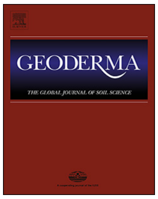
Geoderma 330 (2018) 254–263

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma



Laboratory-based hyperspectral image analysis for predicting soil carbon, nitrogen and their isotopic compositions

Iman Tahmasbian[a,*], Zhihong Xu[a,*], Sue Boyd[a], Jun Zhou[b], Roya Esmaeilani[c], Rongxiao Che[a], Shahla Hosseini Bai[d]

[a] Environmental Futures Research Institute, School of Environment and Science, Griffith University, Nathan, Brisbane, Queensland 4111, Australia
[b] School of Information and Communication Technology, Griffith University, Nathan, Brisbane, Queensland 4111, Australia
[c] Faculty of Computing, University Technology Malaysia 81300, Skudai, Johor, Malaysia
[d] Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, QLD, 4558, Australia

ARTICLE INFO

Handling Editor: A.B. McBratney

Keywords:
Native forest
Chemometrics
Multivariate analysis
PLSR, VNIR, $\delta^{13}C$, $\delta^{15}N$

ABSTRACT

The common methods of determining soil carbon (C), nitrogen (N) and their isotopic compositions ($\delta^{13}C$ and $\delta^{15}N$) are expensive and time-consuming. Therefore, alternative low-cost and rapid methods are sought to address this issue. This study aimed to investigate the potential of hyperspectral image analysis to predict soil total carbon (TC), total nitrogen (TN), $\delta^{13}C$ and $\delta^{15}N$. Hyperspectral images were captured from 96 ground soil samples using a laboratory-based visible to near-infrared (VNIR) hyperspectral camera in the spectral range of 400–1000 nm. Partial least squares regression (PLSR) models were developed to correlate the values of TC, TN, $\delta^{13}C$ and $\delta^{15}N$, obtained from isotope ratio mass spectrometry method, with their spectral reflectance. The developed models provided acceptable predictions with high coefficient of determination (R_c^2) and low root mean square error ($RMSE_c$) of calibration set for TC ($R_c^2 = 0.82$; $RMSE_c = 1.08\%$), TN ($R_c^2 = 0.87$; $RMSE_c = 0.02\%$), $\delta^{13}C$ ($R_c^2 = 0.82$; $RMSE_c = 0.27‰$) and $\delta^{15}N$ ($R_c^2 = 0.90$; $RMSE_c = 0.29‰$). The prediction abilities of the models were then evaluated using the spectra of an external test set (24 samples). The models provided excellent predictions with high R_t^2 and ratio of performance to deviation (RPD) of test set for TC ($R_t^2 = 0.76$; RPD = 2.02), TN ($R_t^2 = 0.86$; RPD = 2.08), $\delta^{13}C$ ($R_t^2 = 0.80$; RPD = 2.00) and $\delta^{15}N$ ($R_t^2 = 0.81$; RPD = 1.94). The results indicated that the laboratory-based hyperspectral image analysis has the potential to predict soil TC, TN, $\delta^{13}C$ and $\delta^{15}N$.

1. Introduction

The biogeochemical cycles of carbon (C) and nitrogen (N) in terrestrial ecosystems have received considerable attention because of their significant role in climate change (Fu et al., 2010; Zaehle et al., 2010). The C and N pools are key indicators of soil quality and it is therefore important that they are monitored in terrestrial ecosystems (Che et al., 2018; Darby et al., 2016; Hosseini Bai et al., 2015a; Nguyen et al., 2017). The soil natural isotopic abundance of C and N ($\delta^{13}C$ and $\delta^{15}N$) have also been widely used as powerful tools to gain insight into C and N cycling in the terrestrial ecosystems (Bai et al., 2017; Che et al., 2017; Fujiyoshi et al., 2017; Hosseini Bai et al., 2017; Hosseini Bai et al., 2015b; Wang et al., 2015). The isotopic compositions, vary over the time due to different processes such as microbial activities, plant uptake and leaching (Farquhar et al., 1989; Högberg, 1997), and need to be analyse frequently. Conventional methods, such as isotope ratio mass spectroscopy, used for analysing soil total C (TC), total N (TN), $\delta^{13}C$ and $\delta^{15}N$, however, are time-consuming and expensive when applied to a large number of samples (Jiang et al., 2017). Hence, the alternative inexpensive and rapid methods are sought to predict soil TC, TN, $\delta^{13}C$ and $\delta^{15}N$.

The visible to near infrared (VNIR) airborne/laboratory-based hyperspectral imaging technique (400–2500 nm), which uses the advantages of both conventional imaging and VNIR spectroscopy, is a physical, rapid, reproducible and low-cost alternative to conventional methods of soil chemical analysis (Gmur et al., 2012; Jung et al., 2015; Stevens et al., 2008; Vaudour et al., 2016). More importantly, hyperspectral imaging permits a quantitative assessment of different soil properties from a single measurement (Ben-Dor and Banin, 1995a; Jia et al., 2016; Viscarra Rossel et al., 2006; Xu et al., 2016). The spatial dimension provided by airborne/laboratory-based hyperspectral imaging technology gives this method an advantage over the VNIR spectroscopy (Manley, 2014). For example, the spatial dimensions are used to obtain the chemical image/map for identification and localisation of

---

* Corresponding authors at: Environmental Future Research Institute, Griffith University, Nathan, Queensland 4111, Australia.
E-mail addresses: iman.tahmasbian@griffithuni.edu.au (I. Tahmasbian), zhihong.xu@griffith.edu.au (Z. Xu).

https://doi.org/10.1016/j.geoderma.2018.06.008
Received 22 September 2017; Received in revised form 20 April 2018; Accepted 11 June 2018
0016-7061/

chemical compounds in the samples with non-homogenous surfaces (Manley, 2014). However, for homogenised samples such as ground soils, the spatial information are used by averaging the several pixels in the spatial dimensions of hyperspectral images (Manley, 2014). Using the average of the pixels increases the signal to noise ratio (Manley, 2014), and therefore improves the accuracy and reproducibility of predictions compared with the single points used in VNIR spectroscopy. One of the main disadvantages of hyperspectral imaging compared with VNIR spectroscopy include the high cost of hyperspectral sensors, especially when wavelengths between 1000 nm to 2500 nm are required (Manley, 2014). Fast computers and substantial data storage capacity are also required due to the large size of hyperspectral images (Manley, 2014). Despite the higher accuracy and capability of laboratory-based sensors to measure both surface and deep soils, they can cover smaller areas compared with airborne hyperspectral systems.

The hyperspectral imaging technique utilises the reflectance measured on different materials using airborne or laboratory-based hyperspectral sensors (Bai et al., 2018; Ben-Dor et al., 1997; Gama et al., 2018; Grahn and Geladi, 2007; Manley, 2014). Multivariate analysis is then used to correlate the spectral reflectance data with their corresponding reference values of the targeted variables measured using the reference chemical methods *e.g.,* isotope ratio mass spectrometry (Manley, 2014; Wijewardane et al., 2016). However, different preprocessing techniques such as wavelength selection and spectral data transformation are usually applied to reduce the collinearity of the spectral data and the impact of artefacts, such as random noise and light scattering (Tahmasbian et al., 2018). The data pre-processing usually increase the signal to noise ratio in the spectral data prior to performing the multivariate analyses (Manley, 2014; Wei et al., 2017). First derivative, second derivative, multiplicative scatter correction (MSC), orthogonal signal correction (OSC), standard normal variate (SNV) and detrending (DT) are the most common transformation techniques applied to hyperspectral reflectance data in wide range of studies (Fearn, 2000; Kamruzzaman et al., 2016b; Lin et al., 2016; Rinnan et al., 2009; Siripatrawan et al., 2011). The best-transformed data are then used to develop the multivariate models using the selected important wavelengths. This step employs an appropriate chemometric algorithm such as partial least squares regression (PLSR).

PLSR is the most commonly used multivariate linear model for chemometrics (Coûteaux et al., 2003; Jiang et al., 2017; Mouazen et al., 2007). PLSR enables us to analyse small data sets with multiple variables, which are strongly correlated and noisy (Höskuldsson, 1988; Wold et al., 1984; Wold et al., 2001). PLSR finds a few linear combinations (latent variables) of the original X-values and uses only those linear combinations in the regression equation. This leads to discarding the irrelevant and unstable information to use the most relevant X-variation for the regression analysis (Höskuldsson, 1988; Wold et al., 1984; Wold et al., 2001). The developed PLSR models are then used to predict the target variables in new samples.

The VNIR airborne hyperspectral imaging and VNIR spectroscopy analyses have been used to analyse soil C and N in a diverse type of soils (Barthès et al., 2006; Brunet et al., 2007; Denis et al., 2014; Stevens et al., 2008; Udelhoven et al., 2003; Wei et al., 2017). For example, soil organic and total C, as well as soil TN, have been analysed using airborne hyperspectral imaging and VNIR spectroscopy in the spectral region of 350–2500 nm; these studies provided results with a wide range of accuracies, R^2 from 0.35 to over 0.90 (Cozzolino et al., 2013; Gmur et al., 2012; Gopal et al., 2015; Jiang et al., 2017; Stevens et al., 2008). Limited information, however, is available on using laboratory-based hyperspectral images for the prediction of TC and TN in soil samples. To the best of our knowledge, there is also no report on using the laboratory-based hyperspectral imaging sensors to predict soil $\delta^{13}C$ and $\delta^{15}N$. The main objective of this study was to assess the potential of using the laboratory-based hyperspectral imaging systems to predict TC, TN, $\delta^{13}C$ and $\delta^{15}N$ in soil samples. The models developed in this study will be used to predict the long-term dynamics of soil TC, TN, $\delta^{13}C$ and $\delta^{15}N$ in real-time after prescribed burning in Toohey Forest, East Queensland, Australia.

The techniques used in airborne hyperspectral imaging and VNIR spectroscopy follow the same principles as those of laboratory-based hyperspectral imaging. Therefore, it was hypothesised that the concentration of TC and TN in soil samples could be predicted using the spectra extracted from laboratory-based hyperspectral images in the spectral region of 400–1000 nm. Furthermore, the featureless (without a specific spectral response) properties in soil samples can also be predicted using their internal correlation with other properties which possess specific spectral responses (Ben-Dor and Banin, 1995b; Ben-Dor et al., 2002). Soil TC, TN and their isotopic compositions, $\delta^{13}C$ and $\delta^{15}N$ have shown to be correlated (Abdullah, 2016; Hobbie et al., 2000; West et al., 2009). Therefore, it was also hypothesised that $\delta^{13}C$ and $\delta^{15}N$ can be predicted using laboratory-based hyperspectral reflectance in the visible to near infrared regions (400–1000 nm) due to their correlation with TC and TN, respectively. We additionally hypothesised that important wavelengths selection would improve the accuracy of the developed models by reducing the collinearity of the spectral data.

2. Materials and methods

2.1. Study site description

Soil samples used in this experiment were collected from Toohey Forest, a suburban native forest located in south-eastern Queensland, Australia (27°30′S, 153°E) (Hosseini Bai et al., 2014a). Toohey Forest is located in a subtropical region, with average minimum temperature (1939–2016) of 14.4 °C, average maximum temperature (1939–2016) of 26.3 °C and mean annual precipitation (1969–2016) of 1163.5 mm.

Geological units of Toohey Forest were diverse, and include argillites, shales, sandstones and shales of the Woogaroo subgroup; soft sedimentary rocks of the Tingalpa formation (Ipswich coal measures); cherts and arsenite of the Neranleigh-Fernvale beds; and sandstones of the Sunnybank formation (Catterall et al., 2001a). The most forest soils were thin lithosols, with a thickness varying between 60 cm and 1 m (Abdullah, 2016). The lithosols contained both stony components and loam to silty loam, which may be rich in humus (Catterall et al., 2001a). The soil pH varied between 4.4 and 5.5. The forest was subjected to prescribed burning since 1993 with different intervals for different sites (Butler et al., 2016; Catterall et al., 2001b).

2.2. Site establishment, soil sample collection and chemical analyses

Two study sites were established randomly in different sectors of Toohey Forest, included Site 1 (27° 32′ 39.42̋″ S, 153° 03' 6.81″ E) and Site 2 (27° 32′ 45.33″ S, 153° 03′ 14.72″ E). The two study sites covered an area of 4000 m^2. Each site had four randomly established circular plots (radius of 12.62 m, area of 500 m^2). Each plot was divided into four sampling areas (*ca.* 125 m^2) for soil collection. The soils samples were collected separately from the depths of 0–5 and 5–10 cm of the sampling areas to increase the variability of data and make a wide-range data set. The soil samples were collected from different locations within the sampling areas at each sampling time.

A total of 120 soil samples were collected from the study sites. Soil samples from Site 1 (56 samples) were collected seasonally (every three months), within an 18-month period, starting in November 2014. Soil samples from Site 2 (64 samples) were collected monthly in the first six months and then seasonally (every three months) for the next six months, within a 12-month period, starting in Jun 2015.

The collected soil samples were carried to the laboratory in separate plastic zipper bags and were air-dried in the Griffith University Soil Laboratory. The air-dried soil samples were sieved with a 2-mm sieve and ground into fine powder. Approximately 20 mg of ground soil samples were transferred into tin capsules for TC, TN, $\delta^{13}C$ and $\delta^{15}N$ analyses using an isotope ratio mass spectrometer (Hosseini Bai et al.,