Abstract
In this work is presented a method based on reflectance spectroscopy Visible-Near Infrared (VIS-NIR) to evaluate soils suspected to be contaminated by heavy metal. The concentration of Cobalt levels on several samples of soil were determined using statistical methods (Partial-Least Square Regression) on measured data.

© 2012 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Organizing Committee of the ENIINVIE-2012. Open access under CC BY-NC-ND license.

Keywords: Difuse spectroscopy reflectance (DSR), Partial-Least Square regression (PLSR), Visible-Near infrared (VIS-NIR).

1. Introduction

The study of soil for agricultural applications has always been of great importance due to the high demand for food to meet the growing needs of the population. Therefore, the analysis methods should be increasingly accurate in predicting, for example, detection of contaminants or nutrients in soils. Chemical methods for soil analysis are very accurate. However, these methods require a laborious laboratory analysis after obtaining the samples in the field. Moreover, the spectral reflectance has shown to be a powerful tool for such studies in agricultural applications because it allows to know the chemical and physical state of the soil, providing results in real-time and on-site of interest due to its portability. Furthermore, that this method can be adjusted to provide results for more than one attribute of the soil with a single analysis.

Using reflectance spectroscopy in the visible and near infrared of the electromagnetic spectrum (VIS-NIR)(400-1000 nm) can predict the existence of common soil contaminants, such as, heavy metals (Ni, Cr, Co, Cu, Pb). Using a predictive model of the spectra measured, the quantitative values of the attributes can be obtained, this is, contaminants. The model is based in statistical analysis of the spectrum signal to be able to obtain the information of interest on the soil studied. For this study we used PLSR (Partial Least Square Regression) in order to obtain quantitative values related to soil chemical condition. As well as, variance analysis of spectral data (variance and standard deviation).

Email address: miranda.daniel@uabc.edu.mx (D. Miranda Salazar)
Latest researches has shown that diffuse reflectance spectroscopy (DRS) using VIS-NIR of the electromagnetic spectrum can provide data low cost prediction on the chemical, physical and biological soil properties [1]. So that (DRS) has become an alternative to soil survey.

NIR spectroscopy has been used for several years in soil science for measuring carbon and nitrogen, metal oxides, cation exchange capacity (CEC), wilting point and particle size. The technique has proved to be simple, quick and not destructive to the environment [2]. It has also proved to be more accurate than conventional methods. For example, the VIS spectroscopy can be more exact that dichromate digestion for analysis of organic carbon [3].

2. Preparation of soil samples

For this study were analyzed 45 samples of soil contaminated with cobalt nitrate hexahydrate \((\text{Co(NO}_3\text{)}_2 \ast 6\text{H}_2\text{O})\) at different concentrations, which were scanned with an optical fiber probe (R200-Angle Reflection Probe of Ocean Optics) to obtain the spectra. The model was developed using statistical process of calibration based on partial least squares regression (PLSR) on the spectral data. It must be mentioned that before developing the model, analysis of variance were conducted to a set of samples with the same concentration in order to find the wavelengths that provided noise spectra. After detecting the wavelengths that provided noise spectra was defined the spectral range scanning. The same set of samples there was performed a test of variance in order to determine the number of scans necessary for obtaining quantitative information from a soil sample. Once the optimal number of captured spectra was calculated, we proceeded to implement the technique PLSR for predicting coefficients pollutant concentration corresponding to each wavelength of the spectrum sample.

2.1. Sieving the soil

The soil samples used for the experiment were extracted from a wineyard area a depth of 35 cm, which had a granular structure, and a sandy loam texture type.

After removing the sample, the sample was transferred to the laboratory and sieved with a No. 30 sieve (< 2 mm opening) and placed in containers of 90 ml, with 100 g of soil in each container. A total of 45 soil samples were prepared using 40 samples to develop the prediction model (Table 1) and 5 samples for model validation (Table 2).

2.2. Preparation of solutions of \(\text{Co(NO}_3\text{)}_2 \ast 6\text{H}_2\text{O}\)

Previous research demonstrated that the technique of spectroscopy using VIS-NIR region, is able to detect heavy metals in concentrations greater than 1 g/kg with great accuracy [4], [5]. So, it was necessary to know the molecular mass of \(\text{Co(NO}_3\text{)}_2 \ast 6\text{H}_2\text{O}\) in order to determine the amounts of mass of this compound to get a gram of Cobalt (Co).

Using the compounds data, there were proposed forty concentrations, computed algebraically. Since for each gram of cobalt, it was necessary to use five grams of compound, as twenty per cent of these five grams, corresponding to cobalt. Therefore, there were prepared forty concentrations in solutions of 13 ml of distilled water (Table 1). Also, five more samples were used to validate the prediction model, covering the entire range of the forty proposals concentrations (Table 2).

2.3. Laboratory procedure

With the proposed concentration levels (Table 1 and 2 ), we proceeded to perform the following task in the chemical laboratory: The mixture was made of solutions with soil samples corresponding to a mass of 100 g each. Subsequently, a homogenate was made of the solution and the sample, which produced a mixture sludge, which was introduced in an oven at a temperature of 125°C for 24 hours. At the end of drying, soil samples were ground using a mortar, and leave then as when sieving was applied since the samples come petrified out of the oven.
Table 1. Concentrations proposals for the development the prediction model

<table>
<thead>
<tr>
<th>Number concentration (samples)</th>
<th>Compound (\text{Co(NO}_3\text{)}_2 \cdot 6\text{H}_2\text{O}) g/13 ml</th>
<th>Cobalt (Co) g/13 ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
<td>0.15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.2</td>
</tr>
<tr>
<td>4</td>
<td>1.25</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>0.3</td>
</tr>
<tr>
<td>6</td>
<td>1.75</td>
<td>0.35</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0.4</td>
</tr>
<tr>
<td>8</td>
<td>2.25</td>
<td>0.35</td>
</tr>
<tr>
<td>9</td>
<td>2.5</td>
<td>0.5</td>
</tr>
<tr>
<td>10</td>
<td>2.75</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 2. Validation concentrations

<table>
<thead>
<tr>
<th>Validation concentration (sample)</th>
<th>Compound (\text{Co(NO}_3\text{)}_2 \cdot 6\text{H}_2\text{O}) g/13 ml</th>
<th>Cobalt (Co) g/13 ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>3.9</td>
<td>0.78</td>
</tr>
<tr>
<td>3</td>
<td>5.8</td>
<td>1.16</td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>1.5</td>
</tr>
<tr>
<td>5</td>
<td>9.9</td>
<td>1.98</td>
</tr>
</tbody>
</table>

3. Calibration and analysis

3.1. Materials and equipment

For spectral analysis was used an optical fiber probe (**R200-Angle Reflection Probe from Ocean Optics**), which has a set of seven fibers, each of them with a diameter of 200μm, one fiber is used for lighting and the other for reading spectral signatures. The system is integrated with a spectrometer (**USB4000 from Ocean Optics**) responsible of converting light signals into digital signals for further acquisition by a PC and then to be processed using the software **Spectra Suite Ocean Optics**. After acquiring the spectra, it was necessary to use some routines written in **MatLab** to import and process data in order to generate the prediction model.

For the spectroscopic analysis we used the experimental diagram shown in (Figure 1).
3.2. Experimentation

Spectroscopic analysis required the use of a material with a diffuse surface as a reference, for this, studies there were conducted a test with different materials. It was shown that the photographic paper is an excellent choice for using as reference. Data from *Spectra Suite Ocean Optics* was observed and stored as the spectral signature of this material, which was used as a reference for capturing and storing the spectral signatures of soil samples.

The data stored by the software *Spectra Suite de Ocean Optics* were exported using *Matlab*.

3.2.1. Wavelength range

The range of the electromagnetic spectrum used in this study was limited by the spectrometer, because of its operating range (345 nm to 1040 nm), with an optical resolution of $\sim 0.3$ nm.

In the first scannings was observed that the spectra contained reflectance values for 3646 wavelengths, therefore, there were captured 100 spectrum with a rate of an spectrum per second. These spectra were stored in an array with dimensions of 100 x 3646 (Figure 2), where rows represent the number of spectra obtained from a single sample, and the columns represent each wavelength.

These data were subjected to statistical analysis that helped identify the wavelengths containing information of the concentration of cobalt. Making an analysis of the data we were able to eliminate the wavelengths that provided noise, which helped on reducing dimensions of the matrix mentioned above, and simultaneously, the amount of data, speeding up the processing information for the development of the prediction model.

The statistical analysis was used to determine and store the standard deviation for each wavelength in a new vector called vector of standard deviation or variance vector ($\sigma^2$), it shown in Figure 2. As 3646 wavelengths are were used, the same amount of values of $\sigma$ were obtained, and as every wavelength corresponds to 100 values for each spectra of the soil sample analyzed.

To find out the wavelengths to be discriminated, the vector was plotted against the standard deviation vector wavelengths as shown in Figure 3, which contains the lengths provided by the spectrometer (345 nm – 1040 nm). Graphically the peaks corresponding to noisy signals, could be seen.
Figure 3 shows a constant region of wavelengths between 450 and 980 nm. So, we can consider that this area is providing the information, or at least does not add noise. Therefore, with this data we were able to confirm the interval at which it worths to. (Figure 4).

For comparison, a spectrum without statistical processing is shown in Figure 5, which means that noisy signals were not eliminated.

3.2.2. Determination of the number of scans

For the elaboration of model prediction and validation, 45 contaminated samples of soil were interrogated with the probe, capturing 100 spectra per sample at a rate of one spectrum per second, so that each
sample was submitted to spectroscopic analysis for approximately two minutes to capture the 100 spectra. Spectra were obtained randomly caught on the surface of the soil samples, contained in a petri-dish container for such analysis.

It was shown that 60 Spectra were enough to obtain all the information for each sample. This was achieved using a statistical analysis similar to that mentioned above with the same data matrix (Figure 2) used for the discrimination of wavelengths.

This analysis consisted in determining variance ($\sigma^2$) and standard deviation ($\sigma$) for 100 groups of spectra as shown in Figure 6.

From the graph of Figure 6 shows that from 40 scans (40 scans are 40 spectra of a sample) we can see a constant interval in the function, which means that after 40 spectra can be obtained more information of the analyzed samples. In this paper we proposed to obtain 60 spectra per sample to ensure that the sample information was reliable.

3.3. Spectroscopic analysis

From the results obtained, we proceeded to analyze spectroscopically the 45 samples to elaborate a data matrix, one for model development and another for validation samples. Finally, 60 spectra were captured for each of the 45 samples with the reflection probe (optical fiber), which was placed normally on the surface of the samples with a separation of $\approx$2 mm on an optical table.

4. Data processing

Captured spectra were imported using Matlab to be processed, in order to perform a model that provided accurate predictions about the quantity of cobalt contained in each soil samples.
4.1. Smoothing of data

Spectra from each sample were stored in a matrix with dimensions of 60 x 2811, where each of rows corresponds to a spectrum and each of columns corresponds to a wavelength. Therefore, there were obtained 45 matrices with the same dimensions, corresponding to each soil sample.

Every array was smoothed to obtain cleaner contours, therefore, accomplishing to obtain this was accomplished spectra as shown in Figures 7 and 8, although this may also mean loss of information by applying smoothing using Matlab. This process was done using programming code to automate the process, and improve system performance.

Fig. 7. Smoothed spectrum

Fig. 8. Unsmoothed spectrum
4.2. Prediction model

With the smoothed spectra, we proceeded to perform a spectrum averaged for each sample. Finally obtaining an information array for each of the 45 contaminated soil samples.

Such vectors of information, were concatenated to generate a data matrix ("Data matrix X") with dimension 40 x 2811, where in each of the rows representing each of the calibration vectors and the columns each of the wavelengths, as can be observed in Figure 9.

![Fig. 9. Data matrix & response vector](image)

In Figure 9 it can be observed a vector called the "response vector Y", which has dimensions of 40 x 1, and each of the rows of this vector contains the values of the concentration of the compound $Co(NO_3)_2 \cdot 6H_2O$ corresponding to concentrations that were mixed with soil samples for model development. The first row contained the value of lower concentration, and the last, the value of higher concentration of the compound. The data in this vector are the same as those contained in Table 1. The data matrix and the vector of responses are needed to determine a relationship between them, and thus be able to make predictions of future concentrations of cobalt (Co) external to the calibration samples.

4.2.1. PLSR (Partial least squares regression)

In order to find out the correlation coefficients between the data matrix $X$ and vector $Y$ response in Figure 9, it was implemented a PLS regression using the Matlab statistics tools, which helped us to calculate a vector $\beta$ with coefficients of relation for each of the wavelengths.

$$Y = [X] \ast [\beta] + [Residuals] \quad (1)$$

From Equation (1) was used the vector $\beta$ to make predictions from of samples used different from those to find the vector relationship, for this, 5 validation samples were used outside of calibration samples. The spectra of these validation samples were stored in a prediction matrix $X^*$ as shown in Figure 10. The product between the prediction coefficients $\beta$ and the matrix $X^*$, provided a new prediction vector $\hat{Y}$ as shown in equation (2).

![Fig. 10. Prediction matrix $X^*$](image)

$$\hat{Y} = [X^*] \ast [\beta] \quad (2)$$
The vector $\hat{Y}$ contains the predictions of the concentrations of these validation samples.

5. Results

The data obtained after the prediction model applied to the validation samples confirmed the results of other investigations [4] and [5], although the model showed problems to predict the concentration of cobalt at 1 g/kg, their behavior was appropriate for samples with higher concentration. Thus demonstrated that heavy metals are detectable with spectroscopy in the VIS-NIR range at concentrations greater than 1 g/kg in soils.

Table 3 shows the predictions ($\hat{Y}$) of the model and the concentrations used in the soil samples that were used for validation. It is noticeable that the values predicted by the model in the samples 2, 3, 4 and 5 are very close to the concentrations of Co prepared in the chemical laboratory.

<table>
<thead>
<tr>
<th>Validation (Samples)</th>
<th>Compound</th>
<th>Cobalt (Co)</th>
<th>Model $\hat{Y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Co(NO$_3$)$_2$.6H$_2$.O</td>
<td>0.12</td>
<td>0.283</td>
</tr>
<tr>
<td>2</td>
<td>3.9</td>
<td>0.78</td>
<td>0.7302</td>
</tr>
<tr>
<td>3</td>
<td>5.8</td>
<td>1.16</td>
<td>1.3278</td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>1.5</td>
<td>1.6150</td>
</tr>
<tr>
<td>5</td>
<td>9.9</td>
<td>1.98</td>
<td>1.7284</td>
</tr>
</tbody>
</table>

6. Conclusion

It was possible to confirm the ability of reflectance spectroscopy (using the visible and near infrared) for the evaluation of heavy metals in soils, which makes it an alternative to assess soil contamination, because pollution (contamination) the problem is growing as the population grows. Therefore, the VIS-NIR spectroscopy is a good option of evaluation, as it a non-destructive analysis method, not harmful to the environment, among other advantages.

Furthermore, statistical techniques have shown to be a powerful tool in spectroscopic analysis. Because they can be used for chemical and physical analysis, either in food, soil, water, etc. Depending on the quality of the optical equipment.

Acknowledgements

The authors wish to acknowledge the support provided by CONACYT and FIAD-UABC for this work.

References


