Inference model to predict heavy metals of Bahr El Baqar soils, Egypt using spectroscopy and chemometrics technique

El-Sayed E. Omran¹

Received: 30 October 2016 / Accepted: 2 November 2016 / Published online: 11 November 2016
© Springer International Publishing Switzerland 2016

Abstract With 35 soil samples gathered from the Bahr El Baqar region, Egypt, the objective of this study was to propose an inference model for prediction of heavy metals (Cd, Cu, Pb, Zn, Ni, Mn, Cr, Co, Fe) with visible near-infrared and short wave infrared (VNIR-SWIR) region (350–2500 nm). For a better understanding of the mechanism that allows the estimation of heavy metals with reflectance spectroscopy, statistical analysis was first made. The fingerprint region of 538, 578, 630, 870, 1900, 2240, and 2376 nm is very useful in recognizing small differences in the heavy metals of Bahr El Baqar region. Partial least squares regression (PLSR) is able to some extent (moderate accuracy) to model heavy metals with laboratory spectra parameters. The best coefficients of determination (R^2) between predicted and chemically analyzed concentrations were for Mn, 0.62; Pb, 0.66; Zn, 0.66, Ni, 0.69, and Cu, 0.60. Iron influences the soil reflectance in the VNIR-SWIR region. This is due to the electronic transition of iron cations (2+, 3+). Mn, Pb, Ni, Cu, and Zn were successfully predicted using PLSR. Only Cr (0.59) and Cd (0.52) content was predicted fairly. The analysis of correlation between heavy metal and soil constituents of Fe_2O_3 , Al_2O_3 and OM, which represented the clay minerals, iron oxides and organic matter, respectively, could support the above predicated binding forms of heavy metals. The order of the correlation coefficients from high to low between metal and wavelength is $\text{Cu} > \text{Zn} > \text{Mn} > \text{Co} > \text{Pb} > \text{Cr} > \text{Ni} > \text{Cd}$. This is almost the same as the order of their correlation coefficients with Fe_2O_3 , Al_2O_3 and OM. These findings also validate a mechanism to predict

heavy metals that have no absorption features in reflectance spectra. The results concerning the relationships between spectral parameters and heavy metals concentration indicate that iron oxides, clay, and OM play an important role in the prediction of soil heavy metals using reflectance spectra. To identify the specific wavelengths for prediction of the studied heavy metals, the regression coefficient with OM, Al_2O_3 , and Fe_2O_3 were used. The most significant peaks for Zn prediction are found in the wavelengths of 430, 570, and 1700 nm; for Cu prediction at 538, 1259, 1500, and 2184 nm; for Pb significant peaks are at 440, 578, 915, and 1894 nm; for Cr at 860, 1001, and 2376 nm; for Ni at 870, 909, and 2240 nm; for Cd at 630, and 1270 nm; for Co at 1900, and 2300 nm. As the current results suggested, large concentrations of heavy metals can be predicted using reflectance spectroscopy. However, it is much more difficult to determine small concentrations of heavy metals in soils of Bahr El Baqar region. Organic matter masks spectral signatures, making it difficult to identify metal quantities at these wavelengths. So, a new concept “inference approach” in light of key soil constituents that absorb VNIR-SWIR radiation would become a powerful tool to enhance the accuracy of heavy metal prediction. Due to the interference effects of organic matter and low heavy metal concentration, the selection of the most sensitive original spectral bands still did not result in highly accurate prediction. Therefore, it is important to predict heavy metals through OM, iron oxides, or clays first. These results indicate that it is possible to predict heavy metals in soils using the rapid and economic reflectance spectroscopy.

✉ El-Sayed E. Omran
ee.omran@gmail.com

¹ Soil and Water Department, Faculty of Agriculture, Suez Canal University, 41522 Ismailia, Egypt

Keywords Bahr El Baqar soil · Reflectance · Spectroscopy · Partial least squares regression · Inference model · Heavy metals

Introduction

The unreasonable utilization of natural (including land) resources by humans caused numerous ecological and environmental issues. Human and anthropogenic activities, such as excessive use of fertilizers, pesticides, sewage irrigation, and discharge of waste affected soil environment significantly. A standout amongst the most polluted drain by anthropogenic activities in Egypt is Bahr El-Baqar (Omran and Abd El Razek 2012). Saad (1997) concluded that 58% of the total drainage water of Bahr El-Baqar drain originate from agricultural drainage, 2% from industrial drainage and 40% from domestic and commercial drainage. Bahr El-Baqar drain conveys the greatest part of untreated wastewater into Lake Manzala through a very densely populated area passing through Qalubya, Sharkia, Ismailia and Port Said Governorates. The soils of Bahr El-Baqar get numerous kinds of pollutants particularly heavy metals such as lead, cadmium, nickel, and mercury, which are considered the most hazardous (Park and Shin 2006). Soil heavy metals have been an exceptionally helpful indicator of environmental and soils quality, which affects the physico-chemical characteristics of soil ecosystems. Although conventional methods of soil heavy metals determination are sufficiently accurate, they are mostly based on wet digestion of soil samples in hot concentrated acids followed by atomic absorption spectrometry or inductively coupled plasma spectrometry. Therefore, these methods are time-consuming, expensive, and require environmentally damaging chemicals and qualified staff (Omran 2016). The cost for the expense for conventional soil characterization figured by the U.S. National Soil Survey Center, is about US\$2500 per pedon and requires 6–12 months to be accomplished (Brown et al. 2006; Rizzo et al. 2014). In this way, development of alternative environmentally friendly method that is accurate, rapid, and inexpensive is of great value.

Visible-near infrared-short wave infrared (VNIR-SWIR) reflectance spectroscopy is relatively less expensive and faster than traditional wet chemical measurements (Summers 2009). The VNIR-SWIR spectral wavelengths (e.g., 350–2500 nm) act as predictor variables, and the soil heavy metals are predicted by a statistical (e.g., partial least squares regression, PLSR) model. Two aspects can largely influence the VNIR-SWIR based determination of soil heavy metals. One is the spectra transformation/pre-processing technique, while the other is soil sample preparation (Stenberg et al. 2010). On one hand, some researchers have looked at the spectral curves of soil samples with various surface roughness and moisture conditions (Wu et al. 2009). The investigation of such changes could be useful as the VNIR-SWIR spectra are known to carry

comprehensive information of both chemical and physical phenomena. On the other hand, the spectral pre-processing technique has been considered an essential part of chemometrics modelling. To date, few studies, however, have assessed the data distribution of each wavelength after different spectral transformation, despite the important role that this plays in the correlation analysis between spectra and soil heavy metals. Moreover, its application permitted the baseline shift, light scattering and non-linearities to be largely removed (Rinnan et al. 2009). Unfortunately, VNIR region are dominated by powerless overtones and combination of vibration bands of atoms with strong molecular bonds containing nitrogen, oxygen and carbon attached to hydrogen and no absorption metals were found in this region. There vibrational modes are modulated by forming complexes with organic molecules containing C–H, N–H and O–H bonds (Malley and Williams 1997).

Heavy metals as soil contaminants commonly occur in low concentrations ($10,000 \text{ mg kg}^{-1}$ dry soil). Inorganic material as a rule does not show characteristic absorption features in the VNIR-SWIR wavelength region (Wu et al. 2005; Schwartz et al. 2012; Pandit et al. 2010; Vohland et al. 2009; Zhang et al. 2010; Yaolin et al. 2011). It is therefore considered difficult to directly detect inorganic soil contaminants in general, and heavy metals in particular (Winkelmann 2005). Peng et al. (2013) inferred that the impact of organic matter (OM) on quantitative estimation of the iron oxide content in soil is critical, especially when OM content exceeds 20 g kg^{-1} . Therefore, the variation in OM content has a vital impact not only on the quantitative estimation of OM itself but also on that of other soil properties. When OM content exceeded 20 g kg^{-1} , it is vital to adopt effective techniques to mitigate this effect (Peng et al. 2013).

So, a new concept “inference approach” in light of key soil constituents that absorb VNIR-SWIR radiation would become a powerful tool to enhance the accuracy of heavy metal prediction. While pure metals do not absorb in the VNIR-SWIR region, they might be detected indirectly through their complexing with OM, association with moieties such as hydroxides, sulfides, carbonates, or oxides, which are detectable, or adsorption to clays that absorb light in this wavelength range (Schwartz et al. 2012). Therefore, the objective of this study is to propose an inference model for heavy metals prediction (Cd, Cu, Pb, Zn, Ni, Mn, Cr, Fe) with VNIR-SWIR range in soil samples collected at Bahr El-Baqar region, Egypt. The present study is gear towards study the feasibility of different pre-processing VNIR-SWIR transformation for soil heavy metals inferences. The present study endeavors to answer two main questions: (1) is it possible to implement VNIR-SWIR spectroscopy as alternative, fast and cost-effective

method for soil heavy metals prediction under low concentrations? and (2) is it possible to establish successful calibration models for heavy metals inference with small number of samples collected at farm scale?

Reflectance spectra of heavy metals

Inorganic material such as heavy metals, which do not show characteristic absorption/reflection features, cannot be directly detected in the VNIR-SWIR region (Schwartz et al. 2012; Pandit et al. 2010; Zhang et al. 2010; Yaolin et al. 2011). Schwartz et al. (2012) proposed that soil contamination may be detected indirectly via association with moieties, which are detectable in the VNIR-SWIR region. Prediction of soil heavy metals from spectra, may be detected indirectly through primary correlations to well recognized absorptions (e.g. those for clay minerals, OM, or iron oxides) and those that rely on secondary correlations with the soil constituents that absorb VNIR-SWIR radiation (such as pH, cation exchange capacity). Various spectroscopic techniques for evaluating heavy metals contamination have been identified (Wu et al. 2007; Choe et al. 2008; Minasny et al. 2009; Pandit et al. 2010). Table 1 summarizes the previous results of soil heavy metals prediction via VNIR-SWIR spectroscopy.

In order to accurately predict soil contamination via a spectral signal, it is first necessary to understand the relationships between spectrally active constituents, which have direct and significant effect on the reflectance of

heavy or trace metals (Wu et al. 2007; Minasny et al. 2009). OM, clay minerals, and iron oxides were principally responsible for predicting a range of heavy metals (Cd, Cu, Fe, Mn, Ni, Pb and Zn) based on multivariate regression. Zhang et al. (2010) has suggested that heavy metals such as Pb can be directly detected and can also be estimated indirectly using reflectance spectroscopy via correlations with iron oxide contents. Kemper and Sommer (2002) found a strong relationship between iron and heavy metals concentrations, including Pb, when utilizing multiple linear regression analysis (MLR) and an artificial neural network. While Wu et al. (2007) found that Fe had a high correlation and high cross validation statistic with heavy metals, Vohland et al. (2009) found a strong correlation between Pb and other heavy metals and Fe and clay fractions. Yaolin et al. (2011) found good correlation between OM, Pb concentration and laboratory based data in the range of 564–624 nm. Moreover, there was a significant relationship between metal levels (Cd, Hg, Pb) within the spectral domain 500–700 nm, which was related to OM (Bray et al. 2009; Ji et al. 2010; Pandit et al. 2010). Some authors reported good predictions of Cd, Co, and Zn (Wu et al. 2007) while others reported poor predictions (Wu et al. 2007; Malley et al. 2004). Contradictory results have also been reported for Cr, Cu, Ni and Pb (Malley et al. 2004; Wu et al. 2007). Figure 1 shows the main characteristic spectral signatures and corresponding soil attributes and provides selected major spectral regions for some soil properties.

Table 1 Summary results of soil heavy metals prediction via VNIR-SWIR spectroscopy

Soil heavy metals predicted	Number of samples	Modeling method	Author
As, Cd, Cu, Fe, Hg, Pb, Zn	214	MLR/ANN	Kemper and Sommer (2002)
As, Pb	346	Polynomial fit	Kemper and Sommer (2003)
Cd, Cu, Fe, Ni, Pb, Zn	70	PLSR	Siebielec et al. (2004)
Hg	120	PCR	Wu et al. (2005)
As, Cd, Co, Cr, Cu, Ni, Pb, Zn	61	PLSR	Wu et al. (2007)
As, Pb, Zn	49	EMLR	Choe et al. (2008)
As, Cu	22	EMLR	Choe et al. (2009)
As, Cu, Fe	33	PLSR	Ren et al. (2009)
Cu, Fe, Mn, Pb, Zn	149	PLSR	Vohland et al. (2009)
As, Cd, Cr, Cu, Hg, Pb, Zn	122	Linear regression	Ji et al. (2010)
Cd, Cu, Mn, Pb, Zn	8	PLSR	Pandit et al. (2010)
Cd, Cu, Fe, Mn, Ni, Pb, Zn, Cr	22	PLSR	Gannouni et al. (2012)
Zn, Cu, Pb, Cr, Ni	124	PLSR	Todorova et al. (2014)
Cu, Mn, Cd, Zn, Fe, Pb, As	103	PLSR	Gholizadeh et al. (2015b)
Ag, Ba, Be, Cd, Co, Cr, Cu, Ni, Pb, Tl, V, Zn	229	PLSR	Sakizadeh et al. (2015)

MLR multiple linear regression analysis, *EMLR* extended multiple linear regression analysis, *ANN* artificial neural networks, *PLSR* partial least squares regression, *PCR* principal component regression, *Pb* lead, *Zn* zinc, *Mn* manganese, *Fe* iron, *Cd* cadmium, *Cu* copper, *Cr* chrome, *Ni* nickel, *As* arsenic, *Hg* mercury, *Co* cobalt

Fig. 1 Reflectance spectra of soil sample collected from Bahr El-Baqar region indicating the spectral features of the most important constituents

Materials and methods

Study area, soil sampling, and analysis

Environmental and land resources in Bahr El-Baqar region are faced critical problems due to the increasing population, demolishing natural resources, heavy metals pollution, as well as others. A total of 35 soil samples was collected in 2014 from the study area which is located in northern Egypt, Bahr El-Baqar region, between 32°05' to 32°16' Longitude and 30°56' to 31°07' Latitude (Fig. 2). As few as 25 samples can result in good predictions at the farm or field scale (Wetterlind et al. 2008). One hundred to 200 samples might be at the lower limit for a model intended to cover a large geographical area with many and diverse soil types (Wetterlind et al. 2013). Outliers were

examined and one samples were removed. 34 samples remained for soil heavy metals modeling and analysis. The samples were air-dried and crushed with a wooden pestle and mortar to pass through a 0.5 mm mesh-screen.

Nine metals are selected based on their: toxicity/environmental context, relation to local geology, irrigation activities, and data completeness and variability. The selected elements are lead (Pb), zinc (Zn), manganese (Mn), iron (Fe), cadmium (Cd), copper (Cu), chrome (Cr), cobalt (Co), and nickel (Ni). The acidity (pH) and soil salinity are also to be considered with the aforementioned parameters since they are important environmental indicators. Clay content and OM are selected as important parameter for indirectly prediction of heavy metals. Soil pH was measured in deionized water (pH_w) and in 0.01 M $CaCl_2$ (pH_{CaCl_2}) (in 1:2.5 suspensions). Total organic carbon (TOC) was determined by wet digestion (Walkley and Black 1934).

Overall methodology

The overall methodology developed includes the following steps: (1) Set-up of a representative sample set from soil sampling strategy and chemical analysis. (2) Analysis of the samples following standard procedures. (3) Elaboration of a procedure for soil sample preparation before spectroscopic measurements. (4) Division of sample set into three groups: training, validation and test. The relationship between the soil heavy metals and the

Fig. 2 Satellite image and sample locations for the study area (Omran 2016)

Fig. 3 The overall methodology used to predict heavy metals of Bahr El Baqar soils using spectroscopy and chemometrics technique

spectroscopy data is found via the training group and simultaneously cross-validated by the validation group. Finally, the model is applied to the test group, independent of the training and validation process. (5) Elaboration of a predictive model for each soil heavy metals (spectral data pre-treatment and calibration). (6) Evaluating the repeatability and reproducibility of the method. Figure 3 shows the overall methodology developed to achieve the study aim and build the most accurate models for predicting soil heavy metals.

The original soil reflectance spectra consisting of 2151 values (at 1 nm interval) between 350 and 2500 nm were averaged at every tenth nanometer wavelength. The new spectrum consisted of reflectance values at 350, 360, 370,... 2500 nm, inclusive. For example, the reflectance at 400 nm was an integrated reflectance values from 396 to 405 nm. This procedure resulted in a new dataset with 214 wavelengths, that was approximately 1/10th the original data volume, while maintaining sufficient spectral information.

Pre-processing and transformation methods

Several pre-processing techniques were utilized to reduce the noise and to enhance the spectra or both. Scatter-corrective methods and spectral derivatives are the most widely used pre-processing techniques in spectroscopy (Rinnan et al. 2009). Scatter-corrective methods incorporate multiplicative scatter correction (MSC), inverse MSC, extended MSC (EMSC), extended inverse MSC, de-trending, standard normal variate (SNV) and normalization. Because of scatter modify for baseline shifts between samples, these techniques are intended to diminish the (physical) variability between samples. MSC is likely the most widely used pre-processing technique followed by SNV and derivation. Each spectrum is then corrected such that all samples seem to have the same scatter level as the reference spectrum (Vohland et al. 2009). Martens et al. (1983) was proposed to utilize just those parts of the spectral axis that do not include relevant information (baseline). It is difficult to decide such regions in practice,

especially in near infrared (NIR) measurements, where the signals from different chemical components are strongly overlapping and correlated, and little or no true baseline is found. This is the motivation behind why, in most cases, the entire spectrum is utilized to locate the scalar correction parameters in MSC. SNV pre-processing is probably the second most applied method for scatter correction of the data (Barnes et al. 1989). MSC and SNV are the same for most practical applications.

The spectral derivation group is represented by Norris-Williams (NW) derivative and Savitzky-Golay (SG) polynomial derivative. Derivatives have the ability to eliminate both additive and multiplicative effects in the spectra. Vasques et al. (2008) found that derivative procedure has the best transformations in pre-processing. Derivatives of the spectra may allow relevant information to be extracted from the near-infrared range (Ertlen et al. 2010). Both derivation techniques use smoothing in order not to reduce the signal-to-noise ratio in the corrected spectra too much. The parameters of the algorithm (interval width, polynomial order) ought to be carefully chosen to avoid amplification of spectral noise (Nicolai et al. 2007). In order to find the derivative at centre point, a polynomial is fitted in a symmetric window on the raw data. When the parameters for this polynomial are calculated, the derivative of any order of this function can undoubtedly be discovered analytically, and this value is subsequently utilized as the derivative estimate for this center point. The number of points used to calculate the polynomial (window size) and the degree of the fitted polynomial are both decisions that should be made. Too small window (moving-window) will lead to the introduction of large artifacts in the corrected spectra and to a reduced signal-to-noise ratio. However, the larger the size of the window, the smaller the distinction between full and moving-window pre-processing.

Pre-processing methods, such as MSC, SNV, Savitzky-Golay, continuum removal (CR) and derivatives are mostly used in the multivariate calibration techniques to decide the best data (Gholizadeh et al. 2015a). The CR will minimize any brightness differences and emphasize the spectra's absorption bands. Spectroscopic measurements were performed in transmission mode, which can be quantified (to absorbance) using Beer's Law (Nicolai et al. 2007). If the dataset does not follow Lambert-Beer's law (a linear relationship between the absorbance and the concentrations), additional factors in PLS regression can compensate for this non-ideal behavior of the spectral predictor (Martens and Næs 1989). The disadvantage of including such additional factors is an increase in model complexity and, in turn, most probably a reduction of the model robustness for future predictions. The utilization of combinations of pre-processing methods is abundant in literature (Rinnan et al. 2009).

Soil heavy metals prediction

The application of VNIR-SWIR spectroscopy for soil heavy metals prediction does not involve any (hazardous) chemicals. It is a non-destructive technique well suited for analyses of some of the essential constituents of the soil (Wetterlind et al. 2013). There are two approaches to predict heavy metals in soils: (1) supervised in which the chromophore absorption band in question is known and under saturation, and (2) unsupervised in which the chromophore absorption band is unknown and is affected by several factors. Whereas the first approach is valid for the infrared (IR) region by using selected wavebands and pre-treatment to prevent saturation, the second represents the case of reflectance spectroscopy across the VNIR-SWIR region. For that purpose, multivariate statistical techniques (chemometrics) are required to extract the information about the quality attributes that is hidden within the spectral information. Essentially, this involves regression techniques coupled with spectral pre-processing.

Pearson correlation coefficient was used to identify correlations between the selected heavy metals (Cd, Cr, Cu, Fe, Mn, Ni, Zn, Co, Pb), iron oxides, clay and OM. Multivariate analyses of the soil spectral data were performed using principal component analysis (PCA) and partial least squares regression (PLSR). PCA identifies patterns in spectral data, allowing the clustering of data and reducing the number of dimensions without the loss of information. PLSR is a technique used in chemometrics (e.g. Conforti et al. 2013) when there are many predictor variables, which are highly collinear. PLSR is based on latent variable decomposition of two sets of variables: the independent, X (spectral reflectance bands) and the dependent variables, Y (heavy metals). The limited number of PLSR factors selected explains most of the variation in both predictors and responses. PLSR is thus a method where the information is commonly explained by the original X variables that is compressed into few variables known as PLS components (latent variables) (Conforti et al. 2013).

PLSR analysis was carried out using The Unscrambler v. 10.4 software (CAMO Software Inc., Woodbridge, NJ, USA). The soil analysis data determined by conventional analysis methods were added to the corresponding spectral files using the WinISI II (Ver. 1.04) software (FOSS NIRSystem/TECATOR, 1999). Matlab (version 8.1.347) and PLS-Toolbox 4.2 were used to fit PLSR calibration with leave-one-out cross-validations. The ENVI 5.3 software (Exelis Visual Information Solutions 2012) was used to perform the continuum removal. TSG software (<http://www.thespectralgeologist.com>), was used for the visualization and also the quantification of this variability.

Table 2 Summary descriptive statistics of representative physico-chemical properties of the Bahr El Baqar soils

Properties	Minimum	Maximum	Range	Mean	Std. deviation	CV	Skewness	Kurtosis
Fe	7.80	87.96	80.16	47.45	25.19	634.38	−0.04	−1.31
Mn	13.73	98.73	85.00	58.98	23.39	547.26	−0.32	−0.81
Cu	7.73	280.30	272.58	65.70	54.68	2990.45	2.27	6.69
Zn	39.67	215.58	175.91	90.56	41.86	1752.64	0.76	0.76
Ni	61.32	88.73	27.41	73.22	7.96	63.34	0.74	−0.41
Pb	24.42	52.40	27.98	36.64	7.04	49.63	0.58	0.05
Cd	10.27	19.07	8.80	14.69	2.31	5.32	0.06	−0.44
Co	70.51	113.82	43.31	89.72	12.58	158.29	0.66	−0.71
Cr	84.92	134.06	49.14	106.96	12.26	150.23	0.16	0.07
OM	0.22	3.55	3.33	1.60	0.79	0.62	0.73	0.69
pH	7.58	8.82	1.24	8.00	0.30	0.09	0.81	0.25
Clay	3.24	80.00	76.76	26.61	24.52	601.14	1.01	−0.51

CV coefficient of variance%, Fe, Mn, Cu, Zn, Pb, Cd, Co, Cr are in mg kg^{−1}; OM and clay are in%

Goodness-of-fit

The following statistics was used to assess the prediction ability of calibration and validation models including, the coefficient of determination in calibration (R^2_c) and validation (R^2_v), the standard error of cross validation (SE_{CV}) in calibration, the standard error of prediction (SE_P) in validation. The satisfying the criteria for establishing the model where $R^2 > 0.70$ (Morón and Cozzolino 2003). The prediction capacity of the models was assed using the ratio performance deviation (RPD) parameter or standard deviation reference of validation (Conzález-Martín et al. 2007). The residual prediction deviation (RPD) in calibration (RPD_c) and validation (RPD_v) models (Pirie et al. 2005). The goodness-of-fit of the prediction inference model is determined by the R^2 and RMSE_P values obtained from plotting observed concentrations vs the predicted concentrations for each variable. The RMSE_P value is the standard deviation of the predicted residuals error and is computed as the square root of (observed-predicted)²/ N (Umetrics 2002). According to Viscarra Rossel (2008), the RPD values were classified in the following way: RPDs between 1.5 and 2.0 indicate poor predictions; RPDs between 2.0 and 2.5 indicate good prediction and RPDs >2.5 very good/excellent model prediction.

Results and discussion

Heavy metal properties of the Bahr El Baqar soils

Errors related to the traditional chemical analysis to which the spectra are correlated will be included in the calibration model. Therefore, a statistical analysis of the soil data before using it for calibration is important. The mean, standard deviation, and range for each heavy metal are

presented in Table 2. The pH average in Bahr El Baqar soils is about 8.00, because of the presence of calcareous parent material. The relatively high content of soil organic matter (average 1.6%) is mainly related to the high organic matter flux to soil due to direct discharge of domestic and industrial wastewaters. The dissolution of humic acid at high pH is responsible for dissolution of Cu and Pb from soil. Organic matter is important for the retention of metals by soil solids, thus decreasing mobility and bioavailability.

The summary statistics results indicate that the mean value of soil Pb, Cu, Cr and Zn concentrations is 36.64, 65.70, 106.96, and 90.56 mg kg^{−1}, respectively and the range between the minimum and maximum values is 27.98, 272.58, 49.14, and 175.91 mg kg^{−1}, respectively, which is large. There are great variations and high skewness for Cu and Zn (2.27 and 0.76). The CV% values reflect the mean variation of each sampling site in the population. The order of the CVs% for each element, from high to low, was Cu > Zn > Fe > Mn > Co > Cr > Ni > Pb > Cd. This result showed that the variation of Cu and Zn in the soil was larger than other metals.

Spectral signatures of Bahr El Baqar soils

Figure 4(left) shows the spectra of all soil samples measured in laboratory, which show the typical soil reflectance shape in each region of wavelength domain. Reflectance is lower in the visible range (350–650 nm) and higher in the near infrared with specific absorbance bands around 1400, 1900 and 2200 nm. A promising relationship between the heavy metal concentrations and reflectance spectra was determined for the samples. Figure 4(right) shows the Bahr El Baqar soils with different heavy metals concentration. The spectral signatures of the soil samples changed significantly as a function of heavy metals concentration. It was hypothesized that the samples with the highest

Fig. 4 Reflectance of soil samples (*left*) and a spectrum with wavelength position of different heavy metals concentration (*right*)

concentrations would have the lowest reflectance, and that the reflectance would increase proportionally as the heavy metal concentrations decreased.

To determine the relationship among soil parameters, Pearson correlation coefficients were calculated (Table 3). Table 3 shows the correlation coefficients of the heavy metals, clay, iron oxides, organic matter, and reflectance at 400, 800, 1200, and 1600 nm for the samples. The results revealed that the Fe is good correlated with Mn and moderately correlated with Cu and Pb ($r = 0.819, 0.488, 0.479$, respectively). Furthermore, there is a well correlation between Ni and Co ($r = 0.966$). There are significant correlations between Cu and each of Zn, Pb, ($r = 0.678, 0.544$, respectively) as shown in Table 3. This result revealed that these metals have the same source of contamination. The Ni, Fe, Mn, Cu, Zn and Co are positively correlated with OM, pH and clay. Clay highly correlated with Fe, Mn, Cu, and Zn (0.773, 0.658, 0.786, and 0.709, respectively). As can be seen in Table 3, there is a negative correlation between pH and each of OM, clay, Cu, and Zn since increasing the clay and OM leads to increasing the total acidity of these soils. By contrast, there is a positive correlation between pH and Cu and Zn. A significant relationship is found between Cd, Cu content and clay content. Considering all soil samples, Pb, Ni and Cr content is not significantly related to clay content.

Overall, Table 3 shows that there is no direct correlations between heavy metals and reflection spectra at 400, 800, 1200, and 1600 nm. The linear relationship between heavy metal and reflectance is significant for 400 and 800 nm. However, this relation is not significant for 1200 and 1600 nm. Therefore, pre-processing spectra and selecting important wavelength from PLSR is important. Earlier research has shown that the calibration models, in which spectra are not pre-processed, are more sensitive to changes in environmental conditions compared to models

for which pre-processing is applied (Moros et al. 2009). So one could still decide to use a pre-processing method to avoid this problem. Due to the numerous and high-correlated spectral data, band selection or band transformation is the prerequisite for heavy metal inference.

Validation the important wavelengths and pre-processing for PLSR

In this study, specific spectra measured in the laboratory were used for predicting metal concentration levels using different pre-processing methods (Table 4). The optimum calibration models were determined by the lowest standard error of calibration (SE_{CV}) and standard error of prediction (SE_P) and the highest correlation coefficient (R^2). The ratio of standard error of performance to standard deviation (RPD) was used to evaluate the accuracy of prediction.

The use of PLSR for the prediction of the variables Ni, Cr, Pb, Zn, Cu, Cd, Mn, and Co based on the specific spectral value leads to the results mentioned in the Table 4. The pre-processing spectra of the samples were used to describe the chemical variables. All these chemical variables with exception of Cu, Cd, Co and Fe significantly have wavelengths for prediction in the following priority order: Mn > Pb > Zn > Cr > Ni according to the model quality (Table 4).

Mn, Zn, Pb, Cu, and Ni, had significant coefficients of determination 0.62, 0.66, 0.66, 0.60, 0.69, respectively for the validation samples, indicating that the partial least-squares calibration model has low predictive power to estimate metal abundance. Based on Table 2, Cu and Zn have larger variation (CV) than other heavy metals. Co, Cr, and Cd had low coefficients of determination (0.55, 0.52, 0.59, respectively) for the validation samples. Based on Table 4, specific spectral pre-processing increases the accuracy of all heavy metals. The application of most of

Table 3 Correlation coefficients of metal concentrations, organic-matter content, clay, and reflectance (Ref.) at 400, 800, 1200 and 1600 nm

Parameters	OM	pH	Clay	Fe ₂ O ₃	Mn	Cu	Zn	Ni	Pb	Cd	Co	Cr	Ref. 400 nm	Ref. 800 nm	Ref. 1200 nm	Ref. 1600 nm
OM	1															
pH	0.220	1														
Clay	0.630**	0.631**	1													
Fe ₂ O ₃	0.522**	0.540**	0.773**	1												
Mn	0.508**	0.435*	0.658**	0.819**	1											
Cu	0.456**	0.490**	0.786**	0.488**	0.401*	1										
Zn	0.438**	0.448**	0.709**	0.471**	0.434**	0.678**	1									
Ni	-0.486**	-0.458**	-0.467**	-0.385*	-0.414*	-0.429*	-0.460**	1								
Pb	0.226	0.392*	0.586**	0.479**	0.388*	0.544**	0.467**	-0.399*	1							
Cd	-0.100	-0.219	-0.424*	-0.337*	-0.265	-0.188	-0.363*	-0.002	-0.582**	1						
Co	-0.460**	-0.413*	-0.396*	-0.341*	-0.381*	-0.415*	-0.434**	0.966**	-0.420*	-0.056	1					
Cr	0.117	0.015	0.201	0.092	0.118	0.076	0.449**	0.263	0.085	-0.536**	0.300	1				
Ref. 400 nm	0.267	0.035	0.373*	0.390*	0.147	0.451**	0.492**	-0.029	0.221	-0.153			-0.027	0.212	1	
Ref. 800 nm	0.245	-0.010	0.318	0.353*	0.129	0.364*	0.482**	-0.027	0.200	-0.164			-0.025	0.204	0.975**	1
Ref. 1200 nm	0.064	0.005	0.163	0.180	0.020	0.273	0.394*	0.026	0.125	-0.248	0.034	0.268	0.763**	0.765**	1	
Ref. 1600 nm	0.026	-0.005	0.138	0.159	0.003	0.252	0.356*	0.042	0.117	-0.252	0.053	0.260	0.742**	0.744**	0.993**	1

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

Table 4 Validation the most important and significant spectra range (nm) for heavy metals prediction

Heavy metals	Original spectra range, nm	Treatment	R^2	RMSE _{CV}	SE _P	SE _{CV}	RPD
Mn	390–550	1st	0.58	16.50	26.50	23.27	1.42
		2nd	0.62	10.50	26.09	21.45	2.23
		CR	0.35	20.12	27.24	25.21	1.16
Cu	1259, 1422, 1629	1st	0.43	42.40	57.87	54.99	1.29
		2nd	0.46	51.52	53.18	50.64	1.06
		CR	0.60	25.45	49.41	48.24	2.15
Zn	430–570, 1151–1164, 1421, 1618, 1721	1st	0.65	38.32	45.29	41.63	1.09
		2nd	0.66	20.42	44.91	41.49	2.05
		CR	0.42	39.11	46.32	43.25	1.07
Ni	870–990, 909, 1411–1436	1st	0.42	7.69	8.34	8.02	1.04
		2nd	0.46	6.80	8.74	7.01	1.17
		CR	0.69	3.72	8.10	6.90	2.15
Pb	440–584, 915, 1065, 1151–1168, 1430, 1632	1st	0.66	3.50	7.90	6.67	2.01
		2nd	0.62	6.12	8.02	6.85	1.15
		CR	0.32	6.91	8.15	7.52	1.02
Cd	907, 1270, 1430, 1528	1st	0.41	1.85	2.47	2.34	1.25
		2nd	0.52	1.12	2.43	2.31	2.10
		CR	0.33	1.45	2.91	2.95	1.59
Co	238, 400–780	1st	0.39	10.52	13.15	12.77	1.20
		2nd	0.35	12.36	11.58	10.58	1.02
		CR	0.55	6.20	10.82	10.01	2.03
Cr	860–1001	1st	0.52	11.23	13.34	12.21	1.09
		2nd	0.59	6.11	12.37	11.83	2.01
		CR	0.21	11.25	13.65	12.20	1.09

1st first derivative, 2nd second derivative, CR continuum removal

the pre-processing methods for these heavy metals increased the accuracy of prediction, which was similar to the conclusion drawn by Ren et al. (2009).

Figure 5 shows validation results of the heavy metals concentration using different pre-processing methods on the spectra. According to the criteria of minimal RMSE_{CV} and maximal R^2 _{CV}, the second derivatives method was considered as the foremost techniques for predicting Mn, Zn, Cd, and Cr, respectively. However, the CR pre-processing method was chosen as the best algorithm for Cu, Ni, and Co. The first derivatives method was considered as the foremost techniques for predicting Pb only. Although the prediction ability with the different pre-processing of specific spectra was improved, it is still under the satisfying threshold.

Heavy metals inference using OM, Al₂O₃, and Fe₂O₃

Due to the interference effects of organic matter and low heavy metal concentrations for all soil samples (Table 2), the selection of the most sensitive original spectral bands did not result in highly accurate prediction (Table 4;

Fig. 5). Organic matter masks spectral signatures, making it difficult to identify metal quantities at these wavelengths (Fig. 6). Therefore, it is important to predict heavy metals through OM, iron oxides (Fe₂O₃), or clays first. Figure 6 shows the heavy metals inference using OM, Al₂O₃, and Fe₂O₃.

To identify the specific wavelengths for prediction of the studied heavy metals, the regression coefficient (Table 5) with OM, Al₂O₃, and Fe₂O₃ were used. The most significant peaks for Zn prediction are found in the wavelengths of 430, 570, and 1700 nm; for Cu prediction at 538, 1259, 1500, and 2184 nm; for Pb significant peaks are at 440, 578, 915, and 1894 nm; for Cr at 860, 1001, and 2376 nm; for Ni at 870, 909, and 2240 nm; for Cd at 630 and 1270 nm; for Co at 1900 and 2300 nm.

Absorptions at 1500 nm are connected with amine N–H vibrations in organic components (Tekin et al. 2012). Absorptions at 1900 nm are connected with absorption of H–O–H bending and O–H stretching vibrations in montmorillonite (Viscarra Rossel et al. 2006). Absorptions at 2170 nm are connected with Al–(OH) bands in albite and at 2240 nm with Fe–(OH) bands in iron minerals (Gaydon

Fig. 5 Validation and prediction results of the pre-processing models for the heavy metals concentration (in mg/kg)

et al. 2009). The absorption near 2300 nm is connected with Mg–OH bending and O–H stretching vibrations in soil minerals, like chlorite, illite and vermiculite (Viscarra Rossel et al. 2006). These significant wavebands support the assumption that measurements of heavy metals with VNIR-SWIR are performed by co-variation with water, clay, organic carbon and some minerals in the soil (Wu et al. 2007).

The heavy metals have different correlation coefficients with spectra, but the strongest correlated bands differ and are indicative of their binding information. The heavy metals display different correlation features (Tables 5, 6) and can be divided into three groups. Group I has Cr, which

is related to iron oxides, clays, and organic matter. Cr shows high spectral correlations at the 2376 nm spectral bands related to Fe (Mg, Al) –OH from iron oxides and clays, and C–H absorption from organic matter, reflecting its binding feature to all these fine soil constituents. Group II has Ni and Mn, which is related to iron oxides and clays. Ni and Mn show high spectral correlations at the 870 and 2240 nm spectral bands, respectively. Group III has Cu, which is related to iron oxides. Cu has strong correlations with bands resulting from Fe³⁺ absorption at 538 nm. This suggest that binding to iron oxides is the most important form of Cu. Group V has Co, which is related to clays. Co shows high spectral correlations at the 1900 nm spectral

Fig. 6 Continuum removal of spectra for heavy metals inference using organic matter (OM), Al_2O_3 , and Fe_2O_3

bands. Group IV has Cd and Pb, which is related to organic matter. Cd and Pb only exhibit good correlation with reflectance (630 and 578, respectively) of the spectral regions between 500 and 700 nm, which are strongly related to the total carbon contents of the soils, suggesting that Cd and Pb are only strongly associated with organic matters, not with iron oxides or clays. Zn displays high correlations with spectral regions at around 1700 nm, which is related to the first overtone of C–H stretch from organic matter; thus binding with organic matter is the dominant form of Zn.

For example: The Pb concentration can be predicted as in the following equations.

$$\text{Pb (ppm)} = 0.0775 \text{ OM} + 35.286 \quad (1)$$

$$\text{OM (\%)} = 1.344 + 0.0197 \text{ TS}_{\text{crDOA600}} \quad (2)$$

$$\text{Pb (ppm)} = 0.0775 \times 1.344 + 0.0197 \text{ TS}_{\text{crDOA600}} + 35.286 \quad (3)$$

Therefore, the proposed “inference model” approach in light of key soil constituents that absorb VNIR-SWIR radiation would become a powerful tool to enhance the accuracy of heavy metal prediction. While pure metals do not absorb in the VNIR-SWIR region, they might be detected indirectly through their complexing with organic matter (OM), association with moieties such as iron oxides, which are detectable, or adsorption to clays that absorb light in this wavelength range (Table 6).

Validation some selected (Cd, Fe and Pb) heavy metals using XRD and SEM

The morphology and composition of the soil samples were examined by scanning electron microscopy (SEM) and x-ray diffraction (XRD). The samples can be divided into two groups, those with small ($\text{OM} \leq 3\%$ e.g., soils 12, 22, 25) and those with large ($\text{OM} > 3\%$ e.g., soils 5, 9, 15) TOC

Table 5 The regression coefficients obtained with partial least squares regression analysis between heavy metals and Fe₂O₃, Al₂O₃, and OM

Heavy metals	Wavelength, nm	Correlation coefficient			Prediction equation of heavy metals, ppm	R^2
		Fe ₂ O ₃	Al ₂ O ₃	OM		
Mn	390	0.357*	0.352*	0.191	$Mn = 56.456e^{-0.004x}$	0.59
	550	0.407*	0.400*	0.247		
	2240	0.465*	0.425*	0.251		
Cu	538	0.603*	0.496*	0.238	$Cu = 0.2376x^2 - 9.3202x + 53.16$	0.64
	1259	0.355*	0.321	0.246		
	2184	0.652*	0.422	0.266		
Zn	430	0.370*	0.366*	0.199	$Zn = 84.346x^{-0.015}$	0.69
	570	0.274	0.207	0.262		
	1700	0.351*	0.325	0.412*		
Ni	870	0.581*	0.558*	0.259	$Ni = 73.088e^{-2E-04x}$	0.65
	909	0.374*	0.348*	0.253		
	2240	0.421*	0.452*	0.243		
Pb	440	0.374*	0.371*	0.203	$Pb = 0.0124x^2 - 0.3577x + 37.897$	0.68
	578	0.268	0.310	0.613*		
	915	0.373*	0.347*	0.253		
Cd	1894	0.352*	0.322	0.266	$Cd = -0.026\ln(x) + 14.757$	0.49
	630	0.382	0.211	0.517*		
	1270	0.354*	0.320	0.246		
Co	1900	0.214	0.654*	0.243	$Co = 86.097x^{0.0122}$	0.56
	2300	0.353	0.634*	0.323		
Cr	860	0.383*	0.361*	0.374*	$Cr = 109.42e^{-0.002x}$	0.67
	1001	0.369*	0.340*	0.240		
	2376	0.632*	0.691*	0.587*		

X = wavelength, nm

* Correlation is significant at the 0.05 level

Table 6 The prediction equations obtained with regression analysis between the specific wavelengths and heavy metals utilized Fe₂O₃, Al₂O₃, and OM equations

Group	Heavy metals	Prediction through	Prediction equations of heavy metals, ppm	R^2
1	Cr	Iron oxides, clays, or organic matter	$Cr = -0.1775 \text{ Iron oxides} + 110.07$	0.67
2	Ni	Iron oxides or clays	$Ni = -0.0197 \text{ Iron oxides} + 73.56$	0.71
	Mn		$Mn = -0.065 \text{ Iron oxides} + 60.117$	0.66
3	Cu	*Iron oxides (mg kg^{-1}) = $5.07 + 66.67 \text{ TS}_{1D425}^{\text{st}} + 77.32 \text{ TS}_{1D1025}^{\text{st}} - 79.38 \text{ TS}_{1D1075}^{\text{st}}$ ($R^2 = 0.68$)	$Cu = -1.0036 \text{ Iron oxides} + 83.264$	0.68
4	Co	*Clay (%) = $28.15 + (-0.092 \text{ TS}_{cr2200})$ ($R^2 = 0.57$)	$Co = -0.0162 \text{ Clay} + 90.002$	0.72
5	Cd	*OM (%) = $1.344 + 0.0197 \text{ TS}_{crDOA600}$ ($R^2 = 0.53$)	$Cd = 0.0003 \text{ OM} + 14.685$	0.70
	Pb		$Pb = 0.0775 \text{ OM} + 35.286$	0.67
	Zn		$Zn = -0.1735 \text{ OM} + 93.598$	0.73

* Adopted from Omran 2017 (personal communication)

 TS_{cr2200} Transformed spectra peak at 2200 nm $TS_{1stD425}$ Transformed spectra using first derivative at 425 nm $TS_{crDOA600}$ The deviation of the arch of the continuum removal spectra at 600 nm

Fig. 7 The morphology and composition of the soil samples, which have Cd, Fe and Pb examined by scanning electron microscopy (SEM) images and XRD

content (Table 2). Samples in the small OM group contain particles composed of Si, Al, Fe and small amounts of K and Ca. As the OM content increases, agglomerates of particles, characteristics of SEM (Méndez et al. 2009), have been identified. Samples in the large OM group have more compact and uniform aggregates (Fig. 7). They are formed of irregularly shaped particles with rounded edges. The XRD shows that the particles contain, together with Mg, Si, Al, Fe and K, elements from soil heavy metals such as Cd and Pb, which are visible in Fig. 7.

Calcium might result from calcite (CaCO_3) in the samples. The presence of microaggregates ($<250\ \mu\text{m}$) is visible in SEM images (Fig. 7). Aggregation is generally facilitated by CaCO_3 and clays (Six et al. 2002), which the chemical analysis showed were both present in the samples. The high pH observed could also mean the presence of cations, such as Ca^{2+} , which form bridges between clay and OM (Bronick and Lal 2005).

Conclusion

VNIR-SWIR spectroscopy is a straightforward and non-destructive analytical method that can be utilized to predict heavy metals, which are spectrally feature-less. Large concentrations of heavy metals can be predicted using reflectance spectroscopy. Heavy metal elements cannot be identified with reflectance spectroscopy at concentrations under $1000\ \text{mg kg}^{-1}$. Therefore, it is harder to determine small concentrations of heavy metals in soils of Bahr El Baqar region. Organic matter masks spectral signatures, making it difficult to identify metal quantities at these wavelengths. The fingerprint region of 538, 578, 630, 870, 1900, 2240, and 2376 nm is very useful in recognizing small differences in the heavy metals of Bahr El Baqar region. A close match between several spectra in this fingerprint region constitutes strong evidence for the identity of heavy metals in the spectra. Several heavy metals (Fe, Cu, Cd, and Pb) have sharp bands between these regions. The majority of single bonds give rise to absorption bands within these frequencies and, due to similarities in their energies, strong interactions occur between neighboring bonds. PLSR is able to some extent to model heavy metals with laboratory spectra parameters. It was possible to predict heavy metals with moderate accuracy. The best coefficients of determination (R^2) between the real and predicted contents were for Mn, 0.62; Pb, 0.66; Zn, 0.66; Ni, 0.69; and Cu, 0.60. Mn, Pb, Ni, Cu, and Zn were

successfully predicted using partial least squares regression, PLSR. To increase prediction accuracy, specific wavelengths were used with OM, iron oxides, and clay. Correlation analysis revealed that most important wavelengths for prediction could be attributed to the absorption features of OM, clay and iron oxides. Group I has Cr, which is related to iron oxides, clays, and organic matter. Group II has Ni and Mn, which is related to iron oxides and clays. Group III has Cu, which is related to iron oxides. Group V has Co, which is related to clays. Group IV has Cd and Pb, which is related to organic matter.

Due to the interference effects of organic matter and low heavy metal concentration, the selection of the most sensitive original spectral bands still did not result in highly accurate prediction. Therefore, it is important to predict heavy metals through OM, iron oxides, or clays first. To identify the specific wavelengths for prediction of the studied heavy metals, the regression coefficient with OM, Al_2O_3 , and Fe_2O_3 were used. The most significant peaks for Zn prediction are found in the wavelengths of 430, 570, and 1700 nm; for Cu prediction at 538, 1259, 1500, and 2184 nm; for Pb significant peaks are at 440, 578, 915, and 1894 nm; for Cr at 860, 1001, and 2376 nm; for Ni at 870, 909, and 2240 nm; for Cd at 630 and 1270 nm; for Co at 1900 and 2300 nm.

These results indicate that it is possible to predict heavy metals in soils using the rapid and economic reflectance spectroscopy. For future investigations, hyperspectral sensor may be useful to optimize the estimation of heavy metals content at large scale. Future research likewise will focus on definitely establishing whether Cd and Pb contamination in soils can be assessed directly or indirectly in the VNIR-SWIR region using artificially soils spiked with Pb in the range 50–10,000 ppm. An urgent research need in Egypt is the development of soil spectral libraries that will improve the predictive ability of VNIR-SWIR spectroscopy for soil minerals and other soil attributes whatever the soil type. Coupling spectral libraries with other diffuse reflectance measurements of soils, such as mid-infrared reflectance spectra, will probably be the next step towards spectral sensing of soil heavy metals.

Acknowledgements The author would like to thank the late Dr. A. Abd El Razek for her assistance in the fieldwork and data collection.

References

- Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc* 43(5):772–777
- Bray J, Rossel R, McBratney A (2009) Diagnostic screening of urban soil contaminants using diffuse reflectance spectroscopy. *Aust J Soil Res* 47:433–442

- Bronick C, Lal R (2005) Soil structure and management: a review. *Geoderma* 124:3–22
- Brown DJ, Shepherd KD, Walsh MG, Dewayne Mays M, Reinsch TG (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132(3–4):273–290
- Choe E, Meer F, Ruitenbeek F, Werff H, De Smeth B, Kim K (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 Sens Environ* 112:3222–3233
- Choe E, Meer F, Ruitenbeek F, Werff H, De Smeth B, Kim K (2009) Qualitative analysis and mapping of heavy metals in an abandoned Au–Ag mine area using NIR spectroscopy. *Environ Geol* 58(3):477–482
- Conforti M, Raffaele F, Giorgio M, Tommaso C, Gabriele B (2013) Potentiality of laboratory visible and near infrared spectroscopy for determining clay content in forest soils: a case study from high forest beech (*Fagus Sylvatica*) in Calabria (Southern Italy). *Environ Qual* 11:49–64
- Conzález-Martín I, Hernández-Hierro JM, Barros-Ferreiro N, Cordon C, García-Villanova RJ (2007) Use of NIRS technology with a remote reflectance fibre-optic probe for predicting major components in bee pollen. *Talanta* 72:998–1003
- Ertlen D, Schwartz D, Trautmann M, Webster R, Brunet D (2010) Discriminating between organic matter in soil from grass and forest by near-infrared spectroscopy. *Eur J Soil Sci* 61:207–216
- 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. *J Geograph Inf Syst* 4:242–253
- Gaydon JW, Glass H, Pascoe R (2009) Method for near infrared sensor-based sorting of a copper ore. *J Near Infrared Spectrosc* 17:177–194
- Gholizadeh A, Boruvka L, Saberioon M, Kozak J, Vasat R, Nemecek K (2015a) Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. *Soil Water Res* 10(4):218–227
- Gholizadeh A, Boruvka Luboš, Vašát R, Saberioon M, Klement A, Kratina J, Tejnecký V, Drábek O (2015b) Estimation of potentially toxic elements contamination in anthropogenic soils on a brown coal mining dumpsite by reflectance spectroscopy: a case study. *PLoS One* 10(2):e0117457
- Ji J, Song Y, Yuan X, Yang Z (2010) Diffuse reflectance spectroscopy study of heavy metals in agricultural soils in the Changjiang River Delta, China. In: 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane
- Kemper T, Sommer S (2002) Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ Sci Technol* 36:2742–2747
- Kemper T, Sommer S (2003) Mapping and monitoring of residual heavy metal contamination and acidification risk after the Aznalcollar mining accident (Andalusia, Spain) using field and airborne hyperspectral data. In: The 3rd EARSeL Workshop on Imaging Spectroscopy, Herrsching
- Malley DF, Williams PC (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(12):3461–3467. doi:10.1021/es970214p
- Malley DF, Martin PD, Ben-Dor E (2004) Application in analysis of soils. In: Roberts CA (ed) Near-infrared spectroscopy in agriculture. *Agronomy Monograph* 44 American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, USA, pp 729–784
- Martens H, Næs T (1989) Multivariate calibration. Wiley, New York
- Martens H, Jensen SA, Geladi P (1983) Multivariate linearity transformations for near infrared reflectance spectroscopy. In: Christie OHJ (ed) *Proc Nordic Symp Applied Statistics*, Stokkland Forlag, Stavanger, Norway, pp 205–234
- Méndez A, Fidalgo J, Guerrero F, Gascó G (2009) Characterization and pyrolysis behaviour of different paper mill waste materials. *J Anal Appl Pyrol* 86:66–73
- Minasny B, Tranter G, McBratney A, Brough D, Murphy B (2009) Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* 153:155–162
- Morón A, Cozzolino D (2003) Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. *J Near Infrared Spectrosc* 11:145–154
- Moros J, de Vallejuelo SFO, Gredilla A, de Diego A, Madariaga JM, Garrigues S, de la Guardia M (2009) 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* 43:9314–9320
- Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* 46(2):99–118
- Omran ESE (2016) Environmental modelling of heavy metals using pollution indices and multivariate techniques in the soils of Bahr El Baqar, Egypt. *Modeling Earth Syst Environ* 2(3):1–17. doi:10.1007/s40808-016-0178-7
- Omran ESE, Abd El Razek AA (2012) Mapping and screening risk assessment of heavy metals concentrations in soils of the Bahr El-Baker Region, Egypt. *J Soil Sci Environ Manag* 6(7):182–195
- Pandit M, Filippelli M, Li L (2010) Estimation of heavy-metal contamination in soil using reflectance spectroscopy and partial least-squares regression. *Int J Remote Sens* 31(15):4111–4123
- Park J, Shin W (2006) Immobilization of Pb contaminated soil using modified clay. *Water Pract Technol* 1:1–10
- Peng Y, Knadel M, Gislum R, Deng F, Norgaard T, Wollesen de Jonge L, Moldrup P, Humlekrog Greve M (2013) Predicting soil organic carbon at field scale using a national soil spectral library. *J Near Infrared Spectrosc* 21:213–222
- Pirie A, Singh B, 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(6):713–721
- Ren HY, Zhuang DF, Singh AN, Pan J, Qiu DS, Shi RH (2009) Estimation of As and Cu contamination in agricultural soils around a mining area by reflectance spectroscopy: a case study. *Pedosphere* 19(6):719–726
- Rinnan A, van den Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal Chem* 28:1201–1222
- Rizzo R, Demattê JAM, da Silva TF (2014) Using numerical classification of profiles based on Vis-Nir spectra to distinguish soils from the Piracicaba Region, Brazil. *Braz R Bras Ci Solo* 38:372–385
- Saad AK (1997) Environmental hydrogeologic impacts groundwater withdrawal in the eastern Nile Delta region with emphasis on groundwater pollution potential. Ph D Thesis, Institute of Environmental Studies Ain Shams Univ, pp 232
- Sakizadeh M, Mirzaei R, Ghorbani H (2015) The Extent and Prediction of Heavy Metal Pollution in Soils of Shahrood and Damghan, Iran. *Bull Environ Contam Toxicol* 95(6):770–776. doi:10.1007/s00128-015-1632-3
- Schwartz G, Eshel G, Ben-Dor E (2012) Reflectance spectroscopy as a rapid tool for monitoring contaminated soil. Porter School of environmental studies, Tel Aviv University PhD thesis

- Siebielec G, McCarty GW, Stuczynski TI, Reeves JB (2004) Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *J Environ Qual* 33(6):2056–2069
- Six J, Feller C, Denef K, Ogle S, de Moraes Sa J (2002) Soil organic matter, biota and aggregation in temperate and tropical soils effects of no-tillage. *Agron, EDP Sci* 22(7–8):755–775
- Stenberg B, Viscarra Rossel R, Mouazen AM, Wetterlind J (2010) Visible and near infrared spectroscopy in soil science. *Adv Agron* 107:163–215
- Summers D (2009) Discriminating and Mapping Soil Variability with Hyperspectral Reflectance Data. Faculty of science, School of earth and environmental science, Adelaide University PhD thesis
- Tekin Y, Tumsavas Z, Mouazen A (2012) Effect of moisture content on prediction of organic carbon and pH using visible and near infrared spectroscopy. *Soil Sci Soc Am J* 76(1):188–198
- Todorova M, Mouazen AM, Lange H, Atanassova S (2014) Potential of near-infrared spectroscopy for measurement of heavy metals in soil as affected by calibration set size. *Water Air Soil Pollut* 225:2036. doi:[10.1007/s11270-014-2036-4](https://doi.org/10.1007/s11270-014-2036-4)
- Umetrics (2002) User Guide to SIMCA P. Umetrics (ed), Umetrics
- Vasques GM, Grunwald S, Sickman O (2008) Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* 146:14–25
- Viscarra Rossel RA (2008) The soil spectroscopy group and the development of a global spectral library. In: 3rd global workshop on digital soil mapping. Utah State University, Logan, Utah, 30 September–3 October 2008
- Vohland M, Bossung C, Frund H (2009) A spectroscopic approach to assess trace-heavy metal contents in contaminated floodplain soils via spectrally active soil components. *Plant Nutr Soil Sci* 172(2):201–209
- Walkley A, Black A (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci Soc Am J* 37:29–38
- Wetterlind J, Bo S, Jonsson A (2008) Near infrared reflectance spectroscopy compared with soil clay and organic matter content for estimating within-field variation in N uptake in cereals. *Plant Soil* 302(1–2):317–327
- Wetterlind J, Stenberg B, Viscarra Rossel, Raphael A (2013) Soil analysis using visible and near infrared spectroscopy. In: FJM Maathuis (ed) *Plant mineral nutrients: methods and protocols*. Humana Press, Springer, New York, pp 95–107 (**Published in serie: Methods in molecular biology, nr 953**)
- Winkelmann KH (2005) On the applicability of imaging spectrometry for the detection and investigation of contaminated sites with particular consideration given to the detection of fuel hydrocarbon contaminants in soil. Unpublished PhD Thesis Brandenburg University of Technology
- Wu Y, Chen J, Wu X, Tian Q (2005) Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. *Appl Geochem* 20:1051–1059
- 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 Sci Soc Am J* 71:918–926
- Wu C, Jacobson A, Laba M, Baveye P (2009) Accounting for surface roughness effects in the near-infrared reflectance sensing of soils. *Geoderma* 152:171–180. doi:[10.1016/j.geoderma.2009.06.002](https://doi.org/10.1016/j.geoderma.2009.06.002)
- Yaolin L, Wei L, Guofeng W, Xinguo X (2011) Feasibility of estimating heavy metal contaminations in floodplain soils using laboratory-based hyperspectral data-a case study along Le'an River, China. *Geo-spatial Information Science* 14(1):10–16
- Zhang Z, Wen J, Zhao D (2010) Band selection method for retrieving soil lead content with hyperspectral remote sensing data. *The Society of Photo-Optical Instrumentation Engineers (SPIE)* 7831