Geoderma 262 (2016) 254–265

Contents lists available at ScienceDirect

Geoderma

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

Large topsoil organic carbon variability is controlled by Andisol properties and effectively assessed by VNIR spectroscopy in a coffee agroforestry system of Costa Rica

Rintaro Kinoshita [a,*], Olivier Roupsard [b,c], Tiphaine Chevallier [d], Alain Albrecht [d], Simon Taugourdeau [e], Zia Ahmed [f], Harold M. van Es [a]

[a] School of Integrative Plant Science, Soil and Crop Sciences Section, Cornell University, Ithaca, NY 14853–1901, USA
[b] CIRAD, UMR Eco&Sols (Ecologie Fonctionnelle & Biogéochimie des Sols et des Agro-écosystèmes), 34060 Montpellier, France
[c] CATIE (Tropical Agricultural Centre for Research and Higher Education), 7170 Turrialba, Costa Rica
[d] IRD, UMR Eco&Sols (Ecologie Fonctionnelle & Biogéochimie des Sols et des Agro-écosystèmes), 34060 Montpellier, France
[e] CIRAD, UMR SELMET (Systèmes d'élevage méditerranéens et tropicaux), 34398 Montpellier, France
[f] CIMMYT (International Maize and Wheat Improvement Center), Dhaka 1207, Bangladesh

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form 8 July 2015
Accepted 16 August 2015
Available online 5 September 2015

Keywords:
Soil organic carbon
VNIR spectroscopy
Random Forest
Co-kriging
Andisols
Allophane
Agroforestry

ABSTRACT

Assessing the spatial variability of soil organic carbon (SOC) is crucial for SOC monitoring and comparing management options. Topsoil (0–5 cm) SOC concentrations were surveyed in a coffee agroforestry watershed (0.9 km^2) on Andisols in Costa Rica with uniform farm management. We encountered high values and large spatial variations of SOC, from 48.1 to 172 g kg^{-1} in the dry combustion set (SOC_{ref}; n = 72) used for calibrating the visible-near-infrared reflectance spectroscopy (VNIRS) samples (SOC_{VNIRS}; 350–2500 nm; n = 520). VNIRS using partial least squares regression was effective in predicting SOC ($R^2 = 0.85$; a root mean square error (RMSE) = 12.3 g kg^{-1}) and proved an effective proxy measurement. We assessed several topographic, vegetation and andic soil property variables, of which only the latter (metal–humus complexes and allophanes) displayed strong correlations with SOC_{ref} concentrations. We compared Random Forest and three geostatistical approaches for the interpolation of SOC in unsampled locations. Ordinary kriging with SOC_{ref} yielded an RMSE of 28.0 g kg^{-1}. Random Forest was successful in incorporating many weakly and non-linearly correlated covariates with SOC (RMSE = 14.7 g kg^{-1}), provided Al_p (the sodium pyrophosphate extractable aluminum), the best predictor of SOC ($r = 0.85$) but also the most costly variable to acquire. Co-kriging with Al_p also showed high reduction in RMSE (16.0 g kg^{-1}). Co-kriging with SOC_{VNIRS} only showed marginal reduction in RMSE to 24.2 g kg^{-1} due to the presence of a high nugget effect.
Local variability of SOC in this volcanic agroforestry watershed was dominated by andic properties whereas topographic or vegetation variables had very little impact. Estimation of SOC variability is recommended using inexpensive proxy measurements like VNIRS (RMSE = 12.3 g kg^{-1}) rather than spatial interpolation techniques.

1. Introduction

Soil organic carbon (SOC) is a fundamental property related to soil physical, chemical and biological quality and is an important component of the global carbon (C) cycle (Magdoff and van Es, 2009). Disruption of sustainable C cycles in agricultural soils has led to diminishing crop yields as well as contributing to further accelerating greenhouse gas (GHG) emissions (Hillel and Rosenzweig, 2010; Lal, 2006; Powlson et al., 2011). At a farm-scale, high spatial variation of SOC may occur, which causes uncertainty when comparing several management practices or when assessing the effectiveness of various soil conservation measures to restore SOC (Minasny et al., 2013).

There is need for accurate approaches to assess the impact of management on SOC at the farm-scale, whatever the inherent variability. Various biotic and abiotic variables have been identified to correlate with SOC at various spatial scales and soil environment, such as past and present land use (Schulp and Veldkamp, 2008), local terrain (Cambule et al., 2014; Thompson and Kolka, 2005), and vegetation (Bou Kheir et al., 2010; Horwath Burnham and Sletten, 2010; Kunkel et al., 2011; Takata et al., 2007). These correlated variables have been used to predict SOC through various methods such as multiple linear regression (Gessler et al., 2000; Thompson and Kolka, 2005), Random Forest (RF; Grimm et al., 2008), boosted regression tree (Razakamanarivo et al., 2011), co-kriging (Terra et al., 2004) and regression kriging

* Corresponding author at: 1007 Bradfield Hall, Cornell University, Ithaca, NY 14853–1901.
E-mail address: rk422@cornell.edu (R. Kinoshita).

http://dx.doi.org/10.1016/j.geoderma.2015.08.026

(Bilgili et al., 2011; Hengl et al., 2004; Simbahan et al., 2006). Any suitable approach may need to be determined in each environment according to the availability of information.

In recent years, visible-near-infrared reflectance spectroscopy (VNIRS), a rapid and cost effective proximal soil sensing method (Barthes et al., 2006; Brunet et al., 2007; Kinoshita et al., 2012) has been used to predict SOC (Bellon-Maurel and McBratney, 2011). It can be used in laboratory or in field by first calibrating the reflectance data with analytically measured SOC information through multivariate statistical or data mining techniques, and predicting the SOC content of a new sample only with the spectra (Barthes et al., 2006; Viscarra Rossel et al., 2006). The predicted SOC contents can be used in geostatistical techniques by directly replacing analytically measured SOC information in ordinary kriging (Lamsal, 2009), or as a covariate in co-kriging and regression kriging as demonstrated in studies using soil organic matter (Bilgili et al., 2011).

Coffee agroforestry is an environmentally and economically important agricultural system in Central America, where Arabica coffee (*Coffea arabica* L., var *Caturra*) is grown under shade trees on the slopes of volcanoes at high elevations (Somarriba et al., 2012). Shade trees are known to improve the size and the quality of coffee beans by buffering unfavorable climatic conditions (Muschler, 2001) and also result in reduced soil erosion and compaction while increasing SOC (Beer et al., 1998). However, past research has shown extremely localized effect (<1 m) of the shade trees on SOC (Payán et al., 2009), which may not be detectable when soil samples from shaded and non-shaded sites are composited (Noponen et al., 2013).

Many of the coffee growing regions in Central America are associated or dominated by Andisols (USDA-NRCS, 2005), which are often evaluated as the most productive arable soils in the tropics especially where their parent material is basaltic (Shoji et al., 1993). Andisols are known to have substantial SOC sequestration potential, containing 1.8% of global total soil C stocks while covering only 0.7% of global ice-free land (Batjes, 1996; Hillel and Rosenzweig, 2010). They contain non-crystalline clay amorphous minerals that originated in volcanic ejecta (Parfitt, 1990) but the mechanisms of SOC stabilization is still debated. Past research has suggested several possible pathways such as Al toxicity against microbial degradation (Boudot, 1992), formation of organo-mineral complexes, physical protection by stabilizing soil aggregates (Huygens et al., 2005) or the fractal structure of allophane (Chevallier et al., 2010). Variable volcanic ash input, mass erosion, soil pH, and moisture regimes cause high spatial variation in the degree of their crystallization and andic soil property distributions (Chesworth, 2008; Nanzyo et al., 1993), leading to high spatial variability of SOC in Andisols. At the landscape scale (140,000 ha), Powers and Schlesinger (2002) have assessed the relationships between soil C and various biotic and abiotic variables across a large elevation gradient (50-to-750 m). They identified elevation and Normalized Difference Vegetation Index (NDVI) to correlate to soil C across the elevation gradient while andic soil properties (allophane and Al–humus complex) showed strong correlation within a similar elevation site. Nevertheless, farm scale variability of SOC on Andisols has rarely been assessed to date and a need exists to explore correlated biotic and abiotic variables at this scale.

This study was conducted in a micro-watershed on Andisols, covered by Arabica coffee agroforestry, a major land use in Central America (Somarriba et al., 2012). The main objectives of this study are 1) to assess the dominant attributes among topographic, vegetation and andic soil properties influencing farm-scale spatial variability of topsoil SOC on Andisols and 2) to determine optimal feature-space and geostatistical interpolation models to predict SOC concentrations at unsampled sites with available predictors.

2. Materials and methods

2.1. Site description

The research site is located in the Central-Caribbean area of Costa Rica. It is part of the Reventazón River Basin on the slope of the Turrialba volcano, which eventually drains into the Caribbean Sea. The research site is located within the Aquiares Coffee Farm (6.6 km^2) called the Mejias creek watershed (Fig. 1a) located between 83° 44′ 39″ and 83° 43′ 35″W, and between 9° 56′ 8″ and 9° 56′ 35″ N. The watershed has an area of 0.9 km^2 with elevation ranging from 1020 to 1280 m a.s.l. and a mean slope of 11.3° (Gómez-Delgado et al., 2011). The climate is tropical humid without a dry season and influenced strongly by the Caribbean Sea (Peel et al., 2007). The mean annual precipitation from 1973 to 2009 was 3014 mm.

According to Mora-Chinchilla (2000), volcanic avalanche deposits form the geology of the area, which was originally produced by the collapse of a 1.3 km wide strip of the southeastern slope of the Turrialba volcano. Indications of lava flows, agglomerates, lahars and ashes are also present. The soils are classified as Andisols and are generally characterized by their high SOC contents, infiltration capacities, and biological activities (Payán et al., 2009). The superficial runoff in the study site is very low even on slopes, and the sediment production by the whole watershed is *ca.* 1 t ha^{-1} yr^{-1} (Gómez-Delgado, 2010; Gómez-Delgado et al., 2011). The steady state infiltrability has been measured as high as *ca.* 1000 mm h^{-1} (Benegas et al., 2014), confirming that the hydrology is dominated by infiltration.

Before the introduction of coffee, the western side of the watershed was under a cardamom (*Elettaria cardamomum*) plantation until 1988 while most of the other parts were under housing and gardens until 1975 (Fig. 1a). The vegetation is now a mixture of homogeneously planted coffee trees and *Erythrina poeppigiana* leguminous shade trees on bare soils. Shade trees have a density of 7.4 trees ha^{-1}, with 15.7% ± 5.5% canopy projection and an average canopy height of *ca.* 20 m (Taugourdeau et al., 2014). The initial planting density of the coffee trees was 6300 locations ha^{-1}, which have been selectively pruned over more than 30 years (old exhausted re-sprouts aged >5–6 yr pruned every year in March, representing ca. 15% of total re-sprouts, resulting in minimum leaf area index (LAI) in March). The canopy openness of the coffee is *ca.* 25% and the average canopy height *ca.* 1.2 m. The Aquiares farm is managed quite intensively (upper conventional mode) in terms of fertilizer application (mean = 214 kg N ha^{-1} yr^{-1}; standard deviation (SD) = 44 kg N ha^{-1} yr^{-1}; 2000–2012), and complies with Rainforest Alliance™ guidelines for its pest and weed management. Weeds are scarce and it is assumed in this research that the soil surface is uncovered except for litter. The yields of green coffee from 1994 to 2011 averaged 1375 kg ha^{-1} yr^{-1} (SD = 341 kg ha^{-1} yr^{-1}).

2.2. Soil sampling strategy

In total, 520 soil samples were systematically collected for VNIRS analysis in the 0.9 km^2 watershed (5.8 samples ha^{-1}) in July–August 2010 from the 0-to-5 cm depth using a push probe after the removal of surface residues. We decided to oversample the study site systematically within an achievable timeframe and budget, because no prior knowledge on the correlation between environmental covariates and SOC could guide sampling (Minasny et al., 2013; Walvoort et al., 2010). The watershed was first divided into 23 grids aligned with the MODIS satellite image (20 nearly-full MODIS pixels, Fig. 1a). Within each MODIS pixel, 6 transects were drawn E–W and 12 N–S (Fig. 1b) and the sampling points were placed on the intersections, following a triangular grid pattern (Gilbert, 1987). The resulting points were 80 m and 40 m apart in E–W and N–S directions, respectively, accounting for higher topographical variations in N-S slopes. Some sampling points were not accessible due to watercourses, roads, and dense vegetation along the rivers and omitted from the sample set.

Among the 520 points, every 8th sample was selected systematically and subjected to additional soil tests (Fig. 1a and b; "reference sample points"). We opted for systematic sampling in order to maximize the chance to capture the overall farm-scale spatial variation of SOC a priori. At each sampling point, four subsamples were taken for the VNIRS set (Fig. 1c), plus four additional subsamples for the reference sample set,