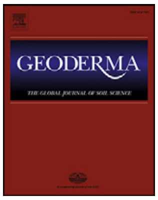
Geoderma 239–240 (2015) 229–239

Contents lists available at ScienceDirect

Geoderma

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



Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy

N.M. Knox [a,b,c], S. Grunwald [c,*], M.L. McDowell [d,1], G.L. Bruland [d,2], D.B. Myers [c,3], W.G. Harris [c]

[a] Earth Observation Division, South African National Space Agency (SANSA), PO Box 484, Silverton 0127, South Africa
[b] University of KwaZulu-Natal, School of Agricultural, Earth and Environmental Sciences, Private Bag X01, Scottsville 3209, Pietermaritzburg, South Africa
[c] Soil and Water Science Department, University of Florida, 2181 McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA
[d] Natural Resources and Environmental Management Department, University of Hawai'i Mānoa, 1910 East–West Rd, Sherman 101, Honolulu, HI 96822, USA

ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form 15 October 2014
Accepted 26 October 2014
Available online 9 November 2014

Keywords:
Partial least squares regression
Random forest regression
Total carbon
Soil organic carbon
Hydrolysable carbon
Recalcitrant carbon
Chemometric modelling

ABSTRACT

Accurate assessment of soil carbon fractions would provide valuable contributions towards monitoring in ecological observatories, assessment of disturbance impacts, global climate and land use change. The majority of chemometric modelling studies have focused on measuring only total soil carbon (C), with only a few evaluating individual soil C pools. Analysis of pools allows for a more detailed picture of ecosystem processes, specifically decomposition and accretion of C in soils. This study evaluated the potential of the visible near infrared (VNIR), mid infrared (MIR) and a combined VNIR–MIR spectral region to estimate and predict soil C fractions. Partial least squares regression (PLSR) and random forest (RF) ensemble tree regression models were used to estimate four different soil C fractions. The soil C fractions analysed included total — (TC), organic — (SOC), recalcitrant — (RC) and hydrolysable carbon (HC). The sample set contained 1014 soil samples collected across the state of Florida, USA. Laboratory analysis revealed the wide range of total and organic C values, from 1 to 523 $g \cdot kg^{-1}$, with only about 10% of the samples containing inorganic C which was therefore omitted from the study. Both PLSR and RF modelling were shown to be effective in modelling all soil C fractions, with as much as 94–96% of the variation in the TC, SOC and RC pools, and 86% of HC being explained by the models. Although both PLSR and RF models were successful in modelling C fractions, RF models appear to target the physical properties linked to the property being analysed, and may therefore be the better modelling method to use when generalising to new areas. This study demonstrates that diffuse reflectance spectroscopy is an effective method for non-destructive analysis of soil C fractions, and through the use of RF modelling a spectral range between 2000 and 6000 nm should suffice to model these soil C fractions.

1. Introduction

Soil carbon (C) sequestration has been targeted by the Intergovernmental Panel on Climate Change (IPCC) as one of the main means to mitigate rise in atmospheric C levels (Bellon-Maurel and McBratney, 2011; Metz et al., 2007). Total soil C can be separated into different soil C fractions based on their turnover rates. Whether the soil is a source or a sink of C depends on the balance between the C recalcitrance (chemical stability of the soils) and evolution of C cycling processes (i.e. the impact of decomposition, accretion, etc.). There has been much debate about the impact of climate warming on soil C (Bellamy et al., 2005; Craine and Gelderman, 2011; Davidson and Janssens, 2006; Fierer et al., 2005; Thornley and Cannell, 2001; Trumbore et al., 1996) and the uncertainty involved in monitoring soil C across large regions is still substantial (Grunwald et al., 2011). There is in addition a need to assess the impact of land use shifts and disturbances (e.g., wildfire) on soils; however a rapid, robust and effective means to monitor changes in soil C at regional, continental and global scales is lacking. As a response to these needs research on spectral-based methods has exploded over the last decade and introduced new avenues including soil C models informed by proximal sensors, such as visible near infrared (VNIR) and mid infrared (MIR) diffuse reflectance spectroscopy (DRS). Importantly, these methods have been shown to yield comparable results to traditional time-consuming and more costly laboratory methods (Bellon-Maurel and McBratney, 2011).

Diffuse reflectance spectroscopy in the VNIR spectral range has been used widely to characterize soil organic carbon (SOC) and soil total

---

Abbreviations: DRS, diffuse reflectance spectroscopy; HC, hydrolysable carbon; MIR, mid infrared; PLSR, partial least squares regression; RC, recalcitrant carbon; RF, random forest regression; RMSE, root mean square error; RPD, residual prediction deviation; SOC, soil organic carbon; SWIR, short wave infrared; TC, total carbon; VNIR, visible near infrared.

[*] Corresponding author.
E-mail address: sabgru@ufl.edu (S. Grunwald).

[1] Present Address: Scitor Corporation, 385 Moffett Park Dr, Sunnyvale, CA 94089, USA.

[2] Present Address: Principia College, Biology and Natural Resources Department, Elsah, IL 62028, USA.

[3] Present Address: Division of Plant Sciences, University of Missouri, 214 Waters Hall, Columbia, MO 65211, USA.

http://dx.doi.org/10.1016/j.geoderma.2014.10.019
0016-7061/

carbon (TC) (Brown, 2007; Ge et al., 2011; Sarkhot et al., 2011; Shepherd and Walsh, 2002; Udelhoven et al., 2003; Vasques et al., 2008, 2009, 2010). Soil OC models derived from VNIR spectra differed widely with studies covering 8.1–19.8 g kg^{-1} (Viscarra Rossel et al., 2006), 6.1–33.0 g kg^{-1} (Reeves et al., 2002), 2.3–55.8 g kg^{-1} (Shepherd and Walsh, 2002), 1.3–285.8 g kg^{-1} (Chang et al., 2001), and 0.1–536.8 g kg^{-1} (Brown et al., 2006) soil C suggesting a more universal applicability of models the wider their C content range. Mid-infrared spectroscopy has been demonstrated to predict SOC and TC, often with better accuracy than VNIR derived models. For example, McDowell et al. (2012a) predicted TC from VNIR and MIR spectral data on Hawaiian soils that are highly diverse and complex in terms of their mineralogy and soil texture. In the study by McDowell et al. (2012b) it was demonstrated that TC predictions derived from VNIR and from MIR spectra yielded robust results with R^2 values of 0.94 or greater, root mean square prediction error (RMSE) values ranged from 2.28%–3.08%, and residual prediction deviation (RPD) of 3.38 or more. In their study the models derived from VNIR and MIR spectra were contrasted, with both spectral regions not being combined to form single spectra. Other studies, such as by McCarty et al. (2002), Reeves et al. (2006), and Reeves (2010) have demonstrated the success of MIR to predict soil C.

The majority of these studies have focused on measuring only total soil C, with only a few evaluating individual soil C pools. Yet these pools allow a more detailed picture of ecosystem processes, specifically decomposition and accretion of C in soils. Thus, pools (i.e., chemically extracted soil C fractions) have unique capabilities to reflect the evolution of soil C as a result of various imposed stressors (e.g., climate and land use change). According to Belay-Tedla et al. (2009) C and nitrogen (N) fluxes are largely controlled by the small but highly bio-reactive, labile pools of these elements in terrestrial soils, while long-term C and N storage is determined by the long-lived recalcitrant fractions. In their study they found that recalcitrant and total C and N contents were not significantly affected by warming, however, labile and microbial biomass C fractions showed significant interactions with warming.

Labile C, although often a small fraction of SOC, significantly affects heterotrophic respiration at short timescales because of its rapid decomposition (Gu et al., 2004). In a study by Bradford et al. (2007) it was shown that there are significant interactive effects when the formation and decomposition responses of different soil C fractions are considered. Hence, the total amount of soil C stocks was dependent on the rate of labile soil C, N and phosphorus (P) inputs to soils. Labile soil C has also been shown to respond to effects in management (Biederbeck et al., 1994). Proximal sensing technology has the potential to infer not only on total soil C content, but also on soil C fractions. However, this research is still in its infancy. For example, Vasques et al. (2009) used VNIR spectra to estimate various soil C fractions and mineralizable C in a subtropical region dominated by sand-rich soils. Also Sarkhot et al. (2011) estimated SOC, inorganic C, and labile C in a small floodplain area on clay-rich soils in Texas. More work needs to still be done in the field of chemometric modelling of C fractions encompassing diverse soil types to determine whether spectral modelling would be a robust method for predicting of soil C fractions.

Recently, Bellon-Maurel and McBratney (2011) critically reviewed both NIR and MIR spectroscopic techniques for soil C studies. Their review highlighted that MIR spectroscopy yielded better (10–40% improvement) models than models developed from NIR spectra. Of the studies reviewed none of them compared NIR and MIR spectroscopy on the same soil samples indicating a critical research gap which is addressed in this study. From the studies they reviewed, they applied either multiple linear regression (MLR) or partial least squares regression (PLSR) for their chemometric models. The simpler MLR models were shown to perform comparably to the PLSR models; however, both methods showed shortcomings in terms of wavelengths selected. Judicious selection of wavelengths is imperative when applying a model to new areas where the soil C range might be similar, but the soil characteristics differ from those where the model was built. Wavelength selection is likely to be an important contributor to increasing the bias in results when the validation dataset is acquired from an “out of field” source. Both methods were found to be biased when applied to new soil samples with different compositions. Recently, the use of non-linear modelling methods, such as ensemble tree regression modelling, has been shown to improve model performances in terms of output error metrics (Vasques et al., 2009, 2010; Viscarra Rossel and Behrens, 2010; Vohland et al., 2011).

The focus of this study was to determine the potential of both the VNIR and MIR spectral regions to estimate soil C fractions. Partial least squares regression (PLSR) and random forest (RF) regression models were derived to estimate soil C fractions. These methods were evaluated firstly to determine which method could best predict soil C fractions, and secondly it was determined if either method produced output that is attributed to the physics of soil C properties. We assume that if the most important model factors (i.e., the importance factors for RF and loadings for PLSR) can be attributed to physical soil C properties the models would be robust if generalised and applied to other areas in the future. The soil C fractions focussed on in this study include total C (TC), soil organic C (SOC), recalcitrant C (RC), and hydrolysable C (HC). VNIR and MIR chemometric models were derived from a wide selection of soil samples collected over the state of Florida, USA. The study objectives were to determine the predictive capabilities of the VNIR, MIR and combined VNIR–MIR spectral datasets for predicting soil C fractions (TC, SOC, RC, and HC). Within this overall objective we evaluated the following sub-objectives:

1. Determine which spectral region performs best.
2. Evaluate which modelling method (PLSR or RF) performs best based on validation datasets.
3. Identify if there is a best method in terms of data preparation for C fraction chemometric analysis.

2. Methods

2.1. Field sampling

A total of 1014 sample sites were located over the state of Florida using a stratified random sampling design. The strata were defined based upon soil suborder data obtained from the Soil Data Mart — Soil Survey Geographic Database (SSURGO) (Natural Resources Conservation Service — http://soildatamart.nrcs.usda.gov/), and a reclassified land use/cover map derived from the 2003 Florida Vegetation and Land Cover Data map prepared by the Florida Fish and Wildlife Conservation Commission, Tallahassee, Florida. Dominant soils in the state are Spodosols (32%), Entisols (22%), Ultisols (19%), Alfisols (13%), Histosols (11%), and Mollisols and Inceptisols together occupy <3% of the state (Vasques et al., 2010).

The soil samples were collected over a period of 1.25 years between 2008 and 2009. Sample sites were located using a differential global positioning system. At each site four 20 × 5.8 cm surface soil cores were collected from within a 2 m diameter area. These four cores were bulked and cooled in the field and then transported and processed in the lab. In addition to the sampled soils, data was collected of the soil suborder and land use practice found at each site.

2.2. Laboratory methods

The bulk samples were air-dried and sieved to retrieve the fine earth fraction (<2 mm), thoroughly mixed and then stored. Subsamples were taken from these dried, sieved and mixed bulk samples for the laboratory analysis. A portion of these subsamples were ball milled for use in some of the laboratory procedures outlined below.