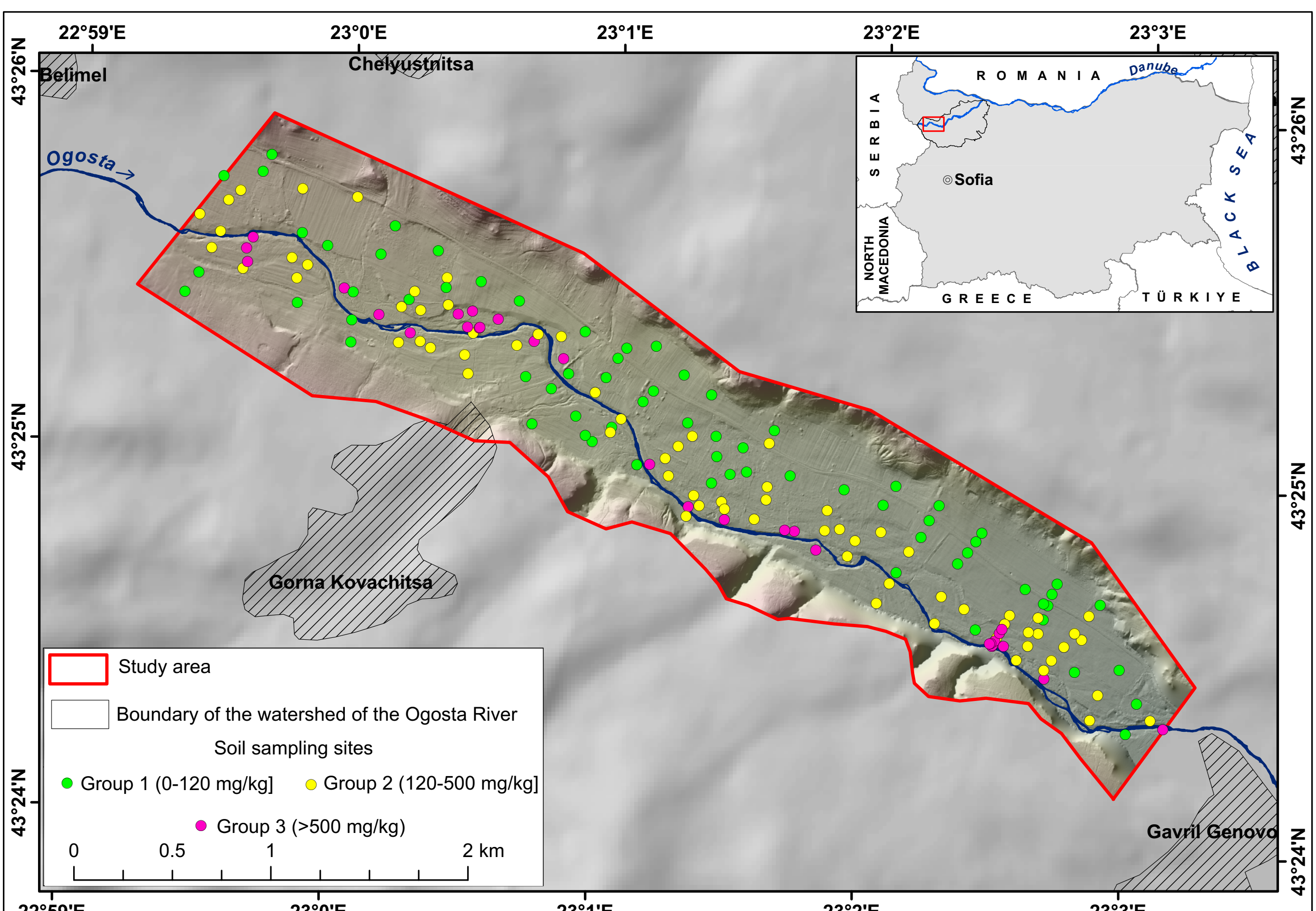

Figure 1. Study area

2.2. Field and laboratory methods

This study uses data from several sampling campaigns conducted in 2010-2020 (Mandaliev et al., 2014; Simmler et al., 2016; Tcherkezova et al., 2019). The obtained results are compatible due to the same soil sampling procedures, sample preparation, and measurement methods.

The sampling considered the morphographic features of the floodplain and was complemented with cross-sections semi-perpendicular to the Ogosta River. The total number of the measured soil samples is 168, of which 141 topsoil samples (0-20 cm) were collected in 2010-2017 and 27 topsoil samples in 2020. The soil material is air-dried, then crushed manually in a porcelain mortar and sieved through a <2 mm wire mesh made of stainless steel. The soil fraction <2 mm was ground with a planetary ball mill to a fine powder. Pellets were then prepared from 4 g soil material and 0.9 g of an amide wax (N.N′-Bisstearoylethylenediamide, Licowax C, Clariant). The Pb and other elements were measured in the pellets by X-ray fluorescence spectrometry (XRF).

2.3. Modeling concept

The spatial distribution of Pb content in the soils of the valley floor was modeled using the Maximum Entropy Species Distribution Modeling (MaxEnt) software, Version 3.4.4 (Phillips et al., 2006). The results of MaxEnt were visualized with ESRI ArcGIS 10.6.1 software product. The MaxEnt model was developed to determine the geographic distribution of biological species based on their ecological requirements for their surroundings using only presence data (Merow et al., 2013; Dai et al., 2022). Like other species distribution models, MaxEnt predicts the probability of the presence of a given species in a particular territory. This probability is relative and is determined by comparing the conditions where the species is present to the rest of the studied area (Phillips et al., 2006; Elith et al., 2011). The concept of the model allows its usage for establishing the spatial distribution of geographical phenomena and objects, e.g., soils with specific concentrations of potentially toxic elements. For such applications of species distribution models, it is necessary to study a sufficiently narrow range of the modeled variable's magnitude so that its values correspond to relatively uniform environmental conditions. Thus, each group of presence data will relate to the environmental settings as an individual biological species to its habitat.

2.4. Predictors

The choice of predictors of contaminated soil distribution is consistent with the main factor for Pb dispersal within the valley floor - flooding from the Ogosta River. The following six parameters explained the environmental settings related to the accumulation of polluted floodplain sediment: elevation, vertical distance to the river channel, lateral distance to the main river, terrain slope, land cover, and morphographic units of topography.

The *elevation* was determined using a digital terrain model (DTM) with a resolution of 1x1 m. It was created from aerial images taken in 2019 with a WingtraOne VTOL mapping drone equipped with a Sony 42 Mpix (DSC-RX1RM2) camera at a flight altitude of 400 m and a focal length of approx 35 mm. We used high-precision GNSS GPS to measure ground control points and Pix4Dmapper photogrammetry software to process the aerial images (Dinkov et al., 2020). The raster layer‘s *vertical distance to the river channel* was created using the GRASS GIS implemented in QGIS (Tcherkezova, 2021). We used the Euclidean Distance tool of Spatial Analyst Tools of ArcMap to generate the layer *lateral distance to the Ogosta River*. The *terrain slope* was calculated from the detailed DTM applying Spatial Analyst Tools – Surface – Slope.

The *land cover* was digitized from the high-quality orthophoto mosaics generated from the aerial photos, allowing us to achieve significantly higher accuracy than satellite images. Therefore, we used CORINE nomenclature for the fourth level, developed for the PHARE countries and corresponding to M 1:50 000. Eleven land cover classes were identified: Discontinuous built-up areas with family houses with gardens; Areas of special installations; Arable land prevailingly without dispersed (line and point) vegetation; Agricultural areas with a significant share of natural vegetation, and with the prevalence of grasslands; Agricultural areas with a significant share of natural vegetation, and with the prevalence of scattered vegetation; Agricultural areas with a significant share of permanent crops, and with the presence of scattered vegetation; Agricultural areas with a significant share of permanent crops, and with the presence of scattered vegetation; Broad-leaved forests with discontinuous canopy, not on the mire; Natural grassland prevailingly without trees and shrubs; Natural grassland with trees and shrubs; Natural young stands and Fresh-water marshes with reeds (Stoyanova, et al., 2020).

Morphographic units of topography were defined by classifying the values of the vertical distance to the main river. The first step involved creating an ECD.file with the ArcToolbox tool – Spatial Analyst Tools – Segmentation and Classification – Train ISO Cluster Classifier. In the second step, the vertical elevation was classified using ArcToolbox – Spatial Analyst Tools – Segmentation and Classification – Classify Raster. As a result of the classification, four primary morphographic levels were outlined on the valley floor at approximately 0.2-0.7 m, 1-2 m, 2-3 m, and 4-7 m above the river. They correspond to the bankfull channel (0-1 m), active floodplain (1-3 m), and high floodplain (3.0-6.5 m), respectively, as defined by Tcherkezova (2015). The active floodplain is also referred to as a low floodplain in the text below.

2.5. Grouping of Pb values

The samples were divided into three groups according to the maximum permissible concentration (120 mg/kg) and the intervention value of Pb (500 mg/kg) in the soils of arable and grasslands according to Bulgarian Regulation on the permissible content of harmful substances in soils (Regulation 3, 2008). The number of samples and intervals of Pb for each group were as follows [mg/kg]: Group 1 (0-120], 68 samples; Group 2 (120-500], 70 samples; Group 3 (>500), 30 samples (Figure 1).

2.6. Model settings

For each group of soil samples, separate modeling was performed with MaxEnt using the set of predictors specified above. The results represent the average values of 10 replicates of the model. The model requires a file containing the coordinates of the samples in .csv format (Comma-separated values) and a directory containing the environmental variables in ASCII format. The MaxEnt model offers four output formats - Gloglog, Logistic, Cumulative, and Raw. We used MaxEnt's logistic output type since it is interpreted as the probability of the presence of the research object with a value from 0 to 1. The software provides Crossvalidate, Bootstrap, and Subsample methods for the validation of the model. We selected the bootstrap method due to the small number of presence data points in the separate groups indicated above. To determine the area occupied by soils with the specific content of Pb, we converted the continuous probability of the presence of these soils into a map of the suitable and unsuitable areas of their occurrence. We used the logistic threshold of maximum test sensitivity plus specificity for the conversion as recommended by Jiménez-Valverde & Lobo (2007). At this threshold, calculated by MaxEnt, the model maximizes the discrimination of presence data from background data.

3. Results

The model's results were evaluated according to the following two criteria: the value of the statistical indicator area under the relative operating characteristic curve (AUC) and the geographic logic of the obtained results. Regarding AUC, there are no generally accepted thresholds for evaluation. In this study, we used those indicated by Araújo et al. (2005). According to the authors, the model can be considered fail if AUC is in the range of 0.5 - 0.6, poor 0.6 - 0.7, satisfactory 0.7 - 0.8, good 0.8 - 0.9, and excellent if AUC>0.90 To evaluate the model, we used the AUC, which was calculated for the test data and referred to as AUC_{test} in the text below.

By the geographic logic of the results, we understand the absence of contradiction between the factors and conditions that control the modeled feature and its spatial distribution produced by the model. For example, we consider it logical if we find high contaminant levels close to the river and low concentrations at the far ends of the valley floor. Vice versa would be illogical because the river is the primary source of contamination in the valley.

3.1. Group 1 - Probability of presence of soil with a concentration of Pb <120 mg/kg

The concentration of Pb in the collected soil samples from group 1 ranges between 19.7 – 117.7 mg/kg with an average of 54.5 mg/kg, a median of 43.9 mg/kg, a mode of 27.6 mg/kg, and a standard deviation of 29.7 mg/kg. The modal value is close to that for floodplain sediments in Europe - 18.0 mg/kg (Salminen et al., 2005), and the median is in the range of the target values for Bulgarian soils, which vary between 40.0 and 50.0 mg/kg depending on the soil texture (Regulation 3, 2008). At the same time, the contaminant concentration exceeds the percentile 90 for floodplain sediments in Europe (62.0 mg/kg) in 35% of the group samples.

The model of group 1 achieved an AUC of 0.774 (Figure 2). Validation of the model with the test data reaches an AUC_{test} value of 0.666, which defined the model for the first group as poor.

The MaxEnt model provides information on the contribution of each of the variables used in its performance. The estimate of the contribution, named permutation importance, is calculated based on the change in AUC due to randomly permuting the values of the variable at the training points. These points include the sampling sites and 10 000 pixels randomly selected by the program from the predictor layer. The more the AUC decreases after permutation, the more significant the impact of the predictor on the model. The spatial distribution of Pb concentrations (0-120] mg/kg depended mainly on the parameters of slope gradient and vertical distance to the river, followed by the lateral distance to the Ogosta River (Table 1). The elevation and morphographic units of topography contributed the least to the model.

Table 1. Permutation importance of predictors to the model of Group 1

Predictors	Code	Permutation importance (%)
Slope	slope	28.5
Vertical distance to the river channel	vdcn	20.1
Lateral distance to the Ogosta River	distance	19.7
Land cover	clc_2019	14.2
Elevation	dtm	9.2
Morphographic units of topography	gmu	8.3

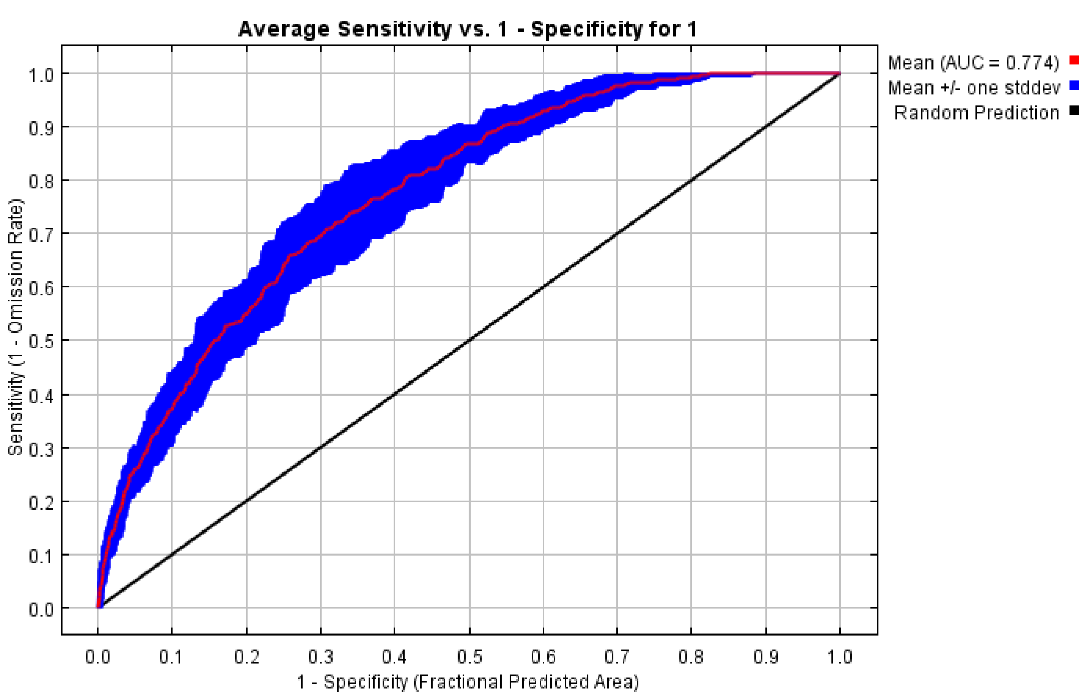

Figure 2. Averaged ROC (relative operating characteristic) curve for the calibrated model of Group 1

The importance of individual variables for model performance can also be assessed using the Jackknife test. In this, several versions of the model are estimated by successively dropping one of the predictors and modeling with the remaining variables. The estimation of these models is shown in light blue in Figure 3. In addition, the program creates models with each variable separately, which are represented by dark blue color. The model with all predictors is indicated by red color. The test confirms the highest significance of the slope gradient and distance to the river. Both variables explain to the greatest extent the distribution of the sampling sites of the group. The smallest contribution is revealed for the elevation and the morphology of the terrain. Dropping the slope from the model reduces its accuracy most significantly. This indicates that the slope parameter caries much information which is not contained in the other predictors. At the same time, the prediction is only slightly degraded when any other variable is dropped. It clearly shows that the rest of the variables do not bring a sufficient amount of new information to the model compared to the parameter of slope gradient.

The model calculated the value 0.4389 for the logistic threshold of maximum test sensitivity plus specificity. The sites with a probability higher than the threshold are considered to have Pb ≤120 mg/kg in the soil, while the area below the threshold is expected to have higher contaminant levels.

The first class of presence probability <0.4389 covers an area of 233.70 ha, which is 58.42% of the valley floor, while the second class >0.4389 includes 166.31 ha or 41.58% (Figure 4).

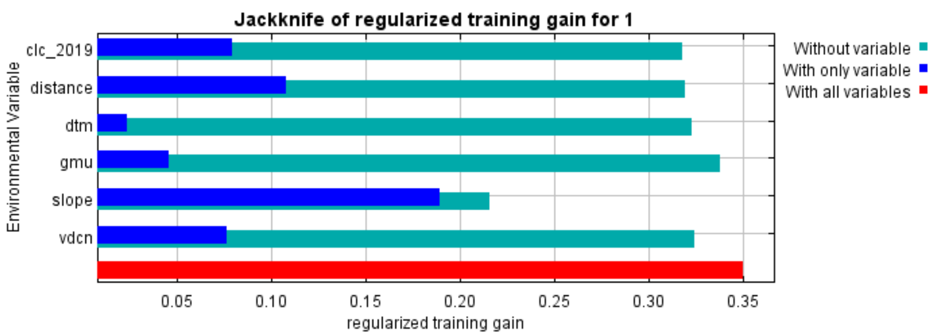

Figure 3. Contribution of individual variables to the model of Group 1

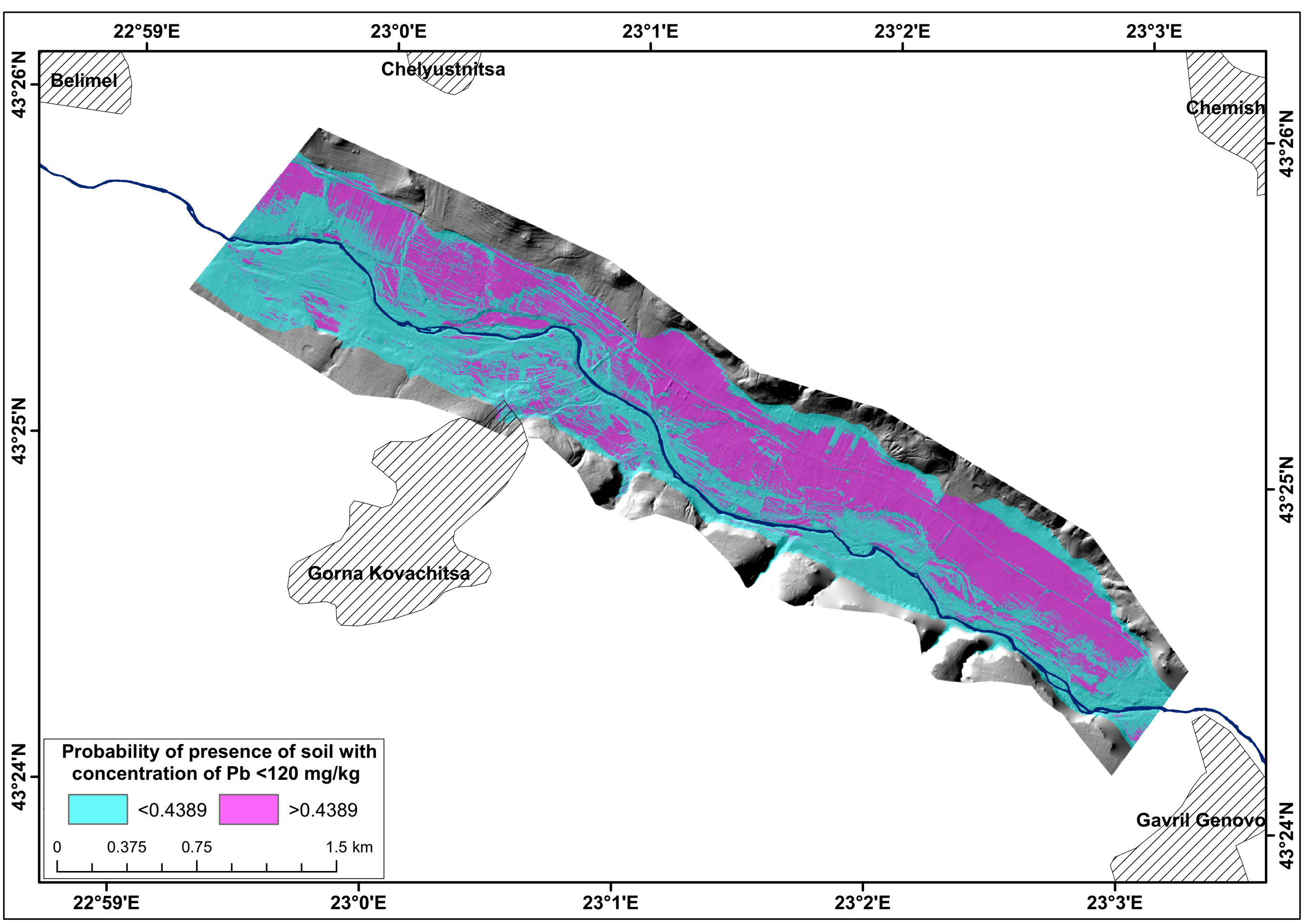

Figure 4. Suitability of the area for soils with a concentration of Pb ≤120 mg/kg. Presence probability >0.4389 indicates more suitable territories, and <0.4389 less suitable

3.2. Group 2 - Probability of presence of soil with a concentration of Pb 120-500 mg/kg

The Pb contents of the soil samples of the second group ranged from 122.6 to 499.6 mg/kg with a mean value of 254.4 mg/kg, a median of 227.0 mg/kg, a mode of 134.6 mg/kg, and a standard deviation of 100.9 mg/kg.

The model of the second group was characterized by an AUC value of 0.829 (Figure 5). Based on AUC_{test} 0.782, the model for the group was evaluated as satisfactory.

The vertical distance to the river channel is the most important contributor to the model of group 2, followed by the slope and the lateral distance to the Ogosta River, both with a twofold smaller importance (Table 2). As in the previous group, the morphographic levels of the terrain are of the slightest use for preparing the predictive forecast.

The Jackknife test confirms the leading role of the vertical distance to the river channel for the model of Group 2. It relegates the lateral distance to the river to second place in the variables' contribution to the prediction. The models built with only one of the vertical or lateral distances have the highest accuracy. In contrast, the forecast with only elevation or land cover is wildly inaccurate. Models without the participation of slope and vertical distance strongly degrade their performance. This fact confirms the informativeness of both indicators for environmental conditions typical of soils with lead contents of (120-500] mg/kg.

Table 2. Permutation importance of predictors to the model of Group 2

Predictors	Code	Permutation importance (%)
Vertical distance to the river channel	vdcn	44.7
Slope	slope	19.4
Lateral distance to the Ogosta River	distance	17.6
Land cover	clc_2019	8.5
Elevation	dtm	7.4
Morphographic units of topography	gmu	2.4

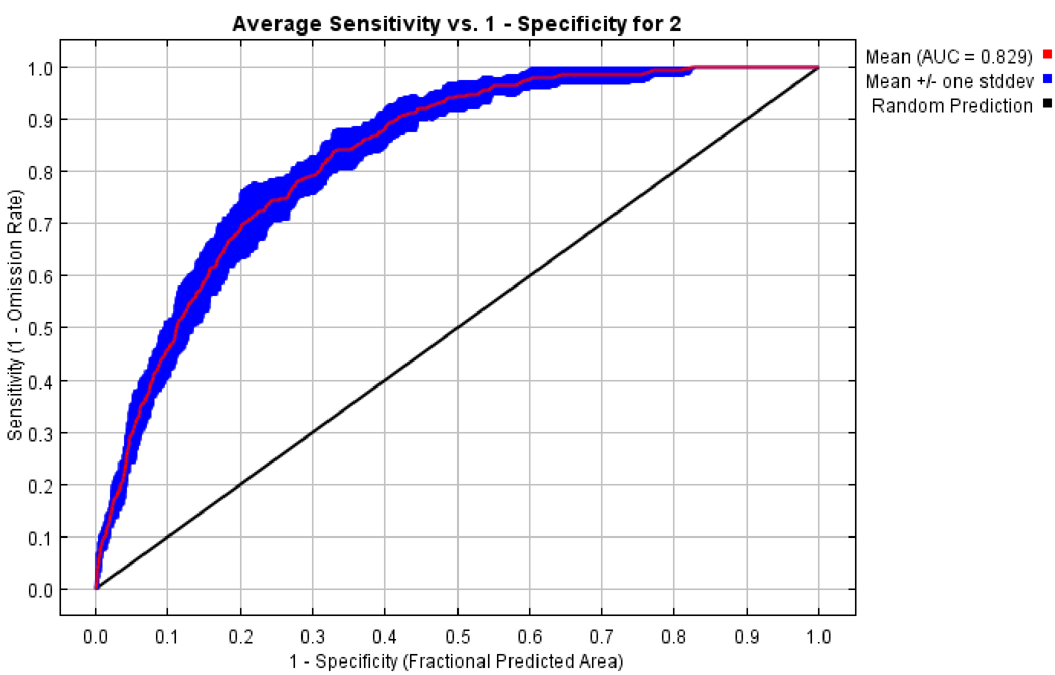

Figure 5. Averaged ROC (relative operating characteristic) curve for the calibrated model of Group 2

The model defines the value 0.3964 as the logistic threshold of the maximum test sensitivity plus specificity. The area with a probability of occurrence below the threshold covers 268.37 ha, occupying two-thirds of the valley floor (67.10%). The suitable territory for the soils of Group 2 extends over 131.64 ha and is twice smaller than the unsuitable territory (32.91%) (Figure 7).

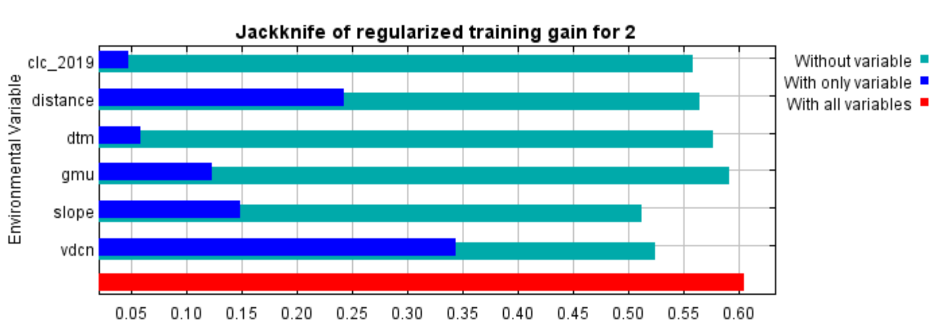

Figure 6. Contribution of individual variables to the model of Group 2

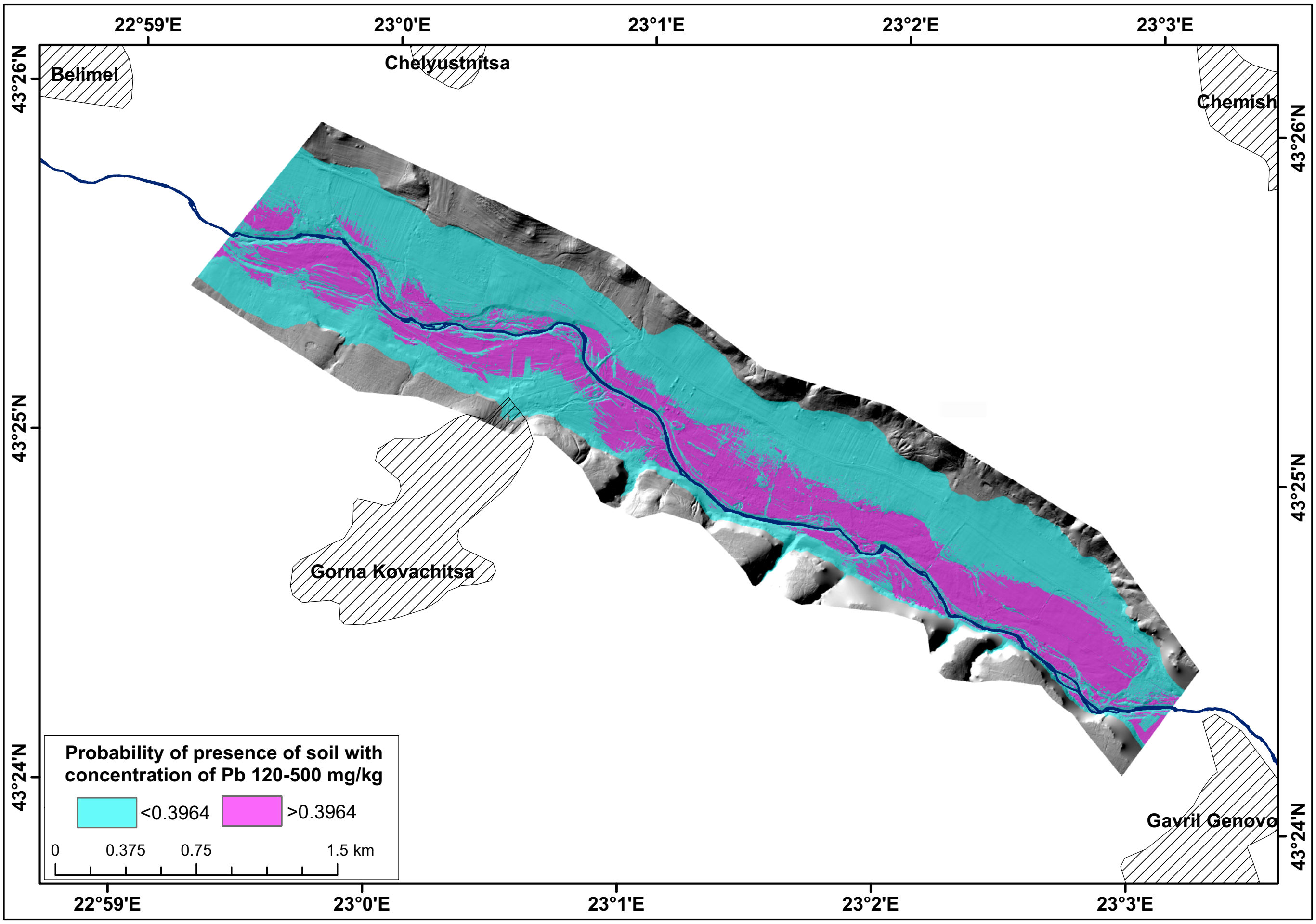

Figure 7. Suitability of the area for soils with a concentration of Pb 120-500 mg/kg. Presence probability >0.3964 indicates more suitable territories, and <0.3964 less suitable

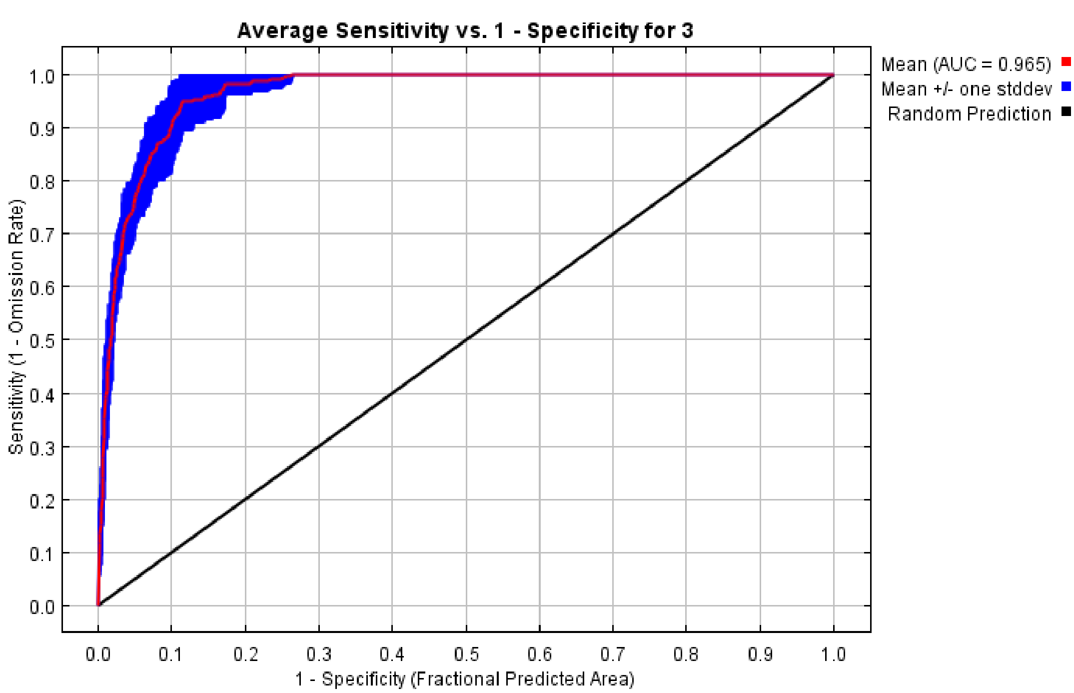

Figure 8. Averaged ROC (relative operating characteristic) curve for the calibrated model of Group 3

3.3. Group 3 - Probability of the presence of soil with a concentration of Pb >500 mg/kg

The content of Pb in the soil samples of the third group ranged from 508.3 to 1550.0 mg/kg with a mean of 784.4 mg/kg, a median of 739.6 mg/kg, mode 553.7 mg/kg and a standard deviation of 240.5 mg/kg.

The AUC value for the model of group 3 is 0.965 (Figure 8) and the AUC_{test} is 0.934, which defines it as excellent.

Vertical distance to the river channel is the essential variable for the model (Table 3). The second variable, lateral distance to the river, has a fourfold smaller contribution. The permutation importance of each of the remaining metrics is a few percent, which determines their minor role in the model prediction.

Table 3. Permutation importance of predictors to the model of Group 3

Predictors	Code	Permutation importance (%)
Vertical distance to the river channel	vdcn	68.7
Slope	slope	16.5
Lateral distance to the Ogosta River	distance	5.7
Land cover	clc_2019	4.4
Elevation	dtm	3
Morphographic units of topography	gmu	1.8

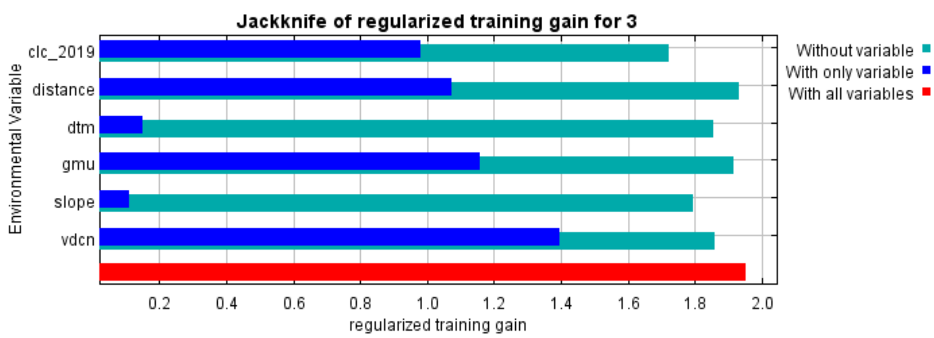

Figure 9. Contribution of individual variables to the model of Group 3

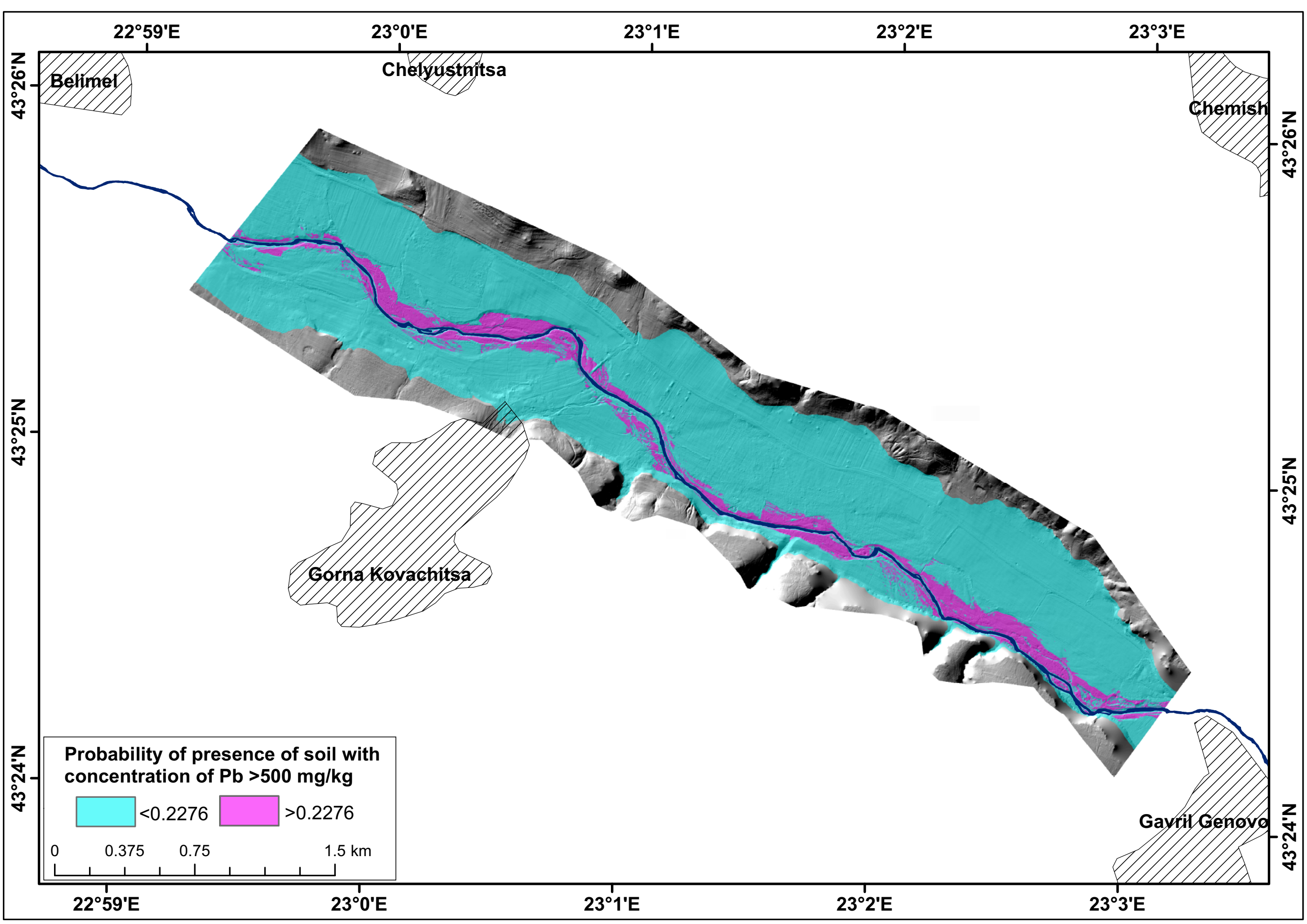

Figure 10. Suitability of the area for soils with a concentration of Pb >500 mg/kg. Presence probability >0.2276 indicates more suitable territories, and <0.2276 less suitable

The results of the Jackknife test show the best performance of the model composed of the vertical distance to the river (Fig. 9). The quality of the model that uses only morphographic units is close to the accuracy of the model with the vertical distance, followed by the individual models with the lateral distance and the land cover. The variables with the most unique information are the land cover and slope, where the difference between the light blue and red bars is the largest. Noteworthy is the minimal degradation of the models that drop the lateral distance or the morphographic unit metrics. It means that most of the information carried by the two variables is also available in the other predictors.

The logistic threshold of maximum test sensitivity plus specificity for the soils of Group 3 has a value of 0.2276. The class with a presence probability above the threshold covers an area of 43.27 ha (Figure 10). About 10.82% of the investigated valley floor is likely to be covered with soils with Pb >500 mg/kg. The class of presence probability below 0.2276 stretches over 356.73 ha (89.18%), where highly contaminated soils of group 3 are not expected to occur.

4. Discussion

Most soil samples in Group 1 have Pb content close to natural levels, but about one-third of the sampling sites have elevated contaminant concentrations. We find low concentrations more frequently in the highest morphographic level of the valley floor populated with more than 80% of the sites of the group (Table 4). The average Pb concentration in these points is 48.42 mg/kg, which is close to the target values for soils in Bulgaria (Regulation 3, 2008). The rest of the presence data of the group falls predominantly in the third morphographic level, and only less than 2% are in the second morpho unit.

The mean Pb concentration in the samples from the active floodplain is 82.73 mg/kg, which is 71% higher than the average value in the high floodplain and exceeds by 33% the percentile 90 for floodplain sediments in Europe. Contaminant concentrations near the natural levels are primarily measured in the highest morphographic unit of the valley floor, while elevated contents can be found most frequently in the third morphographic level. Interviews with local residents reveal the occasional inundation of the active floodplain by the Ogosta River since the mid-20th century to the present. On the contrary, there is no evidence of flooding of the high floodplain except in its lowest sections above the river bed. We assume a different origin of Pb content in the soil of the third and fourth morphographic units. Aesthetic pedogenesis dominates in the higher section and the accumulation of contaminated floodplain sediments in the lower areas.

Bringing together uncontaminated and contaminated soil samples into one pool increases the heterogeneity of environmental settings in the range of the group. Decreased homogeneity is a possible reason for the poor explanation of the spatial distribution of Pb by the predictors, which is confirmed by the low gain of the model for Group 1. Consequently, the model has difficulty distinguishing between sites with concentrations below 120 mg/kg from those with higher concentrations. It is illustrated by the low AUC and AUC_{test} values for Group 1. However, the model correctly locates the suitable area for soils with Pb≤120 mg/kg mainly in the high floodplain (82%), and only 12% of it falls in the adjacent lower morphographic level.

As with Group 1, the sites of Group 2 fall primarily in the third and fourth morphological levels, being almost equally distributed

Table 4. Lead concentration and main morphographic variables by data presence groups

Morphographic unit	Number of samples (N)	Share of the total number of samples [%]	Average concentration of Pb in the soil samples [mg/kg]	Average vertical distance to the river [m]	Average lateral distance to the river [m]	Share of the suitable area for the group [%]
Group 1, N=68						
1	0	0.0	-	-	-	0.0
2	1	1.5	19.7	1.8	22.8	0.0
3	11	16.2	88.5	2.5	75.2	12.0
1+2+3	12	17.7	82.7	2.4	70.8	88.0
4	56	82.4	48.4	6.3	269.9	0.0
Group 2, N=70						
1	0	0.0	-	-	-	0.1
2	13	18.6	337.7	1.6	29.8	15.2
3	21	30.0	256.8	2.6	69.6	37.2
1+2+3	34	48.6	287.7	2.2	54.4	52.5
4	36	51.4	223.0	4.4	165.7	47.5
Group 3, N=30						
1	0	0.0	-	-	-	0.2
2	12	40.0	714.8	1.7	13.0	41.0
3	18	60.0	831.7	2.4	39.5	58.8
4	0	0.0	-	-	-	0.1

between the two morpho units. The average Pb content in the lower morphological level is nearly 30% increased compared to the higher morpho unit. The difference is twice as slight as for Group 1, indicating better uniformity of environmental settings in Group 2. It is probably because most sampling points in the fourth morphographic unit occupy its lower parts, some of which may have been flooded by the river during the mining period. The more homogeneous environmental settings within the group expectedly resulted in an increased gain of the model and higher AUC and AUC_{test} values for Group 2 compared to Group 1. However, the mixing of high and low floodplain locations increases, to a certain extent, the range of variable values. Some of the sites of Group 2 can be found in river-irrigated orchards and vegetable gardens which are located in the higher and more distant parts of the high floodplain, typical of Group 1. Including presence data from contaminated irrigated lands in the periphery of the valley floor increases the variation of the environmental settings within Group 2 additionally.

The presence data of Group 3 fall entirely within the two morphological levels of the active floodplain. The difference between the mean Pb contents in the soils of the second and third morphographic units is only 16%, significantly less than the differences revealed above for the high and low floodplains. Clearly, the selected predictors best describe the prevalence of sites with elevated contaminant concentrations in the low floodplain environment. The better homogeneity of the environment settings covered by the group is the reason for the model's highest gain and AUC_{test} values compared to Group1 and Group 2.

5. Conclusion

This study applied for the first time the MaxEnt model for predictive modeling of the spatial distribution of soils with a specific content of potentially toxic elements. The model's results were assessed by two criteria: the value of the statistical indicator area under the relative operating characteristic curve (AUC) and the geographical logic of the obtained results. AUC_{test} values ranged from 0.666–0.934 for individual models.

The results of the study showed the applicability of MaxEnt to determine the distribution areas of heavy metals and metalloids in alluvial soils contaminated with mine wastes through river inundation. Different combinations of the variables explained the spatial distribution of soils with varying degrees of contamination. Terrain slope, vertical distance to the river channel, and lateral distance to the Ogosta River were the most important predictors for the distribution of soils with low Pb content. At the same time, highly contaminated sites were closely related to the vertical and lateral distance to the river.

The results reveal the main regularities of Pb distribution in soils of the valley floor—the concentrations of the element decrease by the lateral distance to the Ogosta River and by the elevation above the river bed. The inundation frequency and the deposition intensity of polluted river sediments control the distribution pattern of Pb. Its values >500 mg/kg were usually found in low, frequently flooded areas along the river, while contents ≤120 mg/kg could be expected mainly in the higher and more peripheral parts of the valley floor.

It can be concluded that the selected predictors describe well the distribution of highly contaminated soils and define worse the range of soils with lower Pb content. Combining clean and contaminated soil samples into one group is the main reason for the poor performance of MaxEnt for the soils with Pb ≤120 mg/kg. However, the results prove the model's ability to predict not only the spatial distribution of biological species but also the dispersal of hazardous substances in soil.

Acknowledgments

This research was supported by the project "Relationship of the spatial distribution of heavy metals in the soil with the morphology of polluted floodplain river terraces", with the short name TOPOMET, financed by the "Scientific Research" Fund, contract No. KP-06-H24/2 (08.12.2018) with base organization National Institute of Geophysics, Geodesy, and Geography, Bulgarian Academy of Sciences. The authors express their gratitude to Prof. Diana Jordanova, Prof. Dsc. Daniela Jordanova of the National Institute of Geophysics, Geodesy, and Geography - BAS, and Associate Professor Metodi Karadjov and the management of the Geological Institute of the Bulgarian Academy of Sciences for the provided technique for grinding the soil samples. The authors of this paper acknowledge that the original idea of applying Maxent for modeling the geographic distribution of contaminated soils was proposed by Todor Lyubenov in 2012.

References

Balkhair, K.S., Ashraf, M.A. (2016) Field accumulation risks of heavy metals in soil and vegetable crop irrigated with sewage water inwestern region of Saudi Arabia, Saudi J. Biol. Sci., 23 (1), S32-S44. https://doi.org/10.1016/j.sjbs.2015.09.023.

Brown, S.L, Chaney, R.L., Hettiarachchi, G.M. (2016) Lead in Urban Soils: A Real or Perceived Concern for Urban Agriculture? Journal Environmental Quality 45(1): 26-36. https://doi.org/10.2134/jeq2015.07.0376.

Cattle, J., McBratney, A., Minasny, B. (2002) Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination, Journal Environmental Quality, (5):1576-88. https://doi.org/10.2134/jeq2002.1576.

Collin, S., Baskar, A., Geevarghese, D.M., Ali, M. Bahubali, P., Choudhary, R., Lvov, V., Tovar, G., Senatov, F., Koppala, S., Swamiappan, S. (2022) Bioaccumulation of lead (Pb) and its effects in plants: A review. Journal of Hazardous Materials Letters, 3. https://doi.org/10.1016/j.hazl.2022.100064.

Dai X., Wu W., Ji L., Tian S., Yang B. ,4 Guan B., Wu D. (2022) MaxEnt model-based prediction of potential distributions of Parnassiawightiana (Celastraceae) in China, Biodivers. https://doi.org/10.3897/BDJ.10.e81073.

Dinkov, D., Kitev, A., Hristova, D. (2021) High-resolution topographic data of river landscapes using aerial PPK-UAV photogrammetry: a case study in Ogosta River valley, Northwestern Bulgaria. Proceedings of the SGEM International Multidisciplinary Scientific GeoConference, 21, 2.1: 547-554.

Elith, J., Phillips, S., Hastie, T., Dudik, M., Chee, Y., Yates, C. (2011) A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17:43–57.

Glover-Kapfer, P. (2015) A training manual for habitat suitability and connectivity modeling using tigers (Panthera tigris) in Bhutan as example. Technical Report. WWF-Bhutan.

Jiménez-Valverde, A., Lobo, J.M. (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica, 31: 361-369.

Jordanova, D., Goddu, S.R., Kotsev, T., Jordanova, N. (2013) Industrial contamination of alluvial soils near Fe-Pb mining site revealed by magnetic and geochemical studies. Geoderma 192: 237-248.

Kabata-Pendias, A., Pendias, H. (1984) Trace elements in soil and plants. Boca Raton, Florida: CRC Press Inc.

Kan, X., Dong, Y., Feng, L., Zhou, M., Hou, H. (2020) Contamination and health risk assessment of heavy metals in China's lead-zinc mine tailings: A meta-analysis. Chemosphere 267:128909. https://doi.org/10.1016/j.chemosphere.2020.128909.

Kisku, G.C., Barman, S.C., Bhargava, S.K. (2000) Contamination of Soil and Plants with Potentially Toxic Elements Irrigated with Mixed Industrial Effluent and its Impact on the Environment. Water, Air, & Soil Pollution 120, 121–137, https://doi.org/10.1023/A:1005202304584

Laidlaw, M., Filippelli, G., Brown, S., Paz-Ferreiro, J., Reichman, S., Netherway, P., Truskewycs, A., Ball, A., Mielke, H. (2017) Case studies and evidence-based approaches to addressing urban soil lead contamination. Applied Geochemistry 83, 14-30.

Liu X, Wu J, Xu J. (2006) Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS, Environmental Pollutution 141(2):257-64. https://doi.org/10.1016/j.envpol.2005.08.048.

Mandaliev, P., Mikutta, Ch., Barmettler, K., Kotsev, Ts., Kretzschmar R. (2014) Arsenic species formed from arsenopyrite weathering along a contamination gradient in circumneutral river floodplain soils. Environmental Science & Technology 48(1): 208–217.

Maneva, S., Vatchev, T. (2013) Effect of Lead Excess in Soil on the Disease Severity of Fusarium oxysporum f. sp. radicis-lycopersici (FORL) on Tomatoes. Regression Models for Estimatio. Soil science agrochemistry and ecology 2: 59-67.

Mapanda, F., Mangwayana, E.N., Nyamangara, J., Giller, K.E. (2005) The effect of long-term irrigation using wastewater on heavy metal contents of soils under vegetables in Harare, Zimbabwe, Agric. Ecosyst. Environ., 107(2–3), 151-165.

Markus, J. McBratney, A. (2001) A review of the contamination of soil with lead: II. Spatial distribution and risk assessment of soil lead. Environment International, 27, 5: 399-411.

McGrath, D., Zhang, C., Carton, T. (2004) Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland Environmental Pollution 127(2): 239-48. https://doi.org/10.1016/j.envpol.2003.07.002.

Merow C., Smith J., Silander J. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36: 001–012. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Mielke, H.W., Gonzales, C.R., Powell, E.T., Egendorf, S.P. (2022) Lead in Air, Soil, and Blood: Pb Poisoning in a Changing World. Int. J. Environ. Res. Public Health 2022, 19(15), 9500. https://doi.org/10.3390/ijerph19159500.

Pan, L., Fang, G., Wang, Y., Wang, L., Su, B., Li, D., Xiang, B. (2018) Potentially Toxic Element Pollution Levels and Risk Assessment of Soils and Sediments in the Upstream River, Miyun Reservoir, China, Int J Environ Res Public Health, 15(11): 2364. https://doi.org/10.3390/ijerph15112364.

Peng, T., O'Connor, D., Zhao, B., Jin, Y., Zhang, Y., Tian, L., Zheng, N., Li, Xiaoping, Hou, D. (2019) Spatial distribution of lead contamination in soil and equipment dust at children's playgrounds in Beijing, China, Environmental Pollution. https://doi.org/10.1016/j.envpol.2018.11.011.

Phillips S.J. Anderson, R.P. Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions, Ecol. Model., 190: 231-259.

Qi J, Zhang H, Li X, Lu J, Zhang G. (2016) Concentrations, spatial distribution, and risk assessment of soil heavy metals in a Zn-Pb mine district in southern China, Environmental Monitoring Assessment 188(7):413. https://doi.org/10.1007/s10661-016-5406-0.

Regulation No 3 on 1 August 2008 on the permissible content of harmful substances in soil. *State Gazette* 71, issued on August 12, 2008. Available online at: https://lex.bg/laws/ldoc/2135595432 (last accessed on 21/11/2022).

Salminen R. (Chief Editor), Batista M.J., Bidovec M., Demetriades A., De Vivo B., De Vos W., Duris M., Gilucis A., Gregorauskiene V., Halamic J., Heitzmann P., Lima A., Jordan G., Klaver G., Klein P., Lis J., Locutura J., Marsina K., Mazreku A., O'Connor P.J., Olsson S.Å., Ottesen R.T., Petersell V., Plant J.A., Reeder S., Salpeteur I., Sandström H., Siewers U., Steenfelt A. and Tarvainen T. (2005) FOREGS Geochemical Atlas of Europe, Data: FOREGS-EuroGeoSurveys Geochemical Baseline Database: Instructions. Geological Survey of Finland, Espoo. Available online at: http://weppi.gtk.fi/publ/foregsatlas/ForegsData.php (last accessed on 21/11/2022).

Simmler, M., Suess, E., Christl, I., Kotsev, T., Kretzschmar, R. (2016) Soil-to-plant transfer of arsenic and phosphorus along a contamination gradient in the mining-impacted Ogosta River floodplain. Science of Total Environment 572:742-754, ISSN:0048-9697.

Stoyanova, V., Kotsev, Ts., Tcherkezova, E., Zhelezov, G., Koleva, N. (2020) Land cover changes in the period 1993-2019. Proceedings of the SGEM International Multidisciplinary Scientific GeoConference, 20, 2.2: 233-240.

Tcherkezova, E. (2015) GIS-based delineation and regionalization of geomorphographic units in the floodplain of the Ogosta river between the settlement of Gavril Genovo and the "Ogosta" reservoir (NW-Bulgaria). Problems of Geography, 1-2: 114-122.

Tcherkezova, E. (2021) Hydro-geomorphological descriptors for detection of flood-prone areas using high-resolution UAV photogrammetry data. SGEM2021 Conference Proceedings, Photogrammetry and Remote Sensing, 21, 2.1: 301-310.

Tcherkezova, E., Stoyanova, V., Kotsev, Ts. (2019) A Concept of an Integrated Geodatabase for Surface Water, Soil and Groundwater Pollution with Arsenic in the Upper Part of Ogosta Valley, Northwestern Bulgaria. European Journal of Geography, 10, 3: 6-23.

U.S.Environmental protection agency, region III (2020) Lead in Soil - https://www.epa.gov/sites/default/files/2020-10/documents/lead-in-soil-aug2020.pdf [Accepted: 04.11.2022]

Zeng, L., Liao, M., Chen, L., Huang, C. (2007) Effects of lead contamination on soil enzymatic activities, microbial biomass, and rice physiological indices in soil–lead–rice (Oryza sativa L.) system. Ecotoxicology and Environmental Safety, 67, 1: 67-74.

ORCID

https://orcid.org/0000-0002-8562-3164 - TK
https://orcid.org/0000-0001-6632-7514 - VS