International Journal of Environmental Science and Development, Vol. 6, No. 8, August 2015

A Spectroscopic Approach to Assess Potentially Toxic Elements of Reclaimed Dumpsites in the Czech Republic

A. Gholizadeh, L. Borůvka, and M. M. Saberioon

Abstract—**Monitoring of potentially toxic elements (PTEs) in anthropogenic soils on brown coal mining dumpsites requires a large number of samples and cumbersome and time-consuming laboratory measurements. Reflectance spectroscopy within the visible-near infrared (Vis-NIR) region has been used to predict soil constituents due to its rapidity, convenience and accuracy. This study evaluated the suitability of Vis-NIR (350-2500 nm) reflectance spectroscopy for predicting PTEs concentration, using samples collected on large brown coal mining dumpsites in the Czech Republic. partial least square regression (PLSR) and support vector machine regression (SVMR) with cross-validation were used to relate PTEs data to the reflectance spectral data by applying different preprocessing strategies. Overall, SVMR models for the Vis-NIR spectra could be indirectly used for accurate assessment of PTEs' concentrations**

Index Terms—**Partial least square regression, potentially toxic elements, preprocessing, support vector machine regression, visible-near infrared spectroscopy.**

I. INTRODUCTION

Our society and civilization now rely heavily on the mining industry to operate and maintain comfort. Among the various geo-environmental impacts of mining, contamination of soil is by far the most significant effect. Elevated concentrations of PTEs in soils not only dramatically impacts the soil quality, but also due to their persistent nature and long biological half-lives, PTEs concentrations in soils can be measured but their determination is dependent on large-scale sampling and physical or conventional analysis techniques. The negative side to this is that they are time-consuming, less efficient, and expensive when applied at a large scale in contaminated lands [1].

The diffuse reflectance spectroscopy technique is low cost with little or no sample preparation, and has been considered as an alternative to conventional soil analytical methods. Furthermore, this method can be adjusted to provide results for more than one attribute of the soil with a single analysis [2]. This makes it possible to study the characteristics of PTEs in soils using Vis and NIR spectroscopy.

Chemometric methods are often needed to analyze the spectra characteristics of soil. 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) [3], Principle Component Regression (PCR) [4], or Partial Least Squares Regression (PLSR) [5] have been used in the past to build models for estimating the toxic elements content of soil or sediments. All the above-mentioned calibration methods require the creation of robust and generalized models due to their potential tendency to over-fit the data [6]. Therefore; using a method such as Support Vector Machine Regression (SVMR) that can overcome the problems of other calibration methods seems essential. Therefore, this study was conducted to assess selected PTEs, namely manganese (Mn), Cu, Cd, Zn, Fe, Pb and As concentrations in anthropogenic soils on brown coal mining dumpsites using Vis-NIR, to evaluate the feasibility of the technique in the rapid prediction of the above-mentioned contaminants and to compare the performance of PLSR and SVMR methods for multivariate calibrations using soil reflectance spectra.

II. MATERIALS AND METHODS

A. Study Area and Soil Sampling

Three dumpsites Pokrok (50° 60' N; 13° 71' E), Radovesice (50° 54' N; 13° 83' E) and Prunéřov (50° 42' N; 13° 28' E) were located in. All dumpsites are formed by clays. Disturbed and undisturbed soil samples were collected on all dumpsites randomly. 103 soil samples were collected on the Pokrok dumpsite, 40 samples on the Radovesice dumpsite and 48 samples on the Prunéřov dumpsite. Sampling was made in the depth of 0 to 20 cm

B. Soil Analysis

The samples were air-dried and sieved through a 2 mm mesh. All samples were then saved for analyses of PTEs (including Cu, Mn, Cd, Zn, Fe, Pb and As) and reflectance measurements. Target elements were extracted using 2M HNO_3. Arsenic was determined in extracts by flow-through electrochemical coulometry analyser EcaFlow 150GLP (Istran, SK).

C. Reflectance Spectroscopy Measurement

Reflectance was measured in the 350-2500 nm wavelength range by a FieldSpec 3 spectroradiometer with a contact probe. The probe was mounted on a tripod and positioned about 10 cm vertically above the sample. 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. Each sample spectrum was corrected for background absorption by division of the reference spectrum of a

---

Manuscript received July 28, 2014, revised October 18, 2014. This work was supported of the EC Operational Program (Project No.: ESF/MEYS CZ.1.07/2.3.00/30.0040) and Czech University of Life Sciences Prague.

A. Gholizadeh and L. Borůvkaa are with the Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Parague, Kamycka 129, 16521- Suchdol, Praha 6- Prague, Czech Republic (e-mail: gholizadeh@af.czu.cz, boruvka@af.czu.cz*)*.

M. M. Saberioon is with the Smart Farming Technology Research Centre, University Putra Malaysia, Selangor, Serdang, Malaysia (e-mail: saberioon@gmail.com).

DOI: 10.7763/IJESD.2015.V6.659

standardized white $BaSO_4$ panel.

D. Model Construction and Validation

For all samples, an exploratory analysis was carried out to detect outliers before establishing the regression model. The correlation between PTEs concentration and reflectance spectra was determined using Pearson's correlation. In order to calibrate a model that provides accurate predictive performance about the quantity of PTEs contained in each soil sample, the captured soil spectra together with laboratory data of PTEs were imported into R software to be processed. 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 [6]. Then two types of preprocessed spectra were used as made with first and second derivative manipulation. Moreover, PLSR and SVMR models were employed to calibrate spectral data with chemical reference data and to describe the relationship between reflectance spectra and measured PTEs. To determine the optimal number of components the Root Mean Squared Error of Prediction (RMSEP) and Residual Prediction Deviation (RPD), were employed by leave-one-out cross-validation [7].

E. 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. The $RMSEP_{cv}$ and RPD were 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.

$$RPD = \frac{SD}{RMSEP_{cv}}$$

where SD is Standard Deviation value.

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 value between 0.66 and 0.81 indicates approximate quantitative predictions, whereas, an R^2 value between 0.82 and 0.90 reveals good prediction. Calibration models having an R^2 value above 0.91 are considered excellent [8]. Viscarra Rossel *et al*. [9] classified RPD values as follows: RPD $<$ 1.0 indicates a very poor model and their use is not recommended; RPD between 1.0 and 1.4 indicates a poor model where only high and low values are distinguishable; RPD between 1.4 and 1.8 indicates a fair model which may be used for assessment and correlation; RPD values between 1.8 and 2.0 indicate a good model where quantitative predictions are possible; RPD between 2.0 and 2.5 indicates a very good, quantitative model and RPD > 2.5 indicates excellent predictions. This classification system was adopted in this study.

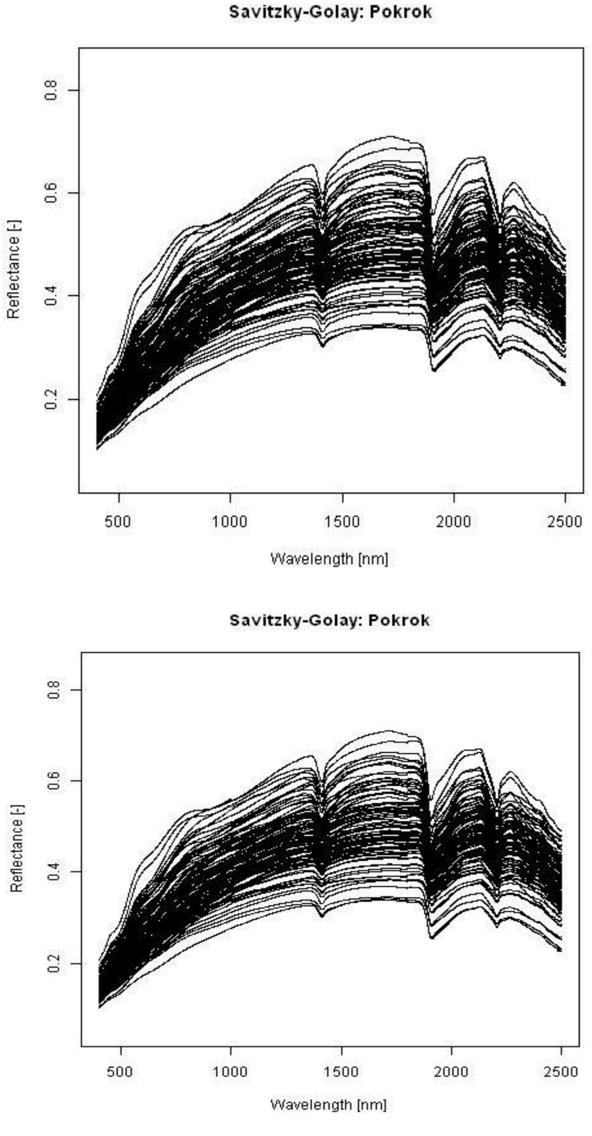
III. RESULTS AND DISCUSSION

A. Soil Samples Descriptive Statistics

General statistical results of PTEs in the three dumpsites are summarized in Table I. The comparison of Coefficients of Variation (C.V.) of different contaminants showed that among all parameters as had the highest C.V., hence, as varied the most as compared to other measured parameters.

TABLE I: DESCRIPTIVE STATISTICS OF PTEs IN THE STUDIED SAMPLE SET ACCORDING TO LOCATION

Item	Cu	Mn	Fe	Cd	Pb	Zn	As
	mg/kg						
Pokrok (n=103)							
Min	5.50	198.3	2503.4	0.01	7.60	8.30	0.49
Max	35.70	869.1	9752.6	0.73	42.40	127.10	19
Mean	13.76	599.4	5418	0.27	18.43	25.26	4.48
Std.	3.58	118.6	1330.1	0.11	5.32	15.77	3.39
C.V. (%)	26	20	25	40	29	62	76
Radovesice (n=40)							
Min	6.42	254.1	1754.4	0.03	4.70	9.38	0.18
Max	22.10	844.1	6876.9	0.30	49.60	66.85	1.30
Mean	14.20	541.3	4489.3	0.17	13.70	21.98	0.67
Std.	3.45	125.1	974.4	0.05	6.40	11.15	0.25
C.V. (%)	24	23	22	30	47	51	38
Prunéřov (n=48)							
Min	8.40	41.6	2105	0.00	0.90	6.60	0.00
Max	92.24	984	9225.4	0.24	24.80	213.11	3.30
Mean	15.81	552.6	5532.5	0.11	14.38	26.83	0.98
Std.	14.36	224.4	1595.5	0.06	4.82	39.32	0.86
C.V. (%)	91	41	29	55	34	147	87



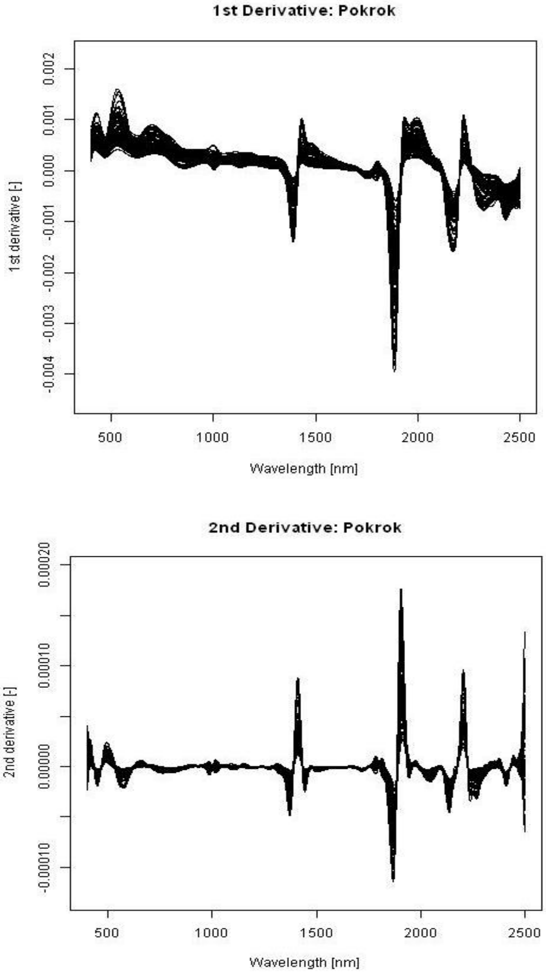


Fig. 1. Raw reflectance spectra, smoothed spectra by Savizky-Golay and preprocessed spectra of soil samples for Pokrok.

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

Smoothed spectra by Savitzky-Golay and first and second derivative spectra of all selected soil were created (Fig. 1). Sets of spectra were characterised qualitatively by observing the positive and negative peaks. Positive peaks are due to the component of interest, while negative peaks correspond to interfering components [10]. 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.

C. Matrix Correlation of PTEs and Reflectance Spectra

An easy approach to visualize spectral implications of PTEs is to plot correlation spectra. Linear correlation coefficients between reflectance and PTEs were moderately high (Fig. 2). This indicates that PTEs do contribute to the reflectance of almost all wavelengths. Fig. 2 also shows that the concentrations of toxic elements in three dumpsite soil samples displayed complex changes in their correlations with the Vis-NIR reflectance of soil spectra. Moreover, it can be seen that each element exhibits its maximum correlation coefficient at a different wavelength. Correlation analysis also indicates that the correlation coefficients of two elements, Cd and Pb, are usually separated from the other elements.

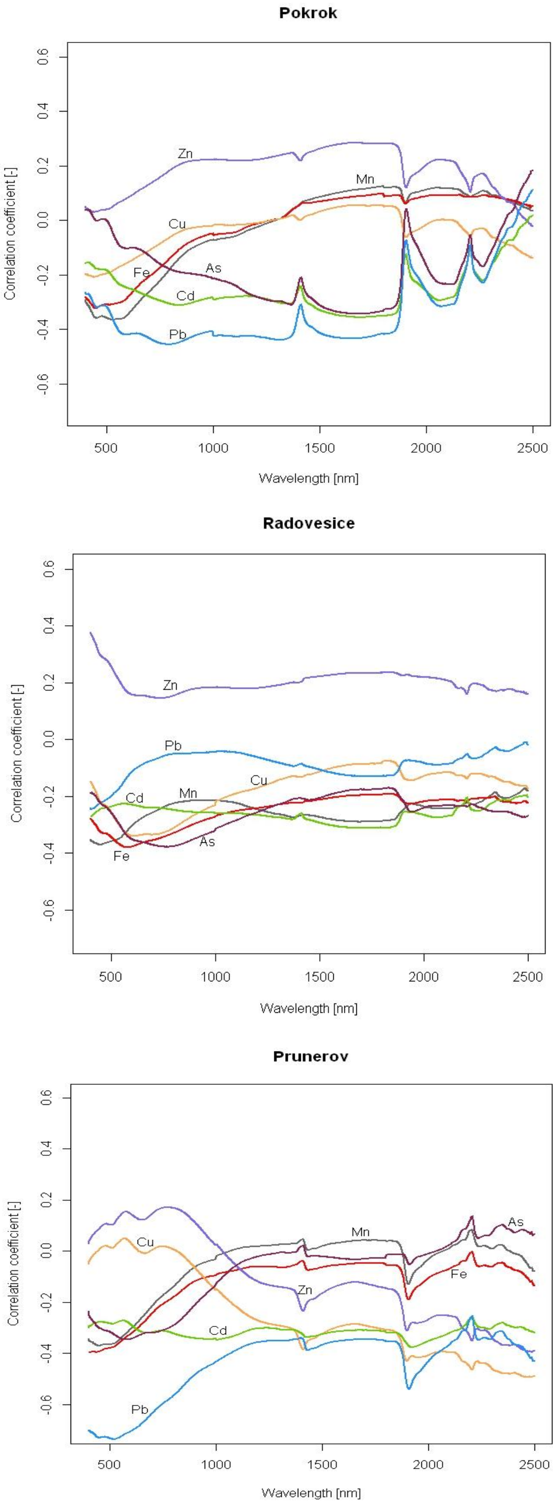


Fig. 2. Correlation between reflectance of Vis-NIR and PTEs in different locations.

In each location the PTEs were categorized into two or three groups according to their behavior and relationships with soil Vis-NIR spectra. This grouping is useful to more easily recognize the prediction ability of Vis-NIR spectroscopy for PTEs with similar behavior. Moreover, it simplifies the estimation of prediction accuracy of each PTE in a group. In Pokrok, the toxic elements were categorized into two groups. The first group of elements (Cd, Pb and As) had stronger negative correlation coefficients with spectral bands than the second group (Cu, Zn, Fe and Mn); the first group displayed high negative spectral correlation at 786 nm for Pb and the second group had the strongest correlation at 1667 nm for Zn. However, correlation coefficient changes of PTEs with Vis-NIR spectra of the 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 correlations for Pb (at 513 nm) and for Zn (at 769 nm).

In Radovesice, the highest positive and also the highest negative spectral correlations can both be seen in the first group of elements (Zn, Cu, As and Fe), in which the strongest positive correlation coefficient related to Zn at 401 nm and Fe represented the lowest spectral correlation at 578 nm arising from Fe^{3+} absorption. These results were similar to results of Ben-Dor [11]. They mentioned that the contribution of the region 390-550 nm is attributed to the spectral absorption features of free iron oxides.

D. 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 PTEs 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 [12].

In general, by comparing the results of the PLSR and SVMR models for the Vis-NIR spectra which can be seen in Table II, one can understand that PLSR have been successfully used to calibrate many soil variables including some PTEs concentrations. In this study, in addition to PLSR which showed fairly good predictions, SVMR provided very good correlations between soil spectra and various PTEs; better prediction was achieved using SVMR and it outperformed the PLSR in calibration and validation both. 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 interaction effects as well as linear combinations of variables, it is able to approximate nonlinear functions between multidimensional spaces [13].

TABLE II: STATISTICS RESULTS FOR CALIBRATION AND CROSS-VALIDATION OF THE VIS-NIR DIFFUSE REFLECTANCE SPECTROSCOPY FOR EACH PTEs

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

IV. CONCLUSION

This study demonstrated the application of laboratory Vis-NIR reflectance spectroscopy for prediction of PTEs including Cu, Mn, Cd, Zn, Fe, Pb and As, using soil samples taken from three brown coal mining dumpsites of the Czech Republic. For each parameter, Vis-NIR calibration models were created by PLSR and SVMR algorithms. 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 PTE concentrations in anthropogenic soils. 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 PTEs, which generally involves conducting large numbers of analyses in a short time.

ACKNOWLEDGMENT

The authors acknowledge the financial support of the EC Operational Program (Project No.: ESF/MEYS CZ.1.07/2.3.00/30.0040) and Czech University of Life Sciences Prague.

REFERENCES

[1] H. Y. Ren, D. F. Zhuang, A. N. Singh, J. J. Pan, D. S. Qid, and R. H. Shi, "Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: A case study," *Pedosphere*, vol. 19, pp. 719-726, 2009.

[2] D. M. Salazara, H. L. Martınez Reyesa, M. E. Martınez-Rosasa, M. M. Miranda Velascoa, and E. A. Ortegaa, "Visible-near infrared spectroscopy to assess soil contaminated with cobalt," *Procedia Eng*, vol. 35, pp. 245-253, 2012.

[3] R. C. Dalal and R. J. Henry, "Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry," *Soil Sci. Soc. Am. J.*, vol. 50, pp. 120-123, 1986.

[4] A. Pirie, B. Singh, and K. Islam, "Ultra-violet, visible, near-infrared, and mid infrared diffuse reflectance spectroscopic techniques to predict several soil properties," *Aust. J. Soil Res.*, vol. 43, pp. 713-721, 2005.

[5] J. Moros, S. F. O. de Vallejuelo, A. Gredilla, A. de Diego, J. M. Madariaga, S. Garrigues, and M. de la Guardia, "Use of reflectance infrared spectroscopy for monitoring the metal content of the estuarine sediments of the Nerbioi-Ibaizabal River (Metropolitan Bilbao, Bay of Biscay, Basque Country)," *Environ. Sci. Technol.*, vol. 43, pp. 9314-9320, 2009.

[6] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

[7] X. Xie, X. Z. Pan, and B. Sun, "Visible and near-infrared diffuse reflectance spectroscopy for prediction of soil properties near a Copper smelter," *Pedosphere*, vol. 22, pp. 351-366, 2012.

[8] P. Williams, "Near-infrared technology-Getting the best out of light," *PDK Projects*, Nanaimo, Canada, 2003.

[9] R. V. V. Rossel, R. N. McGlyn, and A. B. McBratney, "Determining the composition of mineral-organic mixes using UV-Vis-NIR diffuse reflectance spectroscopy," *Geoderma*, vol. 137, pp. 70-82, 2006b.

[10] D. M. Haaland and E. V. Thomas, "Partial least-squares methods for spectral analyses: 1. Relation to other quantitative calibration methods and the extraction of qualitative information," *Anal. Chem.*, vol. 60, pp. 1193-1202, 1988.

[11] E. Ben-Dor, "Quantitative remote sensing of soil properties," *Adv. Agron.*, vol. 75, pp. 173-243, 2002.

[12] A. M. Mouazen, M. R. Maleki, J. De Baerdemaeker, and H. Ramon, "On-Line measurement of some selected soil properties using a VIS–NIR sensor," *Soil Till. Res.*, vol. 93, pp. 13-27, 2007.

[13] A. Stevens, T. Udelhoven, A. Denis, B. Tychon, R. Lioy, L. Hoffmann, and B. van Wesemael, "Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy," *Geoderma*, vol. 158, pp. 32-45, 2010.

A. Gholizadeh was born in Iran in 1979. She got B.Sc. and M.Sc. degrees in agricultural engineering-agronomy from Ferdowsi University of Mashhad, Iran in 2001 and 2004 respectively. She received the Ph.D. degree in precision agriculture from University Putra Malaysia in 2011.

She began her research career in 2007 at University Putra Malaysia as a research assistant and continued to 2011. She also did her job as a postdoctoral research fellow at University Putra Malaysia since 2012 till 2013. Her current job is an educational staff at Czech University of Life Sciences Prague, which has been started since 2013. A selection of her publications during the last year is 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, Models for paddy soil physical properties estimation using visible and near infrared reflectance spectroscopy and Assessment of rice leaf chlorophyll content using visible bands at different growth stages at both the leaf and canopy scale.

Dr. Gholizadeh is also a member of Malaysian Remote Sensing Society as well as International Society of Precision Agriculture.