ORIGINAL PAPER

Mid-infrared spectroscopy and partial least-squares regression to estimate soil arsenic at a highly variable arsenic-contaminated site

N. K. Niazi · B. Singh · B. Minasny

Received: 12 November 2013/Revised: 1 March 2014/Accepted: 18 March 2014/Published online: 29 April 2014 © Islamic Azad University (IAU) 2014

Abstract The potential of mid-infrared spectroscopy in combination with partial least-squares regression was investigated to estimate total and phosphate-extractable arsenic contents in soil samples collected from a highly variable arsenic-contaminated disused cattle-dip site. Principal component analysis was performed prior to mid-infrared partial least-squares analysis to identify spectral outliers in the absorbance spectra of soil samples. The mid-infrared partial least-squares calibration model (n = 149) excluding spectral outliers showed an acceptable reliability (coefficient of determination, $R_c^2 = 0.75$ (P < 0.01); ratio of performance to interquartile distance, $RPIQ_c = 2.20$) to estimate total soil arsenic. For total soil arsenic, the validation of final calibration model using 149 unknown samples also resulted in a good acceptability with $R_v^2 = 0.67$ (P < 0.05) and $RPIQ_{v} = 2.01$. However, the mid-infrared partial leastsquares calibration model based on phosphate-extractable arsenic was not acceptable to estimate the extractable (bioavailable) arsenic content in soil ($R_c^2 = 0.13$ (P > 0.05); $RPIQ_c = 1.37$; n = 149). The results show that the midinfrared partial least-squares prediction model based on total arsenic can provide a rapid estimate of soil arsenic content by

Electronic supplementary material The online version of this article (doi:10.1007/s13762-014-0580-5) contains supplementary material, which is available to authorized users.

N. K. Niazi · B. Singh · B. Minasny Department of Environmental Sciences, Faculty of Agriculture and Environment, The University of Sydney, Sydney, NSW 2006, Australia

Present Address:

N. K. Niazi (🖂)

Institute of Soil and Environmental Sciences, University of Agriculture Faisalabad, Faisalabad 38040, Pakistan e-mail: nabeelkniazi@gmail.com; nabeel.niazi@uaf.edu.pk taking into account the integrated effects of adsorbed arsenic, arsenic-bearing minerals and arsenic associated with organic components in the soils. This approach can be useful to estimate total soil arsenic in situations, where analysis of a large number of samples is required for a single soil type and/ or to monitor changes in soil arsenic content following (phyto)remediation at a particular site.

Keywords Mid-infrared · Partial least-squares · Principal component · Cattle-dip sites · (Phyto)remediation · Prediction model · Contamination

Introduction

Arsenic (As) is an extremely toxic and carcinogenic element, and therefore, contamination of soils with As is a major environmental threat (Smith et al. 1998). The historical application of As-based pesticides at disused cattle-dip sites across northern region of New South Wales (NSW) state in Australia has resulted in the contamination of surrounding soils with As concentrations up to 14,000 mg kg⁻¹ (McLaren et al. 1998; Smith et al. 1998; Niazi et al. 2011b). Arsenic concentration in soil in the vicinity of dip sites varies significantly on a small scale, and analyses of a large number of soil samples are required to estimate the contamination level and distribution of As at a particular site (Kimber et al. 2002; Niazi et al. 2011c, 2012). Estimation of the total As content in soils is imperative to assess the extent of contamination and risks posed to the environment, and to evaluate the impact of remediation strategies (Smith et al. 1998; Mandal and Suzuki 2002).

The existing standard methods for estimating total As concentration in soils are typically based on acid digestion procedures, which requires the use of mixtures of



concentrated acids (HNO₃, HClO₄, HCl, H₂SO₄ and HF) (Huang and Fujii 1996; Chen and Ma 1998). Total As concentration in soil digests is generally measured by hydride generation atomic absorption spectrometer or inductively coupled plasma-atomic emission spectrometer. Although chemical techniques used to determine the total soil As content in soil are highly accurate, the procedures are hazardous and time-consuming. X-ray fluorescence (XRF) spectroscopy is also used to precisely determine the total As content in soil and sediment samples (Hubert 1983), however, the sample preparation in this technique is also time-consuming and access to the equipment is not widely available. There is a need for a rapid, inexpensive and easy-to-use method, in order to routinely estimate total As in large number of soil samples at a given site, to monitor changes in soil As content at a specific site or to estimate variability in total As in soil at a particular site (Kimber et al. 2002; Niazi et al. 2011c, 2012).

Mid-infrared (MIR) spectroscopy in combination with chemometrics has been used as a rapid method to estimate the various soil physicochemical properties such as, pH, cation exchange capacity, clay, sand and silt contents, and sorption coefficient of pesticides in soil (Janik and Skjemstad 1995; Islam et al. 2003; Pirie et al. 2005; Minasny et al. 2008). The MIR spectra of soils contain information about the amount and composition of both organic and inorganic components of the soil. The qualitative information in the MIR spectra of soil is characterized by strong stretching and bending vibrations of functional groups such as, carboxyl, hydroxyl, amide, alkyl and aromatic groups present in soil organic matter (Janik and Skjemstad 1995; Janik et al. 1995; Islam et al. 2003). The vibrational features (e.g. Fe-O, Al-O, Si-O and -OH) of soil minerals also contribute to the MIR spectra (Johnston and Aochi 1996; Cornell and Schwertmann 2003; Jia et al. 2007; Carabante et al. 2010). It is difficult to quantitatively interpret spectral information; however, using chemometrics approaches, particularly partial least-squares (PLS) regression, this problem has been overcome (Janik and Skjemstad 1995; Forouzangohar et al. 2008).

Partial least-squares regression in combination with the spectroscopic techniques has been successfully applied to construct the predictive models; for example, for estimating various soil properties and pesticide sorption coefficients (Geladi 1988; Haaland and Thomas 1988; Bellon-Maurel et al. 2010; Sanghavi and Srivastava 2010). The method is based on the assumption of a multilinear relationship between the predictor variables (e.g. absorbance peaks in the spectra) and the dependent variable of interest (e.g. As concentration in soil). The PLS regression analysis reduces the number of independent variables (i.e. spectra of samples) into a limited number of predictor variables, called PLS factors (Haaland and Thomas 1988). Principal

component analysis (PCA) can provide a basis for the MIR-PLS analysis as it can be useful to depict the existence of any spectral outliers in the absorbance spectral data of soil samples (Forouzangohar et al. 2008). It is a form of latent variable regression in which the spectral data are resolved into a set of orthogonal components, which represent most of the variability in the original data (Mark 1992; Sanghavi et al. 2013a, b).

Infrared spectroscopy has been used to evaluate the adsorption mechanisms of As on the surface of Fe/Al oxides and clay minerals in soils (Goldberg and Johnston 2001; Jia et al. 2007; Carabante et al. 2010). Goldberg and Johnston (2001) used attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy to determine the adsorption mechanism of arsenate on the surface of amorphous Fe and Al oxides. They reported that arsenate sorption bands were present at around \sim 789 and 824 cm⁻¹ on amorphous Fe oxide and at ~850 and 862 cm⁻¹ on the surface of amorphous Al oxide. Jia et al. (2007) investigated the effect of pH on arsenate adsorption on ferrihydrite using FTIR spectroscopy; they attributed spectral bands at ~808, 810, 833, 838 and 876 cm⁻¹ to arsenate– Fe oxides adsorption complexes depending on the pH (3-8) and coverage density of As.

Limited efforts have been devoted to explore the potential of MIR spectroscopy in conjunction with chemometrics (e.g. PLS regression) to predict heavy metal(loid)s in agricultural soils/estuarine sediments and soils contaminated from mining activities (Moros et al. 2009; Niazi et al. 2011b; Song et al. 2012). Song et al. (2012) employed MIR diffuse reflectance spectroscopy to estimate total As and heavy metal (Cr, Cu, Cd and Pb) contents in agricultural soils of Changjiang River Delta, in China. The authors identified significant correlation of Cr, Cu and As with the spectral region associated with Fe oxides, clay minerals and organic matter. Moros et al. (2009) were able to develop multivariate PLS calibration models of acceptable accuracies using both near infrared (NIR) (r = 0.86, RMSEC = 10, RMSEP = 25 and RPD = 1.9) and ATR-MIR (r = 0.85, RMSEC = 17,RMSEP = 26 and RPD = 1.8) spectra. However, a selective particle size fraction (<63 µm) was used in the study. Contrary to this, As prediction was relatively poor $(R^2 = 0.46)$ RMSEC = 1.86, RMSEP = 1.61and RPD = 1.14) in the PLS regression model in agricultural soils of Changjiang River Delta, China (Song et al. 2012). In the current study, we have used a MIR-PLS-based approach to predict total soil As content in a single soil type (Oxisol) spread over a relatively small area $(\sim 100 \text{ m}^2)$ with the hypothesis that MIR spectroscopy and PLS regression-based model will provide estimates of total As in soil with an acceptable reliability. The development of the MIR-PLS-based approach to predict total soil As content might be able to provide a cheaper and timeeffective method compared to conventional chemical procedures. This method could be particularly useful for situations where a large number of soil samples need to be analysed (Niazi et al. 2011c), or repetitive analyses are needed for the samples (Niazi et al. 2012).

The primary objectives of this study were to (1) develop a site-specific predictive model using the combination of MIR spectroscopy and PLS regression to estimate total and phosphate-extractable As contents in soil and (2) apply the predictive model to unknown soil samples (not used in developing the model), in order to estimate As concentration in soils and to evaluate the prediction quality of the model.

The research was carried out at the Department of Environmental Sciences, Faculty of Agriculture and Environment, The University of Sydney, Sydney, Australia (from January 2010 to 31 July 2011).

Materials and methods

Soil samples

Arsenic-contaminated soil samples were taken from a disused cattle-dip site located at Wollongbar as described elsewhere (Niazi et al. 2011a). The soil samples were taken from three depths, i.e. 0–20, 20–40 and 40–60 cm. All soil samples (n = 304) were dried at 40 °C and ground to obtain <200 µm prior to chemical and spectral analyses. The pH of the soil samples at the site was 4.82 (1:5, soil: water), free Fe content 16 %, cation exchange capacity 88 mmol_c kg⁻¹ and organic carbon content 4.5 %.

Soil digestion to measure total As

The soil samples (~ 0.25 g) were digested in a mixture of concentrated hydrofluoric (HF), sulphuric (H₂SO₄), perchloric (HClO₄), nitric (HNO₃) and 1.2 M hydrochloric (HCl) acids (Huang and Fujii 1996). The soil samples were soaked in a mixture HNO₃ and HCl overnight. The soaked samples were then mixed with H₂SO₄, HF and HClO₄ and heated at 120 °C overnight until the mixture was dried. After cooling for 5 min, the digest was dissolved in 25 mL of 6 M HCl and final volume made to 50 mL by adding E-pure[®] water. The digests were analysed for As concentration (mg As kg⁻¹ soil dry weight) using an inductively coupled plasma-atomic emission spectrometer (ICP-AES, Varian[®] Vista AX CCD).

Phosphate-extractable As in soil

Potassium dihydrogen phosphate solution (0.5 M KH_2PO_4 , 1:25 soil:solution ratio, 4 h shaking time at room temperature) was used to extract the specifically sorbed (bioavailable) As in the soil samples (Niazi et al. 2011a). The concentration of phosphate-extractable As (mg As kg^{-1} soil dry weight) in the extracts was measured using a hydride generation atomic absorption spectrometer (HG-AAS, Varian[®] Spectraa 220Z).

The relative standard deviation (RSD) was <2% for the ICP-AES analysis and <1.5% for HG-AAS analysis.

MIR spectroscopy

Finely ground (<200 µm) soil samples were dried at 105 °C overnight before the spectroscopic analysis. Diffuse reflectance (MIR) spectra were obtained using a Fourier transform infrared (FTIR) spectrometer (Bruker TENSOR 37; Ettlingen, Germany) with the Praying MantisTM diffuse reflectance accessory (Harrick Scientific Products, Pleasantville, NY). Approximately 0.4 g of finely ground soil sample was placed in a 4 mm diameter cup without compression and levelled using a stainless steel blade. Scans were collected in the wavenumber range of 4,000–400 cm^{-1} at 8 cm^{-1} resolution; this range corresponds to the frequency range of 2,500-20,000 nm. Potassium bromide (KBr) powder was used as a background reference and assumed to have a reflectance of 1 (100 %). Data in whole spectral range $(4,000-400 \text{ cm}^{-1})$ were used for the principal component analysis (PCA) and partial least-squares (PLS) analysis. The reflectance (R) values were log-transformed $\left[\log(1/R)\right]$ to represent spectra in absorbance units using OPUS[®] software supplied with the spectrometer. Before the chemometric analyses (PCA and PLS), the data were pre-treated by taking first derivative of the absorbance spectra for the baseline removal (see Supplementary Material, Fig. A.1) using the Savitzky-Golay method, with a second-order polynomial and 9 smoothing points. The PCA and PLS regression were performed using the JMP version 8.

Principal component analysis and partial least-squares regression

Principal component analysis was carried out to detect spectral outliers in the data, prior to the development of a prediction model using PLS regression (Fig. 1). Primarily, PCA is a feature reduction mathematical procedure that can be useful to provide basis for the other multivariate linear regression analyses (e.g. PLS regression) (Pirie et al. 2005; Forouzangohar et al. 2008; Kookana et al. 2008). In a PCA, the eigenvectors and their respective eigenvalues are calculated. The data are then reduced to a smaller group of principal components (Haaland and Thomas 1988; Forouzangohar et al. 2008). The loadings in PLS analysis represent the components (factors) in the spectra which are highly correlated with measured data (Haaland and Thomas 1988; Janik and Skjemstad 1995).





Fig. 1 Principal component analysis (PCA) of the first derivative of MIR absorbance spectra of all soil samples (n = 304). The green filled circles are the six spectral outliers in the soil samples. *PC-1* principal component-1, *PC-2* principal component-2

The data set contained 304 soil samples including the 6 spectral outliers identified using PCA (Fig. 1). After excluding the spectral outliers, the remaining 298 samples were randomly split into two sets, each containing 149 number of samples. One set of samples (calibration set; n = 149) was used for constructing MIR-PLS calibration model, and the second set (n = 149) was used for the validation of the calibration model, referred to as the validation set (Fig. 2).

The number of components used in the PLS calibration model was determined using leave-one-out cross validation (LOOCV). The LOOCV procedure estimates the prediction error by removing samples one by one from the calibration samples data and predicting them as unknown samples using the remaining samples in the data.

It is also important to estimate the errors associated with the prediction model using a set of soil samples not used in building the model, i.e. validation set (n = 149) as mentioned earlier. The prediction equation obtained from the calibration model was applied to the validation set in order to predict As concentration in soil samples. To evaluate the efficiency of the prediction model, predicted As concentration values were plotted against the measured soil As concentrations. The prediction ability of calibration and validation models was assessed by determining coefficient of determination in the calibration (R_c^2) and validation (R_y^2) models, standard error of cross validation (SECV) in the calibration model, standard error of prediction (SEP) in the validation model, and ratio of performance to interquartile distance (RPIQ) in the calibration (RPIQ_c) and validation $(RPIQ_v)$ models (Islam et al. 2003; Pirie et al. 2005;







Fig. 2 The frequency distribution (histograms and box-plots) of measured total As concentration in the soil samples used for the MIR-PLS regression analysis for the development of \mathbf{a} final calibration model and for \mathbf{b} validation of the calibration model

Bellon-Maurel et al. 2010). The SECV and SEP were calculated using Eq. 1:

SECV and SEP =
$$\sqrt{\frac{\sum (z_i - \hat{z}_i)^2}{n}}$$
 (1)

where z_i is the measured data, \hat{z}_i is the model predicted data for the sample *i*, and *n* is the number of samples in the calibration or validation set. In the case of calibration model, RPIQ_c is the ratio of interquartile distance $(IQ = Q_3 - Q_1)$ of the measured data in the calibration set to SECV (Eq. 2). While SEP is used to estimate RPIQ_v in the validation model (Eq. 3), $Q_3 - Q_1$ is linked to the measured data in the validation set (Q_1 is the value below which we can find 25 % of the samples; Q_3 is the value below which 75 % of the samples fall) (Bellon-Maurel et al. 2010).

$$\operatorname{RPIQ}_{c} = \frac{(Q_{3} - Q_{1})_{c}}{\operatorname{SECV}}$$
(2)

$$\operatorname{RPIQ}_{v} = \frac{(Q_{3} - Q_{1})_{v}}{\operatorname{SEP}}$$
(3)

Bellon-Maurel et al. (2010) indicated that the RPIQ index is a more reliable and acceptable way for standardizing the SECV/SEP than the ratio of prediction to deviation (RPD), with respect to population spread. The use of RPIQ value is particularly vital for soil sample sets, which usually show a skewed distribution, as found in our study for soil As concentration (Supplementary Material, Fig. A.2; Fig. 2). To calculate the RPIQ value, standard deviation (SD) is replaced by $Q_3 - Q_1$, which accounts much better for the spread of the population (Bellon-Maurel et al. 2010).

Results and discussion

Total As concentration was highly variable in both the $(89-1,761 \text{ mg kg}^{-1})$ calibration and validation $(70-1,486 \text{ mg kg}^{-1})$ sets of the study (Fig. 2). The distribution of As in the two (calibration and validation) data sets was similar with mean As concentrations of 531 and 509 mg kg^{-1} , respectively (Fig. 2). A slightly positive skewness was observed in both the calibration (1.21) and validation (1.02) samples. Phosphate-extractable As concentration data also showed high variability in the calibration and validation sets $(13-141 \text{ and } 7-117 \text{ mg kg}^{-1})$, respectively) (Supplementary Material, Fig. A.2). The variability in the measured soil As content at the site has been attributed to contamination processes, including the cattle-dipping process, pumping out of the dip fluid and disposal of As-containing dip sediment in the vicinity of the cattle-dip (Kimber et al. 2002; Niazi et al. 2011b, c).

MIR spectral features

MIR spectra of the 6 (surface and subsurface) representative first derivative absorbance of the soil samples are shown in Fig. 3a. The spectra show several peaks that were attributed to mineral and organic components in the soil. For example, the presence of hydroxyl (–OH) stretching and bending bands in the spectral region from 3,695 to 3,394 and 764–1,022 cm⁻¹ was due to the presence of structural –OH groups in clay minerals (kaolinite) and Fe/Al oxides (goethite/gibbsite), respectively (Schroeder 2002; Cornell and Schwertmann 2003; Song et al. 2012).

The soil at the site is an Oxisol, and according to the Australian Soil Classification, the soil is a Red Ferrosol (Isbell 2002). The clay fraction of the soil at the site is composed of kaolinite, hydroxy-interlayered vermiculite and metal oxides including: goethite, hematite, gibbsite and anatase (Mckenzie et al. 2004; Niazi et al. 2011b). The spectral peaks observed at 3,695, 3,657 and 3,626 cm^{-1} were attributed to the presence of -OH stretching band for kaolinite (Fig. 3a) (Schroeder 2002; Cornell and Schwertmann 2003; Janik et al. 2007; Song et al. 2012). The bands observed at 413, 428, 470 and 522 cm^{-1} in the spectra of soil samples (Fig. 3a) were attributed to the Fe-O/Al-O bending vibrations (Schroeder 2002; Cornell and Schwertmann 2003; Ibrahim et al. 2008). The presence of a band at 3.142 cm^{-1} in the spectra of samples tested in this study was attributed to -OH stretching in Fe oxides (i.e. goethite) (Cornell and Schwertmann 2003).

The spectral band identified at $\sim 1,720 \text{ cm}^{-1}$ (Fig. 3a) was associated with the presence of carboxylic (-COOH) functional group in the soil organic fraction. The results concur with those obtained by Janik and Skjemstad (1995) and Islam et al. (2003), who observed spectral band at $1,730 \text{ cm}^{-1}$ for the presence of -COOH group in the spectra of soil samples. The carboxylic acids (-COOH) are the low molecular weight organic acids present in the soil organic matter fraction and have the ability to make complexes with the metal cations, such as Fe and Al (Smith et al. 2000; Grafe et al. 2001, 2002; Mandal and Suzuki 2002). Arsenic (as arsenate) adsorption to organic acids/ humic substances is enhanced by the presence of cations, particularly Fe and Al, whereby they act as bridging complexes for arsenate (Lin et al. 2004). The positive contribution of -COOH group of organic fraction to As sorption was revealed by the sharp peak at $\sim 1,720 \text{ cm}^{-1}$ (see Fig. 3b). The bands at 2,902 and 2,989 cm^{-1} were attributed to the aliphatic (-CH₂) stretching vibrations in the soil organic fraction (Fig. 3a, b) (Janik and Skjemstad 1995; Song et al. 2012).

Figure 3b shows the strength of relationship between the model coefficients and the spectral regions of the soil samples. This represents the significance of soil components which were most strongly and positively correlated with the As concentration in soil and explain variation in the prediction model, as described above. The positive peaks at 3,695 and 3,657 cm⁻¹ (Fig. 3b) showed that kaolinite strongly contributed in the MIR-PLS model to estimate As in the soil samples. The presence of very strong and positive sharp peaks (Fig. 3b) at around 413, 428, 470 and 522 cm⁻¹ (as mentioned above) reveals the



Fig. 3 a The mid-infrared (MIR) absorbance spectra of the six representative soil samples collected from the surface (0–20 cm) and the subsurface (20–40 and 40–60) depths, and **b** the MIR partial least-squares (PLS) regression coefficients obtained in the calibration model (n = 149), in relation to the wavenumber (4,000–400 cm⁻¹) used for the MIR spectroscopy



strong contribution of Fe/Al oxide minerals, including goethite, hematite and gibbsite in the PLS regression loadings. The contribution of Fe/Al oxides (goethite, hematite and gibbsite) and kaolinite spectral bands in the PLS loadings (model coefficients) to predict soil As content is attributed to the high adsorption capacity of these minerals, particularly Fe/Al oxides (Goldberg and Johnston 2001; Goldberg 2002; Chabot et al. 2009; Song et al. 2012).

Goldberg and Johnston (2001) used FTIR spectroscopy to investigate arsenate sorption complexes on the surface of synthetic Fe and Al oxides. Arsenate adsorption was identified by the presence of spectral bands at 789 and 824 cm^{-1} on the surface of Fe oxides, and the bands appeared at 850 and 862 cm⁻¹ for arsenate adsorption on the surface of Al oxides. Jia et al. (2007) reported spectral bands in the region from 808 to 838 cm⁻¹ for arsenate adsorption on ferrihydrite. In the present study, a small band at 829 cm⁻¹ could be associated with the existence of Fe–O–As and/or Al–O–As groups due to the adsorption of arsenate on the surface of Fe or Al oxides in the soil (Fig. 3a). The presence of small positive peaks (Fig. 3b) at



Fig. 4 The measured values of total soil As concentration are plotted against the MIR-PLS model predicted As values; **a** final calibration model developed excluding the six spectral outliers (n = 149) and using the calibration set data as shown in Fig. 2; **b** validation of the final calibration model constructed with n = 149 samples. The R_c^2 and R_v^2 are coefficient of determination in the calibration and validation sets, respectively, *SECV* standard error of cross validation, *SEP* standard error of prediction, *RPIQ_c* and *RPIQ_v* represent ratio of performance to interquartile distance in the calibration and validation sets, respectively

around 764, 795, 798, 814 cm^{-1} probably also indicates the contribution of specifically sorbed arsenate on the surface of Fe/Al oxides (goethite, hematite and ferrihydrite/ gibbsite) (Wenzel et al. 2001; Niazi et al. 2011b; Song et al. 2012).

Principal component analysis

Principal component analysis was performed on the (first derivative pre-treated) absorbance spectra of all soil samples (n = 304), which enabled us to detect 6 spectral outliers in the data (Fig. 1, green colour filled circles). The absorbance spectra of these (spectral outlier) samples were included in the preliminary calibration model (data not shown), however, excluded from the PLS analysis performed to develop the final calibration model (Fig. 4a).

Calibration model using MIR spectra and PLS regression for total As

A preliminary calibration model was generated for total As using MIR absorbance spectra of soil samples (n = 155) including the 6 outliers that were identified from the PCA (data not shown), as mentioned earlier. This was done to determine the effect of outliers on the prediction model.

The preliminary calibration model (n = 155) explained 35 % of variation ($R^2 = 0.35$, P > 0.05) between the measured and model predicted total As concentration in the soil samples. The calculated RPIQ and SECV values for this preliminary calibration model were 1.42 and 244 mg kg⁻¹, respectively. The MIR-PLS final calibration model (Fig. 4a) was re-constructed using the first derivative absorbance spectra of 149 samples (excluding the six outliers) in the calibration set (see Fig. 2a). The statistical assessment of the MIR-PLS calibration model showed that the model performed much better than the preliminary calibration model, as demonstrated by the reduced SECV (148) and increased R_c^2 (P < 0.01) and RPIQ_c values (0.75) and 2.20, respectively; Fig. 4a) (Forouzangohar et al. 2008; Bellon-Maurel et al. 2010). The results show that identification of spectral outliers using PCA and their subsequent exclusion from the prediction model substantially increased the efficiency and robustness of the MIR-PLS calibration model for total As. The measured and model predicted As concentration values were uniformly distributed around the reference equation (1:1; dotted) line (Fig. 4a).

Validation of MIR-PLS model for total As

The validity of the resulting preliminary and final calibration models built using the calibration set of 155 and 149 samples, respectively, was tested using a validation set (n = 149; Fig. 2b). The prediction equation obtained from the final calibration model (n = 149; Fig. 4a) was applied to the validation set, in order to calculate the statistical parameters related to the estimation of total As concentrations in unknown soil samples (Fig. 4b). The results showed that 67 % of variation ($R_v^2 = 0.67$, P < 0.05) between the measured and predicted total As concentration values was described by the MIR-PLS calibration model (Fig. 4b). The resultant RPIQ_v value of 2.01 demonstrated an acceptable reliability of the proposed MIR-PLS prediction model (Islam et al. 2003; Bellon-Maurel et al. 2010). The RPIQ value is considered to be highly useful to assess the prediction quality and is much better index to determine the prediction model reliability and acceptability compared to RPD value, as described earlier. Based on the RPIQ value and the purpose of prediction, the



appropriateness of a calibration model can be decided (Bellon-Maurel et al. 2010). For soil analysis such as soil As concentrations in this study, RPIQ values can be divided into three categories: poor <1.6; acceptable 1.6–2.0; and >2.0 as excellent (Bellon-Maurel et al. 2010). The application of the preliminary calibration model (including 6 spectral outliers) on the same validation set explained only 38 % variation ($R^2 = 0.38$, P > 0.05) between the measured and predicted total As concentration values; the SEP increased to 237 mg kg⁻¹ and RPIQ dropped down to 1.46. It is evident from the results that the inclusion of outliers led to development of poor MIR-PLS calibration and validation models, which were limited to estimate total As above ~900 mg kg⁻¹ As concentration in soil at the site (data not shown).

MIR-PLS model based on phosphate-extractable soil As concentration

A calibration model was also developed using the phosphate-extractable As concentration data (data not shown) of soil samples investigated in this study. The calibration model (n = 149) explained no variation $(R_c^2 = 0.13)$, P > 0.05) between the measured and model predicted phosphate-extractable As concentration values. The calculated RPIQ_c and SECV values for phosphate-extractable As based on the MIR-PLS calibration model were 1.29 and 23.02 mg kg⁻¹, respectively, indicating that the proposed MIR-PLS prediction model for phosphate-extractable As estimation in soils is unacceptable. Phosphate-extractable As is considered to represent arsenate sorbed on to the surfaces of Fe/Al oxides (e.g. ferrihydrite and goethite/ gibbsite), as discussed earlier (Niazi et al. 2011c). The poor predictability of phosphate-extractable As could be due to the relatively narrow range for the phosphate-extractable As in the studied samples and long-term ageing of Asadsorbed samples in the field. The site has been contaminated over 40 years, and with ageing (or residence time), adsorbed As must have changed to some stable phases and the extractability of As from such phases is relatively poor compared to relatively freshly adsorbed As on Fe/Al oxides (O'Reilly et al. 2001; Niazi et al. 2011b). O'Reilly et al. (2001) evaluated the effect of residence time on desorption of arsenate from goethite surface using phosphate solution, whereby arsenate sorption was carried out for a period of up to ~ 12 months. The authors observed that over 40 % of arsenate was desorbed from goethite using phosphate suggesting that ageing of arsenate (from 5 to 12 months) led to the formation of stable innersphere (arsenate-goethite) complexes.

ATR-FTIR could be employed to improve spectral signals from As sorption at low concentrations (i.e. phosphateextractable As), in order to develop an acceptable and robust prediction model. The predictability for total As is better than the phosphate-extractable As because the MIR-PLS prediction model based on the total As content took into account spectral regions represented by arsenate adsorbed onto various minerals and organic matter as well as As present in minerals and/or organic components in the soil (Fig. 3a) (Song et al. 2012).

Conclusion

Our results showed that combination of MIR spectroscopy and PLS regression can be useful in estimating total As concentration in soil for the contaminated sites containing highly variable soil As content, such as As in soil adjoining the cattle-dip sites. Significant improvement in the prediction statistics of the MIR-PLS calibration model was attained by removing 6 spectral outliers, which were identified by the PCA of the MIR spectra. The MIR-PLS prediction model has the capability to estimate total As content based on the spectral bands originating from adsorbed As, As present in minerals and organic components of soils. The soil components, particularly Fe and Al oxides and kaolinite, appeared to have pertinent roles in the sorption of As in the soil samples used in this study.

The MIR-PLS-based predictive model has possible application to estimate the total soil As in situations where total As concentration is required for a large number of soil samples, for example (1) for monitoring the effect of phytoremediation technology on soil As content at a particular site (Niazi et al. 2011c) or (2) to estimate the variability in total soil As at contaminated sites, such as cattledip sites (Niazi et al. 2011a). The MIR-PLS model is site specific, and a more robust prediction model is needed for estimating total soil As at sites for other soil types and where As may also be present in different mineral forms, e.g. scorodite and arsenopyrite. The model is not useful for identifying and differentiating the valence forms of As (arsenate, As^{V} /arsenite, As^{III}) in soils.

Acknowledgments The first author gratefully acknowledges the Higher Education Commission of Pakistan for the award of a PhD scholarship. The project has been assisted by the NSW Government through its Environmental Trust. Thanks to Dr Lukas van Zwieten, Stephen Kimber, George Nastase, Joshua Rust, Scott Petty, Desmond Cook and Victor Warren at the Wollongbar Agricultural Research Institute, Wollongbar for field assistance.

References

Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, Roger JM, McBratney A (2010) Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends Anal Chem 29(9):1073–1081

- Carabante I, Grahn M, Holmgren A, Hedlund J (2010) In situ ATR-FTIR studies on the competitive adsorption of arsenate and phosphate on ferrihydrite. J Colloid Interface Sci 351(2):523–531
- Chabot M, Hoang T, Al-Abadleh HA (2009) ATR-FTIR studies on the nature of surface complexes and desorption efficiency of p-arsanilic acid on iron (oxyhydr)oxides. Environ Sci Technol 43(9):3142–3147
- Chen M, Ma LQ (1998) Comparison of four USEPA digestion methods for trace metal analysis using certified and Florida soils. J Environ Qual 27(6):1294–1300
- Cornell RM, Schwertmann U (2003) The iron oxides: structure, properties, reactions, occurrences, and uses (trans: Schwertmann U). 2nd, completely rev. and extended edn. Wiley-VCH, Weinheim
- Forouzangohar M, Kookana RS, Forrester ST, Smernik RJ, Chittleborough DJ (2008) Midinfrared spectroscopy and chemometrics to predict diuron sorption coefficients in soils. Environ Sci Technol 42(9):3283–3288
- Geladi P (1988) Notes on the history and nature of partial least squares (PLS) modelling. J Chemometrics 2:231–246
- Goldberg S (2002) Competitive adsorption of arsenate and arsenite on oxides and clay minerals. Soil Sci Soc Am J 66(2):413–421
- Goldberg S, Johnston CT (2001) Mechanisms of arsenic adsorption on amorphous oxides evaluated using macroscopic measurements, vibrational spectroscopy, and surface complexation modeling. J Colloid Interface Sci 234(1):204–216
- Grafe M, Eick MJ, Grossl PR (2001) Adsorption of arsenate (V) and arsenite (III) on goethite in the presence and absence of dissolved organic carbon. Soil Sci Soc Am J 65(6):1680–1687
- Grafe M, Eick MJ, Grossl PR, Saunders AM (2002) Adsorption of arsenate and arsenite on ferrihydrite in the presence and absence of dissolved organic carbon. J Environ Qual 31(4):1115–1123
- Haaland DM, Thomas EV (1988) Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. Anal Chem 60(11):1202–1208
- Huang PM, Fujii R (1996) Selenium and arsenic. In: Sparks DL (ed) Methods of soil analysis. Part 3, chemical methods. Soil Science Society of America Inc, Wisconsin, pp 793–831
- Hubert AE (1983) Determination of arsenic in geological materials by x-ray fluorescence spectrometry after solvent extraction and deposition on a filter. Talanta 30(12):967–968
- Ibrahim M, Hameed AJ, Jalbout A (2008) Molecular spectroscopic study of River Nile sediment in the greater Cairo region. Appl Spectrosc 62(3):306–311
- Isbell RF (2002) The Australian soil classification, revised edn. CSIRO, Collingwood
- Islam K, Singh B, McBratney A (2003) Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Aust J Soil Res 41(6):1101–1114
- Janik LJ, Skjemstad JO (1995) Characterization and analysis of soils using mid infrared partial least squares. 2. Correlations with some laboratory data. Aust J Soil Res 33(4):637–650
- Janik LJ, Skjemstand JO, Raven MD (1995) Characterization and analysis of soils using mid infrared partial least squares. 2. Correlations with some laboratory data. 1. Correlations with XRF-determined major element composition. Aust J Soil Res 33(4):621–636
- Janik LJ, Merry RH, Forrester ST, Lanyon DM, Rawson A (2007) A rapid prediction of soil water retention using Mid Infrared spectroscopy. Soil Sci Soc Am J 71:507–514
- Jia Y, Xu L, Wang X, Demopoulos GP (2007) Infrared spectroscopic and X-ray diffraction characterization of the nature of adsorbed

arsenate on ferrihydrite. Geochim Cosmochim Acta 71(7):1643–1654

- Johnston CT, Aochi CT (1996) Fourier transform infrared and Raman spectroscopy. In: Sparks DL (ed) Methods of soil analysis, part 3: chemical methods. Soil Science Society of America, Madison, pp 269–321
- Kimber SWL, Sizemore DJ, Slavich PEG (2002) Is there evidence of arsenic movement at cattle tick dip sites? Aust J Soil Res 40(7):1103–1114
- Kookana RS, Janik LJ, Forouzangohar M, Forrester ST (2008) Prediction of atrazine sorption coefficients in soils using midinfrared spectroscopy and partial least-squares analysis. J Agric Food Chem 56(9):3208–3213
- Lin HT, Wang MC, Li GC (2004) Complexation of arsenate with humic substance in water extract of compost. Chemosphere 56(11):1105–1112
- Mandal BK, Suzuki KT (2002) Arsenic round the world: a review. Talanta 58(1):201–235
- Mark H (1992) Data analysis: multilinear regression and principal components analysis. In: Burns DA, Ciurcester EW (eds) Handbook of near infrared analysis. Marcel Dekker, San Diego
- Mckenzie N, Jacquier D, Isbell RF, Katharine B (2004) Australian soils and landscapes: an illustrated compendium. CSIRO, Collingwood
- McLaren RG, Naidu R, Smith J, Tiller KG (1998) Fractionation and distribution of arsenic in soils contaminated by cattle dip. J Environ Qual 27(2):348–354
- Minasny B, McBratney AB, Tranter G, Murphy BW (2008) Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. Eur J Soil Sci 59(5):960–971
- 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(24):9314–9320
- Niazi NK, Bishop TFA, Singh B (2011a) Evaluation of spatial variability of soil arsenic adjacent to a disused cattle-dip site, using model-based geostatistics. Environ Sci Technol 45(24):10463–10470
- Niazi NK, Singh B, Shah P (2011b) Arsenic speciation and phytoavailability in contaminated soils using a sequential extraction procedure and XANES spectroscopy. Environ Sci Technol 45(17):7135–7142
- Niazi NK, Singh B, Van Zwieten L, Kachenko AG (2011c) Phytoremediation potential of *Pityrogramma calomelanos* var. *austroamericana* and *Pteris vittata* L. grown at a highly variable arsenic contaminated site. Int J Phytorem 13(9):912–932
- Niazi NK, Singh B, Zwieten LV, Kachenko AG (2012) Phytoremediation of an arsenic-contaminated site using *Pteris vittata* L. and *Pityrogramma calomelanos* var. *austroamericana*: a longterm study. Environ Sci Pollut Res 19:3506–3515
- O'Reilly SE, Strawn DG, Sparks DL (2001) Residence time effects on arsenate adsorption/desorption mechanisms on goethite. Soil Sci Soc Am J 65(1):67–77
- 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
- Sanghavi BJ, Srivastava AK (2010) Simultaneous voltammetric determination of acetaminophen, aspirin and caffeine using an in situ surfactant-modified multiwalled carbon nanotube paste electrode. Electrochim Acta 55(28):8638–8648
- Sanghavi BJ, Mobin SM, Mathur P, Lahiri GK, Srivastava AK (2013a) Biomimetic sensor for certain catecholamines



employing copper(II) complex and silver nanoparticle modified glassy carbon paste electrode. Biosensors Bioelectron 39(1):124–132

- Sanghavi BJ, Sitaula S, Griep MH, Karna SP, Ali MF, Swami NS (2013b) Real-time electrochemical monitoring of adenosine triphosphate in the picomolar to micromolar range using graphene-modified electrodes. Anal Chem 85(17):8158–8165
- Schroeder PA (2002) Teaching clay science. In: Rule A, Guggenheim S (eds) CMS workshop lectures, vol 11. The Clay Mineral Society, Aurora, pp 181–206
- Smith E, Naidu R, Alston AM (1998) Arsenic in the soil environment: a review. Adv Agron 64:149–195
- Smith AH, Arroyo AP, Mazumder DNG, Kosnett MJ, Hernandez AL, Beeris M, Smith MM, Moore LE (2000) Arsenic-induced skin

lesions among Atacameno people in Northern Chile despite good nutrition and centuries of exposure. Environ Health Perspect 108(7):617–620

- Song Y, Li F, Yang Z, Ayoko GA, Frost RL, Ji J (2012) Diffuse reflectance spectroscopy for monitoring potentially toxic elements in the agricultural soils of Changjiang River Delta, China. Appl Clay Sci 64:75–83
- Wenzel WW, Kirchbaumer N, Prohaska T, Stingeder G, Lombi E, Adriano DC (2001) Arsenic fractionation in soils using an improved sequential extraction procedure. Anal Chim Acta 436(2):309–323

