Chemometric Methods to Predict of Pb in Urban Soil from Port Pirie, South Australia, using Spectrally Active of Soil Carbon

Ali Al Maliki, Gary Owens, Hussain M Hussain, Saadi Al-Dahaan & Nadhir Al-Ansari

To cite this article: Ali Al Maliki, Gary Owens, Hussain M Hussain, Saadi Al-Dahaan & Nadhir Al-Ansari (2018): Chemometric Methods to Predict of Pb in Urban Soil from Port Pirie, South Australia, using Spectrally Active of Soil Carbon, Communications in Soil Science and Plant Analysis, DOI: 10.1080/00103624.2018.1464178

To link to this article: https://doi.org/10.1080/00103624.2018.1464178

Published online: 07 May 2018.
Chemometric Methods to Predict of Pb in Urban Soil from Port Pirie, South Australia, using Spectrally Active of Soil Carbon

Ali Al Maliki\textsuperscript{a}, Gary Owens\textsuperscript{b}, Hussain M Hussain\textsuperscript{c}, Saadi Al-Dahaan\textsuperscript{d}, and Nadhir Al-Ansari\textsuperscript{e}

\textsuperscript{a}Environment Research Centre, Ministry of Science and Technology, Baghdad, Iraq; \textsuperscript{b}University of South Australia, Adelaide, Australia; \textsuperscript{c}Remote Sensing Centre, University of Kufa, Al-Najf, Iraq; \textsuperscript{d}Environment and Geomorphology Department, University of Kufa, Al-Najf, Iraq; \textsuperscript{e}Luleå Tekniska Universitet, Luleå, Sweden

\textbf{Abstract}

A total of 73 soil samples were initially analyzed for lead (Pb) concentration as an indicator of the environment impact of smelter activity in the Port Pirie, South Australia. Chemometric techniques were used to assess the ability of near-infrared (NIR) reflectance spectroscopy to predict soil Pb using spectrally active soil characteristics such as soil carbon (C). The result indicated a strong linear relationship between log-transformed data of soil Pb and spectral reflectance in the range between 500 and 612 nm with $R^2 = 0.54$ and a low root-mean-square error (RMSE$_v = 0.38$) for the validation mode with an acceptable ratio of performance to deviation and ratio of error range (1.6 and 7.7, respectively). This study suggested that NIR spectroscopy based on auxiliary spectrally active components is a rapid and noninvasive assessment technique and has the ability to determine Pb contamination in urban soil to be useful in environmental health risk assessment.

\textbf{Introduction}

While soil may naturally contain varying amounts of lead (Pb) depending on its geological source, soil and dust are believed to be the two main anthropogenic sources and pathways of Pb exposure due primarily to depositions from leaded industrial and manufacturing processes or from smelter emissions (McMichael et al. 1994; Tong et al. 2000; Martley, Gulson, and Pfeifer 2004). Evaluating the environmental impact of soil Pb contamination in urban areas is a complex task that requires a good understanding of mechanisms operating at the interfaces between environmental health, urban geochemistry, historical geography, and spatial distributions (Al Maliki, Bruce, and Owens 2015). This would facilitate reducing the uncertainties of prediction and minimizes investigation (Al Maliki, Bruce, and Owens 2012a).

There have also been some attempts to detect Pb concentration from reflectance spectra directly; Choe et al. (2008) found a significant correlation between the ratios of reflectance at 610–500 nm and Pb concentration. Zhang, Wen, and Zhao (2010) found a strong spectral signature for soil Pb content using bands selected from hyperspectral imaging laboratory measurements, where the soil spectral curve absorption values or reflectance peaks occurred at 838, 1930, and 2148 nm. These bands associated with complementary compounds that relate to Pb. A significant linear relationship between Pb and reflectance was observed, with coefficients of determination ($R^2$) of 0.691 and 0.683 at 800 and 1300 nm, respectively (Pandit, Filippelli, and Li 2010).

\textbf{CONTACT} Nadhir Al-Ansari alyay004@mymail.unisa.edu.au; nadhir.alansari@ltu.se Luleå Tekniska Universitet, Luleå, Sweden.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lcss.

Published with license by Taylor & Francis Group, LLC © 2018 [Ali Al Maliki, Gary Owens, Hussain M Hussain, Saadi Al-Dahaan, and Nadhir Al-Ansari] This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.
Heavy metals (HMs) do not generally absorb in the visible-near-infrared (VNIR) region; however, they can be detected via regression analysis due to covariation (relation) with moieties such as hydroxides, sulfides, carbonates, oxides, or organic matter which are detectable or adsorption to clays that absorb light in the VNIR-short-wave infrared region (Schwartz 2012).

The absorption feature for soil Pb could be attributed to the absorption features of soil constituents such as iron, organic carbon, and clay mineral (Summers 2009; Stenberg et al. 2010; Ji et al. 2010). Therefore, the correlation between Pb and other soil constituents that have spectrally active signatures can be indicative of the Pb contamination levels in soil (Al Maliki, Bruce, and Owens 2012a). The spectral predictive mechanism using organic carbon as a key for an indirect spectroscopy method to detect Pb concentration has been applied in several previously published studies. For example, Ji et al. (2010) found that Pb bound to organic carbon was useful for the prediction of Pb concentration in the same spectral detection regions as organic carbon. Yaolin et al. (2011) obtained the reliable estimates for Pb concentration, from spectral parameters using a correlation between soil organic matter and Pb concentration, where the prediction of Pb concentration was derived via the ratio between the first-order derivative for wavelengths between 564 and 624 nm. Moreover, there was a significant relationship between some metal concentrations [cadmium (Cd), mercury (Hg), Pb] within the spectral domain 500–700 nm, which was related to organic carbon (Bray, Rossel, and McBratney 2009; Ji et al. 2010; Pandit, Filippelli, and Li 2010). This spectral predictive mechanism was also found in some other studies (e.g., Malley and Williams 1997), but differs from the results of Wu et al. (2007). They found that soil Pb contamination can be estimated indirectly using reflectance spectroscopy (RS) via correlations with iron oxide contents.

It is likely that both the aforementioned direct and indirect methods used for estimating spectral reflectance will need to be employed to remotely detect the presence and concentration of Pb. The general objectives of this study are to develop and demonstrate the utility of VNIR techniques to detect soil constituents that help to estimate Pb contamination in areas of Port Pirie, South Australia. This study is the first attempt to use spectral reflectance signatures to predict the Pb concentration in urban soil and to find suitable spectrum for the detection of Pb in soil. This study thus supports ongoing investigation and monitoring of Pb contamination for studying environmental and health impact of Pb contamination which can substantially reduce the time and costs associated with traditional soil contaminant analysis.

### Methods and materials

#### Study area and sampling

The town of Port Pirie is located on the eastern shore of the Spencer Gulf, 230 km north of Adelaide, in South Australia. The study area at Port Pirie included sample sites of occupied houses, vacant lots, abandoned houses, common areas (playground sites), kindergartens, schools, or other nonresidences such as the land east of the Port Pirie river about 500 m east of smelter (Figure 1).

The study area extended to a buffer zone of 5 km radius from the Nyrstar smelter, which was large enough to encompass all residential areas as well as covering a risk area that contained soil and dust with highest levels of lead contamination. This area was divided into a regular 1 × 1-km grid, and initial soil sample locations were taken from the corners of each grid cell, together with one sample in the center of each grid cell. A map of the combined survey area with the sampling grid and sites is shown in Figure 1. A total of 73 samples were collected from urban areas surrounding the large lead smelter as well as two additional background samples from outside the buffer zone of 5 km from the smelter. Prior to field sampling, proper, potential sampling sites were initially selected from digital aerial ortho-photographs.

Aerial color ortho-photographs of 10 cm spatial resolution with Map Grid of Australia zone 54 for Port Pirie location were obtained from AEROMetrax Pty Ltd. and were acquired on 25 February 2011 (http://aerometrex.com.au) and displayed using ArcMap 10.2. At each sampling location, a total of three discrete bare surface soil samples were collected using a hand auger at a depth of 0–5 cm, as solar radiation in the VNIR has limited penetration capabilities. Samples were collected from within a 1 m² area and were thoroughly mixed onsite to generate a single composite soil sample.
for further analysis. A composite soil sample allows for a more representative soil sample to be obtained (Bickel 2010). The objective of this technique was to combine a number of discrete samples collected from each site into a single homogenized sample which was considered to represent average conditions in the sampled body for the purpose of analysis (EPA 2005). This sampling was chosen to characterize soils that had varying levels of Pb (Body et al. 1991; Taylor 2012), making it an ideal and suitable site to assess urban atmospheric deposition of Pb contaminants (Al Maliki, Bruce, and Owens 2012b). Soil samples (n = 73) were transported to the laboratory and desiccated in a fan assisted oven at 60°C until visually dried. As the soil dried, a porcelain mortar was used to gently disaggregate any lumps that had formed. Prior to any analysis, each sample was then sieved to pass through a 2-mm stainless steel sieve.

Soil physiochemical properties

HM contents of the soils including soil Pb and total Fe were determined using Inductively Coupled Plasma-Mass Spectrometry following microwave-assisted digestion with aqua regia. Total carbon (C) and nitrogen (N) contents were determined via dry oxidation (automated combustion) on a LECO CNS analyzer (LECO Corp., St Joseph, MI). Soil pH and electrical conductivity (EC) were measured in 1:5 soils:water suspension using a pH/conductivity meter (Smart CHEM-LAB, TPS, Australia). Soil texture was determined using a micropipette method (Miller and Miller 1987). The summary of descriptive statistics for above soil attributes is displayed in Table 1.

NIR spectroscopy for soil samples

NIR reflectance spectra were obtained between 350 and 2500 nm using a HyLogger™ system (model VSTR-1) with a high intensity halogen lamp as a light source (Mason and Huntington 2012), available at the Glenside Core Library of the South Australian Department for Manufacturing, Innovation, Trade, Resources and Energy (DMITRE). The HyLogger™ system was developed by CSIRO to extract the mineralogy of drill cores using reflectance spectroscopy (Mason and Huntington 2012). As the system
was designed to determine spectral reflectance caused by molecular vibrations indicative of the chemical bonds in crystalline minerals, RS obtained from a range of wavelengths can be used to identify common minerals present in a soil sample. Semiquantitative mineralogy and associated mineralogical parameters such as alteration intensity ‘crystallinity’ and chemistry can also be extracted or inferred from the HyLogging data (Mason and Huntington 2012). Here the HyLogger system was adapted to obtain reflectance of soils contained in plastic Petri dishes, rather than intact soil cores. Ten Petri dishes were placed side by side in one line for each spectral collection and processing episode. For each soil sample, 20 individual reflectance spectra were extracted from points across the surface of the soil within the Petri dish. Subsequently, the individual spectra were averaged to obtain representative reflectance spectra for transformation and processing. Spectral reflectance was measured at a 4-nm bandwidth from 380 to 2500 nm so that a total of 537 data points per spectrum were measured for each sample location in the soil. The Spectral Geologist (TSG) software (www.thespectralgeologist.com) was used for visualization. TSG collected reflectance spectra for hundreds of samples that included rich information for soil constituents. HyLogger data were subjected to multivariate analysis using the Unscrambler X software package version 10.3 (http://www.camo.com) to extract evidence of the underlying qualities and attributes within the spectral information.

Spectral preprocessing

Many preprocessing techniques have been applied successfully in previous studies to spectral reflectance raw data in order to assist in the interpretation of overlapping weak spectral overtones, to remove spectral noise originating from the effects of nonhomogeneous distributions of particle sizes, and to provide optimum predictions. These techniques included averaging, smoothing, absorption (log (1/R), where R is reflectance), transformations, normalization, and derivative processing. In this study, the spectral reflectance at a wavelength of 537 nm as normalized to characterize the soils variables. Normalization was applied to remove spectral noise and to improve the detection of absorption feature at spectral reflectance (Kooistra et al. 2001; CAMO software 2012). In this work, the mean normalization was applied to improve prediction of the reflectance spectra and compared for efficiency of estimation of soil parameters.

Normalize data

The distribution patterns of soil Pb and other variables were assessed using statistical analysis of a histogram and normal QQ-plots. These were used to assess whether a data set was normally distributed (i.e., for a QQ-plot if it was linear, or asymmetric (i.e., far from normal) if the plot was nonlinear). These statistical tools found that raw Pb concentrations across Port Pirie were not normally distributed. Consequently, Pb was log transformed to normalize the data and reduce the skew of the raw data with the aim to reduce error in prediction models. Thus, all subsequent regression processes for predicted soil Pb across the study area were based on log-transformed data. This technique was used in some studies to predict metal concentrations; for example, Choe et al. (2008) in studying pollution in stream sediments indicated the ability to detect HM spectral absorption features by combining geochemistry log scale with field spectroscopy and hyperspectral remote sensing (Choe et al. 2008). Other studies suggested that data transformation to a log scale is an effective way to reduce the effect of outliers (Dao, Morrison, and Kiely 2013) and is a commonly used transform for soil data (Choe et al. 2009; Hengla, Heuvelink, and Stein 2004). HM measurements in general conform to normal distributions when natural log transformations are applied, while lognormal

<table>
<thead>
<tr>
<th>Id</th>
<th>As (mg kg(^{-1}))</th>
<th>Cd (mg kg(^{-1}))</th>
<th>Cu (mg kg(^{-1}))</th>
<th>Mn (mg kg(^{-1}))</th>
<th>Fe (mg kg(^{-1}))</th>
<th>Zn (mg kg(^{-1}))</th>
<th>Pb (mg kg(^{-1}))</th>
<th>pH</th>
<th>EC µS/cm</th>
<th>C%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>46</td>
<td>7371</td>
<td>32</td>
<td>4</td>
<td>5.62</td>
<td>78.6</td>
<td>0.65</td>
</tr>
<tr>
<td>Max</td>
<td>137</td>
<td>26</td>
<td>2495</td>
<td>30168</td>
<td>261386</td>
<td>9704</td>
<td>5002</td>
<td>9.5</td>
<td>21000</td>
<td>8.38</td>
</tr>
<tr>
<td>Mean</td>
<td>25</td>
<td>5</td>
<td>155</td>
<td>1652</td>
<td>31284</td>
<td>1110</td>
<td>592</td>
<td>8.26</td>
<td>3370</td>
<td>2.55</td>
</tr>
<tr>
<td>St. Dev</td>
<td>27</td>
<td>4</td>
<td>333</td>
<td>4818</td>
<td>37534</td>
<td>2005</td>
<td>790</td>
<td>0.71</td>
<td>5625</td>
<td>1.80</td>
</tr>
</tbody>
</table>
distributions have been used to describe soil Pb with a strongly positive skew (Griffith 2002). The results of these studies indicated that a log scale provides a better opportunity to successfully predict HM concentrations compared with non-transformed data.

In addition to spectral preprocessing, described earlier, outlier removing was applied to develop the regression model.

Outlier values may be defined as when the sample is spectrally different than the rest of the samples, and the residual difference is found to be anomalously much greater than found in the other samples. Principal component (PCs) analysis (PCA) was used with a default $P$-value of 0.05 to identify and remove calibration outliers in the spectral data, and new calibrations were then developed. Primarily, PCA is a feature reduction mathematical procedure that can be useful to provide the basis for the other multivariate linear regression analyses (e.g., partial least-squares regression (PLSR)) (Niazi 2011).

**Multivariate analysis and regression techniques**

**Calibration**
The PLSR calibration technique was applied in this study to handle all the spectroscopy data sets in the field of NIRS to get and display chemically relevant information out of measured chemical data. The spectral information was transformed to PLSR components (factors) similarly to the PCs in a PCA, compressing the data and at the same time dealing with the correlations within the reflectance spectral data. A PLSR technique was implemented to establish the relationship between the reflectance spectra and measured soil variables including soil Pb concentration.

The PLSR technique was based on latent variable decomposition of two variable blocks (matrices) $X$ – spectral parameter (predictors) and $Y$ – measured soil variables (dependants), resulting in the smallest potential number of significant components which allowed for the simplification of the relationship between matrices $X$ and $Y$ (Gannouni, Rebai, and Abdeljaoued 2012; Hong-Yan et al. 2009; Kooistra et al. 2001; Vohland, Bossung, and Fründ 2009; Wu et al. 2005). The purpose of the method is to find a small number of latent factors that are predictive for $Y$, and which also use $X$ efficiently (Bilgili et al. 2010; Hong-Yan et al. 2009).

The specific statistical methods used to process and analyze the data can significantly alter the final prediction capabilities of the models. Thus, the efficiency of the predictive models developed was evaluated by comparing correlations of predicted and actual Pb concentrations.

In correlation line loadings plot for PLSR model, the values between dotted lines within the upper and lower bounds represented the components (factors) in the spectra. These factors and the associated scores are used to model the covariance matrix between spectra data and reference measurement. Calibration models in most NIR applications have some samples that fall outside the properties of the main sample range, and present much greater or lower than the rest of the data set. These are known as calibration outliers values, which are from samples that have poorly determined values during the one-out cross-validation process (Siebielec et al. 2004).

**Validation**

Two validation methods were applied in order to validate the prediction results including

*Internal validation.* A full cross-validation of the leave-one-out method was used to verify the prediction capability of the PLSR models.

*External validation.* An extra seven samples with a variety of Pb concentration values (low, medium, high) and not used in the calibration model were used to evaluate the prediction accuracy of the model through many statistical parameters. The use of an external data set for validation is appropriate to check whether the model predicts new samples well, as using cross-validation only can give overly optimistic estimates of prediction ability.
Evaluation

Several prediction accuracy parameters were applied in this work including, the coefficient of determination ($R^2$), the root-mean-square error (RMSE) for both calibration ($\text{RMSE}_C$) and validation based on the cross validation method (leaving one out) ($\text{RMSE}_{\text{CV}}$), coefficient of regression or the slope of the linear regression, the absolute estimation error or standard error (SE) for both calibration ($\text{SE}_C$) and cross validation ($\text{SE}_{\text{CV}}$), as well as the intercept of regression and the bias, which were all taken as relative predictors of the best model (Nduwamungu et al. 2009). Other statistical parameters for prediction accuracy were the ratio of performance to deviation (RPD) which was computed as the ratio of the standard deviation (SD) of reference values to the $\text{RMSE}_{\text{CV}}$ (Equation 1) and the Ratio of Error Range (RER) computed by dividing the range in the reference value (Max-Min) by the $\text{RMSE}_{\text{CV}}$ used in the validation set (Eq. 2).

\[
\text{RPD} = \frac{\text{SD}}{\text{RMSE}_{\text{CV}}} \quad (1)
\]
\[
\text{RER} = \frac{\text{Range}}{\text{RMSE}_{\text{CV}}} \quad (2)
\]

In general, the lowest RMSE, SE, bias, offset (intercept) together with higher RPD $\geq 2$, RER $> 6$, and $R^2$ coefficient of regression close to 1.0 were used as indicators of the most accurate regressions (Bilgili et al. 2010; Brunet et al. 2007; Nduwamungu et al. 2009; Wu et al. 2007).

Results and discussion

Spectra analysis of soil data

The reflectance spectra measured with the HyLogger system from air-dried soil samples exhibited three specific absorption features (Figure 2), two spectral absorption features (1400 and 1900 nm) are common features for water and the other strong spectral feature (around 2200 nm) is related to clay or hydroxyl mineral. Correlation coefficients between soil measurements and spectral data were used to evaluate the spectral response for soil properties. The relationship between the normalized spectral data and measured values for some soil properties in the VNIR region are shown in Table 2. Correlation coefficients results of the above soil properties were in agreement with the often superior predictions of these properties in the literature. For example, the highest correlation with reflectance spectra in about 400 nm of wavelength region indicated the spectral features for soil total C, similar absorption feature due to total organic carbon was found in the VNIR region by Pastor et al. 2008; (Wetterlind 2009). The correlation between total iron and spectral measurements with $r = 0.38$ in region of spectra of 512 nm can be referred to absorption band for iron content, similar information was found by Ben-Dor, Irons, and Epema (1999) who pointed, that the absorption feature of iron...

![Figure 2](image-url)
and transition metals can be exhibited at wavelengths near 500 nm (Ben-Dor, Irons, and Epema 1999). Both of C and iron content can be directly related to absorption in spectral reflectance through a number of fundamental groups such as hydroxyl, carboxyl, and amine group which can be detectable in VNIR (Wetterlind 2009). The carbonate and salt minerals also include many of the molecular bonds absorbing in region about 2300 nm (Clark et al. 1990).

**NIR to detect soil Pb using prediction models**

The summary of some prediction models for soil Pb, total C, iron content, and soil texture (sand and clay fractions) are shown in Table 3, and the confidence level of $P = 0.005$ was used in all prediction models. From the result of using the entire set of soils (calibration set $n = 73$), it was observed that Pb concentrations predicted from spectral reflectance measurements were in poor agreement with the known concentrations of Pb in the soils obtained from soil Pb references (Table 3). However, a better prediction result was found when using log scale of Pb concentration via reflectance. In addition to this result, the PLSR prediction for the Pb concentration model in Figure 3 showed a good correlation ($R^2 = 0.54$) and a low RMSE (RMSE$_v = 0.38$) for the validation mode with acceptable RPD and RER (1.6 and 7.7, respectively).

All of these measures of accuracy indicated a suitable prediction model (Bilgili et al. 2010; Nduwamungu et al. 2009). This result agreement with Hayes 2012 who recommended that the numbers obtained for the RER typically be around four to five times larger than those for the RPD. This indicated that spectral reflectance analysis could successfully predict soil Pb concentrations for a set of Port Pirie soils ($n = 73$) on a log scale.

The NIR is based on molecular vibrations, which need to be robust to have strong bands and to be strong enough in the fundamentals to still have overtone and combination bands (Csorba et al. 2012). Therefore, the Pb concentration predicted values in this experiment may depend on the level of interactions with the spectrally active components throughout the range of 350–2500 nm.

### Table 2: Correlation coefficient ($R$) between Pb concentration and wavelength (nm) and other elements that have distinct spectral signatures.

<table>
<thead>
<tr>
<th>Variables</th>
<th>SR regions with best correlation (nm)</th>
<th>Correlation coefficient ($R$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb (mg kg$^{-1}$)</td>
<td>512</td>
<td>0.334</td>
</tr>
<tr>
<td>Log Pb</td>
<td>512</td>
<td>0.385</td>
</tr>
<tr>
<td>N (%)</td>
<td>1404</td>
<td>0.623</td>
</tr>
<tr>
<td>C (%)</td>
<td>400</td>
<td>0.592</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>1972</td>
<td>0.332</td>
</tr>
<tr>
<td>Fe (mg kg$^{-1}$)</td>
<td>548</td>
<td>0.384</td>
</tr>
</tbody>
</table>

SR: Spectral reflectance, NA: Not applicable

### Table 3: Summary of statistics for chemometric analysis represented prediction models for soil Pb and other spectrally active soil properties

<table>
<thead>
<tr>
<th>PLSR model with normalized treatment for:</th>
<th>No. of samples</th>
<th>No. of components</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$R^2$</td>
<td>RMSE$_c$</td>
</tr>
<tr>
<td>Pb mg kg$^{-1}$</td>
<td>73</td>
<td>7</td>
<td>0.19</td>
<td>598</td>
</tr>
<tr>
<td>Pb log$^c$</td>
<td>73</td>
<td>7</td>
<td>0.63</td>
<td>0.32</td>
</tr>
<tr>
<td>Fe log$^c$</td>
<td>73</td>
<td>7</td>
<td>0.49</td>
<td>0.21</td>
</tr>
<tr>
<td>% Carbon</td>
<td>73</td>
<td>7</td>
<td>0.76</td>
<td>0.86</td>
</tr>
<tr>
<td>% Sand texture</td>
<td>73</td>
<td>7</td>
<td>0.60</td>
<td>11.5</td>
</tr>
<tr>
<td>% Clay texture</td>
<td>73</td>
<td>7</td>
<td>0.29</td>
<td>7.87</td>
</tr>
</tbody>
</table>

The confidence level of $P = 0.005$ was used in all prediction models.

$^a$RPDV is the ratio of the standard deviation (SD) of the reference value to the RMSE in validation, suitable values of 1.4–2

$^b$RERV is the ratio of error range in the reference value (Max-Min) by the RMSE in validation, with suitable values of >6

$^c$Log scale was taken for elements that were atypically distributed
Figure 3. Relationship between predicted and measured Pb concentrations (log mg kg$^{-1}$) for the PLSR models. Construction model (A) and Validation model (B).
general, HMs are not considered spectrally active because they may represent much weaker large-order overtones of the soil on the level of interactions with the spectrally active components throughout the range of 350-2500 nm.

In general, Pb is a spectrally inactive metal; however, Pb concentration can be measured indirectly by correlation with another soil constituent that has the responsibility of RS. Correlation analysis between soil Pb concentration with spectral parameters and soil constituents can be used to detect significant factors for prediction models. In the present study, the correlation coefficient values between Pb concentrations with spectral parameters and some soil components which were expected to have a relationship to spectral data was showed in Table 2. The Pearson correlation coefficient between Pb and the corresponding absorption band values ($\gamma$) showed a best correlation of $R = 0.33$ on a natural scale and $R = 0.33$ (same) on a log scale at a wavelength of 512 nm. While the best correlation for Pb concentration was found with total C, where $R = 0.29$ on a natural scale and $R = 0.44$ on a log scale. The result in Table 4 and Figure 4 shows that the removal of nine outlier samples from the regression data set ($n = 73$) was a good process to improve the accuracy of Pb concentration predictions.

The correlation line for PLSR model highlighted the regions of high importance in the spectra to detect soil Pb (Figure 5). The component values between dotted lines within the upper and lower bounds evident in the loadings plot are highly correlated with the measured data. It is clear from this figure that the wavelength between 500 and 612 nm may be referred to important spectra to detect Pb concentration which was in agreement with the mentioned Pearson correlation.

| Table 4. Comparison between prediction accuracy factors for cross-validation of PLS regression for Pb concentration before and after removable outliers values. |
|-----------------|--------|--------|--------|--------|--------|
| No. of samples  | $R$    | $R^2$  | RMSE   | SLOP   | SECV   |
| All samples     | 0.70   | 0.54   | 0.38   | 0.58   | 0.39   |
| After removal of nine outlier values | 0.75   | 0.57   | 0.36   | 0.64   | 0.36   |

$SECV$: Standard error cross-validation.

Figure 4. Scatterplots of cross-validation model for predicted Pb concentrations after removal of nine outliers.
To validate the regression model, the predicted values should be equal to the reference values. Regression analysis between predicted model and actual concentrations for seven extra samples (which were not used in the data set) indicated that the PLSR model had the predictive power to assess Pb contamination to a good degree as evidenced by the linear regression correlation of $R^2 = 0.60$ (Figure 6). The final regression model is the one that, in general, exhibited the greatest goodness of fit between predicted and actual Pb concentrations (Al Maliki, Bruce, and Owens 2012a).

The final models developed here clearly demonstrated that spectral reflectance could potentially be used to predict soil Pb concentrations in soils contaminated by long-term atmospheric smelter Pb emissions. Nevertheless, there are two critical points to consider. First, the results can be quite site-specific and are unlikely to transfer to another site for two reasons: the calibration set of less than 100 samples was not enough to be generalized and the prediction models generated here do not include independent validation data. Second, Pb prediction models are not standardized because they are based on the feature space of the study area, which may not be applicable to another area. Therefore, the use and interpretation of models by other users requires some additional factors to be addressed such as soil type, particle size, soil texture, and organic matter in order to be truly effective.

**Using spectrally active of soil constituents as indicator to detect soil Pb**

This research showed that soil C regression model has the best detection in wavelengths (Table 3), and it has also shown that the best correlation was found between soil Pb and soil C. The spectral predictive mechanism using C fraction as a key for an indirect spectroscopy method to detect Pb concentration was applied in the VNIR region. High correlation coefficients for C contents were obtained in the visible spectral range (500–700 nm) (Ji et al. 2010). While in the current study, both Pearson correlation result of $R > 0.50$ and correlation loading for regression model observed that the wavelengths between 380 and 550 nm were important to detect soil C. Another study mentioned that the suitable correlation between organic carbon (spectrally active content) with a set of HMs such as Cu, Pb, and Zn (spectrally nonactive)
Figure 6. Evaluation of Pb concentration for seven extra samples (out dataset) using predicted PLSR model.
(with \( R = 0.67, 0.66, \) and 0.65, respectively) was helpful to reveal coherences between these variable and spectral bands than that identified (predicted) of HMs in similar spectral regions of organic carbon (Vohland, Bossung, and Fründ 2009). The organic matter present in soil which constitutes a large fraction of the C content of a soil is considered to be the more spectrally active due to strong molecular vibrations in the NIR region. As a result, the spectral band of Pb at 512 nm was in the range of spectral bands for total C. So, soil C contents may be useful to assist in predicting Pb concentration in a specific spectral range of C due to strong bond that exists between them.

**Conclusion**

This study has shown that NIRS may potentially estimate the total Pb of soil around the Port Pirie smelter. Although Pb is a spectrally inactive metal, a good predictive model for Pb in VNIR was achieved by combining soil RS with soil measurements using a PLSR model. The regression model showed that Pb concentration when converted to a log scale correlated optimally with the modified reflectance spectra. In addition to cross-validation, the accuracy and precision of the PLSR model were determined using an extra set of data that was not used in the original prediction. This data set was evaluated through many prediction accuracy parameters such as \( R^2 \), RMSE, RPD, and RER. The chemometric methods also used correlation loading function to highlight the regions of high spectral importance between 500 and 612 nm for Pb content determination. Total C was more spectrally active and more correlated with Pb than other elements in soil such as iron oxides and clays. In addition to an acceptable correlation between total C and soil Pb, the region associated with Pb detection was also coincident with the region where soil C could be detected. Thus, these results indicated that the absorption features for C content can be utilized to predict soil Pb. While the regression model for Pb concentration developed here was indeed quantitative and accurate, this model was developed for a specific soil and may not be applicable to another soil that has different feature space.

**Acknowledgments**

All authors acknowledge the support of the future Industries Institute (FII), the University of South Australia, while undertaking this research. The first author gratefully acknowledges the financial support of the Iraqi government, Ministry of Higher Education and Scientific Research for a Ph.D. stipend.

**References**


Bickel, M. 2010. Spatial and temporal relationships between blood lead and soil lead concentrations in Detroit, Michigan. M.S diss., Civil and Environmental Engineering, Wayne State University, USA.


