

# Prediction Modeling and Mapping of Soil Carbon Content Using Artificial Neural Network, Hyperspectral Satellite Data and Field Spectroscopy

## Sudheer Kumar Tiwari, Sudip Kumar Saha, Suresh Kumar

Agriculture and Soils, Indiana Institute of Remote Sensing, Dehradun, India Email: <u>sudheeriirs@gmail.com</u>, <u>sksaha@iirs.gov.in</u>, <u>suresh\_kumar@iirs.gov.in</u>

Received 2 March 2015; accepted 16 March 2015; published 23 March 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0/

Abstract

Soil organic carbon (SOC) is an important and reliable indicator of soil quality. In this study, soil spectra were characterized and analysed to predict the spatial soil organic carbon (SOC) content using multivariate predictive modeling technique-artificial neural network (ANN). E01-Hyperion (400 - 2500 nm) hyperspectral image, field and laboratory scale data sets (350 - 2500 nm) were generated which consisted of laboratory estimated SOC content of collected soil samples (dependent variable) and their corresponding reflectance data of SOC sensitive spectral bands (predictive variables). For each data set, ANN predictive models were developed and all three datasets (image-scale, field-scale and lab-scale) revealed significant network performances for training, testing and validation indicating a good network generalization for SOC content. ANN based analysis showed high prediction of SOC content at image ( $R^2 = 0.93$ , and RPD = 3.19), field ( $R^2 = 0.92$  and RPD = 3.17), and lab scale ( $R^2 = 0.95$  and RPD = 3.16). Validation results of ANN indicated that predictive models performed well ( $R^2 = 0.90$ ) with RMSE 0.070. The result showed that ANN methods had a great potential for estimating and mapping spatial SOC content. The study concluded that ANN model was potential tools in predicting SOC distribution in agricultural field using hyper-spectral remote sensing data at image-scale, field-scale and lab-scale.

## **Keywords**

Soil Carbon, Artificial Neural Network, Hyperspectral, Imagine Spectroscopy, Hyperion

How to cite this paper: Tiwari, S.K., Saha, S.K. and Kumar, S. (2015) Prediction Modeling and Mapping of Soil Carbon Content Using Artificial Neural Network, Hyperspectral Satellite Data and Field Spectroscopy. *Advances in Remote Sensing*, **4**, 63-72. <u>http://dx.doi.org/10.4236/ars.2015.41006</u>

#### **1. Introduction**

Soil organic carbon (SOC) is an important and reliable indicator of soil quality. Increasing soil organic carbon concentrations improve soil conditions because of better aggregation, higher infiltration rates and water retention which are conductive to their resistance to erosion. In dry land areas, water is the limiting factor for vegetation growth and influences organic matter production and SOC content. SOC represents a significant fraction of total amount of terrestrial carbon involved in the global carbon cycle. Managing SOC can enhance soil productivity and environmental quality, and can reduce the severity and economic loss of natural disasters. Soil organic matter (SOM) is one of the key soil properties influencing soil physical, chemical and biological processes which are primarily controlling soil fertility and plant growth. In addition, increasing SOC and SOM status can reduce atmospheric CO<sub>2</sub> levels that contribute to climatic change. Sustainable use of soils and protection of environment requires characterization and quantitative estimation of SOC/SOM content and its spatial and temporal variability. Remote sensing technique is now in a strong position to provide meaningful spatial data for use in soil investigations. Spectral signatures for the soil surfaces derived from remote sensing provide precious information for characterization of soils. Four main factors influence the soil reflectance in remote sensing images: mineral composition, soil moisture, organic matter content and soil texture (surface). Size and shape of the soil aggregate also influence the reflectance in the images. The mineral composition of soils affects the reflectance spectrum. Increasing reflectance of soils occurs from the visible to the shortwave infrared-with absorption bands around 1400 nm and 1900 nm related to the amount of moisture in the soil. Organic matter is the third factor that influences soil optical properties. Organic matter may indirectly affect the spectral influence, based on the soil structure and water retention capacity. Hyperspectral remote sensing combines imaging and spectroscopy in a single system which often includes large data sets and requires new processing methods. Hyperspectral data sets are generally composed of about 100 to 200 spectral bands of relatively narrow bandwidths (5 -10 nm). Research studies have shown that SOC and SOM consist of unique spectral fingerprints which can be correlated to their content and composition [1]. Spectroscopy, exploiting the information carried by reflectance in the visible and near-infrared and short wave infra-red part of the EMR from the soil, has demonstrated its capability to accurately determine SOC contents comparatively faster than traditional measuring technique, in the laboratory and field with a portable spectrometer as well as for regional study from aero-space imaging hyperspectral sensors [1]. Additionally, remotely sensed hyperspectral satellite data offer a synoptic view and a repetitive coverage which are two important advantages compared to ground observations and hyperspectral airborne data. Spectrometry (IS) or hyperspectral technology, as an advanced tool that provides high spectral resolution data (near-laboratory quality reflectance and emittance data) for each single picture element (pixel) from a far distance [2], significantly broadens the utility for further mapping of the soil surface from a more precise chemical and physical point of view [3]. With the development of hyperspectral sensors, spectral features related to characteristic absorption bands of soil organic matter can be mapped with more detail [4]. Artificial Neural Network (ANN) multivariate predictive modeling techniques was used by several researchers to predict spatial variation of surface soil properties using air-borne as well as space borne hyperspectral remote sensing data [1]. ANN offers a fundamentally different approach for modeling soil behavior. ANN is an oversimplified simulation of the human brain and composed of simple processing units referred to as neurons. It is able to learn and generalize from experimental data even if they are noisy and imperfect. This ability allows this computational system to learn constitutive relationships of materials directly from the result of experiments. Unlike conventional models, it needs no prior knowledge, or any constants and or assumptions about the deformation characteristics of the geo-materials. ANN use a supervised learning approach, learning from training examples, adjusting weights to reduce the error between the correct result and the result produced by the network. ANN endeavor to develop a general relationship between the inputs and outputs provided. Most of the statistical models based the data distribution and data type, but neural network doesn't depend upon the data type or data distribution it is force to find the relation between the parameters. A better understanding of the spatial variability of SOC is important for refining agricultural management practices and for improving sustainable land use. It provides a valuable base against which subsequent and future measurements can be evaluated. Limited research studies have been carried out to characterize and spatial estimation of SOC content using Hyperspectral satellite data utilizing multi-variate predictive modeling

### 2. Methodology

A flowchart of the methodology is shown in the Figure 1. Hyperion L1 Gst. format data for the month of May,



Figure 1. Flowchart of methodology.

(2009) were obtained and processed. Hyperspectral image processing is very important process of the study and various steps of image processing like bad band removal and atmospheric correction have been done to get the actual reflectance spectra with minimum noise. Atmospheric correction was achieved by using ENVI's Fast Line-of-sight Atmospheric Analysis of Spectral Hyper-cubes (FLAASH) module. FLAASH corrects wavelength in the visible through near-infrared and shortwave infrared region up to 3  $\mu$ m and it incorporate MODTRAN-4 radiation transfer code (FLAASH module user guide, 2006). The total study area was 189 km<sup>2</sup> and 65 soil samples (0 - 10 cm) were collected with field spectra using portable spectroradiometer (Analytical Spectral Devices-ASD) and GPS locations. Specroradiometer was calibrated (Figure 2) using white reference plate (BaSo4 plate) and a 100% reflectance line was available to the user to check the status of the instrument performance and vegetation spectra were tested to verify the performance. The calibrations were repeated every 20 minutes during the spectra collection.

The reflectance values of 65 locations were extracted from the atmospheric corrected Hyperion image using  $3 \times 3$  pixels matrix. The spectral angle mapper (SAM) was used to calculate the spectral similarity between reference spectra (Field and Lab) and image spectra. The spectral similarity between the image spectrum and reference spectrum can be expressed as an angle (Ø) between the two spectra for each channel. The standard range of SAM is 0 - 90 degree. The SAM values were rescaled between 0 to 1. Lower value indicated high similarity and higher SAM value indicates greater dissimilarity. ASD spectra were re-sampled to the Hyperion FWHM (Full width half maximum) and bandwidth. Field and laboratory spectra were evaluated with image spectra for spectral similarity. The soil samples were analysed in the lab to estimate total soil carbon and SOC of soil samples using TOC analyser. Spectral library were generated for field, lab and image spectra with different level of SOC. The highly sensitive reflectance bands found in this analysis were further used for training and prediction of SOC with ANN based predictive model. For ANN modelling, the computer software MatLab and the Neural Network Toolbox were used [5]. An ANN consists of a large number of highly interconnected processing elements, which simulate basic functions of biological neurons. **Figure 3** illustrates a simple one neuron model



Figure 2. Instrument calibration.



Figure 3. Basic artificial neural network structure.

within an artificial network, where an input vector is passed through the neuron for providing an output value. A multiple-layer system consists of input, hidden and output layers were used.

In this study, a multiple-layer feed-forward back-propagation network with three layers was used: input, hidden and output layer. This type of network generally provides better performances in comparison to other type of network [6]. Back-propagation training algorithms calculate error and adjust the weights of the various layers backwards from the output layer all the way back to input layer. This process will run upto minimum error. The number of neurons in the hidden layer was identified using trial-and-error approach as suggested by [7]. The number of neurons in the hidden layer is great importance, as less or too many neurons may cause over fitting problems [8]. Two transfer functions Tan-sigmoid (non linear) and linear function were selected for the hidden and output layers, respectively (Figure 4).

In this study Levenberg-Marquardt algorithm (Demuth & Beale, 1998) was used which provides a fast optimization. The data set was divided in three part training (60%), validation (20%) and testing (20%). To quantify performance of SOC-reflectance models based on ANN methods, various parameters between predicted values (Y') and independent reference measurements (Y) were calculated. The Root Mean Square Error (RMSE) (Equation (1)), coefficient of determination ( $\mathbb{R}^2$ ), Standard deviation of Prediction (Equation (2)) and Ratio of prediction to deviation (RPD) (Equation (3)), were calculated for model performance.

$$RMSE = \sqrt{\frac{\sum (Y' - Y)^2}{n}}$$
(1)

$$SDP = \left\{ \sum Y'^{2} - \left[ \left( \sum Y' \right)^{2} / n \right] / (n-1) \right\}^{1/2}$$
(2)

$$RPD = \frac{SDP}{\left(\sum (Y' - Y)^2 - \left[\sum (Y' - Y)^2 / n\right] / n - 1\right)^{1/2}}$$
(3)



# 3. Results and Discussion

## **3.1. Atmospheric Correction**

The result of atmospheric correction of Hyperion data in the form of changes in reflectance spectra of vegetation on the images before and after FLAASH atmospheric correction was presented in **Figure 5**. After FLAASH atmospheric correction, vegetation showed typical reflectance spectra curve with high reflectance at green band; low reflectance at red band; high reflectance in NIR band.

## 3.2. Laboratory Analysis of soil Samples for Soil Organic Carbon

Collected soil samples were analysed for several soil properties following standard chemical and instrumental analysis method. The histogram (**Figure 6**) is representing data set distribution of all measured SOC value in percentage. The SOC values ranged between 0.312% to 1.42% and mean 0.828% and standard deviation is 0.249%. Maximum soil sample belongs to the SOC range of 0.7% to 0.8% (22%) followed by range 0.8% to 0.9% (18%) and minimum in the range from 1.2% to 1.5% (total 6%). The soil organic matter (SOM) content varies from 0.66% to 2.45%.

#### **3.3. Soil Spectra Collection**

The Analytical Spectral Device (ASD) records spectra in 2151 bands at 1 nm band width. To compare the ground-measured soil spectra collected from ASD with that of atmospherically corrected image spectra, the ASD spectra were re-sampled to the Hyperion FWHM and bandwidth. A spectral library of re-sampled ASD spectra was created. The spectra from the ASD library were used as standard to compare image reflectance spectra extracted using the atmospheric correction of Hyperion satellite data. The collected soil spectra before and after re-sampling are illustrated in Figure 7 and Figure 8, respectively.

#### **3.4. Spectral Similarity Analysis**

SAM was estimated for image vs. field soil spectra and image vs. lab soil spectra for all soil sampling sites. The values varied from 0 to 90. The SAM values were rescaled between 0 and 1. SAM value of 0 indicates a perfect match and higher values indicate greater dissimilarity (Figure 9). Values for image vs. field spectra vary between 0.040 - 0.152 and for Image vs. Lab varies 0.047 to 0.153. SAM analysis data indicated that Image and field soil spectra are matching more than image and lab soil spectra. The results indicated that Hyperion data were still affected by atmospheric effect and field and laboratory soil spectra were least affected by atmosphere

#### **3.5. Identification of SOC Sensitive Bands**

Spectral signatures of materials are characterized by their reflectance or absorbance, as a function of wavelength in the electromagnetic spectrum. Fundamental features in reflectance spectra occur at energy levels that allow molecules to rise to higher vibration states. Absorption features in soil are the result of overlapping bands from different mineral components and organic matter [9]. Spectral bands have correlation coefficient value more than  $-0.80 \text{ R}^2$  were selected for prediction of SOC content with ANN multivariate predictive model. Based criteria, 30 image spectral bands having more than  $-0.80 \text{ R}^2$  and these bands are in electromagnetic spectrum

#### S. K. Tiwari et al.



After

Figure 5. Atmospheric correction of Hyperion data.





ranges from 762 nm to 1649 nm. Similar analysis has also been carried out for the field and lab soil spectral data sets. In case of field and lab data sets 45 spectral bands and 37 spectral bands, respectively were selected for further multivariate analysis. The higher correlation coefficients ( $R^2 >= -0.80$ ) of SOC against spectral reflectance as a function of wavelength vary between 550 nm to 1100nm for field and lab data sets. Similar observations were reported by [10] in the relationships between SOC content and narrow band reflectance in the various region of EMR. SOC is the component of soil organic matter (SOM) which contains biochemical constituents like chlorophyll, oil, cellulose, pectin, starch, lignin and humic acids which mainly influence the reflectance in the visible (400 - 700 nm) and NIR (700 - 1400 nm) region of EMR.







Figure 8. Re-sampled spectrum using FWHM and band width.



Figure 9. Normalized values of spectral angles.

#### 3.6. Artificial Neural Network Training and Modelling

#### 3.6.1. ANN Training

Whole data sets of image, field and lab were subset in three parts—training (60%), testing (20%) and validation (20%). Data were used to train the neural network. In each training stage the network performance was assessed using testing data set and the best performed network was taken for the whole data set simulation. Best performed network requires correct number of neurons in the hidden layer at the time of training. To find out cor-

rect number of hidden neurons, the overall criteria have taken from three methods given by Jeff Heaton in his book "Introduction to Neural Networks", chapter 5: Understanding back propagation.

- The number of hidden neurons should be between the size of input layer and the size of output layer.
- The number of hidden neurons should be 2/3 of input layer size plus the size of output layer.
- The number of hidden neurons should be less than the twice of input layer size.

In this study "Trial and Error approach" suggested by [7] to get the best number of hidden neuron. The trial has taken from 05 to 31 neurons; the 20 neurons in the hidden layer are the best to train the network with image data set in this study. Trial and Error approach works on the basis of R (Figure 10) and Mean Square Error (Figure 11) values and these two parameters were used to find the best number of neuron.

#### 3.6.2. Trained Network Structures and Performance

The ANN trained predictive models developed for each data set and used to estimate SOC content. The performance statistics of ANN trained predictive models are given in Table 1.

The outcome from image spectra and field spectra are more similar than lab spectra set and explanation is that both data set image and field reflectance are captured from the sensors are in natural condition and lab data set measured in control environment. ANN models were trained for all three data sets and which are satisfactory trained with high  $R^2$  value and good RPD. The ANN predictive modeling results indicate that Hyperion hyperspectral imagery can be suitably used for prediction of SOC content mapping over large area provided sufficient ground SOC content data available for training the ANN predictive model.



Figure 10. Selection of best number of hidden neurons based on R value.



Figure 11. Selection of best number of hidden neurons based on MSE.

#### Table 1. Trained network structures and performance.

Data type	Trained network structures and performance			
	Trained network	$\mathbb{R}^2$	RPD	RSME
Image scale	30-20-01	0.93	3.19	0.038
Field scale	45-32-01	0.92	3.17	0.029
Lab scale	37-26-01	0.95	3.61	0.021

#### 3.6.3. Spatial SOC Map Generated by ANN Model

The image reflectance data selected from the hyperspectral image were used in the ANN models previously trained with the image scale data sets to map the spatial distribution of SOC content (Figure 12). A mean filter (window size  $3 \times 3$ ) was applied to the maps of estimated SOC values as a post-classification technique to reduce high frequency variations and to smooth the boundaries. Standing vegetation and settlement masks were applied for generating final map of SOC content distribution. The ANN model estimated SOC content varies 0.08% to 1.9%.

#### 4. Conclusion

ANN predictive modeling technique with combination of hyperspectral performed well and carried out a good result for all three data sets, image, field and lab scales having almost similar values of performance statistics  $R^2$  0.93, 0.92 and 0.95 and RPD 3.19, 3.17 and 3.16 and RMSE 0.0376, 0.0294 and 0.0212 respectively. The Results indicated that ANN predictive modeling and hyperspectral data could be suitably used for prediction of SOC content mapping. This study is also suggesting increasing the research up to field level soil organic carbon prediction using high spatial resolution hyperspectral remote sensing data that would be helpful for precision farming. In their study, [11] also suggested that the spectral unmixing needed to extract the soil spectra from the mixed hyperspectral data when the spatial resolution was 30 m. There is a potential for the use of hyperspectral remote sensing for prediction of soil organic carbon and the use of this technology will facilitate the implementa-



Figure 12. SOC distribution map predicted by ANN model.

tion of digital soil mapping. It is also suggested that researches in SOC prediction from satellite hyperspectral data may be followed through from the German hyperspectral satellite (EnMAP) data which may have a better Signal to Noise Ratio.

#### **References**

- Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L. and Sommer, S. (2009) Using Imaging Spectroscopy to Study Soil Properties. *Remote Sensing of Environment*, **113**, S38-S55. http://dx.doi.org/10.1016/j.rse.2008.09.019
- Goetz, A.F. and and Vane, H.G. (1985) Imaging Spectrometry for Earth Remote Sensing. *Science*, 228, 1147-1153. <u>http://dx.doi.org/10.1126/science.228.4704.1147</u>
- [3] Ben-Dor, E., Patkin, K., Banin, A. and Karnieli, A. (2002) Mapping of Several Soil Properties Using DAIS-7915 Hyperspectral Scanner Data: A Case Study over Soils in Israel. *International Journal of Remote Sensing*, 23, 1043-1062. <u>http://dx.doi.org/10.1080/01431160010006962</u>
- [4] Cloutis, E.A. (1996) Hyperspectral Geological Remote Sensing: Evaluation of Analytical Techniques. International Journal of Remote Sensing, 17, 2215-2242. <u>http://dx.doi.org/10.1080/01431169608948770</u>
- [5] Demuth, H. and Beale, M. (1998) Neural Network Toolbox for Use with MATLAB, User's Guide, Version 3. The MathWorks Inc., Natick.
- [6] Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*, 2, 359-366. <u>http://dx.doi.org/10.1016/0893-6080(89)90020-8</u>
- [7] Chang, Y.M., Chang, L.C. and Chang, F.J. (2004) Comparison of Static Feedforward and Dynamic-Feedback Neural Networks for Rainfall Runoff Modeling. *Journal of Hydrology*, 290, 297-311. <u>http://dx.doi.org/10.1016/j.jhydrol.2003.12.033</u>
- [8] Huang, W. and Foo, S. (2002) Neural Network Modelling of Salinity Variation in Apalachicola River. *Water Research*, 36, 356-362. <u>http://dx.doi.org/10.1016/S0043-1354(01)00195-6</u>
- Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G. and Vergo, N. (1990) High Spectral Resolution Reflectance Spectroscopy of Minerals: *Journal of Geophysical Research*, 95, 12653-12680.
- [10] Bartholomeus, H.M., Schaepman, E.M., Kooistra, L., Stevens, A., Hoogmoed, B.W. and Spaargaren, O.S.P. (2008) Spectral Reflectance Based Indices for Soil Organic Carbon Quantification. *Geoderma*, 145, 28-36. http://dx.doi.org/10.1016/j.geoderma.2008.01.010
- [11] Gomez, C., Rossel, R.A.V. and McBratney, A.B. (2008) Soil Organic Carbon Prediction by Hyperspectral Remote Sensing and Field Vis-NIR Spectroscopy: An Australian Case Study. *Geoderma*, **146**, 403-411.