15. Toes, M. J., Jones, A. L. and Prescott, L., Drug interactions with paracetamol. *Am. J. Ther.*, 2005, **12**, 56–66.
16. Chahiyan, H., Gharib, F. and Farajtabar, A., Thermodynamic studies on solubility and protonation constant of acetaminophen at different ionic strengths and various temperatures. *J. Mol. Liq.*, 2014, **199**, 137–142.
17. Faraji, M., Farajtabar, A. and Gharib, F., Determination of water–ethanol mixtures autoprotolysis constants and solvent effect. *J. Appl. Chem. Res.*, 2009, **9**, 7–12.
18. Farajtabar, A. and Gharib, F., Solvent effect on protonation constants of salicylic acidin mixed aqueous organic solutions of DMSO. *Monatsh. Chem.*, 2010, **141**, 381–386.
19. Farjtabar, A. and Gharib, F., Solvent effect on protonation constants of 5, 10, 15, 20-tetrakis(4-sulfonatophenyl)porphyrin in different aqueous solutions of methanol and ethanol. *J. Sol. Chem.*, 2010, **39**, 231–244.
20. Gran, G., Determination of the equivalence point in potentiometric titrations, Part II. *Analyst*, 1952, **77**, 661–671.
21. Nernst, W., Dielectric constant and chemical equilibrium. *Z. Phys. Chem.*, 1984, **13**, 531–536.
22. Beck, M. T. and Nagypal, I., *Chemistry of Complex Equilibria*, Ellis Harwood, New York, USA, 1990.
23. Kilic, E. and Aslan, N., Determination of autoprotolysis constants of water–organic solvent mixtures by potentiometry. *Microchim. Acta*, 2005, **151**, 89–92.
24. Farajtabar, A., Naderi, F. and Gharib, F., Autoprotolysis in water/methanol/NaCl ternary systems. *J. Serb. Chem. Soc.*, 2013, **78**, 1561–1567.
25. Maleki, N., Haghighi, B. and Safavi, A., Evaluation of formation constants, molar absorptivities of metal complexes, and protonation constants of acids by nonlinear curve fitting using Microsoft Excel Solver and user-defined function. *Microchem. J.*, 1999, **62**, 229–236.
26. ACD/Labs 6.00 C. ACD pKa DB Advanced Chemistry Development Inc., Ontario, Canada.
27. Wan, H., Holmen, A. G., Wang, Y., Lindberg, W., Englund, M., Nagard, M. B. and Thompson, R. A., High-throughput screening of pK_a values of pharmaceuticals by pressure-assisted capillary electrophoresis and mass spectrometry. *Rapid Commun. Mass Spectrom.*, 2003, **17**, 2639–2648.
28. Barbosa, J., Barron, D., Beltran, J. L. and Buti S., On the role of solvent in acid–base equilibria of diuretics in acetonitrile–water mixed solvents. *Talanta*, 1998, **45**, 817–827.
29. Barbosa, J., Toro, I. and Sanz-Nebot, V., Acid–base behaviour of tripeptides in solvents used in liquid chromatography. Correlation between pK values and solvatochromic parameters of acetonitrile–water mixtures. *Anal. Chim. Acta*, 1997, **347**, 295–304.
30. Puranik, S. M., Kumbharkhane, A. C. and Mehrota, S. C., The static permittivity of binary mixtures using an improved bruggeman model. *J. Mol. Liq.*, 1994, **59**, 173–177.
31. Taft, R. W., Abboud, J. L. M and Kamlet, M. J., Linear solvation energy relationships. An analysis of Swain's solvent 'acity' and 'basity' scales. *J. Org. Chem.*, 1984, **49**, 2001–2005.
32. Kamlet, M. J., Abboud, J. L. M., Abraham, M. H. and Taft, R. W., Linear solvation energy relationships. A comprehensive collection of the solvatochromic parameters, .pi.*, .alpha., and .beta., and some methods for simplifying the generalized solvatochromic equation. *J. Org. Chem.*, 1983, **48**, 2877–2887.
33. Marcus, Y., The use of chemical probes for the characterization of solvent mixtures. Part 2. Aqueous mixtures. *J. Chem. Soc. Perkin Trans.*, 1994, **2**, 1751–1758.
34. Farajtabar, A. and Gharib, F., Spectral analysis of naringenin deprotonation in aqueous ethanol solutions. *Chem. Pap.*, 2013, **67**, 538–545.

Received 8 September 2018; revised accepted 16 March 2020

doi: 10.18520/cs/v119/i2/374-377

Soil organic carbon prediction using visible–near infrared reflectance spectroscopy employing artificial neural network modelling

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

Agriculture and Soils Department, Indian Institute of Remote Sensing (ISRO), 4 Kalidas Road, Dehradun 248 001, India

Visible–near infrared (VNIR) spectroscopy is a relatively fast and cost-effective analytical technique for estimating soil organic carbon (SOC). The present study was undertaken for predicting SOC 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, India, and analysed for SOC. The reflectance spectra of soil samples was measured using a 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 utilized for predictive modelling of SOC using the ANN model. This model could predict SOC values with R^2 of 0.92 and MSE value of 0.24, indicating that this technique can be used to predict SOC in a spatial domain when coupled with high-resolution hyperspectral satellite/airborne data.

Keywords: Artificial neural network model, reflectance spectroscopy, soil organic carbon, visible and near infrared region.

SOIL organic carbon (SOC) plays a fundamental role in determining the physical, chemical and biological properties of the soil. It is beneficial for maintaining soil productivity, water-holding capacity as well as carbon sequestration for alleviating the ill-effects of greenhouse gases and thus climate change[1–3]. SOC holds about 4.5 times the amount of the biotic carbon pool and 3.3 times the amount of atmospheric carbon pool[4]. Thus reliable estimation of SOC is vital for understanding the human-induced effects on the global carbon cycle and associated climate change[5]. Conventional methods for SOC determination in soil laboratories are costly, time-consuming and may be environmentally hazardous[6]. Thus there is an urgent need for the development of fast, accurate and non-destructive methods (thus reducing the number of soil chemical analyses) for SOC estimation, which will help in generating high-resolution soil property maps of large areas at modest costs.

*For correspondence. (e-mail: justin@iirs.gov.in)

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

Artificial neural network (ANN), an effective data mining technique, was found to give better results in predicting SOC content, clay content and soil pH than other regression and statistical techniques[16]. This may be due to its strong connectivity of networks that could possibly draw out an effective relationship between soil properties and spectral features. Also, the soil map derived using ANN was found to be more accurate than all other conventional methods of soil mapping[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 the network channel leading to inputs, which can be further used reversely with known inputs to derive unknown outputs[18]. Multilayer feedback propagation algorithms facilitate ANN training in a more effective way[19], and these were utilized for estimating soil nutrients like phosphorus using DEM-derived terrain attributes, which proved the suitability of ANN in quantitative prediction of soil nutrients[20]. Therefore, the present study was undertaken to verify the possibility of employing ANN modelling for predicting SOC using VNIR reflectance spectroscopy.

The study was conducted in a watershed located in the mid-Himalayan region of Mandi district, Himachal Pradesh (lat. 32°4′35.04″–32°1′3.8964″N and long. 76°39′49.60″–76°44′15.84″E), India, covering a total geographical area of 1000 ha (10 km^2). Nearly 80% of the watershed is comprised of agricultural fields where rice and wheat are the major crops grown. Figure 1 shows the complete methodology adopted in the study. To properly represent soil spatial variability in the watershed and ensure unbiased and precise sampling, grid sampling approach was used for soil sample collection, with a grid size of 250 m × 250 m 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 for the generation of reflectance spectra in the laboratory. Preprocessed and 0.2 mm sieved soil samples were used for the estimation of SOC (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 that the soils are devoid of soluble salts/salinity. 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 SOC content, with majority of the values greater than 0.5%. The SOC values ranged from 0.26 to 2.71 g kg^{-1} soil, with a mean value of 1.36 g kg^{-1} soil. The soils belonged to moderate to well drainage class.

The surface reflectance spectra of the preprocessed soil samples in the wavelength range 350–2500 nm were

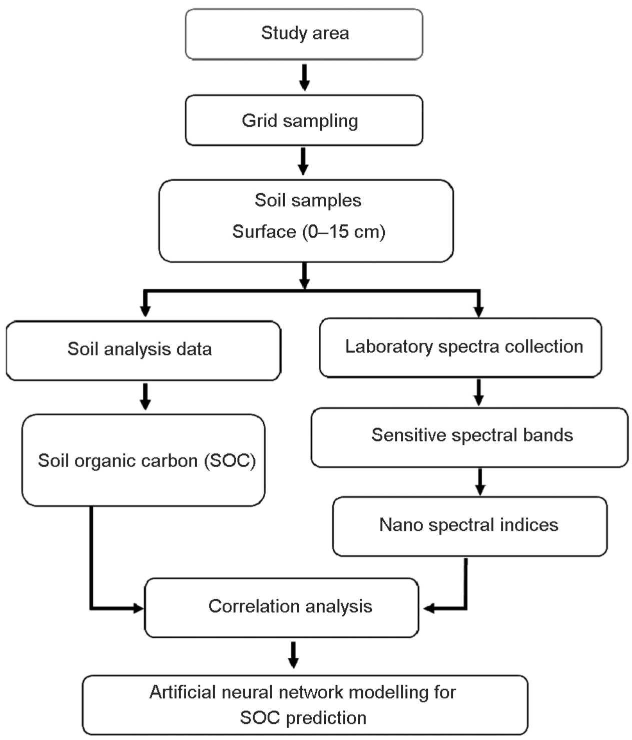

Figure 1. Overall methodology adopted in the present study.

measured in the laboratory using a spectroradiometer (ASD FieldSpec Pro Spectroradiometer), under controlled dark-room conditions. Precautions were taken to avoid stray light as well as dark current generated within the instrument. Also, the instrument was calibrated using white reference panel at the start as well as after every five successive reflectance measurements. 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 generating spectral libraries. The spectra collected at 1 nm interval were also resampled to 5 nm interval using the spectral resampling tool in ENVI 5.0 for further analysis.

Sensitive spectral bands specific to SOC were studied through a literature survey[15,21–23]. Various bands identified for their sensitivity to SOC were 400, 441, 520, 907, 960, 1100, 1720, 1744, 1870, 2052, 2180 and 2309 nm. 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 the most sensitive bands within each of the three regions. For rest of the spectra, only bands identified from the 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) (eqs 1–5). These are image colour indices, which are sensitive to bare soil properties[12,24]. Since variations in SOC content influence soil colour, these indices were chosen for the present study. The indices make use of blue, green and red wavelength regions in the visible range; hence single nano-bands in these regions having highest correlation coefficient (r) were selected for developing the indices. Reflectance values 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 SOC. The bands in the shortwave IR region, especially 2180 and 2309 nm, were also found to have higher correlation with SOC values (Table 1). Owing to the significance of the NIR region in predicting SOC, bands in this region along with the first derivative, second derivative and logarithmic values, reciprocal of logarithmic values and reciprocals of derivatives were also subjected to correlation analysis.

$$\mathrm{BI} = \mathrm{Sqrt}[(\mathrm{Blue}^2 + \mathrm{Green}^2 + \mathrm{Red}^2)/3], \tag{1}$$

$$\mathrm{CI} = (\mathrm{Red} - \mathrm{Green})/(\mathrm{Red} + \mathrm{Green}), \tag{2}$$

$$\mathrm{HI} = (2 * \mathrm{Red} - \mathrm{Green} - \mathrm{Blue})/(\mathrm{Green} - \mathrm{Blue}), \tag{3}$$

$$\mathrm{RI} = \mathrm{Red}^2 + (\mathrm{Blue} * \mathrm{Green}^3), \tag{4}$$

$$\mathrm{SI} = (\mathrm{Red} - \mathrm{Blue})/(\mathrm{Red} + \mathrm{Blue}), \tag{5}$$

where 451, 520 and 690 nm wavelengths represent the blue, green and red bands respectively.

The ANN modelling for SOC prediction was executed using Matlab software Ver 2015. The nano-spectral indices that showed good correlation with SOC were used for analysis (Table 2). In addition, the reciprocal of first derivative of reflectance value at wavelength 2309 nm and reflectance value of the 2180 nm band showed high correlation coefficient and were also taken as input to the ANN. The data preparation and subsequent database generation were done in Excel and the neural network was executed using Neural Network Tool in Matlab. The entire soil dataset was separated into two datasets – calibration (50 nos) and validation (25 nos) datasets using random numbers. During the model training and development phase, the calibration set was further internally separated into 3 sub-datasets, i.e. training (68%), validation (16%) and testing (16%) datasets.

A multilayer 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 nonlinear nature. A tan–sigmoid function known to proximate nonlinear interactions between inputs and outputs was chosen as the transfer function for the hidden and output layers[25]. Overall network was trained using the Levenberg–Marquardt algorithm[26] and its performance was evaluated using R values from regression plots. The developed ANN model was

Table 1. Correlation coefficient (r^2) of spectral bands with soil organic carbon (SOC)

Sensitive bands	r^2
451	0.49
520	0.40
690	0.52
907	0.37
1720	0.37
2180	0.77
2309	0.67

Table 2. Correlation coefficient of spectral indices with SOC

	Spectral indices				
	Brightness index	Colouration index	Hue index	Redness index	Saturation index
SOC	0.31	0.21	0.29	0.35	0.37

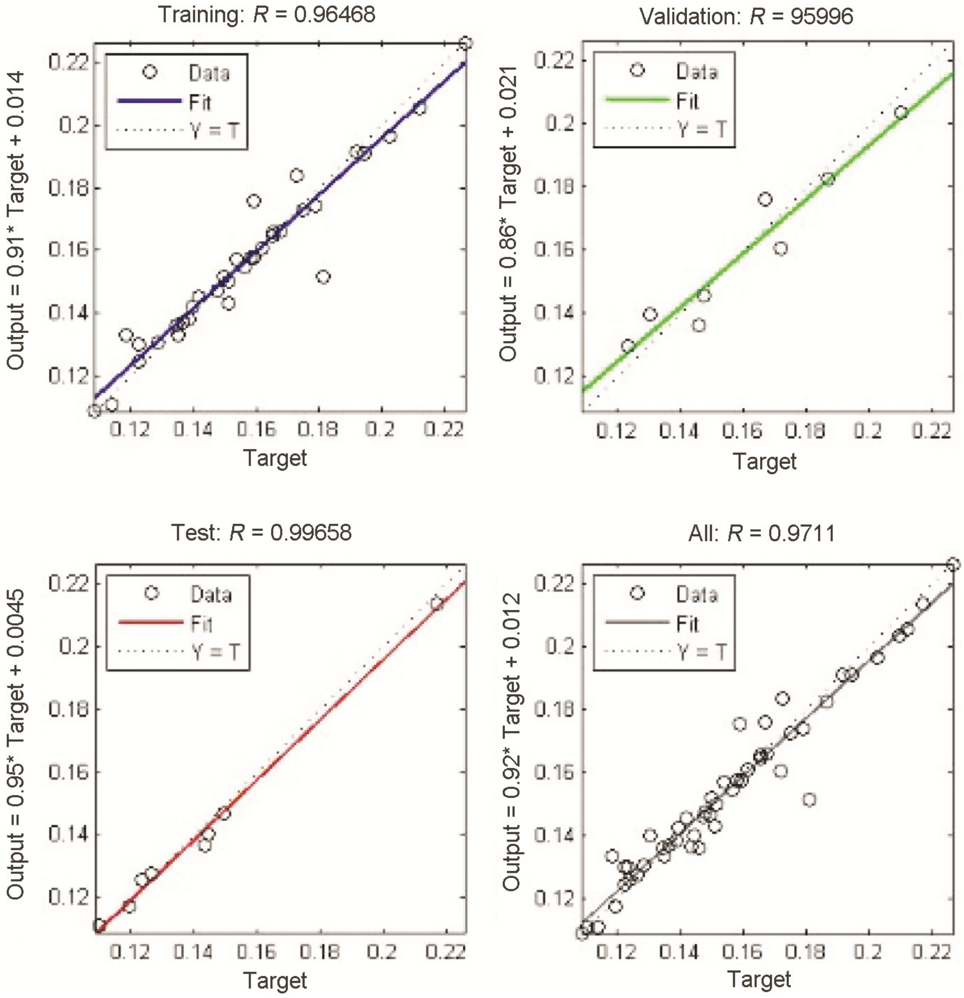

Figure 2. Regression plot of the accepted artificial neural network model.

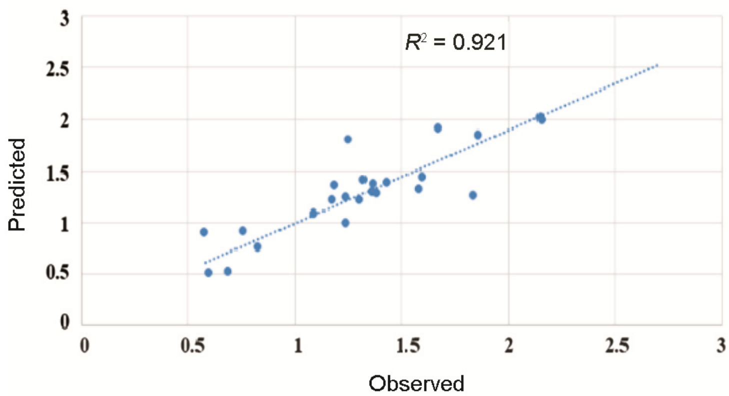

Figure 3. Observed versus predicted plot of soil organic carbon values.

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 (R^2) and MSE (mean square error) values.

Based on the results of Pearson correlation test, seven variables (CI, BI, HI, SI, RI, R_{2180} and reciprocal of the first derivative of R_{2309}) were selected for predictive modelling of SOC using ANN. Thus, the designed ANN network had seven input layers corresponding to the various predictive variables. Based on the trial and error method suggested by Chiang *et al.*[27] the 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 (Figure 2). Network training using the data was done until *R*-values of all the three sub datasets were greater than 0.9 or close to 1. The well-calibrated model was employed for the prediction of SOC values using the validation dataset. Using the well-trained ANN model, we were able to precisely predict SOC values with an R^2 of 0.92 and MSE of 0.24 (Figure 3).

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

1. Prescott, C. E. and Maynard, D. G. R., Humus in northern forests: friend or foe? *For. Ecol. Manage.*, 2000, **133**, 23–36.
2. Munson, S. A. and Carey, A. E., Organic matter sources and transport in an agriculturally dominated temperate watershed. *Appl. Geochem.*, 2004, **19**, 1111–1121.
3. Six, J. and Paustian, K., Aggregate-associated soil organic matter as an ecosystem property and a measurement tool. *Soil Biol. Biochem.*, 2014, **68**, A4–A9.
4. Lal, R., Soil carbon sequestration to mitigate climate change. *Geoderma*, 2004, **123**, 1–22.
5. Kumar, S. and Lal, R., Mapping the organic carbon stocks of surface soils using local spatial interpolator. *J. Environ. Monit.*, 2011, **13**, 3128–3135.
6. Araujo, S. R., Dematte, J. A. M., Franceschini, M. H. D., Rizzo, R., Stenberg, B. and Wetterlind, J., Improving the predictive performance of a national vis–NIR spectroscopic library by comparing clustering data transformation, and data-mining calibration techniques. Annals of XVI Brazilian Remote Sensing Symposium – SBSR, Foz do Iguacu, Brazil, 13–18 April 2013.
7. Williams, P. C. and Norris, K. H., Variables affecting near infrared spectroscopic analysis. In *NIR Technology in the Agricultural and Food Industries* (eds Williams, P. and Norris, K.), American Association of Cereal Chemists, St Paul, MN, USA, 2001, pp. 171–185.
8. Chang, C. W. and Laird, D. A., Near-infrared reflectance spectroscopy analysis of soil C and N. *Soil Sci.*, 2002, **167**, 110–116.
9. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M. and Wetterlind, J., Visible and near infrared spectroscopy in soil science. *Adv. Agron.*, 2010, **107**, 163–215.
10. Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J. and Skjemstad, J. O., Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 2006, **131**, 59–75.
11. Gomez, C., Rossel, R. A. V. and McBratney, A. B., Soil organic carbon prediction by hyperspectral remote sensing and field vis–NIR spectroscopy: an Australian case study. *Geoderma*, 2008, **146**, 403–411.
12. Zhuo, L., Yaolin, L., Jian, W. and Wang, J., Quantitative mapping of soil organic material using field spectrometer and hyperspectral remote sensing. *Int. Arch. Photogramm., Remote Sensing Spat. Inf. Sci.*, 2008, **27**, 901–906.
13. Ben-Dor, E., Chabrillat, S., Dematte, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L. and Sommer, S., Using imaging spectroscopy to study soil properties. *Remote Sensing Environ.*, 2009, **113**, 38–55.
14. Cecillon, L. *et al.*, Assessment and monitoring of soil quality using near infrared reflectance spectroscopy (NIRS). *Eur. J. Soil Sci.*, 2009, **60**, 770–784.
15. Gmur, S., Vogt, D., Zabowski, D. and Moskal, L. M., Hyperspectral analysis of soil nitrogen, carbon, carbonate, and organic matter using regression trees. *Sensors*, 2012, **12**, 10639–10658.
16. Viscarra Rossel, R. A. and Behrens, T., Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 2010, **158**, 46–54.
17. Zhu, A. X., Mapping soil landscape as spatial continua: the neural network approach. *Water Resour. Res.*, 2000, **36**, 663–677.
18. Huang, Y., Lan, Y., Thomson, S. J., Fang, A., Hoffmann, W. C. and Lacey, R. E., Development of soft computing and applications in agricultural and biological engineering. *Comput. Electron. Agric.*, 2010, **71**, 107–127.
19. Bagheri, B. M. *et al.*, Digital soil mapping using artificial neural networks and terrain-related attributes. *Pedosphere*, 2015, **25**(4), 580–591.
20. Keshavarzi, A., Sarmadian, F., Omran, E. S. E. and Iqbal, M., A neural network model for estimating soil phosphorus using terrain analysis. *Egypt. J. Remote Sensing Space Sci.*, 2015, **18**, 127–135.
21. Daninnel, K. W., Tripathi, N. K. and Honda, K., Artificial neural network analysis of laboratory and *in situ* spectra for the estimation of macronutrients in soil of Lop Buri (Thailand). *Aust. J. Soil Res.*, 2003, **41**, 47–59.
22. Shepherd, K. D. and Walsh, M. G., Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.*, 2002, **66**, 988–998.
23. Dalal, R. C. and Henry, R. J., Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Sci. Soc. Am. J.*, 1986, **50**, 120–123.
24. Lagacherie, P., McBratney, A. B. and Voltz. M., In *Digital Soil Mapping: An Introductory Perspective*, Elsevier, Amsterdam, The Netherlands, 2007, pp. 389–399.
25. Haykin, S., *Neural Networks and Learning Machines* (*Third Edition*), Prentice Hall, New Jersey, USA, 1999.
26. Demuth, H. and Beale, M., Neural Network toolbox, for use with MATLAB. The Math Works, Inc, Massachusetts, USA, 2004.
27. Chiang, Y. M., Chang, L. C. and Chang, F. J., Comparison of static-feed forward and dynamic-feedback neural networks for rainfall–runoff modelling. *J. Hydrol.*, 2004, **290**, 297–311.

ACKNOWLEDGEMENTS. We thank the Indian Space Research Organisation for financial support under the Earth Observation Applications Mission project (ISRO/DOS) on 'Mountain ecosystem processes and services'.

Received 30 May 2017; accepted 1 May 2019

doi: 10.18520/cs/v119/i2/377-381