

VISIBLE AND NEAR-INFRARED SPECTROSCOPY FOR MONITORING POTENTIALLY TOXIC ELEMENTS IN RECLAIMED DUMPSITE SOILS OF THE CZECH REPUBLIC

A. Gholizadeh, L. Borůvka, and R. Vašát

*Department of Soil Science and Soil Protection
Faculty of Agrobiolgy, Food and Natural Resources
Czech University of Life Sciences Prague
Prague, Czech Republic*

M.M. Saberioon

*Smart Farming Technology Research Center
Faculty of Engineering
University Putra Malaysia
Serdang, Malaysia*

ABSTRACT

Due to rapid economic development, high levels of potentially harmful elements and heavy metals are continuously being released into the brown coal mining dumpsites of the Czech Republic. Elevated metal contents in soils not only dramatically impact the soil quality, but also due to their persistent nature and long biological half-lives, contaminant elements can accumulate in the food chain and can eventually endanger human health. Conventional methods for investigating potentially harmful element contamination of soil based on raster sampling and chemical analysis are time consuming and relatively expensive. Visible and Near-Infrared (Vis-NIR) diffuse reflectance spectroscopy provides a rapid and inexpensive tool to simultaneously and accurately predict various soil properties. In this study concentrations of Manganese (Mn), Copper (Cu), Cadmium (Cd), Zinc (Zn), Iron (Fe), Lead (Pb) and Arsenic (As) in soil samples from fields near the brow coal mining dumpsites in the Czech Republic were chemically analyzed and the suitability of Vis-NIR diffuse reflectance spectroscopy for predicting their occurrence was evaluated. Soil spectral reflectance was measured with an ASD FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Inc., USA) under laboratory conditions and the correlations between seven toxic elements and soil diffuse reflectance spectra were studied. Partial Least Square Regression (PLSR) and Support Vector Machine Regression (SVMR) models were constructed to relate soil contaminants data to the reflectance spectral data by applying first and second derivatives preprocessing strategies. Then, the performance of Vis-NIR calibration models was evaluated by

Residual Prediction Deviation (RPD) and coefficients of determination (R^2). Based on the correlation patterns with reflectance spectra, the seven studied potentially toxic elements were categorized into two or three groups. Moreover, according to the criteria of minimal RPD and maximal R^2 , the first derivative and SVMR models provided more accurate prediction models for soil contaminants than PLSR models which were more feasible to predict the toxic metal levels in agricultural soils. Overall, this study indicated that the Vis-NIR reflectance spectroscopy technique combined with a continuously enriched soil spectral library as well as a suitable chemometric indicator could be a nondestructive alternative for monitoring of the soil environment. Because soil properties in contaminated areas generally show strong variation, a comparatively large number of calibrating samples, which are variable enough and uniformly distributed, are necessary to create more accurate and robust Vis-NIR reflectance spectroscopy calibration models. Future studies with real-time remote sensing data and field measurements are also strongly recommended.

Keywords: Soil contamination, Visible and Near-Infrared, Partial least square regression, Support vector machine regression, Data preprocessing.

INTRODUCTION

Among the various geo-environmental impacts of mining, heavy metal contamination of soil is by far the most significant effect. Elevated metal contents in soils not only dramatically impact on soil quality, but also due to their persistent nature and long biological half-lives, heavy metals can accumulate in the food chain and can eventually influence human health (Xie et al., 2012). Although the adverse effects of heavy metals have long been known and exposure to heavy metals continues and is even increasing in some areas, are now abandoned without particular safety measures and their environmental impact has been poorly studied. Consequently, monitoring of soil near metal smelter and mines is essential for environmental risk assessment and remediation.

Depending on large-scale sampling and physical or conventional analysis techniques, heavy metals concentrations in soils can be measured, but the negative point is that they are time-consuming, less efficient, and much more expensive when applied at a large scale in contaminated lands (Ren et al., 2009). In practice, sampling density and analytic diversity are frequently less than sufficient due to significant cost of analyses.

Diffuse reflectance spectroscopy technique is low cost with little or no sample preparation, and has been considered as an alternative to complement conventional soil analytical methods (Gholizadeh et al., 2013). It has shown to be a powerful tool for such studies in agricultural applications because it allows knowing the state of soil, providing results in real-time and on-site of interest due to its portability. Some researchers used Visible (Vis) and Near Infrared (NIR), ranging from 350 nm to 2500 nm to analyze the spectrally active properties of

sediment and soil samples. Toxic elements in soils can often be absorbed or bound by these spectrally active constituents (Song et al., 2012). This makes it possible to study the characteristics of metals in soils using Vis and NIR spectroscopy (Wu et al., 2005). Kemper and Sommer (2002) successfully used reflectance spectroscopy to estimate Arsenic (As), Iron (Fe), Mercury (Hg), Lead (Pb), Sulfur (S) and Antimony (Sb) contents in the Aznalcollar Mine area in Spain. Bray et al. (2009) also used Vis, NIR and Mid-Infrared (MIR) reflectance spectroscopy calibrated by ordinal logistic regression for the screening of either contaminated or uncontaminated soil at different thresholds for Copper (Cu), Zinc (Zn), Cadmium (Cd) and Pb. They believe that Vis-NIR can substantially decrease both the time and cost associated with screening for heavy metals.

Chemometrics methods are often needed to analyze the spectra characteristics of soil (Martens and Naes, 1989). Using a set of well-known calibration methods makes this process feasible. Choosing the most robust calibration technique can help to achieve a more reliable prediction model. Multiple Linear Regression (MLR) (Dalal and Henry, 1986), Principle Component Regression (PCR) (Pirie et al., 2005) or Partial Least Squares Regression (PLSR) (Song et al., 2012) have been used in the past to build models for estimating the toxic elements content of soil or sediments. All above-mentioned calibration methods require the creation of robust and generalized models due to their potential tendency to over-fit the data (Gholizadeh et al., 2013). Therefore; using a method such as Support Vector Machine Regression (SVMR) that can overcome the problems of other calibration methods seems essential.

To the best of our knowledge, SVMR technique has not yet been commonly used to analyse soil contamination, in the spectral domain. Therefore, this study was conducted to assess selected heavy metals namely, Manganese (Mn), Cu, Cd, Zn, Fe, Pb and As concentrations in anthropogenic soils on brown coal mining dumpsites, to evaluate the feasibility of Vis-NIR technique in the rapid prediction of above-mentioned contaminants in the soils and to compare the performance of PLSR and SVMR methods for multivariate calibrations using soil reflectance spectra. It was envisaged that this rapid and inexpensive method for obtaining accurate information of heavy metals would be valuable to provide reference data for soil environment monitoring by remote sensing.

MATERIALS AND METHODS

Study Area and Soil Sampling

Six dumpsites in Czech Republic were selected: Pokrok, Radovesice, Březno, Merkur, Prunéřov and Tumerity. All are formed by clays. On a part of each dumpsite, cover with natural topsoil was spread in an amount of approximately 2500 to 3000 t per ha one year before sampling. The topsoil material originated from humic horizons of natural soils of the region, particularly Vertisols, partly also Chernozems (clayic and haplic). The topsoil was not mixed with the dumpsite material. Disturbed and undisturbed soil samples were collected on all dumpsites randomly. Approximately half of the sampling points were located on the area with natural topsoil cover, half of the points were on the area without the

cover. Sampling was made in the 0 to 20 cm layer. This depth corresponds to common depth of ploughing soil layer, as these soils should be used as arable land in future. The depth of the topsoil cover, where it was applied, was also at least 20 cm.

Soil Analysis

The samples were air-dried and sieved through a 2 mm mesh. All samples were then saved for analyses of heavy metals (including Cu, Mn, Cd, Zn, Fe, Pb and As) and reflectance measurements. Total concentrations of heavy metals were determined by digesting soil samples (< 0.149 mm fraction) with a mixture of concerned hydrochloric and nitric acids (4:1, v:v) (Xie et al., 2012) and then analyzed by inductively coupled plasma mass spectrometry.

Reflectance Spectroscopy Measurement

The reflectance was measured in 350-2500 nm wavelength range by a FieldSpec 3 spectroradiometer (Analytical Spectral Devices Inc., USA) with contact probe. The spectral resolution of the spectroradiometer was 3 nm for the region 350-1000 nm and 10 nm for the region 1000-2500 nm. A fiber-optic probe with 8° field of view was used to collect reflected light from the sample. The probe was mounted on a tripod and positioned about 10 cm vertically above the sample. The sample dish was over-filled with soils and then leveled off using a blade to ensure a flat surface flush with the top of the dish. The final spectrum was an average based on 20 iterations from 4 directions with 5 iterations per direction to increase the signal-to-noise ratio.

Model Construction and Validation

For all samples, an exploratory analysis was carried out to detect outliers before establishing the regression model (Wu et al., 2005).

The correlation between heavy metals concentration and spectra reflectance was determined using Pearson's correlation. Univariate regression models were built for monitoring Mn, Cu, Cd, Zn, Fe, Pb and As, using reflectance in the Vis-NIR region.

In order to calibrate a model that provides accurate predictive performance about the quantity of heavy metals contained in each soil sample, the captured soil spectra together with laboratory data of heavy metals were imported into R software (R Development Core Team, 2011) to be processed. From a total of 264 samples taken for laboratory analysis, mostly subsets were used to determine the content of heavy metals due to cost constraints. Number of samples subjected for individual analysis was then as follows: the entire data were tested for Mn and Fe; 148 samples were tested for Pb; 115 for Cu and Zn and 104 samples for Cd and As. Spectral preprocessing techniques are a variety of mathematical methods for correcting light scattering in reflectance measurements and data enhancement before the data was used in calibration models. The first derivative is very effective for removing baseline offset; the second derivative is very effective for

both the baseline offset and linear trend from a spectrum (Rinnan et al., 2009). In this study, prior to all further spectra treatments, the noisy part of the spectra range (350-399 nm) was cut out and then the spectra were subjected to Savitzky-Golay smoothing with a second-order polynomial fit and 11 smoothing points (Ren et al., 2009; Song et al., 2012) in order to remove the artificial noise caused by the spectroradiometer instrument. The predictive models were fitted based on smoothed raw spectra first, and then using two types of preprocessed spectra as made with first and second derivative manipulation which were calculated using the Savitzky-Golay algorithm as well. Moreover, PLSR and SVMR models were employed to calibrate spectral data with chemical reference data and to establish relations between reflectance spectra and measured heavy metals.

Based on Viscarra Rossel et al. (2006a), PLSR handles multicollinearity, it is robust in terms of data noise and missing values, and in spite of PCR it balances the two objectives of explanation response and predictor variation (thus calibrations and predictions are more robust) and it presents the decomposition and regression in a single step. PLSR models were fitted with the *pls* R package (Mevik and Wehrens, 2007) using the classical orthogonal scores algorithm. To determine the optimal number of components the Root Mean Squared Error of Prediction (RMSEP) and Residual Prediction Deviation (RPD), ratio of the standard deviation and RMSEP, were minimized by leave-one-out cross-validation (Xie et al., 2012).

The concept of SVMR follows a different approach of supervised learning. Its algorithm is based on the statistical learning theory (Vohland et al., 2011). It has been known to strike the right balance between accuracy attained on a given finite amount of training patterns and the ability to generalize to unseen data. The most valuable properties of SVMs are their ability to handle large input spaces efficiently, to deal with noisy patterns and multi-modal class distributions, and their restriction on only a subset of training data in order to fit a non-linear function (Gholizadeh et al., 2013). For SVMR prediction we used radial basis function kernel contained in *e1071* R package (Meyer et al., 2012).

Accuracy Assessment of Techniques

Assessment of the prediction accuracy of the models was carried out using a leave-one-out cross-validation approach (R^2_{cv} and $RMSEP_{cv}$) and also by the values of R^2 and RPD. Root Mean Square Error of Cross-Validation ($RMSEP_{cv}$) is the factor, by which the prediction accuracy has been increased, compared to the mean composition for all samples (Mouazen et al., 2010). The $RMSEP_{cv}$ was computed as follows:

$$RMSEP_{cv} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2}$$

where y'_i is the predicted and Y_i is the observed value.

3. RESULTS AND DISCUSSION

Soil Samples Descriptive Statistics

General statistical results of heavy metals in the six dumpsites are summarized in Table 1.

Table 1. Descriptive statistics of heavy metals in the studied sample set according to location

Item	Cu	Mn	Fe	Cd	Pb	Zn	As
mg/kg							
<i>Pokrok</i>							
<i>n</i>	103	103	103	103	103	103	103
Min	5.50	198.30	2503.41	0.01	7.60	8.30	0.49
Max	35.70	869.14	9752.64	0.73	42.40	127.10	19
Mean	13.76	599.42	5418.02	0.27	18.43	25.26	4.48
Std.	3.58	118.61	1330.14	0.11	5.32	15.77	3.39
C.V.(%)	26	20	25	40	29	62	76
<i>Radovesice</i>							
<i>n</i>	40	40	40	40	40	40	40
Min	6.42	254.12	1754.37	0.03	4.70	9.38	0.18
Max	22.10	844.12	6876.94	0.30	49.60	66.85	1.30
Mean	14.20	541.31	4489.27	0.17	13.70	21.98	0.67
Std.	3.45	125.13	974.41	0.05	6.40	11.15	0.25
C.V.(%)	24	23	22	30	47	51	38
<i>Březno</i>							
<i>n</i>	25	25	25	25	25	25	25
Min	9.01	473.34	2398.50	0.00	10.90	11.49	0.49
Max	38.81	885.84	31281.81	0.37	21.60	200.27	5.89
Mean	14.37	680.98	9967.19	0.16	14.17	41.50	1.12
Std.	5.95	105.93	103.58.19	0.11	2.97	41.62	1.04
C.V.(%)	41	16	104	64	21	100	93
<i>Merkur</i>							
<i>n</i>	38	38	38	38	38	38	38
Min	7.29	317.99	2361.80	0.04	9.30	6.95	0.33
Max	16.76	787.28	8047.71	0.27	55.90	32.22	9.57
Mean	12.22	590	4852.73	0.16	17.53	13.56	0.97
Std.	1.77	100.74	1355.59	0.06	7.23	4.19	1.45
C.V.(%)	14	17	28	39	41	31	149
<i>Pruněřov</i>							
<i>n</i>	48	48	48	48	48	48	48
Min	8.40	41.59	2105.06	0.00	0.90	6.60	0.00
Max	92.24	984	9225.44	0.24	24.80	213.11	3.30
Mean	15.81	552.62	5532.46	0.11	14.38	26.83	0.98
Std.	14.36	224.37	1595.54	0.06	4.82	39.32	0.86
C.V.(%)	91	41	29	55	34	147	87
<i>Tumery</i>							
<i>n</i>	10	10	10	10	10	10	10
Min	12.29	496.78	4163.75	0.00	9.50	15.50	0.37
Max	20.34	1027.64	8484.29	0.20	14.50	48.56	0.51
Mean	15.03	753.14	6702.30	0.12	12.25	25.61	0.42
Std.	2.40	192.30	1426.63	0.05	1.38	10.32	0.05
C.V.(%)	16	26	21	44	11	40	12

The comparison of Coefficients of Variation (C.V.) of different contaminants showed that among all parameters As had the highest C.V., especially in Merkur

area (149%) that means it has varied the most as compared to other measured parameters. Likewise, Pb in Tumerity (11%) had the lowest C.V. which shows it is more homogenous than the other properties.

In the study area, the estimated mean concentration of Cd (0.27 mg/kg), Pb (18.43 mg/kg) and As (4.48 mg/kg) in Pokrok was higher than other locations that might be a serious threat for agronomic practices on the soils beside the mining area. The mean concentration of Fe (9967.19 mg/kg) in the studied soil sample of Březno was also high, which was probably related to the iron oxide-rich characteristic of the soil type and its forming.

Vis-NIR Reflectance Spectroscopy of the Soil Samples and Data Preprocessing

A visual inspection of the spectra allowed detecting some spectral readings possibly affected by measurement errors. These were removed and the final spectral library had a total of 264 soil spectra. Raw reflectance, smoothed spectra by Savitzky-Golay and first and second derivative spectra of all selected soil samples in one of the locations (Pokrok) which had the most samples, are shown in Fig. 1. Other locations also showed the same trend.

Sets of spectra were characterized qualitatively by observing the positive and negative peaks (Fig. 1), which occur at specific wavelengths (Viscarra Rossel et al., 2006a). Due to the presence of the same spectrally active properties in all locations, the Vis-NIR spectra of all soil sample sets were similar. The characteristic wavebands of reflectance spectra were only around 1400, 1900, and 2200 nm. However, there were more features of high variability at around 460-550, 1400, 1900-2000 and 2200 nm in the first derivative. The second derivative more or less showed the same spectrum in all locations. Stenberg et al. (2010) also indicated that the first and second derivatives were by far the most popular spectral preprocessing techniques for soil property prediction using the Vis and NIR spectroscopy.

Generally, from the Fig. 1, it can be understood that three essential absorption bands are evident throughout all the compressed spectra (around 1400, 1900 and 2200 nm). Also, the general shape and slopes of the all curves are similar. The regions around 1400 and 1900 nm were related to vibrational frequencies of OH groups in the water and hydroxyl absorption, and the features around 2000-2500 nm were related to the characteristics of soil organic matter and clay minerals (Ren et al. 2009). Thus, preprocessed spectra highlighted more spectral features in the Vis region and made reflectance spectra narrower and sharper in the NIR region. Based on Song et al. (2012), although intense bands in the Vis-NIR spectra are not directly associated to the presence of metals or other constituents, it is clear that metals can interact with the main spectrally active components of soil. Based on this phenomenon, chemometrics models can be developed for soil samples in order to screen their toxic elements concentrations. Similar results were reported by Wu et al. (2005), Ren et al. (2009) and Song et al. (2012).

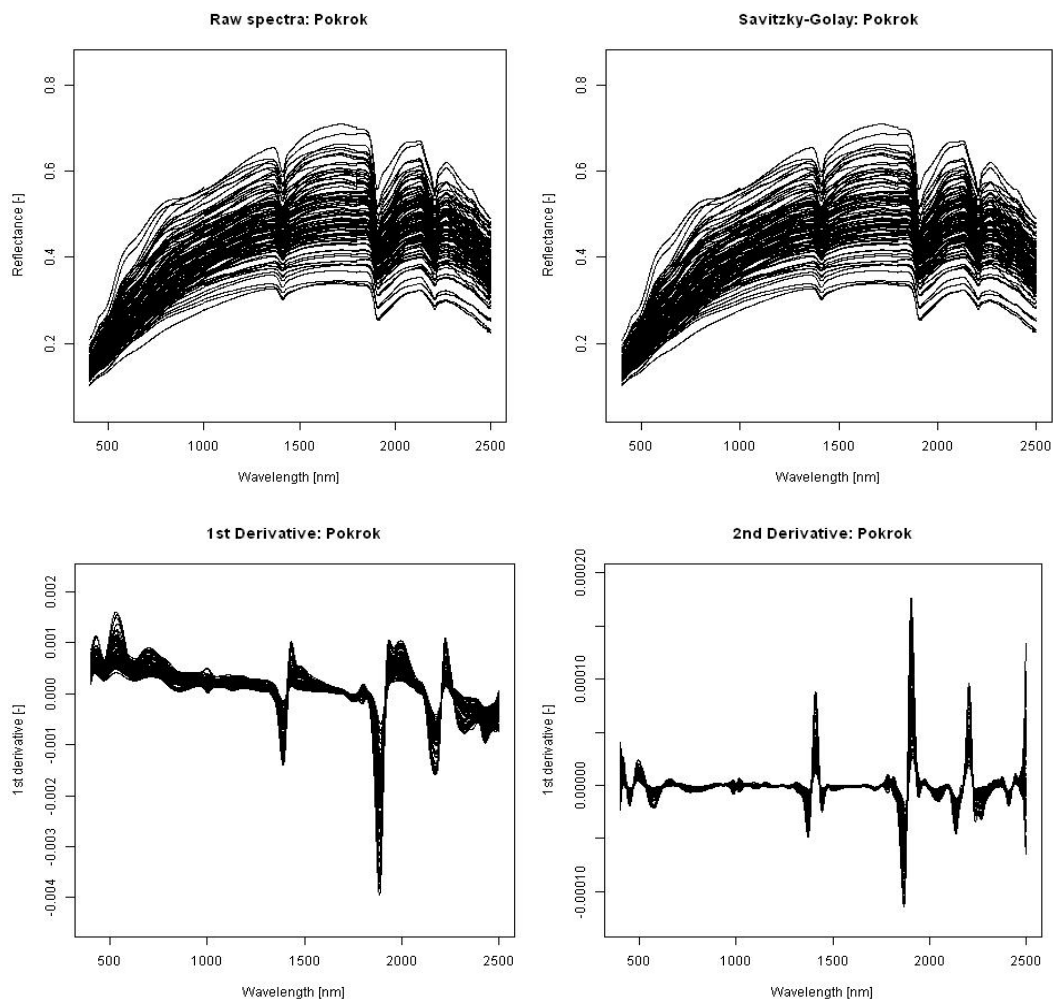


Fig. 1. Raw reflectance spectra, smoothed spectra by Savitzky-Golay and preprocessed spectra of soil samples for Pokrok

Matrix Correlation of Heavy Metals and Reflectance Spectra

Same as Song et al. (2012), linear correlation coefficients between reflectance and heavy metals were moderately high especially in Březno ($-0.6 < r < 0.6$) throughout the Vis and NIR regions (Fig. 2). This indicates that heavy metals do contribute to the reflectance of almost all wavelengths. Fig. 2 also shows that the concentrations of toxic elements in six dumpsites soil samples displayed complex changes in their correlations with the Vis-NIR reflectance of soil spectra. Moreover, it can be seen that each metal exhibits its maximum correlation coefficient at a different wavelength. Correlation analysis also indicates that the correlation coefficients of the two elements Cd and Pb are usually separated from the other elements.

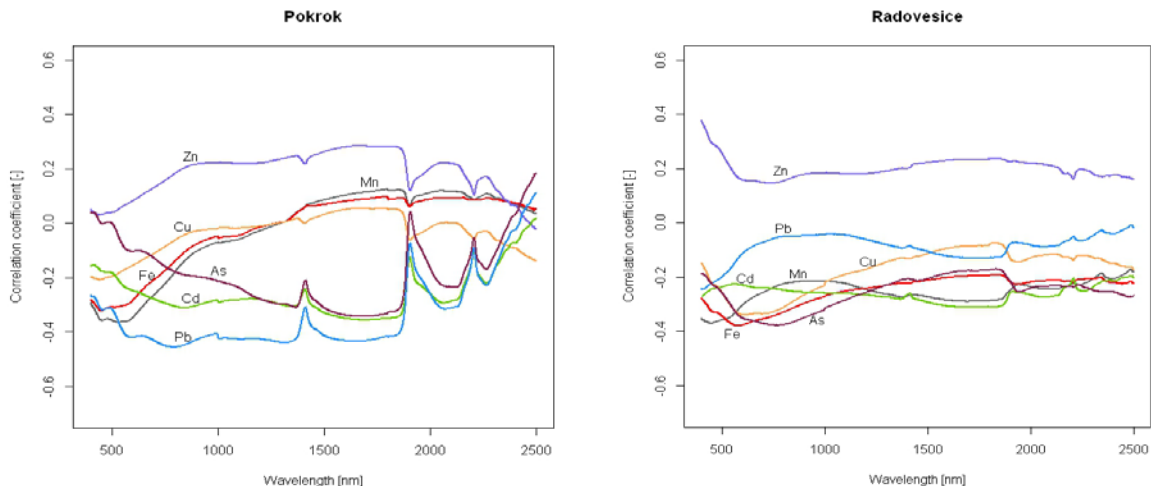
In each location the heavy metals could be categorized into two or three groups according to their behavior and relationships with soil Vis-NIR spectra. In Pokrok, the toxic elements categorized to two groups. The first group of metals

(Cd, Pb and As) had stronger negative correlation coefficients with spectral bands than the second group (Cu, Zn, Fe and Mn), which first group displaying the highest negative spectral correlation at 786 nm in Pb and the second group had the strongest correlation at 1667 nm in Zn (Fig. 2). However, correlation coefficients changes with Vis-NIR spectra of Prunéřov dumpsite, in contrast to dumpsite Pokrok categorized into three groups namely (Mn, As and Fe), (Zn and Cu) and (Cd and Pb), but it also displayed the strongest negative correlation in Pb (at 513 nm) and strongest correlation in Zn (at 769 nm).

In Radovesice, the highest positive and also the highest negative spectral correlations both can be seen in the first group of elements (Zn, Cu, As and Fe), which the strongest positive correlation coefficient related to Zn at 401 nm and Fe represented the lowest spectral correlation at 578 nm arising from Fe³⁺ absorption. These results were similar to results of Ben-Dor (2002). They mentioned that the contribution of the region 390-550 nm is attributed to the spectral absorption features of free iron oxides. Březno dumpsite also exhibited the highest positive and negative correlation coefficients in Vis region, at 401 nm (Zn) and 433 nm (Cd), respectively. Correlation coefficients changes of all metals in Merkur dumpsite exhibited similar behavior to Vis-NIR spectra and categorized in one group. Similar to results of Vohland et al. (2009), the order of the correlation coefficients between the metals in this dumpsite was Cd > Zn > As > Pb > Mn > Cu > Fe, and the highest positive and negative correlation coefficients belonged to Vis Region again (561 nm to 651 nm, respectively).

Clearly, correlation changes in the Vis region of the first group of heavy metals (Fe, As, Cu and Zn) in Tumerity dumpsite fell into 410-540 nm and correlation changes of Cd which categorized in the second group of heavy metals (Cd and Pb) fell into the NIR region with the highest negative correlation coefficient at 1913 nm (around water absorption band). Stenberg et al. (2010) also reported the same results.

The earlier report of Song et al. (2012) indicated the similar results in the agricultural soils of Changjiang River Delta, China. These findings provide support for the use of diffuse reflectance spectra in predicting the heavy metal contents of soil samples.



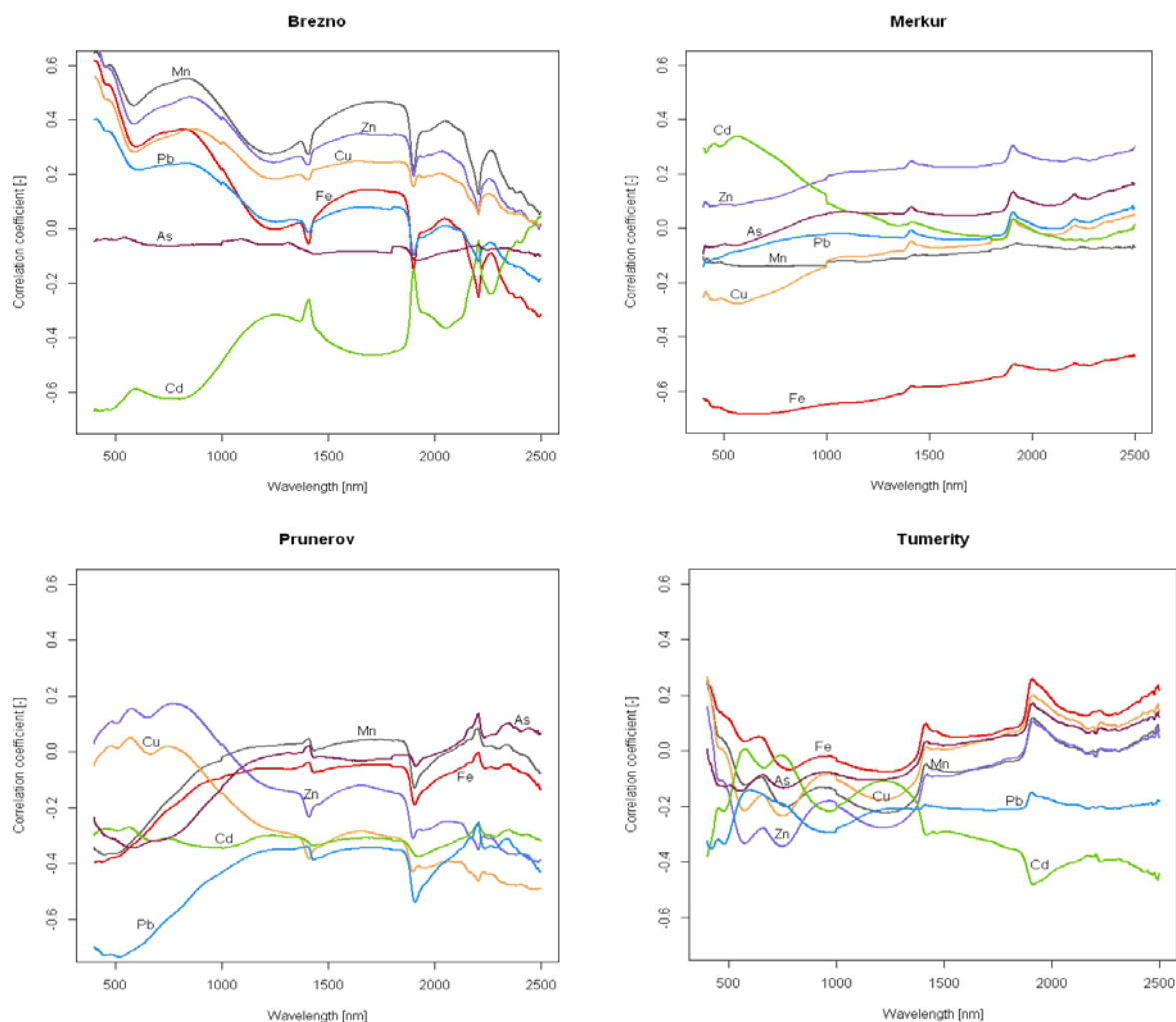


Fig. 2. Correlation between reflectance of Vis-NIR and heavy metals in different locations

Multivariate Analysis Using PLSR and SVMR and Validation Test

First derivative technique was selected as the most suitable preprocessing technique. Multivariate calibration techniques such as PLSR and SVMR have been used to extract soil heavy metals calibration models from the reflectance spectra of soils in the Vis and NIR. The adequacy of each calibration model was evaluated based on the value of R^2 and the RPD (Mouazen et al., 2007). In fact, R^2 indicates the percentage of the variance in the Y variable that is accounted for by the X variables. An R^2 value between 0.50 and 0.65 indicates that more than 50% of the variance in Y is accounted for by variable X, so that discrimination between high and low concentrations can be made. An R^2 between 0.66 and 0.81 indicates approximate quantitative predictions, whereas, an R^2 between 0.82 and 0.90 reveals good prediction. Calibration models having an R^2 above 0.91 are considered to be excellent (Williams, 2003). Moreover, RPD is the factor, by

which the prediction accuracy has been increased compared to using the mean composition for all samples (Mouazen et al., 2010). Viscarra Rossel et al. (2006b) classified RPD values as follows: $RPD < 1.0$ indicates very poor model and their use is not recommended; RPD between 1.0 and 1.4 indicates poor model where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates fair model which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicate good model where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates very good, quantitative model and $RPD > 2.5$ indicates excellent predictions. This classification system was adopted in this study.

As can be seen in Table 2, the two modeling strategies considered in this study provide different prediction accuracy of the studied heavy metals. For the PLSR calibration set, R^2 values ranged between 0.48 and 0.68. Good and excellent prediction R^2 ($R^2 > 81$ and $R^2 > 90$, respectively) were not obtained for none of the seven elements. The best predictive models were obtained for As ($R^2 = 0.65$), following by Cd ($R^2 = 0.60$) and inadequate models ($R^2 < 0.50$) were obtained for Fe. Furthermore, according to the Viscarra Rossel et al. (2006b) classification, it is clear that the prediction of As can be evaluated as good ($1.8 < RPD < 2.0$) and prediction of Fe ($1.0 < RPD < 1.4$) is not satisfactory, whereas the predictions model of Cu, Mn, Cd, Pb and Zn can be introduced as a fair model. The large variability of the sample set (colour and texture of samples from different dumpsites) used in this study also affects the accuracy of PLSR calibration models developed for elements.

Malley and Williams (1997) first predicted heavy metals in freshwater sediment samples using NIR calibration method, with 119 samples selected as a calibration set. They reported very good predictions for Cu ($R^2 = 0.91$, $RPD = 3.35$), Pb ($R^2 = 0.81$, $RPD = 2.45$) and Zn ($R^2 = 0.93$, $RPD = 3.80$), but the prediction for Cd ($R^2 = 0.63$, $RPD = 1.74$) was less satisfactory. Kemper and Sommer (2002) reported excellent prediction for Pb ($R^2 = 0.94$, $RPD = 5.89$) and bad predictions for Cu, Zn, and Cd ($R^2 = 0.22-0.49$). Malley et al. (2004) reported acceptable prediction for Cu ($R^2 = 0.69$, $RPD = 1.8$) in 829 samples from ago-Manitoba. In the study of Xie et al. (2012), the models provided fairly accurate predictions for Fe ($R^2 > 0.80$, $RPD > 2.00$), less accurate but acceptable for screening purposes for Cu, Pb, and Cd ($0.50 < R^2 < 0.80$, $1.40 < RPD < 2.00$) and poor accuracy for Zn ($R^2 < 0.50$, $RPD < 1.40$). Due to this variability researchers tended to develop calibration models for each field they measured with Vis and NIR spectroscopy (Imade Anom et al., 2000; Mouazen et al., 2005). Moreover, Dunn et al. (2002) indicated that the poor predictive ability of Vis and NIR for many soil constituents might result from a poorly distributed sample set with a small range, rather than the inability of Vis and NIR to predict the soil property. Besides sample variation, sample distribution and sample size are all critical to a successful Vis-NIR calibration (Xie et al., 2012).

To the best of our knowledge, the SVMR technique has not yet been commonly used to analyse and prediction of heavy metals, in the spectra domain. In the current work, SVMR was also used to develop prediction models. The results of the SVMR model for Cu, Mn, Fe, Cd, Pb, Zn and AS in Vis-NIR spectra are shown in Table 2. Among heavy metals studied, As is the most

accurately measured with SVMR ($R^2 = 0.95$, $RPD = 2.63$). This prediction accuracy is classified to be excellent. The calibration results for Mn was not as good as the results of the other elements ($R^2 = 0.65$, $RPD = 1.75$) and indicates a fair model. Results obtained with SVMR for Cu, Fe, Cd, Pb and Zn is classified to be good or very good, although the prediction accuracy of Fe and Pb is slightly lower than those of Cu, Cd and Zn.

Table 2. Statistics results for calibration and cross-validation of the Vis-NIR diffuse reflectance spectroscopy for each heavy metal

Item	<i>n</i>	PLSR					SVMR				
		R^2	RMSEP	R^2_{cv}	RMESP _{cv}	RPD	R^2	RMSEP	R^2_{cv}	RMESP _{cv}	RPD
Cu	115	0.56	5.89	0.50	6.28	1.45	0.84	3.47	0.78	4.08	2.29
Mn	264	0.54	102.81	0.44	116.43	1.45	0.65	90.26	0.58	101.25	1.75
Fe	264	0.48	1616.42	0.48	1619.03	1.32	0.77	1053.89	0.71	1141.08	2.04
Cd	104	0.60	0.06	0.57	0.11	1.68	0.85	0.03	0.78	0.08	2.31
Pb	148	0.58	2.58	0.51	3.12	1.50	0.73	1.75	0.66	2.24	1.97
Zn	115	0.54	17.64	0.45	21.84	1.42	0.80	10.23	0.71	14.51	2.16
As	104	0.68	2.13	0.61	2.98	1.81	0.95	1.13	0.89	1.89	2.63

Table 2 indicates that in validation procedure, cross-validation R^2 of PLSR ranged between 0.44 for Mn and 0.61 for As, while this range for SVMR was between 0.58 for Mn to 0.89 again for As. Based on R^2_{cv} and RMESP_{cv}, in both calibration and validation, the best estimates were clearly obtained for As prediction. Generally, R^2_{cv} and RMESP_{cv} for both methods were satisfactory but same as calibration, SVMR results were more reliable which emphasizes the need of using more flexible techniques such as SVMR.

In total, by comparing the results of the PLSR and SVMR models for the Vis-NIR spectra, it can be seen that PLSR have been successfully used to calibrate many soil variables including some heavy metal concentrations (Xie et al., 2012; Song et al., 2012). In this study, in addition to PLSR which showed fairly good predictions, SVMR provided very good correlations between soil spectra and various heavy metals; better prediction was achieved using SVMR and it outperformed the PLSR. From a practical point of view, the prediction accuracies obtained with these two methods generally seem to be acceptable for a number of agricultural applications including soil science research. Support Vector Machine Regression (SVMR)' superior performance over PLSR can be explained by the inclusion of nonlinear and interactions effects as well as linear combinations of variables, it is able to approximate nonlinear functions between multidimensional spaces (Stevens et al, 2010, Gholizadeh et al., 2013).

CONCLUSION

This study demonstrated the application of laboratory Vis-NIR reflectance spectroscopy for prediction of heavy metals including Cu, Mn, Cd, Zn, Fe, Pb and As, using soil samples taken from six brown coal mining dumpsites of Czech Republic. For each parameter, Vis-NIR calibration models were created by PLSR and SVMR algorithms. Correlation analysis revealed that heavy metals do contribute to the reflectance of almost all wavelengths and the correlation

coefficients of the two elements Cd and Pb are usually separated from the other elements. The results showed obvious differences in predictability and accuracy of PLSR and SVMR. Soil spectroscopy in the Vis-NIR region with a SVMR model is shown to be a very promising method for the determination of metal concentrations in anthropogenic soils. The best predictability of Vis-NIR reflectance spectroscopy was obtained by SVMR for As ($R^2 > 0.90$; RPD > 2.5), followed by Cd, Cu, Zn, Pb and Mn. Generally, our results confirmed that Vis-NIR reflectance spectroscopy combined with first derivative and SVMR methods have a great potential for site-specific soil monitoring in high-risk regions and lead to overoptimistic performance in the assessment of heavy metals, which generally involves conducting large numbers of analyses in a short time. For future investigations, hyperspectral sensors may be useful and have to be explored for fitting specific spectral regions and for models to optimize the estimation of heavy metals content.

REFERENCES

- Ben-Dor, E. 2002. Quantitative remote sensing of soil properties. *Adv. Agron.* 75:173-243.
- Dalal, R.C., and Henry, R.J. 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Sci. Soc. Am. J.* 50:120-123.
- Dunn, B.W., Beecher, H.G., Batten, G.D., and Ciavarella, S. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis-a case study from the Riverine Plain of south-eastern Australia. *Aust. J. Exp. Agr.* 42: 607-614.
- Gholizadeh, A., Boruvka, L., Saberioon, M.M., and Vasat, R. 2013. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Appl. Spectrosc.* 67:1349-1362.
- Imade Anom, S.W., Shibusawa, S., Sasao, A., Sato, H., Hirako, S., and Blackmore, S. 2000. Moisture, soil organic matter and nitrate nitrogen content maps using the real-time soil spectrophotometer. pp. 307-312. In: *IFAC Bio-robotics, information technology and intelligent control for bio-production systems*. Sakai, Osaka, Japan.
- Kemper, T., and Sommer, S. 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ. Sci. Technol.* 36:2742-2747.
- Malley, D.F., and Williams, P.C. 1997. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environ. Sci. Technol.* 31:3461-3467.

- Malley, D.F., Martin, P.D., and Ben-Dor, E. 2004. Application in analysis of soils. pp. 729-784. In: Roberts, C.A., Workman Jr., J. and Reeves III, J. B. (eds.), Near-infrared spectroscopy in agriculture. ASA-CSSA-SSSA, Madison, WI.
- Martens, H., and Naes, T. 1989. Multivariate calibration. p. 419. New York: John Wiley and Sons.
- Mevik, B.H., and Wehrens, R. 2007. The pls package: principal component and partial least squares regression in R. *J. Stat. Softw.* 18:1-24.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F., 2012. e1071: Misc Functions of the Department of Statistics (e1071), TU Wien. R package version 1.6-1.
- Mouazen, A.M., De Baerdemaeker, J., and Ramon, H. 2005. Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. *Soil Till. Res.* 80:171-183.
- Mouazen, A.M., Maleki, M.R., De Baerdemaeker, J., and Ramon, H. 2007. On-Line measurement of some selected soil properties using a VIS–NIR sensor. *Soil Till. Res.* 93:13-27.
- Mouazen, A.M., Kuang, B., De Baerdemaeker, J., and Ramon, H. 2010. Comparison between principal component, partial least squares and artificial neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma.* 158:23-31.
- Pirie, A., Singh, B., and Islam, K. 2005. Ultra-violet, visible, near-infrared, and mid infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Aust. J. Soil Res.* 43:713-721.
- Ren, H.Y., Zhuang, D.F., Singh, A.N., Pan, J.J., Qid, D.S., and Shi, R.H. 2009. Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: A case study. *Pedosphere.* 19:719-726.
- Rinnan, A., van den Berg, F., and Engelsen, S.B. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trend. Anal. Chem.* 28:1201-1222.
- Song, Y., Li, F., Yang, Z., Ayoko, G.A., Frost, R.L., and Ji, J. 2012. Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China. *Appl. Clay Sci.* 64:75-83.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., and Wetterlind, J. 2010. Visible and near infrared spectroscopy in soil science. 107:163-215. In: Donald L. Sparks, (ed.), *Advances in Agronomy*. Burlington, VT: Elsevier.

- Stevens, A. Udelhoven, T., Denis, A., Tychon, B., Liroy, R., Hoffmann, L., and Van Wesemael, B. 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*. 158:32-45.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., and Skjemstad, J.O. 2006a. Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*. 131:59-75.
- Viscarra Rossel, R.V., McGlynn, R.N., and McBratney, A.B. 2006b. Determining the composition of mineral-organic mixes using UV-Vis-NIR diffuse reflectance spectroscopy. *Geoderma* 137:70-82.
- Vohland, M., Bossungl, B., and Frund, H.C. 2009. A spectroscopic approach to assess trace-heavy metal contents in contaminated floodplain soils via spectrally active soil components. *J. Plant Nutr. Soil Sci.* 172:201-209.
- Vohland, M., Besold, J., Hill, J., and Frund, H.C. 2011. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma*. 166:198-205.
- Williams, P. 2003. Near-infrared technology-Getting the best out of light. PDK Projects, Nanaimo, Canada.
- Wu, Y., Chen, J., Wu, X., Tian, Q., Ji, J., and Qin, Z. 2005. Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. *Appl. Geochem.* 20:1051-1059.
- Xie, X.L., Pan, X.Z., and Sun, B. 2012. Visible and near-infrared diffuse reflectance spectroscopy for prediction of soil properties near a Copper smelter. *Pedosphere*. 22:351-366.