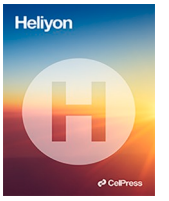
Heliyon 9 (2023) e18511

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



A coupled geostatistical and machine learning approach to address spatial prediction of trace metals and pollution indices in sediments of the abandoned gold mining site of Bekao, Adamawa, Cameroon

Reynolds Yvan Abende Sayom [a,*], Martin Luther Mfenjou [a], Mouhamed Ayiwouo Ngounouno [a], Michele Maguy Cathya Etoundi [a], William André Boroh [a], Luc Leroy Mambou Ngueyep [a,b], Arsene Meying [a]

[a] School of Geology and Mining Engineering, University of Ngaoundere, P.O. Box 115, Meiganga, Cameroon
[b] Laboratory of Mechanics and Materials of Civil Engineering (L2MGC), CY Cergy Paris University, 5 Mail Gay Lussac, Neuville sur Oise, F-95031, Cergy-Pontoise Cedex, France

ARTICLE INFO

Keywords:
Abandoned gold mining site
Sediments
Spatial variability
Trace metals
Pollution indices
Kriging
Artificial neural network
Prediction

ABSTRACT

Trace metals present in high amounts in aquatic systems are a perpetual concern. This study applied geostatistical and machine learning models namely Ordinary Kriging (OK), Ordinary Cokriging (OCK) and Artificial Neural Network (ANN) to assess the spatial variability of trace metals and pollution indices in surface sediments along the Lom River in an abandoned gold mining site at Bekao (Adamawa Cameroon). For this purpose, thirty-one (31) surface sediment samples are collected in order to determine the total concentrations of As, Cr, Cu, Fe, Mn, Ni, Pb, Sn and Zn. These trace metals are used to compute pollution indices as the sediment pollution index (SPI), the Nemerow index (NI), the modified contamination degree (mCD), and the potential ecological risk assessment (RI). OK, OCK and ANN models are compared to determine the best model performance. The best models are selected based on the values of the root mean square error (RMSE), the coefficient of determination (R^2), the scatter index (SI) and the BIAS. Results showed that the sequence of trace metal mean concentrations in the sediments is Fe > Mn > Cu > Ni > Sn > Cr > Zn > Pb > As. The mean concentrations of Ni, Cu, Zn and Sn are above the average shale values (ASV) and the pollution status is globally moderate to significant with a low potential ecological risk. The spatial dependency obtained with semivariogram models are moderate to weak for Mn, Fe, Ni, Pb, SPI, NI, mCD, RI As, Cr, and Sn and strong for Cu and Zn. According to cross-validation parameters, ANN model is the best method for the prediction on trace metal concentrations and pollution indices in surface sediments along the Lom River in the abandoned gold mining site of Bekao.

1. Introduction

Trace metal pollution affects a variety of environmental elements, including soil, sediment, water and air, and has been linked to

* Corresponding author.
E-mail address: rabendesayom@gmail.com (R.Y. Abende Sayom).

https://doi.org/10.1016/j.heliyon.2023.e18511
Received 6 April 2023; Received in revised form 7 July 2023; Accepted 19 July 2023
Available online 20 July 2023

the mining of metallic minerals. As part of the mining process, the entire soil mass is excavated and exposed to weathering, degradation, and transport agents, which causes soil erosion and seriously contaminates the area around it [1]. Mining activity has started to have a more significant negative impact on the environment due to the constant rise in demand for metals.

Sediments are a crucial indicator of metal contamination when evaluating the health of ecosystems. Heavy metals frequently find their way into bottom sediments in aquatic environments. Sediments are a crucial indicator of metal contamination when evaluating the health of ecosystems. Heavy metals frequently find their way into bottom sediments in aquatic environments [2,3]. When conditions become favorable, these metals can by various processes be released in water and aquatic organisms could be exposed to high doses of heavy metals. Benthic organisms, in particular, are in direct contact with sediments. Trace metals can accumulate in their tissues and therefore can enter the human food chain which can cause potential risks to human health [4–6].

The Lom River is well known for the numerous artisanal and small scale gold mining operations that take place there. These operations transport significant amounts of sediments, frequently rich in trace metals, which may have an adverse effect on the rich aquatic ecosystem and fauna of the two regions. In recent years, due to a continuous increase in the demand for metals, the Adamawa region has experienced a strong expansion of alluvial gold mining along the Lom River and its tributaries, generating sometimes irreversible damage to water resources. In this context of pollution, it is relevant to evaluate the distribution of trace elements, their bioavailability and toxicity in this area.

In view of the multiple health and environmental challenges generated by artisanal and semi-mechanized exploitation, mastering the spatial variation of trace metal contents in sediment is primordial for issues of protection and remediation for sustainable planning and management. Due to spatial heterogeneity, monitoring trace metals distributions involving large areas is one of the most difficult challenges. Laboratory analysis limits the spatial visualization of trace metals' distribution at sampled points, thus limiting the accuracy of information on mobility, origin, extent, association, and biological availability along the stream [7]. In addition, the use of Environmental quality indices have been recognized as a powerfool tool for contamination assessment, helpful for determining whether pollutants present in the sediment may have an impact on biological processes and for comparing the quantity of contaminants in the sediment with the related quality standard [8].

During recent years, much research has been carried out on geostatistics to analyze the spatial variability of trace metals content in soils, water and sediments. Numerous geostatistical methods as simple kriging, universal kriging, empirical Bayesian kriging (EBK), regression kriging (RK) are widely used to predict the spatial distributions of trace metals in soils and sediments in soil, sediments and environmental sciences [9–13]. Ordinary Kriging (OK) in particular is used to determine the values of the studied parameters at non-sampled locations as well as to estimate the spatial variability of physicochemical parameters or trace metals in soils and sediments [10,13–17]. Although being one of the most used geostatistical methods to predict the physicochemical parameters and the spatial distribution of trace metals [16,18]. OK tends to smooth estimates and limit their realism. It is therefore interesting to use Ordinary Cokriging (OCK) which makes it possible to take into account other variables related to the parameter studied which could refine the estimate by using the information they provide [19].

Nowadays, due to the rise and expansion of artificial intelligence, several machine learning models have shown their effectiveness for prediction. Machine learning methods represent a large class of nonlinear data-driven algorithms that are used in all fields of science, including soil science [20]. These methods are widely used to assess the spatial variability in sediment and soil properties. Several studies have used Artificial Intelligence methods to assess soil, sediment or water quality parameters [21–25]. Multivariable adaptive regression spline (MARS) and least square-support vector machine (LS-SVM) were used to predict indices of the five-day biochemical demand (BOD5) and chemical oxygen demand (COD) in water [26]. Support vector Regression (SVR) was used to predict cadmium and chromium content in soils and to detect lead in soils [27–29] and a novel Multiple-Kernel Support Vector Regression was proposed to estimate the hard-to-measure water quality parameters [30]. Random Forest (RF) was used to assess the spatial distribution of arsenic in a contaminated site, identify sources of contamination and predict spatial variations of heavy metals in soils and some properties of soils [28,31,32]. Other studies have used artificial neural network (ANN) to assess soil, sediment contamination [21,22,33–35]. ANN is a type of artificial intelligence (computer system) that attempts to mimic the way the human brain processes and stores information. The ANNs are considered as standard nonlinear estimators, and their abilities have been verified in a variety of fields [36]. The main difference between ANN to other artificial intelligence methods is due to the algorithm used for reducing the generalization error. The advantages of artificial neural network over traditional statistical techniques as explained by Peng and Wen [37] include: (1) Neural network is more accurate than statistical techniques especially when dealing with incomplete data records; (2) As the neural network can develop its own weighting scheme, so it is faster than other statistical techniques and (3) There is no need for prior knowledge during ANN's modeling, so it is a more flexible and powerful tool, while besides the main redeeming features of artificial neural network.

To the best of our knowledge, in Cameroon, machine learning methods have not been used for the prediction of sediments quality in mining areas. The objectives of this paper were to: (a) assess the sediment contamination by trace metals and pollution indices in the Lom River located in Bekao gold mining sites (Adamawa Cameroon); (b) predict the sediments contamination (trace metals and pollution indices) using ordinary kriging (OK), ordinary co-kriging (OCK) and artificial neural network (ANN) and (c) compare the performance of these three models, accordingly.

2. Materials and methods

2.1. Study area

The study area is located at Bekao in the Mbéré Department and Adamawa region of Cameroon. It is an abandoned artisanal and

semi-mechanized gold mining site in the western part of the village of Bekao. The Lom, Kaya, Yoyo, Foum, and Goumba rivers are present in the hydrographic network, which is not particularly significant. The study area is characterized by a relief composed of a plateau carved out by U-shaped valleys and V-shaped valleys and an equatorial transition climate characterized by two seasons: 9 months of rainy season from March to November and 3 months dry season from December to February. The abandoned semi-mechanized mining is drained by the Lom River. The geographical location of the study area is presented in Fig. 1.

2.2. Sampling and analytical techniques

Thirty-one (31) surface sediment samples from S1 to S31 were collected during the month of June 2022 along the Lom River in the abandoned artisanal and semi-mechanized gold mining site of Bekao. The sampling map of sediments is presented in Fig. 2. For reasons of accessibility and representativeness, the samples were taken following a step of 100 m. The samples were taken upstream (S1 and S2), inside, (S3 to S29) and downstream (S30 and S31) of the abandoned gold mining area.

The collected sediment samples were placed in plastic bags to avoid contamination, then stored and transported to the laboratory at 4 °C for further analysis. Before assessing the trace metals in the sediments, the preliminary step involved pressing sediment samples to obtain sediments in granular form. The pellet was irradiated and its fluorescence was measured using Skyray Instrument EDX Pocket III brand with detection limit of 0.001–0.01%. The trace metals determined in sediment samples were As, Cr, Cu, Fe, Mn, Ni, Pb, Sn and Zn. They were analyzed at laboratory of the Framework of Support for the Promotion of Mining Handicrafts '(CAPAM)' in Yaounde (Cameroon).

To evaluate the precision and accuracy of measurement, the Marlborough District Council (MDC) instructions for analysis of soil and sediment samples using a portable X-ray Fluorescence Spectrometry (XRF) [38] was employed. The portable XRF equipment is calibrated at the factory, so generally the user is not required to do so. However, it is essential to verify the calibration of the device using a certified reference material (CRM). Before the investigation began, the certified reference material (XRF Scientific 1201080) containing the investigated elements had been analyzed for this purpose. For Mn, Fe, Cu, Sn, Pb, and As, the percentage difference (% D) between the measured values and the certified values of the certified reference material was within 8%, and for Cr, Ni, and Zn, it was between 9 and 11%. This indicates that the analytical instrument is considered accurate (%D $\leq$ 20%). Three analyses of each sediment sample were conducted and the results were presented as mean. For all the studied elements, the relative standard deviations (RSDs) of the three measurements were under 5%.

2.3. Trace metals contamination investigations

Four tools were used to assess the state of metallic pollution of the sediments. These tools were the sediment pollution index (SPI),

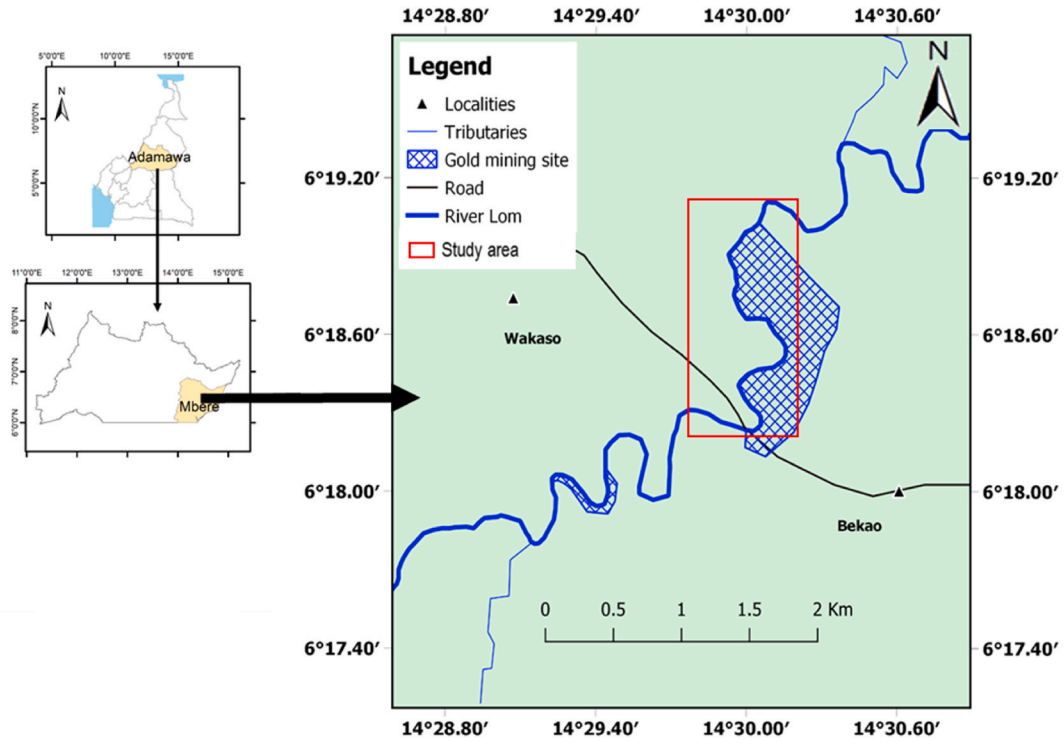


Fig. 1. Location map of the study area.

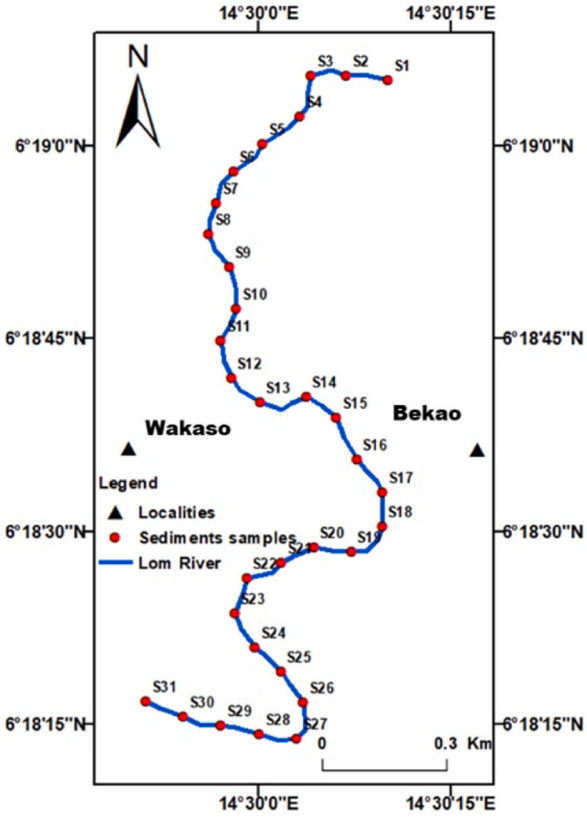


Fig. 2. Sediment sampling map.

the Nemerow index (NI), the modified contamination degree (mCD) and the potential ecological risk assessment (RI).

2.3.1. Sediment pollution index (SPI)

The SPI is a multi-metal technique established for an overall assessment of sediment quality in terms of heavy metal concentrations, taking into account metal toxicity [39]. The calculated formula is given at Eq. (1):

Table 1
Classification of pollution indices.

Index	Range quality	Range quality	References
SPI	<2	Natural	[39]
	2–5	Lowly polluted	
	5–10	Moderately polluted	
	10–20	Highy polluted	
	>20	Dangerous	
NI	≤1	Not polluted	[40]
	1–2	Slightly polluted	
	2–3	Moderately polluted	
	>3	Heavily polluted	
mCD	<1.5	Very low	[41]
	1.5–2	Low	
	2–4	Moderate	
	4–8	High	
	8–16	Very high	
	16–32	Extremely high	
	>32	Ultra high	
RI	≤150	Low risk	[42,43]
	150–300	Moderate risk	
	300–600	Considerable risk	
	≥600	High risk	

$$SPI = \frac{\sum(Cf_I * W_i)}{\sum W_i}, \tag{1}$$

where W_i is the toxicity weight of metal *i*. Toxicity weight 1 was assigned for Cr, Mn, and Zn; 2 was assigned for Ni and Cu, 5 for Pb [39]. The five categories of SPI are presented in Table 1.

2.3.2. Nemerow index (NI)

The NI was used to estimate the overall metal quality of sediments [44] and it is calculated by the formula given at Eq. (2)

$$NI = \sqrt{\frac{(Cf)^2_{max} + (Cf)^2_{mean}}{2}}, \tag{2}$$

where Cf_{max} and Cf_{mean} are respectively, are the maximum and average values, respectively, of the single-element pollution index. The NI can be classified into four categories (see Table 1).

2.3.3. Modified contamination degree (mCD)

The mCd developed by Abrahin and Parker [45] is a relatively index used to quantify overall accumulation of a group of metals in sediment the mCD is computed by Eq. (3) and Eq. (4).

$$mCD = \frac{\sum_{i=1}^{i=n} Cf_i}{n}, \tag{3}$$

$$Cf_i = \frac{C_n}{B_n}, \tag{4}$$

where C_n is the concentration of a given metal in sediments, B_n refers to thebackground value of heavy metals given by Turekian and Wedepohl [46] and n the number of analyzed metals. On the basis of mCD, Sediments contamination can be classified into seven categories [41] which are presented in Table 1.

2.3.4. Potential ecological risk assessment (RI)

The ecological hazards of the studied trace metals were determined by determining RI. Because large concentrations of trace metals pose major concerns for the overall environment, determining the possible RI of the samples is critical to determining the level of ecological risk. Eq. (5) was used to calculate the RI.

$$RI = \sum ERI = \sum T_i * Cf_i, \tag{5}$$

where T_i is the toxic-response factor (also known as biological toxicity factor) of each trace metal. The T_i values of the elements are given by Guo et al. [47]: As = 10; Ni, Cu, and Pb = 5; Cr = 2; Mn and Zn = 1). Four classes of RI are obtainable (see Table 1).

2.4. Exploratory data analysis

This involves determining the essential spatial features to describe the data. Exploratory data analysis (EDA) is the final test of data quality. It includes statistical analysis comparisons of the data by analyzed elements, calculated indices and their spatial distributions necessary to find the optimal approaches for integrating different data sets into a single, coherent set. The descriptive statistics of the concentrations of the nine (09) trace metals, and pollution indices in the sediments considered in this study consist mainly in the determination of the minimum, the maximum, the mean, the standard deviation (SD) the coefficient of variation (CV), the skewness and kurtosis.

The multivariate statistical analysis (MSA) was applied through the principal component analysis (PCA) to the concentration of the trace studied metals and pollution indices to determine the associations between them among the groups and to investigate the similarity between all the parameters. Statistical analyses were performed using XLSTAT 2014 and STATISTICA softwares.

2.5. Geostatistical modeling of trace metals and pollution indices

The spatial prediction and distribution of trace metals and pollution indices in sediments with geostatistical modeling were performed using kriging approaches. Kriging is a spatial interpolation method, considered the most accurate from a statistical point of view, which allows a linear estimation based on statistical quantities of the spatialized data [19]. In this study, values at unsampled sites are predicted using kriging and cokriging. Their general formulae are relatively given by Eqs. (6) and (7).

$$\widehat{Z}(S_0) = \sum_{i \forall (S_0)} \lambda_{i,} Z(S_i), \tag{6}$$

$$\widehat{Z}(S_0) = \alpha + \sum_{i \forall (S_0)} \lambda_{i,0}, Z(S_i) + \sum_{j=1}^{q} \sum_{i \in (S_0)} \lambda_{i,j} \omega_j \left(S_{i,j}\right), \tag{7}$$

where $\widehat{Z}$ is the predicted value at an unsampled location x_0; Z is a measured value at the sampled location x_i; n is the number of considered samples within the interpolation neighborhood, and $\lambda_{i,}$ is the weight assigned to the measured sample i for predicting $\widehat{Z}$.

Kriging differs only in the choice of weights λ_i. The weight assigned to each data item is determined by analyzing the structure of the spatial distribution of the data. The spatial distributions of trace metals and pollution indices was assessed by calculating simples and crossed semivariograms (Eq. (8) and Eq. (9)), that represents the average variance between the observations separated by a certain distance, and describes the structure of the spatial variability of the investigated variable. Once the experimental variogram is calculated, it will be fitted to a predefined model depending on the characteristics of the data.

$$\gamma(h) = \left(\frac{1}{2 * |N(h)|}\right) * \sum_{N(h)} [z(x_i + h) - z(x_i)]^2, \tag{8}$$

$$\gamma(h) = \left(\frac{1}{2 * |N(h)|}\right) * \sum_{N(h)} [z(x_i + h) - z(x_i)].[y(x_i + h) - y(x_i)], \tag{9}$$

where $\gamma(h)$ is the experimental semivariogram, $N(h)$ is the number of pairs of samplings sites separated by h, $z(x_i)$ and $z(x_i + \mathrm{h})$ are the numerical values of the observed variable at points x_i, and $x_i + h$, $y(x_i)$ and $y(x_i + h)$ the numerical values of the covariate at points x_i, and $x_i + h$.

The predicted maps of studied trace metals and pollution indices were generated using Ordinary Kriging (OK) and Ordinary Cokriging (OCK) through geostatistical analyst in ArcGIS 10.5. To determine the trace metals at the unsampled points, ordinary kriging utilizes a weighted average of known neighboring values, which is dependent on their proximity, grouping, and values. According to Ophélie and Nicolas [19], Cokriging allows to improve the estimations obtained by kriging by using the information provided by secondary variables. It uses multivariate analysis methods that focus on the observation and simultaneous processing of several statistical variables in order to obtain relevant synthetic information.

2.6. Prediction using artificial neural network (ANN)

ANN was used to evaluate the spatial variability of trace metal concentrations and pollution indices of sediments in the Lom River at Bekao. The nftools tool of Matlab 2019 software was used. ANN is an approach for modeling complex systems, which are particularly difficult to model using classical statistical methods. Its design is schematically inspired by the functioning of biological neurons in the human brain given the large capacity for storing information and the inter-neuron connection called synaptic weight to store knowledge [48,49]. An ANN model is produced by connecting numerous neurons in a predetermined configuration. The three primary elements characterizing the neural network are the distributed representation of information, local operations, and non-linear processing. The learning process, also known as training, creates the connections between neurons and is carried out by providing the ANN with known inputs and outputs in an organized fashion. In order to produce the desired output for a known input pattern, the strength of these connections is modified using an error convergence method [50].

Neural networks are organized in layers. These consist of a number of interconnected neurons that contain an activation function. Inputs (X_1, X_i, …, X_n) are presented to the network through the input layer, which communicates them to the hidden layers where processing takes place using weighted connections. Then, the hidden layers transmit the response to the output layer (E). The connections between the neurons are made by weights (W_1, W_i, …, W_n). MultiLayers Perceptron were adopted in the development of the neural network. This network comprises one or more layers of artificial neurons to capture the inputs, one or more hidden layers (MLP or MultiLayers Perceptron) and a layer of artificial neurons to emit the outputs of the model. Each layer contains calculation units (neurons) connected to other neurons via weights (W_{ij} and W_{jk}) [49,51,52]. The activation function also called transfer function is interested in the relationship between the inputs and the outputs of the nodes of the network. The most common choices for the activation function are binary functions (threshold activation functions), linear functions, identity function, hyperbolic or sigmoid functions. The inputs Xi are presented to the input layer of neurons then they are propagated to the hidden layer and finally to the output layer according to the equation Eq. (10):

$$E_j = f\left(\sum_{i=1}^{n} X_i W_{ij} + b_j\right), \tag{10}$$

where E_j is the sum of the weights between the inputs of the jth neuron of the hidden layer; X_i the output value of the i th neuron of previous layer; W_{ij} the synaptic weight of neuron i of the input layer to neuron j of the hidden layer and b_j is the BIAS or the activation threshold of neuron j.

For this study, the ANN architecture used consisted of a single-layer feed-forwarded network (simple perceptron) with ten neurons in the hidden layer. The training algorithm used was the Levenberg-Marquardt algorithm.

2.7. Cross validation

The crossvalidation technique is used in assessing how the results of the statistical analysis will generalize to an independent data set. Cross-validation technique was adopted for evaluating and comparing the performance of geostatistical interpolations (OK and OCK), and Artificial Neural Network (ANN) in predicting the studied trace metal concentrations and the four pollution indices in sediments. These performances were assessed using four criteria (1) the root mean square error (RMSE); (2) the coefficient of determination (R^2); (3) the scatter index (SI) and (4) the BIAS. The best models were selected based on the least RMSE, SI, and BIAS as well as the highest R^2 [25]. The R^2, RMSE, SI and BIAS were calculated using the equations Eqs. (11)–(14) respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(X_k - Y_k)^2}. \tag{11}$$

$$R^2 = 1 - \frac{\sum_{1}^{n}(X_k - Y_k)^2}{\sum_{1}^{n}(X_k - \overline{X})^2}. \tag{12}$$

with $\bar{y} = \frac{1}{N}\sum_{1}^{N} X_k$.

$$SI = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}((Y_k - \overline{Y_k}) - (X_k - \overline{X_k})}}{\frac{1}{N}\sum_{i=1}^{N} X_k}. \tag{13}$$

$$BIAS = \frac{\sum_{i=1}^{N}(Y_k - X_k)}{N}. \tag{14}$$

where X_k and Y_k are the measured concentration of the trace metal and predicted concentration of the trace metal by the selected model/formula, respectively and N the total number of data sets. The calculation of the SI involves dividing the RMSE by the average of the observations at each grid point. This produces a percentage representation of the deviation of the RMSE from the mean observation, effectively indicating the expected error percentage for the parameter.

3. Results and discussions

3.1. Data analysis and spatial distribution of trace metals and pollution indices

The descriptive statistics of studied trace metals and pollution indices in sediment samples are reported in Table 2 and their spatial variations are presented in Fig. 3.

3.1.1. Trace metals

Trace metal concentrations decrease in the following order: Fe > Mn > Cu > Ni > Sn > Cr > Zn > Pb > As. The means concentrations of trace metals in sediments were also compared with Average Shale Values (ASVs) reported by Turekian and Wedepohl [46] frequently utilized as background values in the assessment of sediment contamination. The Fe concentrations varied from 10,849.8 to 10,494 mg/kg. The average concentration of Fe (31,121.5 mg/kg) was below the ASV value and 9.64% of the samples were above the value set by the ASVs. The highest Fe concentration was found in the sampling point S15. Mn concentrations ranged from 0.00 to

Table 2
Descriptive statistics of trace metals and pollution indices.

Trace metals	Min	Max	Median	Mean	SD	CV (%)	Skewness	Kurtosis
As	0.0	37.2	0.0	5.2	10.4	197	1.8	1.9
Cr	0.0	109.7	70.8	65.9	23.1	34	−1.1	1.2
Cu	117.9	416.3	200.2	212.8	75.5	35	0.8	0.0
Fe	10849.9	104946.0	25585.1	31121.5	18473.6	58	2.5	7.0
Mn	0.0	676.7	356.9	331.5	181.6	54	−0.5	−0.2
Ni	0.0	387.1	133.3	142.0	101.1	70	0.3	−0.5
Pb	0.0	113.7	6.0	15.6	22.9	145	2.8	8.9
Sn	55.0	198.9	124.2	125.4	24.9	20	0.1	2.2
Zn	0.0	139.2	18.5	33.1	39.7	118	1.9	2.1
SPI	0.8	3.4	1.4	1.6	0.6	37	1.3	1.3
NI	6.6	23.8	14.9	15.0	3.0	19	0.1	2.3
mCD	1.9	5.7	3.3	3.4	0.7	19	1.1	3.1
RI	21.6	82.2	40.6	44.2	14.8	33	0.9	0.6

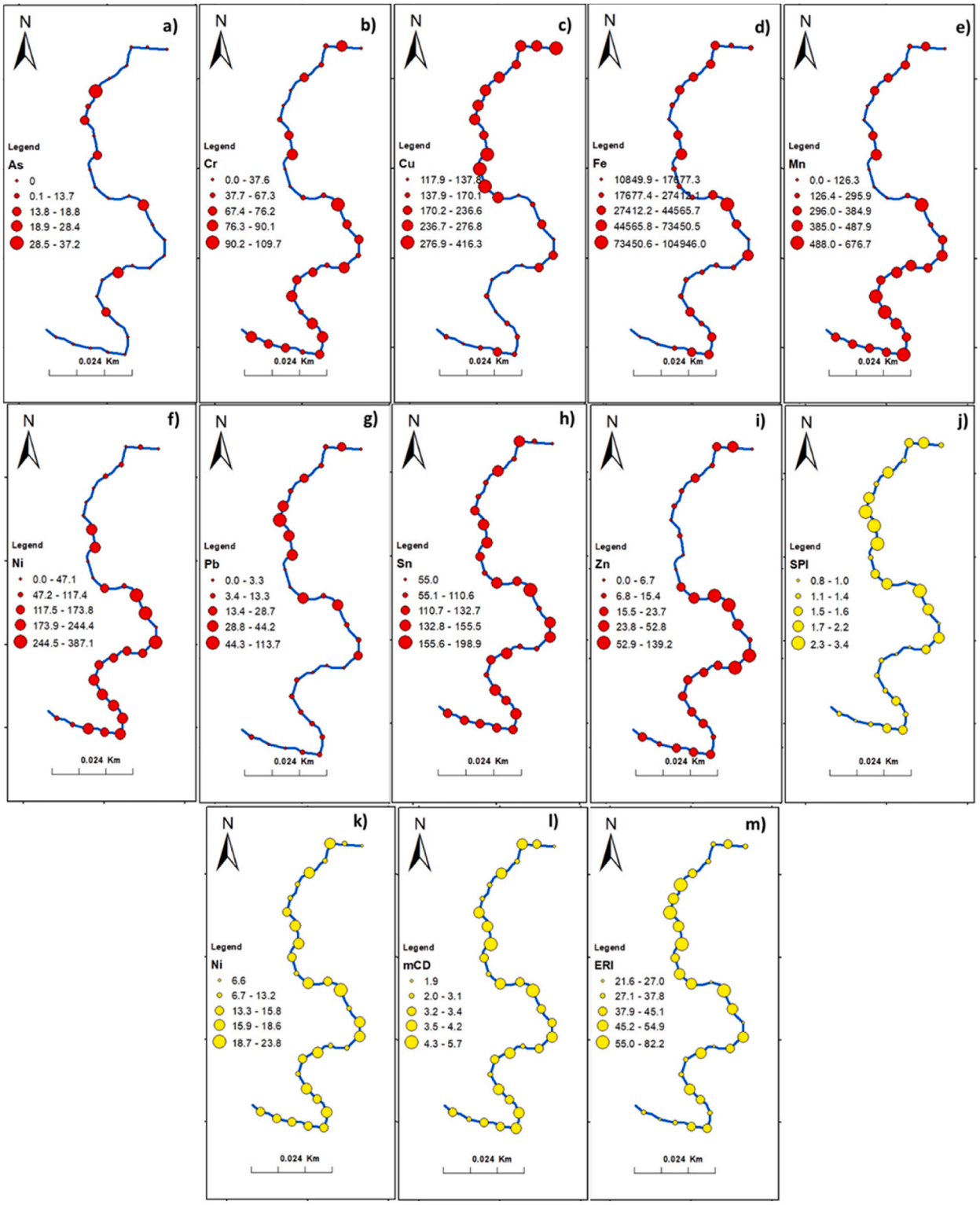


Fig. 3. Spatial distributions of trace metals and pollution indices: a) **As**; b) **Cr**; c) **Cu**; d) **Fe**; e) **Mn;** f) **Ni**; g) **Pb**; h) **Sn**; i) **Zn**; j) **SPI**; k) **NI**; l) **mCD** and m**) RI.**

676.7 mg/kg, with an average concentration of 331.5 mg/kg. All the Mn concentrations were below the allowable limit of 850 mg/L set by the ASVs. The lowest and the higher concentrations of Mn were respectively found in the sampling points S1 and S24. The concentrations of Cu and Sn ranged from 117.9 to 416.3 mg/kg and 55.0–198.9 mg/kg with average values of 212.8 and 212.8 mg/kg, respectively. S12 and S15 were the sampling points with highest value of Cu and Sn, respectively. As and Pb are dangerous metals.

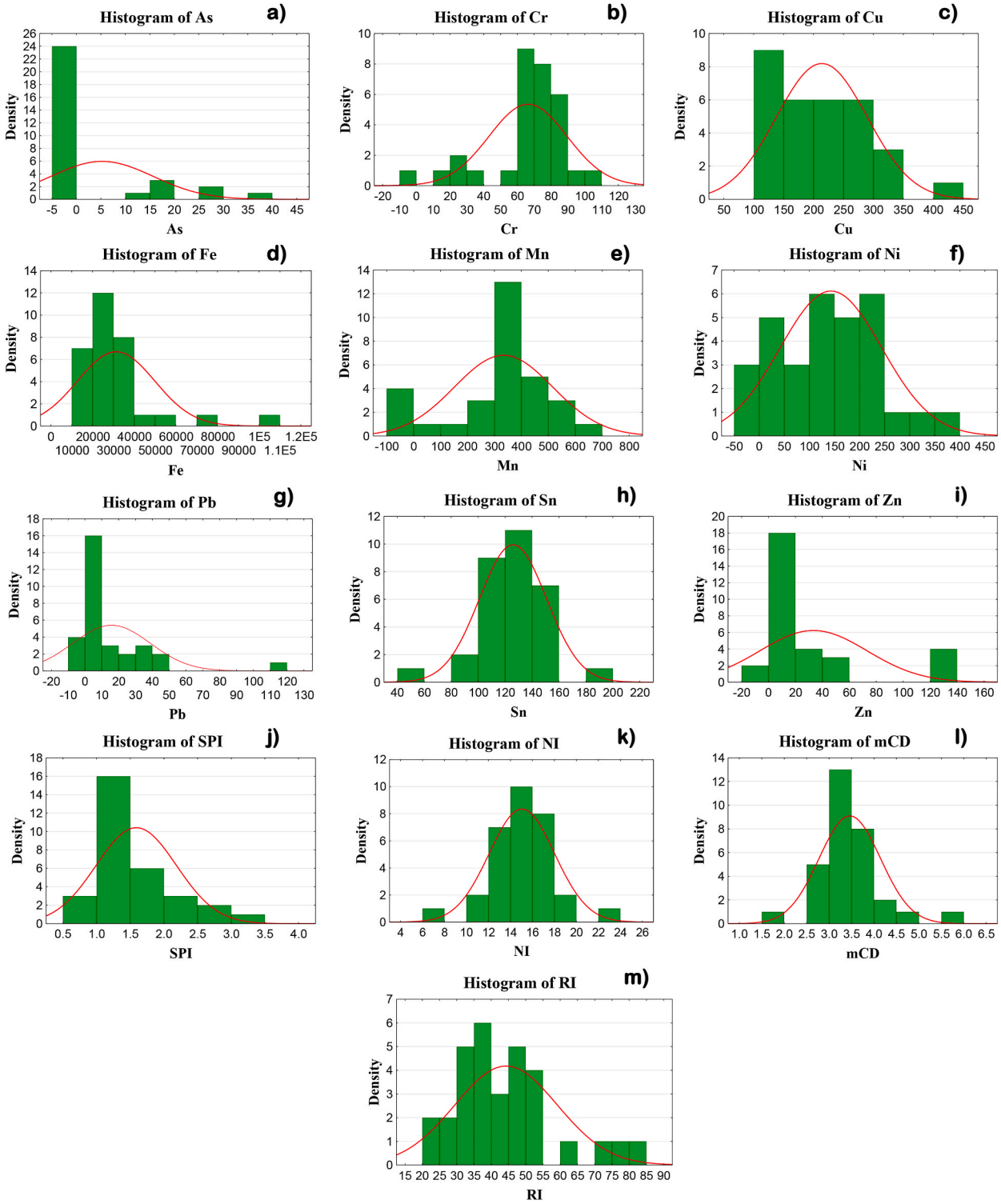


Fig. 4. Histograms of trace metals and polution indices: a) **As**; b) **Cr**; c) **Cu**; d) **Fe**; e) **Mn;** f) **Ni**; g) **Pb**; h) **Sn**; i) **Zn**; j) **SPI**; k) **NI**; l) **mCD** and m) **RI.**

Their concentrations in sediments ranged from 0 to 37.2 mg/kg and 0–113.7 mg/kg with mean values of 5.2 and 15.6 mg/kg, respectively. 22.58% and 25.80% of the sediment samples had As and Pb concentrations above the ASVs, respectively. S6 and S8 were the sampling points with higher values of As and Pb, respectively. The concentration of Zn and Ni were 0–139.2 and 0–387.1 mg/kg, with average of 33.1 and 142 mg/kg, respectively. Only 13% of samples had a concentration of Zn above the ASVs values versus 74.19% for the Ni.

The coefficient of variation (CV) was used to assess the variability of trace metals concentrations in surface sediments. Wilding [53] and Oku et al. [54] classified the variability based on the coefficient of variations (CVs) as least (lower 15%), moderate (15–35%) and mostly (superior 35%). The concentrations of, As, Pb, Zn Ni, Fe, Mn in the studied sediment samples had a great degree of variability, indicated by high coefficient of variation (CV) values of 197%, 145%, 118%, 70%, 58%, and 54%, respectively. However, Cu, Cr and Sn showed a moderate degree of variability with CV values of 35%, 34% and 20%, respectively. According to Tajudin et al. [16] and Jianshu et al. [55], low CV values for trace metals indicate natural input while very high values suggesting significant anthropogenic involvement.

Graphical methods of data analysis have the advantage of presenting the data in a comprehensive way. For this purpose, frequency histograms were plotted to appreciate the distribution of the data. Fig. 4 shows the distributions of the nine studied trace metals. The bell shapes highlighted by the frequency histograms, confirmed by the skewness ($\leq$0.5) of Cr, Sn Mn and NI indicated their approximate normal distributions [56] while the other elements present non-normal distributions. Except for NI, whose distribution showed an asymmetrical and bimodal fashion, all the other studied elements showed asymmetrical and unimodal distributions. In order to improve prediction performance, elements that did not have normal distributions have been normalized using a log-decimal transformation.

3.1.2. Pollution indices

SPI, NI, mCD and RI were computed to assess the overall level of pollution in sediments of the study area. Their respective statistics are presents in Table 3 and their spatial variation in Fig. 3. SPI, NI, mCD, and RI values ranged from 0.817 to 3.385, 6.617 to 23.792, 1.854 to 5.721 and 21.563 to 82.242, with average values of 1.583, 14.980, 3.447 and 44.152, respectively. The higher value of SPI were located at the sampling point S8. 80.60% of the samples were considered natural and 19.40% lowly polluted. For the NI, the higher value were found at the sampling point S15 and all the sample were considered heavily polluted. According to mCD, 3.22% of samples were judged lowly contaminated, 83.87% moderately contaminated and 12.9% highly contaminated. The highest mCD were observed at the sampling point S15. The calculation of RI showed that the highest value were observed at the sampling point S10 and all the samples had a low potential ecological risk. SPI, NI, mCD and RI showed respectively CV values of 37%, 19%, 19% and 33% indicating a relatively moderate variability.

All the computed indices demonstrated that the surface sediments of the Lom River at Bekao were moderately to significantly contaminate by trace metals. These results were similar to those observed in the sediments in the active gold mining site of Gankombol [57]. The highest values of the pollution indices and ecological risk assessment were observed at the sampling points inside the abandoned gold mining site, and the lowest values were observed in the samples taken upstream of the gold mining site.

3.1.3. Principal component analysis (PCA)

The relationship between all the studied trace metals and pollution indices was determined by using the PCA (Fig. 5). It is also used to determine the different covariates for OCK. Fig. 5 shows two factors (F1 and F2) explaining 69.24% of the total variance. Factor 1 represents 44.27% of this total variance and Factor 2 represents 24.97% of the total variance. According to their relationships, three main groups and an independent element were observed. The first group was composed of Pb, As, SPI, and RI, the second group was composed of Fe, Sn, NI and mCD. the third one was composed of Zn, Ni, Cr and Mn and the independent element was Cu. It is located alone in its quadrant as shown in Fig. 5.

Table 3
Characteristics of semivariogram models (ordinary kriging) of trace metals and pollution indices.

Trace metals	Nugget (C_0)	Partial Sill (C)	Range	Sill C + C_0	C_0/Sill	Spatial dependency
As	92.908	27.84	0.0097654	120.748	77%	Weakly
Cr	349.96	42.629	0.008482	392.589	89%	Weakly
Cu	633.43	7176.9	0.0082987	7810.33	8%	Strongly
Fe	2.64E+08	1.98E+08	0.0024019	461030000	57%	Moderately
Mn	15731	26739	0.0084159	42470	37%	Moderately
Ni	5278.5	7035.5	0.0084159	12314	43%	Moderately
Pb	271.09	421.4	0.0097094	692.49	39%	Moderately
Sn	444.86	92.193	0.0024019	537.053	83%	Weakly
Zn	338.88	2009.7	0.0082285	2348.58	14%	Strongly
SPI	0.1934	0.29185	0.0097094	0.48525	40%	Moderately
NI	5278.5	7035.5	0.0024019	12314	43%	Moderately
mCD	0.30998	0.27185	0.0024019	0.58183	53%	Moderately
RI	139.72	142.91	0.0097242	282.63	49%	Moderately

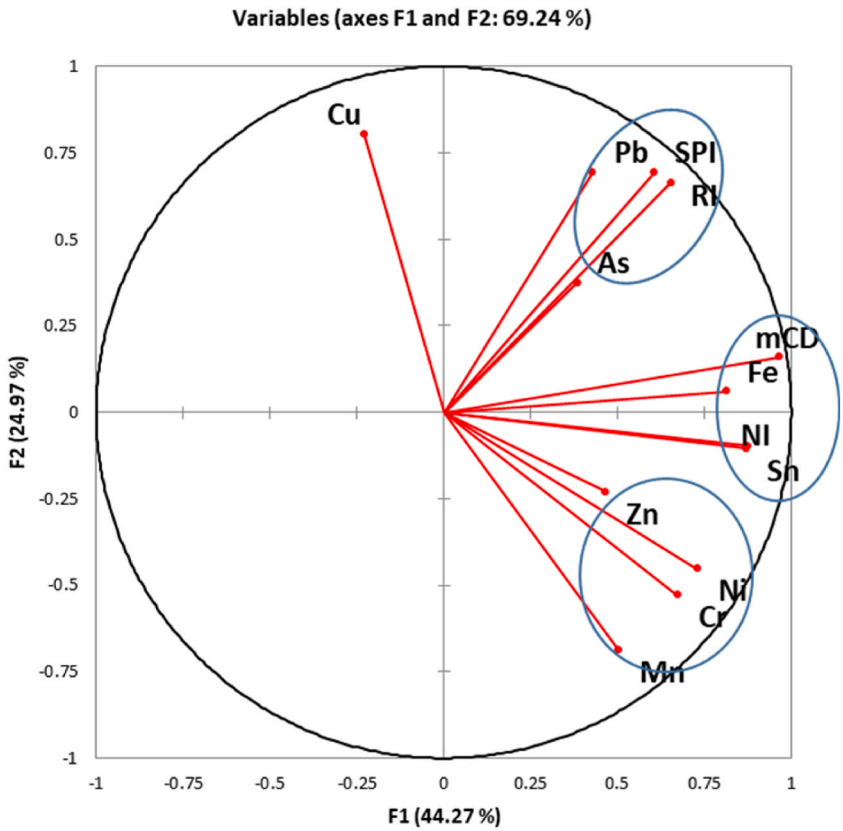


Fig. 5. Principal components analysis of trace metals and pollution indices.

3.2. Geostatistical modeling and prediction

3.2.1. Semivariogram analysis

Semivariograms are analytical geostatistical instruments that describe the spatial distribution of continuous and categorical data. Kriging takes advantage of the spatial correlation between observations to predict parameter values at unsampled locations using information related to one or more parameters. OK and OCK were used in this study. The experimental semivariograms and the modeling of the spatial dependency of the studied trace metals and pollution indices obtained with OK and OCK methods are presented in Figs. 6 and 7 and their characteristics are presented in Tables 4 and 5, respectively. These characteristics consist in the sill that represents the variance of the variable and nugget that could represent the measurement error or some variation at small scale. The nugget/sill ratio (N/S) ratio is a crucial indicator to reflect the extent of spatial autocorrelations of environmental factors, low ratio (<25%) implies a strong spatial dependence, moderate ratio (25–75%) a moderate spatial dependence and a high ratio (>75%) indicates weak spatial dependency. Spherical model defined by Di Virgilio et al. [58], as the most commonly used were used in this study to characterize spatial variability of trace metals and pollution indices.

3.2.1.1. Ordinary kriging (OK). According to Table 4, strong spatial dependency was observed for Cu and Zn, moderate spatial dependency was observed for Mn, Fe, Ni, Pb, SPI, NI, mCD, and RI, and finally weak spatial dependency for As, Cr, and Sn. A strong spatial correlation may indicate a natural factor contribution, a moderate spatial correlation may indicate both natural and anthropogenic contributions, and a weak spatial correlation may indicate a randomly spatial dependence and a high proportion of anthropogenic sources of contamination [55,59]. However, the notion of anthropogenic or natural contribution remains very difficult to define as far as heavy metals are concerned because of the complexity of the internal factors that govern their distribution (pH, organic matter, topography). In addition, in this region, the gold recovery process initially requires that excavations be created to reach the mineralized gravel; the gravel is then washed to extract the gold and stored, exposing the trace metal-bearing minerals to weathering agents, suggesting that these trace metals in sediments originate from the combination of anthropogenic and natural factors, such as the alteration of the various rocks in the study area. On the other hand, the numerous gold mining operations located along the Lom River, upstream of Bekao, contribute to the input of trace metals into the aquatic system, suggesting that the Lom River serves as a transporter for the pollution produced by these mining operations.

3.2.1.2. Ordinary Cokriging (OCK). Cokriging is a multivariate estimation method that allows the simultaneous estimation of several

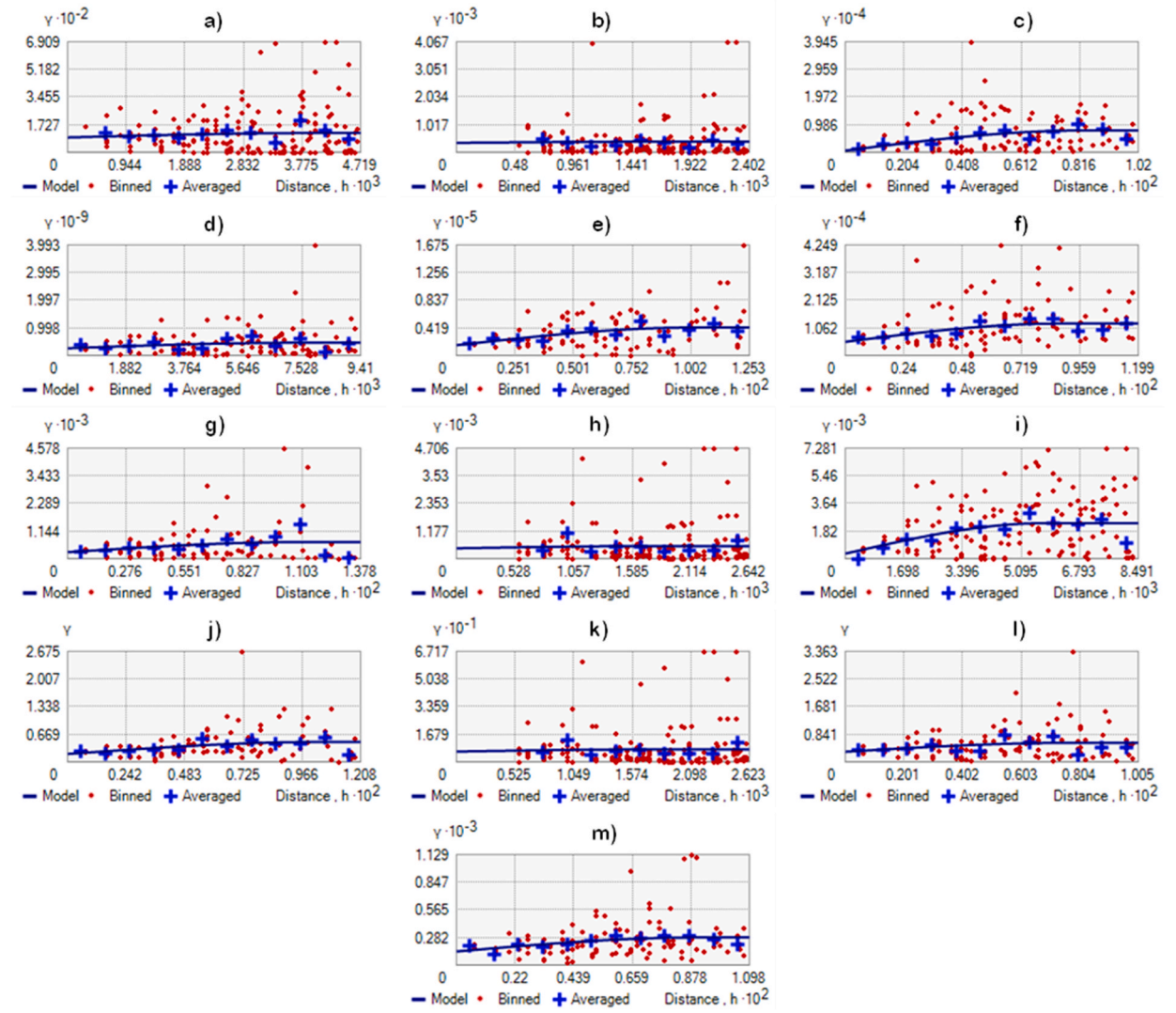


Fig. 6. Experimental semivariograms of trace metals and pollution indices: a) **As**; b) **Cr**; c) **Cu**; d) **Fe**; e) **Mn;** f) **Ni**; g) **Pb**; h) **Sn**; i) **Zn**; j) **SPI**; k) **NI**; l) **mCD** and m**) RI.**

variables by taking into account structural links between them or by introducing additional information through auxiliary variables. Thus in this study, the three groups highlighted by the PCA (Fig. 5) constituted groups of associated covariates that were used for the estimation of each element. Fig. 7 shows the experimental cross-semivariograms obtained for each trace metals and pollution indices. The characteristics of the crossed semivariogram for each variable are presented in Table 5. The results obtained were not very far from those of the simple semivariogram. Moderate spatial dependence was observed for Cr, Mn, Ni, Zn, Pb, SPI and RI. In another hand, Fe, As, Sn, NI and mCD were demonstrated the weakness of their spatial dependence. The spatial dependence obtained from semi-variogram analysis showed that Fe, Sn, NI and mCD, Mn, Zn, Ni, Pb, SPI and RI were respectively aligned with the PCA results confirming for each case similarities in their potential origins.

3.2.2. Predicted maps

In this study, the kriging methods (OK and OCK) were employed to generate the prediction maps of the 09 studied trace metals and the four pollution indices. Figs. 8 and 9, present the kriged maps obtained by OK and OCK, respectively.

In Figs. 8 and 9, As concentrations ranged from 1.0 to 13.3 mg/kg for OK and 1.2–12.3 mg/kg for OCK. Cr concentrations ranged from 40.2 to 82.4 mg/kg for OK and 37.3–85.2 mg/kg OCK. Cu concentrations ranged from 134.7 to 375.0 mg/kg for OK. Fe concentrations ranged from 20,137.6 to 50,780.4 mg/kg for OK and 19,842.9–49,394.5 mg/kg for OCK. Mn concentrations ranged from 142.2 to 507.0 mg/kg for OK and 90.4–516.6 mg/kg for OCK. Ni concentrations ranged from 46.4 to 226.6 mg/kg for OK and 16.1–260.1 mg/kg for OCK. Pb concentrations were between 2.8 and 52.1 mg/kg for OK and between 0.4 and 59.8 mg/kg for OCK. Sn concentrations were between 104.1 and 153.8 mg/kg and between 108.3 and 147.5 mg/kg. Zn concentrations were between 3.2 and

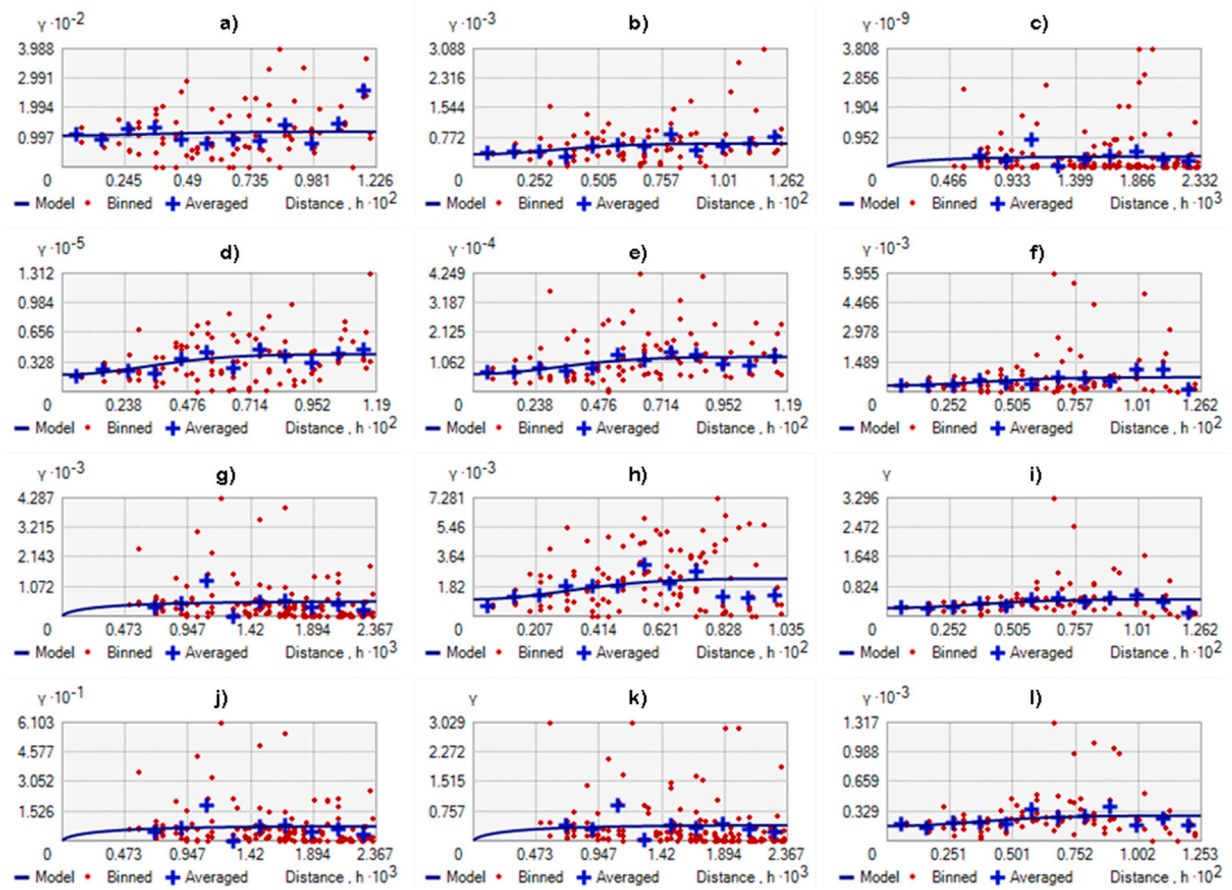


Fig. 7. Experimental crossed semivavariograms of trace metals and pollution indices: a) **As**; b) **Cr**; c) **Fe**; d) **Mn**; e) **Ni**; f) **Pb**; g) **Sn**; h) **Zn**; i) **SPI**; j) **NI**; k) **mCD** and l) **RI.**

Table 4
Characteristics of crossed semivariogram models (ordinary cokriging) of trace metals and pollution indices.

Trace metal	Nugget (C_0)	Partial Sill (C)	Range	Sill C + C_0	C_0/Sill	Spatial dependency
As	103.57	15.444	0.0097654	119.014	87%	Weakly
Cr	286.05	312.97	0.008482	599.02	48%	Moderately
Cu	–	–	–	–	–	–
Fe	3.21E+08	0	0.0024019	321200000	100%	Weakly
Mn	14682	25784	0.0084159	40466	36%	Moderately
Ni	5197.3	6999.7	0.0084159	12197	43%	Moderately
Pb	252.1	445.96	0.0097094	698.06	36%	Moderately
Sn	493.9	0	0.0024019	493.9	100%	Weakly
Zn	762.16	1532.4	0.0082285	2294.56	33%	Moderately
SPI	0.19294	0.2694	0.0097094	0.46234	42%	Moderately
NI	7202.5	0	0.0024019	7202.5	100%	Weakly
mCD	0.3715	0	0.0024019	0.3715	100%	Weakly
RI	146.55	129.82	0.0097242	276.37	53%	Moderately

118.4 mg/kg for OK and between 9.8 and 99.9 mg/kg for OCK. According to pollution indices, the range of variation with OK and OCK for SPI, NI, mCD and RI were respectively 1.1 to 2.3 and 1.1 to 2.5, 12.6 to 18.3 and 13.0 to 17.6, 2.8 to 4.1 and 3.0 to 4.1 and 32.5 to 59.4 and 32.34 to 64.12). The OK and OCK model predictions maps globally exhibit the same trends. For all the studied trace metals and pollution indices, the highest values were located inside the mining area. Except for Cr, Mn and Ni, the concentrations tend to increase from upstream to the mining area and then decrease downstream. As with metals, the pollution indices have their highest values in the mining area. All of this highlights the role played by gold mining in contaminating sediments, which facilitates the leaching of mineralized rocks through excavations.

Table 5
Cross-validation statistics.

Models	parameters	Cr	Mn	Fe	Ni	Cu	Zn	As	Sn	Pb	SPI	NI	mCD	RI
OK	RMSE	23.93	158.64	19176.97	90.63	46.10	31.13	10.95	26.54	20.50	0.55	3.16	0.70	14.24
	R^2	0.06	0.40	0.008	0.005	0.41	0.045	0.002	0.038	0.004	0.09	0.04	0.07	0.03
	SI	0.34	0.42	1.49	0.95	0.56	1.25	2.07	0.19	2.11	0.46	0.19	0.19	0.39
	BIAS	1.03	1 .55	7595.03	−0.04	−1 .11	−0.19	0.61	1.05	−0.01	0.002	0.13	0.01	0.30
OCK	RMSE	20.60	137.16	19761.54	78.13	–	32.14	10.41	25.75	18.75	0.49	3.06	0.70	12.87
	R^2	0.17	0.41	0.005	0.1	–	0.047	0.013	0.004	0.06	0.04	0.004	0.006	0.003
	SI	0.31	0.41	0.64	0.92	–	1.23	1.97	0.20	1.85	0.45	0.20	0.20	0.37
	BIAS	0.01	0.41	287.781	0.85	–	−0.15	0.44	1.04	0 .15	−0.01	0.12	0.04	0.13
ANN	RMSE	18.52	53.67	14931.02	67.81	58.77	3.18	8.22	12.40	8.33	0.45	2.14	0.43	12.06
	R^2	0.91	0.95	0.76	0.94	0.88	0.97	0.81	0.92	0.81	0.85	0.67	0.61	0.68
	SI	0 .15	0.55	0.19	0.32	0.98	1.2	1.08	0.98	0 .97	0 .99	0.86	0.99	0.99
	BIAS	4.99		2279.4	13.63	−15.4	3.12	0.1	−1.52	−0.6	0.01	−2.6	−.005	−0.3

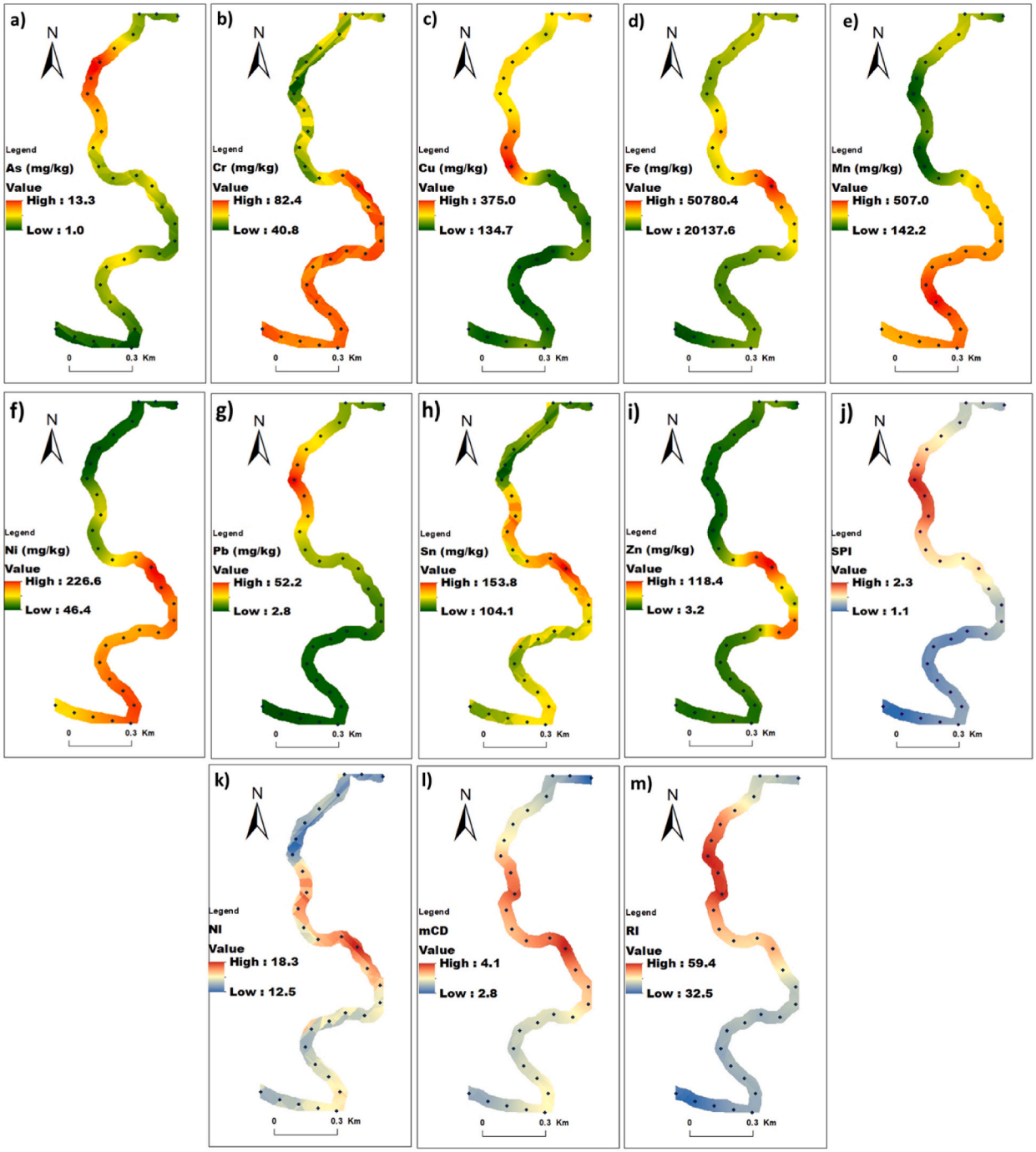


Fig. 8. Predicted maps of trace metals and pollution indices using Ordinary Kriging: a) **As**; b) **Cr**; c) **Cu**; d) **Fe**; e) **Mn;** f) **Ni**; g) **Pb**; h) **Sn**; i) **Zn**; j) **SPI**; k) **NI**; l) **mCD** and m**) RI.**

3.3. Artificial neuronal network models and prediction

The spatial distribution of studied metals and computed pollution indices produced by ANN are presented in Fig. 10. For the trace metals, all the maps showed three degrees of contamination, low contamination represented in green, medium contamination in Yellow and high contamination represented in red. The high contamination were observed inside the mining site. Except for Fe, Ni and Mn, the area located upstream of the mining site has a relatively low level of metal contamination. Regarding the pollution indices, they follow almost the same trends as those of trace metals. According to Fig. 10, the range of variation for OK and OCK models were

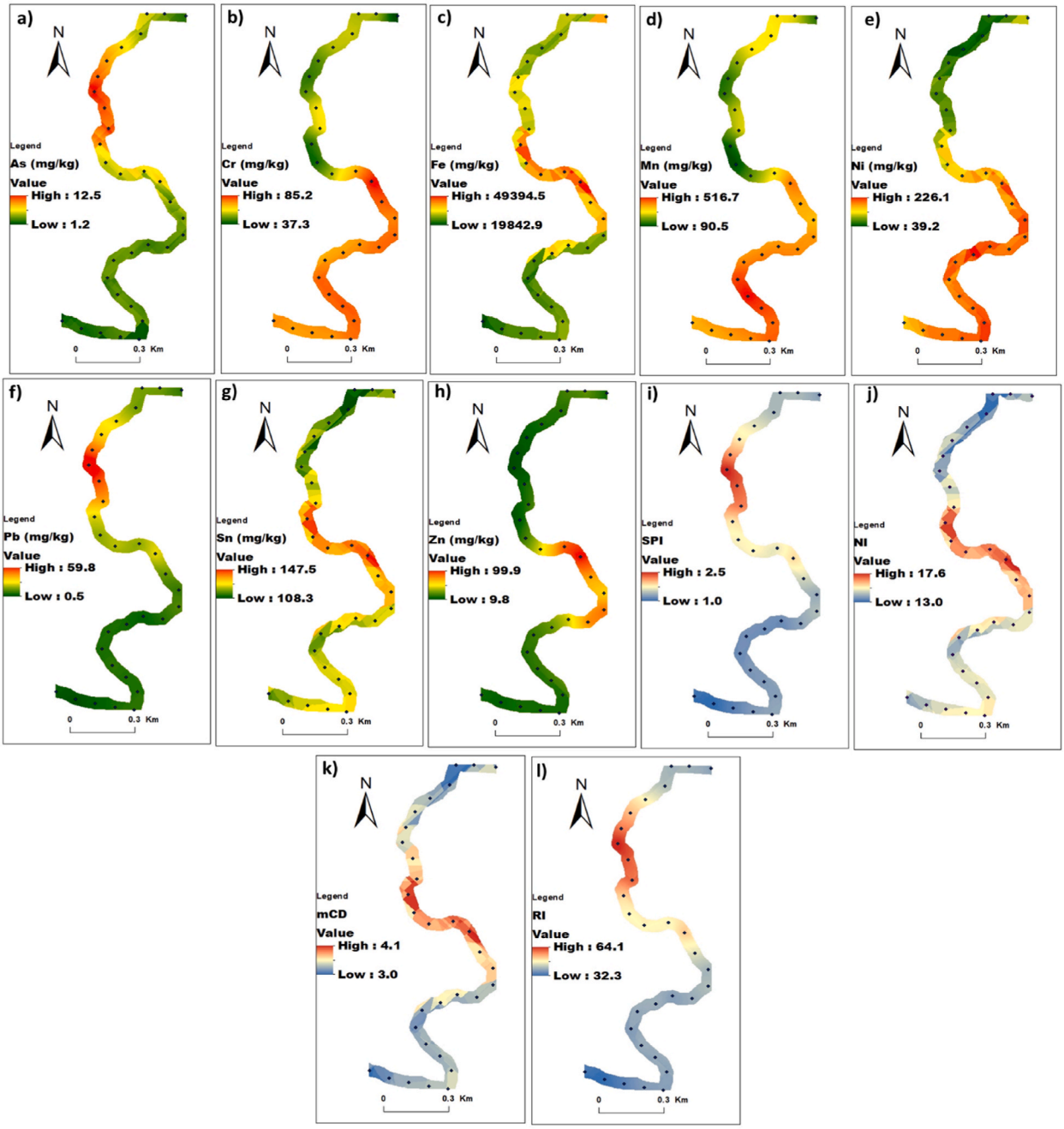


Fig. 9. Predicted maps of trace metals and pollution indices using Ordinary Cokriging: a) **As**; b) **Cr**; c) **Cu**; d) **Fe**; e) **Mn;** f) **Ni**; g) **Pb**; h) **Sn**; i) **Zn**; j) **SPI**; k) **NI**; l) **mCD** and m**) RI.**

respectively As (0.00–68.99 mg/kg), Cr (0.00–183.45 mg/kg), Cu (277.40–653.61 mg/kg) Fe (0.00–27,647 mg/kg), Mn (0.00–443.46 mg/kg), Ni (0.00–52.52 mg/kg), Pb (0.00–77.76 mg/kg) Sn (55.36–232.74 mg/kg), Zn (0.00–47.12 mg/kg), SPI (1.27–2.94), NI (2.68–15.38), mCD (087–3.46) and RI (0.00–76.50).

3.4. Cross-validation

The accuracy of OK, OCK and ANN models for the prediction of the studied trace metals and the pollution indices was assessed using RMSE, R^2, SI and BIAS (Table 5). The RMSE values clearly demonstrated that except for Cu, ANN is the model with the best performance for the prediction of the studied parameters followed by OCK while OK showed the highest RMSE values. As for the RMSE,

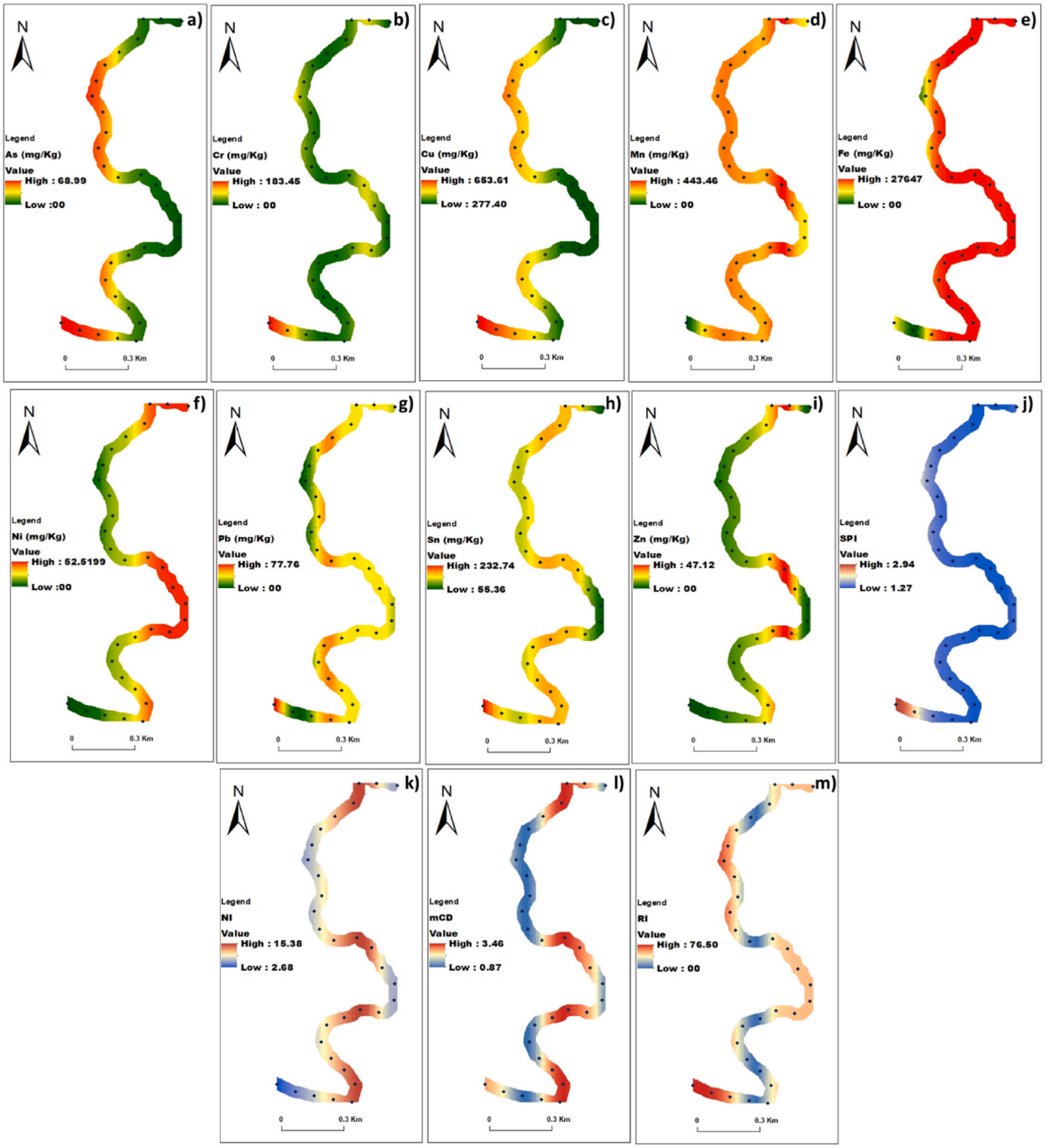


Fig. 10. Predicted maps of trace metals and pollution indices using ANN: a) **As**; b) **Cr**; c) **Cu**; d) **Fe**; e) **Mn;** f) **Ni**; g) **Pb**; h) **Sn**; i) **Zn**; j) **SPI**; k) **NI**; l) **mCD** and m**) RI.**

the R^2 values confirmed that ANN was the model with the best performance. Regarding to OK and OCK, the R^2 values showed that except for Fe, Sn, SPI, NI, mCD and RI, OCK showed better performance than OK. Overall, OK and OCK showed relatively very low R^2 values. Indeed, exceptions made for Mn and Cu ($R^2 \geq 0.40$), the values of $R^2 \leq 0.17$, reflecting poor performances for the studied parameter. The SI values calculated for the various models showed that except for Cr, Fe, Ni, As and Pb, the values predicted by ANN had a larger dispersion than those obtained by OK and OCK. Those obtained by OCK, except for Sn, NI and mCD had a smaller dispersion than those predicted by OK. To better understand the accuracy of the developed models, the BIAS has been computed to understand if the selected model offers overall under- or over-predictions. According to Table 5, except for Fe and As, the models obtained by ANN showed the greatest values of BIAS. This indicated that, for each case, under- or over-predictions compared to OK and

OCK. Finally, except for Ni, Pb, SPI and mCD, OCK has the smallest values of BIAS. Cross-validation has shown that for RMSE and R^2, although OK and OCK can represent both trace metal and pollution index prediction techniques, ANN is a more robust predictor despite having a relatively high BIAS.

4. Conclusions

The spatial variability of trace metals and pollution indices in the surface sediments of the abandoned gold mine sites along the Lom River at Bekao (Adamawa Cameroon) was assessed in this paper. In addition, the study was compared the performance of OK, OCK and ANN models in the prediction of trace metals and pollution indices. For this, thirty one (31) surface sediments samples were collected and then analyzed in order to determine the concentrations of As, Cr, Cu, Fe, Mn, Ni, Pb, Sn and Zn. The trace metal concentrations obtained were compared with ASVs then used to compute pollution indices as SPI, NI, mCD and RI. The following conclusions are reached from the overall research findings:

- The sequence of trace metal mean concentrations was Fe > Mn > Cu > Ni > Sn > Cr > Zn > Pb > As and the mean concentrations of Ni, Cu, Zn and Sn were above the ASVs;
- mCD, NI, SPI, and RI indices suggested that the level of pollution were globally moderate to significant with a low potential ecological risk;
- According to OK and OCK, moderate spatial dependency was observed for Mn, Fe, Ni, Pb, SPI, NI, mCD, RI As, Cr, and Sn indicating an anthropogenic input, while Cu and Zn showed a strong spatial dependency;
- All the prediction maps issued for OK, OCK and ANN revealed that the low concentrations of trace metals were most observed upstream of the mining area and the highest concentrations inside the mining area, showing the potential impact of gold mining exploitation;
- Cross validation demonstrated that although OK and OCK can represent techniques for predicting trace metals and pollution indices, ANN was a more robust and better predictor than both.

Therefore, ANN appears to be a more accurate and robust approach for predicting sediment contamination in unsampled locations. It is important to mention that a larger database could improve the model's performance.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contribution statement

Reynolds Yvan ABENDE SAYOM: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Martin Luther Mfenjou: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Mouhamed AYIWOUO NGOUNOUNO: Performed the experiments; Analyzed and interpreted the data.

Michele Maguy Cathya ETOUNDI: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

William André BOROH: Performed the experiments; Analyzed and interpreted the data.

Luc Leroy MAMBOU NGUEYEP: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Arsene MEYING: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Data availability statement

Data will be made available on request.

Declaration of competing interest

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

Acknowledgements

The authors thank the Eco-materials and Environment Laboratory of the School of Geology and Mining Engineering, University of Ngaoundéré, Cameroon, for its support during fieldwork and analysis. The authors also wish to thank the anonymous reviewers and the editor for their helpful suggestions and enlightening comments.

References

[1] A.I. Tsafe, L.G. Hassan, D.M. Sahabi, Y. Alhassan, B.M. Bala, Assessment of heavy metals and mineral compositions in some solid minerals deposit and water from a gold mining area of Northern Nigeria, Int. Res. J. Geol. Min. 2 (9) (2012) 254–260.

[2] A.W. Rate, A.E. Robertson, A.T. Borg, Distribution of heavy metals in near-shore sediments of the Swan River estuary, Western Australia, Water, Air, Soil Pollut. 124 (2000) 155–168.

[3] L. Giusti, Heavy metal contamination of brown seaweed and sediments from the UK coastline between the Wear river and the Tees river, Environ. Int. 26 (4) (2001) 275–286.

[4] H.A. Ghregat, Y. Abu-Rukah, M.A. Rosen, Application of geoaccumulation index and enrichment factor for assessing metal contamination in the sediments of Kafrain Dam, Jordan, Environ. Monit. Assess. 178 (2011) 95–109.

[5] B. Cui, Q. Zhang, K. Zhang, X. Liu, H. Zhang, Analyzing trophic transfer of heavy metals for food webs in the newly-formed wetlands of the Yellow River Delta, China, Environ. Pollut. 159 (2011) 1297–1306.

[6] C. Wang, S. Liu, Q. Zhao, L. Deng, S. Dong, Spatial variation and contamination assessment of heavy metals in sediments in the Manwan Reservoir, Lancang River, Ecotoxicol. Environ. Saf. 82 (2012) 32–39.

[7] P. Yang, R. Mao, H. Shao, Y. Gao, The spatial variability of heavy metal distribution in the suburban farmland of Taihang Piedmont Plain, China, Comptes rendus de biologie 332 (6) (2009) 558–566. https://doi:10.1016/j.crvi.2009.01.004.

[8] K.L. Speneer, C.L. Macleod, Distribution and partitioning of trace metals in estuarine sediment cores and implications for the use of sediment quality standards, Hydrol. Earth Syst. Sci. 6 (2002) 989–998, https://doi.org/10.5194/hess-6-989-2002.

[9] R. Webster, M.A. Oliver, Geostatistics for Environmental Scientists, second ed., John Wiley & Sons Ltd, England, 2007.

[10] S.K. Reza, U. Baruah, S.K. Singh, T.H. Das, Geostatistical and multivariate analysis of soil heavy metal contamination near coal mining area, northeastern India, Environ. Earth Sci. 73 (9) (2015) 5425–5433.

[11] Y. Jiang, S. Chao, J. Liu, Y. Yang, Y. Chen, A. Zhang, H. Cao, Source apportionment and health risk assessment of heavy metals in soil for a township in Jiangsu Province, China, Chemosphere 168 (2017) 1658–1668, https://doi.org/10.1016/j.chemosphere.2016.11.088.

[12] A. Gribov, K. Krivoruchko, Empirical Bayesian kriging implementation and usage, Sci. Total Environ. 722 (2020), 137290, https://doi.org/10.1016/j.scitotenv.2020.137290.

[13] N.M. Ayiwouo, N.L.L. Mambou, A.W. Boroh, T.K. Sifeu, I. Ngounouno, Spatial variability of trace metals in sediments along the Lom River in the gold mining area of Gankombol (Adamawa Cameroon) using geostatistical modeling methods, Model. Earth Syst. Environ. 9 (2022) 313–329, https://doi.org/10.1007/s40808-022-01500-9.

[14] M.A. Salam, S.C. Paul, F.I. Shaari, A.E. Rak, R.B. Ahmad, W.R. Kadir, Geostatistical distribution and contamination status of heavy metals in the sediment of Perak River, Malaysia, Hydrology 6 (2) (2019) 30.

[15] K.M. Jolanta, S.B. Andrzej, Geostatistical modelling of soil contamination with arsenic, cadmium, lead, and nickel: the Silesian voivodeship, Poland case study, AIMS Geosci. 6 (2) (2020) 135–148, https://doi.org/10.3934/geosci.2020009.

[16] N.S. Tajudin, M. Zulkifli, M.F. Miskon, M.I. Anuar, Z. Hashim, F. Faudzi, N.M.A. Jamaluddin, Integrated approach of heavy metal evaluation using geostatistical and pollution assessment index in soil of bauxite mining area, Pertanika J. Sci. Technol. 30 (2) (2022).

[17] M.M. Njayou, N.M. Ayiwouo, L.L. Mambou, I. Ngounouno, Using geostatistical modeling methods to assess concentration and spatial variability of trace metals in soils of the abandoned gold mining district of Bindiba (East Cameroon), in: Modeling Earth Systems and Environment, 2022, pp. 1–15, https://doi.org/10.1007/s40808-022-01560-x.

[18] S. Eljebri, M. Mounir, A.T. Faroukh, A. Zouahri, R. Tellal, Application of geostatistical methods for the spatial distribution of soils in the irrigated plain of Doukkala, Morocco, Model. Earth Syst. Environ. 5 (2018) 669–687, https://doi.org/10.1007/s40808-018-0558-2.

[19] O. Lemarchand, N. Jeannée, Méthodes de cartographie et approche géostatistique.La cartographie de la pollution au dioxyde d'Azote en Alsace, Cahier des thèmes transversaux ArScAn 9 (2009) 203–214.

[20] A.M.C. Wadoux, A.B. Minasny B McBratney, Machine learning for digital soil mapping: applications, challenges and suggested solutions, Earth Sci. Rev. 210 (2020), 103359.

[21] D.A. Tarasov, A.N. Medvedev, A.P. Sergeev, A.V. Shichkin, A.G. Buevich, A hybrid method for assessment of soil pollutants spatial distribution, AIP Conf. Proc. 1863 (2017), 050015, https://doi.org/10.1063/1.4992212.

[22] C. Li, C. Zhang, T. Yu, X. Liu, Y. Yang, Q. Hou, Z. Yang, X. Ma, L. Wang, Use of artificial neural network to evaluate cadmium contamination in farmland soils in a karst area with naturally high background values, Environ. Pollut. 304 (2022), https://doi.org/10.1016/j.envpol.2022.119234.

[23] M. Najafzadeh, A. Ghaemi, S. Emamgholizadeh, Prediction of water quality parameters using evolutionary computing-based formulations, Int. J. Environ. Sci. Technol. 16 (2019) 6377–6396, https://doi.org/10.1007/s13762-018-2049-4.

[24] M. Najafzadeh, F. Homaei, H. Farhadi, Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: integration of remote sensing and data-driven models, Artif. Intell. Rev. 54 (2021) 4619–4651, https://doi.org/10.1007/s10462-021-10007-1.

[25] M. Najafzadeh, S. Basirian, Evaluation of river water quality index using remote sensing and artificial intelligence models, Rem. Sens. 15 (9) (2023) 2359, https://doi.org/10.3390/rs15092359.

[26] M. Najafzadeh, A. Ghaemi, Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods, Environ. Monit. Assess. 191 (6) (2019) 380, https://doi.org/10.1007/s10661-019-7446-8.

[27] L. Jie, Hyperspectral remote sensing estimation model for cd concentration in rice using support vector machines, Yingyong Kexue Xuebao/J. Appl. Sci. 30 (1) (2012) 105–110.

[28] G. Liu, X. Zhou, Q. Li, Y. Shi, G. Guo, L. Zhao, C. Zhang, Spatial distribution prediction of soil as in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data, Environ. Pollut. 267 (2020), 115631.

[29] Y. Huang, J. Lin, X. Lin, W. Zheng, Quantitative analysis of Cr in soil based on variable selection coupled with multivariate regression using laser-induced breakdown spectroscopy, J. Anal. At. Spectrom. 36 (11) (2021) 2553–2559.

[30] M. Najafzadeh, S. Niazmardi, A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters, Nat. Resour. Res. 30 (2021) 3761–3775, https://doi.org/10.1007/s11053-021-09895-5.

[31] H. Zhang, S. Yin, Y. Chen, S. Shao, J. Wu, M. Fan, C. Gao, Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China, J. Clean. Prod. 273 (2020), 122858.

[32] A. Suleymanov, I. Tuktarova, L. Belan, R. Suleymanov, I. Gabbasova, L. Araslanova, Spatial prediction of soil properties using random forest, k-nearest neighbors and cubist approaches in the foothills of the Ural Mountains, Russia, Model. Earth Syst. Environ. (2023) 1–11, https://doi.org/10.1007/s40808-023-01723-4.

[33] R. Olawoyin, A. Nieto, R.L. Grayson, F. Hardisty, S. Oyewole, Application of artificial neural network (ANN)–self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions, Expert Syst. Appl. 40 (9) (2013) 3634–3648, https://doi.org/10.1016/j.eswa.2012.12.069.

[34] F. Dai, Q. Zhou, Z. Lv, W. Wang, G. Liu, Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau, Ecol. Indicat. 45 (2014) 184–194, https://doi.org/10.1016/j.ecolind.2014.04.003.

[35] M. Sakizadeh, R. Mirzaei, H. Ghorbani, Support vector machine and artificial neural network to model soil pollution: a case study in Semnan Province, Iran, Neural Comput. Appl. (2022), https://doi.org/10.1007/s00521-016-2231-x.

[36] H. Yoon, S.H. Jun, Y. Hyun, G.O. Bae, K.K. Lee, A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer, J. Hydrol. 396 (2011) 128–138.

[37] C. Peng, X. Wen, Recent applications of artificial neural networks in forest resource management: an overview, in: Environmental Decision Support Systems and Artificial Intelligence, Aaai Workshop, 1999, pp. 15–22.

[38] Marlborough District Council, Analysis of Soil Samples Using a Portable X-Ray Fluorescence Spectrometry (XRF), 2013.

[39] M. Singh, G. Müller, I.B. Singh, Heavy metals in freshly deposited stream sediments of rivers associated with urbanization of the Ganga Plain, India, Water, Air, Soil Pollut. 141 (2002) 35–54.
[40] Z. Yang, W. Lu, Y. Long, X. Bao, Q. Yang, Assessment of heavy metals contamination in urban topsoil from Changchun City, China, J. Geochem. Explor. 108 (1) (2011) 27–38.
[41] S. Jahan, V. Strezov, Comparison of pollution indices for the assessment of heavy metals in the sediments of seaports of NSW, Australia, Mar. Pollut. Bull. 128 (2018) 295–306.
[42] L. Hakanson, An ecological risk index for aquatic pollution control: a sedimentological approach, Water Res. 14 (1980) 975–1001.
[43] G.L. Yuan, T.H. Sun, P. Han, J. Li, X.X. Lang, Source identification and ecological risk assessment of heavy metals in topsoil using environmental geochemical mapping: typical urban renewal area in Beijing, China, J. Geochem. Explor. 136 (2014) 40–47.
[44] Y.A. Hu, X.P. Liu, J.M. Bai, K.M. Shih, E.Y. Zeng, H.F. Cheng, Assessing heavy metal pollution in the surface soils of a region that had undergone three decades of intense industrialization and urbanization, Environ. Sci. Pollut. Res. 20 (2013) 6150–6159.
[45] G.M.S. Abrahim, R.J. Parker, Assessment of heavy metal enrichment factors and the degree of contamination in marine sediments from Tamaki Estuary, Auckland, New Zealand, Environ. Monit. Assess. 136 (2008) 227–238.
[46] K. Turekian, K. Wedepohl, Distribution of the elements in some major units of the earth's crust, Geol. Soc. Am. Bull. 72 (1961) 175–192.
[47] W. Guo, X. Liu, Z. Liu, G. Li, Pollution and potential ecological risk evaluation of heavy metals in the sediments around Dongjiang Harbor, Tianjin, Procedia Environ Sci 2 (2010) 729–736.
[48] S. Haykin, Feedforward neural networks: an introduction, in: Nonlinear Dynamical Systems: Feed Forward Neural Network Perspectives, 2001, pp. 1–16.
[49] M. Djekoun, R. Djebar, S. Bensoltane, L. Ghrieb, Contribution des réseaux de neurones artificiels (RNA) a la caractérisation d'un stress oxydatif chez saccharomyces cerevisiae induit par le cadmium, in: Mise en évidence d'un bio-marqueur potentiel (cas du malondialdehyde), Sciences & Technologie. C, Biotechnologies, 2011, pp. 16–22.
[50] A. Sarkar, P. Pandey, River water quality modelling using artificial neural network technique, Aquat. Procedia 4 (2015) 1070–1077.
[51] Y. Najjar, H. Ali, On the Use of BPNN in Liquefaction Potential Assessment Tasks. Artificial Intelligence and Mathematical Methods in Pavement and Geomechanical Systems, vol. 75, Attoh-Okine (Editor), 1998, pp. 55–63.
[52] Y. Najjar, X.C. Zhang, Characterizing the 3D stress-strain behavior of sandy soils: a neuro-mechanistic approach, in: Numerical Methods in Geotechnical Engineering, 2000, pp. 43–57.
[53] L.P. Wilding, Spatial variability: its documentation, accomodation and implication to soil surveys, in: Soil Spatial Variability, Las Vegas NV, 30 November-1 December 1984, 1985, pp. 166–194.
[54] E. Oku, A. Essoka, E. Thomas, Variability in soil properties along an Udalf toposequence in the humid forest zone of Nigeria, Agric. Nat. Resour. 44 (4) (2010) 564–573.
[55] L.V. Jianshu, Y. Liu, Z. Zhang, J. Dai, Factorial kriging and stepwise regression approach to identify environmental factors influencing spatial multiscale variability of heavy metals in soils, J. Hazard Mater. 261 (1) (2013) 387–397, https://doi.org/10.1016/j.jhazmat.2013.07.065.
[56] F. Rosemary, S.P. Indraratne, R. Weerasooriya, U. Mishra, Exploring the spatial variability of soil properties in an Alfisol soil catena, Catena 150 (2017) 53–61.
[57] N.M. Ayiwouo, N.L.L. Mambou, S.T. Kingni, I. Ngounouno, Spatiotemporal variation and assessment of trace metal contamination in sediments along the Lom River in the gold mining site of Gankombol (Adamawa Cameroon), Environ. Earth Sci. 81 (2022) 379, https://doi.org/10.1007/s12665-022-10501x.
[58] N. Di Virgilio, A. Monti, G. Venturi, Spatial variability of switchgrass (Panicum virgatum L.) yield as related to soil parameters in a small field, Field Crops Res. 101 (2007) 232–239.
[59] E. Venteris, N. Basta, J. Bigham, R. Rea, Modeling spatial patterns in soil arsenic to estimate natural baseline concentrations, J. Environ. Qual. 43 (3) (2014) 936–946.