Spatial Prediction of Heavy Metal Soil Contents in Continental Croatia Comparing Machine Learning and Spatial Interpolation Methods

Radočaj, Dorijan; Jurišić, Mladen; Župan, Robert; Antonić, Oleg

UDK 351.777.6:517.518.85:004.85:332.36
Original scientific paper / Izvorni znanstveni članak

Spatial Prediction of Heavy Metal Soil Contents in Continental Croatia Comparing Machine Learning and Spatial Interpolation Methods

Dorijan RADOČAJ, Mladen JURIŠIĆ – Osijek[1], Robert ŽUPAN – Zagreb[2], Oleg ANTONIĆ – Osijek[3]

ABSTRACT. Soil contamination caused by heavy metals presents a potential long-term issue to human health and biodiversity due to the bioaccumulation effect. Previous research at the micro level in Croatia detected soil contamination caused by heavy metals above maximum permitted values, which also implied the necessity of their current spatial representation at the macro level in Croatia. The aim of this study was to provide a spatial prediction of six heavy metals considered as contaminants of soils in continental Croatia using two approaches: a conventional approach based on interpolation and a machine learning approach. The prediction was performed on the most recent available data on cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), lead (Pb) and zinc (Zn) concentrations in soils, from the Ministry of environment and energy. The conventional prediction approach consisted of the interpolation using the ordinary kriging (OK) in case of input data normality and stationarity, alongside the inverse distance weighting (IDW) method. For the machine learning approach, random forest (RF) and support vector machine (SVM) methods were used. IDW outperformed RF and SVM prediction results for all soil heavy metals contents, primarily due to sparse soil sampling. Soil Cr contents were predicted above the maximum allowed limit, while elevated soil contamination levels in some parts of the study area were detected for Ni and Zn. The highest soil contamination levels were observed in the urban areas of generalized land cover classes, indicating a necessity for its monitoring and the adjustment of land-use management plans.

Keywords: soil contamination, random forest, support vector machine, ordinary kriging, inverse distance weighting, land cover.

[1] Dorijan Radočaj, mag. ing. geod. et geoinf., Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, HR-31000 Osijek, Croatia, e-mail: dradocaj@fazos.hr, Prof. Mladen Jurišić, PhD, Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, HR-31000 Osijek, Croatia, e-mail: mjurisic@fazos.hr,

[2] Assoc. Prof. Robert Župan, PhD, Faculty of Geodesy, University of Zagreb, Kačićeva 26, HR-10000 Zagreb, Croatia, e-mail: robert.zupan@geof.unizg.hr,

[3] Prof. Oleg Antonić, PhD, corresponding author, Department of Biology, Josip Juraj Strossmayer University of Osijek, Cara Hadrijana 8/A, HR-31000 Osijek, Croatia, e-mail: oleg.antonic@biologija.unios.hr.

1. Introduction

Bioaccumulation of heavy metals located in the soil presents a long-term danger to human health and vegetation stress, which affects biodiversity and enters the food chain through plant-based food consumption (Hu et al. 2020). Soil contamination by heavy metals has been a concern worldwide, with the detected contamination on agricultural areas (Atafar et al. 2010), urban areas (Hu et al. 2013) and forests (Yan et al. 2015). A previous study by Sollitto et al. (2010) explored soil contamination caused by heavy metals on a micro scale of Croatia, detecting multiple locations of critical soil anthropogenic contamination in Zagreb. The recent information about heavy metal soil content spatial distribution at the macro scale in Croatia was not found during a literature review, which disables a prompt soil contamination management in case of severe contamination. The analysis of soil contamination caused by heavy metals according to land cover classes is an indicator of the dominant exposure of heavy metals, enabling the adjustment of land-use management and the soil remediation planning (Zhao et al. 2012). Soil contamination by heavy metals is also a side-effect of growing urbanization globally and presents an issue for human health due to higher population density (Hu et al. 2013).

An accurate soil mapping presents a foundation for sustainable soil remediation and overall land-use management (Ottesen et al. 2008). The prediction of the spatial distribution of heavy metals in the soil is traditionally performed by spatial interpolation methods in a geographic information system (GIS). This process allows the determination of the continuous distribution of heavy metals over the entire study area. The collection of discrete soil sampling points requires extensive field and laboratory work. Sampled points allow a limited representation of the distribution of heavy metals on agricultural land, making the interpolation methods necessary for their monitoring (Radočaj et al. 2020a). Various geostatistical and deterministic interpolation methods were traditionally applied in the prediction of soil variables with high accuracy (Qiao et al. 2018). Geostatistical methods, consisting of kriging variations, were generally superior to deterministic methods for the majority of applications regarding interpolation accuracy and user subjective impact (Šiljeg et al. 2019). Since the introduction of machine learning algorithms in the GIS environment, their application in the prediction of soil properties became increasingly popular in soil-related studies (Keskin et al. 2019). These methods present a potential improvement of the conventional approach using interpolation methods regarding processing time-efficiency and prediction accuracy. The machine learning algorithms were successfully applied for the analysis of the spatial distribution of soil heavy metals in multiple studies (Khanal et al. 2018, Sergeev et al. 2019).

The primary aim of this study was to evaluate heavy metal soil contents in continental Croatia using the conventional approach based on the geostatistical and deterministic interpolation, as well as the machine learning approach using multiple independent predictors. The created soil contamination maps would represent one of the few studies of heavy metal spatial distribution research in Croatia. The spatial distribution of these heavy metals according to land cover classes was the secondary aim of the research in order to evaluate the most contaminated sites in the study area and to identify potential contamination sources.

2. Materials and Methods

The study area covers the continental biogeoregion of Croatia, a 30 864 km^2 area classified according to the European Environment Agency data on a European Union level (URL 1) (Fig. 1). A total of 469 soil samples collected during 2016 in the study area were used from the Ministry of environment protection and energy, Department of environment and nature protection web feature service (WFS). The soil content data of six heavy metals considered as contaminants in Croatia were extracted for further analysis, according to Ordinance on the protection of agricultural land from contamination (URL 2). These data consist of total soil contents of cadmium (Cd), chromium (Cr), copper (Cu), nickel (Ni), lead (Pb) and zinc (Zn), expressed in mg kg^{-1}. Training and test datasets were created using the Simple Random Sampling method of input soil sample data in a ratio of 80% for training (375 samples) and 20% for test dataset (94 samples), according to Tayebi et al. (2017). Land cover classes were downloaded from CORINE Land Cover (CLC) 2018 data, created by the classification of satellite multispectral Sentinel-2 and Landsat 8 images within the European Space Agency (ESA) Copernicus program. Land cover is represented using the five generalized land cover classes: artificial surfaces, agricultural areas, forests with semi-natural areas, wetlands and water bodies (URL 3). Agricultural areas were the dominant land cover class, covering 54.0% of the study area, followed by forests (40.7%) and artificial surfaces (3.5%). Water bodies were excluded from further analysis, as sampling for heavy metals was performed exclusively for soil. Terrain elevation data was collected from the ESA Copernicus program digital elevation model EU-DEM v1.1, which is natively available at 25 m spatial resolution on the EU level (URL 4).

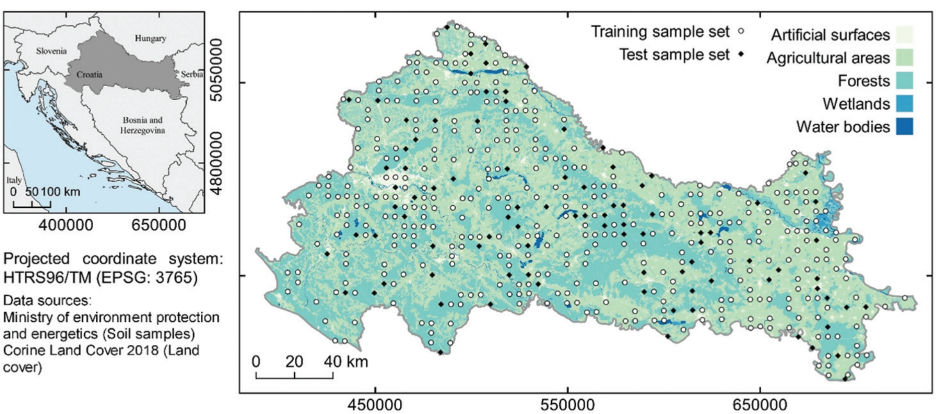

Fig. 1. *Continental biogeoregion of Croatia.*

The spatial prediction of heavy metal soil content in the continental biogeoregion of Croatia was performed using two different approaches. The selected projected coordinate system for the analysis in the GIS environment was the Croatian terrestrial reference system (HTRS96/TM). The spatial resolution of 250 m for the interpolated rasters was selected based on the modified Inspection density method using the finest legible resolution, which integrated the surface of the study area and soil sampling count (Hengl 2006). Other input rasters were resampled and processed at 250 m spatial resolution for the standardization with the interpolated rasters. Open-source GIS software, SAGA GIS v7.4.0 and QGIS v3.8.3, were used for the research.

The first approach was based on using machine learning algorithms for classification based on multiple predictors which affect the accumulation and mobility of heavy metals in soil. Spatial predictors for the classification of heavy metal soil content were selected based on the data availability and their impact on the availability of soil metals in recent soil-related studies (Table 1). All soil sample values which served as predictors, natively imported as point vector data, were interpolated to raster data type in the pre-processing. The optimal interpolation method for predictor values was determined according to data normality and stationarity from the descriptive statistics. Machine learning algorithms implemented in the first prediction approach were Random Forest (RF) and Support Vector Machine (SVM), both allowing high accuracy and resistance to data overfitting in similar studies (Khanal et al. 2018, Keskin et al. 2019). RF performs classification based on the aggregation of multiple randomized decision trees and is thoroughly described in (Belgiu and Drăguț 2016). SVM represents a robust machine learning classification method of the non-linear statistical approach, that performs well in case of limited input training data (Elbisy 2015). The selection of RF and SVM parameters was performed in the iterative procedure using the training data and selecting the parameters that acquired the highest interpolation accuracy based on the test dataset. Both training and test datasets were buffered to fixed 250 m distance from the original points for the conversion to training and test polygons, as a requirement of RF and SVM. Optimal parameters for the RF classification were set to maximum tree depth of 30, maximum sample count of 5 and a maximum number of categories of 10. SVM was performed using the c-support vector classification method, with polynomial kernel type and gamma value of 1. A detailed description of the parameters used in RF and SVM classification is available in a study by Kranjčić and Medak (2020).

Table 1. *Predictors used for the heavy metal soil content calculation.*

Independent predictor variable	Effects on heavy metal accumulation	References
Soil clay content (Clay)	Affects solubility and availability to vegetation, finer soil texture (higher clay content) indicates high immobilization and lower drainage of heavy metals to groundwater	Laghlimi et al. 2015
Soil silt content (Silt)		
Soil sand content (Sand)		
Soil pH (pH)	Affects soil heavy metal solubility and availability to vegetation, lower pH indicates higher solubility	Rees et al. 2014
Soil carbon-to-nitrogen ratio (C/N)	Represents soil organic matter, which affects the binding of heavy metals with organic compounds	Mohamed et al. 2010
Soil calcium carbonate content ($CaCO_3$)	Affects accumulation and mobility of heavy metals in the soil	Sungur et al. 2014
Digital elevation model (DEM)	Affects micro variations of topography and precipitation flow	Shi et al. 2018
Terrain slope (Slope)	Represents terrain waterlogging, which affects fractionation of heavy metals in soil	Zheng et al. 2012

The second approach for the prediction of heavy metal soil contents was based on conventional spatial interpolation methods. Ordinary kriging (OK) is the most commonly used geostatistical method in similar soil-related studies, being considered as the best unbiased spatial predictor (Li and Heap 2008, Negreiros et al. 2010). OK was founded on a presumption of spatial dependence between the variance of soil sample values with the distance between them. The empirical variogram represents the relationship of these properties, while the variance of soil sample values is being aggregated in individual lags, which represent specific distance intervals in the search range of OK (Robinson and Metternicht 2006). The interpolation of the input soil variables at the unknown locations was performed using the mathematical models that were fitted to an empirical variogram, used for the approximation of the variance of values according to the distance to the soil samples. OK was performed according to 20 nearest neighboring points in one sector and 12 lags, each covering 5 800 m distance. Tested mathematical models in this study were linear, square root, Gaussian and spherical, whose formulas were explained in a study by Ver Hoef (2018). The selection of the best-fitting mathematical model to the empirical variogram was performed according to the highest fitting coefficient of determination values of repeatedly evaluated mathematical models per soil parameter.

The limitations of kriging are the interpolation inefficiency in cases of non-normal and non-stationary input sample data (Hengl et al. 2004). The low sample count also disables the fitting of some or all mathematic functions to the empirical variogram. Deterministic methods overcome these limitations by the implementation of the uniform functions according to the distance of the unknown predicted location to the soil samples. Inverse distance weighting (IDW) is one of the most commonly used deterministic interpolation methods, which calculates value weights according to the distance between soil samples and predicted unknown locations (Robinson and Metternicht 2006, Li and Heap 2008). IDW also outperformed OK in the case of low sample count (Radočaj et al. 2020a) and with the presence of extremely high values of soil heavy metal content (Qiao et al. 2018). IDW was performed with the power parameter of 3 with the Inverse distance to a power method, as the primary interpolation parameters for IDW. This combination of parameters reduced the local variability in case of low soil sample data stationarity. IDW was performed using the 20 nearest neighboring points in one sector, with the maximum search range of 30 000 m.

The threshold values for the evaluation of data normality and stationarity were based on the descriptive statistics using the two conditions proposed by Radočaj et al. (2020b). The first condition was a 0.500 threshold value for the coefficient of variation (CV), where the higher CV values in combination with the second condition of skewness and kurtosis deviations higher than 0.500 from the baseline values of 0.000 and 3.000 indicate the absence of data normality and stationarity.

Accuracy assessment of prediction results by both approaches, as well as for the pre-processing of predictors that were natively available as point vector data, was performed using the coefficient of determination (R^2) and root mean square error (*RMSE*). These values provide complementary information in accuracy assessment for the prediction of soil values and were calculated according to the equations (1) and (2) (An et al. 2016):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}, \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \tag{2}$$

in which y_i is the soil sample value, $\hat{y}_i$ s the predicted value, $\bar{y}_i$ is the average of the soil sample value y, n is the number of values. The higher R^2 and lower *RMSE* values indicate higher prediction accuracy of the evaluated method.

The analysis of heavy metal soil content distribution per land cover class was performed to identify the areas of greatest soil contamination and to estimate potential primary contamination sources. CLC 2018 subclasses were reclassified and dissolved to four generalized land cover classes, which correspond to artificial surfaces, agricultural areas, forests and wetlands. Mean and RMSE values were calculated for six evaluated heavy metals of particular prediction model per land cover class for the estimation of the contamination levels.

3. Results and Discussion

The descriptive statistics of point vector predictor data during the pre-processing indicated a moderate level of data normality and stationarity, as in all cases only one condition of CV values higher than 0.500 and skewness and kurtosis 0.500 deviation from ideal values applied (Table 2). Finer soil texture fractions were observed in the study area, which is in accordance with United States Department of Agriculture (USDA) specifications of soil texture classes in the continental Croatia (Radočaj et al. 2020b). These values indicated the selection of maximum permitted values of heavy metal soil content for silty-loamy soils from the Ordinance on the protection of agricultural land from contamination, which were used in the further analysis (URL 2). According to this source, maximum permitted heavy metal soil contents for Cd, Cr, Cu, Ni, Pb and Zn were 1, 80, 90, 50, 100 and 150 mg kg^{-1}, respectively. Sand soil content, soil C/N and $CaCO_3$ achieved the lowest overall normality and stationarity scores of applied predictors, as CV and skewness values deviated from ideal values higher than the selected thresholds.

Table 2. *Descriptive statistics of soil samples for predictor values.*

	Clay (%)	Silt (%)	Sand (%)	pH	C/N	$CaCO_3$ (mg kg^{-1})
mean	30.96	56.67	12.38	6.22	9.47	3.71
minimum	7.56	9.83	0.47	4.19	1.92	0.00
maximum	73.67	87.89	75.75	8.23	26.60	37.40
CV	0.347	0.226	1.168	0.192	1.289	1.225
skewness	0.837	–0.776	1.869	0.227	1.594	1.788
kurtosis	0.955	0.539	3.358	–1.357	2.573	3.305

These observations supported the application of OK interpolation for predictor data, whose parameters and interpolation accuracy are displayed in Table 3. The moderate spatial dependence was achieved for all predictors, while square root model allowed the highest spatial dependence, which did not reflect on the high interpolation accuracy in both cases it was used. Soil pH and soil texture fractions resulted in the highest interpolation accuracies, partially as they are slightly susceptible to long-term variations in the field, compared to soil C/N, which can have high variability annually (Giles et al. 2012). The lowest interpolation accuracy based on the R^2 values was observed for sand soil content, soil C/N and $CaCO_3$, which also indicated the lowest data normality and stationarity during the analysis of descriptive statistics. *RMSE* values showed the similar trend, indicating that silt, sand and $CaCO_3$ produced the lowest interpolation accuracy in that regard.

Table 3. *OK interpolation parameters and accuracy assessment for predictor values.*

Soil property	model	nugget	sill	nugget/sill	range (m)	R^2	*RMSE*
Clay	Gaussian	0.752	1.134	0.663	39141.5	0.649	3.711
Silt	Gaussian	1.132	1.766	0.641	25682.0	0.531	4.584
Sand	spherical	1.301	1.852	0.703	35046.5	0.496	5.025
pH	square root	0.679	1.817	0.374	36912.0	0.716	0.407
C/N	square root	1.127	1.956	0.576	29529.5	0.394	1.724
$CaCO_3$	linear	0.608	0.912	0.667	33205.5	0.389	3.856

All soil sample values of heavy metal soil content resulted in very high kurtosis values, indicating a highly skewed data distribution due to several extremely large values (Table 4). Cu, Zn and Pb achieved the highest kurtosis values, having maximum values 25.1, 12.1 and 12.3 times larger than the mean values, respectively. As the prerequisites of input data normality and stationarity for OK interpolation were not present, it was excluded from the further prediction of heavy metal soil contents distribution in the study. A total of three prediction models were used, containing RF and SVM in the first approach, as well as IDW from the second prediction approach. Finer soil textures resulted in the entire study area with the exception of the northern part, while the highest values of soil C/N and $CaCO_3$ were observed at the most western part of the study area (Fig. 2). The highest soil pH values resulted in the dominant wetland area in the Nature part Kopački rit in the northeast part of the study area, while high values were obtained for the major part in the proximity of rivers Dunav, Sava and Drava. DEM and Slope showed minor variabilities, as lowland soil covers most of the study area.

Table 4. *Descriptive statistics of soil samples for heavy metals.*

	Cd (mg kg^{-1})	Cr (mg kg^{-1})	Cu (mg kg^{-1})	Ni (mg kg^{-1})	Pb (mg kg^{-1})	Zn (mg kg^{-1})
mean	0.32	80.37	28.70	39.71	32.80	94.78
minimum	0.00	0.00	0.00	0.00	0.00	0.00
maximum	4.80	237.50	720.60	195.25	404.65	1147.50
CV	1.279	0.346	1.361	0.555	1.045	0.931
skewness	6.499	1.055	13.164	2.955	7.893	8.614
kurtosis	54.684	6.462	217.113	15.452	69.357	86.692

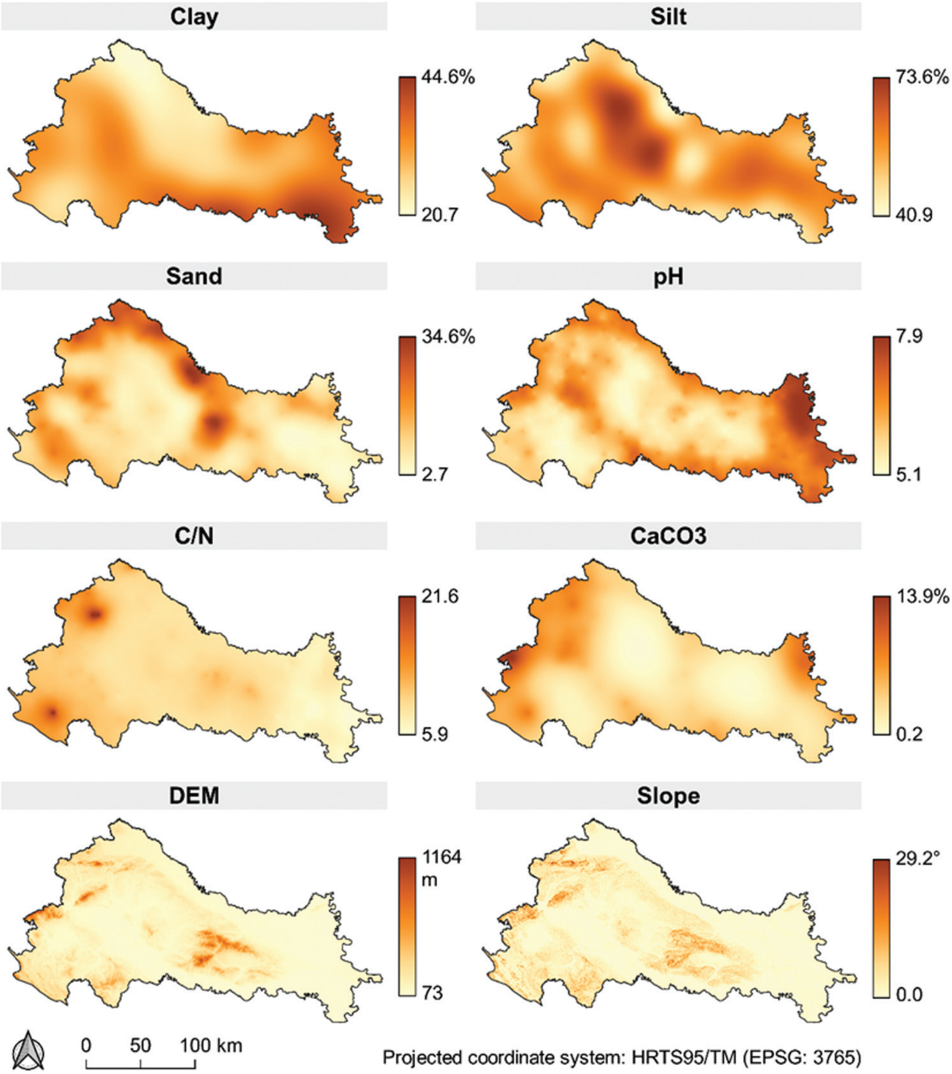

Fig. 2. *Predictor input rasters in RF and SVM methods after pre-processing.*

IDW resulted in the highest interpolation accuracy, having the highest R^2 values for all evaluated heavy metals except for Ni and the lowest *RMSE* values (Table 5). RF achieved similar accuracy trend as IDW for input data with a lower accuracy. SVM was ranked as the third-best method in all cases. High prediction accuracy of RF and SVM for heavy metal soil contents in similar studies (Keskin et al. 2019) implies that the lack of independent predictors, as well as low sample count with low normality and stationarity, potentially caused lower accuracy for RF and SVM. The high amount of extreme values specific for soil heavy metal contamination due to the proximity to contamination sources implies a necessity for denser soil sampling. The sampling density of data used in this study of one sample per 65.8 km^2 allows only for the prediction on a macro level, comparatively to Ballabio et al. (2016), with the sampling density of one sample per 199 km^2. Based on these observations, the IDW method offers a stable prediction choice for the heavy metal soil content data, performing similarly regardless of input data properties and sampling density (Kravchenko 2003). The integration with Sentinel-2 or Landsat 8 multispectral satellite images series used as predictors could significantly increase the accuracy of RF and SVM prediction, based on Castaldi et al. (2019).

Table 5. *Accuracy assessment for the prediction of the spatial distribution of soil heavy metals.*

Heavy metal	RF		SVM		IDW	
	R^2	*RMSE*	R^2	*RMSE*	R^2	*RMSE*
Cd	0.505	0.043	0.438	0.042	0.533	0.041
Cr	0.518	6.229	0.410	4.214	0.630	3.225
Cu	0.524	8.564	0.465	8.930	0.404	8.312
Ni	0.729	7.405	0.429	9.052	0.726	4.372
Pb	0.461	4.920	0.449	4.994	0.534	4.002
Zn	0.404	7.224	0.470	8.089	0.561	6.481

The predicted heavy metal soil contents for three applied prediction methods are displayed in Fig. 3. IDW retained local heterogeneity for most of the applied methods, also having the highest predicted value ranges for all heavy metals. Both RF and SVM resulted in the smooth, continuous values with high homogeneity. This is particularly valid for SVM, which retained the spatial variability of DEM and Slope predictors. RF mostly enabled the detection of similar contaminated areas as IDW, particularly for Cd and Zn. The presence of soil Cr and Zn was the highest for all prediction models, but for the most of the study area the results were below the maximum permitted heavy metal soil contents. Cr soil content predicted by IDW showed potentially dangerous levels at the southern and western parts of the study area, while Ni and Zn were predicted as close to the maximum permitted values in the proximity of Zagreb and southern part of the study area. High Cr values were continously distributed in the study area, which indicates low anthropogenic contamination sources and more likely geogenic origin of these values (Šorša et al. 2018). Contrary to soil Cr values, the distribution of Cd, Cu and Pb contamination sites indicate the presence of discrete point sources, which could originate from industrial sources in the proximity of larger settlements. Elevated Ni values were observed alongside the flow of Sava river, whose sources most probably are industrial sites in Slovenia, spreading downstream through the study area (Milačič et al. 2010).

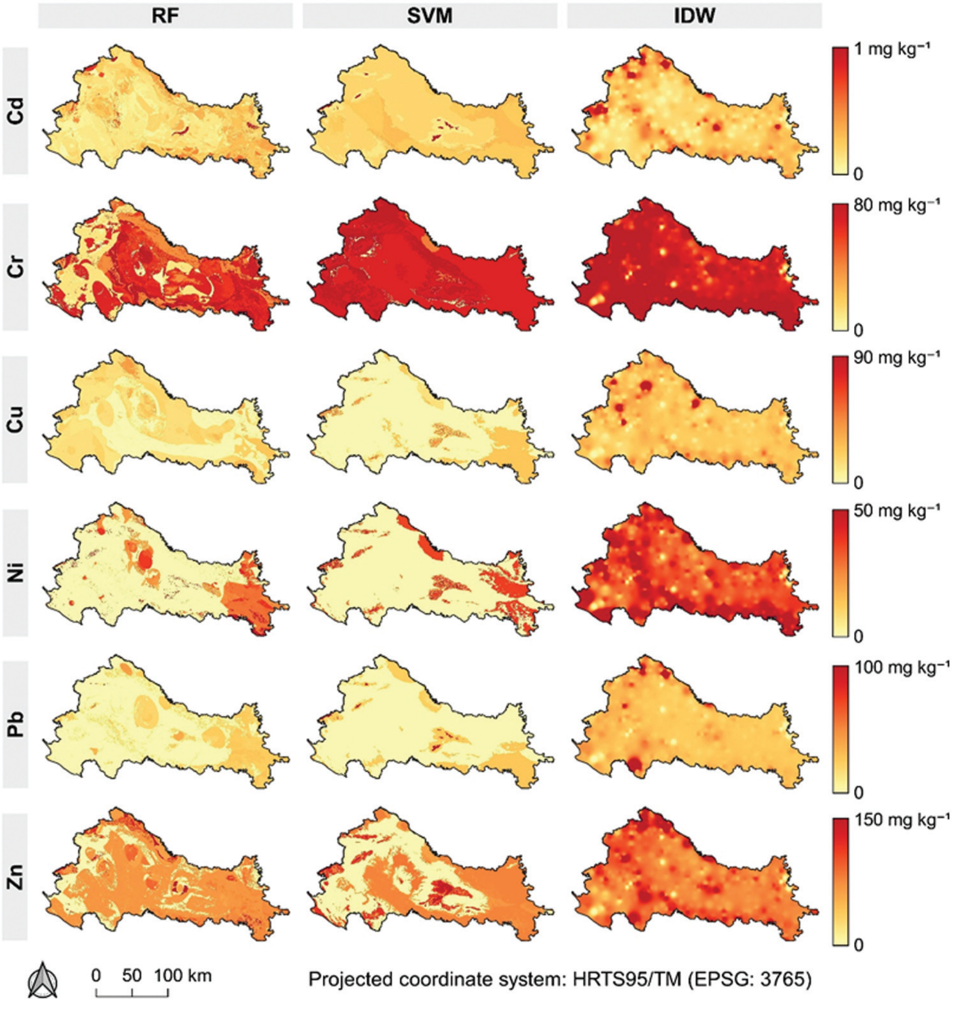

Fig. 3. *Predicted heavy metal soil contents.*

The highest soil contamination by heavy metals was observed at artificial surfaces, dominantly representing urban areas, followed by similar moderate distribution on agricultural areas and forests (Fig. 4). Heavy metal soil contents on agricultural areas resulted in a high correlation with forests, which implies that common contaminants in agriculture, such as fertilizers and pesticides, were not the primary sources of contamination in the study area, which requires further research (Atafar et al. 2010). A high local heterogeneity of IDW prediction can be observed through the highest mean values for all six heavy metals in all observed land cover classes. The predicted value ranges were higher for RF and SVM, which resulted in lower mean values, but higher SD in some instances, especially for Cu and Ni prediction in all land cover classes.

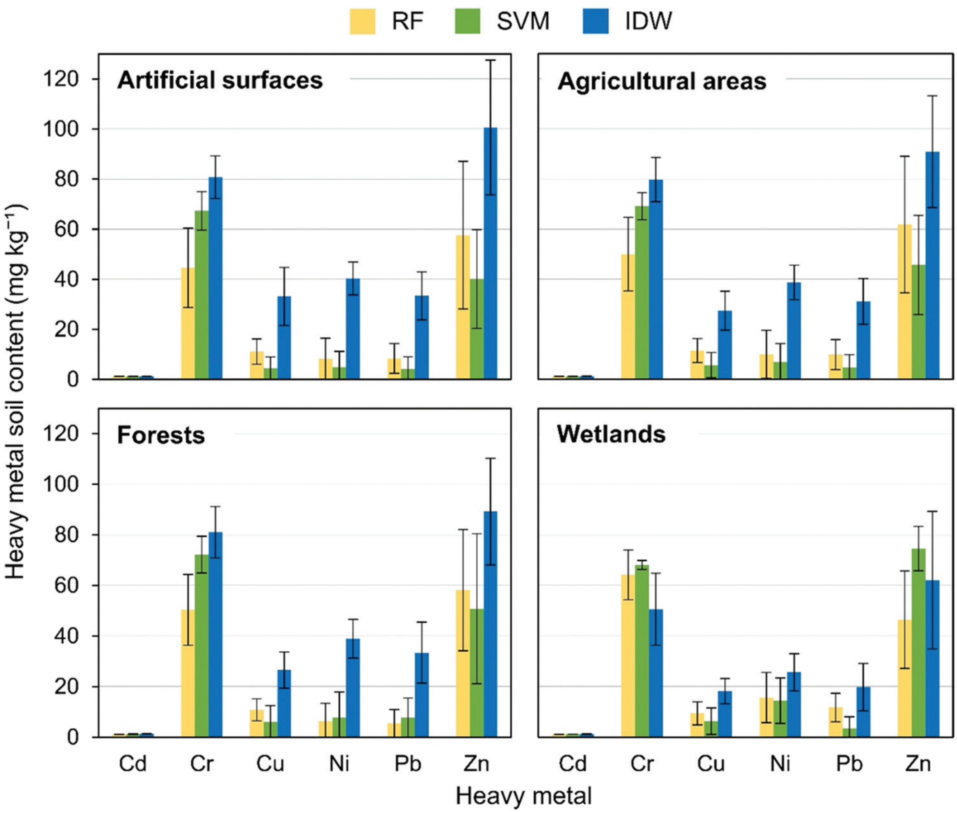

Fig. 4. *Heavy metal soil content distribution according to CLC 2018 land cover classes.*

4. Conclusions

The application of both conventional interpolation and machine learning methods for the precision of six heavy metal soil contents allowed efficient mapping on a macro scale in continental Croatia. The availability of soil sample data provided by WFS from the Ministry of environment protection and energy was successfully applied for the evaluation of prediction methods and mapping on a macro scale with moderate prediction accuracy. Open-source GIS software is as well available for widespread use of the interpolation and machine learning methods in the research of heavy metal soil contents, allowing high processing possibilities with no cost for the user. For the analyses of heavy metal soil content spatial distribution on a micro scale at the city or municipality level, higher soil sampling density is required for the prediction with the same accuracy. IDW method performed the best in the specific conditions of low sampling density and moderately low input data normality and stationarity. Based on the existing literature, with the availability of more abundant input soil sample data and more independent predictors, machine learning methods are expected to produce significantly higher prediction accuracy. The highest presence of soil heavy metals was detected at artificial surfaces land cover class, which primarily consists of the urban area. As urbanization is a growing issue both globally and in Croatia, more population is exposed to the negative health impacts of heavy metals and should be managed to restrict their further input in the environment. Based on the relative comparison of heavy metal soil content in agricultural areas and forests, no major deviation was observed for all six analyzed heavy metals, which implies that fertilizer and pesticides were not a primary contamination source in agricultural areas. Dangerously high levels of soil Cr, Ni and Zn showed that soil contamination of heavy metals is a present concern in Croatia and should be a subject of future research and the adjustment of land-use management.

ACKNOWLEDGEMENTS. This work was supported by the Faculty of Agrobiotechnical Sciences Osijek as a part of the scientific project "AgroGIT – technical and technological crop production systems, GIS and environment protection".

References

An, D., Zhao, G., Chang, C., Wang, Z., Li, P., Zhang, T., Jia, J. (2016): Hyperspectral field estimation and remote-sensing inversion of salt content in coastal saline soils of the Yellow River Delta, International Journal of Remote Sensing, 37(2), 455–470.

Atafar, Z., Mesdaghinia, A., Nouri, J., Homaee, M., Yunesian, M., Ahmadimoghaddam, M., Mahvi, A. H. (2010): Effect of fertilizer application on soil heavy metal concentration, Environmental monitoring and assessment, 160(1–4), 83.

Ballabio, C., Panagos, P., Monatanarella, L. (2016): Mapping topsoil physical properties at European scale using the LUCAS database, Geoderma, 261, 110–123.

Belgiu, M., Drăguț, L. (2016): Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B. (2019): Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands, ISPRS Journal of Photogrammetry and Remote Sensing, 147, 267–282.

Elbisy, M. S. (2015): Support vector machine and regression analysis to predict the field hydraulic conductivity of sandy soil, KSCE Journal of Civil Engineering, 19(7), 2307–2316.

Giles, M. E., Morley, N. J., Baggs, E. M., Daniell, T. J. (2012): Soil nitrate reducing processes-drivers, mechanisms for spatial variation, and significance for nitrous oxide production, Frontiers in microbiology, 3, 407.

Hengl, T. (2006): Finding the right pixel size, Computers & geosciences, 32(9), 1283–1298.

Hengl, T., Heuvelink, G. B., Stein, A. (2004): A generic framework for spatial prediction of soil variables based on regression-kriging, Geoderma, 120(1–2), 75–93.

Hu, Y., Liu, X., Bai, J., Shih, K., Zeng, E. Y., Cheng, H. (2013): Assessing heavy metal pollution in the surface soils of a region that had undergone three decades of intense industrialization and urbanization, Environmental Science and Pollution Research, 20(9), 6150–6159.

Hu, B., Xue, J., Zhou, Y., Shao, S., Fu, Z., Li, Y., Chen, S., Qi, L., Shi, Z. (2020): Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning, Environmental Pollution, 114308.

Keskin, H., Grunwald, S., Harris, W. G. (2019): Digital mapping of soil carbon fractions with machine learning, Geoderma, 339, 40–58.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S. (2018): Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield, Computers and electronics in agriculture, 153, 213–225.

Kranjčić, N., Medak, D. (2020): Evaluating Different Machine Learning Methods on RapidEye and PlanetScope Satellite Imagery, Geodetski list, 74(1), 1–18.

Kravchenko, A. N. (2003): Influence of spatial structure on accuracy of interpolation methods, Soil Science Society of America Journal, 67(5), 1564–1571.

Laghlimi, M., Baghdad, B., El Hadi, H., Bouabdli, A. (2015): Phytoremediation mechanisms of heavy metal contaminated soils: a review, Open journal of Ecology, 5(8), 375.

Li, J., Heap, A. D. (2008): A review of spatial interpolation methods for environmental scientists, Geoscience Australia, Canberra, Australia.

Milačič, R., Ščančar, J., Murko, S., Kocman, D., Horvat, M. (2010): A complex investigation of the extent of pollution in sediments of the Sava River. Part 1: selected elements, Environmental monitoring and assessment, 163(1–4), 263–275.

Mohamed, I., Ahamadou, B., Li, M., Gong, C., Cai, P., Liang, W., Huang, Q. (2010): Fractionation of copper and cadmium and their binding with soil organic matter in a contaminated soil amended with organic materials, Journal of Soils and Sediments, 10(6), 973–982.

Negreiros, J., Painho, M., Aguilar, F., Aguilar, M. (2010): Geographical information systems principles of ordinary kriging interpolator, Journal of Applied Sciences, 10(11), 852–867.

Ottesen, R. T., Alexander, J., Langedal, M., Haugland, T., Høygaard, E. (2008): Soil pollution in day-care centers and playgrounds in Norway: national action plan for mapping and remediation, Environmental Geochemistry and Health, 30(6), 623–637.

Qiao, P., Lei, M., Yang, S., Yang, J., Guo, G., Zhou, X. (2018): Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing, Environmental Science and Pollution Research, 25(16), 15597–15608.

Radočaj, D., Jurišić, M., Gašparović, M., Plaščak, I. (2020a): Optimal Soybean (Glycine max L.) Land Suitability Using GIS-Based Multicriteria Analysis and Sentinel-2 Multitemporal Images, Remote Sensing, 12(9), 1463.

Radočaj, D., Jurišić, M., Zebec, V., Plaščak, I. (2020b): Delineation of Soil Texture Suitability Zones for Soybean Cultivation: A Case Study in Continental Croatia, Agronomy, 10(6), 823.

Rees, F., Simonnot, M. O., Morel, J. L. (2014): Short-term effects of biochar on soil heavy metal mobility are controlled by intra-particle diffusion and soil pH increase, European Journal of Soil Science, 65(1), 149–161.

Robinson, T. P., Metternicht, G. (2006): Testing the performance of spatial interpolation techniques for mapping soil properties, Computers and electronics in agriculture, 50(2), 97–108.

Sergeev, A. P., Buevich, A. G., Baglaeva, E. M., Shichkin, A. V. (2019): Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals, Catena, 174, 425–435.

Shi, T., Guo, L., Chen, Y., Wang, W., Shi, Z., Li, Q., Wu, G. (2018): Proximal and remote sensing techniques for mapping of soil contamination with heavy metals, Applied Spectroscopy Reviews, 53(10), 783–805.

Sollitto, D., Romic, M., Castrignanò, A., Romic, D., Bakic, H. (2010): Assessing heavy metal contamination in soils of the Zagreb region (Northwest Croatia) using multivariate geostatistics, Catena, 80(3), 182–194.

Sungur, A., Soylak, M., Ozcan, H. (2014): Investigation of heavy metal mobility and availability by the BCR sequential extraction procedure: relationship between soil properties and heavy metals availability, Chemical Speciation Bioavailability, 26(4), 219–230.

Šiljeg, A., Barada, M., Marić, I., Roland, V. (2019): The effect of user-defined parameters on DTM accuracy – development of a hybrid model, Applied Geomatics, 11(1), 81–96.

Šorša, A., Peh, Z., Halamić, J. (2018): Geochemical mapping the urban and industrial legacy of Sisak, Croatia, using discriminant function analysis of topsoil chemical data, Journal of Geochemical Exploration, 187, 155–167.

Tayebi, M., Naderi, M., Mohammadi, J., Tayebi, M. H. (2017): Comparing different statistical models for assessing Fe-contaminated soils based on VNIR/SWIR spectral data, Environmental Earth Sciences, 76(21), 734.

Ver Hoef, J. M. (2018): Kriging models for linear networks and non-Euclidean distances: Cautions and solutions, Methods in Ecology and Evolution, 9(6), 1600–1613.

Yan, W., Mahmood, Q., Peng, D., Fu, W., Chen, T., Wang, Y., Li, S., Chen, J. Liu, D. (2015): The spatial distribution pattern of heavy metals and risk assessment of moso bamboo forest soil around lead-zinc mine in Southeastern China, Soil and Tillage Research, 153, 120–130.

Zhao, H., Xia, B., Fan, C., Zhao, P., Shen, S. (2012): Human health risk from soil heavy metal contamination under different land uses near Dabaoshan Mine, Southern China, Science of the Total Environment, 417, 45–54.

URLs

URL 1: European Environment Agency – Biogeographical Regions, https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3, (13.10.2020).

URL 2: Ordinance on the protection of agricultural land from contamination (N.N. 32/10),
http://www.propisi.hr/print.php?id=6842, (13.10.2020).

URL 3: Corine Land Cover 2018 Technical Guidelines,
https://land.copernicus.eu/user-corner/technical-library/clc2018technical-guidelines_final.pdf, (13.10.2020).

URL 4: European Space Agency Copernicus program – EU-DEM v1.1,
https://land.copernicus.eu/user-corner/technical-library/eu-dem-v1-1-user-guide/at_download/file, (13.10.2020).

Prostorna predikcija udjela teških metala u tlima kontinentalne Hrvatske usporedbom metoda strojnog učenja i prostorne interpolacije

SAŽETAK. Onečišćenje tla uzrokovano teškim metalima uzrokuje potencijalno dugoročnu opasnost za zdravlje ljudi i biološku raznolikost zbog učinka bioakumulacije. Prethodna istraživanja na mikro razini u Hrvatskoj otkrila su onečišćenje tla teškim metalima iznad maksimalno dopuštenih vrijednosti, što je ujedno impliciralo potrebu poznavanja njihove trenutne prostorne zastupljenosti na makro razini u Hrvatskoj. Cilj ovog istraživanja bio je provesti prostorno predviđanje šest teških metala u tlu koji se smatraju onečišćujućima u kontinentalnoj Hrvatskoj koristeći dva pristupa: konvencionalni pristup zasnovan na interpolaciji i pristup strojnog učenja. Predviđanje je provedeno na najnovijim dostupnim uzorcima tla kadmija (Cd), kroma (Cr), bakra (Cu), nikla (Ni), olova (Pb) i cinka (Zn), prikupljenim od strane Ministarstva zaštite okoliša i energetike. Konvencionalni pristup predviđanja sastojao se od interpolacije korištenjem uobičajenog kriginga (OK) u slučaju normalnosti i stacionarnosti ulaznih podataka, zajedno s metodom inverzne udaljenosti (IDW). Za pristup strojnog učenja korištene su metoda slučajnih šuma (RF) i metoda vektora podrške (SVM). IDW je nadmašio rezultate predviđanja RF i SVM za sve sadržaje teških metala u tlu, prvenstveno zbog nedovoljno gustog uzorkovanja tla. Sadržaj Cr u tlu predviđen je iznad najveće dopuštene granice, dok su za Ni i Zn utvrđene opasne razine onečišćenja tla na dijelovima istraživanog područja. Najveće razine onečišćenja tla zabilježene su u urbanim područjima generaliziranih klasa zemljišnog pokrova, što ukazuje na potrebu za njegovim praćenjem i prilagođavanjem planova upravljanja korištenjem zemljišta.

Ključne riječi: *onečišćenje tla, metoda slučajnih šuma, metoda vektora podrške, obični kriging, metoda inverzne udaljenosti, pokrov zemljišta.*

Received / Primljeno: 2020-11-25

Accepted / Prihvaćeno: 2020-12-11