

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Geoderma

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

Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy

Meryl L. McDowell^{a,*}, Gregory L. Bruland^{a,1}, Jonathan L. Deenik^b, Sabine Grunwald^c, Nichola M. Knox^{c,2}

^a Natural Resources and Environmental Management Department, University of Hawai'i Mānoa, 1910 East-west Rd, Sherman 101, Honolulu, HI 96822, USA

^b Tropical Plant and Soil Science Department, University of Hawai'i Mānoa, 3190 Maile Way, Honolulu, HI 96822, USA

^c Soil and Water Science Department, University of Florida, 2169 McCarty Hall, P.O. Box 110290, Gainesville, FL 32611-0290, USA

ARTICLE INFO

Article history:

Received 22 November 2011

Received in revised form 7 June 2012

Accepted 11 June 2012

Available online xxxx

Keywords:

Diffuse reflectance spectroscopy

Agriculture

Hawaii

Mid-infrared

Soil carbon

Visible near-infrared

ABSTRACT

Accurate assessment of total carbon (C_t) content is important for fertility and nutrient management of soils, as well as for carbon sequestration studies. The non-destructive analysis of soils by diffuse reflectance spectroscopy (DRS) is a potential supplement or alternative to the traditional time-consuming and costly combustion method of C_t analysis, especially in spatial or temporal studies where sample numbers are large. This alternative technique has been utilized in several locations in the United States and elsewhere, but has not been tested on the unique and diverse tropical soils in Hawaii. This study investigated the feasibility of DRS for C_t prediction of Hawaiian agricultural soils by creating visible, near-infrared (VNIR) and mid-infrared (MIR) spectral libraries and developing chemometric models with partial least squares regression (PLSR) and random forests (RF) ensemble tree regression. The sample set contained 305 soils from across the five main Hawaiian Islands, representing 10 soil orders and over 100 soil series. The C_t of these samples measured by dry combustion ranged from <1% to 56%, greater than that of most previously published studies. The VNIR spectra of the soils commonly exhibited features associated with OH^- and H_2O , iron oxides, phyllosilicates, and organic molecules. The numerous features in the MIR spectra can be attributed to OH^- , organic molecules, and a variety of silicate minerals, such as phyllosilicates and quartz. Excellent results were obtained from both PLSR models using VNIR and MIR spectra. The models had R^2 values of 0.95/0.94 (VNIR/MIR), root mean squared error (RMSE) values of 2.80%/3.08%, residual prediction deviation (RPD) values of 4.25/3.91, and ratio of performance to inter-quartile distance (RPIQ) values of 3.66/3.38. The RF VNIR and MIR models were also of high quality, with $R^2 = 0.95/0.96$ (VNIR/MIR), $\text{RMSE} = 2.82\%/2.28\%$, $\text{RPD} = 4.27/5.28$, and $\text{RPIQ} = 3.68/4.56$. An evaluation of the different wavelength ranges and chemometric methods determined that for this sample set all were capable of robust predictions of C_t . These results indicate that DRS is an alternative technique for analysis of C_t in Hawaiian soils.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Soil total carbon (C_t) is a major component of fertility in subtropical and tropical soils (Oenema et al., 2006; Paustian et al., 1997; Tiessen et al., 1994). Loss of C_t in these soils due to cultivation has been shown to lead to declines in soil fertility, water retention, and structure, as well as being associated with the formation of surface crusts and soil compaction (Feller et al., 1996). Soil carbon is also the driving force of biological activity, serving as the primary source of energy and nutrients for many soil organisms (Craswell and Lefroy, 2001). Improving the facility to quantify soil C_t is a critical component of managing soils sustainably.

Traditional methods for quantifying soil C_t , such as chromate oxidation (Walkley and Black, 1934) and combustion (Allison et al., 1965)

are expensive and slow (McCarty et al. 2002; Nelson and Sommers, 1996; Watson et al., 2000). Chromate oxidation also generates toxic wastes that must be properly disposed (Craswell and Lefroy, 2001). Another technique, loss-on-ignition (Ball, 1964), though relatively cheap and rapid, has been shown to be inaccurate because certain mineral fractions are also decomposed by heating at the high temperatures required by this method (Lal et al., 2001; Nelson and Sommers, 1996). This is especially problematic with kaolinite, poorly crystalline aluminosilicates (i.e., allophane and imogolite), and iron oxyhydroxides (i.e., ferrihydrite) that are common in Hawaiian soils. Furthermore, there is increasing interest in the spatial and temporal variability of C_t , and the disadvantages of these traditional laboratory analysis methods are compounded by the large number of samples required for accurate assessment in these studies (McCarty and Reeves, 2006).

Thus, there is a growing demand for new techniques to measure C_t that are faster and cheaper than traditional methods (McBratney et al., 2006; Shepherd and Walsh, 2002). Visible, near-infrared (VNIR) and mid-infrared (MIR) diffuse reflectance spectroscopy (DRS) are promising technologies that have the potential to revolutionize soil monitoring

* Corresponding author. Tel.: +1 405 209 0200; fax: +1 808 956 6539.

E-mail address: mcdowell@hawaii.edu (M.L. McDowell).

¹ Present address: Principia College, 1 Maybeck Pl, Elsah, IL 62028-9799, USA.

² Present address: South African National Space Agency (SANSA), Earth Observation, P.O. Box 484, Silverton, 0127, South Africa.

by allowing for samples to be scanned rapidly, inexpensively, and non-destructively (Janik et al., 1998; Viscarra Rossel et al., 2006). Soil spectra are related to laboratory-based measurements of C_t with chemometric modeling, allowing for the prediction of C_t for measured soil spectra. Thus, DRS enables the collection and processing of greater numbers of soil samples in order to better characterize how changes in management practices affect fertility and soil carbon sequestration at increasingly finer spatial and temporal scales.

Reflectance spectroscopy has been used to characterize minerals and soils for several decades (e.g., Hunt, 1977; Stoner and Baumgardner, 1981). Once combined with chemometrics, the use of DRS for predictive analysis of soil properties began to grow in popularity with soil scientists such as Krishnan et al. (1980), Dalal and Henry (1986), Williams (1987), Janik and Skjemstad (1995), and Janik et al. (1995). Many of the recent studies for C_t obtained excellent results, with $R^2 > 0.80$ (Mouazen et al., 2010; Sarkhot et al., 2011; Vasques et al., 2008, 2009; Viscarra Rossel and Behrens, 2010). Several studies have focused on soil carbon, but others have investigated a variety of other soil properties as well, such as pH, cation exchange capacity, sand, silt, and clay fractions, total N, total P, and organic matter (e.g., Cohen et al., 2005; Shepherd and Walsh, 2002; Terhoeven-Urselmans et al., 2010; Viscarra Rossel et al., 2006). In addition to the standard linear regression techniques such as partial least squares regression (PLSR), chemometric techniques have grown to include neural networks (Dardenne et al., 2000; Janik et al., 2009), regression tree analysis (Cohen et al., 2005; Vasques et al., 2008), and multivariate adaptive regression splines (Shepherd and Walsh, 2002). The majority of studies focus on specific geographic regions, which have included many areas in the United States, such as Maryland (Reeves et al., 2001, 2002), and Florida (Vasques et al., 2008, 2009), as well as Australia (Dalal and Henry, 1986; Janik and Skjemstad, 1995; Janik et al., 1995; Viscarra Rossel et al., 2006) and several countries in Africa (McCarty et al., 2010; Shepherd and Walsh, 2002) and Europe (Mouazen et al., 2010; Sørensen and Dalsgaard, 2005; Stenberg, 2010; Udelhoven et al., 2003). Brown et al. (2006), Sankey et al. (2008), and Terhoeven-Urselmans et al. (2010) have also explored the idea of global predictive models that cover a wide soil attribute space.

The notably diverse and unique soils of the Hawaiian Islands (Deenik and McClellan, 2007; Gavenda et al., 1998) have not been included in previous studies, however. Not only is the parent material of different ages on each island, but within one island there is dramatic variation in factors such as elevation and weather patterns that affect rainfall (Gavenda et al., 1998). Circumstances like these have created high spatial variability in soil types. Additionally, Hawaii has a range of soil types that have received little attention in past soil prediction studies. For example, Andisols are the most extensive soil order in Hawaii but are not as common in other agricultural regions. These soils are derived from volcanic tephra and include large proportions of amorphous and poorly crystalline materials, as well as organic matter, giving this soil type different characteristics compared to the more common agricultural soils in continental settings. Soils with high iron oxide abundances are also prevalent in Hawaii, such as the highly weathered Oxisols and Ultisols on the islands of Kauai and Oahu. The spectra from soils of this type have been shown to differ from those of less-weathered soils (Stoner and Baumgardner, 1981). The diversity in Hawaiian soils increases the need for soil studies as well as the number of samples required, which also increases the expense in terms of money and time, especially with standard soil characterization procedures at the U.S. National Soil Survey Center costing approximately \$2500 per pedon with a processing time of 6–12 months (Brown et al., 2006). Visible, near-infrared and MIR DRS offer a new way to process soil samples more rapidly and for a lower cost, helping to better characterize the variability of soil C_t that exists across the Hawaiian Islands.

Additionally, more convenient methods of C_t analysis are needed particularly in Hawaii because Hawaiian soils have the potential to be important in C sequestration efforts. Approximately $\frac{1}{4}$ of the global

C_t pool is located in the tropics (Craswell and Lefroy, 2001). In these tropical areas, conversion of natural ecosystems to agroecosystems has been shown to deplete the soil C_t pool by 75% or more (Lal, 2004). Some tropical soils have lost as much as 20–80 tons of $C\ ha^{-1}$, most of which is released as carbon dioxide (CO_2) into the atmosphere (Lal, 2004). As many of these tropical and subtropical soils are now relatively C-depleted, they represent a potential C sink if part of the lost C can be sequestered (Paustian et al., 1997). Soil C sequestration depends on the land use, management, soil texture, soil profile characteristics, and climate (Lal, 2004). Evaluating change in soil C requires multiple C_t analyses, which becomes more feasible with the use of DRS.

The VNIR and MIR DRS method of C_t analysis must be evaluated for Hawaiian soils before it can be implemented in important C_t studies with these unique soils. Thus, the objectives of this study were as follows: (1) to employ VNIR and MIR DRS to characterize Hawaiian soils representing a wide range in C content; (2) to investigate and model relationships between VNIR and MIR DRS spectra and laboratory-measured soil C_t ; and (3) to compare the predictive ability of partial least squares chemometric models with random forests ensemble tree approaches for soil C_t prediction. This research also tests the overarching hypothesis that fundamental wavelength regions in the MIR range will produce better predictions of C_t than combination or overtone regions in the VNIR range.

2. Materials and methods

2.1. Sample collection

The sample set consisted of 305 soil samples from the islands of Kauai, Oahu, Molokai, Maui, and Hawaii. We gathered 216 of the samples from those stored in the soil archive at the Natural Resources Conservation Service (NRCS) National Soil Survey Center in Lincoln, Nebraska. One hundred and forty five of these soil samples were from the surface A or O horizon. The corresponding subsurface horizons of 17 of these samples made up the other 71 of the 216 NRCS archive samples; these were included to describe the variability in C_t from surface through subsoil. These NRCS archive samples were collected in Hawaii between 1981 and 2007. The other 89 soils in the sample set were collected in the summer and fall of 2010. These samples were taken from the upper 0–20 cm of the soil profile with a stainless-steel piston corer and circular plastic sleeve insert of 4.8 cm diameter. The sampling locations for the 2010 field campaign were chosen to supplement the spatial distribution and soil orders of the NRCS archive samples, as well as by the availability of and permission to access the land. Fig. 1 illustrates the distribution of the full set of soil samples across the five main Hawaiian Islands.

Much of the diversity of Hawaiian soils is captured in the sample set. In total, over 100 series from 10 soil orders are represented (Fig. 1). The set includes soils from more than 25 crop types, such as sugar cane (*Saccharum* spp.), pineapple (*Ananas* spp.), dryland taro (*Alocasia* spp.), coffee (*Coffea* spp.), corn (*Zea* spp.), and watermelon (*Citrullus* spp.), and both conventional and organic management practices. Though the study focuses on agricultural soils, we have also included some soils from other land uses, such as pasture and agroforestry.

2.2. Sample preparation

The NRCS archive soils were air dried and sieved to 2 mm before storage in the archive. The newly-collected Hawaii field samples were oven dried at 105 °C for 24 h and then passed through a 2 mm sieve. To test comparability between the drying processes of the samples, a selection of the NRCS archive air dried samples was subjected to the oven drying process as well, and no differences in spectral properties were observed in the air dried and oven dried subsamples for either the VNIR or MIR regions. A portion of each <2 mm sample was ball milled

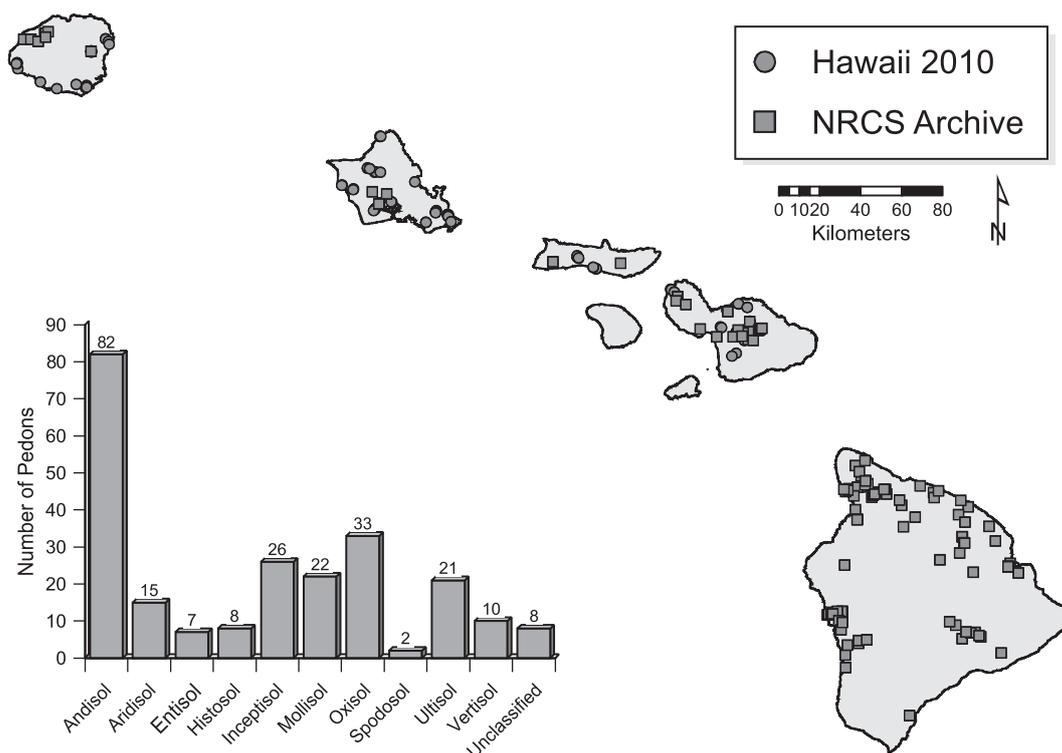


Fig. 1. Distribution of collection sites for soil samples in the data set from the Natural Resources Conservation Service (NRCS) archive and Hawaii 2010 field campaign, along with the number of samples per soil order.

to <250 μm . A subset of the 2 mm sieved samples was used for VNIR DRS analysis and subsets of the ball milled samples were used for MIR DRS analysis and C_t analysis by dry combustion.

2.3. Laboratory analysis

2.3.1. Traditional total carbon analysis

Total carbon values were determined for the soil samples through dry combustion using a LECO CN2000™ combustion gas analyzer (AOAC International 1997) at the Agricultural Diagnostic Services Center (ADSC) at the University of Hawaii Mānoa. Many of the NRCS archive samples were analyzed for C_t by dry combustion at the time of their sampling, and those were not included in this step. However, we selected a small batch of the samples with previous C_t measurements ranging from 0.79 to 54.66%C to be re-analyzed with the rest of the samples at the ADSC to provide a cross-check of values obtained from the different laboratories. We found differences in C_t between the previous and current results of <1% for samples with C_t <10% and <7% for samples with C_t >10%.

2.3.2. Visible, near-infrared diffuse reflectance spectroscopy

We collected VNIR diffuse reflectance spectra with an AgriSpec VNIR spectrometer and muglight light source (Analytical Spectral Devices, Inc., Boulder, CO). A 3.2 cm diameter sample cup with transparent bottom is placed on top of the muglight, which directs light upward to encounter the sample. The AgriSpec contains three detectors which provide a spectral range of 350–2500 nm. The sampling interval is 1 nm and the spectral resolution varies from 3 nm (at 700 nm) to 10 nm (at 1400 nm). A Spectralon (Labsphere, North Sutton, NH) white reference was scanned as a reference spectrum at the beginning of each session and every 30 min or less thereafter. Each spectrum was the result of 30 internal scans. Three spectra were collected for each sample, with the sample cup rotated 20° between each measurement. The three replicate spectra were then averaged to produce one spectrum per sample to be used for data analysis.

Because of a slight offset in the reflectance between the VNIR and SWIR1 detectors, we removed the small region of 990–1010 nm in the spectrum from further analysis.

2.3.3. Mid-infrared diffuse reflectance spectroscopy

Mid-infrared diffuse reflectance spectra were collected from ball milled subsamples in neat form using a Scimitar 2000 FTIR spectrometer (Varian, Inc., now Agilent Technologies, Santa Clara, CA) outfitted with a diffuse reflectance infrared Fourier transform (DRIFT) accessory and 6 mm diameter sample cup. The instrument measures reflectance over the range of 400–6000 cm^{-1} with a sampling interval of 2 cm^{-1} and spectral resolution of 4 cm^{-1} . Each spectrum was the result of 32 internal scans. A separate scan of KBr powder provided a background spectrum that was subtracted from each sample scan to correct for atmosphere and instrument effects. A new background spectrum was collected at the beginning of a session and between every seven samples. Despite this correction, two narrow regions of the spectrum were affected by fluctuating atmospheric conditions and continued to exhibit residual atmospheric features. Therefore, the regions of 1350–1419 cm^{-1} and 2281–2449 cm^{-1} were excluded from further analysis and interpretation.

2.4. Data analysis

Two types of regression methods were evaluated for the prediction of C_t from spectral data. Partial least squares regression is a standard regression technique that has been used extensively in chemometrics. Random forests (RF) ensemble tree regression is a newer, non-parametric classification and regression tree approach (Breiman, 2001; Liaw and Wiener, 2002). The spectral range included in these regression analyses was reduced slightly to exclude any increase in noise at the limits of the range. The range of 425–2450 nm was used for the VNIR spectra and 489–5300 cm^{-1} for the MIR spectra. No samples were considered outliers or excluded from the analyses. The sample set was randomly divided into two groups; 70% of the samples were used for

Table 1

Descriptive statistics of soil total carbon (%) in the data sets used for partial least squares regression (PLSR) and random forests (RF) ensemble tree regression.

	<i>n</i> ^a	Mean	Std. Dev.	Minimum	Maximum	Median
Full dataset	307	10.75	13.89	0.15	55.29	4.59
Calibration	215	11.54	14.54	0.24	55.29	4.78
Validation	92	8.89	12.12	0.15	53.63	3.73

^a Number of samples.

calibration of the models and 30% were used for validation of the models. The Levene's test for equality of variances, the Student's *t*-test for equality of means, and the Kolmogorov–Smirnov test for equality of probability distributions were applied to ensure that C_t values of the calibration and validation subsets were representative of the full sample set.

2.4.1. Partial least squares regression

Various pre-processing transformations to the spectra were explored in the initial PLSR modeling to find the best way to highlight the C_t information contained in the spectra from our data set. These pre-treatments included mean normalization, smoothing, derivative, absorbance (i.e., $\log(1/R)$ where *R* is reflectance) and all combinations between them. All PLSR models were developed using mean-centered data.

Partial least squares regression was conducted with the Unscrambler X Software (CAMO Software Inc., Woodbridge, NJ). We determined the optimal number of factors to use in the final models from the explained variance of the initial models, with the goal of maximizing the explained variance but minimizing the likelihood of over-fitting. The quality of the models was evaluated through parameters such as the coefficient of determination (R^2), root mean squared error (RMSE), residual prediction deviation (RPD) (Williams, 1987), and ratio of performance to interquartile distance (RPIQ) (Bellon-Maurel et al., 2010). The latter two parameters were defined as $RPD = SD/SEP$ and $RPIQ = IQ/SEP$, where *SD* is the standard deviation of the validation set, *SEP* is the standard error of prediction, and *IQ* is the interquartile distance of the validation set ($IQ = Q3 - Q1$). For the purposes of this study, we consider one model to be better than another if it has higher R^2 , RPD, and RPIQ values, and a lower RMSE value.

2.4.2. Random forests ensemble tree regression

The RF models were tested with the raw spectral data and twice again with two pre-processing transformations. For one set of models the transformation applied was the same used in the final PLSR models.

The other pre-processing transformation evaluated was the Norris Gap derivative for both the VNIR and MIR spectra; Vasques et al. (2008) found this pretreatment more effective than others for tree-based regression methods.

Random forests ensemble tree regression was implemented using the randomForest package (Liaw and Wiener, 2002) in the R statistical language (R Development Core Team, 2011). Multiple models were created to test different model parameters and combinations that reduced model error, which included varying the number of trees grown from 500, 1000, and 1500, and the number of random variables used to decide the best split at each node from 1/3 the number of predictors, half that number, and twice that number. The RF model produces its own measures of error, the mean of squared residuals and % variance explained, from the aggregation of the prediction of the samples left out of the subset used to grow each tree. However, we also calculated the R^2 , RMSE, RPD, and RPIQ for the separate prediction of our reserved validation set to facilitate comparison with the PLSR output.

3. Results and discussion

3.1. Laboratory analysis of total carbon

The C_t values of the samples in the full data set ranged from <1% to 56% (Table 1). The distribution over this range was weighted toward lower values of C_t , as might be expected for agricultural soils, but the inclusion of many high C_t values reflects the large variation in soil types, climatic and environmental conditions, and land uses in the Hawaiian Islands. Numerous previous studies of C_t with DRS have been limited to the lower end of the range. McCarty et al. (2002), Sarkhot et al. (2011), and Reeves et al. (2001, 2002) all contained soils with C_t of 10.4% or less; the C_t range of soils used in Vasques et al. (2008) was slightly higher at up to ~28%. Studies of organic carbon have used a range similar to these previous C_t studies (e.g., McCarty et al., 2010; Mouazen et al., 2010; Vasques et al., 2009; Viscarra Rossel and Behrens, 2010), with the exception of Vasques et al. (2010), which contained soils with organic carbon values as high as ~57%.

3.2. Qualitative description of spectra

Fig. 2a demonstrates the variability of the VNIR spectra in the sample set with several examples of individual soil spectra. The overall shape of the VNIR spectra were generally similar, with most displaying some degree of the steep red slope between ~400 and 700 nm characteristic of Fe^{3+} , and many having broad minima near ~660 nm and/or

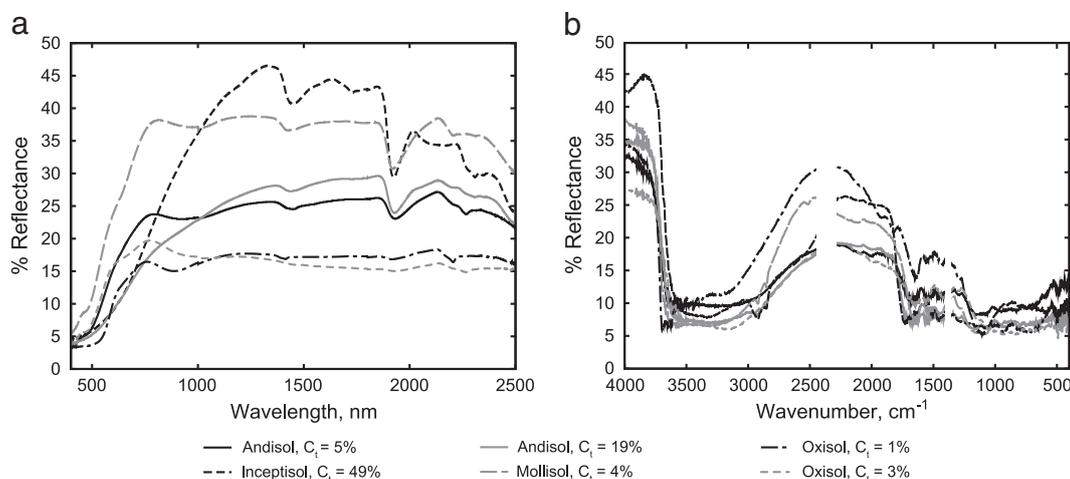


Fig. 2. Example spectra selected to illustrate the variation in a) visible, near-infrared and b) mid-infrared diffuse reflectance spectra in the Hawaiian soils data set. The soil order and total carbon (C_t) value for each spectrum are also shown.

~880–900 nm attributable to iron oxide minerals such as hematite and goethite (Burns, 1993; Gaffey et al., 1993; Morris et al., 1985). Minima at ~1400–1450 nm and ~1910–1930 nm, commonly indicative of H₂O and OH⁻ (Clark et al., 1990; Gaffey et al., 1993; Hunt, 1977), were present in almost all spectra, but the shape of these features ranged from a well-defined single minimum or doublet to a shallow and indistinct minimum. A minimum observed in some of the spectra at ~1725 nm could be associated with C–H bonds (Gaffey et al., 1993; Stuart, 2004). In almost all cases the spectra displayed one or more small minima between ~2200 and 2400 nm. In soil spectra, these features may be due to organic molecules (e.g., CH₂, CH₃, and NH₃), Si–OH bonds, cation–OH bonds in phyllosilicate minerals (e.g., kaolinite, montmorillonite), CO₃ ion in carbonate minerals (e.g., calcite), or any combination of these components (Clark et al., 1990; Gaffey et al., 1993; Hunt, 1977).

The spectral range of the MIR measurements for this study overlapped with the VNIR measurements from 6000–4000 cm⁻¹ (~1700–2500 nm), and therefore we focus on describing the remaining 4000–400 cm⁻¹ portion of the MIR spectra here. The MIR spectra had more numerous and well-defined features than the VNIR spectra, as illustrated by the selected example spectra in Fig. 2b. The most striking feature, which was present in all spectra, was the steep-sided minimum near 3625 cm⁻¹, or in some cases at 3700 cm⁻¹ with a smaller minimum at 3625 cm⁻¹, related to the stretching vibration of O–H (Clark et al., 1990; Gaffey et al., 1993; Hunt, 1977; Nguyen et al., 1991; Ryskin, 1974; Stuart, 2004). Many of the features in the infrared region, particularly between ~3400 and 1500 cm⁻¹, can be attributed to bonds within organic molecules, such as aliphatic hydrocarbons, aromatic compounds, carboxylic acid, amines, amides, and some phosphorus and sulfur compounds. Each type of bond leads to features in certain broad regions, but the specific location within that region may depend on the other elements and the structure of the compound. Some spectra had features at ~3530, 3450, and 3385 cm⁻¹ possibly linked to the N–H bond or the mineral gibbsite (Gaffey et al., 1993; Nguyen et al., 1991; Stuart, 2004). The minima at ~2925 and 2850 cm⁻¹ are characteristic of C–H bonds in alkyl groups (Gaffey et al., 1993; Stuart, 2004). Band assignments between ~2000 and 1500 cm⁻¹ can be ambiguous; both organic components, such as amide, carboxyl, and others with C–H, C=O, and C≡O bonds, and the common silicate minerals quartz and kaolinite have overlapping features in this region (Nguyen et al., 1991; Stuart, 2004). The fingerprint region (~1500–600 cm⁻¹) is where features from the fundamental vibrations of silicate minerals are located. Our spectra differed from many other soil spectra in their lack of strong quartz features in this region (e.g., Janik et al., 1995; Nguyen et al., 1991; Viscarra Rossel et al., 2006). Instead, the region was generally low in reflectance with several small minima, including those at ~1300, 1120, 1162 and 1074–1048, 1100, 1150 and 1110, 1020, 950–915, 815–800, 730–673, 500–480, and 430 cm⁻¹, likely related to clay minerals such as illite, smectite, and kaolinite, and minimal amounts of silica (Farmer, 1974; Moenke, 1974; Nguyen et al., 1991; Van der Marel and Beutelspacher, 1976).

When grouped according to soil order, neither the VNIR nor MIR spectra of the soil samples conformed to a single, consistent shape or set of features for each order. A comparison between the spectra of soils from different orders showed an equivalent degree of spectral variability within soil orders as among soil orders. The spectra from two Andisols and two Oxisols shown in Fig. 2 provide examples of this variability. Because taxonomic classification of soils is determined by criteria that include parameters with no direct influence on spectral character, and properties that have spectral influence (e.g., mineralogy) may overlap orders, variability in spectra of soils from the same order is not surprising. Interestingly, all but a couple of the Oxisol and Ultisol VNIR spectra displayed obvious features from iron oxide minerals, which gave them similar spectral shapes. This differed from the other soil orders in which some, but not all, of the spectra exhibited iron oxide features.

Evaluation of any trends in spectral character with soil C_t value is of interest given our ultimate objective of predicting C_t value from soil spectra. The MIR spectra of all soils used in this study do not exhibit any pattern in spectral features or overall shape based on measured C_t. Features and overall shape of the VNIR spectra also lack a consistent trend from low C_t to high C_t, but there is one notable observation with regard to C_t value. All VNIR spectra are similar for those samples with C_t greater than ~30%, whereas there is variation in the spectra from samples with lower C_t values. These spectra from samples with higher C_t have a shallower slope from the short wavelengths to a reflectance maximum around 1300 nm, such as the spectrum of the Inceptisol in Fig. 2a, rather than the steep slopes at shorter wavelengths observed in many of the lower C_t sample spectra, which are attributable to iron oxide minerals. The spectral features caused by iron oxides influence the spectral shape, causing an increase in reflectance at short wavelengths and a depression in reflectance at slightly longer wavelengths. An explanation of this behavior may be that highly oxidized (i.e., weathered) soils tend to be less likely to contain high amounts of carbon, resulting in the lack of spectra with iron oxide features in the high C_t portion of the sample set. Additionally, these higher C_t spectra also have a higher maximum reflectance than the majority of the lower C_t spectra, but within this range there is no positive or negative trend of maximum reflectance with C_t value. The relatively high reflectance for samples with high amounts of carbon may be counter-intuitive but could be related to several factors, including the partial dependence of overall reflectance on the complex interactions of particle size, residual moisture content, and even other less obvious soil properties such as cation exchange capacity (CEC) (e.g., Baumgardner et al., 1985). Also, it may be related to variation in the specific type of organic matter in the soil; Stoner and Baumgardner (1981) showed that the spectra of samples with minimally-decomposed organic fibers have higher overall reflectance relative to those containing highly decomposed organic matter.

3.3. Partial least squares regression models

Our pre-treatment trials showed that mean normalization of the spectra prior to PLSR model development provided the best performance for our VNIR spectra, whereas transformation to the Savitzky–Golay 1st derivative was more suitable for our MIR spectra. Using these pre-processing transformations, the PLSR models to predict C_t from both the VNIR and MIR spectra performed well, with R² values of 0.95/0.94 (VNIR/MIR) and RMSE values of 2.80%/3.08% for validation (Table 2). The high RPD and RPIQ values (4.25/3.91 and 3.66/3.38, respectively, for VNIR/MIR) were another indication of model quality. Fig. 3 compares the measured and predicted C_t values from the validation set for each of these two models. Visual inspection of these measured versus predicted plots reveals a relatively larger spread in the data at lower C_t contents, particularly for the MIR model. The results for the VNIR and MIR PLSR models were similar, but these fit and error parameters were slightly better for the VNIR model.

Table 2

Calibration and validation results to predict total soil carbon (in %) for partial least squares regression models using visible, near-infrared (VNIR) and mid-infrared (MIR) diffuse reflectance spectra.

	Calibration			Validation				
	n ^a	R ^{2b}	RMSE (%) ^c	n	R ²	RMSE (%)	RPD ^d	RPIQ ^e
VNIR	215	0.94	3.52	92	0.95	2.80	4.25	3.66
MIR	215	0.95	3.19	92	0.94	3.08	3.91	3.38

^a Number of samples.

^b Coefficient of determination.

^c Root mean squared error.

^d Residual prediction deviation.

^e Ratio of performance to inter-quartile distance.

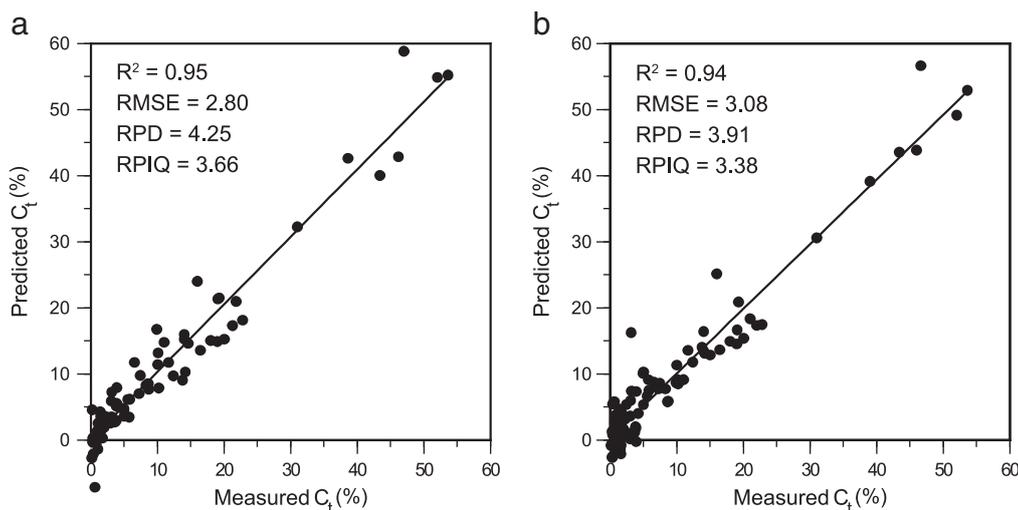


Fig. 3. Measured vs. predicted soil total carbon (C_t) values for the validation set of a) visible, near-infrared and b) mid-infrared diffuse reflectance spectra as determined by partial least squares regression. Coefficient of determination (R^2), root mean squared error (RMSE), residual prediction deviation (RPD), and ratio of performance to inter-quartile distance (RPIQ) are also given.

Our PLSR models built from VNIR DRS compare well to previous work by other researchers. Studies by Sarkhot et al. (2011), Vasques et al. (2008), and McCarty et al. (2002) reported C_t prediction models with lower R^2 and RPD values. A higher R^2 value (0.97) was produced in the work by Reeves et al. (2002). Models developed for organic C are more numerous, but the R^2 values achieved in these studies are similar (McCarty et al., 2010; Mouazen et al., 2010; Vasques et al., 2009, Vasques et al., 2010; Viscarra Rossel and Behrens, 2010). The majority of RMSE values given in these previous studies are lower than the RMSE value of our VNIR PLSR model. This difference may be influenced by the much larger range of C values in our model (e.g., a maximum of ~56% compared to 3.39% in Reeves et al. (2002), ~10% in McCarty et al. (2002) and Sarkhot et al. (2011), and ~27% in Vasques et al. (2008)).

Fewer studies have been conducted using MIR DRS to predict C_t . The results from our MIR PLSR model are comparable to those from Reeves et al. (2001) ($R^2 = 0.93$) and McCarty et al. (2002) ($R^2 = 0.95$) but are not quite as good as in Reeves et al. (2002) ($R^2 = 0.98$). The MIR model for prediction of organic C from McCarty et al. (2010), with an R^2 value of 0.94, is comparable to the MIR model in this study, as well. As with the VNIR studies mentioned above, the RMSE values of these previous MIR models are lower than for our study and likely reflect their lower C values.

3.4. Random forests ensemble tree regression models

We found that the best combination of RF model parameters for our VNIR data was 1000 trees and 335 random variables in the subset for each split; the best combination for our MIR data was 500 trees and 396 random variables, respectively. Of the pre-processing transformations, the best RF models were produced using the Norris Gap derivative for our VNIR data and the Savitzky–Golay 1st derivative (same pre-treatment as in the PLSR model) for our MIR data. These RF models to predict C_t also produced good results with both the VNIR and MIR data (Table 3). The R^2 and RMSE values for validation were 0.95 and 2.82% for the VNIR model and 0.96 and 2.28% for the MIR model. The measured versus predicted C_t values from the validation set for these models are shown in Fig. 4. The RPD and RPIQ values of 4.27/5.28 (VNIR/MIR) and 3.68/4.56 (VNIR/MIR) suggest high quality models. In this case, the MIR model slightly outperformed the VNIR in each of these parameters.

We have found one study similar to ours where RF was used to predict soil C from DRS data in the laboratory. Viscarra Rossel and Behrens

(2010) developed a RF model to predict organic C using VNIR DRS and reported an R^2 of 0.71 and RMSE of 1.23%. Both the RF models presented here performed considerably better than this previous study with regard to R^2 , and the difference in RMSE may be due at least in part to a smaller C range (<13.90%) in the Viscarra Rossel and Behrens (2010) work. More studies have been conducted for the prediction of C_t and organic C using committee trees (Sarkhot et al., 2011; Vasques et al., 2008, 2009, Vasques et al., 2010), but the quality and error parameters for these models are also poorer than those from our RF models.

3.5. Variable significance in regression models

The relative importance of the variables (i.e., wavelengths and wave-numbers) to the C_t prediction models are plotted in Fig. 5. For the PLSR models we examined the weighted regression coefficients incorporating the first three factors, where larger absolute value indicates larger variable significance. The RF method ranks variable importance by the increase in error when absent from the model. Variables with larger values are considered more important.

The VNIR PLSR model (Fig. 5a) focused mainly on the region around 670 nm, corresponding to the region of iron oxide features in the spectra. This region was somewhat important to the VNIR RF model as well (Fig. 5b), but the RF model relied more on the portion of the spectrum influenced by H_2O and OH (~1413 and ~1918 nm) and especially cation–OH interactions (~2045 and 2268 nm). The PLSR model used this general region (~1920, ~2124, and ~2305 nm), but it was of lesser importance than the iron oxide region.

Table 3

Calibration and validation results to predict soil total carbon (in %) for random forests ensemble tree regression models using visible, near-infrared (VNIR) and mid-infrared (MIR) diffuse reflectance spectra.

	Calibration			Validation				
	n^a	Mean of squared residuals	% variance explained	n	R^{2b}	RMSE (%) ^c	RPD ^d	RPIQ ^e
VNIR	215	19.06	90.94	92	0.95	2.82	4.27	3.68
MIR	215	11.53	94.52	92	0.96	2.28	5.28	4.56

^a Number of samples.

^b Coefficient of determination.

^c Root mean squared error.

^d Residual prediction deviation.

^e Ratio of performance to inter-quartile distance.

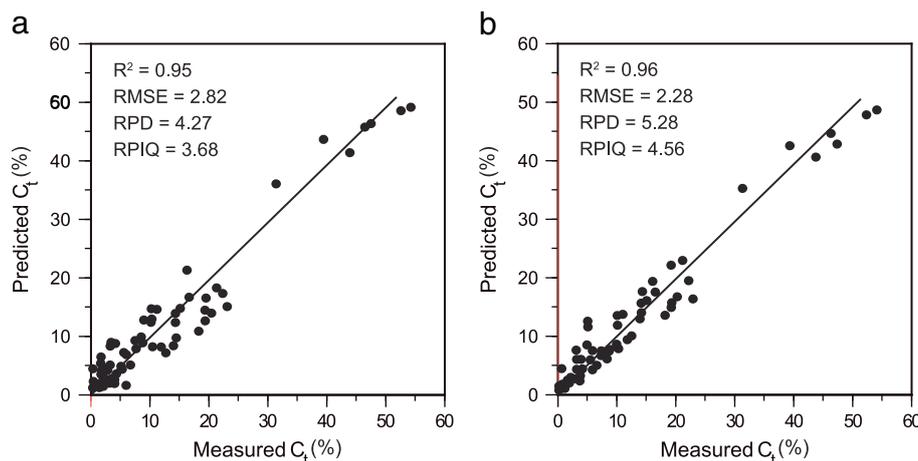


Fig. 4. Measured vs. predicted soil total carbon (C_t) values for the validation set of a) visible, near-infrared and b) mid-infrared diffuse reflectance spectra as determined by random forests ensemble tree regression. Coefficient of determination (R^2), root mean squared error (RMSE), residual prediction deviation (RPD), and ratio of performance to inter-quartile distance (RPIQ) are also given.

Many variables stood out as important for the MIR PLSR model (Fig. 5c), including some from the region where the MIR spectra overlap into the VNIR wavelength range. These variables are $\sim 5255\text{ cm}^{-1}$ ($\sim 1903\text{ nm}$), related to H_2O and OH^- features, and $\sim 4440\text{ cm}^{-1}$ ($\sim 2252\text{ nm}$), $\sim 4320\text{ cm}^{-1}$ ($\sim 2315\text{ nm}$), and $\sim 4245\text{ cm}^{-1}$ ($\sim 2356\text{ nm}$), likely related to cation–OH features. No variables from the region of overlap were identified as important in the MIR RF model (Fig. 5d). The region around $\sim 3675\text{ cm}^{-1}$, which corresponds to the intense feature from the O–H bond in the spectrum, was used in the PLSR model but not the RF model. Variables located in the region influenced by the C–H bond (~ 3000 and 2933 cm^{-1}) were important to the RF model but not the PLSR model. Both MIR models relied on the spectral region containing features related to multiple organic components and silicate minerals. These particular variables are 1820 , 1705 , and 1520 cm^{-1} for the PLSR model and 1792 , 1717 , and 1613 cm^{-1} for the RF model. The fingerprint region, corresponding to predominantly silicate features, also was important for the RF model (1300 , 1230 , and 955 cm^{-1}).

3.6. Comparison of regression methods and wavelength ranges

A comparison of our PLSR and RF results showed that the latter method performed slightly better than PLSR on MIR spectral data. Results for the two VNIR models were almost identical to one another. Although the specific wavelengths that the VNIR and MIR models found most important differed between the methods, both combinations generated models with similar accuracy. The small differences in the final results suggest that either method is capable of producing high quality models with VNIR or MIR spectra. Each method has its advantages and disadvantages to be weighed depending on the data set and circumstances. Partial least squares regression is well known and the results are easily interpreted, but the appropriate number of factors must be determined to avoid over-fitting, and models may be adversely affected if the data has a highly non-normal distribution. Because RF is a non-parametric method, it requires little adjustment of model input parameters and is not affected by non-normal data distribution, but it is computationally intensive, and further interpretation of results beyond the model's built in error and variable importance output is difficult because the model structure is complex and less transparent. In addition, soil scientists, lab managers, and growers may be less familiar with the RF method and it is not yet widely available in software packages outside of R. In either case, pre-processing transformation of the spectral data may be required for the best models.

For the soils in this study, the PLSR model developed from the VNIR wavelength range slightly outperformed the MIR model, whereas the opposite was observed for the RF models. It must also be noted that

the specific results of the models do vary when different groupings are used for the calibration and validation subsets. The alternative models are of comparable quality, but slight changes in the similar fit and error parameters mean that the relative performance of one wavelength range over the other may not always hold. In past studies the MIR wavelength range produced the better model for C_t prediction (McCarty et al., 2002; Reeves, 2010; Reeves et al., 2002; Viscarra Rossel et al., 2006), but we demonstrate that the VNIR wavelength range can perform just as well in some cases. The common improvement in models using MIR spectra has led to the idea that the features from fundamental vibrations in the MIR are more suited to C_t prediction than the features in the VNIR. For these Hawaiian soils, the model results are not different enough to conclusively state that the features present in one wavelength range are better for C_t prediction than those of the other. Variation in the degree of difference between VNIR and MIR models in DRS studies, including the current work, is likely a reflection of the different sets of soil spectra used in these studies and the high dependence of model quality on the individual data sets. The potential increase in accuracy with a MIR model must also be weighed with the advantages and disadvantages for each wavelength range, as pointed out by Viscarra Rossel et al. (2006). Ball milling to a fine homogenous powder must be included in the sample preparation for collection of MIR spectra, adding time and cost, whereas VNIR spectra are collected from 2 mm sieved soils.

3.7. Implications for the application of diffuse reflectance spectroscopy to Hawaiian soils

This study provided the first spectral library of Hawaiian agricultural soils. The VNIR and MIR spectra displayed the expected features from common soil constituents as seen in other soil spectra; the proportion of VNIR spectra with prominent iron oxide features may have been greater than in many soil studies because of the large extent of weathered Oxisols and Ultisols in Hawaii. Along with these important Oxisols and Ultisols, the sample set included a large number of Andisols and the seven other soil orders present in Hawaii, providing a representative set of Hawaiian agricultural soils. Soils from the orders common in Hawaii have been under-represented in the literature and this data set helps to expand our knowledge of the spectral characteristics of these soils. The diversity in soil type did not impede the creation of good prediction models and actually helped to produce robust models that should be appropriate for most, if not all, agricultural soils found in the Hawaiian Islands.

Our results from both PLSR and RF prediction models, particularly the R^2 values, suggest that the use of VNIR and MIR DRS has potential

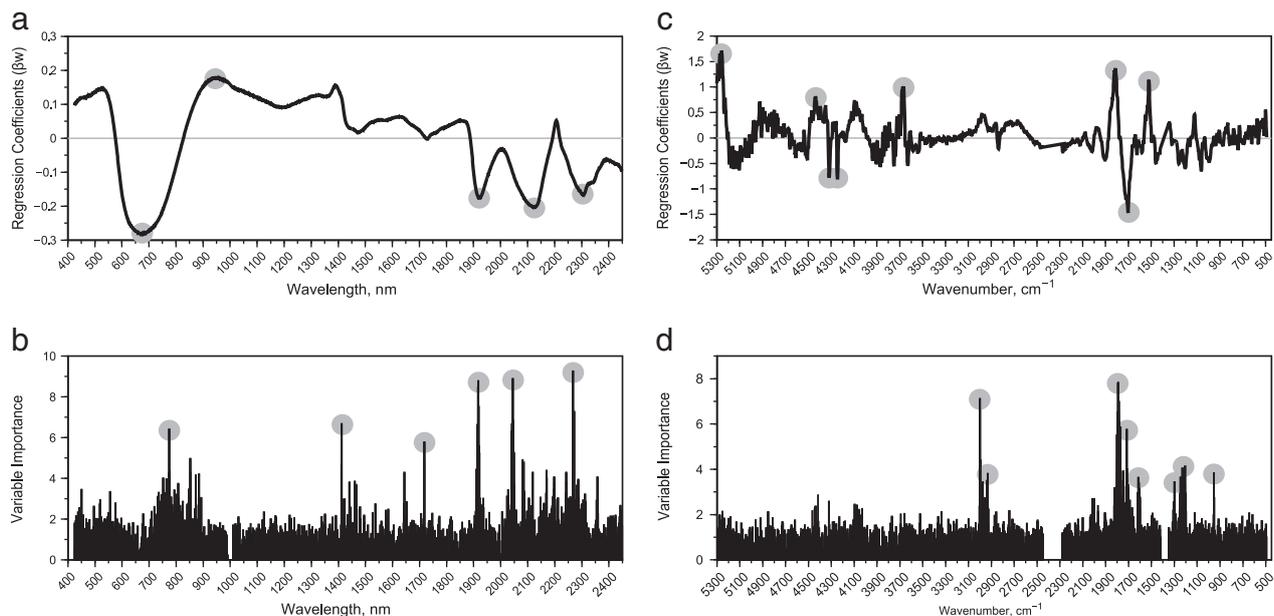


Fig. 5. Important variables for the prediction of C_t in the a) visible, near-infrared (VNIR) partial least squares regression (PLSR) model, b) VNIR random forests (RF) ensemble tree regression model, c) mid-infrared (MIR) PLSR model, and d) MIR RF model. The shaded circles highlight some of the most significant variables for each model from visual estimation.

as a method for determining C_t in Hawaiian soil studies. Root mean squared error values of 2.28% to 3.08% may indicate sufficient prediction accuracy for usage with moderate to high C_t soils, but specific errors at low C_t values must be investigated further to determine if it is reasonable for usage with very low C_t soils. The bias component of the RMSE from these models is low, ranging from 0.072% to 0.325%, suggesting that the remaining larger portion of the error could be decreased by the use of replicates in future predictions, as outlined in Bellon-Maurel et al. (2010). Transitioning from this initial evaluation to practical application of the method will require further model testing and refinement to decrease prediction error as much as possible, especially for samples with low C_t values. According to the models from this data set, the choice of VNIR or MIR spectra should not have a significant effect on the accuracy of C_t prediction for agricultural soils in Hawaii.

4. Conclusions

This work investigated the use of DRS for C_t analysis in Hawaiian soils. We collected VNIR and MIR spectra from primarily agricultural soils with a wide range of C_t content and mineralogy throughout the Hawaiian Islands and presented the first characterization of diffuse reflectance spectra from such a diverse set of Hawaiian soils. Both wavelength ranges, VNIR and MIR, and both chemometric methods, PLSR and RF ensemble tree regression, produced high quality C_t prediction models from this sample set, with R^2 values of 0.94 or greater, RMSE values less than 3.08%, and RPD and RPIQ values of 3.38 or more. These excellent results from model development and prediction showed that soil C_t prediction by DRS may be a viable and beneficial alternative technique for use in Hawaii. After such promising results with DRS for C_t analysis on these Hawaiian soils, further study into the prediction of other soil properties from VNIR and MIR spectra is warranted to take advantage of the technique's ability to provide information on multiple parameters from a single spectrum.

Acknowledgements

This research was supported by USDA CSREES TSTAR Project 2009-34135-20183, CTAHR Hatch Project HA-154, and by the UHM University Research Council (URC). We would like to thank J. Hempel, L. West, T.

Reinsch, L. Arnold, and R. Nesser of the NRCS National Soil Survey Center in Lincoln, NE for help with access, sampling and scanning of the archived samples; L. Muller and A. Quidez for help with scanning of samples at UHM; and Drs. G. Uehara, R. Yost, and D. Beilman of UHM for support of this project. We also appreciate the Hawaii Landowners, Managers, and Extension Agents that gave us access to their fields for collecting soil samples. These include from Kauai: R. Yamakawa and J. Gordines (CTAHR), S. Lupkes (BASF), and Grove Farms; from Oahu: R. Corrales, A. Umaki, and J. Grzebek (CTAHR), Hoa Aina, MAO Organic Farm, Nii Nursery, J. Antonio and M. Conway (Dole), C. and P. Reppun, L. Santo, T. Jones, and N. Dudley (HARC), and A. Sou (Aloun Farms); from Molokai: A. Arakaki (CTAHR), K. Duvchelle (NRCS), and R. Foster (Monsanto); and from Maui: J. Powley and D. Oka (CTAHR), M. Nakahata and M. Ross (HC&S), T. Callender (Ulupono), and B. Abru.

References

- Allison, L.E., Bollen, W.B., Moodie, C.D., 1965. Total Carbon. In: Black, C.A. (Ed.), *Methods of Soil Analysis, Part 2*. Am. Soc. of Agron., Inc., Madison, WI, pp. 1346–1366.
- AOAC International, 1997. *Official Methods of Analysis of AOAC International*, 16th ed. AOAC International, Arlington, VA.
- Ball, D.F., 1964. Loss-on-ignition as an estimate of organic matter and organic carbon in noncalcareous soils. *Journal of Soil Science* 15, 84–92.
- Baumgardner, M.F., Silva, L.F., Biehl, L.L., Stoner, E.R., 1985. Reflectance properties of soils. *Advances in Agronomy* 38, 1–44.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical reviews of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry* 29, 1073–1081.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
- Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D., Reinsch, T.G., 2006. Global soil characterization using a VNIR diffuse reflectance library and boosted regression trees. *Geoderma* 132, 273–290.
- Burns, R.G., 1993. Origin of electronic spectra of minerals in the visible to near-infrared region. In: Pieters, C.M., Englert, P.A. (Eds.), *Remote geochemical analysis: elemental and mineralogical composition*. Cambridge University Press, Cambridge, UK, pp. 3–29.
- Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., 1990. High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research* 95, 12653–12680.
- Cohen, M.J., Prenger, J.P., DeBusk, W.F., 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. *Journal of Environmental Quality* 34, 1422–1434.
- Craswell, E.T., Lefroy, R.D.B., 2001. The role and function of organic matter in tropical soils. *Nutrient Cycling in Agroecosystems* 61, 7–18.
- Dalal, R.C., Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Science Society of America Journal* 50, 120–123.

- Dardenne, P., Sinnaeve, G., Baeten, V., 2000. Multivariate calibration and chemometrics for near infrared spectroscopy: which method? *Journal of Near Infrared Spectroscopy* 8, 229–237.
- Deenik, J., McClellan, A.T., 2007. Soils of Hawaii. *Soil and Crop Management* 20, 1–12.
- Farmer, V.C., 1974. The layer silicates. In: Farmer, V.C. (Ed.), *The infrared spectra of minerals*. Bartholomew Press, Surrey, UK, pp. 331–363.
- Feller, C., Albrecht, A., Tessier, D., 1996. Aggregation and organic matter storage in kaolinitic and smectitic tropical soils. In: Carter, M.R., Stewart, B.A. (Eds.), *Structures and Organic Matter Storage in Agricultural Soils*. CRC Lewis Publishers, Boca Raton, FL, pp. 309–360.
- Gaffey, S.J., McFadden, L.A., Nash, D., Pieters, C.M., 1993. Ultraviolet, visible, and near-infrared reflectance spectroscopy: laboratory spectra of geologic materials. In: Pieters, C.M., Englert, P.A. (Eds.), *Remote geochemical analysis: elemental and mineralogical composition*. Cambridge University Press, Cambridge, UK, pp. 43–77.
- Gavenda, R., Smith, C., Vollrath, N., 1998. *Soils, Atlas of Hawaii*, 3rd edition. University of Hawaii Press, Honolulu, HI.
- Hunt, G.R., 1977. Spectral signatures of particulate minerals in the visible and near-infrared. *Geophysics* 42, 501–513.
- Janik, L.J., Skjemstad, J.O., Raven, M.D., 1995. Characterization and analysis of soils using mid-infrared partial least squares. I. Correlations with XRF-determined major element composition. *Australian Journal of Soil Research* 33, 621–636.
- Janik, L.J., Skjemstad, J.O., 1995. Characterization and analysis of soils using mid-infrared partial least-squares. II. Correlations with some laboratory data. *Australian Journal of Soil Research* 33, 637–650.
- Janik, L.J., Merry, R.H., Skjemstad, J.O., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture* 38, 681–698.
- Janik, L.J., Forrester, S.T., Rawson, A., 2009. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemometrics and Intelligent Laboratory Systems* 97, 179–188.
- Krishnan, P., Alexander, J.D., Butler, B.J., Hummel, J.W., 1980. Reflectance technique for predicting soil organic matter. *Soil Science Society of America Journal* 44, 1282–1285.
- Lal, R., Kimble, J.M., Follett, R.F., Stewart, B.S., 2001. *Methods of Assessment of Soil Carbon*. CRC Press, Boca Raton, FL.
- Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. *Science* 304, 1623–1627.
- Liaw, A., Wiener, W., 2002. Classification and regression by random forest. *R News* 2, 18–22.
- McBratney, A.B., Minasny, B., Viscarra-Rossel, R.A., 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. *Geoderma* 131, 59–75.
- McCarty, G.W., Reeves, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal* 66, 640–646.
- McCarty, G.W., Reeves, J.B., 2006. Comparison of IR and MIR diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. *Soil Science* 171, 94–102.
- McCarty, G.W., Reeves III, J.B., Yost, R., Doraiswamy, P.C., Doumbia, M., 2010. Evaluation of methods for measuring soil organic carbon in West African soils. *African Journal of Agricultural Research* 5 (16), 2169–2177.
- Moenke, H.H.W., 1974. Silica, the three-dimensional silicates, borosilicates, and beryllium silicates. In: Farmer, V.C. (Ed.), *The infrared spectra of minerals*. Bartholomew Press, Surrey, UK, pp. 365–382.
- Morris, R.V., Lauer Jr., H.V., Lawson, C.A., Gibson Jr., E.K., Nace, G.A., Stewart, C., 1985. Spectral and other physicochemical properties of submicron powders of hematite (α -Fe₂O₃), maghemite (γ -Fe₂O₃), magnetite (Fe₃O₄), goethite (α -FeOOH), and lepidocrocite (γ -FeOOH). *Journal of Geophysical Research* 90, 3126–3144.
- Mouazen, A.M., Kuang, B., De Baerdemaeker, J., Ramon, H., 2010. Comparison among principal component, partial least squares, and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* 158, 23–31.
- Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. In: Sparks, D.L., Page, A.L., Lemke, P.A., Loeppert, R.H., Softanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sommers, M.E. (Eds.), *Methods of Soil Analysis, Part 3. SSSA-ASA, Madison, WI*, pp. 983–997.
- Nguyen, T.T., Janik, L.J., Raupach, M., 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. *Australian Journal of Soil Research* 29, 49–67.
- Oenema, O., Janssen, B.H., Smaling, E., Hoffland, E., 2006. Nutrient management in tropical agroecosystems. *Agriculture, Ecosystems and Environment* 116, 1–3.
- Paustian, K., Andren, O., Janzen, H.H., Lal, R., Smith, P., Tian, G., Tiessen, H., Van Noordwijk, M., Woomer, P.L., 1997. Agricultural soils as a sink to mitigate CO₂ emissions. *Soil Use and Management* 13, 230–244.
- R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org>.
- Reeves III, J.B., McCarty, G.W., Reeves, V.B., 2001. Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *Journal of Agricultural and Food Chemistry* 49, 766–772.
- Reeves III, J., McCarty, G., Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environmental Pollution* 116, S277–S284.
- Reeves III, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done? *Geoderma* 158, 3–14.
- Ryskin, Ya.I., 1974. The vibrations of protons in minerals: hydroxyl, water, and ammonium. In: Farmer, V.C. (Ed.), *The infrared spectra of minerals*. Bartholomew Press, Surrey, UK, pp. 137–181.
- Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. *Geoderma* 148, 149–158.
- Sarkhot, D.V., Grunwald, S., Ge, Y., Morgan, C.L.S., 2011. Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. *Geoderma* 164, 22–32.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66, 988–998.
- Sørensen, L.K., Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy. *Soil Science Society of America Journal* 69, 159–167.
- Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma* 158, 15–22.
- Stoner, E.R., Baumgardner, M.F., 1981. Characteristic variations in reflectance of surface soils. *Soil Science Society of America Journal* 45, 1161–1165.
- Stuart, B., 2004. *Infrared spectroscopy: fundamentals and applications*. John Wiley & Sons Ltd., West Sussex, UK.
- Terhoeven-Urselmans, T., Vagen, T.-G., Spaargaren, O., Shepherd, K.D., 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Science Society of America Journal* 74, 1792–1799.
- Tiessen, H., Cuevas, E., Chacon, P., 1994. The role of soil organic matter in sustaining fertility. *Nature* 371, 783–785.
- Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-squares regression: a feasibility study. *Plant and Soil* 251, 319–329.
- Van der Marel, H.W., Beutelspacher, H., 1976. *Atlas of infrared spectroscopy of clay minerals and their admixtures*. Elsevier Scientific Publishing Company, Amsterdam.
- Vasques, G.M., Grunwald, S., Sickman, J.O., 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* 146, 14–25.
- Vasques, G.M., Grunwald, S., Sickman, J.O., 2009. Modeling of soil organic carbon fractions using visible – near-infrared spectroscopy. *Soil Science Society of America Journal* 73, 176–184.
- Vasques, G.M., Grunwald, S., Harris, W.G., 2010. Spectroscopic models of soil organic carbon in Florida, USA. *Journal of Environmental Quality* 39, 923–934.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75.
- Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158, 46–54.
- Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science* 37, 29–38.
- Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D.J., Dokken, D.J., 2000. *Land Use, Land-Use Change, and Forestry, A Special Report of the IPCC*. Cambridge University Press, Cambridge, UK, pp. 53–126.
- Williams, P.C., 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: Williams, P., Norris, K. (Eds.), *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, St. Paul, MN, pp. 143–167.