7th International Conference
"New Horizons Towards Sustainable Development"
6-7 November 2023, Dina Al-Maadawi Hotel, Egypt
International Journal of Environmental Studies and Researches (2023), 2 (4):52-67

Assessment of Soil Pollution in The Industrial Zone in South Jeddah Using Pollution Indices and Machine Learning Model

Adel M. M., Salama, Mohamed F. Azzazy, Ezzat A. Elfadaly

Environmental Studies and Research Institute, University of Sadat City

Abstract

In order to reduce high concentrations of toxic elements in polluted soils, an accurate assessment of the heavy metal concentrations in the industrial city of south Jeddah is required.In this study, the contamination risks for 14 heavy metals, including As, Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Zn, Al, Se, and V in the soil, were evaluated using the contamination degree (CD), pollution load index (PLI), potential ecological risk index (RI), and geoaccumulation index (I geo).Support vector machine regression's effectiveness was used to predict the CD, PLI, and RI based on data for the fourteen heavy metals in the soil,. The results showed thatthere were wide variations in the values of the fourteen heavy metals in soil samples, and they are much polluted at this area of study. The I-geo values indicated non-pollution andpollutionby heavy metals.The soil samples were unpolluted (Igeo< 0) by As, Cd, and Se. In contrast, those samples are strongly polluted (Igeo< 3) by Cu, Pb, and Zn.All of the soil samples under investigation were found to be highly contaminated by the examined elements, per CD, RI, and PLI values. The calibration (Cal.) models of support vector machine regression (SVMR) performed the best in predicting the CD and R1 based on trace elements, with R^2value of 0.99. The validation (Val.) models performed the best in predicting the CD and RI based on data for f trace elements, with high R^2values (0.98 -0.99).

Keywords: heavy elements, Jeddah, pollution Indices, soil, support vector machine regression.

Introduction

Heavy elements in the soil system is increasingly becoming a global issue at both the private and governmental levels, particularly because soil is such an important component of both rural and urban ecosystems (**Asio et al., 2009**), and it is an important "ecological crossroad" in the landscape (**Abu Khatita, 2011**). Concern over preventing heavy metal accumulation in soil, water, and vegetables has significantly shifted as a result of growing awareness of the health risks correlated with environmental chemicals. Toxic metals have been found to have serious health consequences, including the promotion of carcinogenesis-induced tumours (**Nagajyoti et al., 2010; Wang et al., 2017**). Due to their non-biodegradability and lengthy biological half-lives, heavy metals and trace elements are also a cause for concern. Significant concentrations of hazardous heavy metals like Cd, Cu, Zn, Cr, Ni, Pb, and Mn are carried in surface soil by wastewater from industry and other sources, which makes it difficult to use agricultural soil safely and rationally (**Tchounwou et al., 2012; Hu et al., 2017; Panagos et al., 2018; Pourret et al., 2018**). Long-term irrigation with industrial or municipal wastewater is known to significantly affect the amount of heavy and trace elements in surface soil, including Cd, Cu, Zn, Cr, Ni, Pb, and Mn (**Panagos et al., 2018**).

Uncontrolled activity is primarily responsible for the deposition of several hazardous substances into the soil, which ultimately leads to soil degradation and health risks for humans. The primary human factors contributing to the deposition of soil-bearing solids are agricultural practices, specifically land usage, pesticides, inorganic and petrochemical fertilizers for organic matter (bio-solids, animal manure, and organic fertilizers), (**Wang et al., 2017**). Additionally, human activities have a significant effect on the overall quality of soil (**Tchounwou et al., 2012**). The geochemical characteristics of agricultural soils, particularly the concentrations of heavy elements, affect the soil's quality (**Panagos et al., 2018; Pourret et al., 2018**). Today, a wide range of soil contamination indices and measuring techniques, like the I-geo and PLI, are available for evaluating soil contamination. The majority of studieshave focused on the distribution of trace elements in soil and how they move through it (**Kamani et al., 2015; Zhao et al., 2015**). I-geo is a robust tool for measuring and analyzing the quantity of trace elements in soil because it naturally captures both the effects of human activity on these elements and their effects on the environment (**Shui et al., 2020**).

The management of the ecosystem at a safe level is critically impacted by the forecast of pollution indices. In this field, several deterministic models have already been applied during the past few decades (**Sarkar & Pandey, 2015**). However, these cutting-edge models' statistical effectiveness is often low since real-world natural ecosystems are frequently too complex for them. SVMR might provide straightforward and reliable methods for creating models that estimate various pollution indices. SVMR can resolve difficult problems by generalizing non-linear patterns seen in a particular dataset **(Isiyaka et al., 2019).** These data-driven approaches can be used to address extremely nonlinear problems **(Sarkar Adnan et al., 2019).** They have been successfully utilized to assess the precision of the predicted soil constituents **(Šilji´c et al., 2018).** In the industrial city of south Jeddah, there is a shortage of information regarding the effectiveness of SVMR models when used with components for calculating CD, PLI, and RI of soil. There have been some recent attempts to evaluate the effectiveness of ANNs, PLSR, and MLR techniques for assessing soil pollution indices.

In this context, the current study's particular goals were to (i) Assess the ecological risk of soil by calculating Igeo, CD, PLI, and RI; (ii) evaluatethe accuracy of using the SVMR models in quantifying the pollution indices of the soil.

Materials and Methods

Study area and experimental framework

In this study, the sampling area is selected. Surface soil samples were collected in a pollutedarea inside the industrial city of south Jeddah. The industrial area was divided into 5 sectors, and 5 replicates were taken from each sector. All the sampling sites located in location station at Jeddah from roadsides at N 21°26'21.25"E 39°13'13.97"latitude and longitude, second at N 21°25'55.60" E 39°13'41.91"latitude and longitude, third sampling station at N 21°25'39.27" E 39°13'56.32"latitude and longitude,the fourth sampling station was at Jeddah, located near residential complexes at N 21°25'16.53" E 39°14'18.93" latitude and longitude, and the fifth sampling station was obtained at Jeddah at N 21°24'43.53" E39°15'2.70". These points are shown on the Fig. 1.

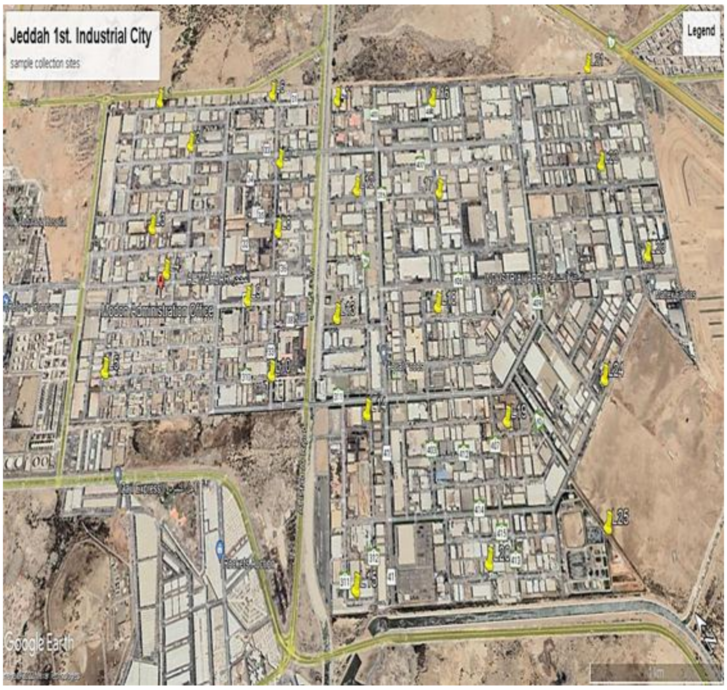

Fig. 1. Map sampling locations of the study

Sampling and laboratory analysis

Soil is critical in supplying vegetative resources for human and ecological existence and wellbeing. As a result, the kingdom of Saudi Arabia and the city of Jeddah must safeguard and monitor soil quality (**UNEP, 2011**). Prior to sampling activities, every sampling glass were cleansed with (HNO_3) and rinsed with deionized water to ensure there were no pollutants in the bottle that could affect the samples. All samples were takento the laboratory in sealed polyethylene bags. The materials were mixed, homogenised, air-dried at 25 to 35^oC, crushed, and sieved to 2 mm. Sample preparations and analysis by ICP Avio in this study was performed as illustrated in Fig. 2. To determine the total element concentration, 1g of powdered soil was digested with aqua regia (HNO3: HCl, 1:3). The concentration of heavy metals such as As, Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Zn, Al, Se, and V were determined**.**

Fig. 2. Sample preparations and analysis by ICP Avio 200

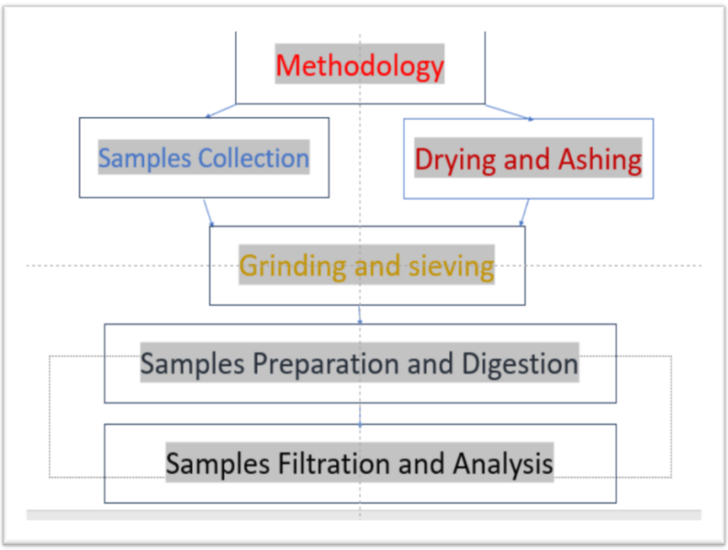

Fig. 3. The experimental framework of this Study

Pollution Assessment Indices

Pollution monitoring indices were developed to assess the degree of contamination with potentially harmful elements. The CD, PLI, and RI were used to measure the overall risk of several factors, while I geo methods were used to estimate single elements.

Hkanson (1980) proposed using the CD in the processes of total contamination evaluation. It is determined using the algorithm in Table 1; four levels of contamination are presented in Table 1 based on the contamination level values.

The PLI is a great tool for assessing overall pollution levels found at different locations (Table 1). The PLI was calculated for all elements in each sample as the nth root of the CF multiplications (Harikumar et al., 2009). The sample's PLI was computed by taking the nth root of the PLI samples.

The RI was used to estimate the environmental danger of heavy metals. According to Hkanson (1980), the RI is divided into four classes, as presented in Table 1.

Müller's (1969) Igeo is a useful tool for detecting whether or not a sample has been affected by anthropogenic heavy metals. The Igeo index was calculated using the algorithm in Table 1. The soil levels of pollution can be classified using the Igeo index values on a scale ranging from 1 to > 5 to the seven groups given in Table 1.

Table 1. The pollution indices, equations, indices criteria, their classifications and reference.

Pollution Indices	Equation	Indices Criteria	Classes	Reference
CD	$D_c = \sum_{i=1}^{i=n} CF$ CF: the contamination factor of each analyzed element in the sample. i: the number of analyzed elements.	<8	Low CD	Håkanson1980
		8 < CD < 16	Moderate CD	
		16< CD < 32	Considerable CD	
		CD > 32	Very high CD	
		>50	Extremely severe enrichment	
I_{geo}	$I_{geo} = \log_2(\frac{Cn}{1.5Bn})$ Cn: the measured concentration of heavy metal n in the sampled sediment. Bn: the geochemical background of the element that is adapted from the literature.	Igeo ≤ 0	Unpolluted	Müller 1969
		0 < Igeo ≤ 1	Unpolluted to Moderately polluted	
		1 < Igeo ≤ 2	Moderately polluted	
		2 < Igeo ≤ 3	Moderately to strongly polluted	
		3 < Igeo ≤ 4	Strongly polluted	
		4 < Igeo ≤ 5	Strongly to extremely polluted	
		5 < Igeo	Extremely polluted	
PLI	PLI = (CF1 × CF2 × CF3 ×...× CFn)$^{1/n}$	1> PLI	Unpolluted	Harikumar et al., 2009
		1 < PLI	Polluted	
RI	$RI = \sum Er$ $Er = Tr \times CF$ Er: the potential ecological risk factor of an individual element. Tr: the toxic response factor. CF: the contamination factor.	RI < 150	Low ecological risk	Håkanson1980
		150 < RI < 300	Moderate ecological risk	
		300 < RI < 600	Considerable ecological risk	
		600 < RI	Very high ecological risk	

Support Vector Machine Regression (SVMR) Model

SVM is a well-known machine learning method based on mathematical learning theory. It can classify enormous volumes of data, detect features, and do regression analysis. SVR attempts to generate functions using the datasets (x, y) provided, where x represents the input vector (where x comprises parameters) and y represents the outcome (y represents the projected pollution indices).The following is a description of the SVM regression function:

$$f(x) = \omega^T \varphi(x) + b \qquad (1)$$

where f(x) indicates the output of the SVM and $\varphi(x)$ indicates a non - linear mapping function. The weighting array ϖ and bias factor b, correspondingly, are to be adjusted utilizing the following regularized functions:

$$\begin{cases} \min R(\omega, \xi, \xi^*, \varepsilon) = \frac{1}{2}\|\varpi\|^2 + C\left[\nu\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)\right] \\ \quad \text{subject to}: \; y_i - \varpi^T \varphi(x_i) - b \le \varepsilon + \xi_i \\ \qquad \varpi^T \varphi(x_i) + b - y_i \le \varepsilon + \xi_i \\ \qquad \xi_i, \varepsilon \ge 0 \end{cases} \qquad (2)$$

where C represents the adjustment value required to balance component and overfitting and the the model normalization component $\|\varpi\|^2$. $\xi_i$and ξ_i^* are the positive slack variables. Based on Lagrange multipliers, the SVR model is resolved.

$$\begin{cases} \max R(\alpha_i, \alpha_i^*) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \\ \quad \text{subjective to}: \; \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \qquad 0 \le \alpha_i, \alpha_i^* \le \frac{C}{l} \\ \qquad \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \le C.\nu \end{cases} \qquad (3)$$

In this case, the kernel functionis $K(x_i, x_j)$, and the positives Lagrange multipliers are $\alpha_i$and α_i^*, accordingly.

The parameters of the SVM are eventually defined after obtaining the desired outcome for the objective function, thus the following regression formula is used to represent an input vector x.

$$f(x) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)\, K(x_i, x_j) + b \qquad (4)$$

Model Evaluation

Some of the following statistical indicators were used to assess the efficacy of a regression model: RMSE and R^2.

Each parameter is described in detail as follows: F_{act} denotes the actual value determined using laboratory procedures. F_p is the forecast or simulated value, F_{ava} is the mean value, and N is the amount of data points in total.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_{act} - F_p)^2} \tag{5}$$

$$\mathrm{R}^2 = \frac{\sum(F_{act} - F_p)^2}{\sum(F_{act} - F_{ave})^2} \tag{6}$$

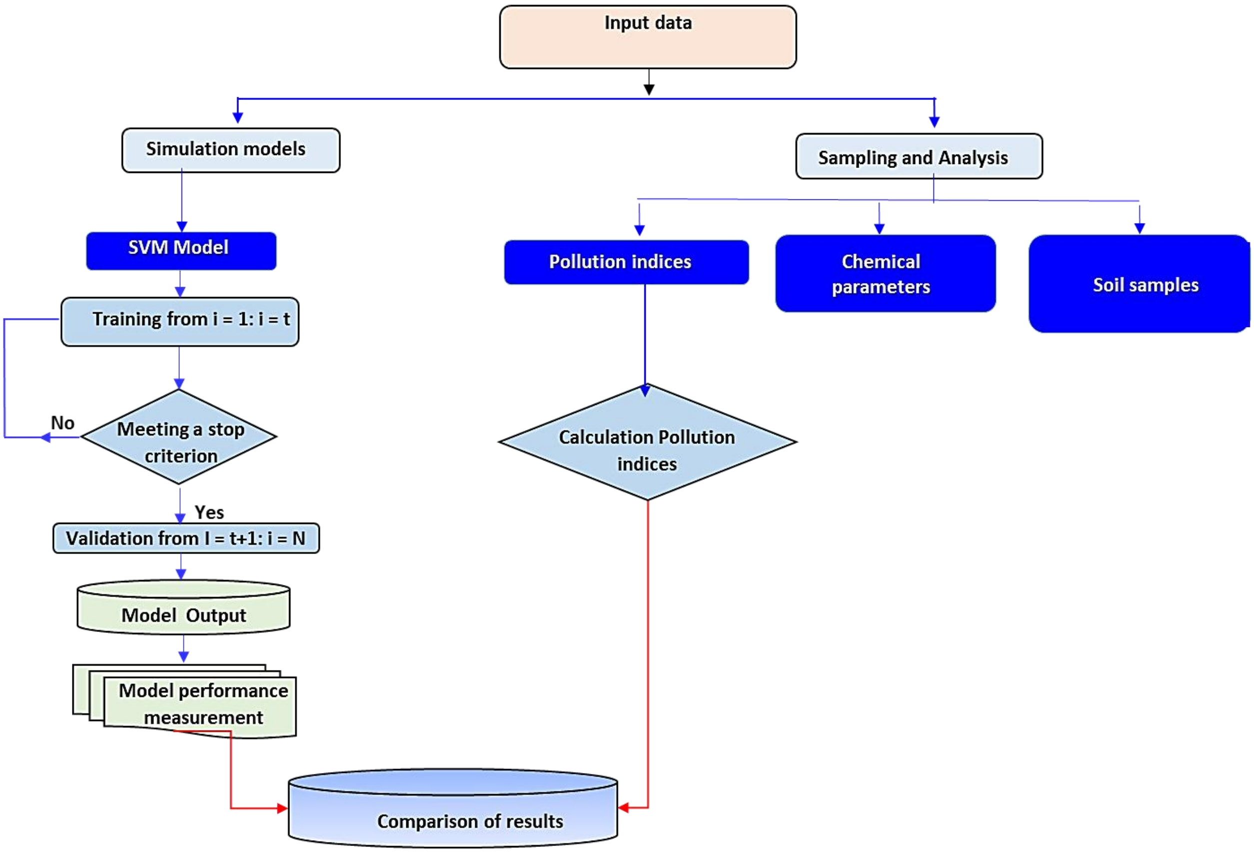

Fig. 4. Flowchart of the SVMR procedure that was utilized topollution indices of soil samples

Results and Discussion

The variation of fourteen heavy metals in soil samples

Table 2 presents the lowest, greatest, mean and standard deviation (SD) values of the heavy metals in soil samples. There were wide variations in the values of the fourteen heavy metals in the soil samples. For example, the As varied from -80.09 to -32.79 with mean value of -49.47, Co varied from 4.13 to 8.97 with mean value of 6.16, Fe varied from 7756.65 to 16568.93 with mean value of 12558.14, and Pb varied from 1.81 to 1950.49 with mean value of 144.15. The findings revealed that the soil samples in this location are highly contaminated. Water contamination is one of the world's most serious environmental problems. Untreated sewage and industrial waste harm the environment. Because the sewage concentration pipes are not connected to the inadequate water pipework, sewage flows directly into the water source. An industrial waste is a point source of pollution, whereas agriculture and urban runoff are non-point sources (**Othman &AlaaEdin 2012**). In this inquiry, the sampling area is identified. The surface dirt was collected in a contaminated area of south Jeddah's industrial city. Poorly treated wastewater discharged by the ceramics, steel and alloys, and other metal processing sectors is a major source of Ni in the soil. Long-term Ni exposure via the food supply may cause skin allergies, dermatitis, rhinitis, nasal sinusitis, lung damage, and nasal mucosal injury (**Yaylali-Abanuz, 2011**). Zn and Cr are heavy metals with human sources in industrial settings (**Yaylali-Abanuz, 2011**). The non-ferrous metal industry and agricultural practices are the primary anthropogenic sources of Zn (**Kabata-Pendias, 2000; Mondol, 2011**). Zinc is a highly mobile element. Toxic and carcinogenic effects of high Zn dosages include neurologic and haematological problems, hypertension, and renal and liver function issues (Roa et al., 2001).

Table 2. Minimum, maximum, mean and standard deviation (SD) of the heavy metals values.

	Minimum	**Maximum**	**Mean**	**SD**
As	-80.09	-32.79	-49.47	12.60
Cd	-2.84	-0.21	-1.47	0.51
Co	4.13	8.97	6.16	1.22
Cr	15.47	81.09	27.61	12.89
Cu	26.23	3631.89	215.40	719.25
Fe	7756.65	16568.93	12558.14	2513.70
Mn	129.23	312.24	223.75	46.02
Ni	10.07	35.15	16.65	6.25
Pb	1.81	1950.49	144.15	388.98
Zn	37.07	2155.38	568.01	613.87
Al	5194.35	15094.98	9304.48	2327.49
Se	-312.14	-99.30	-165.88	46.84
V	0.76	30.93	11.19	6.35

Correlation Matrix

Pearson's correlation matrix was used to assess the rate of similarity and investigate the interrelationships between the elements under consideration (Adimallaet al., 2019). The r values provided suggested a high degree of positive correlation and a significant linear relationship between distinct metal pairings. Table 3 displays the Pearson's correlation analysis results for heavy metals in soil samples collected from25 sites in Jeddah area. The results showed strong correlations between As with Cd, Co, Fe, Mn, Al, and Se with r values of 0.90, -0.68, -0.97, -0.82, -0.73, and 0.94, respectively. A strong correlations between Mn with Al, and Se with (r = 0.86, and -0.67, respectively). The highest strong relationship was found between Fe and Se with r of-0.97. The lowest strong relationship was found between Cu and Pb as well as between Cr and Se with Ni and between Cr and Cu with r of -0.01, -0.01 and 0.00, respectively. The close correlations imply that their occurrence may have originated from a common source, which is probably through other industrial (chemical, paint) activities. Significant correlations between the aforementioned elements are regarded to show that they may accumulate in the same way or emanate from the same contaminated source. It can originate from a single source, such as industrial waste.

The variation in the values of CD, PLI, and RI

In Fig. 5 and Table 4 present CD, PLI, and RI of soil samples for each site. There were wide variations in the values of the three pollution indices. For example, the CD varied from -5791 to -1817 with mean value of -3107, PLI varied from -4.17 to 0 with mean value of -2.3704, RI varied from 3140 to 11686 with mean value of 5183. There were clear differences in the values of CD, PLI, and RI between the sites as shown in Table 4. The results showed that CD, PLI, and RI indicated highly contaminated conditions.

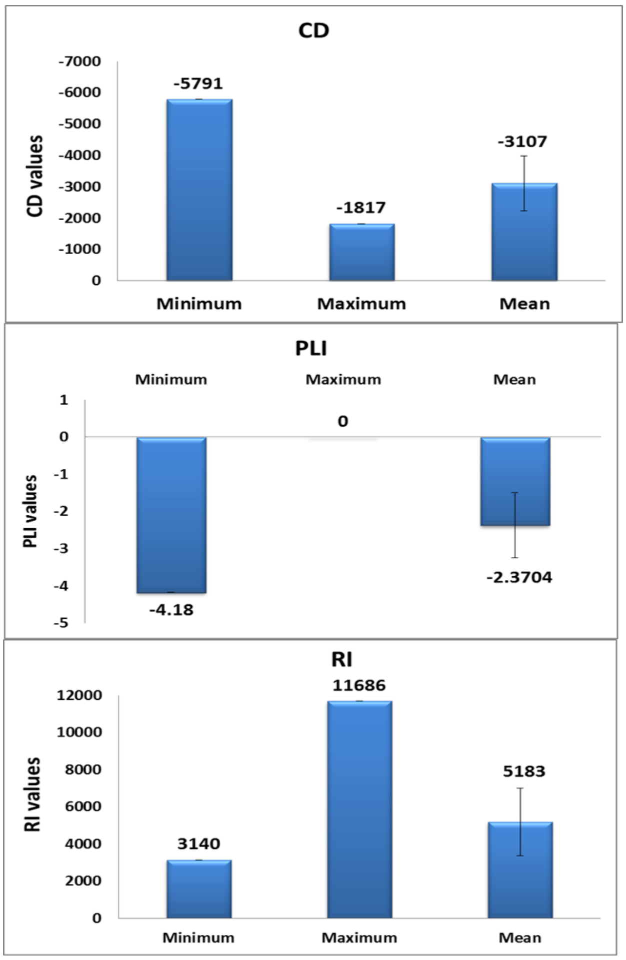

Fig. 5. The statistical description of CD, PLI, and RI.

Table 3. Correlation coefficient of the relationships between the soil trace elements.

	As	**Cd**	**Co**	**Cr**	**Cu**	**Fe**	**Mn**	**Ni**	**Pb**	**Zn**	**Al**	**Se**	**V**
As	1.00												
Cd	0.90**	1.00											
Co	-0.68**	-0.62**	1.00										
Cr	-0.13	0.20	0.04	1.00									
Cu	0.08	0.20	-0.10	0.00	1.00								
Fe	-0.97**	-0.87**	0.81**	0.04	-0.10	1.00							
Mn	-0.82**	-0.73**	0.85**	-0.01	-0.18	0.85**	1.00						
Ni	-0.35	-0.32	0.51**	-0.10	0.57**	0.39	0.32	1.00					
Pb	0.19	0.45*	-0.27	0.87**	-0.01	-0.27	-0.27	-0.19	1.00				
Zn	-0.33	-0.27	-0.06	0.06	0.56**	0.10	0.07	0.31	-0.03	1.00			
Al	-0.73**	-0.59**	0.87**	0.11	0.00	0.82**	0.86**	0.30	-0.27	0.06	1.00		
Se	0.94**	0.92**	-0.53**	-0.07	0.08	-0.97**	-0.67**	-0.32	0.22	-0.43*	-0.57**	1.00	
V	-0.36	-0.24	0.39	0.23	0.21	0.30	0.36	0.10	0.10	0.28	0.39	-0.28	1.00

*,**. Correlation is significant at the 0.05 level and the 0.01 level (2-tailed) respectively.

Table 4. The values of CD, PLI, and RI of soil samples for each site.

Sites	Pollution indices		
	CD	PLI	RI
Field1	-1940.58	-1.66	3435.04
Field2	-2718.64	-2.29	4158.74
Field3	-3880.69	-3.69	5843.92
Field4	-1817.42	-2.40	4612.94
Field5	-1869.36	-1.31	3140.70
Field6	-2633.95	-2.23	4837.17
Field7	-2844.22	-2.16	4598.16
Field8	-2562.52	-2.25	3624.64
Field9	-2740.16	-1.69	3372.13
Field10	-2578.02	-3.98	4278.00
Field11	-2783.46	-2.40	4058.29
Field12	-2453.14	-2.93	5430.20
Field13	-3260.12	-2.36	5193.52
Field14	-3140.31	-1.43	4277.80
Field15	-3670.09	-3.00	6402.80
Field16	-2422.37	-1.74	3912.92
Field17	-3166.90	-2.20	4079.05
Field18	-2424.28	-2.11	3837.86
Field19	-5791.15	0.00	11686.27
Field20	-3654.95	-2.44	5609.49
Field21	-4005.30	-2.99	7525.34
Field22	-3609.05	-2.31	5522.72
Field23	-3967.80	-2.47	6286.16
Field24	-3837.31	-3.04	6582.24
Field25	-3912.02	-4.18	7291.32

The variation of geoaccumulation index (I_{geo}) for assessing soils contamination

The pollution of soils was examined using the I-geo in this investigation. The I-geovalues indicated pollution by As, Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb, Zn, Al, Se, and V in Tables 5 &6.I-geo (As) varied from -8.90 to -3.64, I-geo (Cd) varied from -5.68 to -0.43, I-geo (Cr) varied from 0.05 to0.24, and I-geo (Cu) varied from 0.10 to 13.21 (Table 5). There were clear differences inthe values of I-geo for each heavy metal between sites of thesoil samples, as shown in Tables5 &6.The results show that trace element I-geo levels suggested polluted circumstances.I-geo is excellent for monitoring andanalyzingheavy metals levels in the soil because it intuitively represents the effects of human activities on heavy metals as well as the environmental influence of trace elements (Shui et al., 2020). The Igeo values of the examined soil samples fluctuated between unpolluted to moderately to extremely polluted. The soil samples were unpolluted (Igeo< 0) by As, Cd, and Se. In contrast, those samples very strongly polluted (Igeo< 3) by Cu, Pb,and Zn.

Table 5. Geoaccumulation index (I_{geo}) values of each trace element in soil samples.

Field	As	Cd	Co	Cr	Cu	Fe	Mn	Ni	Pb	Zn	Al	Se	V
1	-3.76	-1.73	0.09	0.05	0.20	0.04	0.03	0.11	0.80	0.61	0.013	-414.15	0.01
2	-5.49	-2.77	0.13	0.07	0.14	0.05	0.05	0.18	0.45	0.51	0.023	-575.59	0.02
3	-7.16	-3.66	0.13	0.12	0.24	0.07	0.06	0.18	3.96	3.81	0.027	-827.66	0.02
4	-3.76	-0.43	0.09	0.24	0.20	0.04	0.03	0.10	26.01	0.56	0.016	-417.98	0.02
5	-3.68	-1.83	0.10	0.09	0.10	0.03	0.04	0.11	0.21	0.33	0.018	-397.21	0.00
6	-5.21	-2.88	0.13	0.07	0.15	0.05	0.05	0.14	0.28	0.48	0.025	-561.67	0.03
7	-5.44	-2.97	0.13	0.06	0.26	0.05	0.06	0.13	0.24	0.55	0.026	-603.48	0.01
8	-4.36	-2.37	0.11	0.06	0.18	0.04	0.05	0.14	0.88	1.22	0.020	-542.83	0.02
9	-3.95	-2.44	0.08	0.05	0.17	0.04	0.03	0.10	0.19	0.65	0.016	-576.98	0.01
10	-4.73	-1.80	0.10	0.08	13.21	0.05	0.04	0.33	1.80	6.16	0.022	-566.86	0.02
11	-4.45	-2.62	0.11	0.06	0.45	0.05	0.05	0.13	0.45	1.14	0.021	-590.03	0.02
12	-4.91	-2.96	0.12	0.07	0.30	0.05	0.05	0.16	5.57	0.94	0.022	-533.17	0.02
13	-5.47	-2.98	0.13	0.09	0.24	0.05	0.05	0.14	0.25	0.61	0.030	-691.90	0.02
14	-5.20	-3.33	0.15	0.05	0.11	0.05	0.05	0.35	0.02	0.11	0.019	-662.08	0.00
15	-6.70	-3.67	0.15	0.11	0.22	0.06	0.06	0.19	0.69	1.06	0.033	-782.01	0.02
16	-4.50	-2.43	0.11	0.07	0.10	0.05	0.04	0.12	0.16	0.39	0.018	-513.87	0.02
17	-5.59	-3.18	0.12	0.07	0.12	0.05	0.05	0.15	0.22	2.95	0.024	-669.11	0.00
18	-3.64	-2.09	0.11	0.05	0.17	0.04	0.04	0.12	0.28	4.37	0.019	-518.76	0.01
19	-8.90	-5.68	0.11	0.10	0.16	0.00	0.06	0.17	2.71	6.03	0.022	-1248.57	0.02
20	-6.18	-3.50	0.12	0.07	0.16	0.06	0.06	0.15	0.43	0.89	0.023	-774.73	0.01
21	-7.72	-4.24	0.18	0.09	0.17	0.07	0.07	0.22	0.51	0.69	0.038	-855.09	0.01
22	-6.05	-3.36	0.13	0.07	0.12	0.06	0.05	0.15	0.65	0.33	0.024	-764.60	0.01
23	-6.39	-3.64	0.14	0.09	0.17	0.07	0.05	0.20	0.17	1.08	0.025	-842.19	0.01
24	-7.15	-3.08	0.14	0.09	0.19	0.07	0.07	0.17	0.45	2.19	0.031	-817.89	0.03
25	-7.03	-4.10	0.17	0.11	2.04	0.06	0.06	0.21	0.68	2.89	0.031	-839.35	0.05

Table 6. The statistical description of geoaccumulation index (I_{geo}) of each trace element of soil samples.

	Minimum	Maximum	Mean	SD
As	-8.90	-3.64	-5.50	1.40
Cd	-5.68	-0.43	-2.95	1.02
Co	0.08	0.18	0.12	0.02
Cr	0.05	0.24	0.08	0.04
Cu	0.10	13.21	0.78	2.62
Fe	0.00	0.07	0.05	0.01
Mn	0.03	0.07	0.05	0.01
Ni	0.10	0.35	0.17	0.06
Pb	0.02	26.01	1.92	5.19
Zn	0.11	6.16	1.62	1.75
Al	0.01	0.04	0.02	0.01
Se	-1248.57	-397.21	-663.51	187.34
V	0.00	0.05	0.02	0.01

Assessment of CD, RI and PLI in sites of soil samples

Table 7 shows the descriptive statistical results of CD, RI and PLI in sites of soil samples.According to the CD, RI, and PLI the findings presented in Table 7, all of the studied soil samples were significantly polluted by the examined elements. The CD, PLI and RI outcomes (Table 7) showed that 100% of the soil samples are polluted. These findings are consistent with those of Elbehiry et al. (2019), who assessed the hazards of four trace elements in Nile Delta soils near the study region using the PLI. They discovered that the PLI revealed pollution in the soil.

Table 7. Assessment of CD, RI and PLI in sites of soil samples.

Indices	Classes	Soil Samples (%)
Contamination Degree (CD)	Low	100%
	Moderate Dc	0%
	Considerable Dc	0%
	Very high Dc	0%
Pollution Load Index (PLI)	Unpolluted	100%
	Polluted	0%
Ecological Risk Index (RI)	Low ecological risk	0%
	Moderate ecological risk	0%
	Considerable ecological risk	0%
	Very high ecological risk	100%

Performance of SVMRmodels based on several elements to assess pollution indices

PLIs of soil sites can be calculated with high precision using mathematical procedures (Jorfi et al. 2017). However, these procedures take time since they involve several mathematical formulas in order to translate a large number of chemical element information into just one number corresponding to contaminants in the soil levels.In contrast, the SVMR strategies are simple and do not require several phases to calculate the CD, PLI, and RI. In recent years, multivariate regression approaches, such as SVMR, have become popular as alternative ways to forecast contamination scores based on data for multiple heavy metals.With $R^2 = 0.99$, the SVMR calibration (Cal.) models scored exceptionally well when forecasting the CD and R1 according to trace element data (Table 8). The validation (Val.) models predicted the CD and RI according to heavy metal information the best, with $R^2 = 0.98$ -0.99 (Table 8 & Figure 6). In overall, the Cal. and Val. of SVMR models predicted the CD and R1 better. The R^2 and RMSE for the CD and R1 were substantially greater in the Cal. and Val. of SVMR models.

Table 8. Results of validation models of SVMR of the association between observed and predicted CD, RI and PLI in soil samples at different sites.

Variable	Calibration		Validation	
	R^2	RMSE	R^2	RMSE
CD	0.999***	25.81	0.989***	116.82
PLI	0.307**	0.71	0.187*	0.782
RI	0.995***	125.9	0.980***	269.29

*,*** Statistically significant at P ≤ 0.05 and P ≤ 0.001, respectively.

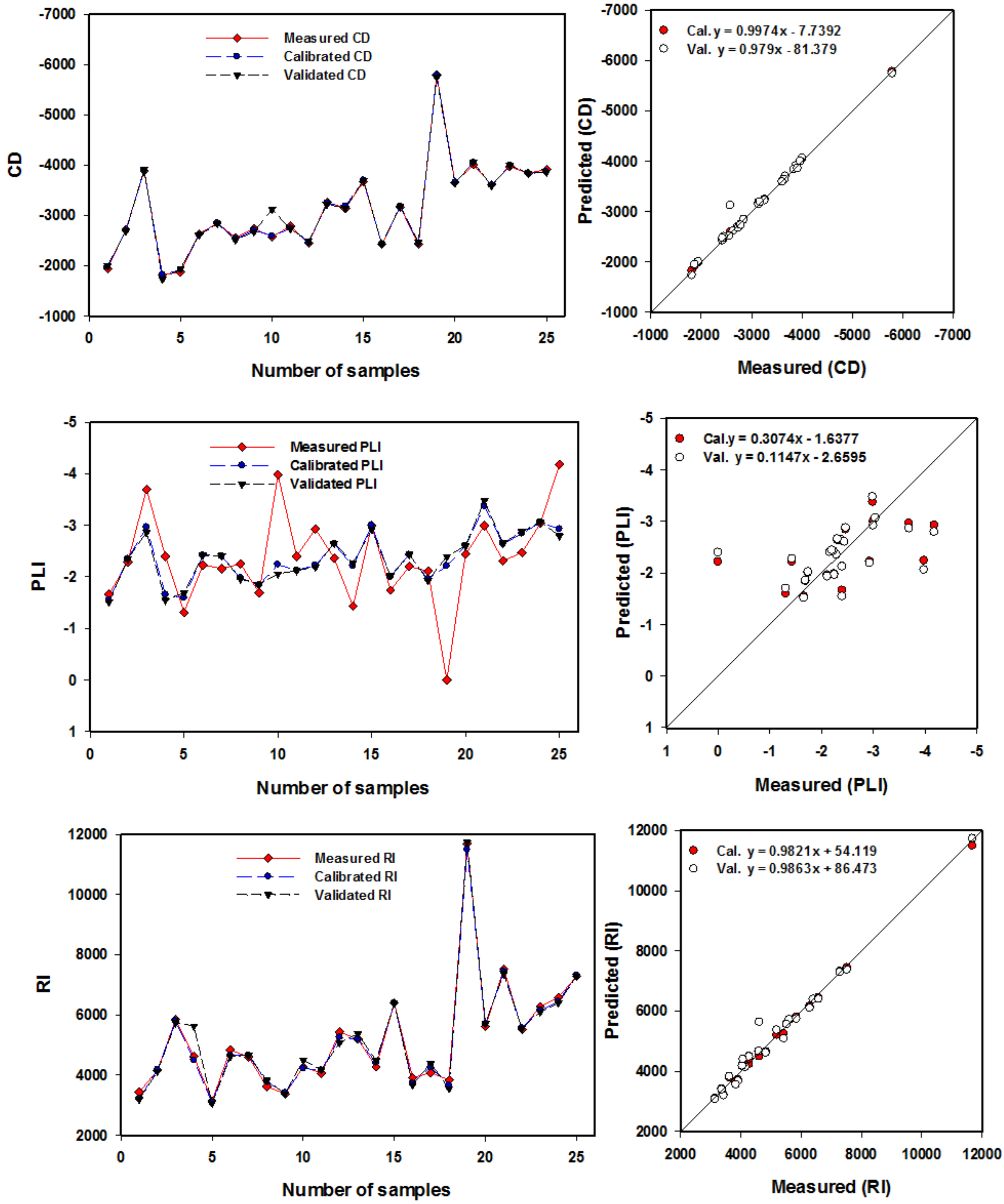

Fig. 6. Comparison of measuring and validating datasets of CD, RI, and PLI at soil sample using the SVMR models. The results of the statistical analysis weredisplayed in Table 8.

Conclusions

The results revealed that the values of the fourteen heavy metals in the soil samples varied greatly. For example, Fe ranged from 7756.65 to 16568.93, with a mean of 12558.14, and Pb ranged from 1.81 to 1950.49, with a mean of 144.15.The CD, PLI, and RI readings revealed highly polluted environments. In general, the Cal. and Val. of SVMR models were efficientin predictingthe CD and RI. The R^2 and RMSE for the CD and RI were greater in the Cal. and Val. of SVMR models. In conclusion, combining the I-geo, CD, PLI and RI, and SVMR models is a useful and practical method for determining the risk of heavy metal contamination. The SVMR models could also be used to assess soil pollution indices by using chemometric techniques.

References

Abu Khatita, A.M. 2011. Assessment of soil and sediment contamination in the middle Nile Delta area (Egypt)—geo-environmental study Using combined sedimentological, Geophysical and Geochemical Methods. Ph.D. Thesis, Alexander-Universität Erlangen Nürnberg zur Erlangung, Erlangen, Germany.

Adimalla, N., Qian, H., Wang, H. Assessment of heavy metal (HM) contamination in agricultural soil lands in northern Telangana, India: An approach of spatial distribution and multivariate statistical analysis. 2019. Environ. Monit. Assess. (191), 246.

Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M., Kisi, O. 2019. Daily streamflow prediction using optimally pruned extreme learning machine. J. Hydrol. (577), 123981.

Asio, V.B., Jahn, R., Perez, F.O., Navarrete, I.A., Abit, S.M. 2009.A review of soil degradation in the Philippines. Ann. Trop. Res. 31 (2), 69–94.

Elbehiry, F., Elbasiouny, H., El-Ramady, H., Brevik, E.C. 2019.Mobility, distribution, and potential risk assessment of selected trace elements in soils of the Nile Delta, Egypt. Environ. Monit. Assess. (191), 713.

Håkanson, L. An ecological risk index for aquatic pollution control: A sedimentological approach.1980. Water Res., 14 (2), 975–1001.

Müller, G.1969. Index of geoaccumulation in sediments of the Rhine River.GeoJournal. 2, 108–118.

Harikumar, P.S., Nasir, U.P., Rahman, M.P.M. 2009. Distribution of heavy metals in the core sediments of a tropical wetland system. Int. J. Environ. Sci. Technol.,(6), 225–232.

Hu, B., Jia, X., Hu, J., Xu, D., Xia, F., Li, Y. 2017. Assessment of heavy metal pollution and health risks in the soil-plant-human system in the Yangtze River Delta, China. Int. J. Environ. Res. Public Health14(9), 1042.

Isiyaka, H.A., Mustapha, A., Juahir, H., Phil-Eze, P. 2019. Water quality modelling using artificial neural network and multivariate statistical techniques. Model Earth Syst. Environ. (5), 583–593.

Jorfi, S., Maleki, R., Jaafarzadeh, N., Ahmadi, M. 2017. Pollution load index for heavy metals in Mian-Ab plain soil, Khuzestan, Iran. Data Brief. (15), 584–59.

Kabata-Pendias, A. 2000. Trace Element in Soils and Plants; CRC Press: Boca Raton, FL, USA, p. 413.

Kamani, H., Ashrafi, S.D., Isazadeh, S., Jaafari, J., Hoseini, M., Mostafapour, F.K., Bazrafshan, E., Nazmara, S., Mahvi, A.H. 2015. Heavy metal contamination in street dusts with various land uses in Zahedan, Iran. Bull. Environ. Contam.Toxicol.,(94), 382–386.

Mondol, M.N., Chamon, A.S., Faiz, B., Elahi, S.F. 2011. Seasonal variation of heavy metal concentrations in Water and plant samples around Tejgaon industrial Area of Bangladesh. J. Bangladesh Acad. Sci., 35 (1), 19–41.

Müller, G. 1969.Index of geoaccumulation in sediments of the Rhine River.GeoJournal.,(2), 108–118.

Nagajyoti, P.C., Lee, K.D., Sreekanth, T.V.M. 2010. Heavy metals, occurrence and toxicity for plants: A review. Environ. Chem. Lett.,(8), 199–216.

Panagos, P., Ballabia, C., Lugato, E., Jones, A., Borrelli, P., Scarpa, S., Orgiazzi, A., Montanarella, L. 2018.Potential sources of anthropogenic copper inputs to European agricultural soils.Sustainability. (10), 2380.

Pourret, O. 2018.On the necessity of banning the term "Heavy Metal" from the scientific literature.Sustainability.(10), 2879.

Roa, M.S., Gopalkrishnan, R., Venkatesh, B.R., 2001. Medical geology-An emerging field in environmental science. In National Symposium on Role of Earth Sciences, Integrated and Related Societal Issues, Lucknow, India, 2–4 November 2001; pp. 213–222, GSI Spl. Pub. No. 65.

Sarkar, A., Pandey, P. 2015. River water quality modelling using artificial neural network technique.Aquat.Procedia.,(4), 1070–1077.

Shui, L., Pan, X., Chen, X., Chang, F., Wan, D., Liu, D., Hu, M., Li, S. 2020. Wang, Y. Pollution characteristics and ecological risk assessment of heavy metals in sediments of the three gorges reservoir.Water.(12), 1798.

Šilji´cTomi´c, A., Antanasijevi´c, D., Risti´c, M., Peri´c-Gruji´c, A., Pocajt, V. 2018. Application of experimental design for the optimization of artificial neural network-based water quality model: A case study of dissolved oxygen prediction. Environ. Sci. Pollut. Res. Int. (25), 9360–9370.

Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J. 2012.Heavy Metal Toxicity and the Environment.Galanin.,(101), 133–164.

Yaylali-Abanuz, G. 2011.Heavy metal contamination of surface soil around Gebze industrial area, Turkey.Microchem. J.,(99), 82–92.

Kabata-Pendias, A. Trace Element in Soils and Plants; CRC Press: Boca Raton, FL, USA, 2000; p. 413.

Wang, J., Liu, G., Liu, H., Lam, P.K. 2017. Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui, China. Sci. Total Environ.,(583), 421–431.

Zhao, Y., Mao, G., Han, S., Gao, L. 2015. Effect of namomaterials on heavy metal transport in alkaline soil.Soil Sediment Contam. Int. J.,(24), 694–703.