

6

Mid infrared reflectance spectroscopy and partial least-squares regression to predict total As content in soil

ABSTRACT

The potential of mid infrared (MIR) spectroscopy in combination with partial least-squares (PLS) regression was investigated to estimate the total As content in soil samples collected from a highly variable As-contaminated cattle-dip site. The principal component analysis (PCA) was performed prior to MIR-PLS regression analysis that identified spectral outliers in the absorbance spectra of soil samples. The MIR-PLS (final) calibration model excluding ($n = 149$) the spectral outliers showed an acceptable reliability (coefficient of determination, $R^2_c = 0.73$; residual prediction deviation, $RPD_c = 1.94$) to estimate total soil As. The validation of final calibration model using a validation set ($n = 149$) of unknown samples also resulted in a good acceptability with $R^2_v = 0.63$ and $RPD_v = 1.66$. The MIR-PLS based model can provide a rapid estimate of total soil As content by taking into account the integrated effects of mineralogical and organic components of soils. This approach can be useful to estimate total soil As in some conditions where analysis of a large number of soil samples is required.

6.1 INTRODUCTION

Arsenic (As) is an extremely toxic and carcinogenic element, 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 New South Wales (NSW), Australia has resulted in the contamination of surrounding soils with high (up to 14,000 mg kg⁻¹) As concentrations. Arsenic concentration in soil in the vicinity of dip sites varies on small scale, and it requires analysis of a huge number of soil samples to estimate the level of As contamination at a particular site (Kimber et al. 2002; Niazi et al. 2011). Estimation of the total As content in soils is imperative to evaluate the extent of contamination and assess the risk posed to the environment (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, these use mixtures of concentrated acids (HNO₃, HClO₄, HCl, H₂SO₄, HF) (Huang and Fujii 1996; Chen and Ma 1998). Total As concentration in the digests is generally measured by hydride-generation atomic absorption spectrometer or inductively couple plasma atomic emission spectrometer. Although chemical techniques to determine the total soil As content in soil are highly accurate, however, these 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 the equipment is not widely available. There is a need for a rapid, inexpensive and easy-to-use method to estimate total As content in soil in certain situations. For example, for a routine estimation of total As in large number of soil samples at a given site, to monitor changes in soil As content at a given site or to estimate variability in total As in soil at a particular site.

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 the soil samples contain information about the amount and composition of both organic and mineral components in soils. The qualitative information in the MIR spectra of soils is characterised 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). Also the presence of various vibrational features (e.g. Fe–O, Al–O, Si–O and –OH) of the soil mineral components can contribute in the MIR spectra of the soils (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.

Partial least-squares regression in combination with the spectroscopic technique was first introduced by Haaland and Thomas (1988) to construct predictive models. The method is based on the assumption of a linear relationship between the dependent variable of interest (e.g. As concentration in soil) and the predictor variable (e.g. absorbance peaks in the spectra). The PLS regression analysis reduces the number of independent variables (i.e. spectra of samples) into a limited number of predictor variables, called eigenspectra (i.e. PLS loadings or factors) (Haaland and Thomas 1988). The PCA can provide 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 the soils 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. These components (principal components) represent most of the variability in the original data with a reduced amount of random measurements noise (Mark 1992).

Infrared spectroscopy has been used to delineate 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 Fourier transform infrared (FTIR) spectroscopy to determine the adsorption mechanism of arsenate (As^{V}) on the surface of amorphous Fe and Al oxides. They reported that arsenate sorption bands were present between 817 and 861 cm^{-1} on Fe oxides; and between 856 and 866 cm^{-1} on the surface of Al oxides. Jia et al. (2007) investigated the effect of pH on arsenate adsorption on ferrihydrite using FTIR spectroscopy; they depicted that the spectral region from 808–838 cm^{-1} was attributed to the As^{V} –Fe oxides adsorption complexes depending on the pH (3–8) and coverage density of As.

Previous spectroscopic studies, however, have not estimated As contamination in soils. The development of the MIR-PLS based approach to predict total soil As content might be able to provide a cheaper and time-effective method compared to conventional chemical procedures. Method could be particularly useful for situations where a large number of samples needs to be analysed or repetitive analyses are needed for the samples.

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 As content in soil based on the hypothesis that MIR spectroscopy has the capability to capture the spectral features of

adsorbed and mineral forms of As in soils, and (2) apply the predictive model to unknown soil samples to estimate As concentration in soils and to evaluate the prediction quality of the model.

6.2 MATERIALS AND METHODOLOGY

6.2.1 Soil samples

Arsenic-contaminated soil samples were taken from the disused cattle-dip site located at Wollongbar as described earlier (Chapter 3). The soil samples used in this study were taken from three depths, i.e. 0–20, 20–40 and 40–60 cm as described in Chapters 3 and 4. All soil samples ($n = 304$) were dried at 40 °C and ground to obtain < 200 μm fraction prior to acid digestion.

6.2.2 Soil digestion to measure total As

The samples (~0.25 g) were digested in a mixture of concentrated hydrofluoric (HF), sulphuric (H_2SO_4), perchloric (HClO_4), nitric (HNO_3) and 1.2 M hydrochloric (HCl) acids (Huang and Fujii 1996) (Section 4.2.4, Chapter 4). Briefly, the samples were soaked in HNO_3 and HCl overnight. The soaked samples were then mixed with H_2SO_4 , HF and HClO_4 and heated at 120 °C overnight until the mixture was dried. After cooling for 5 minutes, the digest was dissolved in 25 mL of 6 M HCl and final volume for made to 50 mL by adding E-pure[®] water. The digests were analysed for As using an inductively coupled plasma atomic absorption spectrometer (ICP-AES, Varian[®] Vista AX CCD) as described earlier in Chapter 4.

6.2.3 MIR spectroscopy

Soil samples were finely ground ($< 200\ \mu\text{m}$) and dried at $105\ ^\circ\text{C}$ overnight before the spectroscopic analysis. Diffuse reflectance (MIR) spectral measurement of the soil samples was performed 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 sharp-edge stainless steel blade. Scans were collected in the wavenumber (frequency) range of $4000\text{--}400\ \text{cm}^{-1}$ ($2500\text{--}20000\ \text{nm}$) at $8\ \text{cm}^{-1}$ resolution. The potassium bromide (KBr) powder was used as a background reference and assumed to have a reflectance of 1 (100%). The data in whole (wavenumber) range ($4000\text{--}400\ \text{cm}^{-1}$) were used for the principal component analysis (PCA) and partial least square (PLS) analysis. The reflectance (R) values were log-transformed ($\log(1/R)$) 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 Figure 6.1) 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.

6.2.4 Principal component analysis and partial least-squares

Principal component analysis was carried out to detect the presence of any spectral outliers in the spectral data, prior to develop a prediction model using PLS regression (see Figure 6.2). 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; Forouzanoghar et al. 2008; Kookana et al. 2008).

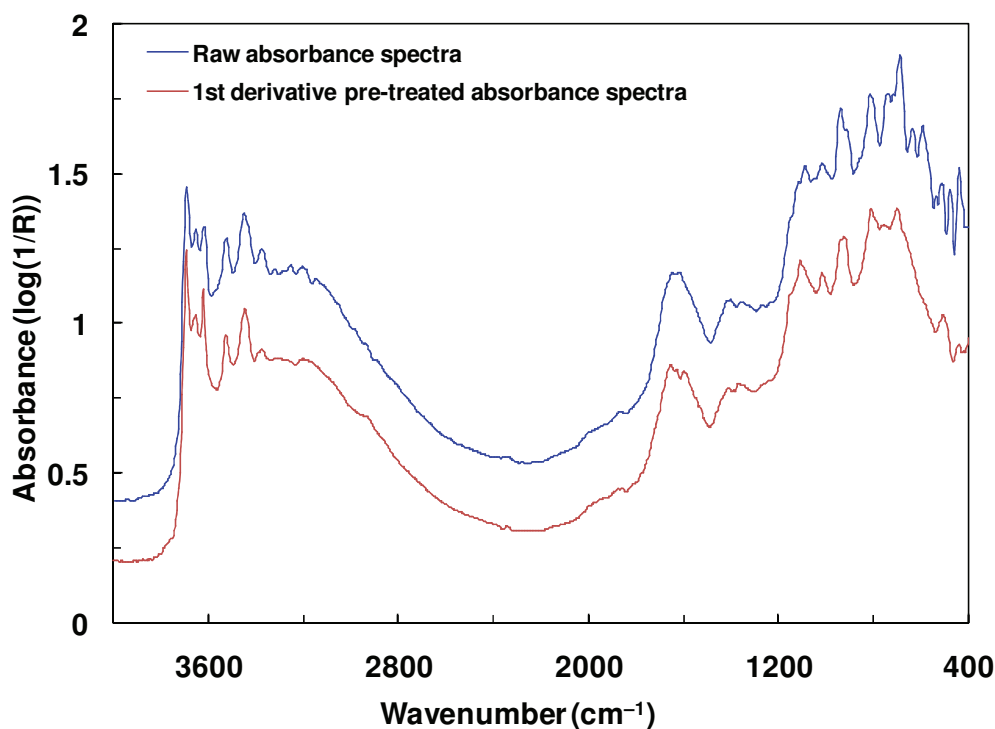


Figure 6.1 The absorbance spectra of a soil sample in the MIR spectral region (4000–400 cm^{-1}). The blue colour line represents the raw absorbance spectrum; red colour spectrum is first derivative of the raw spectrum after removing the baseline.

It reduces the dimensionality of the data and forms $n \times p$ covariance matrix, where n is the number of spectra of the soil samples analysed and p represents the absorbance values (Pirie et al. 2005). In PCA, the eigenspectra and their respective eigenvalues are calculated. The number of components in the data are then reduced to a smaller group of principal components, termed as PCA loadings (Haaland and Thomas 1988; Forouzangohar et al. 2008).

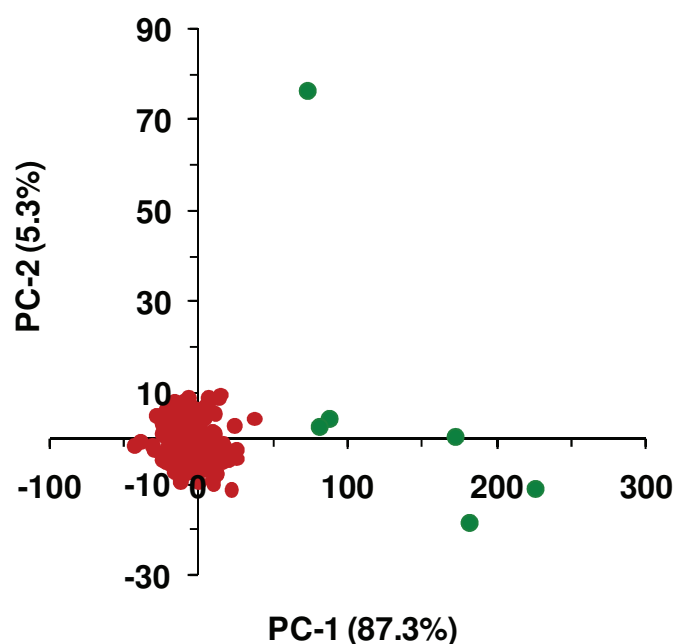
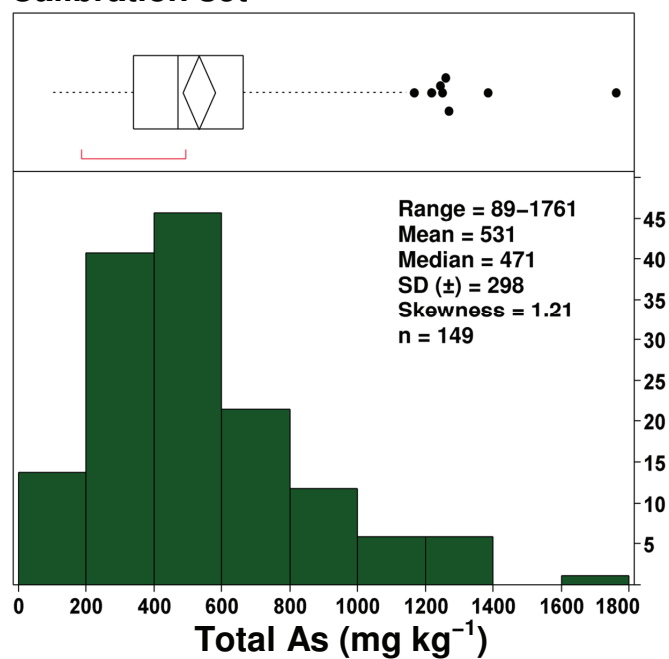


Figure 6.2 Principal component analysis (PCA) performed on the (first derivative pre-treated) absorbance spectra of all soil samples ($n = 304$). The green filled circles are the 6 spectral outliers in the soil samples. PC-1 = principal component-1; PC-2 = principal component-2.

The loadings in PLS analysis represent the components (factors) in the spectra which are highly correlated with measured data. In a PLS regression model, model factors and the associated scores are used to model the covariance matrix between measured data and spectra (Haaland and Thomas 1988; Janik and Skjemstad 1995).

The data set contained 304 soil samples including the 6 spectral outliers identified using PCA (Figure 6.2). After excluding the spectral outliers, the remaining 298 samples were randomly split into two sets, each containing 149 number of samples. One (calibration set) set of samples ($n = 149$) 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 validation set (Figure 6.3).

(a) Calibration set



(b) Validation set

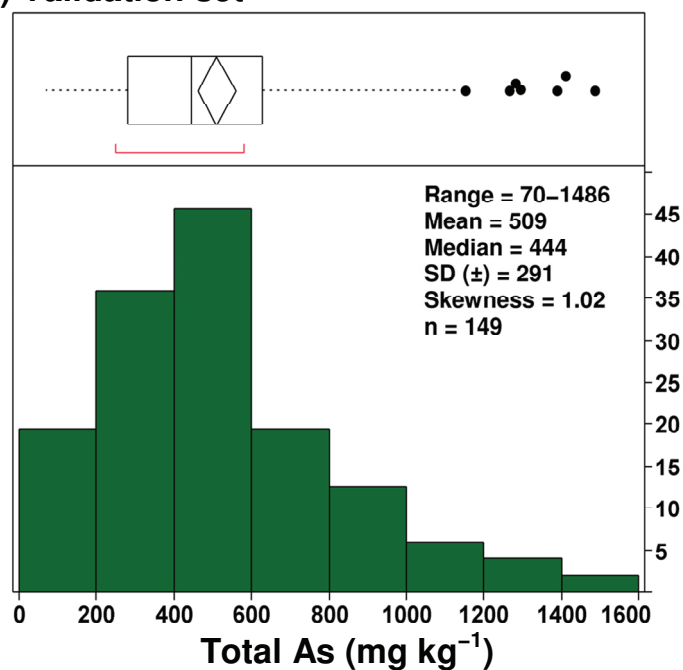


Figure 6.3 Frequency distribution (histograms and box-plots) of the measured total As concentration in soil samples used for the MIR-PLS regression analysis for the development of (a) final calibration model; and for (b) validation of the calibration model.

The calibration model was developed using leave-one-out cross validation (LOOCV) method in PLS regression analysis. 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. The PLS model training process continues until the minimum prediction residual sum of squares (PRESS) is attained, which is used to choose the optimal number of PLS factors in order to diminish the possibility of overfitting the model (Haaland and Thomas 1988).

It is also important to estimate the true errors in the prediction model and validate it using a set of unknown soil samples, 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 the concentration of As in soil in these samples. To evaluate the efficiency of the prediction-model, predicted As concentration values were plotted against the measured soil As concentrations. 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 (SECV) in calibration, the standard error of prediction (SEP) in validation, and the residual prediction deviation (RPD) in calibration (RPD_c) and validation (RPD_v) models (Pirie et al. 2005; Islam et al. 2003). The SECV and SEP are calculated using eq 1:

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

where z_i is the measured data and \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, RPD_c is the ratio of standard deviation (SD) of the measured data in the calibration set to

SECV (eq 2). While SEP is used to estimate RPD_v in the validation model (eq 3), and SD is linked to the measured data in the validation set.

$$RPD_c = \frac{SD_c}{SECV} \quad (2)$$

$$RPD_v = \frac{SD_v}{SEP} \quad (3)$$

The RPD value accounts for the natural variation in the measured data to the size of prediction errors obtained in the model and is useful to interpret the prediction efficiency of the model (Janik and Skjemstad 1995; Islam et al 2003; Pirie et al. 2005)

6.3 RESULTS AND DISCUSSION

Total As concentrations were highly variable in the studied soil samples in both calibration (89–1761 mg kg⁻¹) and validation (70–1486 mg kg⁻¹) sets (Figure 6.3a,b). The distribution of As in both (calibration and validation) data sets seems to be similar with nearly identical mean As concentration values (Figure 6.3). A slightly positive skewness was observed in the calibration (1.21) and validation (1.02) set samples. The variability in the measured soil As content in the soil samples is inherent and considered to be associated with the cattle dipping process, pumping-out of the dip fluid and disposal of the As-containing dip sediment in the vicinity of the cattle-dip (Chapter 3) (Kimber et al. 2002; Niazi et al. 2011).

6.3.1 MIR spectral features

MIR spectra of the 6 (surface and subsurface) soil samples are shown in Figure 6.4a. The spectra show several stretching and bending bands which were attributed to the mineral and organic components in soil. For example, the presence of hydroxyl (–OH) stretching and

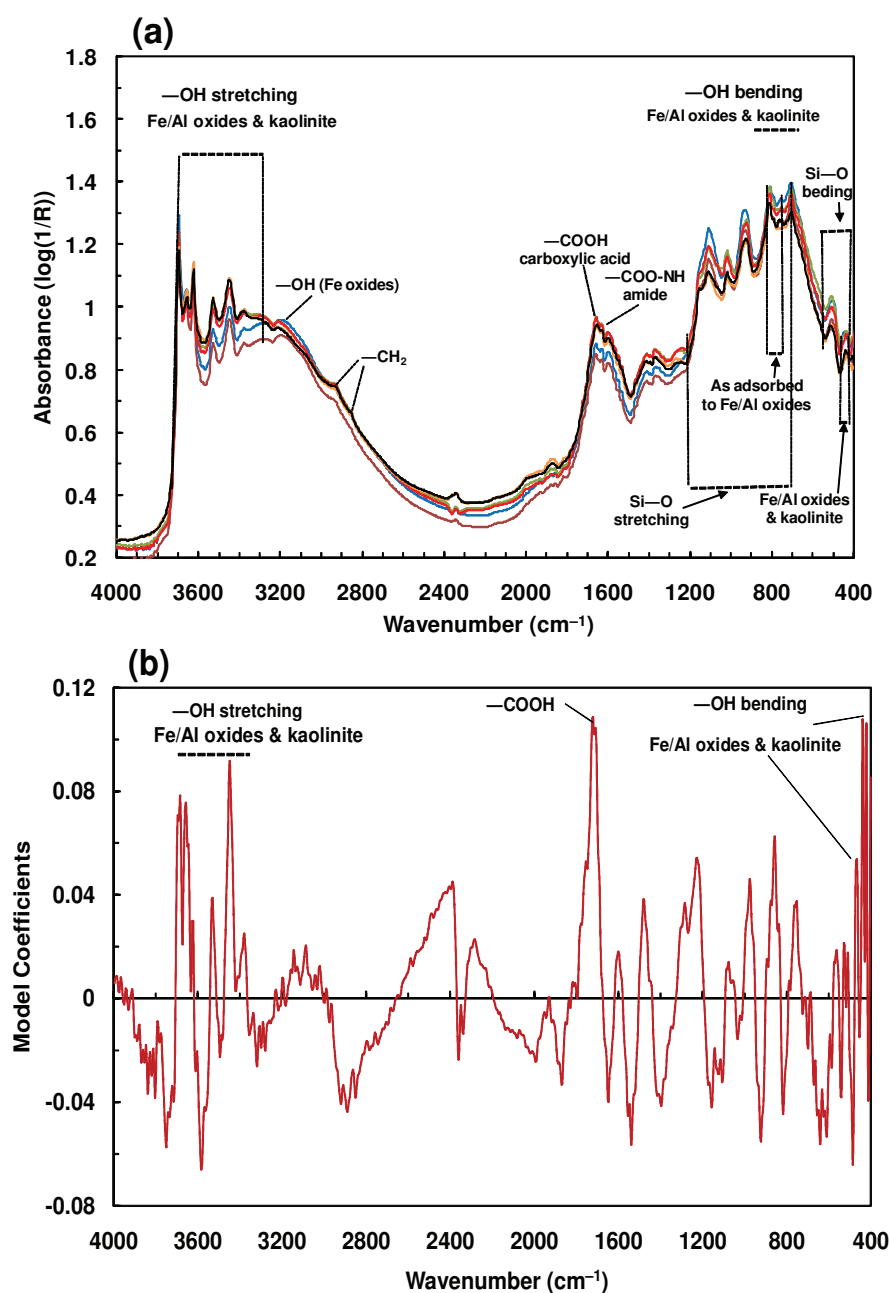


Figure 6.4 (a) The mid infrared (MIR) absorbance spectra of the six soil samples collected from the surface (0–20 cm) and the subsurface (20–40, 40–60 cm) depths; and (b) the MIR partial least-squares (PLS) coefficients obtained in the calibration model ($n = 149$), in relation to the wavenumber (4000–400 cm^{-1}) used for the MIR spectroscopy.

bending bands in the spectral region from 3750–3300 cm^{-1} and 950–820 cm^{-1} was possibly due to the presence of structural –OH groups in clay minerals and metal (Fe/Al) oxides (Johnston and Aochi 1996). Phyllosilicates can be distinguished into the –OH stretching (3750–3400 cm^{-1}), –OH bending (950–600 cm^{-1}), Si–O stretching (1200–700 cm^{-1}) and Si–O bending (600–400 cm^{-1}) vibrations (see Figure 6.4a) (Haberhauer et al. 1998). The presence of a band at 3145 cm^{-1} in the spectra of samples tested in this study could possibly be attributed to –OH stretching in Fe oxides, as reported by Haberhauer et al. (1998) who also attributed the presence of –OH stretching band at 3140 cm^{-1} to Fe oxides.

The soil at the site is a Red Ferrosol (Chapters 3 and 4) (Isbell 2002). These soils have high amounts of Fe/Al oxides and mainly contain kaolinite. Thus the contribution of other phyllosilicates (e.g. smectite, illite) would be negligible or none in the absorbance spectra of the studied soil samples. The spectral peaks observed in Figure 6.4a in the region between 3695 and 3622 cm^{-1} could be attributed to the presence of –OH stretching band for kaolinite (Janik et al. 2007). The bands observed between 428 and 470 cm^{-1} in the spectra of soil samples were attributed to the Fe–O/Al–O bending vibrations (see Figure 6.4a) (Cornell and Schwertmann 2003; Ibrahim et al. 2008).

The spectral bands in the region from 1510–1230 cm^{-1} and at 1730 cm^{-1} have been reported to be associated with the presence of aromatic and carboxylic (–COOH) functional groups (Janik and Skjemstad 1995; Haberhauer et al. 1998; Pirie et al. 2005). The spectral band identified in the analysed soil samples at 1720 cm^{-1} in Figure 6.4a could be associated with the presence of –COOH group in soil organic fraction. The bands in the spectral region from 2928–2852 cm^{-1} could be attributed to the aliphatic (–CH₂) stretching vibrations present in organic fraction in the studied soil samples (see Figure 6.4a). This is in agreement with those

reported by Janik and Skjemstad (1995) and Haberhauer et al. (1998) for the presence of $-\text{CH}_2$ functional group in spectra of soils (i.e. bands at $2930\text{--}2851\text{ cm}^{-1}$).

Figure 6.4b shows the strength of relationship between the model coefficients and the spectral regions in the absorbance spectra of soil samples. This represents the significance of soil components which were most strongly and positively correlated with the As concentration in soil and explain the variation in prediction model. The positive peaks in the region between 3695 and 3622 cm^{-1} (see Figure 6.4b) showed that kaolinite strongly contributed in the MIR-PLS model to estimate As in soil samples. The presence of very strong and positive sharp peaks in Figure 6.4b in the spectral region from $428\text{--}470\text{ cm}^{-1}$ (as mentioned above) revealed the strong contribution of Fe/Al oxides in the PLS regression loadings. The contribution of Fe/Al oxides and kaolinite spectral bands in the PLS loadings (model coefficients) in relation to predicting As content in the soil is attributed to the high capacity of these minerals, particularly Fe oxides for adsorption of As in the soil (Goldberg and Johnston 2001; Goldberg 2002; Jia et al. 2007; Chabot et al. 2009).

In a study, Goldberg and Johnston (2001) used FTIR spectroscopy to investigate sorption complexes of arsenate on the surface of synthetic Fe and Al oxides. Arsenate adsorption was identified due to the presence of spectral bands at 789 cm^{-1} and between 817 and 861 cm^{-1} due to the adsorption of arsenate on the surface of Fe oxides; and the bands appeared at $856\text{--}866\text{ cm}^{-1}$ on the surface of Al oxides. In another study, Jia et al. (2007) reported that the spectral bands appeared in the region from $808\text{--}838\text{ cm}^{-1}$ due to the adsorption of arsenate on ferrihydrite. In the present study, the appearance of (small) bands in the region between 795

and 829 cm^{-1} could be associated with the existence of Fe–O–As or Al–O–As groups due to the adsorption of As (as arsenate) on the surface of Fe or Al oxides in soil (Figure 6.4a).

The carboxylic acids ($-\text{COOH}$) are the low molecular weight organic acids present in the soil organic matter fraction and have the ability to make complexes with the metal ions, such as Fe (Grafe et al. 2001; Grafe et al. 2002; Mandal and Suzuki 2002). Arsenate can make complexes either directly or by binding through Fe with these organic groups. The positive contribution of $-\text{COOH}$ group of organic fraction was revealed by the sharp peak at 1720 cm^{-1} as shown in Figure 6.4b.

6.3.2 Principal component analysis

Previous studies indicate that PCA is a useful tool for the identification of spectral outliers in the absorbance spectra of the samples and can be employed to increase the quality of the prediction-model (Pirie et al. 2005). A PCA was performed on the (first derivative pre-treated) absorbance spectra of all soil samples ($n = 304$) which enabled to detect 6 spectral outliers in the data (see Figure 6.2, 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 a final calibration model (as described next).

6.3.3 Calibration model using MIR spectra and PLS regression

A preliminary calibration model was generated using MIR absorbance spectra of soil samples ($n = 155$) including the 6 samples which were identified as spectral outliers using PCA (Figure 6.5a), as mentioned earlier. This was done to determine the effect of outliers on the prediction model (see Figure 6.5a vs 6.5b; on next page).

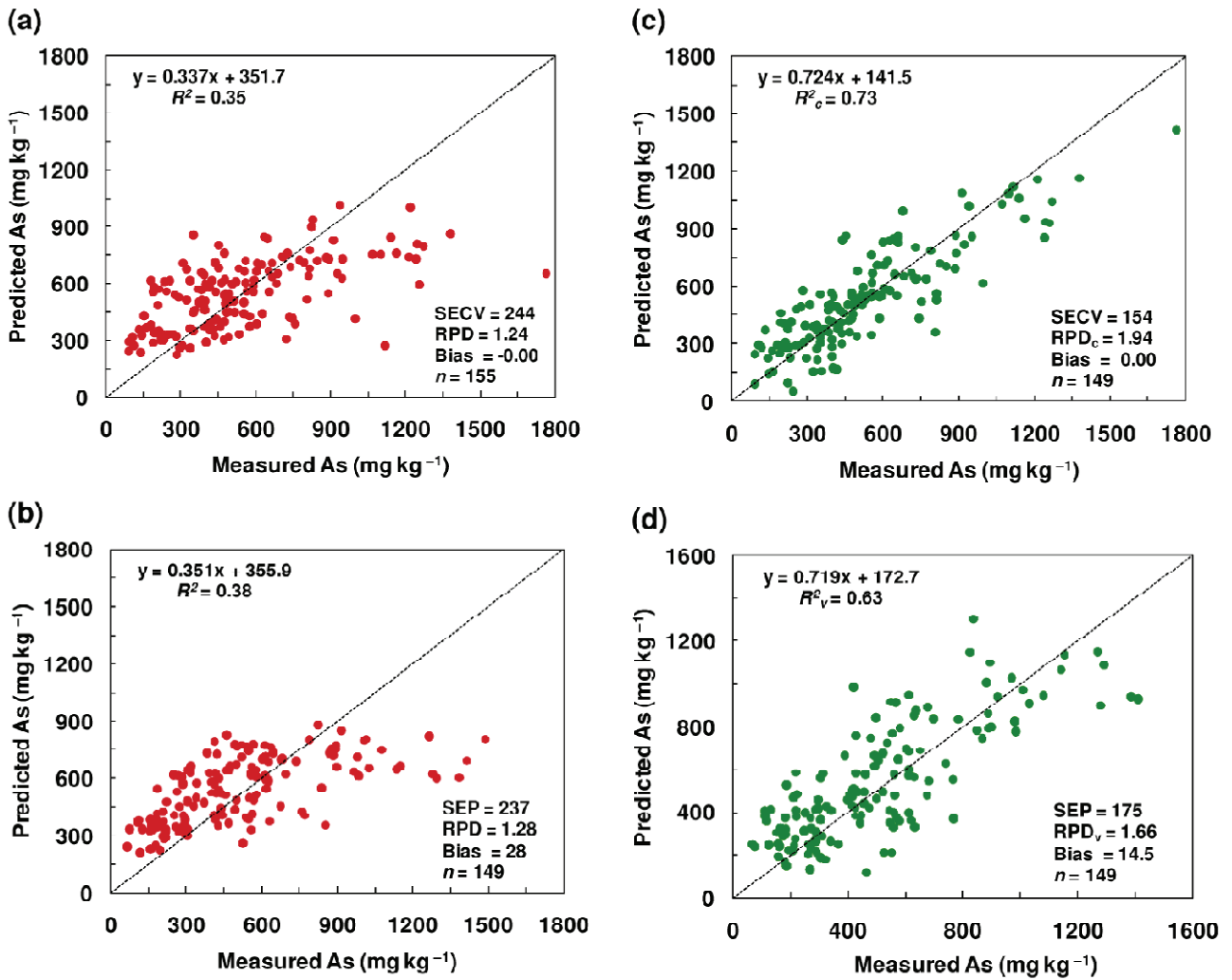


Figure 6.5 The measured values of total soil As concentration are plotted against the MIR-PLS model predicted As values; (a) preliminary calibration model including the 6 spectral outliers ($n = 155$); (b) validation of the preliminary calibration model; (c) final calibration model developed excluding the 6 spectral outliers spectral data ($n = 149$) using the calibration set data as shown in Figure 6.3a; (d) validation of the calibration model constructed with $n = 149$ samples. The R^2_c and R^2_v are coefficient of determination in the calibration and validation sets, respectively; SECv = standard error of cross validation; SEP = standard error of prediction; RPD_c and RPD_v = residual prediction deviation in the calibration and validation sets, respectively.

The preliminary calibration model ($n = 155$) explained 35% of variation ($R^2 = 0.35$) between the measured and model-predicted As concentration values (Figure 6.5a). The calculated RPD

and SECV values for this preliminary calibration model were 1.24 and 244, respectively. The MIR-PLS final calibration model (Figure 6.5c) was re-constructed using absorbance spectra of the 149 samples in the calibration set as shown in Figure 6.3a. 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 (154) and increased R^2_c and RPD_c values (0.73 and 1.94; Figure 6.5c) (Pirie et al. 2005; Forouzangohar et al. 2008). This shows that detection of the spectral outliers using PCA and their subsequent exclusion increased the efficiency and robustness of the MIR-PLS calibration model. The measured and predicted As values were uniformly distributed around the reference equation (dotted) line as shown in Figure 6.5c.

6.3.4 Validation of MIR-PLS model

The validity of the resulting preliminary and final calibration models built on the calibration set of 155 and 149 samples, respectively, was tested using a validation set ($n = 149$; in Figure 6.3b) of unknown soil samples (see Figure 6.5b,d). The prediction equation obtained from the final calibration model ($n = 149$; Figure 6.5c) was applied to the validation set in order to calculate the statistical parameters related to the estimation of As concentrations in unknown soil samples (Figure 6.5d). The results in Figure 6.5d showed that 63% of variation (R^2_v) between the measured and predicted As concentration values could be described by the MIR-PLS calibration model built excluding the spectral outliers ($n = 149$). The resultant RPD_v value of 1.66 demonstrated an acceptable reliability of the proposed MIR-PLS prediction model (Chang et al. 2001; Islam et al. 2003). The RPD value is considered to be highly useful to assess the prediction quality. Based on the RPD value and the purpose of prediction, the appropriateness of a calibration model can be decided (Islam et al. 2003; Pirie et al. 2005; Kookana et al. 2008). For example, Chang et al. (2001) considered RPD values between 1.4

and 2.0 as acceptable, and > 2.0 as excellent in soil characteristics estimation. Similarly, Dunn et al. (2002) suggested three categories of RPD values: poor < 1.6 ; acceptable 1.6–2.0; and > 2.0 as excellent for soil analysis. Conversely, application of preliminary calibration model (including 6 spectral outliers in Figure 6.5a) on the same validation set explained only 38% variation (R^2) between the measured and predicted values (see Figure 6.5b); the SEP increased to 237 and RPD dropped down to 1.28.

MIR-PLS model using phosphate-extractable As concentration in soil

A calibration model was developed using the phosphate-extractable As concentration data for three depths (see Figures 3(A) and 3(B) in Appendix 3); the soil samples used for the (phosphate-extractable As) MIR-PLS model were similar to those used for total As MIR-PLS model. This calibration model ($n = 149$) explained 13% of variation ($R_c^2 = 0.13$) between the measured and model-predicted phosphate-extractable As concentration values (Figure 3(B)a in Appendix 3). The calculated RPD_c and SECV values for the calibration model were 1.07 and 23.02, respectively. The results showed an *unacceptable* reliability of the proposed MIR-PLS prediction model using phosphate-extractable As concentration (Chang et al. 2001; Islam et al. 2003), i.e. indicated by the RPD value which is vital to assess the prediction quality, as described earlier in Section 6.3.2. The phosphate-extractable As is thought to represent the sorbed arsenate in soil on the surfaces of Fe/Al oxides. The spectral bands of arsenate adsorption complexes with Fe/Al oxides in the MIR spectra could possibly exist in the region between 750 and 866 cm^{-1} (see Figure 6.3a) (Gold berg and Johnston 2001; Jia et al. 2007). Hence, only this spectral region could account for the MIR-PLS prediction model developed using phosphate-extractable. However, MIR prediction model presented earlier, based on total As content could possibly take into account a range of (MIR) spectral regions representing As-binding with various mineral and/or organic components in soil (described earlier in

Section 6.3.1). These differences could explain the poor prediction ability of MIR-PLS model formulated using phosphate-extractable As concentration.

6.4 CONCLUSIONS

The data 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. The significant improvement in MIR-PLS calibration model in terms of prediction statistics was attained by removing 6 samples identified as spectral outliers, which was only possible including the MIR-PCA combination as a part of MIR-PLS model approach. The MIR-PLS prediction model has the capability to give direct estimation of total As content in soil by taking into account the integrated effects of mineralogical and organic components of soils. The soil components, particularly Fe/Al oxides and kaolinite appeared to have pertinent role in the sorption of As in the soil samples used in this study.

The MIR-PLS based predictive model proposed in this study has the ability to estimate the total soil As content in various conditions; such as, estimation of total As in large number of soil samples required (i) for monitoring effect of phytoremediation on soil As content at a particular site (Chapter 5) or (ii) to estimate total As in soil at highly variable sites, such as cattle-dip sites (Chapters 3,4). Arsenic was mainly present in adsorbed form at the studied site. Hence, a robust prediction model is needed for estimating total soil As at different sites where As may also be present in different mineral forms, e.g. scorodite, arsenopyrite. Also, this approach may not be useful to determine the valence forms of As ($\text{As}^{\text{V}}/\text{As}^{\text{III}}$) in soil.

6.5 REFERENCES

- 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.
- Chang CW, Laird DA, Mausbach MJ, Hurburgh CRJ (2001) Near-infrared reflectance spectroscopy-principal components regression analysis of soil properties. *Soil Sci Soc Am J* 65:480-490.
- 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. Completely rev. and extended edn. Wiley-VCH, Weinheim.
- Dunn BW, Beecher HG, Batten GD, Ciavella S (2002) The potential of near-infrared reflectance for soil analysis-a case study from the Riverine Plain of south-eastern Australia. *Aust J Soil Res* 42:607-614.
- 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.
- 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.
- Haberhauer G, Rafferty B, Strebl F, Gerzback MH (1998) Comparison of the forest soil litter derived from three different sites at various decompositional stages using FTIR spectroscopy. *Geoderma* 83:331-342.
- 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, Victoria.

- 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, Skjemstad 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, WI, USA, 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, Forouzanoghar M, Forrester ST (2008) Prediction of atrazine sorption coefficients in soils using mid-infrared spectroscopy and partial least-squares analysis. *J Agric Food Chem* 56 (9):3208-3213.
- 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.
- 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.
- Niazi NK, Singh B, Van Zwieten L, Kachenko AG (2011) 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.
- 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.
- Smith E, Naidu R, Alston AM (1998) Arsenic in the soil environment: A review. Adv Agron 64:149-195.