Geopersia **7** *(1), 2017, PP. 71-84* *DOI: 10.22059/geope.2017.223390.648290*

Spectroscopic Based Quantitative Mapping of Contaminant Elements in Dumped Soils of a Copper Mine

Vahid Khosravi[1]*, Faramarz Doulati Ardejani[2], Saeed Yousefi[3]
[1] *Faculty of Mining, Petroleum and Geophysics, Shahrood University of Technology, Shahrood, Iran*
[2] *School of Mining, College of Engineering, University of Tehran, Tehran, Iran*
[3] *Nehbandan mining college, University of Birjand, Nehbandan, Iran*
**Corresponding author, e-mail: v.khosravi@shahroodut.ac.ir*
(*received: 01/01/2017 ; accepted: 29/04/2017*)

Abstract
Possibility of mapping the distribution of Arsenic and Chromium in a mining area was investigated using combination of (VNIR) reflectance spectroscopy and geostatistical analysis. Fifty five soil samples were gathered from a waste dump at Sarcheshmeh copper mine and VNIR reflectance spectra were measured in a laboratory. Savitzky- Golay first derivative was used as the main pre-processing method before developing Genetic Algorithm Partial Least Squares Regression (GA-PLSR) and PLSR models for predicting toxic elements concentrations. Physicochemical mechanism that allows the prediction with reflectance spectroscopy was also investigated and it was found that, elements sorption by spectrally active Fe and clay contents of soil was the major mechanism helping the prediction of spectrally featureless As and Cr. Positive relationships were observed between performance of predicting models and iron and clay contents of the samples. Comparing to PLSR, higher prediction performances of both toxic elements concentrations were obtained by applying GA-PLSR model. Furthermore, similar spatial patterns for soil pollution hotspots were observed by geostatistical interpolation (kriging) of chemically measured and models' predicted values. Results demonstrated that the amount and spatial variability of arsenic and chromium can be determined using VNIR spectroscopy and geostatistics in Sarcheshmeh mine's waste dump.

Keywords: *Mapping Elements, Spectroscopy, Genetic Algorithm, PLSR, Geostatistics*

Introduction

Exposure to hazardous materials resulting from industrial and mining activities, urban disposal and incidental accumulations is one of the most critical environmental challenges facing human and ecological communities (Shi *et al.*, 2014). Due to their persistent nature and long biological half-lives, high levels of toxic elements do not only denudes the soil health, but can affect the food chain, causing significant harm to human health and the long-term sustainability of the local ecosystem (Chakraborty *et al.*, 2015; Gholizadeh *et al.*, 2015).

Amount, type and spatial distribution of toxic elements in the soil have to be accurately determined before planning for expensive and time consuming remediation processes (Mulligan *et al.*, 2001; Lewandowski *et al.*, 2006). Spatial distribution has a great importance and conventionally is investigated based on collecting numerous soil samples in field and extensive chemical analysis in laboratory (Cattle *et al.*, 2002; Motelay-Massei *et al.*, 2004; Horta *et al.*, 2015). Although chemical methods are accurate and well documented by the literatures, they are expensive, laborious and time intensive while needing chemical agents and qualified staff. Developing quick, inexpensive and precise alternative measurement methods is of great interest by the researchers.

As a feasible technique in both laboratory and in situ conditions (Rossel *et al.*, 2006), Visible and Near-Infrared (VNIR) reflectance spectroscopy is a rapid and non-invasive analysis technique which needs less sample preparation (only drying and crushing), minimal reagent consumption (McCarty *et al.*, 2002) and several soil attributes can be estimated from a single spectrum (McBratney *et al.*, 2006). It has been used for predicting, spatial analysis and mapping the spectrally active properties of soil samples such as moisture (Gill *et al.*, 2006), organic matter (Conforti *et al.*, 2015), clay (Gomez *et al.*, 2008; Nawar *et al.*, 2016), Fe (Hong-Yan *et al.*, 2009), carbonate and salt content (Ben-Dor and Banin, 1990; Farifteh *et al.*, 2007).

It deems to be possible to monitor and predict the toxic element concentration in sediments and soils since they can be absorbed or bound by above mentioned spectrally active constituents. Wu and colleagues investigated the possibility of predicting Ni, Cr, Cu, Hg, Pb, Zn in suburban soil using PLSR approach associated with reflectance spectra (Wu *et al.*, 2005). They also conducted a research with the

aim of exploring the physicochemical mechanism helping estimation of heavy metals using reflectance spectroscopy (Wu *et al.*, 2007).

The spectra of soil might be broad and non-specific mainly owing to overlapping of soil properties (Rossel and Behrens, 2010). In order to overcome the problem, the spectral characteristics are analyzed using some chemometric methods (Martens and Naes, 1992). Partial least squares regression (PLSR) (Rossel and Behrens, 2010; Araújo *et al.*, 2014; Adeline *et al.*, 2017), support vector machine (SVM) (Araújo *et al.*, 2014; Gholizadeh *et al.*, 2015), principal component analysis (PCA) (Rossel and Behrens, 2010), principal components regression (PCR) (Chakraborty *et al.*, 2015), stepwise multiple linear regression (SMLR) (Vasques *et al.*, 2008; Shi *et al.*, 2013) and random forest (RF) (Wijewardane *et al.*, 2016; Chakraborty *et al.*, 2017) are some of the most preferred methods among researchers to characterize spectra and develop models for predicting the elemental content of the soil.

Some publications have already focused on the spectral-based estimation of toxic elements in soils and water polluted by mining activities. Most of them were devoted to study on homogeneous areas like sediments of mine sites (Choe *et al.*, 2008), agricultural soils around mining areas and croplands beside mines (Hong-Yan *et al.*, 2009; Zhuang, 2009), areas being contaminated after a mining accident (Kemper and Sommer, 2002), areas around abandoned mines (Choe *et al.*, 2009; Shamsoddini *et al.*, 2014), reclaimed mining areas (Wu *et al.*, 2011) and soil profiles near large copper smelters (Xian-Li *et al.*, 2012).

Mine waste dumps have high mineralogical, physical and geochemical heterogeneity. Little studies have been carried out on heterogeneous mining environments like dump site soils. For instance, Gholizadeh *et al.* (2015) evaluated the suitability of reflectance spectroscopy for predicting concentrations of potentially toxic elements on brown coal mine dumpsites in the Czech Republic using partial least square regression and support vector machine. In another study, Gannouni *et al.* (2012) reported the potential use of reflectance spectroscopy in identifying Fe and clay minerals as well as quantitative characterization of toxic metals including Mn, Pb, Zn, Ni, Cr, Fe, Cu, and Cd for the mine waste of Jalta and Bougrine in the North of Tunisia.

However, based on our reviews, no reports have been published to date on study of the metal sulfide mines waste dumps by spectroscopy. Accelerated interactions of certain dumped sulfide minerals (of which pyrite is the most abundant) with water and oxygen, produce acidic sulfur-rich wastewaters, called acid mine drainage (AMD). Such effluents pose serious environmental damages by the fact that they often contain elevated dissolved toxic elements including metals and metalloids. Flowing and increasing pH and salinity of AMD could precipitate toxic elements and inflict environmental hazards to water, sediments and soils of vicinal areas.

Sarcheshmeh is one of the most typical sulfide mines in Iran producing over 400 million tons of mining wastes per year. Acid mine drainage resulted from mining wastes and tailings are discharge directly into the environment causing migration of heavy metals to surrounding regions. In this case study, soil Arsenic and Chromium pollution on a waste dump of Sarcheshmeh copper mine was estimated using VNIR technique. Our objectives were as follows: i) Reveal the binding mechanism by which to predict toxic elements using spectrally active properties, like aluminum oxide (Al_2O_3) representing clay minerals and iron oxides (Fe_2O_3), ii) identify the effect of GA on PLSR calibration method by comparing the estimation accuracy of soil As and Cr content; and iii) Obtaining the spatial distribution and pollution hotspots of As and Cr at regional scale using VNIR and geostatistics.

Materials and methods

Study area

The mining area of Sarcheshmeh which is situated 160 Km southwest of Kerman city, Kerman province, Iran was selected as the study area (Fig. 1). Sarcheshmeh is the biggest porphyry copper mine in Iran and one of the largest Oligo-Miocene deposits in the world. The area has a moderate climate with an annual average rainfall of about 450 mm mostly in winter (Ghaderian and Ravandi, 2012). As a complex intrusive body, the Sarcheshmeh stock has an outcrop of about 1.1 × 2.2 km and consists of three different igneous phases, including diorite/granodiorite, dacitic and related porphyries, and andesite and related dikes (Tahmasebi & Hezarkhani, 2010).

The oval-shaped ore body has dimensions of 2,000 m by 900 m containing 1,200 million tons of ore with average grades copper, molybdenum,

silver and gold of 1.13 %, 0.03 %, 3.9 and 0.11 ppm, respectively (Waterman & Hamilton, 1975). Chalcopyrite, Covellite, Chalcocite, Molybdenite and Bornite are some of the main sulfide minerals of this deposit while Pyrite – the main acid producing source- is the main gangue mineral. More than 35 years of open pit mining has generated 31 active and inactive waste dumps in the vicinity of the mine site with environmental concerns surrounding them.

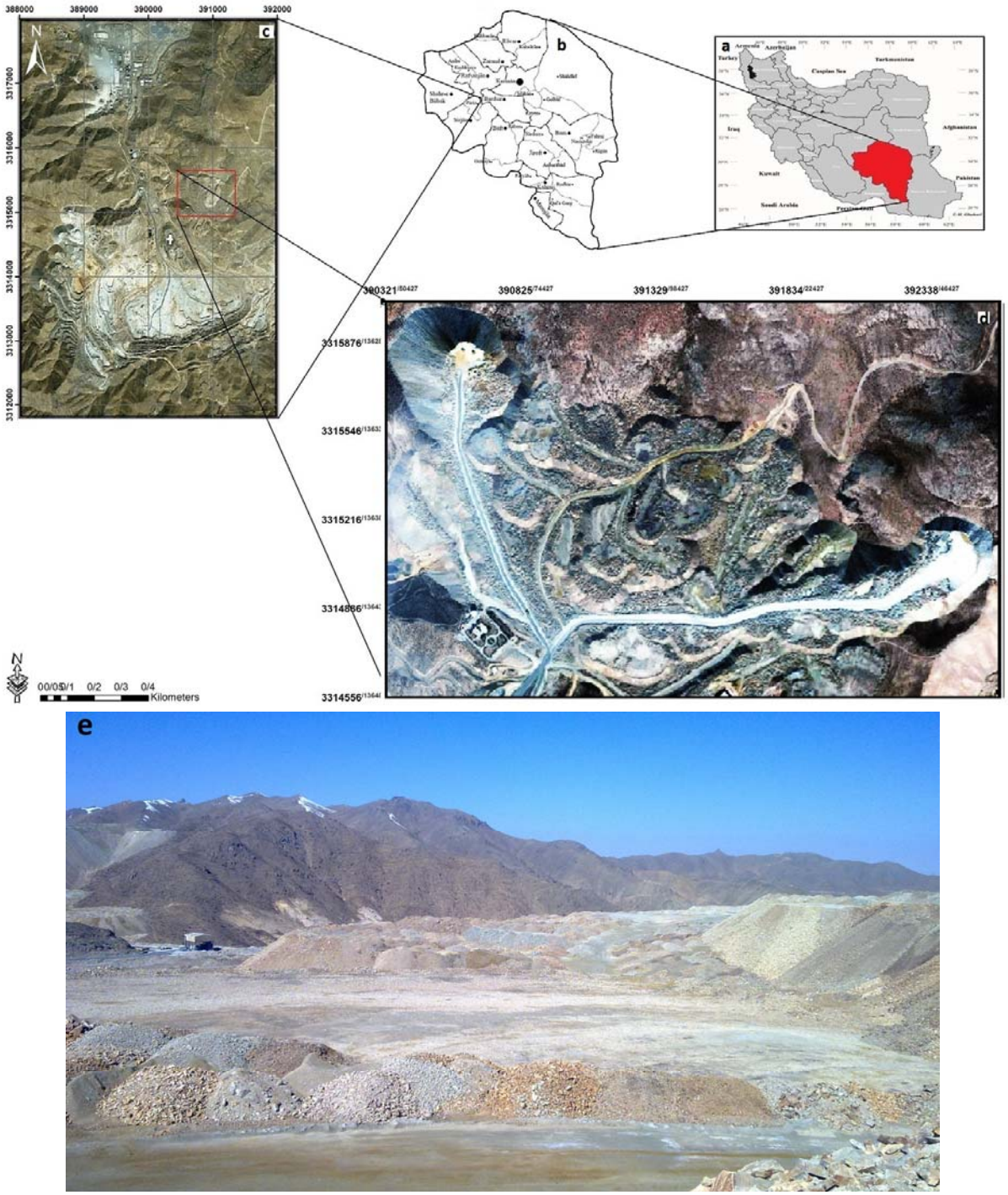

Figure 1. Study area a: In Iran. b) In Kerman province. c: In Sarcheshmeh copper complex. d: Dump number 31. e: A view from part of of the dump surface

Sampling and analysis

Waste dump number 31 locating at the north eastern part of the main pit was the sampling site in this study. (Fig. 1). Fifty five topsoil samples were collected from the dump surface while using a global positioning system (GPS) to record the longitude and latitude of the sampling locations. After air drying at 40 °C, collected samples were screened through a 2 mm sieve to remove stones and coarse fragments before analysis. samples were then ground to 80-mesh size (200 µm) to minimize the impacts of particle-size on soil spectral reflectance (Niazi *et al.*, 2015). Each soil sample was split into two subsamples. One part was used for spectroscopic measurements and the other for investigating soil chemical properties.

Chemical methods

Chemical analysis subsamples were subjected to a variety of chemical analyses. Inductively Coupled Plasma Mass Spectrometry (ICP-MS) method was used to measure concentrations of arsenic and chromium at the LabWest Minerals Analysis Pty Ltd., Australia. Spectrally active properties were evaluated by X-ray diffraction for iron oxides (Fe_2O_3) and aluminum oxide (Al_2O_3), representing clay minerals, at Iran Mineral Processing Research Centre (IMPRC). For each sample, the measurement of pH was carried out using a pH electrode inserted in the slurry containing 50 g of soil in 50 ml of distilled water**.**

Spectroscopic measurements

Before spectral measurements, samples were ground to sieve through a 0.2 mm sieve and then oven-dried at 105 °C overnight. Reflectance spectra of the soil samples were recorded using an Analytical Spectral Device (ASD) Fieldspec® 3 portable spectroradiometer (Analytical Spectral Devices, Inc., USA) in a laboratory. It covers a spectral range of 350–2500 nm and collects data at 10 scans per second. The soil samples were poured in a glass dish with the height and diameter of 2 and 10 cm, respectively (Mouazen *et al.*, 2010). The reflectance of each sample relative to a white $BaSO_4$ panel was then entered directly onto a laptop for three times. Average of the measurements was calculated and used as the final spectrum for pre-processing and constructing models.

Chemical and spectral Preprocessing

Normality of soil Attributes concentration was assessed based on Kolmogorov–Smirnov statistical test at the 5% significance level. Data which did not meet the requirements of a normal distribution were subjected to logarithm transformation.

Prior to the chemometrics analysis, signal to noise ratio was improved by removing the spectral ranges of 350–399 nm and 2451–2500 nm from the measured spectra. For the purpose of reducing the nonlinearities and scattering effects, remaining reflectance spectra were transformed into the absorbance (log10 (1/R), R is reflectance) (Conforti *et al.*, 2015). In order to remove the baseline shift effect, the absorbance spectra were then subjected to the Savitzky–Golay first derivative calculated by fitting a first order polynomial to a spectral window size of 7 data points. Applying multiplicative scatter correction was not necessary due to homogeneous and fine-grained structure of soil samples which attenuated the multiplicative scatter effects. First derivative absorbance spectra were PCA transformed with a varimax rotation using R software version 3.1.0 (R Core Team, 2014). The score plot of the first three PCs was provided to show the relation between data. Data points lying far from the others were considered as outliers and eliminated.

Prediction mechanism

The mechanism of the prediction was investigated by plotting Pearson correlation coefficients between toxic elements content of the samples and spectral variables. By this procedure, the most effective and useful spectral wavelengths were explored. In order to obtain more reliable results, Pearson correlation coefficients between toxic elements and active spectral features content of the soil samples were also calculated.

Along with correlation analysis, the relationship between Fe and clay minerals content of the samples and prediction accuracy of proposed models was investigated in this study. All samples were sorted ascendingly for two times, first by Fe_2O_3 and then Al_2O_3 measured content of the soils. Based on the measured contents of Fe_2O_3 and Al_2O_3, samples were then divided into appropriate equal interval classes. The absolute relative error between predicted and measured As and Cr concentrations of every sample was calculated and then the mean absolute relative error (MARE) was obtained for each class (Eq. 1). The lower the MARE value of each class, the higher the accuracy of prediction for that class. Close relations between

the prediction accuracy of elements and content of each active spectral feature shows the importance of that feature in the mechanism of prediction by VNIR reflectance spectroscopy.

$$MARE = \frac{\sum \left| \frac{predicted\ value - meausred\ value}{meausred\ value} \right|}{n} \quad (1)$$

where n is the number of soil samples

Developing PLSR and GA-PLSR models

Efficiency of wavelength selection using Genetic algorithm (GA) was evaluated by comparing PLSR and GA-PLSR models. PLSR models both the X- and Y-matrices simultaneously to find the latent (or hidden) variables in X that will best predict the latent variables in Y. These PLS components, referring as factors, are similar to principal components. GA is a stochastic optimization algorithm and population based search technique patterned after mechanics of natural selection and natural genetics in biological evolution. It is based on the survival of the fittest theory with genetic operator inspired from the nature. Search, optimization and machine learning are some of the topics that GA successfully deals with (Wang, 1991). Readers are referred to following literatures for detailed theory of PLSR and GA (Holland, 1975; Geladi and Kowalski, 1986; Golberg, 1989; Van Huffel, 1997).

General and GA-PLSR models were established by performing leave-one-out-cross validation (LOOCV) approach on the entire soil samples. In order to calibrate the models and estimate their predictive performances, LOOCV is the best method when facing with small data sets (40–120 samples) (Martens & Dardenne, 1998). The optimum number of latent variables (LVs), was also determined by leave-one-out cross-validation procedure. In order to obtain an optimum GA-PLSR model, different values for GA parameters were tested, resulting in: population Size 30, window width (15 nm), Mutation rate 0.4, Crossover rate 0.1, max generations (150) and replicate runs (5). Root mean square error (RMSE) was used as the fitness criterion.

Prediction quality assessment

Models assessment was based on the coefficient of determination (R^2), root mean squared error (RMSE) (Eq. 2) and the ratio of prediction to deviation (RPD). RPD, the ratio of the SD to the RMSE (Eq. 3) is a factor evaluating the generalizing capability of each model. A five level description of RPD was used including: Useless calibration for values below 1.5, possibility to distinguish between high and low values for RPD between 1.5 and 2, possibility of approximate quantitative predictions for values between 2.0 and 2.5 and finally good and excellent prediction for values between 2.5 and 3.0, and above 3.0, respectively (Saeys *et al.*, 2005).

$$RMSE = \sqrt{\frac{1}{N}\sum_{I=1}^{N}(y'_i - y_i)^2} \quad (2)$$

$$RPD = \frac{SD}{RMSE_p} \quad (3)$$

where y_i' is the predicted value, y_i denotes the observed value, N represents the number of samples and SD is referred to the standard deviation.

Spatial distribution

Spatial distributions of measured and predicted values for As and Cr were visualized by contour mapping of the concentration data using SURFER program (Golden Software Inc., USA). Plots obtained using predicted and measured concentrations were then compared to assess the potential of VNIR spectroscopy in mapping toxic elements. The kriging geostatistical gridding method, was used in order to produce an optimal prediction of the not sampled points. Using variogram models, Kriging creates an accurate grid from unevenly spaced data and identifies trends embedded within (Cressie, 2015). A flowchart of the research procedure is presented in Fig. 2.

Results and discussion

Chemical and Spectral Properties

The score values of the first three PCs provided two and three-dimensional plots. Four soil samples were removed due to their abnormal reflectance spectra (Fig. 3a). Table 1 shows the chemical analysis data for remaining 51 samples. Arsenic average concentration was 42.16 $mg.kg^{-1}$ ranging between the minimum and maximum values of 4.6 $mg.kg^{-1}$ and 178 $mg.kg^{-1}$ respectively.

Meanwhile, Cr content of the soil had average of 49.16 $mg.kg^{-1}$ with variation range of 118 $mg.kg^{-1}$. Both elements showed positively skewed distributions and high coefficient of variations which proved the highly heterogeneous environment of the dump. Because of logarithmic distribution of both elements concentrations, all further analysis were performed on log-transformed values (log10).

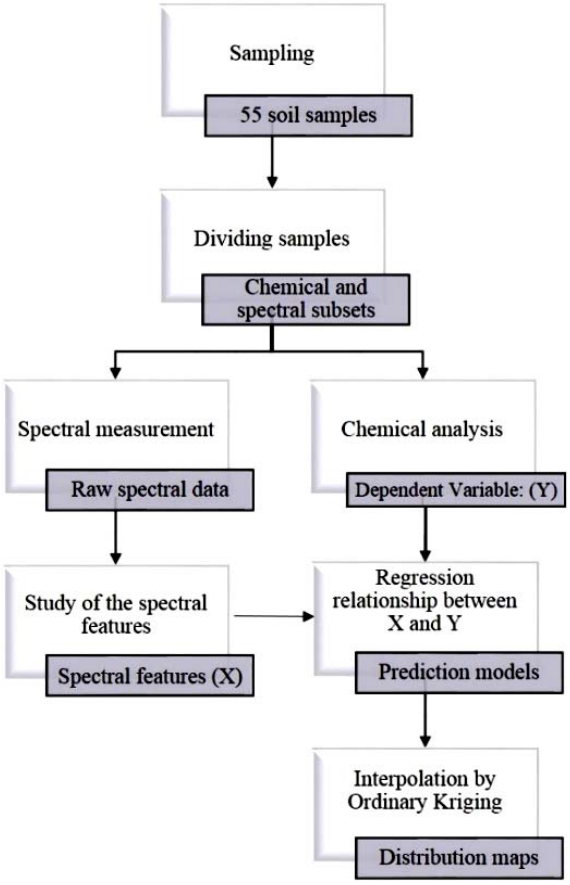

Figure 2. Flowchart of the study

Frequency distribution histograms of normalized arsenic and chromium concentration values are shown in fig. 3b and c.

The paste pH values ranged from 2.081 to 7.526, i.e., strong acid to neutral. The mean value of pH was 4.784; suggesting the acidity of soil in the study area.

Raw, first and second derivative absorbance spectra of four selected representative samples are shown in Fig. 4. The main spectral region for Iron oxy/hydroxides is visible. Accordingly features at 430 nm, 500 nm, 530 nm and 650 nm in the visible region are mainly due to the electronic transitions of of the Fe^{3+} in oxy/hydroxides including goethite (FeOOH) and hematite (Fe_2O_3) (Song *et al.*, 2012).

Clay minerals are mostly responsible for occurring peaks in the NIR region (Kooistra *et al.*, 2003). Peaks at ~1400 and ~2200 nm are mainly related to O-H bonds in the hydroxyl or clay minerals, such as kaolinite, illite and smectite (White, 1971; Nayak and Singh, 2007). The absorption peak at ~1900 nm is due to the O-H bond in water (Clark *et al.*, 1990).

Mechanism of the prediction

Considering Fig. 5, each element displays its maximum correlation coefficient at a different wavelength across the spectral range. Regarding the relationship with soil VNIR spectra, As and Cr had two different behaviors. Arsenic had the highest spectral correlation at 530 nm while the strongest correlation for chromium was at 560 nm both arising due to Fe^{3+} absorption.

In accordance with Fe_2O_3 spectral features, correlation in the visible region of the As varied between 410 and 660 nm. ~460 and ~560 nm relating to the spectral response wavebands of hematite, and ~930 nm relating to the spectral response location of goethite was the other important wavelengths (Fig. 5).

Some studies have reported the important role of Fe oxy/hydroxides in controlling As (arsenate and arsenite) sorption in soil (Alloway, 1990). A positive correlation was found between the contents of As and the Fe oxy/hydroxides in floodplains along the river Rhine in Netherlands (Kooistra *et al.*, 2001). In a goethite dominated area, similar results were reported for prediction of soil toxic elements including As while mentioning the importance of Fe oxides (Wu *et al.*, 2007).

The high correlation coefficients at ~530 and ~550 nm were corresponded to the spectral response wavebands of clay showing that the internal relation between clay and As may be another important prediction mechanism.

Table 1. Statistical description of the soil properties ($mg.kg^{-1}$)

Parameter	Min	Max	Mean±SD	Skewness	Kurtosis	C.V.(%)
Arsenic	4.6	178.1	42.16±39.85	1.997	3.604	94.522
Log (Arsenic)	0.663	2.251	1.481±0.348	0.302	0.014	23.513
Chromium	15	133	49.16±25.686	1.34	1.69	52
Log (Chromium)	1.18	2.12	1.64±0.32	0.12	-0.12	19.51

SD: Standard deviation, C.V.: coefficient of variation

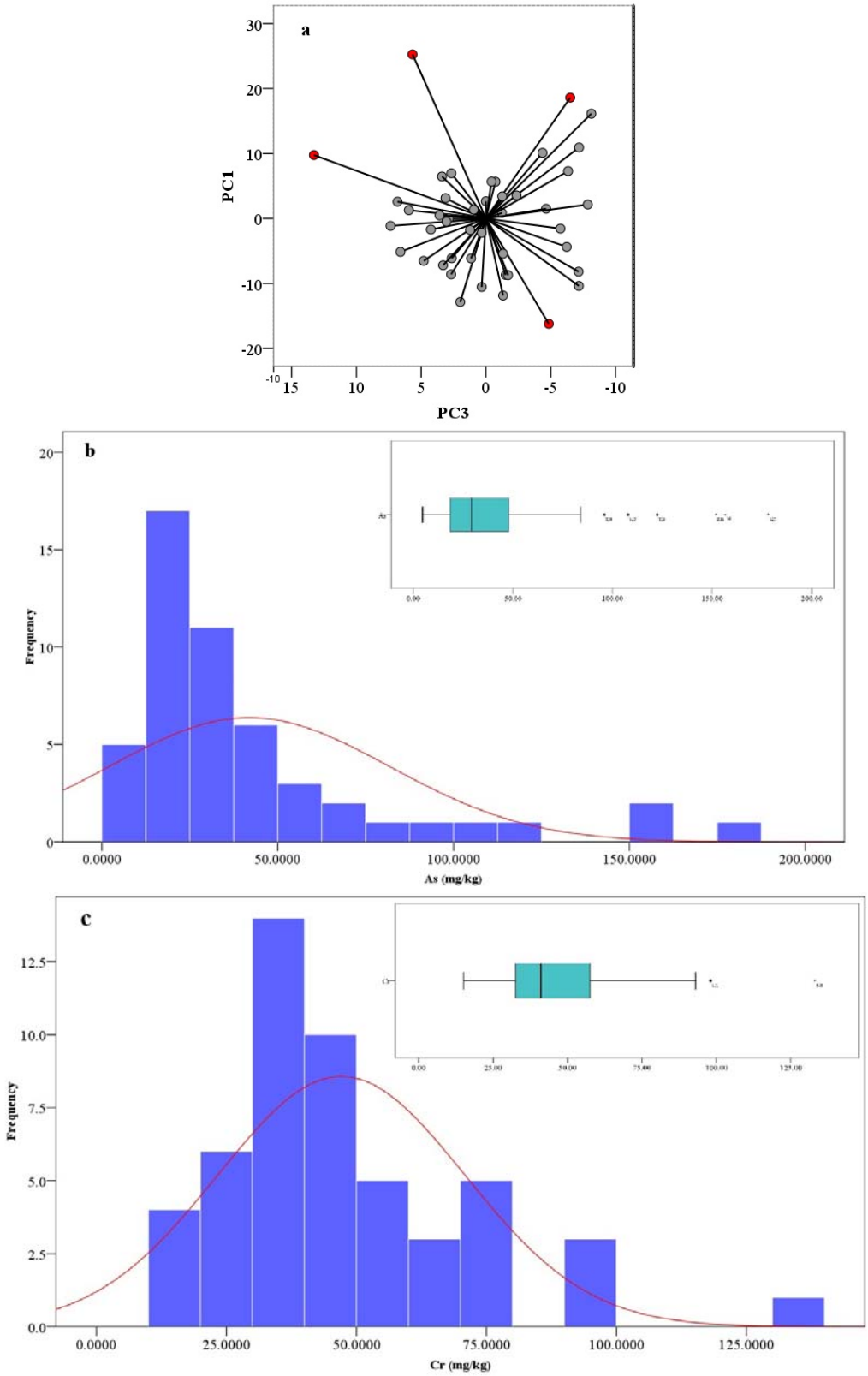

Figure 3. a: 2D score plot of two principal component analysis derived PCs (Red filled circles are outliers), frequency distribution histograms and box plots of b: arsenic and c: chromium contents of the soil samples

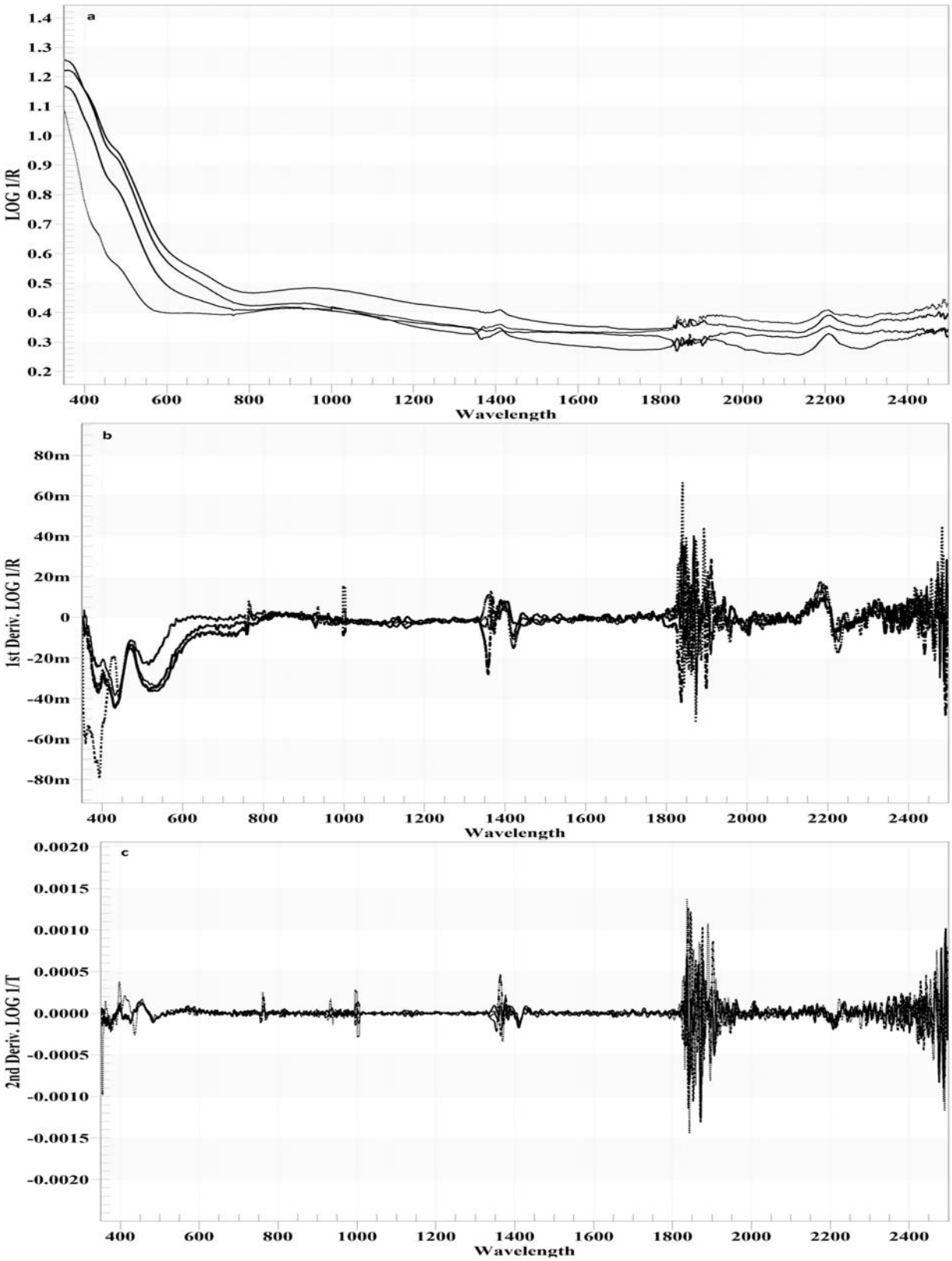

Figure 4. a) Raw b) first derivative and c) second derivative absorbance spectra of four representative soil samples (unit of the wavelengths: nm)

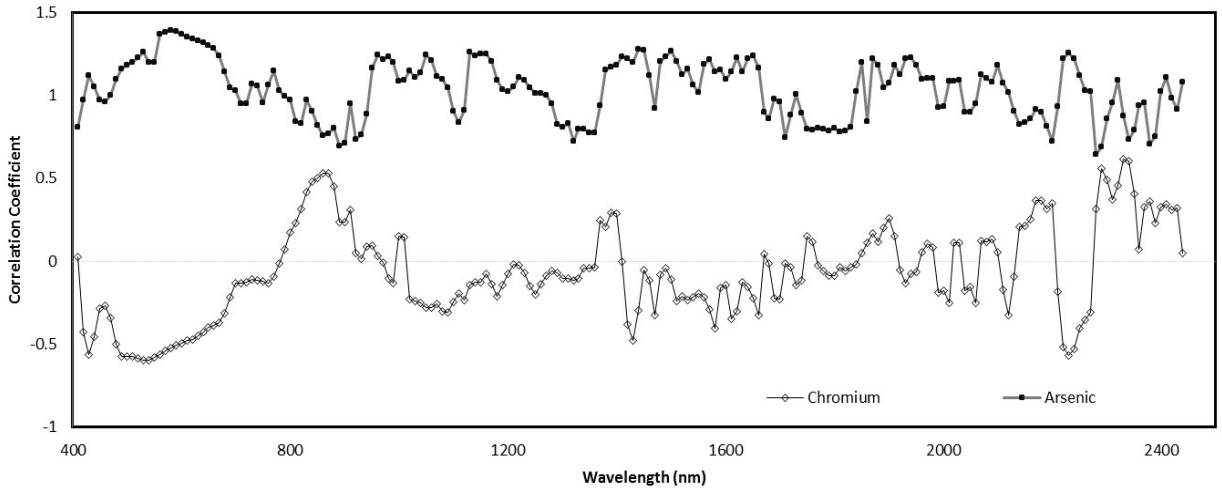

Figure 5. Correlation pattern between VNIR first derivative absorbance spectra and As and Cr contents of soil samples

In a study conducted by Fordham and Norrish, the adsorption capacity of clay minerals were investigated for adsorbing As in the absence of Fe oxy/hydroxides on their surface (Fordham & Norrish, 1983). Strong negative correlation coefficients between As and the spectral bands attributable to the absorption features of clay were also reported by (Song *et al.*, 2012).

Most of the important correlations between Chromium and VNIR first derivative absorbance spectra were observed at below 700 nm with the highest correlation coefficient at around 560 nm. This spectral range is compatible with the spectral response wavebands of Iron oxides, hence showing a strong relationship between Cadmium and iron content of the soils (Fig. 5). High correlation coefficients between Cr and spectral response wavebands of clay was also observed proving the internal relation between clay and Cr. Our observation was confirmed by calculating correlation coefficients between As, Cr and spectral features of the samples. As shown in table 2, both elements had significant correlations with iron and clay. The order of correlation with Fe_2O_3 was As>Cr while a reverse order was observed for the correlation with Al_2O_3, thought, both elements had higher correlations with Fe_2O_3. In a study of monitoring potentially toxic elements in the agricultural soils using diffuse reflectance spectroscopy, strong negative correlation coefficients between Cr and As and the spectral bands attributable to the absorption features of clay and organic matter was reported (Song *et al.*, 2012).

Samples were sorted twice based on the content of the measured Fe_2O_3 and Al_2O_3 and then graded into five and four groups respectively. For each modeling method, the MAREs between the measured and predicted concentrations of the elements in each group was then calculated (Table 3). Generally speaking, for both elements the MAREs values of all groups were lower using GA-PLSR showing higher performance comparing to PLSR. Furthermore, MAREs of As and Cr decreased by increasing Fe_2O_3 and Al_2O_3 contents indicating a positive relationship between performance of both predicting models and iron and clay contents of the samples.

Evaluation of the Prediction models

Parameters of the PLSR and GA-PLSR models are presented in table 4. The PLSR models were constructed on entire wavelengths (400–2450 nm) while fewer number of wavelengths was selected by GA (i.e. 397 and 431 for As and Cr respectively).

Based on the obtaining results, selection of variables by GA, improved the prediction accuracies for both toxic elements. Accordingly, GA-PLSR models of As and Cr increased the R^2 values by 14 and 29 percent respectively. Meanwhile, the models provided lower RMSE values of 0.149 and 0.161 for As and Cr respectively.

The higher the prediction accuracy, the closer the predicted and measured values, i.e. scatter points of predicted values vs. measured values were closer to the 1:1 line. For both toxic elements, scatter plots of PLSR and GA-PLSR models are shown in Figure 6. As can be seen, GA-PLSR outperformed the general PLSR in terms of predicting both toxic elements concentrations. One may conclude that the inherent ability of genetic algorithm in optimum selection of spectral variables was the main reason for better results obtained by GA-PLSR.

Table 2. Correlation coefficients between toxic elements and iron and clay contents of the soil samples

Element	Correlation Coefficient	
	Al_2O_3	Fe_2O_3
As	0.51	0.68
Cr	0.56	0.59

Table 3. Mean absolute relative error (MARE) for evaluating the function of spectral features in As and Cr assessment using GA-PLSR and PLSR models

Spectral Features	Group	n	MARE for PLSR		MARE for GA-PLSR	
			As	Cr	As	Cr
Fe_2O_3	#1	16	0.23	0.27	0.19	0.25
	#2	10	0.25	0.23	0.17	0.21
	#3	13	0.19	0.24	0.22	0.23
	#4	8	0.16	0.20	0.14	0.17
	#5	4	0.17	0.19	0.11	0.18
Al_2O_3	#1	21	0.31	0.27	0.28	0.25
	#2	13	0.28	0.24	0.23	0.17
	#3	9	0.25	0.24	0.21	0.23
	#4	8	0.22	0.20	0.22	0.18

n: Number of soil samples in each group

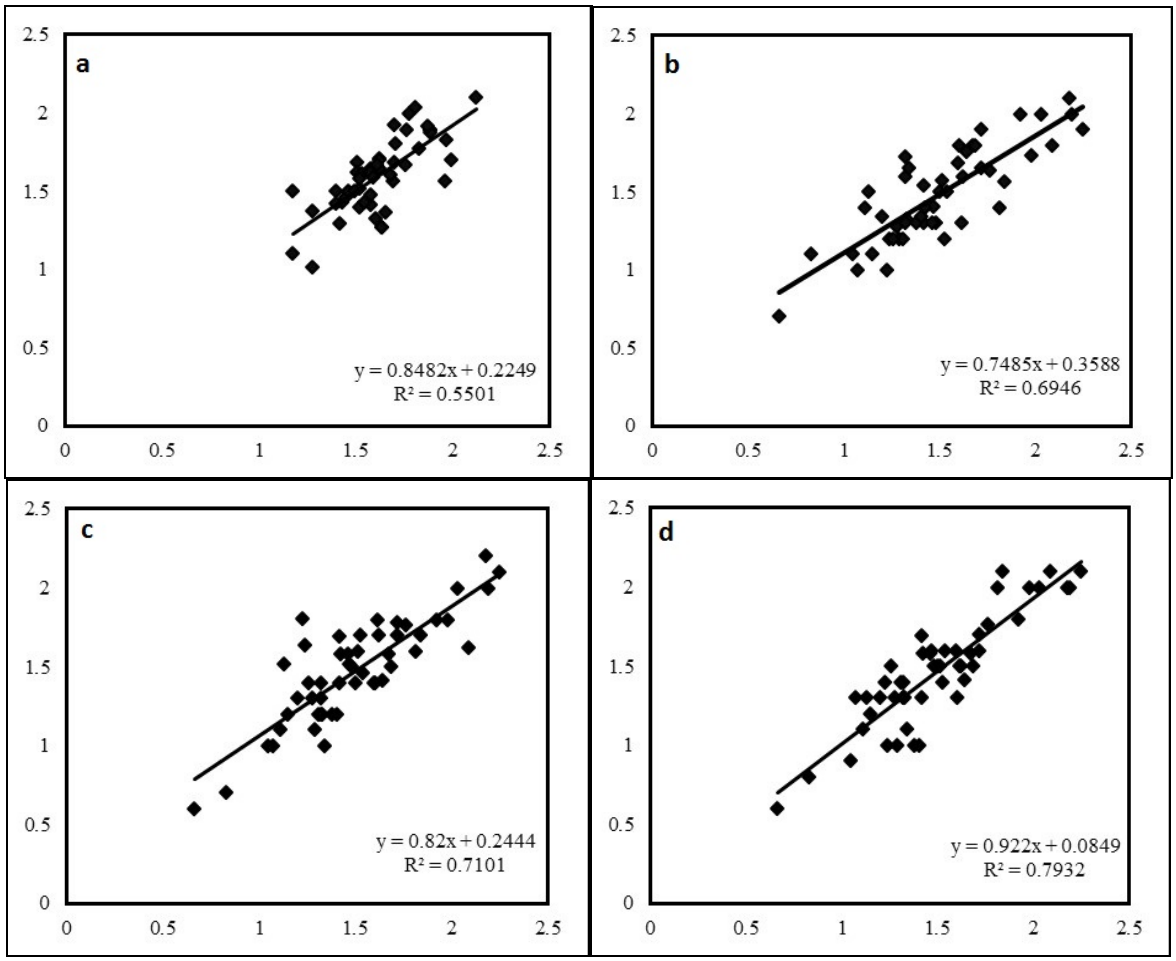

Figure 6. Scatter plots of the predicted versus measured a) log (Cr), b) log (As) using PLSR and c) log (Cr), d) log (As) using GA-PLSR

Spatial distribution

Geostatistics is an effective tool for studying the change and variance of soil properties and determining their regional distribution (Chen *et al.*, 2015). Spatial distribution maps of soil As and Cr concentrations, obtained by ordinary kriging using chemically measured and GA-PLSR predicted values are presented in Fig. 7.

Table 4. Summary Statistics of PLSR and GA-PLSR Models

Element	PLSR				GA-PLSR				
	LV	R^2	RMSE ($mg.kg^{-1}$)	RPD	Number of GA Selected Bands	LV	R^2	RMSE ($mg.kg^{-1}$)	RPD
As	4	0.69	0.165	2.12	397	4	0.79	0.149	2.33
Cr	5	0.55	0.182	1.78	431	4	0.71	0.161	2.01

LVs: Latent variables; R^2, Coefficient of Determination; RMSE, Root Mean Square Error, RPD: Ratio of Prediction to Deviation

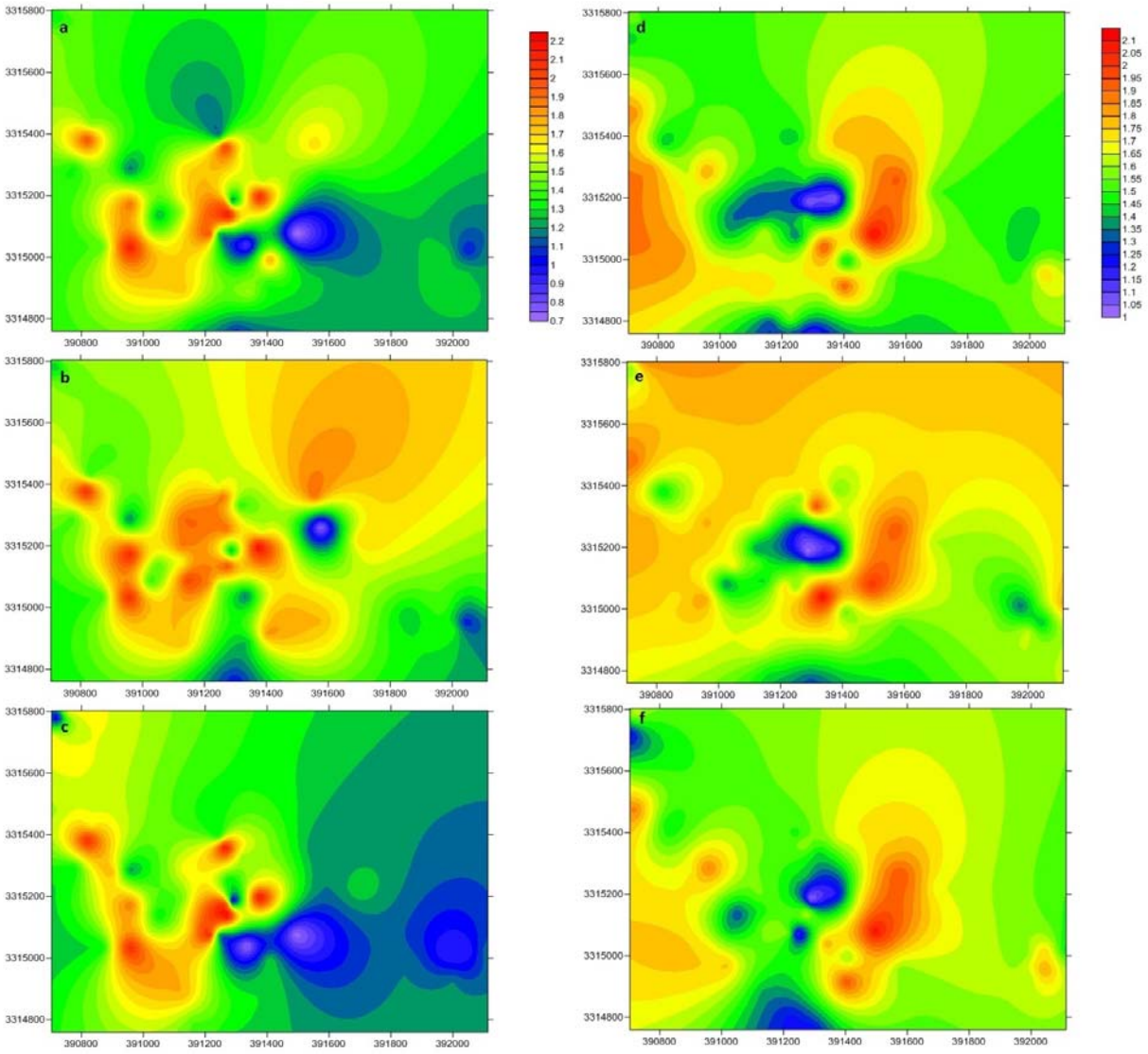

Figure 7. Spatial distribution of a) As measured, b) As PLSR predicted, c) As GA-PLSR predicted, d) Cr measured, e) Cr PLSR predicted and f) Cr GA-PLSR predicted concentrations

Similar geographical trends between measured and predicted plots were observed for both of the soil elements, although there were very slight differences. Considering Fig. 7a, b and c, similar patterns were observed as the measured and predicted As concentrations were high in the center to the left part of the dump, though the map of GA-PLSR model was much similar to that of the measured data. For Cr, high concentrations were observed in center and left of the measured and predicted plots, though, there were more areas containing high quantities of Cr in the PLSR predicted map (Fig. 7c, d and f). Regarding the obtaining plots, GA-PLSR was shown to be an effective method for determining the spatial distribution of As and Cr elements using VNIR spectra.

Methods used in this research can also be tested on other toxic elements or pollutants for future studies. Meanwhile, applying spectral parameters used in this study to multi/hyperspectral imagery can be a potential subject for further investigations on broader areas –in Sarcheshmeh and other places– for quantitative mapping of toxic elements distributions in broader affected areas. It is noteworthy to mention that spectroscopic based models have just been considered as beneficial supplements till now and further studies are needed to enhance and use them as an alternative to traditional methods.

In future studies, more reliable results could be obtained by increasing sample numbers or with other advanced spectral pre-processing techniques, such as wavelet. Applying spectral wavelengths to the multi and hyperspectral images for mapping contamination levels through remote sensing techniques could be another interesting topic for researchers.

Conclusions

As one of the very few studies, quantitative modeling of toxic elements in a mining environment was applied using spectroscopy. General and GA-PLSR models were developed for prediction of As and Cr concentrations in a dumpsite soil using rapid and cost-effective VNIR spectroscopic method. Selecting wavelengths by GA had an important role in enhancing the models resulting in higher R^2_{CV} values of 0.79 and 0.71 for As and Cr comparing to general PLSR model results. Meanwhile the $RMSE_{CV}$ values of GA was decreased about 10 and 12 percent for As and Cr respectively. The GA-PLSR model gave RPD values ranging from 2.01–2.33 for both elements while the general PLSR model provided RPD values of 1.78–2.12 using the entire soil samples. Therefore, GA-PLSR was more successful for predicting the toxic elements concentrations by means of VNIR reflectance. Mechanism allowed prediction of soil As and Cr concentrations by VNIR was mainly based on intercorrelation between toxic elements and active soil components including Fe oxides and clay content of the soil samples. Efficiency of the proposed method was also proven when similar distribution patterns were observed between spatial distribution maps of soil As and Cr contents retrieved by interpolating the measured and predicted values. As the spectroscopic modeling results seem to be in adequate agreement with measured data, further studies are needed to determine the feasibility of applying ground-derived spectral wavelengths to multi and hyper spectral images.

References

Adeline, K.R.M., Gomez, C., Gorretta, N., Roger, J.M., 2017. Predictive ability of soil properties to spectral degradation from laboratory vis-nir spectroscopy data. Geoderma, 288: 143-153.

Alloway, B.J., 1990. Heavy metals in soils. Blackie & Son Lt., 614 pp.

Araújo, S.R., Wetterlind, J., Demattê. J.A.M., Stenberg, B., 2014. Improving the prediction performance of a large tropical vis-nir spectroscopic soil library from brazil by clustering into smaller subsets or use of data mining calibration techniques. European Journal of Soil Science, 65: 718-729.

Ben-Dor, E., Banin, A., 1990. Near-infrared reflectance analysis of carbonate concentration in soils. Applied Spectroscopy, 44: 1064-1069.

Cattle, J.A., McBratney, A., Minasny, B., 2002. Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. Journal of environmental quality, 31: 1576-1588.

Chakraborty, S., Weindorf, D.C., Deb, S., Li, B., Paul, S., Choudhury, A., Ray, D.P., 2017. Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. Geoderma, 289: 72-81.

Chakraborty, S., Weindorf, D.C., Paul, S., Ghosh, B., Li, B., Ali, M.N., Ghosh, R.K., Ray, D., Majumdar, K., 2015. Diffuse reflectance spectroscopy for monitoring lead in landfill agricultural soils of india. Geoderma Regional, 5: 77-85.

Chen, T., Chang, Q., Clevers, J., Kooistra, L., 2015. Rapid identification of soil cadmium pollution risk at regional scale based on visible and near-infrared spectroscopy. Environmental pollution, 206: 217-226.

Choe, E., van der Meer, F., van Ruitenbeek, F., van der Werff, H., de Smeth, B., Kim, K.W., 2008. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the rodalquilar mining area, se spain. Remote Sensing of Environment, 112: 3222-3233.

Clark, R.N., King, T.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth, 95: 12653-12680.

Conforti, M., Castrignanò. A., Robustelli. G., Scarciglia. F., Stelluti. M., Buttafuoco. G., 2015. Laboratory-based vis–nir spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content. Catena, 124: 60-67.

Cressie N. 2015. Statistics for spatial data. John Wiley & Sons, 904 pp.

Farifteh, J., Van der Meer, F., Atzberger, C., Carranza, E., 2007. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (plsr and ann). Remote Sensing of Environment, 110: 59-78.

Fordham, A., Norrish, K., 1983. The nature of soil particles particularly those reacting with arsenate in a series of chemically treated samples. Soil Research, 21: 455-477.

Gannouni, S., Rebai, N., Abdeljaoued, S., 2012. A spectroscopic approach to assess heavy metals contents of the mine waste of jalta and bougrine in the north of tunisia. Journal of Geographic Information System, 4: 242.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: A tutorial. Analytica chimica acta, 185: 1-17.

Ghaderian, S.M., Ravandi, A.A.G., 2012. Accumulation of copper and other heavy metals by plants growing on sarcheshmeh copper mining area, iran. Journal of Geochemical Exploration, 123: 25-32.

Gholizadeh, A., Borůvka, L., Saberioon, M.M., Kozak, J., Vašát, R., Němeček, K., 2015. Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. Soil and Water Research, 10(4): 218-227.

Gholizadeh, A., Borůvka, L., Vašát, R., Saberioon, M.M., Klement, A., Kratina, J., Tejnecký, V., Drábek, O., 2015. Estimation of potentially toxic elements contamination in anthropogenic soils on a brown coal mining dumpsite by reflectance spectroscopy: A case study. PloS one, 10: e0117457.

Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M., 2006. Soil moisture prediction using support vector machines. JAWRA Journal of the American Water Resources Association, 42: 1033–1046.

Golberg, D.E., 1989. Genetic algorithms in search, optimization, and machine learning. Addion wesley Publishing Company, 412 pp.

Gomez, C., Lagacherie, P., Coulouma, G., 2008. Continuum removal versus plsr method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma, 148: 141-148.

Holland, J. H., 1992. Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. MIT Press, 211 pp.

Hong-Yan, R., Zhuang, D.F., Singh, A., Jian-Jun, P., Dong-Sheng, Q., Run-He, S., 2009. Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: A case study. Pedosphere, 19: 719-726.

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T., McBratney, A., Pallasser, R., Pozza, L., 2015. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. Geoderma, 241: 180-209.

Kemper, T., Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. Environmental science & technology, 36: 2742-2747.

Kooistra, L., Wanders, J., Epema, G., Leuven, R., Wehrens, R., Buydens, L., 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. Analytica Chimica Acta, 484: 189-200.

Kooistra, L., Wehrens, R., Leuven, R., Buydens, L., 2001. Possibilities of visible–near-infrared spectroscopy for the assessment of soil contamination in river floodplains. Analytica Chimica Acta, 446: 97-105.

Lewandowski, I., Schmidt, U., Londo, M., Faaij, A., 2006. The economic value of the phytoremediation function – assessed by the example of cadmium remediation by willow (salix ssp). Agricultural Systems, 89: 68-89.

Martens, H.A., Dardenne, P., 1998. Validation and verification of regression in small data sets. Chemometrics and intelligent laboratory systems, 44:.99-121.

Martens, H., Naes, T., 1992. Multivariate calibration. John Wiley & Sons, 423 pp.

McBratney, A.B., Minasny, B., Rossel, R.V., 2006. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. Geoderma, 136: 272-278.

McCarty, G., Reeves, J., Reeves, V., Follett, R., Kimble, J., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Science Society of America Journal, 66: 640-646.

Motelay-Massei, A., Ollivon, D., Garban, B., Teil, M., Blanchard, M., Chevreuil, M., 2004. Distribution and spatial trends of pahs and pcbs in soils in the seine river basin, france. Chemosphere, 55: 555-565.

Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. Geoderma, 158: 23-31.

Mulligan, C.N., Yong, R.N., Gibbs, B.F., 2001. Remediation technologies for metal-contaminated soils and groundwater: An evaluation. Engineering Geology, 60: 193-207.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M., 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-nir diffuse reflectance spectroscopy. Soil and Tillage Research, 155: 510-522.

Nayak, P. S., Singh, B., 2007. Instrumental characterization of clay by XRF, XRD and FTIR. Bulletin of Materials Science, 30: 235-238.

Niazi, N.K., Singh, B., Minasny, B., 2015. Mid-infrared spectroscopy and partial least-squares regression to estimate soil arsenic at a highly variable arsenic-contaminated site. International Journal of Environmental Science and Technology, 12: 1965-1974.

Pandit, C.M., Filippelli, G.M., Li, L., 2010. Estimation of heavy-metal contamination in soil using reflectance spectroscopy and partial least-squares regression. International Journal of Remote Sensing, 31: 4111-4123.

Rossel, R.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma, 158: 46-54.

Rossel, R.V., Walvoort, D., McBratney, A., Janik, L.J., Skjemstad, J., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties, Geoderma. 131: 59-75.

Saeys, W., Mouazen, A.M., Ramon, H., 2005. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. Biosystems Engineering, 91: 393-402.

Shamsoddini, A., Raval, S., Taplin, R., 2014. Spectroscopic analysis of soil metal contamination around a derelict mine site in the blue mountains, australia. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2: 75.

Shi, T., Chen, Y., Liu, Y., Wu, G., 2014. Visible and near-infrared reflectance spectroscopy-an alternative for monitoring soil contamination by heavy metals. Journal of hazardous materials, 265: 166-176.

Shi, T., Cui, L., Wang, J., Fei, T., Chen, Y., Wu, G., 2013. Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. Plant Soil, 366: 363-375.

Song, Y., Li, F., Yang, Z,, Ayoko, G.A., Frost, R.L., Ji, J., 2012. Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of changjiang river delta, china. Applied Clay Science, 64: 75-83.

Tahmasebi, P., Hezarkhani, A., 2010. Application of adaptive neuro-fuzzy inference system for grade estimation; case study, sarcheshmeh porphyry copper deposit, kerman, iran. Australian journal of basic and applied sciences, 4: 408-420.

Van Huffel, S., 1997. Recent advances in total least squares techniques and errors-in-variables modeling. Society for Industrial and Applied Mathematics, Philadelphia, 379 pp.

Vasques, G., Grunwald, S., Sickman, J., 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma, 146: 14-25.

Wang, Q., 1991. The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. Water resources research, 27: 2467-2471.

Waterman, G. C., Hamilton, R., 1975. The Sarcheshmeh porphyry copper deposit. Economic Geology, 70: 568-576.

White, W.B., 1971. Infrared characterization of water and hydroxyl ion in the basic magnesium carbonate minerals. American Mineralogist, 56: 46-53.

Wijewardane, N.K., Ge, Y., Morgan, C.L., 2016. Moisture insensitive prediction of soil properties from vnir reflectance spectra based on external parameter orthogonalization. Geoderma, 267: 92-101.

Wu, Y., Chen, J., Ji, J., Gong, P., Liao, Q., Tian, Q., Ma, H., 2007. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. Soil Science Society of America Journal, 71: 918-926.

Wu, Y., Chen, J., Wu, X., Tian, Q., Ji, J., Qin, Z., 2005. Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. Applied Geochemistry, 20: 1051-1059.

Wu, Y., Zhang, X., Liao, Q., Ji, J., 2011. Can contaminant elements in soils be assessed by remote sensing technology: A case study with simulated data. Soil Science, 176: 196-205.

Xian-Li, X., Xian-Zhang, P., Bo, S., 2012. Visible and near-infrared diffuse reflectance spectroscopy for prediction of soil properties near a copper smelter. Pedosphere, 22: 351-366.

Zhuang, D-f., 2009. Analysis of visible and near-infrared spectra of as-contaminated soil in croplands beside mines. Spectroscopy and Spectral Analysis, 29: 114-118.