Prediction of Nickel Concentrations in Urban and Peri-Urban Soils: Application of a Hybridized Empirical Bayesian Kriging and Support Vector Machine Regression Approach

Prince Chapman Agyeman (agyeman@af.czu.cz)
Czech University of Life Sciences Prague
Ndiye Michael Kebonye
Czech University of Life Sciences Prague
Kingsley JOHN
Czech University of Life Sciences Prague

Research Article

Keywords: Hybridization, empirical bayesian kriging, support vector machine regression, self-organizing map, Nickel

Posted Date: June 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-616432/v1

Prediction of nickel concentrations in urban and peri-urban soils: application of a hybridized empirical bayesian kriging and support vector machine regression approach

Prince Chapman Agyeman[1*], Ndiye Michael Kebonye[1], Kingsley JOHN [1].

[1]Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, 16500 Prague, Czech Republic

[*]Correspondence E-mail: agyeman@af.czu.cz (P.C. Agyeman)

Abstract

Soil pollution is a big issue caused by anthropogenic activities. The spatial distribution of potentially toxic elements (PTEs) varies in most urban and peri-urban areas. As a result, spatially predicting the PTEs content in such soil is difficult. A total number of 115 samples were obtained from Frydek Mistek in the Czech Republic. Calcium(Ca), magnesium(Mg), potassium(K), and nickel (Ni) concentrations were determined using Inductively Coupled Plasma Optical Emission Spectroscopy. The correlation matrix between the response variable and the predictors revealed a satisfactory correlation between the elements. The prediction results indicated that support vector machine regression (SVMR) performed well although its estimated root mean square error (RMSE) (235.974) and mean absolute error (MAE) (166.946) were higher when compared with the other methods applied. Conversely, the hybridized model of empirical bayesian kriging -multiple linear regression (EBK-MLR) performed poorly as indicated by the measured coefficient of determination value below 0.1. The empirical bayesian kriging-support vector machine regression (EBK-SVMR) model was the best model, with low RMSE (95.479) and MAE (77.368) values and a high coefficient of determination (R^2 = 0.637). EBK-SVMR modeling technique was visualized using self-organizing map. The clustered neurons of the hybridized model CakMg -EBK-SVMR component plane showed a diverse color pattern predicting the concentration of Ni in the urban and peri urban soil. The results proved that combining EBK and SVMR is an effective technique for predicting Ni concentrations in urban and peri-urban soil.

Keywords. Hybridization; empirical bayesian kriging; support vector machine regression; self-organizing map; Nickel

Highlights

- Empirical Bayesian kriging was used as a base model for hybridization.
- Support vector machine regression was the best algorithm to hybridized with.
- The EBK-SVMR hybrid model outperforms the other models in predicting Ni.
- The performance of the hybrid model was visualized using a self-organizing map.

INTRODUCTION

Nickel (Ni) is regarded as a micronutrient for plants due to its contribution to atmospheric nitrogen(N) fixation as well as urea metabolism, both of which are needed for the germination of seed.[1] Apart from its contribution to seed sprouting, Ni also acts as an inhibitor for fungi and bacteria and promote substantially plant development. The deficiency of Ni in the soil for plants to uptake results in leaves showing chlorosis symptoms. Cowpeas and green beans, for example, require the application of Ni-based fertilizer to optimize N fixation.[2] The continuous application of Ni-based fertilizer to enrich the soil and increase the potency of the leguminous plant to fix N in the soil successively increases Ni concentration in the soil. Even though Ni serves as a micronutrient for plants, its excesses in the soil cause more harm than good, the toxicity of Nickel in the soil minimizes the pH level in the soil and impedes iron uptake as an essential nutrient for plant growth.[1] According to Liu[3] Ni has been discovered as the 17th essential element that is required for plant development and growth. Apart from Ni playing a role in plant development and growth, it also needed by humans for various applications and uses. Ni use in various industrial sectors is neded for, electroplating, nickel-based alloy production, and for the automobile industries for ignition device and spark plug manufacturing.[4] Furthermore, Ni-based alloys and plated items have been utilized extensively in kitchen wares, fittings for ballroom, for goods in the foods industry, electricals wires and cables, turbines for jets, implants for surgical, textiles and building ships.[5] Enriched Ni levels in the soil (i.e. surface soil) are attributed to anthropogenic and natural sources, but primarily, Ni is of natural source than anthropogenic.[4] In recent time, Ni pollution in soil has become critical due to the incessant release of Ni via steel industries, the application of phosphate fertilizers, mining, processing of metals and sewage sludge discharge. According to Freedman and Hutchinson [6]and Manyiwa et al., [7] the main topsoil pollution source immediate a vicinity and adjacent environments is principally caused by Ni-Cu based smelter and mines.

Soil pollution assessment is prevalent in the recent era because of the wide range of health-related issues that arise from soil-plant relationships, soil and soil organism relationships, ecological degradation and environmental impact assessment related issues. Hitherto, spatial prediction of potentially toxic elements (PTEs) such as Ni in the soil using the conventional

means has been laborious and time-consuming. The advent of digital soil mapping (DSM) and its success chalked [8] in this present time has improved predictive soil mapping (PSM) tremendously. Predictive soil mapping, or DSM, according to Minasny and McBratney, [9] has proven to be a prominent soil science subdiscipline. Lagacherie and McBratney, 2006 defines DSM as "the *creation and population of spatial soil information systems by the use of field and laboratory observational methods coupled with spatial and non-spatial soil inference systems*". McBratney et al.,[10] outlined that DSM or PSM in contemporary time is the utmost effective technique to foretell or map the spatial distribution of PTEs, types of soil and soil properties. Geostatistics and Machine learning algorithm (MLA) are modelling techniques in DSM that employ large and minimal data via computer aid to create a digitized map.

Deutsch, [11] and Olea, [12] defines geostatitics "as *a collection of numerical techniques that deal with the characterization of spatial attributes, employing primarily random models in a manner similar to the way in which time series analysis characterizes temporal data*." Particularly, geostatitics involves the assessment of variograms, that allows to quantify and define the dependency of spatial values from every sort of dataset.[13] Gumiaux et al.,[13] further illustrated that, the assessment of variogram in geostatistics is based on the 3 principles including (a) to compute the data correlation scale, (b) to identify and calculate the anistoppies in the disparity of the dataset and (c) estimate the area effects in addition to intrinsic errors that takes in accounts measured data that is segregated from local effects. On the basis of these concepts, various interpolation techniques such as universal kriging, cokriging, ordinary krigging, empirical bayesian kriging, simple kriging, and other well-known interpolation techniques used within geostatistics are used to map or predict PTEs, soil properties, and soil types.

Machine learning algorithms (MLA) are a relatively new technique that employs the use of larger nonlinear data classes propelled by algorithms that are primarily used for data mining, identifying data patterns, and repeatedly applied to classification and regression tasks in scientific fieldswadoux et al. [14] such as soil science. Substantial research papers have relied on MLA models to predict PTEs in soil, such as Tan et al., [15](random forest for heavy metal estimation in agricultural soil), Sakizadeh et al.,[16](application of support vector machine and artificial nueral network to model soil pollition). Furthermore, Vega et al., [17](CART for modelling

heavy metal retention and sorption in soil) Sun et al., [18] (application of cubist is the distrbution of Cd in the soil) and other algorithms like *k*-nearest neighbors, generalized boosted regression and boosted regression tree also applied MLA to predict PTEs in the soil.

Application of DSM algorithms in prediction or mapping comes with several challenges. Many authors have argued the superiority of MLA to geostatitics and contrariwise. Even though one is superior to the other, the combination of the two has increased the accuracy level in mapping or prediction in DSM.[8] Woodcock and Gopal[19] Finke [20] ; Pontius and Cheuk, [21]and Grunwald[22] have commented on the imperfection and some error that exist in predictive soil mapping. Soil scientists have tried a variety of techniques to optimize the effectiveness, precision of mapping and prediction in DSM. The incorporation of uncertainty and validation is one of the many different facets that have been integrated into DSM to optimize effectiveness as well as decrease the imperfection. Nevertheless, Agyeman et al.,[8] outlined that the act of validation and the uncertainty that come with creation of map and prediction should be validated independently to enhance map quality. Recently there have been a new trend in DSM that foster the combination of geostatitics and MLA in mapping and prediction. Several soil scientist and authors such as Sergeev et al.,[23]; Subbotina et al.,[24]; Tarasov et al., [25]and Tarasov et al., [26] have harness accurate qualities in geostatistics and machine learning to generate hybrid models that increase the efficiency and quality of the prediction as well as mapping. Some of these hybrization or combined algrorithmic models are artificial nueral network-kriging (ANN-RK), multi-layer perceptron residual kriging (MLP-RK), generalized regression neural network residual kriging (GR-NNRK) [25]and artificial nueral network-kriging- multilayer perceptron (ANN-K- MLP) .[26]

According to Sergeev et al.,[23] the act of combining various modelling techniques has the potential to eliminate flaws and increase the efficiency of the hybrid model produced over the single models from which it was developed. Against this backdrop, this new paper deem it necessary to apply a combined algorithm from geostatistic and MLA to develop the best hybridized model to predict the enrichment of Ni in the urban and peri-urban area. This research will lean on empirical Bayesian kriging (EBK) as the base model and hybridizing it with support vector machine (SVM) as well as multiple linear regression(MLR) model. The

hybridization of EBK with any MLA is uncharted. The plurality of hybrid models seen is a combination of ordinary, residual, regression kriging and MLA. EBK is a geostatistical interpolation technique that involves a spatial stochastic process characterized by localize being as non-stationary/stationary random field with a defined localize parameter on the field that gives way to vary spaces across.[27] EBK has being applied in diverse research such as investigation the distribution of organic carbon in agrogray soils (Samsonova et al., 2017) [28]soil contamination assessment [29]and mapping soil properties(John et al., 2021).[30]

On the other hand self organisinising map (SeOM) is a learning algoritm that has been applied in diverse articles such as Li et al.,[31] Wang et al,[32] Hossain Bhuiyan et al.,[33] and Kebonye et al.,[34] to determine the spatial attributes and grouping of elements. Wang et al., [32] outlined that SeOM is a vigorous learning technique known for its capacity in grouping and imagining that is allowed to deal with nonlinear problems. SeOM unlike other pattern recognition techniques such principal componet analysis, fuzzy clustering, hierachical clustering and multiple criteria decision making performs better in organization and recognising the pattern of PTEs. Accroding to Wang et al.,[32] SeOM is capable of spatially grouping the distribution of related neurons and providing high-resolution data visualization. SeOM will be used to visualize Ni prediction data for the best model that was developed in order to characterize the results for easy interpretation.

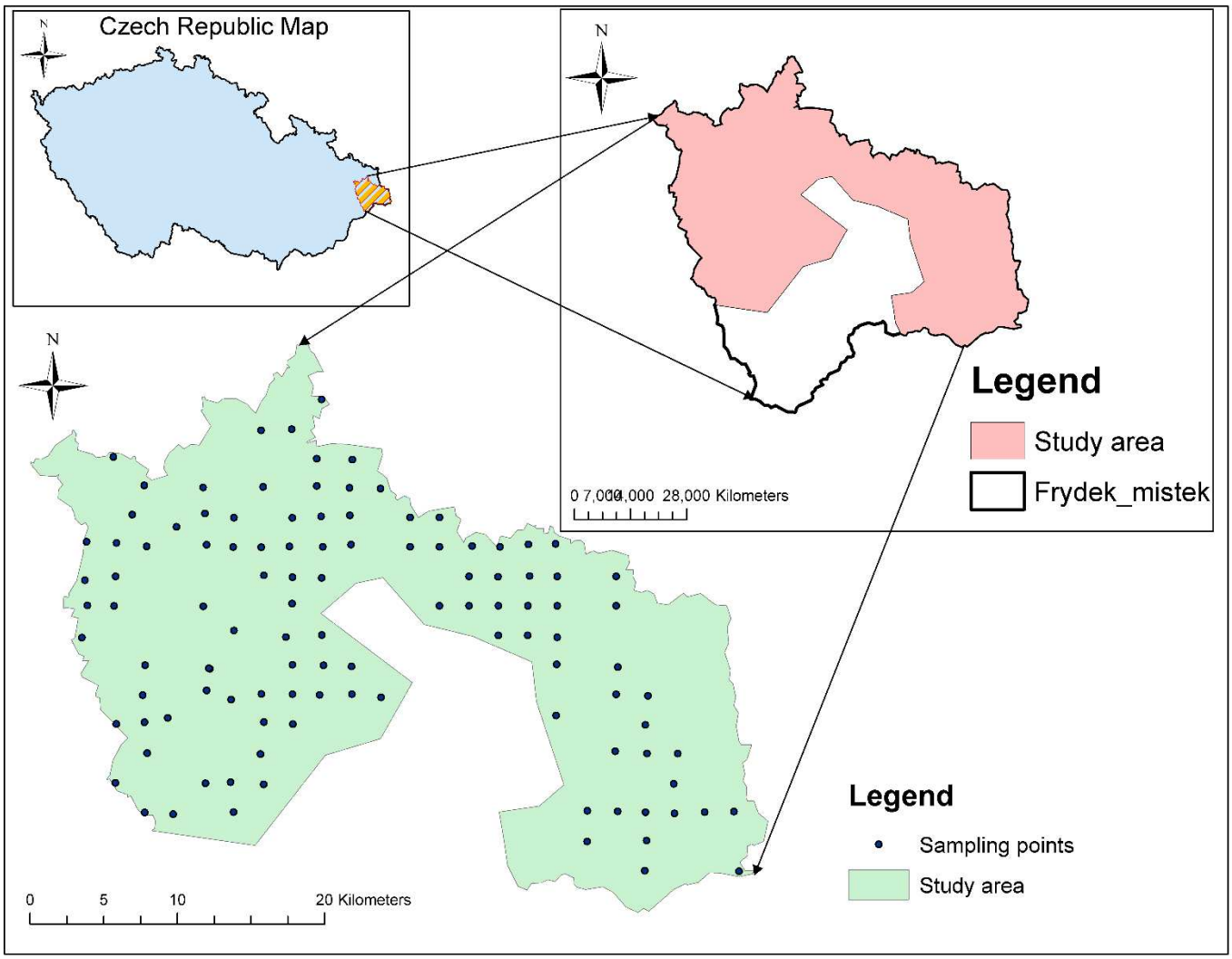
This paper intends to generate a robust mapping model with elevated resolution that predicts Ni content in urban and peri- urban soil. We hypothesized that the dependability of the hybridized model primarily relies on the influence of the other model attached to the base model. We acknowledge the challenges in DSM, and while these challenges are being addressed on several fronts, the combination of geostatitics and MLA model progression appears to be sluggish, so we will attempt to answer the research question that will potentially generate a hybrid model. Nevertheless, how accurate is the model in predicting the targeted elements? What is the efficiency assessment level based on validation and accuracy assessment? Therefore, the specific objectives of this research are (a) to create a combined hybrid model using EBK as the base model against SVMR or MLR, (b) compare the models generated (c) propose the best hybrid model to predict the concentration of Nickel in urban or

peri-urban soil and (d) to apply SeOM to create high-resolution spatial variability maps of Nickel.

MATERIALS AND METHODS

Study area

The research is being conducted in the Czech Republic, specifically in the Frydek Mistek district of the Moravian-Silesian Region. (Figure 1). The geomorphology of the study area is a very rugged landscape that is largely part of the Moravian-Silesian Beskydy region, which is a portion of the outer Carpathian mountain range. The study area falls within latitude 49° 41'0' North and longitude 18 ° 20'0' East at an altitude varying between 225 and 327 m above sea level; however, the Koppen classification system of the area's climatic situation is rated as Cfb= temperate oceanic climate with a high amount of rainfall even in dry months. Temperatures usually range between 24 ° F and 75 ° F throughout the year, seldom falling below 8 ° F or rising above 86 ° F, and annual precipitation ranges between 685 and 752 mm. [35] The district's area survey is projected to be 1208 km2, with 39.38 percent of the land area under cultivation and 49.36 percent under forest cover. The region used for this study, on the other hand, is approximately 889.8 km2. In and around the Ostrava neighbourhood, the steel industry and metal works are active. The soil properties are easily differentiated from the soil's color, structure, and carbonate content. The soil's texture is medium to fine, and it is derived from parent materials. They are either colluvial, alluvial, or aeolian in nature. Some soil areas show mottles in the top and subsoil, which are usually accompanied by concrete and bleaching. However, cambisols and stagnosols are the most common soil types in the area (Kozák, 2010).[36] With elevations ranging from 455.1 to 493.5 m, cambisols is predominate in the Czech Republic (Vacek et al., 2020).[37]



227

228 **Figure 1:** Study area map

229

Soil sampling and analysis

Topsoil samples totaling 115 were obtained from urban and peri-urban soil in the Frydek Mistek district. The sample pattern used was the regular grid, and the soil sample intervals were 2× 2 km using a handheld GPS device (Leica Zeno 5 GPS) at a depth of 0 to 20 cm for topsoil. The samples were wrapped in Ziploc bags, labeled appropriately, and transported to the laboratory. The samples were air-dried to produce a pulverized sample, crushed by a mechanical system (Fritsch disk mill pulverize), and sieved (2 mm). A gram of the dried, homogenized, and sieved soil sample (sieve size 2 mm) was placed in a Teflon bottle that was clearly labeled. In each Teflon container, 7 ml of 35% HCl and 3 ml of 65% HNO_3 were dispensed (using automatic dispensers—one for each acid) and the cap was gently closed to allow the sample to remain overnight for reactions (aqua regia procedure). The supernatant was put on a hot metal plate for 2 hours to promote the digestion process of the sample before being allowed to cool. The supernatant was transferred to a 50 ml volumetric flask and diluted to 50 ml with deionized water. After that, the diluted supernatant was filtered into 50 ml PVC tubes. In addition, 1 ml of the diluted solution was diluted with 9 ml of deionized water and filtered into a 12 ml test tube prepared for PTE pseudo-concentration in this study. ICP-OES (inductively coupled plasma optical emission spectrometry) (Thermo Fisher Scientific, USA) was used to calculate metal concentrations following standard procedures and protocols. The quality control and quality assurance process were ensured and each study by testing the reference criteria. To ensure that the errors were reduced, duplicate analysis was performed.

Empirical Bayesian kriging

Empirical Bayesian kriging (EBK) is one of the numerous geostatistical interpolation techniques used in modelling in diverse fields such as soil science. Unlike the other kriging interpolation techniques, EBK varies from conventional kriging methods by considering the error of the semivariogram model estimation.[38] In EBK interpolation, several semivariogram models are calculated during the interpolation instead of a unitary semivariogram. The interpolation technique makes way for uncertainties associated,hereby plotting semivariogram and programmed the highly complex parts of composing a sufficient kriging approach(Samsonova et al., 2017b). The interpolation process of EBK follows 3 criteria as proposed by Krivoruchko, [38](a)

the model estimate semivariogram from the input dataset (b) based on the generated semivariogram a new prediced is value against each inputted dataset location and (c) finally a model is computed from the simulated dataset. The bayesian equation rule is give as posterior

$$Prob(A,B) = Prob\left(\frac{A}{B}\right)Prob(B) = Prob(B,A) = P\left(\frac{B}{A}\right)P(A) \ equation\ 1$$

Where the $Prob(A)$ represents the prior, $Prob(B)$ marginal probability in the most instances there they are ignored, $Prob\ (B,A)$ the posterior. The semivariogram calculation is based on the Bayes rule, which exhibits the proclivity that the observed dataset can be created from the semivariogram. Krivoruchko, (2012) explains that, during the computation of semivariogram in step 1, a set of data are utlized to stimulate a new location input and however, step 2 and 3 are replicated.

Support vector machine regression

Support vector machine is a machine learning algorithm that generates an optimal disengaging hyperplane to differentiate identical but not linearly independent categories. Vapnik, [39] created the algorithm for intent classification, but it has recently been used to solve regression-oriented problems. According to Li et al., [40] SVM is one of the best classifier techniques and has been used in a variety of fields. The regression component of SVM is used in this analysis (support vector machine regression-SVMR). Cherkassky and Mulier, [41] pioneered SVMR as a regression based on kernel, and its computation was performed using a linear regression model with a multinational space function. John et al.,[42] reported that the SVMR modelling employs a hyperplane linear regression, which creates a nonlinear relationship and allows for the space function. According to Vohland et al., [43] epsilon (ε)-SVMR uses a trained dataset to obtain a represented model as an epsilon -insensitive functi used to map data independently with the optimum epsilon-deviation from dependent data training. The preset distance error within is ignored from the actual value, and if the error is larger than the epsilon (ε), the soil prop erty compensates for it. The model also reduces the intricacy of training data to a broader subset of support vectors. The equation as proposed by Vapnik, [39] is given as

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x_, x_k) + b_, \text{ equation 2}$$

In which the b represents the scalar threshold, $K(x_, x_k)$ representing the kernel function, α denoting the Lagrange multiplier, N symbolizing the number dataset, x_k representing the data input and y is the data output. One of the critical kernels used is the SVMR operation with is the gaussian radial basis function (RBF). The RBF kernel was applied to ascertain the optimum SVMR model that is essential to procure the finest penalty set factors C and the kernel parameters gamma (γ) for the PTE training data. We assessed the set of training and then tested the validation set's model performance.

Multiple linear regression

Multiple Linear Regression Model (MLR) is a regression model that embodies the relationship between a response variable and numerous predictor variables by employing linearly incorporated parameters that are computed using the least-squares method. In MLR, the least square model is a prediction function that is directed toward a soil property following the selection of an explanatory variable. Nickel was used as response variables in order to create a linear relationship using the explanatory variable. The PTE was used as the response variables which was used to establish the linear relationship utilizing the explanatory variable. The MLR equation is given as

$$y = a + \sum_{i-1}^{n} b_1 \text{ X } x_i \pm \varepsilon_i \ equation\ 3$$

In which the y represents the response variable, a denotes the intercept, n signifies the number of predictors, b_1 denotes the partial regression of coefficient, x_i implies the predictors or the explanatory variables and the ε_i signifies the error in the model which is also called residual. The model was utilized in R.

Hybrid modelling

The hybrid models were obtained by sandwiching the EBK as the base model with SVMR and MLR. This was done by extracting of predicted values from the EBK interpolation. The predicted values obtained from interpolated Ca, K and Mg was passed through permutation process to

obtain new variables such as CaK, CaMg and KMg. The elements Ca, K and Mg was then combined to obtain the fourth variable CaKMg. Overall, the variables obtained were Ca, K, Mg, CaK, CaMg, KMg and CaKMg. These variables became our predictors that will aid in predicting Nickel concentration in urban and peri-urban soil. The predictors will be subjected through SVMR algorithm obtain a hybrid model Empirical bayesian kriging – Support vector machine (EBK_SVM). Similarly, the variables will be piped through MLR algorithm to likewise obtain a hybrid model Empirical bayesian kriging -multiple linear regression (EBK_MLR). Generally, the variables Ca, K, Mg, CaK, CaMg, KMg and CaKMg were used as covariates which served as predictors in predicting the Nickel content in urban and peri-urban soil. The finest model (EBK_SVM or EBK_MLR) obtained will then be visualized using the self-organizing map. The workflow of the study is presented in Figure 2.

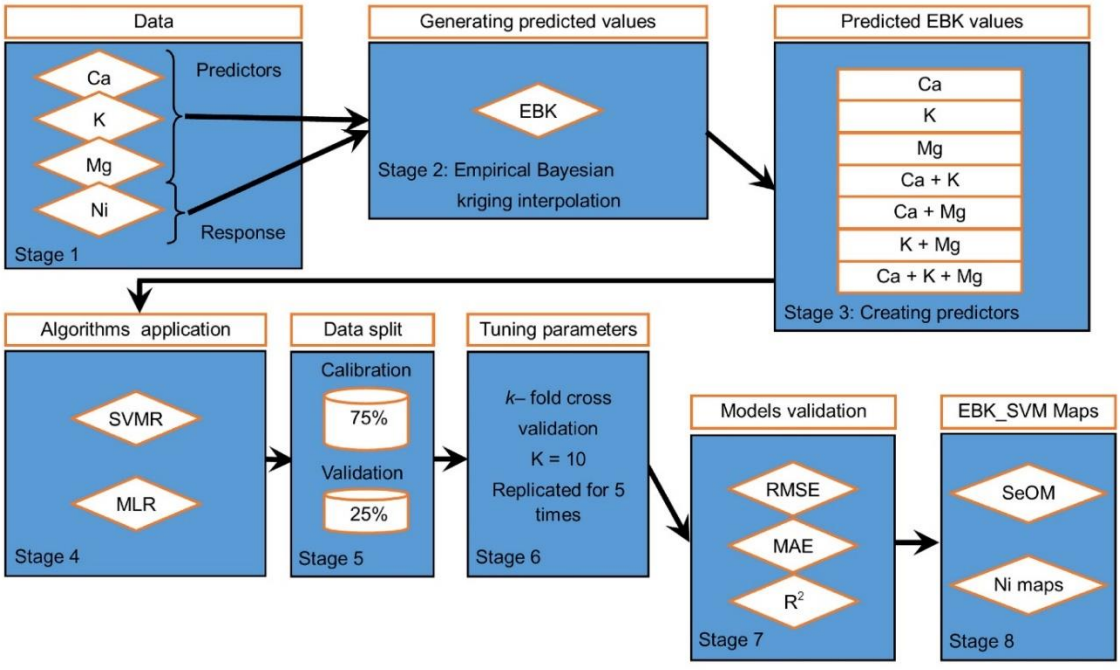


Figure 2: Flowchart of the study

Self-organizing maps (SeOM)

Using SeOM has become a popular tool used in diverse fields for the organizing, assessment and prediction of data in the financial sector, medical sector, industrial sector, in statistics, soil science, etc. 1 developed SeOM using an artificial neural network for organization, evaluation, and prediction, as well as unsupervised learning techniques. In this study, SeOM was used to visualize the concentration of Ni based on the finest model used in the prediction of Ni in urban and peri-urban soil. The data treated in SeOM assessment serves as an n input dimensional vector variable . [44] [45] Melssen et al., [46] delineated that an input vector is connected to an output vector with a single weight vector by a single input layer into a neural network. The output generated from SeOM comes out as a two-dimensional map made up of diverse neurons or nodes knitted together into either a hexagonal, circular or square topological plot based on their proximity.[31] Map sizes were compared baed on metrics, quantization error (QE) and topographic error (TE), and a SeOM model with 0.086 and 0.904 respectively was chosen which was a 55-map unit (5×11). The neuron structure was chosen based on empirical equation node number given as

$$m = 5 \times \sqrt{n}$$

In which the m denotes the quantity of SeOM map neurons, n representing the input data quantity.

Data partitioning

The number data used in this study is 115 samples A random method was employed to dissect the data into a test data (25% for validation) and a training dataset (75% for calibration). The training dataset was used to produce the regression models (calibration), and the test dataset was used to validate generalization capabilities.[47] This was done to evaluate the appropriateness of the diverse models that is being used to predict nickel content in the soil. All the models used were subjected to a 10-fold cross-validation process that was replicated five times. The variables generated from EBK interpolation were used as the predictors or explanatory variable to predict you target variables (PTEs). The modelling process was performed in R.

Model and accuracy assessment

A variety of validation parameters were used to determine the best and finest model suitable for the prediction of nickel concentration in the soil and evaluating the accuracy of the model as well as its validation. The hybridized models were assessed using mean absolute error (MAE), root mean square error (RMSE), and R square, or coefficient determination (R^2). R^2 defines the variance of the proportion in the answer and is represented by the regression model. The RMSE and the magnitude of the variance within the independent measurement describe the model prediction power, while MAE determines the true quantitative value. The R^2 value must be high to evaluate the best hybridized model using the validation parameters, and the closer the value is to 1, the higher the accuracy. According to Li et al., [48]an R^2 criteria value of 0.75 or greater is considered a good prediction, from 0.5 to 0.75 is acceptable model performance and below 0.5 is unacceptable model performance. A lower obtained value is sufficient and considered best for the selection of a model using the RMSE and MAE validation criteria evaluation methods. The following equation describes the validation methods.

Mean absolute error

$$MAE = \frac{1}{n}\sum_{i=1}^{n} Y_i - \hat{Y}_i$$

R square

$$\mathrm{R}^2 = 1 - \frac{\sum(\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i})^2}{\sum(\mathrm{Y_i} - \hat{\mathrm{Y}}_\mathrm{i})^2}$$

Root mean square error

$$RMSE(mg/kg) = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

Whereby n represents the size of the observations Y_i represents the measured response and the $\hat{Y}_i$ also stated as the predicted response values, accordingly, for the ith observation term.

RESULTS AND DISCUSSION

Statistical description

The statistical description of the predictors and the response variables are shown in Table 1, displaying the mean, standard deviation (SD), coefficient of variability (CV), minimum value, maximum value, kurtosis and skewness. The elements minimum and maximum values descend in this Mg< Ca < K < Ni and Ca< Mg< K< Ni order, respectively. The concentration of the response variable (Ni) sampled from the study area ranged from 4.86 mg/kg to 42.39 mg/kg. Comparing Niwith the world average value (29 mg/kg) and the European average value (37 mg/kg) indicates that the overall computed geometric mean of the study area is under tolerable limits. Nevertheless, comparing the mean concentration of Nickel (Ni) in this current study to the agricultural soils in Sweden, as indicated by Kabata-Pendias, [49] exhibits that the current mean concentration of Ni is higher. Similarly, the mean concentration of in the current study (Ni 16.15 mg/kg) of the urban and peri-urban soil in Frydek Mistek is higher than the permissible limit for Ni in urban soil in Poland as reported by Różański et al., [50] (10.2 mg/kg). Furthermore, the mean nickel concentration in Tuscan urban soil recorded by Bretzel and Calderisi, [51] (1.78 mg/kg) is very low in comparison to the current study. Jim, [52] also identified a low Ni concentration in Hong Kong urban soil (12.34 mg/kg), which was lower than the current research. Birke et al., [53] reported a Ni mean concentration of 1.45 mg/kg in an old mining and urban industrial area in Sachsen Anhalt, Germany, which was 1.45 mg/kg higher than the Ni mean concentration in the current study. The concentration of Ni in some parts of the study area's urban and peri-urban soil that exceeds the allowable limit might be attributed largely to steel industries and metal works. This is inline with Khodadoust et al., [54] study that steel industries and metal works are major sources of nickel pollution in the soil. However, the predictor variables also ranged from 538.70 mg/kg to 69161.80 mg/kg for Ca, 497.51 mg/kg to 3535.68 mg/kg for K and 685.68 mg/kg to 5970.05 mg/kg for Mg. Jakovljevic et al., [55] investigated the total content of Mg and K in central Serbian soil. They found that the total concentration (410 mg/kg and 400 mg/kg, respectively) was lower than in the current analysis. Indistinguishably, in eastern Poland,Orzechowski and Smolczynski, [56] assessed the total content of Ca, Mg and K and the results suggested that the mean concentration Ca (1100 mg/kg), Mg

(590mg/kg) and K (810 mg/kg) in the topsoil were lower the respective element in this present study. A recent study conducted by Pongrac et al., [57] revealed that Ca total content analyzed in 3 different soil in Scotland Uk (Mylnefield soil, Balruddery soil and Hartwood soil) suggested the Ca content of the present study is higher.

The dataset distribution of the elements exhibited different skewness due to the differences in the measured concentration of the elements sampled. The skewness and the kurtosis of the elements ranged from 1.53 to 7.24 and 2.49 to 54.16 correspondingly. All the computed skewness and kurtosis level of the elements were above +1 and it thus indicates that the data distribution is irregular skewed in the right direction and leptokurtic. The estimated CV of the elements also suggested that K, Mg and Ni showed a moderate variability, whereas Ca extremely high variability. The CV of K, Ni and Mg explained that they are homogeneously distributed. Moreover, Ca distribution is non-homogeneous and its level of enrichment might be influenced by an external source.

Table 1: Statistical description of predictors and response

	Ca	K	Mg	Ni
	◄	———	►	
		mg/kg		
		Predictor s		Response
Mean	3624.83	1289.75	1981.91	16.15
Standard deviation	7969.72	446.87	666.69	6.78
Coefficient of Variability	219.86	34.65	33.64	41.97
Minimum value	538.70	497.51	685.68	4.86
Maximum value	69161.80	3535.68	5970.05	42.39
Kurtosis	54.16	4.82	11.67	2.49
Skewness	7.24	1.54	2.48	1.63

Correlation between response and predictor variable

The correlation of the predictors against the response element suggested a satisfactory correlation among the elements (see Figure 3). The correlation suggested that CaK showed a moderate correlation with r value = 0.53 and CaNi similarly displayed moderate correlation. Even though Ca and K showed moderate nexus among each other but researchers such as Kingston et

al., [58]and Santo, [59] have suggested that their content in the soil is inversely proportional. However, Ca and Mg are antagonistic to K but CaK correlated very well. This might be due to the application of fertilizer such as potassium carbonate that is 56% richer in potassium. Potassium correlated moderately with magnesium (KMg r = 0.63) and in the fertilizer industry these two elements have a history of strong relationships due to the application of potassium magnesium sulfate, potassium magnesium nitrate and muriate of potash to the soil to enhance its deficiency level. Nickel correlated moderately with Ca, K and Mg with r values = 0.52, 0.63 and 0.55 respectively. The relationships involving calcium, magnesium, and PTE such as nickel are complicated but notwithstanding magnesium inhibits calcium absorption, calcium decreasing the effects of excess magnesium, and both magnesium and calcium reduce the toxicity effects of the nickel in the soil.

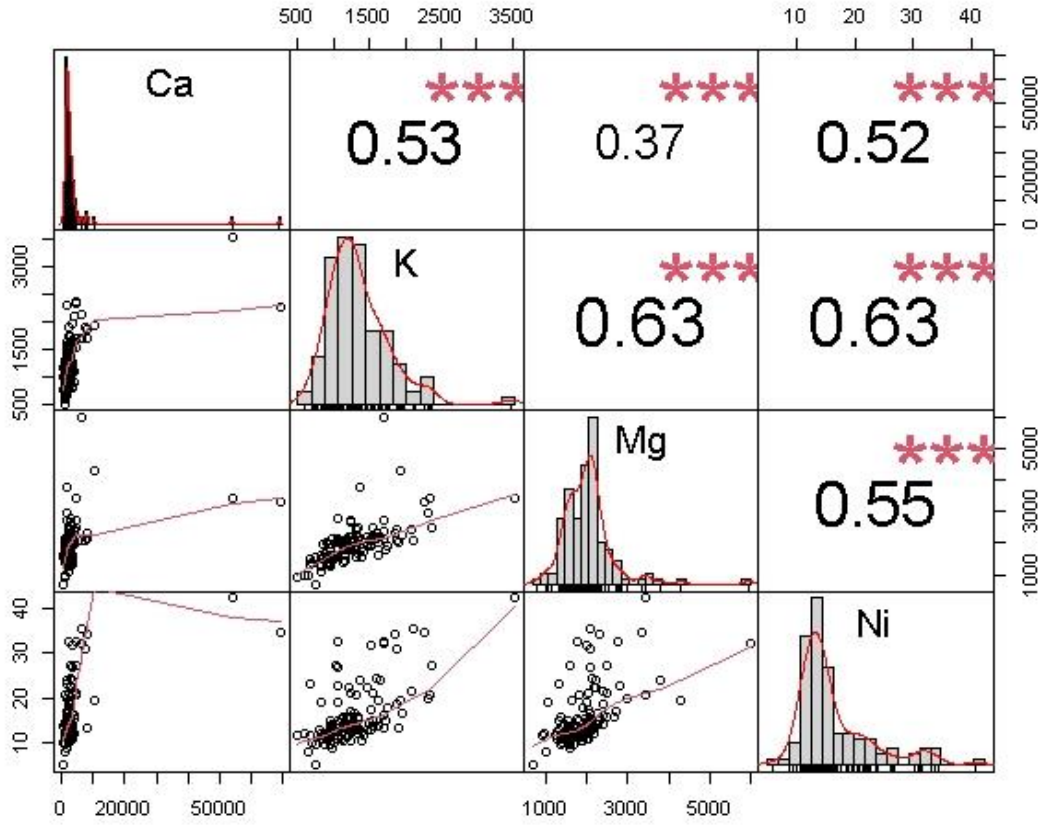


Figure 3: Correlation matrix of the elements showing the relationship between predictors and response (Note: The plot includes scatter plots between the element, and the significance levels is based on p < 0,001)

Spatial distribution of the elements

The distribution of the predictors and the response variables was interpolated using empirical Bayesian kriging. According to Burgos et al., [60] application of spatial distribution is a technique used to quantify anhighlight hot spots of polluted areas. The enricnment level of Ca in Figure 4 can be seen in the northwestern part of the spatial distribution map. The map shows moderate to high hotspots of Ca enrichment. Calcium enrichment in the northwestern part of the map might be due to the application of quicklime (Calcium oxide) to reduce soil acidity and its application in steel plants as basic oxygen in steel making process. On the other hands, other farmers prefer to use calcium hydroxide in acidic soil to neutralize the pH level, which also surges the calcium content of the soil.[61] Potassium exhibited hot spots in the northwestern part of the map and eastern part as well. The Northwestern part is the predominantly agrarian community, and a moderate to a high pattern of K might be due to the application of NPK and muriate of potash. This is coherent with other authors such as Madaras and Lipavský, [62] Madaras et al., [63] Pulkrabová et al., [64]Asare et al., [65] who observed that using muriate of potash and NPK for soil stabilization and treatment resulted in high K content in the soil. Potassium enrichment in the northwestern part of the spatial distribution map might be due to applying potassium-based fertilizers such as potassium chloride, potassium sulphate, potassium nitrate, sylvinite, and kainit in increase the deficient soil to increase its k content. Zádorová et al., [66] and Tlustoš et al., [67] outlined that the application of potassium-based fertilizer increases the potassium level in the soil and, by a long effect will significantly upsurge soil nutrient content, especially K. Magnesium showed a hot spot in the northwestern part of the map and relatively moderate hotspot in the southeastern part of the map. Colloid fixation in soil depletes the concentration of magnesium in the soil. Its deficiency in the soil causes plants to portray interveinal chlorosis of yellowish colouration. Magnesium-based fertilizers, such as potassium magnesium sulphate, magnesium sulphate and Kieserite, treat deficiency syndrome (purple, red or borwn coloration of plants indicating lack magnnesium) in soils with normal pH ranges 2. The accumulation of Nickel on the surface of the urban and peri-urban soil might be due to anthropogenic activities such as agriculture and NNickel importance in stainless steel production 3.

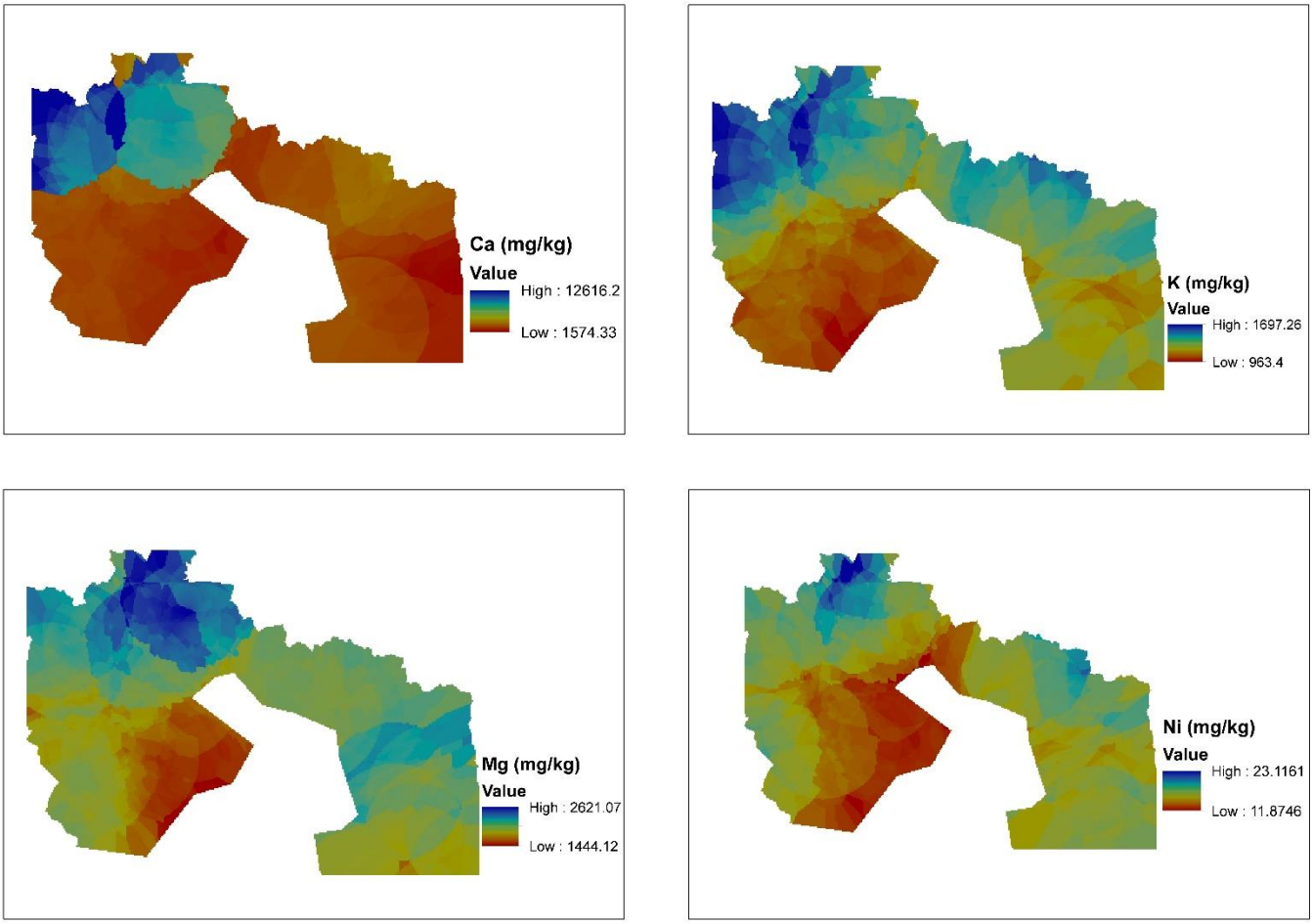


Figure 4: Spatial distribution of the elements

Performance of models

The performance of individual approaches for predicting Ni content in urban and peri-urban soil was assessed using the models' performance (Table 2). Table 2 comparison assessment compares support vector machine regression (SVMR), empirical Bayesian kriging support vector machine regression (EBK SVMR), and multiple linear regression -support vector machine regression (MLR SVMR). Model validation and accuracy assessment confirmed that the Ca_ Mg_ K predictors and the EBK SVMR model yielded the optimal performance. The R^2, the root means square error (RMSE) and the mean absolute error (MAE) of the calibrated model Ca_Mg_K- EBK_SVMR model obtained 0.637 (R^2), 95.479 mg/kg (RMSE) and 77.368 mg/kg (MAE) as against 0.663 (R^2), 235.974 mg/kg (RMSE) and 166.946mg/kg (MAE) for Ca_Mg_K-SVMR. Despite that, Ca_Mg_K-SVMR (0.663 mg/kg R^2), and Ca_Mg-EBK_SVMR (0.643 =R^2)

obtaining a good R^2 their RMSE and MAE results were higher than that of Ca_Mg_K-EBK_SVMR (see Table 2). Moreover, the RMSE and MAE of the Ca_Mg-EBK_SVMR model are 17.5 and 13.4, bigger than that of Ca_Mg_K-EBK_SVMR. Similarly, the RMSE and MAE of Ca_Mg-K SVMR model is equally bigger than Ca_Mg_K-EBK_SVMR RMSE and MAE by a margin of 2.5 and 2.2 respectively. The cross-validation accuracy assessment Ca_Mg_K-EBK_SVMR hybrid model predicts Ni content in the urban and peri- urban soil 63.70% accuracy level. This level accuracy, according to Li et al., [48] is an acceptable model performance rate.

The current results compared to a previous study by Tarasov et al.,[25] whose hybridize model created MLPRK (multi-layer perceptron residual kriging) to the current study EBK_SVMR accuracy assessment indices reported with regards, RMSE(210), and MAE(167.5) were higher than the results we had in the current study (RMSE 95.479,MAE 77.368). However, when the R^2 (0.637) of the current study is compared to the R^2(0.544) of Tarasov et al., [25] it is clear that the coeffiecient of determination (R^2) in this hybrid model is higher. The maginal errors (RMSE and MAE) of the hybrid model (EBK SVMR) are two times lower. Similarly, Sergeev et al., [23] recorded 0.28 (R^2) for the hybrid model developed (multi-layer perceptron residual kriging), compared to 0.637 (R^2) for Ni in the current study. The prediction accuracy level of this model (EBK SVMR) is 63.7% as opposed to 28 percent Sergeev et al.[23]

Table 2: Model comparison using diverse prediction models

	Models								
	SVMR			**EBK _SVMR**			**EBK_MLR**		
				mg/kg					
Predictors	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**	**R2**	**RMSE**	**MAE**
Ca	2.78E-01	1478.87	1015.300	6.85E-05	2710.6	1402.38	2.160E-02	6.246	4.425
K	0.306	323.602	228.251	0.079	144.949	120.139	6.000E-02	6.123	4.324
Mg	0.161	418.771	324.731	0.002	162.293	122.298	1.350E-01	5.872	4.296
Ca_Mg	0.416	346.868	259.569	0.643	1664.64	1031.49	1.480E-01	5.829	4.314
Ca_K	0.584	312.264	242.796	0.533	113.175	91.920	6.100E-02	6.119	4.320
K_Mg	0.511	281.239	196.885	0.523	114.201	90.926	1.350E-01	5.871	4.302
Ca_Mg_K	0.663	235.974	166.946	**0.637**	**95.479**	**77.368**	1.500E-01	5.823	4.320

Visualization of predicted Nickel via EBK_SVMR model using self-organizing map

Presented in Figure 5 is the PTEs concentrations as component planes comprising of individual neurons. No component plane exhibited the same color pattern as shown. However the appropriate number of neurons per plotted map was 55. The SeOMs were made using various colors, and the more similar the color pattern, the more comparable the sample attributes are. The single elements (Ca, K, and Mg) displayed a similar color pattern with single high neurons and majority low neuron according to the its color precise scale. Consequently, CaK and CaMg shared some similarities with very high-level neurons and low to moderate colour patterns as well. Both models predicted the concentration of Nickel in the soil by displaying moderate to high shades of colors such as red, orange, and yellow. The KMg model showed a lot of high color pattern according to the precise scale as well as low to moderate patches of colours. The component plane distribution patterns of the models revealed high colour pattern according to the precise color scale ranging from from low to high, indicating the potential concentration of Nickel in the soil (see Fig 4). The CakMg model component plane showed a diverse colour pattern from low to high according to the precise color scale. In additament, this model's prediction of nickel content (CakMg) is similar to the spatial distribution map of Nickel shown in Figure 3. Both maps revealed high, moderate, and low proportional Nickel concentrations in urban and peri-urban soil.

Figure 6 depicts the silhouette method in k-mean groupings on the maps, which are divided into three clusters based on the predicted values in each model. The silhouette method indicated the optimal clustering number. Cluster 1 obtained the most soil samples,74, out of the 115 collected. Cluster 2 received 33 samples, while Cluster 3 received 8 samples.

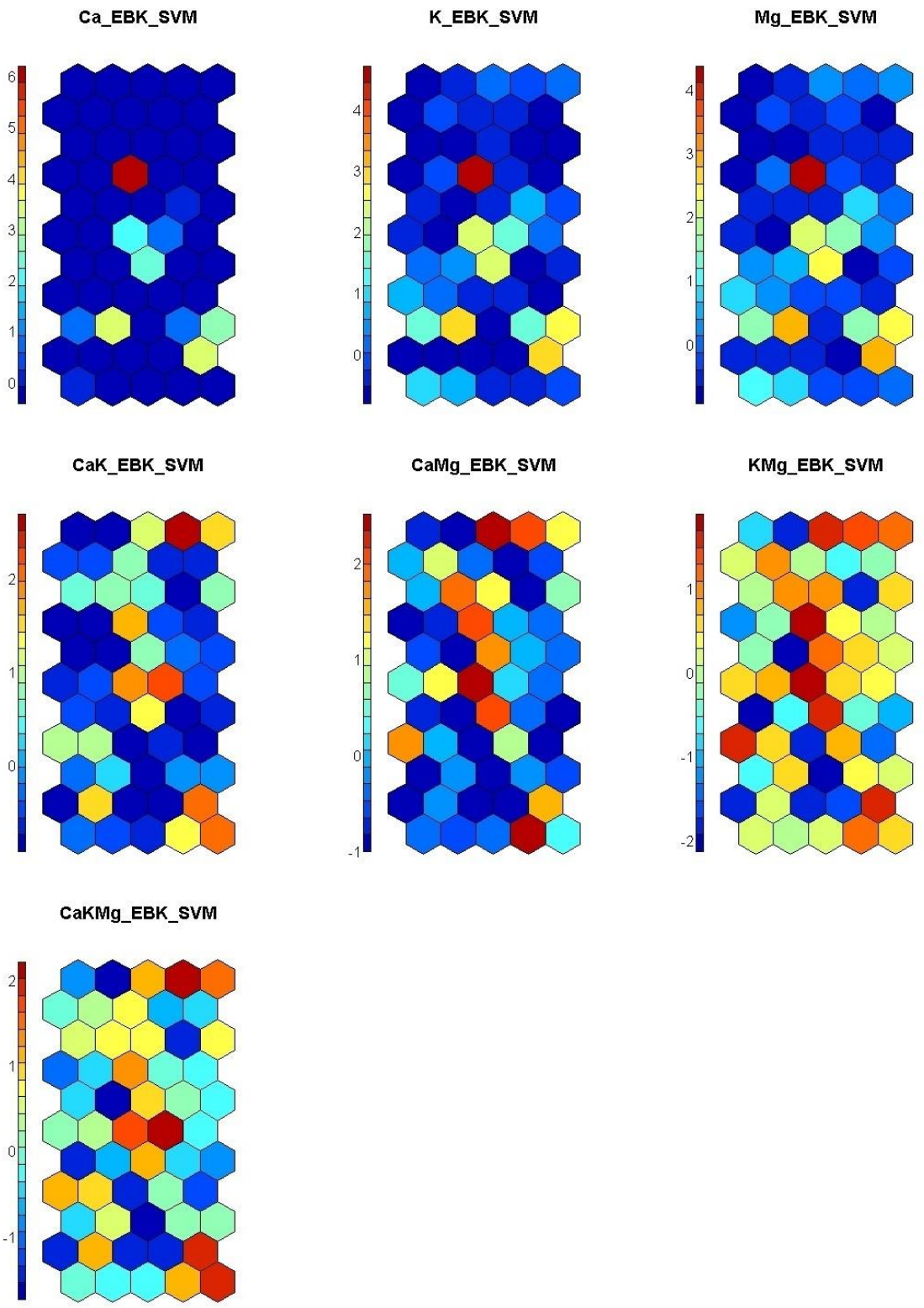


Figure 5: Component planes for each empirical bayesian kriging -support vector machine (EBK_SVM_SeOM) variable output.

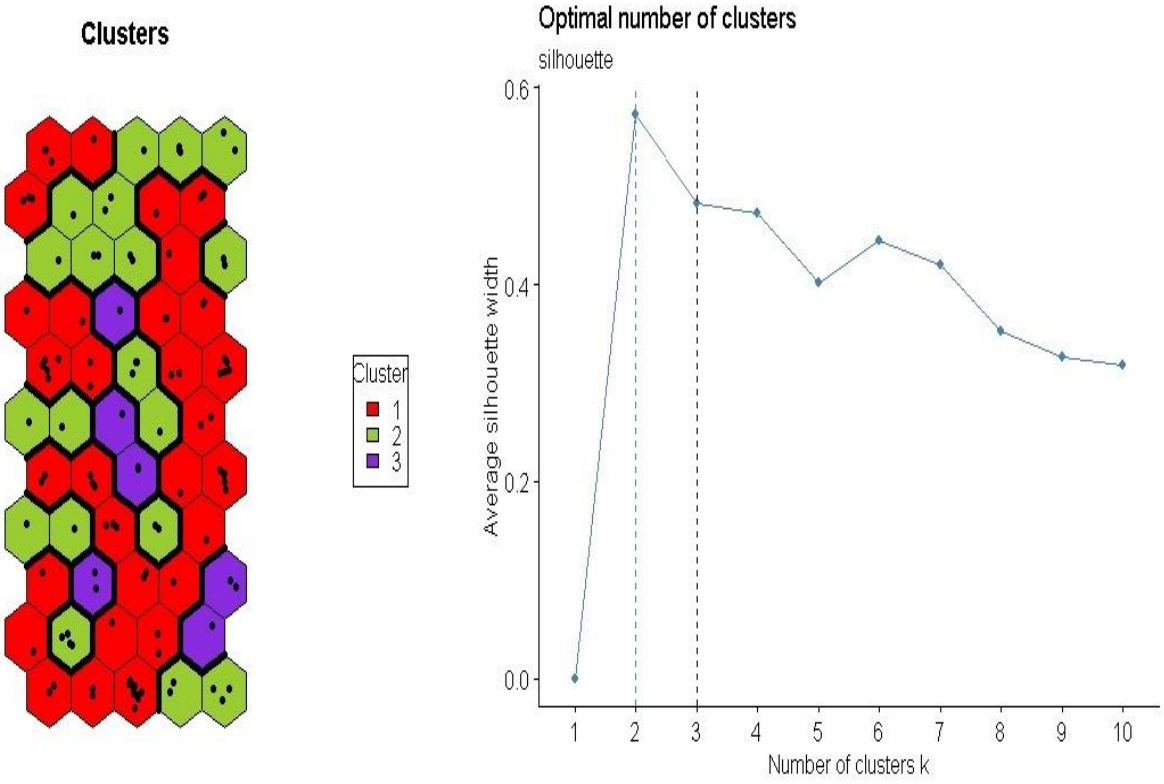


Figure 6: Different clusters classification components

Conclusion

The current research clearly illustrates a modeling technique for nickel concentration in urban and peri-urban soil. The study tested different modeling techniques, combining elements with modeling techniques to obtain the best method for predicting nickel concentration in soil. The results indicated that the support vector machine regression model (Ca Mg K-SVMR) predicted the concentration of Ni in the soil as a unitary model, but validation and accuracy evaluation parameters revealed that the error in terms of RMSE and MAE was very high. The modeling technique employed utilizing EBK_MLR models, on the other hand, was similarly deficient due to the low coefficient of determination (R^2) values. The use of EBK SVMR and combined elements (CaKMg) resulted in good results with low RMSE and MAE error and a 63.7 percent accuracy level. The results proved that combining the EBK algorithm with a machine learning algorithm can generate a hybrid algorithm that can predict the concentration of PTEs in soil. The study demonstrated the EBK model's potential to minimize error levels and increase the accuracy level of spatial distribution models of soils in urban or peri-urban soil. Generally, we suggest the application of EBK-SVMR model for assessing and predicting PTEs in the soil, moreover, hybridization using EBK with various machine learning algorithms is also recommended.

Acknowledgement

This study was supported by an internal Ph.D. grant no. 21130/1312/3150 of the Faculty of Agrobiology, Food and Natural Resources of the Czech University of Life Sciences Prague (CZU). The support from the Ministry of Education, Youth and Sports of the Czech Republic (project No. CZ.02.1.01/0.0/0.0/16_019/0000845) is also acknowledged. Finally, The Centre of Excellence (Centre of the investigation of synthesis and transformation of nutritional substances in the food chain in interaction with potentially risk substances of anthropogenic origin: comprehensive assessment of the soil contamination risks for the quality of agricultural products, NutRisk Centre).

Declaration of Competing Interest

The authors declare that they have no known competing personal interests or relationships that could have appeared to influence the scientific work in this manuscript.

620 Reference

621 (1) PlantProbs.net. Nickel in plants and soil
622 https://plantprobs.net/plant/nutrientImbalances/sodium.html (accessed Apr 28, 2021).

623 (2) Guodong Liu, E. H. Simonne, and Y. L. Nickel Nutrition in Plants | EDIS. *EDis* **2011**.

624 (3) Liu, G. D. "A New Essential Mineral Element–Nickel." *Plants Nutr. Fertil. Sci.* **2001**.

625 (4) Kabata-Pendias, A.; Mukherjee, A. *Trace Elements from Soil to Human*; 2007.

626 (5) Kasprzak, K. S. Nickel. Advances in Modern Environmental Toxicology. *Environ. Toxicol.* **1987**, *11*,
627 145–183.

628 (6) Freedman, B.; Hutchinson, T. C. Pollutant Inputs from the Atmosphere and Accumulations in Soils
629 and Vegetation near a Nickel-Copper Smelter at Sudbury, Ontario, Canada. *Can. J. Bot.* **1980**, *58*
630 (1), 108–132. https://doi.org/10.1139/b80-014.

631 (7) Manyiwa, T.; Ultra, V. U.; Rantong, G.; Opaletswe, K. A.; Gabankitse, G.; Taupedi, S. B.; Gajaje, K.
632 Heavy Metals in Soil, Plants, and Associated Risk on Grazing Ruminants in the Vicinity of Cu–Ni
633 Mine in Selebi-Phikwe, Botswana. *Environ. Geochem. Health* **2021**.
634 https://doi.org/10.1007/s10653-021-00918-x.

635 (8) Agyeman, P. C.; Ahado, S. K.; Borůvka, L.; Biney, J. K. M.; Sarkodie, V. Y. O.; Kebonye, N. M.;
636 Kingsley, J. Trend Analysis of Global Usage of Digital Soil Mapping Models in the Prediction of
637 Potentially Toxic Elements in Soil/Sediments: A Bibliometric Review. *Environmental Geochemistry*
638 *and Health*. Springer Science and Business Media B.V. 2020. https://doi.org/10.1007/s10653-
639 020-00742-9.

640 (9) Minasny, B.; McBratney, A. B. Digital Soil Mapping: A Brief History and Some Lessons. *Geoderma*
641 **2016**, *264*, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017.

642 (10) McBratney, A. B.; Mendonça Santos, M. L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**,
643 *117* (1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

644 (11) Deutsch.C.V. Geostatistical Reservoir Modeling,... - Google Scholar
645 https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=C.V.+Deutsch%2C+2002%2C+Geos
646 tatistical+Reservoir+Modeling%2C+Oxford+University+Press%2C+376+pages.+&btnG= (accessed
647 Apr 28, 2021).

648 (12) Olea, R. A. Geostatistics for Engineers & Earth Scientists. *Stoch. Environ. Res. Risk Assess.* **2000**, *14*
649 (3), 207–209. https://doi.org/10.1007/pl00009782.

650 (13) Gumiaux, C.; Gapais, D.; Brun, J. P. Geostatistics Applied to Best-Fit Interpolation of Orientation
651 Data. *Tectonophysics* **2003**, *376* (3–4), 241–259. https://doi.org/10.1016/j.tecto.2003.08.008.

652 (14) Wadoux, A. M. J. C.; Minasny, B.; McBratney, A. B. Machine Learning for Digital Soil Mapping:
653 Applications, Challenges and Suggested Solutions. *Earth-Science Reviews*. Elsevier B.V. November
654 1, 2020, p 103359. https://doi.org/10.1016/j.earscirev.2020.103359.

655 (15) Tan, K.; Wang, H.; Chen, L.; Du, Q.; Du, P.; Pan, C. Estimation of the Spatial Distribution of Heavy
656 Metal in Agricultural Soils Using Airborne Hyperspectral Imaging and Random Forest. *J. Hazard.*
657 *Mater.* **2020**, *382*, 120987. https://doi.org/10.1016/j.jhazmat.2019.120987.

(16) Sakizadeh, M.; Mirzaei, R.; Ghorbani, H. Support Vector Machine and Artificial Neural Network to Model Soil Pollution: A Case Study in Semnan Province, Iran. *Neural Comput. Appl.* **2017**, *28* (11), 3229–3238. https://doi.org/10.1007/s00521-016-2231-x.

(17) Vega, F. A.; Matías, J. M.; Andrade, M. L.; Reigosa, M. J.; Covelo, E. F. Classification and Regression Trees (CARTs) for Modelling the Sorption and Retention of Heavy Metals by Soil. *J. Hazard. Mater.* **2009**, *167* (1–3), 615–624. https://doi.org/10.1016/j.jhazmat.2009.01.016.

(18) Sun, H.; Guo, Z. X.; Guo, Y.; Yuan, Y. Z.; Chai, M.; Bi, R. T.; Yang, J. Prediction of Distribution of Soil Cd Concentrations in Guangdong Province, China. *Huanjing Kexue/Environmental Sci.* **2017**, *38* (5), 2111–2124. https://doi.org/10.13227/j.hjkx.201611006.

(19) Woodcock, C. E.; Gopal, S. Fuzzy Set Theory and Thematic Maps: Accuracy Assessment and Area Estimation. *Int. J. Geogr. Inf. Sci.* **2000**, *14* (2), 153–172. https://doi.org/10.1080/136588100240895.

(20) Finke, P. A. Chapter 39 Quality Assessment of Digital Soil Maps: Producers and Users Perspectives. *Dev. Soil Sci.* **2006**, *31* (C). https://doi.org/10.1016/S0166-2481(06)31039-2.

(21) Pontius, R. G.; Cheuk, M. L. A Generalized Cross-Tabulation Matrix to Compare Soft-Classified Maps at Multiple Resolutions. *Int. J. Geogr. Inf. Sci.* **2006**, *20* (1), 1–30. https://doi.org/10.1080/13658810500391024.

(22) Grunwald, S. Multi-Criteria Characterization of Recent Digital Soil Mapping and Modeling Approaches. *Geoderma* **2009**, *152* (3–4), 195–207. https://doi.org/10.1016/j.geoderma.2009.06.003.

(23) Sergeev, A. P.; Tarasov, D. A.; Buevich, A. G.; Subbotina, I. E.; Shichkin, A. V.; Sergeeva, M. V.; Lvova, O. A. High Variation Subarctic Topsoil Pollutant Concentration Prediction Using Neural Network Residual Kriging. *AIP Conf. Proc.* **2017**, *1836* (June 2017). https://doi.org/10.1063/1.4981963.

(24) Subbotina, I. E.; Buevich, A. G.; Shichkin, A. V.; Sergeev, A. P.; Tarasov, D. A.; Tyagunov, A. G.; Sergeeva, M. V.; Baglaeva, E. M. Multilayer Perceptron, Generalized Regression Neural Network, and Hybrid Model in Predicting the Spatial Distribution of Impurity in the Topsoil of Urbanized Area. *AIP Conf. Proc.* **2018**, *1982*. https://doi.org/10.1063/1.5045410.

(25) Tarasov, D. A.; Buevich, A. G.; Sergeev, A. P.; Shichkin, A. V. High Variation Topsoil Pollution Forecasting in the Russian Subarctic: Using Artificial Neural Networks Combined with Residual Kriging. *Appl. Geochemistry* **2018**, *88*, 188–197. https://doi.org/10.1016/j.apgeochem.2017.07.007.

(26) Tarasov, D.; Buevich, A.; Shichkin, A.; Subbotina, I.; Tyagunov, A.; Baglaeva, E. Chromium Distribution Forecasting Using Multilayer Perceptron Neural Network and Multilayer Perceptron Residual Kriging. In *AIP Conference Proceedings*; American Institute of Physics Inc., 2018; Vol. 1978, p 440019. https://doi.org/10.1063/1.5044048.

(27) Gribov, A.; Krivoruchko, K. Empirical Bayesian Kriging Implementation and Usage. *Sci. Total Environ.* **2020**, *722*. https://doi.org/10.1016/j.scitotenv.2020.137290.

(28) Samsonova, V. P.; Blagoveshchenskii, Y. N.; Meshalkina, Y. L. Use of Empirical Bayesian Kriging for Revealing Heterogeneities in the Distribution of Organic Carbon on Agricultural Lands. *Eurasian Soil Sci.* **2017**, *50* (3), 305–311. https://doi.org/10.1134/S1064229317030103.

(29) Fabijańczyk, P.; Zawadzki, J.; Magiera, T. Magnetometric Assessment of Soil Contamination in Problematic Area Using Empirical Bayesian and Indicator Kriging: A Case Study in Upper Silesia, Poland. *Geoderma* **2017**, *308*, 69–77. https://doi.org/10.1016/j.geoderma.2017.08.029.

(30) John, K.; Afu, S. M.; Isong, I. A.; Aki, E. E.; Kebonye, N. M.; Ayito, E. O.; Chapman, P. A.; Eyong, M. O.; Penížek, V. Mapping Soil Properties with Soil-Environmental Covariates Using Geostatistics and Multivariate Statistics. *Int. J. Environ. Sci. Technol.* **2021**, 1–16. https://doi.org/10.1007/s13762-020-03089-x.

(31) Li, T.; Sun, G.; Yang, C.; Liang, K.; Ma, S.; Huang, L. Using Self-Organizing Map for Coastal Water Quality Classification: Towards a Better Understanding of Patterns and Processes. *Sci. Total Environ.* **2018**, *628–629*, 1446–1459. https://doi.org/10.1016/j.scitotenv.2018.02.163.

(32) Wang, Z.; Xiao, J.; Wang, L.; Liang, T.; Guo, Q.; Guan, Y.; Rinklebe, J. Elucidating the Differentiation of Soil Heavy Metals under Different Land Uses with Geographically Weighted Regression and Self-Organizing Map. *Environ. Pollut.* **2020**, *260*. https://doi.org/10.1016/j.envpol.2020.114065.

(33) Hossain Bhuiyan, M. A.; Chandra Karmaker, S.; Bodrud-Doza, M.; Rakib, M. A.; Saha, B. B. Enrichment, Sources and Ecological Risk Mapping of Heavy Metals in Agricultural Soils of Dhaka District Employing SOM, PMF and GIS Methods. *Chemosphere* **2021**, *263*. https://doi.org/10.1016/j.chemosphere.2020.128339.

(34) Kebonye, N. M.; Eze, P. N.; John, K.; Gholizadeh, A.; Dajčl, J.; Drábek, O.; Němeček, K.; Borůvka, L. Self-Organizing Map Artificial Neural Networks and Sequential Gaussian Simulation Technique for Mapping Potentially Toxic Element Hotspots in Polluted Mining Soils. *J. Geochemical Explor.* **2021**, *222*, 106680. https://doi.org/10.1016/j.gexplo.2020.106680.

(35) Weather Spark. Average Weather in Frýdek-Místek, Czechia, Year Round - Weather Spark https://weatherspark.com/y/83671/Average-Weather-in-Frýdek-Místek-Czechia-Year-Round (accessed Sep 14, 2020).

(36) Kozák, J. Soil Atlas of the Czech Republic. **2010**, 150.

(37) Vacek, O.; Vašát, R.; Borůvka, L. Quantifying the Pedodiversity-Elevation Relations. *Geoderma* **2020**, *373*, 114441. https://doi.org/10.1016/j.geoderma.2020.114441.

(38) Krivoruchko, K. *Empirical Bayesian Kriging*; 2012; Vol. Fall 2012.

(39) Vapnik, V. The Nature of Statistical Learning Theory. *Technometrics* **1995**, *38* (4), 409. https://doi.org/10.2307/1271324.

(40) Li, Z.; Zhou, M.; Xu, L. J.; Lin, H.; Pu, H. Training Sparse SVM on the Core Sets of Fitting-Planes. *Neurocomputing* **2014**, *130*, 20–27. https://doi.org/10.1016/j.neucom.2013.04.046.

(41) Cherkassky, V.; Mulier, F. *Learning from Data: Concepts, Theory, and Methods: Second Edition*; 2006. https://doi.org/10.1002/9780470140529.

(42) John, K.; Isong, I. A.; Kebonye, N. M.; Ayito, E. O.; Agyeman, P. C.; Afu, S. M. Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. *Land* **2020**, *9* (12), 1–20. https://doi.org/10.3390/land9120487.

(43) Vohland, M.; Besold, J.; Hill, J.; Fründ, H. C. Comparing Different Multivariate Calibration Methods for the Determination of Soil Organic Carbon Pools with Visible to near Infrared Spectroscopy. *Geoderma* **2011**, *166* (1), 198–205. https://doi.org/10.1016/j.geoderma.2011.08.001.

(44) Fraser, S. J.; Dickson, B. L. *A New Method for Data Integration and Integrated Data Interpretation: Self-Organising Maps*; 2007.

(45) Li, T.; Sun, G.; Yang, C.; Liang, K.; Ma, S.; Huang, L. Using Self-Organizing Map for Coastal Water Quality Classification: Towards a Better Understanding of Patterns and Processes. *Sci. Total Environ.* **2018**, *628–629*, 1446–1459. https://doi.org/10.1016/j.scitotenv.2018.02.163.

(46) Melssen, W. J.; Smits, J. R. M.; Buydens, L. M. C.; Kateman, G. Using Artificial Neural Networks for Solving Chemical Problems. Part II. Kohonen Self-Organising Feature Maps and Hopfield Networks. *Chemometrics and Intelligent Laboratory Systems*. Elsevier May 1, 1994, pp 267–291. https://doi.org/10.1016/0169-7439(93)E0036-4.

(47) Kooistra, L.; Wanders, J.; Epema, G. F.; Leuven, R. S. E. W.; Wehrens, R.; Buydens, L. M. C. The Potential of Field Spectroscopy for the Assessment of Sediment Properties in River Floodplains. *Anal. Chim. Acta* **2003**, *484* (2), 189–200. https://doi.org/10.1016/S0003-2670(03)00331-3.

(48) Li, L.; Lu, J.; Wang, S.; Ma, Y.; Wei, Q.; Li, X.; Cong, R.; Ren, T. Methods for Estimating Leaf Nitrogen Concentration of Winter Oilseed Rape (Brassica Napus L.) Using in Situ Leaf Spectroscopy. *Ind. Crops Prod.* **2016**, *91*, 194–204. https://doi.org/10.1016/j.indcrop.2016.07.008.

(49) Kabata-Pendias. Kabata-Pendias A. 2011. Trace elements in soils and... - Google Scholar https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Kabata-Pendias+A.+2011.+Trace+elements+in+soils+and+plants.+4th+ed.+New+York+%28NY%29%3A+CRC+Press&btnG= (accessed Nov 24, 2020).

(50) Różański, S. Ł.; Kwasowski, W.; Castejón, J. M. P.; Hardy, A. Heavy Metal Content and Mobility in Urban Soils of Public Playgrounds and Sport Facility Areas, Poland. *Chemosphere* **2018**, *212*, 456–466. https://doi.org/10.1016/j.chemosphere.2018.08.109.

(51) Bretzel, F.; Calderisi, M. Metal Contamination in Urban Soils of Coastal Tuscany (Italy). *Environ. Monit. Assess.* **2006**, *118* (1–3), 319–335. https://doi.org/10.1007/s10661-006-1495-5.

(52) Jim, C. Y. Urban Soil Characteristics and Limitations for Landscape Planting in Hong Kong. *Landsc. Urban Plan.* **1998**, *40* (4), 235–249. https://doi.org/10.1016/S0169-2046(97)00117-5.

(53) Birke, M.; Rauch, U.; Chmieleski, J. Environmental Geochemical Survey of the City of Stassfurt: An Old Mining and Industrial Urban Area in Sachsen-Anhalt, Germany. In *Mapping the Chemical Environment of Urban Areas*; John Wiley and Sons, 2011; pp 269–306. https://doi.org/10.1002/9780470670071.ch18.

(54) Khodadoust, A. P.; Reddy, K. R.; Maturi, K. Removal of Nickel and Phenanthrene from Kaolin Soil Using Different Extractants. *Environ. Eng. Sci.* **2004**, *21* (6), 691–704. https://doi.org/10.1089/ees.2004.21.691.

(55) Jakovljevic, M.; Kostic, N.; Antic-Mladenovic, S. *The Availability of Base Elements (Ca, Mg, Na, K) in Some Important Soil Types in Serbia*; 2003. https://doi.org/10.2298/zmspn0304011j.

(56) Orzechowski, M.; Smolczynski, S. *IN SOILS DEVELOPED FROM THE HOLOCENE DEPOSITS IN*

*NORTH-EASTERN POLAND**; -, 2007; Vol. 15.

(57) Pongrac, P.; McNicol, J. W.; Lilly, A.; Thompson, J. A.; Wright, G.; Hillier, S.; White, P. J. Mineral Element Composition of Cabbage as Affected by Soil Type and Phosphorus and Zinc Fertilisation. *Plant Soil* **2019**, *434* (1–2), 151–165. https://doi.org/10.1007/s11104-018-3628-3.

(58) Kingston, G.; Anink, M. C.; Clift, B. M.; Beattie, R. N. *Potassium Management for Sugarcane on Base Saturated Soils in Northern New South Wales*; 2009; Vol. 31.

(59) Santo, L. T., Nakahata, M. H., & Schell, V. P. Santo LT, Nakahata MH, Ito GP and Schell VP (2000).... - Google Scholar https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Santo+LT%2C+Nakahata+MH%2C+Ito+GP+and+Schell+VP+%282000%29.+Calcium+and+liming+trials+from+1994+to+1998+at+HC%26S.+Technical+supplement+to+Agronomy+Report+83%2C+Hawaiian+Agricultural+Research+Centre. (accessed May 16, 2021).

(60) Burgos, P.; Madejón, E.; Pérez-de-Mora, A.; Cabrera, F. Horizontal and Vertical Variability of Soil Properties in a Trace Element Contaminated Area. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10* (1), 11–25. https://doi.org/10.1016/j.jag.2007.04.001.

(61) Olinic, T.; Olinic, E. The Effect of Quicklime Stabilization on Soil Properties. *Agric. Agric. Sci. Procedia* **2016**, *10*, 444–451. https://doi.org/10.1016/j.aaspro.2016.09.013.

(62) Madaras, M.; Lipavský, J. *Interannual Dynamics of Available Potassium in a Long-Term Fertilization Experiment*; 2009; Vol. 55. https://doi.org/10.17221/34/2009-pse.

(63) Madaras, M.; Koubova, M.; Lipavský, J. Stabilization of Available Potassium across Soil and Climatic Conditions of the Czech Republic. *Arch. Agron. Soil Sci.* **2010**, *56* (4), 433–449. https://doi.org/10.1080/03650341003605750.

(64) Pulkrabová, J.; Černý, J.; Száková, J.; Švarcová, A.; Gramblička, T.; Hajšlová, J.; Balík, J.; Tlustoš, P. Is the Long-Term Application of Sewage Sludge Turning Soil into a Sink for Organic Pollutants?: Evidence from Field Studies in the Czech Republic. *J. Soils Sediments* **2019**, *19* (5), 2445–2458. https://doi.org/10.1007/s11368-019-02265-y.

(65) Asare, M. O.; Horák, J.; Šmejda, L.; Janovský, M.; Hejcman, M. A Medieval Hillfort as an Island of Extraordinary Fertile Archaeological Dark Earth Soil in the Czech Republic. *Eur. J. Soil Sci.* **2021**, *72* (1), 98–113. https://doi.org/10.1111/ejss.12965.

(66) Zádorová, T.; Penížek, V.; Šefrna, L.; Drábek, O.; Mihaljevič, M.; Volf, Š.; Chuman, T. Identification of Neolithic to Modern Erosion-Sedimentation Phases Using Geochemical Approach in a Loess Covered Sub-Catchment of South Moravia, Czech Republic. *Geoderma* **2013**, *195–196*, 56–69. https://doi.org/10.1016/j.geoderma.2012.11.012.

(67) Tlustoš, P.; Hejcman, M.; Kunzová, E.; Hlisnikovský, L.; Zámečníková, H.; Száková, J. Nutrient Status of Soil and Winter Wheat (Triticum Aestivum L.) in Response to Long-Term Farmyard Manure Application under Different Climatic and Soil Physicochemical Conditions in the Czech Republic. *Arch. Agron. Soil Sci.* **2018**, *64* (1), 70–83. https://doi.org/10.1080/03650340.2017.1331297.