Title:

Soil organic carbon (SOC) prediction using VNIR reflectance spectroscopy employing artificial neural network (ANN) modelling

Authors:

Justin George K*, Suresh Kumar*, Arya Raj R*

*Agriculture and Soils Department (ASD),
Indian Institute of Remote Sensing (IIRS), ISRO
4 Kalidas Road, Dehradun, Uttarakhand, India – 248001

* Corresponding Author:
Justin George K
Agriculture and Soils Department (ASD),
Indian Institute of Remote Sensing (IIRS), ISRO
4 Kalidas Road, Dehradun, Uttarakhand, India – 248001
Phone: 0135-2524139
Email: justin@iirs.gov.in

Co-authors:
Suresh Kumar: suresh_kumar@iirs.gov.in
Arya Raj R: aarthiraj9@gmail.com
Soil organic carbon (SOC) prediction using VNIR reflectance spectroscopy employing artificial neural network (ANN) modelling

Abstract

Visible-Near Infrared (VNIR) spectroscopy has been identified and explored as a relatively fast and cost effective analytical technique for estimating soil organic carbon (SOC). The present study was undertaken for predicting soil organic carbon using VNIR reflectance spectroscopy employing Artificial Neural Network (ANN). Surface soil samples (0-15 cm) were collected from 75 georeferenced locations through grid sampling approach in a hilly watershed of Himachal Pradesh and analyzed for SOC. The reflectance spectra of soil samples was measured using ASD Spectroradiometer in the wavelength range of 350 - 2500 nm. Various spectral indices were generated using the sensitive bands in the visible region. The SOC sensitive spectral indices and reflectance transformations were used for predictive modelling of SOC through Artificial Neural network (ANN) model. ANN model could predict the SOC values with an $R^2$ of 0.92 and MSE value of 0.24 and indicated that this technique can be used to predict the SOC in a spatial domain when coupled with high-resolution hyperspectral satellite/airborne data.

Keywords: Soil organic carbon (SOC), VNIR reflectance spectroscopy, artificial neural network model (ANN)

Soil organic carbon (SOC) plays a fundamental role in determining the physical, chemical as well as biological properties of soil. SOC has been known to be beneficial for maintaining soil productivity, water holding capacity as well as carbon sequestration for alleviating the ill effects of greenhouse gases (GHGs) and thus climate change. SOC holds about 4.5 times the amount of the biotic carbon pool and 3.3 times the amount of the atmospheric carbon pool. Thus reliable estimation of SOC is vital for understanding the human induced effects on global carbon cycle and the associated climate change. Conventional methods for SOC determination in soil laboratories are costly, time taking and may be environmentally hazardous. Thus there is an urgent need for the development of fast, accurate and non-destructive methods (thus reducing the number of soil chemical analysis) for
SOC estimation as an alternative to the traditional methods (laborious and time consuming) which help us in generating high-resolution soil property maps of large areas at modest costs.

Visible and Near Infrared (VNIR) spectroscopy has been identified and explored as a relatively fast and cost effective analysis technique for various agricultural paradigms by Williams and Norris\textsuperscript{7}. Spectroscopy has proved its potential to accurately determine soil organic carbon (SOC) in laboratory as well as in field where nondestructive imaging can be performed\textsuperscript{8,9}. Studies by Viscarra Rossel \textit{et al} \textsuperscript{10} had proven that NIR spectroscopy can estimate several primary as well as secondary soil properties like SOC, Total N and Ca, K, and Mn respectively. The electromagnetic range of 400 - 2500 nm in spectroscopy enables accurate prediction of soil organic carbon due to the influence of SOC on the shape and magnitude of reflectance spectrum\textsuperscript{11}. Zhuo \textit{et al}\textsuperscript{12} mapped soil organic carbon using spectrometer and also used spectra from laboratory to map physical and chemical properties of soil with special emphasis on characteristic absorption bands of soil. Also the potentiality of spectroscopy in mapping and quantifying a wide range of soil properties was reviewed by Ben-Dor \textit{et al}\textsuperscript{13}, as an effective tool for advanced soil mapping studies. Near infrared spectroscopy using field and space borne sensors enables direct mapping of soil properties and assessment of soil quality.. This capability of field or lab spectra was reviewed by researchers and proved to be ideal for soil related studies\textsuperscript{14}. ASD spectrometer was also successfully used in estimating soil carbon, nitrogen, carbonate and organic matter in various horizons of soil profile from different sites in Washington and Oregon at 400 to 1000 nm using regression tree model\textsuperscript{15}.

Artificial neural network (ANN), a potent data mining technique, was found to give better results in predicting soil organic carbon content, clay content and soil pH than other regression and statistical techniques\textsuperscript{16}, may be because of its strong connectivity of networks that could possibly draw out effective relationship between soil properties and spectral features. Also there are reports that soil map derived out of ANN was found to be more accurate than all other conventional methods of soil mapping\textsuperscript{17}. ANN uses various supervised networks like feed-forward neural network (FFN) with Backpropagation network architecture, which acts as a channel where the output defines the pattern and function of network channel, leading to inputs which can be further used in a reverse manner with
known inputs to derive unknown outputs. Multilayer feedback propagation algorithms facilitate ANN training in a much effective manner and was used for estimating soil nutrients like phosphorus using DEM derived terrain attributes, which proved suitability of ANN in quantitative prediction of soil nutrients. Therefore, the present study was undertaken to verify the possibility of employing Artificial Neural network modelling for predicting soil organic carbon using VNIR reflectance spectroscopy.

The study was conducted in a watershed located in the mid Himalayan region of Mandi district Himachal Pradesh (latitude 32° 4' 35.04'' N to 32° 1' 3.8964'' N and longitude 76° 39' 49.60'' E to 76° 44' 15.84'' E). The study area covered a total geographical area of 1000 ha (10 km²). Nearly 80% of the watershed is comprised of agricultural fields where rice and wheat are the major crops grown. The complete methodology implemented in the study is given in Fig 1. To properly represent the soil spatial variability in the watershed and ensure unbiased and precise sampling, grid sampling approach was adopted for soil sample collection, with a grid size of 250m x 250m on ground. Georeferenced surface soil samples (0-15 cm) were collected from 75 locations in the watershed during November 2015, when the fields were fallow. The soil samples were air dried, preprocessed and divided into two parts, one part for analysis of SOC as well as basic soil physio-chemical characteristics and other part for the generation of reflectance spectra in the laboratory. Preprocessed and 0.2mm sieved soil samples were used for the estimation of soil organic carbon (SOC) using TOC analyzer. Based on analytical results soils in the watershed were found to be predominantly acidic in nature with pH values ranging from 4.34 to 5.61 (mean value of 4.85). The electrical conductivity (EC) values were very low, indicating the soils are devoid of soluble salts/salinity. The soils in the area were found to be loamy in texture, predominantly belonging to silt loam and sandy loam textural classes. The samples were found to have medium to high soil organic carbon content (SOC), with majority of the values greater than 0.5%. The SOC values ranged from 0.26 to 2.71, with a mean value of 1.36 g kg⁻¹ soil. The soils belonged to moderate to well drainage class.

The surface reflectance spectra in the wavelength range of 350 – 2500 nm, of preprocessed soil samples were measured using Analytical Spectral Device (ASD) FieldSpec Pro Spectroradiometer in laboratory, under controlled dark room conditions. Precautions were taken to avoid stray light as well as dark
current generated within the instrument and instrument was calibrated using white reference panel at
the start as well as after every five successive reflectance measurement. The spectra were collected as
an average of 25 readings for each sample, with a spectral resolution of 1 nm, after real time viewing.
For each sample data, five spectra were collected and averaged out for creating spectral libraries. The
spectra collected at 1 nm interval were also resampled to 5 nm interval using spectral resampling tool
in ENVI 5.0 for further analysis.

Sensitive spectral bands specific to soil organic carbon were studied through literature survey\textsuperscript{15, 21, 22, 23}. Various bands identified for their sensitivity to SOC were 400, 441, 520, 907, 960, 1100, 1720, 1744,
1870, 2052, 2180 and 2309 nm (Gmur et al\textsuperscript{15}, Daniel et al\textsuperscript{21}, Shepherd and Walsh\textsuperscript{22}, Dalal and Henry\textsuperscript{23}).
Correlation between selected spectral bands and SOC data was done and the sensitive bands were
identified. In the visible range (400-700 nm), spectra were divided into three regions i.e., blue (400-500
nm), green (500-600 nm) and red (600-700 nm), for analysis and identification of most sensitive bands
within each of these three regions. For rest of the spectra, only the bands identified from literature were
used for analysis.

The spectral bands selected through correlation analysis were used to generate spectral indices. Five
nano-spectral indices were developed namely Brightness Index (BI), Colouration Index (CI), Saturation
Index (SI), Redness Index (RI) and Hue Index (HI) (Equations 1 to 5). The indices are image color
indices, which were reported to be sensitive to bare soil properties\textsuperscript{12, 24}. Since variations in SOC content
influences the soil color, these indices were chosen for the current study. These indices make use of
blue, green and red wavelength regions of visible range, hence single nano bands in these regions having
highest correlation coefficient (r) were selected for developing indices. Reflectance at 451 nm, 520 nm
and 690 nm representing blue, green and red regions respectively were selected (Table 1). The indices
derived were then used for correlation analysis with the SOC. The bands in the shortwave IR region
especially 2180 and 2309, were also found to have higher correlation with SOC values (Table 1). Owing
to the significance of NIR region in predicting soil organic carbon, the bands in NIR region along with
its first derivative, second derivative and their logarithmic values, reciprocal of logarithmic values and
reciprocals of derivative were also subjected to correlation analysis.
Brightness Index:

\[ BI = \sqrt{\left(Blue^2 + Green^2 + Red^2\right)/3} \]  \hspace{1cm} \text{Eq. 1}

Colouration Index

\[ CI = \frac{Red - Green}{Red + Green} \]  \hspace{1cm} \text{Eq. 2}

Hue Index

\[ HI = \frac{2 \times Red - Green - Blue}{Green - Blue} \]  \hspace{1cm} \text{Eq. 3}

Redness Index

\[ RI = Red^2 + \left(Blue \times Green^2\right) \]  \hspace{1cm} \text{Eq. 4}

Saturation Index

\[ SI = \frac{Red - Blue}{Red + Blue} \]  \hspace{1cm} \text{Eq. 5}

Where, 451 nm, 520 nm and 690 nm wavelengths chosen as blue, green and red bands respectively

ANN model for soil organic carbon prediction was carried out using Matlab software Ver 2015. The nano spectral indices that showed good correlation with SOC were used for analysis (Table 2). In addition to the nano spectral indices, the reciprocal of first derivative of reflectance value at wavelength 2309 and reflectance value of band 2180 showed high correlation coefficient and were also taken as input to the ANN. The data preparation and subsequent data base generation was done in Excel and the neural network was executed using Neural Network Tool in Matlab. The entire soil dataset was separated into two, as calibration (50 no.) and validation (25 no.) data sets using random numbers. During the model training and development phase, the calibration set was further internally separated into 03 sub datasets i.e., training (68 %), validation (16 %) and testing (16 %) datasets.

A multilayer Feed-forward Neural network (FFN) consisting of input, hidden and output layers known for its better performance in solving intricate input-output relationship was used for the prediction of SOC. To circumvent overtraining issue coupled with FFN, we used back propagation algorithm for network training, which is recognized for its easy execution. ANNs are known for their capability to model and represent functions of non-linear nature. A Tan-Sigmoid function known to proximate non-linear interactions between inputs and outputs was chosen as the transfer function, for the hidden and output layers. Overall network was trained using Levenberg-Marquardt algorithm and the network
performance was evaluated using R values from the regression plots. The developed ANN model was further employed for prediction of SOC values corresponding to the input variables of validation dataset. The SOC prediction accuracy of the model was evaluated using the coefficient of determination (R2) and MSE (Mean Square Error) values.

Based on the results of Pearson correlation test, 07 variables (CI, BI, HI, SI, RI, R\textsubscript{2180} and reciprocal of 1\textsuperscript{st} derivative of R\textsubscript{2309}) were selected for predictive modeling of SOC using ANN. Thus, the designed ANN network had 07 input layers corresponding to the various predictive variables. Based on the trial and error method suggested by Chiang et al 27, number of neurons in the hidden layer of the network for prediction of SOC was optimized to be 12. The training of the network was assessed using the regression plots of training, validation, testing (three sub datasets) and total calibration dataset, exhibiting R-values (Fig 2). Network training using the data was done until R-values of all the three sub datasets reached values greater than 0.9 or close to 1. The well-calibrated model was saved and used for the prediction of SOC values using the validation dataset. Using the well trained ANN model, we were able to attain a very precise prediction of SOC with an R\textsuperscript{2} value of 0.92 and the MSE of 0.24 (Fig 3).

This study demonstrated that VNIR spectroscopy can be an effective tool for quantitative prediction of various soil nutrients especially, SOC. The SOC prediction model developed using soil samples from a mid - Himalayan watershed could be used for the effective characterization and soil nutrient prediction in other regions. This nondestructive prediction model could help us in the real time evaluation of SOC, when coupled with the high resolution hyperspectral satellite/airborne data, enabling the farmers for adopting precision farming for sustainable agricultural production.

Acknowledgements

Authors sincerely thank the financial support received from Indian Space Research Organization (ISRO) under Earth Observation Applications Mission (EOAM) project (ISRO/DOS) on ‘Mountain Ecosystem Processes and Services’ for conducting this study.
References


25. Haykin, S., Neural Networks and Learning Machines (Third edition) Prentice Hall, New Jersey, USA, 1999


Tables (02 nos)

<table>
<thead>
<tr>
<th>SI No</th>
<th>Sensitive Bands</th>
<th>( r^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>451</td>
<td>0.49</td>
</tr>
<tr>
<td>2</td>
<td>520</td>
<td>0.40</td>
</tr>
<tr>
<td>3</td>
<td>690</td>
<td>0.52</td>
</tr>
<tr>
<td>4</td>
<td>907</td>
<td>0.37</td>
</tr>
<tr>
<td>5</td>
<td>1720</td>
<td>0.37</td>
</tr>
<tr>
<td>6</td>
<td>2180</td>
<td>0.77</td>
</tr>
<tr>
<td>7</td>
<td>2309</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Table 1: Correlation Coefficient (\( r^2 \)) of spectral bands with soil organic carbon (SOC)

<table>
<thead>
<tr>
<th>Spectral Indices</th>
<th>SOC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BI</td>
</tr>
<tr>
<td>SOC</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table 2: Correlation Coefficient (\( r^2 \)) of spectral indices with soil organic carbon (SOC)
Figure Legends

Fig 1: Overall methodology adopted in the study

Fig 2: Regression plot of accepted ANN model

Fig 3: Observed vs Predicted plot of SOC values

Figures (03 nos)
Fig 1: Overall methodology adopted in the study

Fig 2: Regression plot of accepted ANN model
Fig 3: Observed vs Predicted plot of SOC values