Research Article

The prediction of heavy metals lead (Pb) and zinc (Zn) contents in soil using NIRs technology and PLSR regression method

Husnizar[1*], Wahyu Wilopo[2], Ahmad Tawfiequrrahman Yuliansyah[3]

[1] Master of System Engineering, Faculty of Engineering, Gadjah Mada University, Indonesia
[2] Department of Geological Engineering, Faculty of Engineering, Gadjah Mada University, Indonesia
[3] Department of Chemical Engineering, Faculty of Engineering, Gadjah Mada University, Indonesia

*corresponding author: husnizar83@gmail.com

Abstract : The case of soil contamination by heavy metals in environment shows an increasing amount along with the constructions and development period that has been done. The identification of heavy metal content in the field is extremely hard to be done without a laboratory activity. Therefore, it needs a quick and non-destructive approach or method to identify the metal content of the soil in the field. The application of Near Infrared Reflectance Spectroscopy (NIRs) technology is a method that non-destructively able to detect the heavy metal content in the soil by using Partial Least Square Regression (PLSR). Pretreatment spectrum which is done using the Multiplicative Scatter Correction (MSC) can improve the results of the prediction models of PLSR. The results of MSC pretreatment spectrum can repair and improve the accuracy of the predictions of Lead (Pb) and Zinc (Zn) in the soil. Eight samples were used for analysis of each of Pb and Zn content. The measured data were pre-treated by MSC. It was obtained that value of r = 0.98, R^2 = 0.97 and RPD = 6.46 for the Pb content measurement. Meanwhile, the measurement for Zn obtained the value of r = 0.98, R^2 = 0.97 and RPD = 6.28. Therefore, it can be inferred that the NIRs is one of technologies which is worth reckoned as the right and quick means to predict the content of heavy metals in soil in a non-destructively and environmentally friendly way.

Keywords: heavy metals, NIRs, PLSR, solid, standar solution

Introduction

Soil is the main medium for plants, animals and human beings to grow and develop; above that soil, all the living things do their activities (Shi et al., 2014). A wide range of human activities such as mining, transportation, waste disposal and fertilization could also be the cause of the occurrence of threats to the soil health (Wang et al., 2014). In addition, consuming crops such as maize, rice and soybeans grown on heavy metal contaminated soils greatly increases the potential risk of food safety and health to human. Basically, there are several heavy metals that are minor needed by plants such as Cu and Zn. But, if the number had passed the threshold, this type of heavy metal becomes very dangerous when absorbed by the plants. Heavy metal also will not decompose on its own, so it would be very dangerous if it is continuously accumulated (Zhuang et al., 2009). The determination of heavy metals concentration in soils is very necessary, because by monitoring the health condition of the soil can be understood and the proper way in preventing and rehabilitating the contaminated soil can be determined (Schwartz et al., 2011; Wang et al., 2014; Shi et al., 2014). Basically, several conventional methods are already available for detecting and measuring the heavy metals content in the soil. However, the methods, in addition to the expensive cost, complicated sample preparation, take a long time, the results of testing with the methods still have chemical

waste, and can describe specific location only (Liu et al., 2011; Wang et al., 2014). Near Infrared Reflectance Spectroscopy (NIRs) is a method which is very promising in terms of analyzing some things, whether agricultural products and soil. The advantages of this technology along with simple sample preparation, it can analyze quickly and non-destructively, and it is free from the use of the chemical compound, so it is friendly to the environment (Liu et al., 2010). The research related to the use of NIRs technology still largely examines the content of organic material, only some of which already conduct research of heavy metals in the soil. Research related to the use of NIRs technology for detecting heavy metal is performed by Moros et al. (2008) by building a prediction model to detect arsenic (As) and lead (Pb) metal in a sample of red paprika powder. Luce et al. (2017) conducted a research using NIRs technology to detect the content of heavy metals in soils in the paper mills. Another study about the composition of the organic contents was made by Magawa et al. (2016) about the development model of NIRs for rapid quantification of protein content at sweet potato. In addition, Alfia et al. (2016) conducted a research on the determination of fat content non-destructively on cocoa powder by using NIRs technology. The aim of this study was to examine the potential and the reliability of NIRs technology to detect of lead (Pb) and zinc (Zn) contents in agricultural soil.

Materials and Methods

This research had been carried out by sampling the soil in the village of Naga Umbang in the district of Aceh Besar, Indonesia. Treatments of the experimental soil samples were conducted in the Laboratory of Soil and Water Engineering, Department of Agricultural Engineering. Meanwhile, the spectrum acquisition of land was conducted in the Laboratory of Instrumentation and Energy, Department of Agricultural Engineering, Faculty of Agriculture, Syiah Kuala University, Banda Aceh.

Research method

Before the soil was tested to the basic content of the heavy metals, it was left for 2-3 days in the laboratory so that the temperature of the soil is homogenous. Then, the soil samples were cleaned from the residual grass, gravel and plants roots. Soil samples were mixed by using a mortal and filtered with a size of 20 mesh. The testing of natural content of heavy metals in the soil was carried out using the method of Atomic Absorption Spectrometer (AAS) which was

conducted in the laboratory of Institute for Industrial Research and Standardization of Banda Aceh (BARISTAND). The soil samples were experimented with adding 1,000 ppm standard solution into samples for each of the heavy metals Pb and Zn with compositions of 250 ppm, 500 ppm, 750 ppm, and 1,000 ppm by using the formula (1) of disbursements as follows :

$$V_1 M_1 = V_2 M_2 \ ....(1)$$

where:
V_1 : the volume of the solution before dissolving
M_1 : the molarity of the solution before dissolving
V_2 : the volume of the solution after the dissolving
M_2 : the molarity of the solution after dissolving

Spectrum acquisiton

Spectrum soil was taken by using NIRs technology with workflow configuration tool that was built by using integrated software Thermo Integration® and for running was done by Thermo® operation. The wavelength interval used was between 1000 and 2500 nm with intervals of approximately 2 nm. In one scanning, there were 32 times data, averaging that data, and then saved in the file i.e. SPA, JDX and CSV. The spectrum acquisition was committed to the paddy fields and fields samples with samples of $\pm$ 10 to 20 grams. The soil sample was inserted into sample curvet, and then it was placed above the sample seat. The data were obtained by transforming the reflectant value into the form of Log (1/R). The pretreatment method of spectrum used was the Multiplicative Scatter Correction (MSC).

The development of calibration model

The development of calibration model used in this study was the method of Partial Least Square Regression (PLSR) to detect the content of heavy metals Pb and Zn in soils of paddy fields and the fields. Setting the data prediction with NIRs (into variable X) compared to the data of laboratory testing results by using AAS method (into variable Y) so we got the optimum wavelength of each heavy metal contents of Pb and Zn. The determination of the reliability and validity of a model which has already built can be seen from the value of the coefficient of correlation (r), the coefficient of determination (R^2), the root mean square errors (RMSE) at the time of calibration (RMSEC) and the ratio between the standard

deviation with RMSEC (RPD). So, a good Model is if it has a high value of r, the smaller value of RMSEC than the standard deviation (SD) of the data and the high value of the RPD (Naes et al., 2004; Sinelli et al., 2008). The formula used to calculate RMSEC and RPD is as follows:

$$RMSEC = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\bar{y}_i - y_i)^2} \quad ....(2)$$

$$RPD = \frac{SDp}{RMSEC} \quad ....(3)$$

where :

$\bar{y}_i$: The parameter values from the alleged early models up to-i

Y_i : The parameter laboratory measurement results for samples to -i

N : The number of samples in the calibration, validation or predictions

SDp : Standard deviation for the prediction

Out of the two methods using pretreatment and without using any pretreatment were performed the comparison and the election method which was the best method to predict the content of heavy metals in soils of paddy fields and the fields.

Results and Discussion

Prior to experiment soil samples, chemical testing to the content of heavy metals has been done in advance using AAS to know the natural content of heavy metals in soil samples. Table 1 shows the natural heavy metals in the soil. Then, an experiment was conducted by adding standard solution of 1,000 ppm into the soil sample for each of the heavy metals Pb and Zn with a composition of 250, 500, 750 and 1,000 ppm. Therefore, there were 16 paddy fields and lea samples that have been contaminated with heavy metals lead (Pb) and zinc (Zn).

In Tables 2 and 3 show the concentration experimental results of metals content in soil samples. The result of the soil spectrum resulting from the NIRs application is the raw spectrum, where wavelengths produced ranged from 1,000 to 2,500 nm with an interval of + 2 nm. The lapse of the wavelength ranging from 780 to 999 nm was not used, because it usually describes colour from a sample only. Figure 1 shows a raw soil spectrum resulted from the NIRs application. The results of pretreatment spectrum using Multiplicative Scatter Correction (MSC) Method can be seen in Figure 2.

The resulting wavelength range remained the same i.e. 1.000 up to 2,500 nm. However, the difference with the raw spectrum soil was the subtlety, the distance between each of the wavelengths of the wave spectrum samples produced. Because the MSC method attempts to remove the effects of scattering by linearizing each spectrum to an '*ideal*' spectrum of the sample, with is normally corresponds to the average spectrum.

Table 1. The results of the laboratory testing using AAS method

No	Soil Type	Heavy Metals (ppm)		Testing Method
		Pb	Zn	
1	Paddy/Rice Fields	1.12	46.82	AAS
2	Lea	25.04	32.39	

Table 2. The concentration experiment of heavy metals content in paddy fields samples

No	Heavy Metals	Samples (ppm)			
		I	II	III	IV
1	Pb	251.12	501.12	751.12	1001.12
2	Zn	296.82	546.82	796.82	1046.82

Table 3. The concentration experiment of heavy metals content in lea samples

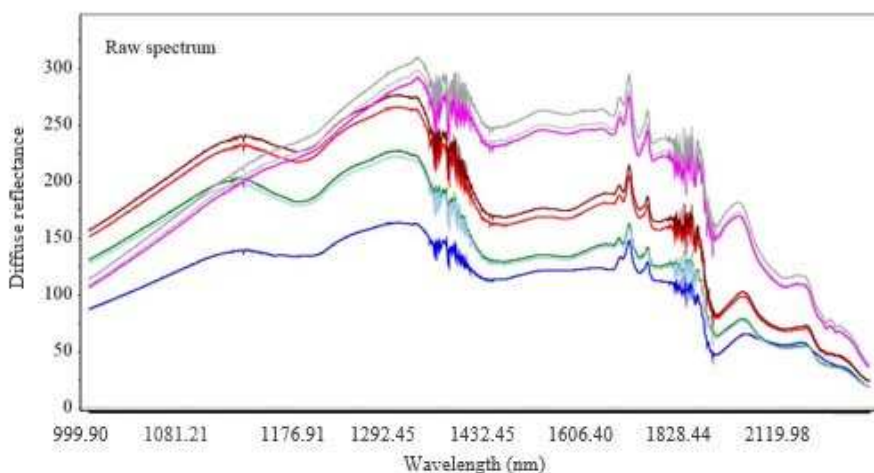| No	Heavy Metals	Samples (ppm)			
		I	II	III	IV
1	Pb	275.04	525.04	775.04	1025.04
2	Zn	282.39	532.39	782.39	1032.39

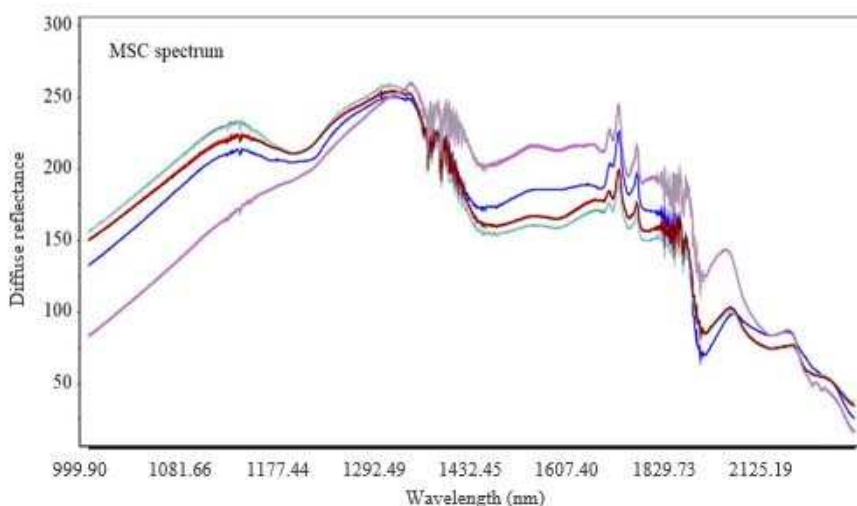Figure.1 Raw Paddy Fields and Lea Spectrum without pretreatment

Figure 2. Paddy fields and lea spectrum with pretreatment MSC method

Callibration model using PLSR

Calibration model using partial least square regression (PLSR) is one of the mathematical models that can be used to predict the content of the chemical composition. One of these is the content of the heavy metals lead (Pb) and zinc (Zn) in soil that is non-destructively done with NIRs application. To improve the predicted performance from PLSR, a pretreatment spectrum using the Multiplicative Scatter Correction (MSC) need to be performed. In this study, we compared the best processing method between the raw spectrums using MSC.

Heavy metal lead (Pb)

The processing of calibration data by using the method of PLSR obtained the values of coefficient of correlation (r) = 0.92, residual predictive deviation (RPD) = 2.68, coefficient of determination (R^2) = 0.84, and RMSEC = 111.54 (Table 4). This model was the raw spectrum or without using pretreatment spectrum, it can be seen in Figure 3, but the results obtained were already eligible for a minimum of the feasibility and reliability of model predictions that were built because some parameters produced high values of r > 0.9, RPD > 2.5, and R^2 > 0.8.

Table 4. The calibration results of prediction models of heavy metals Pb

Spectrum	Statistical Parameter			
	R^2	r	RMSEC	RPD
Row	0.84	0.92	111.54	2.68
MSC	0.97	0.98	46.28	6.46

Furthermore, the processing of the data calibration using PLSR and pretreatment of spectrum MSC method obtained the values of coefficient of correlation (r) = 0.98, residual predictive deviation (RPD) = 6.46, coefficient of determination (R^2) = 0.97, and RMSEC = 46.28 (Table 4).
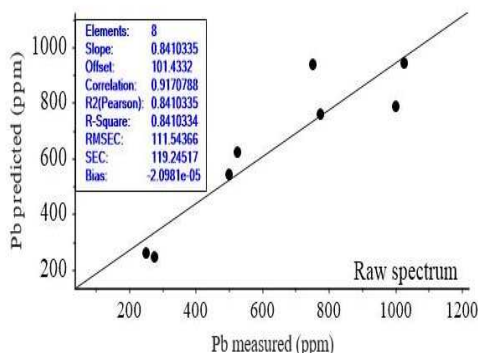

Figure 3. The plot of the laboratory measured data and the predicted NIRs of heavy metal Pb Content without pretreatment

The results of several parameters illustrated that using the pretreatment of spectrum method MSC obtained the values of r, the RPD and R^2 higher, and the RMSEC value lower compared to without pretreatment of spectrum. So, it can be inferred that the model which was built with the pretreatment of spectrum MSC method is better in reliability in predicting the content of heavy metals lead (Pb) in a soil sample (Figure 4).
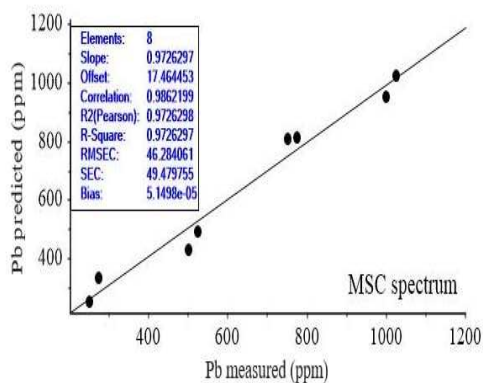

Figure 4. The plot of the laboratory measured data and the predicted NIRs of heavy metal Pb content with pretreatment

From the development of the predictions of heavy metals lead (Pb) model which were built using pretreatment spectrum MSC as well as raw spectrum, the following statistical parameters were obtained

Heavy metal zinc (Zn)

The data processing calibration by using the method of *partial least square regression* (PLSR) obtained the values of coefficient of correlation (r) = 0.97, residual predictive deviation (RPD) = 5.07, coefficient of determination (R^2) = 0.95, and RMSEC = 58.92 (Table 5). This model is the raw spectrum or without using pretreatment spectrum, it can be seen in Figure 5.

Table 5. The calibration results of prediction models of heavy metals Zn

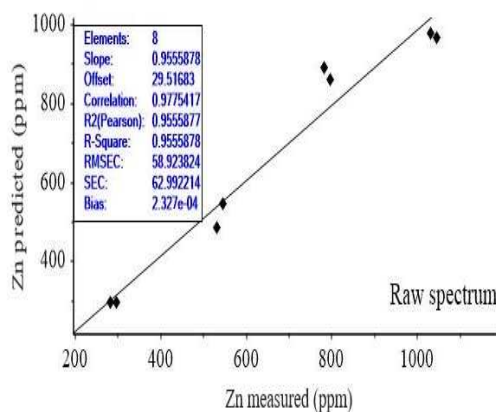| Spectrum	Statistical Parameter			
	R^2	r	RMSEC	RPD
Raw	0.95	0.97	58.92	5.07
MSC	0.97	0.98	47.57	6.28

Figure 5. The plot of the laboratory measured data and the predicted NIRs of heavy metal Zn content without pretreatment

The results obtained were already eligible for a minimum of the feasibility and reliability of model predictions that were built, because some of the parameters resulted in high values of r > 0.9, RPD > 3, and R^2 > 0.9. Further processing was data calibration by using the PLSR method in the presence of additional pretreatment spectrum method of Multiplicative Scatter Correction (MSC). From the prediction model, we obtained the values of coefficient of correlation (r) = 0.98, residual predictive deviation (RPD) = 6.28, coefficient of determination (R^2) = 0.97, and RMSEC = 47.57 (Table 5). Some parameters described that by using pretreatment spectrum MSC methods the values of r, R^2 and RPD were higher than without pretreatment spectrum although the differences in the resulting numbers were not significant. Therefore, it can be inferred that the model built with pretreatment spectrum MSC gave better reliability in predicting the content of heavy metals zinc (Zn) in soil samples,

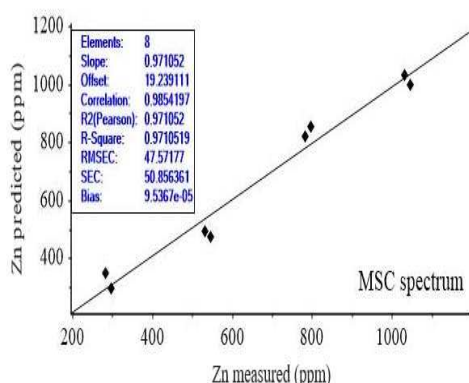and the value of RPD obtained in the group of excellent model performance (Figure 6).

Figure 6. The plot of the laboratory measured data and the predicted NIRs of heavy metal Zn content with MSC pretreatment

From the result of the prediction model development of heavy metal zinc (Zn) which was built by using pretreatment spectrum MSC and raw spectrum, the following statistical parameters were obtained. The bulding of prediction model by using the regression method (PLSR) in predicting the heavy metals content in the soil such as lead (Pb) and zinc (Zn) is very good. It can be seen from the values of coefficient of correlation (r) > 0.90, coefficient of determination (R^2) > 0.80, and RPD > 2.5. It also compared in accordance with the existing references that the value of the RPD was in the between 1.5 and 2 indicating that the model performance is sufficient which can be said that it might be considered as rough quantitative predictions, whereas for RPD = 2.5 and RPD > 3 indicated that the model is in the categories of good and excellent model performance (Nicolai et al., 2007). Pretreatment method using Multiplicative Scatter Correction (MSC) in predicting the content of heavy metals Pb and Zn greatly influenced reliability prediction model which was built primarily to heavy metal Pb content. This can be seen from some statistical parameters obtained for heavy metals Pb, where the values using method without pretreatment were r = 0.92, R^2 = 0.84 and RPD = 2.68. While the values obtained using the method were r = 0.98, R^2 = 0.97 and RPD = 6.46, which increased significantly. So also in predicting metal content of Zn, where the values with the method without pretreatment were r = 0.97, R^2 = 0.95 and RPD = 5.07, while values obtained by the method of MSC pretreatment were r = 0.98, R^2 = 0.97 and RPD = 6.28. Eventhough it did not increase significantly it still improved on each of the parameters of the resulting statistics.

The determination of the relevant wavelength for heavy metal Pb and Zn contents

The content of heavy metals in soil samples of paddy fields and fields, namely metal lead (Pb), produced vibrations in the range of 1,310, 1,412, 1,710, 1,753, and critical area contained on 1,907 until 2003 nm. So, it can be concluded that the relevant wavelength range for estimation of heavy metal Lead (Pb) content can be seen at 6 peak wavelengths with 1 critical area between 1,907 until 2003 nm. For more details, we can see the cluster results of heavy metal Pb content from the soil spectrum in Figure 7.
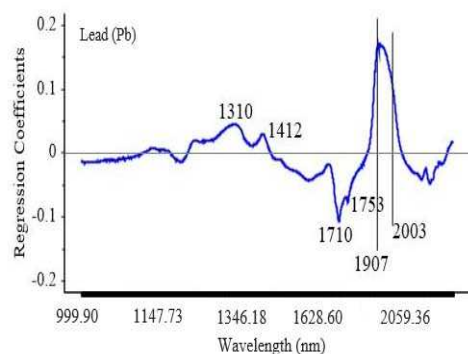

Figure 7. The relevant wavelength plot for heavy metals lead (Pb) content in soil spectrum

The content of heavy metal zinc (Zn) found in a sample of paddy fields and fields produced vibrations between 1,416, 1,721, 1903 until 1,995 and 2,304 nm. For area 1,903 until 1,995 was a critical area for the prediction of heavy metal Zn content. Therefore, it can be concluded that the relevant wavelength range for prediction of heavy metal Zinc (Zn) content occurred on the 5 peaks with one critical area. For more details, we can see the cluster results of heavy metal Zn content from the soil spectrum in Figure 8.
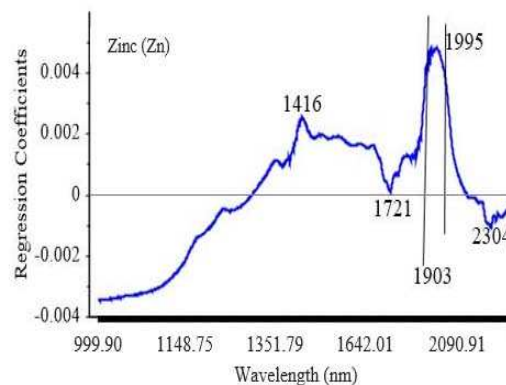

Figure 8. The relevant wavelength plot for heavy metals zinc (Zn) content in soil spectrum

Conclusion

From the results of research that has been done, it can be concluded that the model prediction for the heavy metals content in soils of paddy fields and fields with NIRs using partial regression last square (PLSR) was very accurate. Some parameters such as values of r > 0.9, R^2 > 0.8, and RPD > 2.5 already met the reliability and feasibility of the prediction model built. Pretreatment Spectrum method by using Multiplicative Scatter Correction (MSC) was also very influential towards the reliability of model predictions built, which increased the value of the RPD for heavy metal contents of Pb and Zn. There were 6 peaks with 1 critical area between 1,907 and 2,003 for the relevant wavelength of soil spectrum for heavy metal lead (Pb). Whereas the relevant wavelengths of soil spectrum for heavy metal zinc (Zn) were 5 peaks with 1 critical area between 1,903 and 1,995 nm. The expected suggestion for further research is to conduct a further research with additional parameters such as spectrum correction method spectrum or pretreatment such as Standard Normal Variate (SNV), Mean Normalization (MN) and De-Trending (DT). Then, conducting an addition of the number of samples either the experiment of soil samples, different soil textures and other heavy metals such as copper (Cu), arsenic (As) so that independent validation of a model can be done.

Acknowledgements

References

Alfia, M., Zulfahrizal., Munawar, A.A. 2016. Non-destructive determination of cocoa bean fat content by NIRS (comparison between mean normalization and de-trending). *Jurnal Ilmiah Mahasiswa Pertanian Unsyiah* 1 (1): 1027-1036 *(in Indonesian)*.

Liu, Y., Sun, X. and Ouyang, A. 2010. Nondestructive measurement of soluble solid content of navel orange fruit by visible–NIR spectrometric technique with PLSR and PCA-BPNN. *LWT - Food Science and Technology* 43 (4): 602–607.

Liu, Y.L., Li, W., Wu, G.F. and Xu, X.G. 2011. Feasibility of estimating heavy metal contaminations in floodplain soils using laboratory-based hyperspctral data—a case study along Le'an River, China. *Geo-spatial Information Science* 14 (1): 10–16.

Luce, M.St., Ziadi, N., Gagnon, B. and Karam, A. 2017. Visible near infrared reflectance spectroscopy prediction of soil heavy metal concentrations in paper mill biosolid-and liming by-product-amended agricultural soils. *Geoderma* 288: 23–36

Magwaza, L.S., Messo Naidoo, S.I., Laurie, S.M., Laing, M.D. and Shimelis, H. 2016. Development of NIRS model for rapid quantification of protein content in sweet potato [*Ipomoea batatas* (L.) LAM.]. *LWT – Food Science and Technology* 72: 63-70

Moros. J, Llorca. I, Cervera. M.L, Pastor. A, Garrigues. S. and de la Guardia. M. 2008. Chemometric determination of arsenic and lead in untreated powdered red paprika by diffuse reflectance near-infrared spectroscopy. *Analytica Chimica Acta* 613: 196-206.

Naes, T., Isaksson, T., Fearn, T. and Davies, T. 2004. A User-Friendly Guide to Multivariate Calibration and Classification. Chichester (UK): NIR publications.

Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I. and Lamertyn, J. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology* 46: 99-118.

Schwartz, G., Eshel, G. and Ben-Dor, E. 2011. Reflectance spectroscopy as a tool for monitoring contaminated soils, in: S. Pascucci (Ed.), Soil Contamination, InTech, Rijeka, 2011, pp. 67–90.

Shi, T., Yiyun, C., Yaolin, L. and Guofeng, L. 2014. Visible and near-infrared reflectance spectroscopy: An alternative for monitoring soil contamination by heavy metals. *Journal of Hazardous Materials* 265: 166–176.

Sinelli, N., Spinardi, A., Di Egidio, V., Mignani, I. and Casiraghi, E. 2008. Evaluation of quality and nutraceutical content of blueberries (*Vaccinium corymbosum* L.) by near and mid-infrared spectroscopy. *Postharvest Biology and Technology* 50: 31-36.

Wang, J., Lijuan, C., Wenxiu, C., Tiezhu, S., Yiyun, C. and Yin, G. 2014. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma* 216: 1-9.

Zhuang, P., McBride, M.B., Xia, H., Li, N. and Li, Z. 2009. Health risk from heavy metals via consumption of food crops in the vicinity of Dabaoshan mine, South China. *Science of the Total Environment* 407 (5): 1551–1561.