Spatial Distribution of Lead (Pb) in Soil: A Case Study in a Contaminated Area of the Czech Republic

Nicolas Francos^{1*}, Asa Gholizadeh² and Eyal Ben-Dor¹

¹ The Remote Sensing Laboratory, Tel Aviv University, Zelig 10, Tel Aviv 69978, Israel

² Department of Soil Science and Soil Protection, Faculty of Agrobiological Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Suchbátka, 16500, Prague, Czech Republic

*Contact e-mail: nicolasf@mail.tau.ac.il

Abstract

For decades, the Příbram district in the Czech Republic has been affected by industrial and mining activities. These activities are important sources of heavy metal pollutants that are detrimental to soil quality. A recent study examined visible–near-infrared (VNIR), shortwave-infrared (SWIR) and X-ray fluorescence (XRF) spectroscopy to model soil lead (Pb) content in a selected area located in the Příbram district. Following that study, and using the same chemical and geographical data, we examined the spatial distribution of Pb content in the soil, with a combination of different traditional spatial analyses (Moran's I, hotspot analysis, and Kriging) that were significantly validated. One of the novel points of this work is the use of the Getis–Ord hotspot analysis before the execution of a Kriging interpolation model to better emphasize clustering patterns. The results indicated that Pb was a spatially dependent soil property and through extensive in situ sampling, it was possible to generate a very accurate Kriging interpolation model. The high-Pb hotspots coincided with topographic obstacles that were modeled using topographic profiles extracted from the open-source Google Earth platform, indicating that Pb content does not always exhibit a direct relationship with topographic height as a result of runoff, due to the contribution of topographic steps. This observation provides a new perspective on the relationship between Pb content and topographic patterns.

Keywords: lead, Kriging, hotspot analysis, interpolation, spatial distribution.

Highlights:

- Different spatial analyses were executed to examine the spatial distribution of Pb
- Extensive in-situ sampling provided an accurate Kriging interpolation model of Pb
- Pb hotspots with field soil data were mapped with a high degree of certainty
- The Pb hotspots coincided with topographic obstacles modeled using Google Earth

1. Introduction

Soil contaminants, including potentially toxic elements (PTEs), have been increasing as a result of anthropogenic influences such as urbanization, industrialization, as well as population and agricultural growth (Kicińska, 2019; Kicińska and Wikar, 2021). Over the last few decades, soils in several regions of the Czech Republic have also been heavily affected by industrial activities.

In a recent study by Gholizadeh et al. (2021), visible–near-infrared (VNIR) shortwave infrared (SWIR) and X-ray fluorescence spectroscopic techniques were used to assess several PTEs in the Příbram district, which is one of the most heavily polluted areas in the Czech Republic, and Pb was the most accurately predicted element using both individual and fused spectral data; however, the spatial distribution of this element was not examined. Highlighting the spatial distribution of Pb hotspots in the soil (i.e., very high and very low levels) could facilitate the development of the initial steps to efficiently and accurately manage and rehabilitate the area. Therefore, in this study, we aimed at detecting Pb hotspots using surface soil samples in order to map Pb content with a high degree of certainty. In this context, three different spatial analysis techniques were tested to determine whether Pb was a spatially dependent contaminant. The study results contribute by providing additional information to the preliminary findings of Gholizadeh et al. (2021). In addition, spatial evaluation of PTEs, including Pb, provides a way to delineate the distribution of PTEs, their content, occurrence, spatial dependency, and to better understand their source.

Similar to the study by Gholizadeh et al. (2021), in this study, several samples were collected; we also extracted the optimal distance for collecting Pb samples in our study area, because beyond a given distance, the autocorrelation between samples and their neighbor samples decreases. By identifying this distance using the variogram extracted from the Kriging interpolation method (Oliver and Webster, 1990, 2014), we suggest important information for efficient collection of Pb samples in similar study areas. Thus, in future studies, it will be possible to obtain efficient interpolation models collecting fewer field samples and in larger areas than in this study and that of Gholizadeh et al. (2021).

Before executing the interpolation model, the Pb values were preprocessed by Getis–Ord hotspot analysis (Getis and Ord, 1992). This approach, i.e., combining the two methods, was adopted from

Ngaruiya and Ngigi (2014), who mapped hotspots of sewer chokes in Nairobi, Kenya. Nevertheless, as the study of Ngaruiya and Ngigi (2014) is more related to urban planning, we decided to check this novel approach to examine if it can be used to track after Pb contamination trajectories. Although Ngaruiya and Ngigi (2014) used the inverse distance weighted (IDW) method, we adopted the Kriging interpolation method because it can be optimized with the variogram. In addition, in the present work, the topography of the study area was also taken into account to determine its possible contribution to the presence of Pb hotspots, pursuant to Chen et al. (2012) who showed that heavy metals can be transported by rainfall and runoff, and that there is a relationship between heavy metals and topographic altitude.

2. Materials and Methods

To achieve the goal of this study, 157 soil samples were collected using both grid and transect sampling designs with manual augers (Gholizadeh et al., 2021) from a selected study area in the district of Příbram (49°71' N and 14°01' E), in the Czech Republic (shown in Figure 1). The Příbram region is famous for its smelting activities and historic Pb-Ag ore mining (Vaněk et al., 2008). The occurrence of flood events between 1932 and 1952 resulted in pollutant deposition in the environment from waste dumps and the rupturing of tailing ponds within the area (Kotková et al., 2019).

Figure 1. Location of the study area and sampling points. Sampling point colors indicate different sampling strategies (executed at different times): yellow – new sampling points; red – previously analyzed, but newly collected and reanalyzed sampling points (Gholizadeh et al., 2021).

The pseudo-total content of Pb was extracted using aqua regia (ISO 11466:1995). The solution was 10-fold diluted with deionized water (conductivity 18.2 MW) and then filtered through a 0.45- μ m nylon membrane disc filter (Thermo Fisher Scientific, Waltham, MA, USA) prior to analysis. The extract was analyzed by ICP-OES iCAP 7000 (Thermo Fisher Scientific, Waltham, MA, USA) for Pb content. The concentration measurement was subjected to quality control (QC) the SRM 2711 (Montana II soil) internal reference material (National Institute of Standards and Technology, Gaithersburg, MD, USA). The acquired value was in good agreement with the reference data; the recovery differences were less than 10% ($n = 3$). As suggested by

Hewavitharana (2011), the samples and standards were matrix-matched to compensate for matrix influences on the analytical response. All of the analyses were performed in triplicate.

It is important to emphasize that both the limit of detection (LOD) and the limit of quantification (LOQ) of the aqua regia method were reported by Shahbazi and Beheshti (2019) to be very low: LOD = 7.11 mg/kg and a LOQ = 21.33 mg/kg. These values appeared to be even lower when we examined the boxplot (Figure 2) and the descriptive statistics (Table 1) of the Pb (mg/kg) values. This is because the distribution of the values in this study is quite large (minimum = 72.68 mg/kg, maximum = 6880.36 mg/kg) and the LOD and LOQ values are much lower than 72.68 mg/kg.

Figure 2. Boxplot of the Pb values (mg/kg).

Table 1. Descriptive statistics of Pb values (mg/kg).

Parameter	Value
Number of samples	157
Mean	1972.22
Median	1681.75
Minimum	72.68
Maximum	6880.36
Standard deviation	1198.53
Skewness	1.55
Kurtosis	3.49

For the purpose of spatial analysis, first, the spatial autocorrelation was examined through Moran's I statistic (Getis and Ord, 1992), where positive values indicate spatial dependency. Hence, it is possible to examine, objectively and rapidly, whether clustering patterns exist before executing further spatial analyses. Nevertheless, the Moran's I statistic is a relative measure and can only be interpreted within the context of its computed z -score or p -value (Mitchell, 2005). Thus, obtaining z -scores over 1.96 in combination with significant p -values, the spatial distribution of the high and low z -score values in the dataset were more spatially clustered than would be expected if the underlying spatial processes were random (Hughes and Haran, 2013; Tiefelsdorf and Griffith, 2007).

Afterward, a Getis–Ord hotspot analysis was executed (Getis and Ord, 1992) on the Pb measurements of the samples. The Getis–Ord hotspot analysis was easy to interpret because looking at each entity within the context of neighboring features provided z -scores and p -values to determine significantly high or low contents. Finally, a simple Kriging interpolation (Oliver and Webster, 1990, 2014) was applied on the z -score values extracted from the Getis–Ord hotspot analysis, because it can be fitted against a variogram, which provides information on the spatial autocorrelation of datasets. With this information, the model can be optimized using the samples with the highest autocorrelation versus each entity (Oliver and Webster, 2014).

None of the Kriging hyperparameters, including the optimal range for the variogram, were modified, following the default suggestion of ArcGIS. The obtained map was validated using a full leave-one-out cross validation (LOOCV) of the generated Kriging model. These analyses were performed using ArcGIS 10.3 software. Finally, to evaluate the contribution of the topographic behavior of the study area to the Pb accumulation, we used the topographic profile tool of Google Earth for a graphic representation.

3. Results and Discussion

3.1. Moran's I Spatial Autocorrelation

Figure 3 illustrates the spatial autocorrelation report extracted from the ArcGIS software. It can be seen that there is a significant spatial autocorrelation for Pb in the study area, given a positive Moran's I statistic (0.43), a very high z -score (7.95), and a very low p -value (< 0.01). These results are quite important because they encourage the evaluation of other spatial analyses and interpolation methods for mapping Pb content and also suggest that the study area is characterized by significant clusters of low and or high Pb content. It should be noted that the Pb values were not preprocessed at this stage.

Figure 3. Moran's I spatial autocorrelation report extracted from ArcGIS software.

3.2. Hotspot Analysis and Kriging Interpolation

As shown in Figure 4a, the Getis–Ord hotspot analysis identified important clusters of low and high contents of Pb (mg/kg) with a high degree of certainty. These results strengthen the conclusions extracted from the Moran’s I autocorrelation report (shown in Figure 3). Figure 4b highlights a map with the extracted z -scores in each field sample after the Getis–Ord hotspot analysis and the execution of a Kriging spatial interpolation executed to the same z -scores. This map demonstrates that on the one hand, there is a low content of Pb in the southern zone of the study area. On the other hand, in the center and in the northern areas, the Pb contents are relatively high.

Figure 4. (a) The p -value plotted against the generated z -scores for each field sample after the Getis–Ord hotspot analysis; (b) Kriging interpolation from the extracted z -scores.

According to Figure 5, the full LOOCV of the Kriging interpolation provided a low root mean square error (RMSE) of 0.44 (z -score), a very high R^2 (0.97), and a slope close to 1 (0.94), therefore, it can be considered a reliable interpolation model for mapping the Pb content in this study area. The full LOOCV analysis also presented a very low mean error (bias) of 0.0006 (z -

score), considering the range of values, which were from 4.29 to 5.92 (z-score). Thus, the previous z-score transformation through the Getis–Ord hotspot analysis contributed to obtaining a satisfactory LOOCV stage. This is because the Getis–Ord hotspot analysis optimizes and emphasizes the clustering patterns.

Figure 5. (a) The variogram; (b) the LOOCV predictions; (c) the measured error after the LOOCV.

The variogram, presented in Figure 5a, demonstrates the spatial relationship as a function of distance. This variogram shows that, as the distance between each entity in relation to the others

increases, the correlation between the measured Pb against the neighboring samples decreases. When the distance exceeds 0.002 degrees, the correlation becomes asymptotic, indicating that there is no more autocorrelation. Therefore, a higher distance range may be problematic for correct interpolation in the study site in question. After 0.002 degrees of distance, the variance begins to stabilize showing constant values close to zero and the sample values are not related to each other. Thus, the Kriging interpolation model was optimized considering this range as the default suggestion of the ArcGIS software.

As several interpolation methods with many hyper-parameters can be configured, the number of interpolation models that can be extracted could be infinite. Although it is not really possible to extract the best model among all of these combinations, we can always validate our outputs. Therefore, the validation stage is the most important task for this evaluation because we cannot expect to obtain the best model, but we can find a very good one that can efficiently represent our study area.

3.3. Topographic Profile

The main issue of our findings is the fact that the locations of the smelter and mining activities are in the south of the study site (Figure 6), although according to our maps, we obtained higher Pb contents in the northern areas than in the southern areas of the study site, which was closer to the possible pollutant sources (the smelter and mining locations). As shown in Figure 6, the locations of the smelter (500 m) and the site of the historical mining activities (539 m) are located at higher altitudes than the study site (441 m). Therefore, runoff (and flooding) events may deposit Pb sediments at lower altitudes.

Figure 6. A Google Earth representation of the study area, the smelter, and the historical mining activity site.

To interpret the above-mentioned issue, using the Google Earth topographic profile tool, we found three topographic steps (Figure 7), where the contaminated runoff probably accumulated. At all these topographic steps, we found the highest Pb contents and a reasonable explanation for this phenomenon could be that the runoff that accumulated at each topographic step left Pb residuals after evaporation of the water. Thus, as shown in this study, it is essential to consider not only the height but also the topographic steps. These obstacles may have a considerable effect on the accumulation of PTE (here Pb) sediments that are carried by the runoff.

Figure 7. The topographic profile and the three topographic steps where we found the highest Pb contents: (a) The first topographic step; (b) the second topographic step; (c) the third topographic step.

Summary and Conclusions

In this study, the Moran's I statistic indicated, with high significance, that Pb in the study area was spatially correlated. Moreover, interpolation methods such as Kriging in combination with the Getis–Ord hotspot analysis could be successfully used to generate maps representing Pb content with a high degree of certainty. This was because the Getis–Ord hotspot analysis identified significantly hot and cold spots of Pb that were later mapped using the Kriging interpolation method. Then the results of the generated interpolation model were further validated with a full LOOCV showing very low errors. The variogram extracted from the Kriging interpolation model suggested that the maximum distance to sample Pb in the study area should not be higher than 0.002 degrees and the model was optimized using this distance range. In future work, this distance might be considered a reference distance for collecting Pb samples for interpolation in similar

study areas. The use of the Getis–Ord hotspot analysis before executing interpolation models should be further investigated as it emphasizes the clustering patterns and contributes to the validation stage of the interpolation models.

The extracted interpolation model also highlighted the possible contribution of the topographic steps to the accumulation of Pb content. Moreover, regarding the relationship between topography and heavy metals due to transportation in the runoff (Chen et al., 2012), this study showed that topographic obstacles must also be considered, because runoff accumulates at topographic steps and leaves residual Pb after the water has evaporated. Hence, the relationship between topographic height and heavy metals is not a direct one. The insights of this study are very important for optimizing long-term contamination monitoring, land rehabilitation, and environmental management practices.

Certainly, we investigated the spatial correlation of the Pb in the soil using the conventional aqua regia method (ISO 11466:1995), but as the study of Gholizadeh et al. (2021) showed high relationship between VNIR–SWIR spectroscopy and the conventional method, it is assumed that VNIR–SWIR spectral analysis can also be used to predict Pb content to then create Pb spatial distribution map with reliable accuracy. We strongly recommend to investigate this direction in future studies.

References

- Chen, Y., Liu, Y., Liu, Y., Lin, A., Kong, X., Liu, D., Li, X., Zhang, Y., Gao, Y., Wang, D., 2012. Mapping of Cu and Pb contaminations in soil using combined geochemistry, topography, and remote sensing: a case study in the Le'an River Floodplain, China. *International Journal of Environmental Research and Public Health* 9(5), 1874–1886. <https://doi.org/10.3390/ijerph9051874>
- Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24, 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>

- Gholizadeh, A., Coblinski, J.A., Saberioon, M., Ben-Dor, E., Drábek, O., Demattê, J.A.M., Borůvka, L., Němeček, K., Chabrilat, S., Dajčl, J., 2021. vis-NIR and XRF data fusion and feature selection to estimate potentially toxic elements in soil. *Sensors* 21, 2386. <https://doi.org/10.3390/s21072386>
- Hewavitharana, A.K., 2011. Matrix matching in liquid chromatography–mass spectrometry with stable isotope labelled internal standards—Is it necessary? *Journal of Chromatography A* 1218(2), 359–361. <https://doi.org/10.1016/j.chroma.2010.11.047>
- Hughes, J., Haran, M., 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society Series B* 75, 139–159. <https://doi.org/10.1111/j.1467-9868.2012.01041.x>
- Kicińska, A., 2019. Environmental risk related to presence and mobility of As, Cd and Tl in soils in the vicinity of a metallurgical plant – Long-term observations. *Chemosphere* 236, 124308. <https://doi.org/10.1016/j.chemosphere.2019.07.039>
- Kicińska, A., Wikar, J., 2021. Ecological risk associated with agricultural production in soils contaminated by the activities of the metal ore mining and processing industry - example from southern Poland. *Soil and Tillage Research* 205, 104817. <https://doi.org/10.1016/j.still.2020.104817>
- Kotková, K., Nováková, T., Tůmová, Š., Kiss, T., Popelka, J., Faměra, M., 2019. Migration of risk elements within the floodplain of the Litavka River, the Czech Republic. *Geomorphology* 329, 46–57. <https://doi.org/10.1016/j.geomorph.2018.12.010>
- Mitchell, A., 2005. *The ESRI Guide to GIS Analysis: Spatial Measurements and Statistics*. ESRI Press.
- Ngaruiya, J., Ngigi, M., 2014. Analysis of sewer chokes using GIS: a case study Nairobi City Western Region. *International Journal of Engineering Science Invention* 3, 57–65.
- Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems* 4, 313–332. <https://doi.org/10.1080/02693799008941549>
- Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. *CATENA* 113, 56–69. <https://doi.org/10.1016/j.catena.2013.09.006>
- Shahbazi, K., Beheshti, M., 2019. Comparison of three methods for measuring heavy metals in calcareous soils of Iran. *SN Applied Sciences* 1(12), 1541. <https://doi.org/10.1007/s42452-019-1578-x>
- Tiefelsdorf, M., Griffith, D.A., 2007. Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A* 39, 1193–1221. <https://doi.org/10.1068/a37378>
- Vaněk, A., Ettler, V., Grygar, T., Borůvka, L., Šebek, O., Drábek, O., 2008. Combined chemical and mineralogical evidence for heavy metal binding in mining-and smelting affected alluvial soils. *Pedosphere* 18(4):464–478. [https://doi.org/10.1016/S1002-0160\(08\)60037-5](https://doi.org/10.1016/S1002-0160(08)60037-5)