Short-range-order minerals as powerful factors explaining deep soil organic carbon stocks distribution: the case of a coffee agroforestry plantation on Andosols in Costa Rica

Tiphaine Chevallier[1], Kenji Fujisaki[1], Olivier Roupsard[1,2], Florian Guidat[3], Rintaro Kinoshita[3,5], Elias de Melo Viginio Filho[3], Peter Lehner[4], Alain Albrecht[1]

[1] Eco&Sols, IRD, CIRAD, INRA, Montpellier SupAgro, Univ Montpellier, Montpellier, France
[2] CIRAD, UMR Eco&Sols, LMI IESOL, B.P. 1386 CP 18524, Dakar, Senegal
[3] CATIE, Tropical Agricultural Centre for Research and Higher Education, 7170 Turrialba, Costa Rica
[4] Cafetalera Aquiares S.A., PO Box 362-7150, Turrialba, 7150, Costa Rica
[5] Research Center for Global Agromedicine, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan

Correspondence to: Tiphaine Chevallier (tiphaine.chevallier@ird.fr)

Abstract. Soil organic carbon (SOC) constitutes the largest terrestrial C stock, particularly in the Andosols of volcanic areas. Quantitative information on distribution of SOC stocks is needed to construct a baseline for studying temporal changes in SOC. The spatial variation of soil short-range-order minerals such as allophane usually explains the variability of topsoil SOC contents, but SOC data for deeper soil layers are needed. We found that within a 1-km^2 Costa Rican basin covered by coffee agroforestry, SOC stocks in the upper 200 cm of soil were highly variable (24 to 72 kgC m^{-2}). Topsoil SOC stocks were not correlated with SOC stocks present in deeper layers. Diffuse-reflectance mid-infrared (MIR) spectroscopy made possible the analysis of a large number of samples (69 soil profiles, *i.e.* 598 soil samples) for ammonium-oxalate and sodium-pyrophosphate extractable forms of Al, Fe and Si, as well as SOC content and bulk density. Using the MIR spectra, we identified two different soil materials, which were identified as *allophanic* and *halloysitic* soil material. *Allophanic* soil occurred on top of the *halloysitic* soil. The thickness of the *allophanic* soil material, rich in SRO minerals and related to young andic A horizon, explained the variability of SOC. This study illustrates that knowledge of topography and pedogenesis are needed to understand and extrapolate the distribution of SOC stocks at landscape scales.

1 Introduction

Soil organic carbon (SOC) not only contributes importantly to soil fertility and productivity, but is a larger pool of C than the world's vegetation and atmosphere combined (Lal, 2004). Those facts suggest that SOC is a potential sink for atmospheric CO$_2$, especially in soils whose formerly high levels of SOC have become depleted through land use. Therefore, many benefits may accrue from quantitative research on spatial patterns of SOC stocks at scales ranging from landscapes down to individual experimental plots.

Among the many factors that affect those patterns are soil type, climate, topography, and vegetation biomass (Batjes, 2014; Jobbagy and Jackson, 2000). High spatial variations in SOC content can occur even at small scales (m) (Gessler et al., 2000; Chevallier et al., 2000) and such variations increase the uncertainty of comparisons among SOC stocks under different land-management practices (Costa Junior et al., 2013). In addition, most calculations of global-level C-budgets have not taken deep SOC into account

because SOC at those levels is not considered to contribute much to the exchange of C between soil and the atmosphere. However, authors are now paying increased attention to dynamics and estimated storage capacities of deep SOC that underlies different ecosystems (Bounouara et al., 2017; Cardinael et al., 2015; Mathieu et al., 2015; Rasse et al., 2006; Shi et al., 2013). In a global review for tropical regions, Batjes (2014) estimated SOC stocks in the 0-200 cm depth range at 616 to 640 PgC, versus 201-213 PgC in the 0-30 cm range. Attempting to decrease experimental uncertainty by measuring SOC at smaller spatial intervals can be impractical because of the time and expense of standard SOC analyses. Therefore, development of accurate, low-cost techniques are needed for quantifying SOC contents at the necessary spatial, horizontal and vertical, scales.

As a soil type on which to test those techniques, Andosols have three attractive traits: they store a disproportionate amount of soil carbon; their SOC patterns are spatially complex both vertically and horizontally (Mora et al., 2014); and their soil constituents associated with SOC in Andosols might be used as proxies for quantifying SOC contents via diffuse-reflectance mid-infrared spectroscopy (MIRS) (Kinoshita et al. 2016). Although Andosols represent only about 0.84% of the terrestrial soils, they store approximately 5% of the global soil C (Matus et al. 2014). Derived from volcanic material, they have high levels of short-range-order (SRO) minerals, *e.g.* Allophane or imogolite; high SOC concentrations (Batjes, 2014; Feller et al., 2001; Torn et al., 1997); high water retention; and low bulk densities (Shoji et al. 1996). They can store up to three times as much SOC as non-Andosols. Clear correlations have been found between SOC content and Allophane content (Basile-Doelsch et al., 2005) or Aluminium organic matter complexes (Al_p) (Percival et al., 2000; Shen et al., 2018). As explanations for the high SOC contents, most authors have posited that SOC in Andosols is stabilized against decomposition by some combination of (i) acidic condition; (ii) Al toxicity; (iii) SOC adsorption on the mineral surfaces (Mayer and Xing, 2001); (iv) complexation, precipitation, and formation of organo-metal (Al/Fe) complexes, also called Al/Fe humus complexes (Percival et al., 2000; Scheel et al., 2007; Torn et al., 1997); and (v) entrapment in the mesoporosity (Mayer, 1994) with a particular network structure (Chevallier et al., 2010; Mayer et al., 2004; McCarthy et al., 2008).

SOC in the deeper levels of Andosols has not been studied extensively. Volcanic regions with high percentages of Andosols (compared to the other soil types) showed some of the lowest degrees of vertical stratification of SOC stocks, but with a high degree of uncertainty. The ratio of SOC stocks at 0-30 cm those at 0-100 cm has been evaluated as 0.48, with a coefficient of variation of 29% (Batjes 2014). The SRO mineral distribution in volcanic regions can be complex both vertically and horizontally (Churchman et al. 2016) especially in humid areas where (i) an active volcano produces thin, intermittent ash deposits, (ii) soil erosion causes movement of soil materials (Zehetner et al., 2003), and (iii) the ash weathers on slopes in which zones from which Si is leached away (thereby enabling SRO minerals to form and persist) alternate with zones in which Si accumulates (thereby causing the formation of aluminosilicates like halloysite). Factors (ii) and (iii), especially, can combine to produce terrains in which older, SRO-depleted soils become overlain by newer, SRO-rich topsoils.

Previous work (Kinoshita et al., 2016) at the study site described in this article (a 1 km² volcanic micro-watershed) showed that spatial variations in organo-Al complex and allophane contents mainly explained the high spatial variation of SOC contents of topsoils (from 48 to 172 gC kg⁻¹soil at 0-5 cm depth). Kinoshita et al. did not sample deeper layers of soil at the site. However, Kinoshita et al.'s correlations between SRO and SOC, together with Mora et al.'s (2014) data on the potentially complex distributions of SRO minerals in volcanic soils, suggested that deep-soil SOC stocks may be not related to topsoil SOC stocks. Thus, we hypothesized that SOC stocks would be highly variable as SRO minerals content along soil profiles and that deep SOC stocks may be not correlated to topsoil SOC stocks. Testing those hypotheses required the analysis of hundreds of soil samples

from soil profiles (0-200 cm) at widely distributed locations within the site. The results provided a large database with which to test an additional hypothesis: that signatures of SRO minerals in MIR spectra of soil samples would be useful proxies to predict the type of soil material (Andic vs Halloysitic soil material), SOC and bulk density (Bd).

75   Several authors have shown that diffuse-reflectance MIR spectroscopy is a time- and cost-effective analysis to quantify SOC contents. Therefore, MIR spectroscopy has become increasingly popular for spatial mapping of SOC (Ben-Dor et al., 2009; Clairotte et al., 2016; Nocita et al., 2015; Visacarra Rossel et al., 2016). Especially in the MIR region, each of a soil's mineral constituents affects spectra in a characteristic way. For example, absorbance peaks of allophane and imogolite, two SRO minerals that are specific to Andosols are near 1000 cm^{-1}. In contrast, the absorbance peaks of polymerized silicates are near 350 cm^{-1}

80   (Parfitt, 2009). As SRO minerals control soil Bd (Shoji et al. 1996) and SOC content (Torn et al., 1997), SRO signatures in MIRS spectra may be useful proxies for soil Bd and SOC. Therefore, MIR spectroscopic analysis could replace soil extractions (Janik et al., 1998) and be used for classifying soil samples as Andosols or non-Andosols. Based upon those classifications, MIRS could be appropriate for spatial mapping of SROs and SOCs in volcanic areas where soil age, type, and andic properties vary. In summary, the three hypotheses tested were (1) Spatial distribution of SOC stocks at depths down to 200 cm can vary dramatically in volcanic

85   areas, even within a small watershed; (2) Surface SOC stocks in volcanic areas are not reliable predictors of stocks that might exist down to depths of 200 cm; and (3) MIRS is an effective and reliable technique for classifying soil materials according to some characteristics of andic soils associated with contents of SRO minerals.

2 Materials and Methods

2.1 Site description

90   Previous authors have studied the site and described it in detail (Kinoshita et al. 2016, Gómez-Delgado et al., 2011). It is a 0.9 km^2 watershed located in the Central-Caribbean area of Costa Rica. More specifically, it is located within the Aquiares Coffee Farm (83° 43' 35"W; 9° 56' 35" N), in the Reventazón river basin on the slope of the Turrialba volcano. The site's elevation ranges from 1020 to 1280 m a.s.l., and the mean slope is 11.31°. Mean annual precipitation was 3014 mm between 1973 and 2009. The climate is tropical humid without a dry season; the mean annual temperature is 19.5°C. According to Mora-Chinchilla (2000), volcanic

95   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. The Turrialba deposits are mainly andesitic volcanic ashes. Indications of lava flows, agglomerates, lahars, and ashes are also present. Thin ash deposits still occur regularly; the last was in mid-2016. The site's soils are classified as Andosols (Payan et al., 2009), with a pH 5 at the surface and 6 below 1 m (data not shown).

Before the introduction of the coffee in 1975, the watershed's lands were occupied by housing and gardens-including a cardamom

100   plantation-all of which had replaced a pristine forest. Since 1975, the vegetation has consisted of coffee trees (*Coffea arabica L.*, var Caturra, 6,300 trees ha^{-1}) and *Erythrina poeppigiana* leguminous shade trees at a density of 7.4 trees ha^{-1}. The Aquiares farm is intensively managed, with regular pruning and applications of fertilizer (214 ± 44 kg N ha^{-1} yr^{-1}). Its management of pests and weeds complies with Rainforest Alliance^{TM} standards. Coffee yields were 1375 ± 341 kg green coffee ha^{-1} yr^{-1} (1994-2011).

2.2 Soil sampling

2.2.1 Soil-profile samples for MIRS and chemical analyses

Taking into account Kinoshita et al.'s (2016) data on the spatial variation of SOC contents in the study site's topsoils, we chose 69 sampling points, all of them within inter-row of coffee plants (Fig. 1). At each location, after removing surface plant residues, we used a hand-driven, 5-cm-diameter steel auger to extract soil profile down to the lithic contact or 200 cm. Each soil profile thus consisted of four to ten 20-cm-thick soil samples, depending on soil depth at each location. The soil samples (a total of 598) were oven-dried at 40°C for 72 hours and ground to <200 µm for MIRS and chemical analyses.

2.2.2 Samples for Bd determination

Because digging soil pits for bulk density is tedious, and because the spatial variability of Bd may be less than that of SOC contents, at least in non-Andosols (Don et al., 2007), we collected samples for Bd determination at only seven of the 69 locations identified in Sect. 2.2.1. The seven locations were chosen because of their locations and the spectral characteristics of their corresponding MIRS samples (See Sect. 2.4.1.) For logistical reasons, the soil pits for bulk density could not be dug at exactly the same points where MIRS samples had been obtained. However, every pit was dug at a distance of less than 100 cm from the corresponding MIRS-sample location. Each pit was dug at depth intervals of 20 cm to a depth of 200 cm. Four replicates were sampled from each soil layer of each pit.

2.3 Acquisition and pre-processing of MIR spectra

For each of the 598 soil samples, we acquired a reflectance spectrum in the MIR region at 934 wavenumbers between 4000 and 400 cm^{-1}, at 3.86 cm^{-1} intervals. The spectrophotometer (a Fourier transform Nicolet 6700, by Thermo Fischer Scientific, Madison, WI, USA) was equipped with a silicon carbide source, a Michelson interferometer as dispersive element, and a deuterated triglycine sulphate detector. Soil samples were placed in a 17-well plate, then scanned using an auto-sampler (soil surface area scanned: ca. 10 mm²). The resulting MIR spectrum for each sample was the sum of 32 co-added scans. The body of the plate was used as reference standard; it was scanned once per plate. Reflectance was converted to absorbance. Twenty wavenumbers were removed due to noisy spectra. The data set that resulted from acquisition and pre-processing consisted of 598 mean MIR spectra.

2.4 Laboratory soil analysis

2.4.1 Chemical analyses

Of the 69 soil profiles described in Sect. 2.2.1, we selected ten spatially dispersed 200-cm soil profiles (comprising a total of 98 soil samples), as being representative of the total sample set from a spectral viewpoint (Fig. 2A). The 98 soil samples were analysed for SOC contents and extractable Al, Si, and Fe. The SOC contents were determined with a CHN elemental analyser (Carlo Erba NA 2000, Milan, Italy). Al, Si, and Fe associated with active amorphous constituents and organo-metal ligands (Al_o, Si_o, and Fe_o) were extracted with ammonium oxalate solution (Blakemore et al., 1981) as follows. Soil samples (0.5 g) were shaken for 4 h in the dark in 50 mL of a solution of 0.2 M ammonium oxalate and 0.2 M oxalic acid at pH 3. Then, after centrifuging and filtration, we measured the Al_o, Si_o, and Fe_o in the filtrate by ICP-AES.

Al and Fe associated with soil organic matter (Al_p and Fe_p) were estimated by extraction with sodium pyrophosphate solution (Blakemore et al., 1981). We caution that pyrophosphate has been reported to extract different forms of Fe, some of which are not specifically related to organo-Fe complexes (Parfitt and Childs, 1988). For this analysis, soil samples (0.5 g) were shaken for 16 h in 50 mL of a 0.1 M sodium pyrophosphate solution at pH 10. After centrifuging and filtration, sodium-pyrophosphate-extractable Al_p and Fe_p in the filtrate were determined by ICP-AES.

The data set that resulted from this task consisted of the 98 samples' respective contents of SOC; Si_o; Al_o and Al_p; and Fe_o and Fe_p.

2.4.3 Bulk density (Bd)

Bulk densities were determined according to the soil-core method, using a bevelled cylinder (98 mL). The four replicates (Sect. 2.2.2) of each soil sample were oven-dried for 48 h at 105 °C and sieved at 2 mm to remove coarse fragments such as stones and living roots. Those fragments represented less than 1% of each soil sample mass, and therefore were considered negligible. The Bd of each soil sample was calculated as the average of the Bds of the sample's four replicates.

The data set that resulted from this task consisted of 66 Bd values. One value for each of the ten soil depths at the seven sampling locations, except where the sampling pit could not be dug to the full 200-cm depth.

2.5 Development of models relating MIR spectra to Bd, and contents of Al_o, Al_p, Si_o, Fe_o, and SOC

2.5.1 Models for contents of Al_o, Al_p, Si_o, Fe_o, and SOC

From the data sets for MIRS (Sect. 2.3) and laboratory analyses (Sect. 2.4.1), we developed predictive models for contents of Al_o, Al_p, Si_o, Fe_o, and SOC. One model was developed for each constituent, for a total of five models. All of the models were based upon 69 samples, from seven soil profiles, that were common to both data sets. The other 29 samples that were common to both data sets were used for validating the models, as described below. The models were developed by fitting the samples' MIR absorbance spectra (with no mathematical pre-treatments) to the samples' measured contents of each of the five constituents. Fitting was done via modified partial least-squares regressions. The accuracy of each prediction model was determined by external validation using laboratory analyses and MIRS spectra of the above-mentioned 29 samples, which were from three different soil profiles. The accuracy was quantified by computing (i) the coefficient of determination (R^2), (ii) the root mean square error (RMSE) between predicted and measured values, and (iii) the ratio (denoted as RPD) of the standard deviation of the value set to RMSE.

We also tested four additional calibration models for predicting SOC contents. These models were based on two spectral classes and two conventionally measured data classes. One model was built for each combination of spectral class and soil-type class. The two spectral classes and two soil-type classes were defined via two analyses. The first was a principal-component analysis (PCA) and a K-means clustering analysis of the 598 soil spectra. The second was a principal-component analysis (PCA) and a K-means clustering analysis of the 98 samples conventionally analysed data: Al_o, Al_p, Si_o, Fe_o, $Al_o+0.5\ Fe_o$, Allophane, and $(Al_o-Al_p)/Si_o$ (Terra et al., 2018). Because these models were built on a small number of soil samples (about 30 or 60 soil samples in one cluster), the accuracy of the prediction models was determined by cross validation, using the leave-one-out method.

2.5.2 Model for Bd

This model was based upon the 66 measured Bd values (Sect. 2.4.3), and the MIR absorbance spectra of soils from the seven corresponding sites. To build the model, we used modified least-squares regressions to fit the MIR absorbance spectra to the Bd values. Our Bd-prediction model (built on spectral signatures of soil samples passed through a 200 µm-sieve) did not pretend to predict Bd *sensu stricto*, which is a physical property determined by soil constituents and their arrangements; *i.e.* the soil structure. Instead, it used specific vibrational processes of the SRO minerals and organo-metal complexes as proxies for predicting Bd. Because only 66 Bd were measured, we could not build two different models based on clusters (as we had done for SOC contents). Hence, we assessed the accuracy of the Bd-prediction model by cross validation, using the leave-one-out method.

2.6 Data analysis

2.6.1 Data calculations

In the case of samples whose contents of Al_o, Al_p, Si_o, Fe_o, and SOC, and Bds had not been measured analytically, we estimated those contents from the samples' MIR spectra, using the models described in Sect. 2.5. Then for each of the 598 soil samples, we calculated:

- The degree of weathering of the volcanic glass, calculated as $Al_o + 1/2\ Fe_o$, with Al_o and Fe_o in grams per 100 g of soil (Shoji et al., 1996)
- Grams of Allophane per 100 g of soil, calculated as $100\ Si_o\ /[23.4 - 5.1(Al_o - Al_p) \div Si_o]$ with Si_o, Al_o, and Al_p given in grams per 100 g of soil (Mizota and Van Reewijk, 1989),
- The atomic Al:Si ratio in the soil aluminosilicates, calculated as $Al:Si = (Al_o - Al_p)/Si_o$, with Si_o, Al_o, and Al_p given in grams per 100 g of soil (Parfitt and Wilson, 1985 quoted in Parfitt and Childs, 1988). A ratio of Al:Si about 2 is characteristic of imogolite, proto-imogolite, and Al-rich Allophane. Ratios lower than 2 indicate Si-rich Allophane and the presence of significant amounts of crystallized minerals (Levard et al., 2012).
- SOC stock (kgC per m^{-2}), calculated as $SOC\ content\ (gkg^{-1}) \times Bd\ (gcm^{-3}) \times sample\ thickness\ (cm) \div 100$ with "sample thickness" which refers to the 20-cm thickness of each of the 598 soil samples.

2.6.2 Statistical analyses

Effects of spectral cluster and soil depth on SRO minerals, organo-metal complexes and SOC contents, Bd, and SOC stocks were analysed with linear mixed models that considered soil profile a random effect. The T-test was used to assess the effect of spectral cluster on the variable to be explained for each depth. A random forest regression model was used to evaluate and order the importance of SRO minerals and organic-metal complexes, spectral cluster, and soil depth on SOC contents. Random forest is a machine-learning technique based on randomly built decision trees. At each node, a subset of covariates is also randomly chosen. Random forest was used instead of multiple-linear-regression methods because it allows use of both categorical and numeric covariates, collinearity between covariates, and non-linear relationships between covariates and the variable to be explained. %IncMSE was used to assess the relative importance of covariates in explaining variability of SOC content. For a given covariate, %IncMSE is the difference between mean standard error (MSE) of the model with permutation of this covariate, and model without that permutation. The larger the %IncMSE, the more important this covariate in predicting SOC content. We used R software (R Core Team, 2018) for the statistical analyses. Results were given for the 598 soil samples, but statistics on the conventionally-measured variables can be found in the supplementary materials (S1, S2, S3).

2.6.3 Spatial maps

Golden Software Surfer V 8.0 was used for spatial mapping of the predicted SOC stocks, Allophane contents, and thickness of
205 *allophanic* soil material layer (see Sect. 3.1.2 for definition). The Digital elevation model (DEM) used in some of the mapping was
created from a vertical-resolution, digital-terrain model obtained from the TERRA-1998 project (scale $\approx$ 1: 25000; CENIGA,
1998). The vertical and vertical resolutions of that model were 5 m and 10 m respectively.

3 Results and discussion

3.1 Extractable Al, Si, Fe

210 **3.1.1 Prediction of extractable Al, Si, Fe contents (SRO minerals and organo-metal complexes)**

The five models for predicting contents of extractable Al, Si, and Fe from MIR spectra gave accurate results for the whole set of
soil samples (Table 1) with RPD>2 (Chang et al., 2001). Our attempts to develop a prediction model for Fe_p content proved
unsuccessful, probably because pyrophosphate extraction does not discriminate sufficiently among forms of Fe (Parfitt and Childs,
1988). The model for predicting Al_p contents from MIR spectra gave negative values (mean= -0.19 $\pm$ 0.08 g Al_p 100 g^{-1} soil) for
215 nine samples, all of which were among the 500 that were not analysed conventionally. Most of those nine samples were from two
specific soil profiles (i.e., two locations). By convention, we set the Al_p contents of those nine to zero when calculating their other
soil characteristics. All other predicted contents of extractable Al, Si and Fe contents were positive.

In previous published studies, Soriano-Disla et al. (2014) found that contents of oxalate-extractable forms could be predicted from
MIR spectra, but Misnany et al. (2009) found otherwise. Our results indicated that MIR spectroscopy seems to be a promising tool
220 for predicting contents of extractable Al, Si, and Fe in volcanic soils, and thus contents of SRO minerals and organo-Al complexes,
providing that the specific prediction models are based upon conventional analyses of a large set of representative samples. That
was likely due to the specific peaks of imogolite and allophane in the MIR range (Parfitt, 2009). However, further studies are
needed, especially on pyrophosphate-extractable elements.

3.1.2 Clustering according to soil type based upon MIR spectra, versus conventional analyses

225 MIR spectra for the 598 soil samples formed two distinct clusters ("blue" and "orange" in Fig2B), as did the conventionally
analysed data of the 98 samples ("circle" and "square" in Fig 2B). The data from the measured cluster represented by circle showed
andic properties and was called *Allophanic* measured cluster by comparison to the other measured cluster represented by square
and called *Halloysitic* measured cluster. A comparison of the clusterings of the 98 samples that were analysed both conventionally
and by MIRS is revealing. For the most part, samples whose MIR spectra are in the "blue" spectral cluster (see Fig. 2B) fell in the
230 *Allophanic* measured cluster, while samples whose spectra are in the "orange" spectral cluster fell in the *Halloysitic* measured
cluster. That correlation held true for all but 18 of the 98 samples. The conventionally analysed data of all 18 of those samples are
in the measured *Allophanic* cluster, but their spectra fall in the "orange" spectral cluster. In Fig. 2B, those samples are identifiable
as the ones whose andic-property symbol is a circle that surrounds an "orange" spectral symbol. All of those occurrences along the
border between the two spectral clusters reveal that the *Allophanic* and *Halloysitic* measured clusters overlap slightly. This study
235 of the clusterings shows that MIR spectra depended upon some of the characteristics of andic soil samples, and also shows a
graduation of the two soil materials. Because of the above-described strong correlations between spectral clusters and measured

clustered, we will denote the orange spectral cluster as the *Halloysitic* spectral cluster, and the blue spectral cluster as the *Allophanic* spectral cluster from this point onward.

Further analyses confirmed that the variations in andic characteristics of all 598 soil samples were essentially explained by the spectral clustering and secondarily by the soil depth. The two spectral classifications (*Allophanic* and *Halloysitic*) were powerful for organising and describing the two categories of soil materials (Table 2). *Allophanic* soil materials, corresponding to the soils in the *allophanic* spectral cluster, rich in organo-Al complexes (0.42 ± 0.12 g Al_p 100 g^{-1} soil), and SRO minerals: allophane (15.8 ± 4.4 g allophane 100 g^{-1} soil), $Al_o+0.5$ Fe_o (5.3 ± 1.2 g $100g^{-1}$ soil) and amorphous Al, Si and Fe (4.5 ± 1.1 g Al_o 100 g^{-1} soil, 1.6 ± 0.4 g Si_o 100 g^{-1} soil, 1.5 ± 0.2 g Fe_o 100 g^{-1} soil). These soils have Al_p:Al_o ratios of about 1.0 ± 0.4, and Al:Si ratios of about 2.6 ± 0.2. In contrast, *halloysitic* soil materials, corresponding to the soils in the *halloysitic* spectral cluster are poor in organo-Al complexes (0.18 ± 0.11 g Al_p 100 g^{-1} soil), and SRO minerals: allophane (5.3 ± 2.9 g allophane 100 g^{-1} soil), $Al_o+0.5$ Fe_o (1.9 ± 0.9 g $100g^{-1}$ soil), and amorphous Al, Si and Fe (1.5 ± 0.8 g Al_o 100 g^{-1} soil, 0.7 ± 0.3 g Si_o 100 g^{-1} soil, 0.9 ± 0.3 g Fe_o kg^{-1} soil). Their Al_p:Al_o ratios are highly variable (about 1.7 ± 1.5), and their Al:Si ratios are about 1.7 ± 0.6. In a given cluster, the Al_p and the Fe_o content, the Al_p:Al_o ratio decrease with depth, but not the allophane content, Al:Si ratio, nor the quantity $Al_o+0.5$ Fe_o.

Chemical extraction provides no structural information, but it does help distinguish between crystalline and SRO minerals. The literature (Levard et al., 2012) notes that SRO minerals like Al-rich allophanes, proto-imogolite, and imogolite predominate in volcanic-ash soils with $Al_o+0.5Fe_o \gg 2\%$ and Al:Si ratios >2, such as in the *Allophanic* cluster. Surprisingly, and especially in the surface soil, ratios of Al_p to Al_o are greater than 1 with high allophane contents (from 5 to 20 g allophane per 100 g soil), indicating a co-existence of allophane and Al-organo complexes; which may be due to some combination of the soil pH (approximately 5; *i.e.* near the boundary between allophane and Al/Fe-organo complexes) and the regular inputs of organic materials and ashes from the surface (Mizota and Van Reewijk, 1989). On the contrary in the *Halloysitic* cluster, a lower Al:Si ratio should sign a high Si activity in solution and likely formation of halloysite (Parfitt et al., 1997). However, in both types of soil materials, the high Al_p:Al_o ratio revealed that the majority of active Al groups were present mainly as organo-Al complexes, especially near the surface.

Even though the soil reflectance and andic characteristics varied continuously through soil samples (Fig. 2B), we observed that a two-cluster classification of MIR spectra was powerful for describing the horizontal and vertical variation of andic soil properties. However, MIR does not provide such definitive results for soils in other contexts and at a global scale (Viscara Rossel et al., 2016) because most of the world's soil types are complex mixtures of materials from diverse origins, and because the criteria for global soil taxonomic classifications include parameters with no direct influence upon spectral character. When applied to soil materials sharing the same weathering and pedogenesis processes, vis-NIR spectra (Terra et al., 2018) or MIR spectra seem to be useful tools for soil survey and classification. In volcanic areas where strong variations in mineralogy occur (e.g. SRO minerals may dehydrate to form phyllosilicates), MIR spectra seem to be especially well adapted. These difference in mineralogy and proportion of SRO minerals in the *Allophanic* and *Halloysitic* materials cause changes MIR spectra in reflectance intensity and absorption features in particular at the bands near 3620 cm^{-1} and 3700 cm^{-1}, which are characteristic of halloysite (Hidalgo et al., 2010; Fig. 2C). Hence, the effects of mineral weathering or pedogenesis pathways upon MIR spectra of volcanic-derived soil samples were clear enough for building two different soil-material clusters (*allophanic* vs. *halloysitic*) using PCA and K-means algorithms based upon spectral data distance.

3.1.3 Three types of soil profiles in the watershed

Both types of material were encountered at every soil depth, but *allophanic* materials predominated near the surface (n=57 on 68 samples at 0-20 cm), while *halloysitic* material predominated at depth (n=24 on 42 soil samples at 180-200 cm) (Table 2). These
275 results confirmed the common observation that allophane is associated with halloysite (Ross and Kerr, 1934 quoted in Parfitt 2009). Both minerals may be derived directly from volcanic glass and halloysite can also form via weathering of Allophane (Parfitt, 2009; Torn et al., 1997; Wada, 1989). The allophane: halloysite ratio depends upon the chemistry and regularity of ash depositions as well as upon the concentration of Si in solutions which depend on precipitation and drainage (Churchman et al. 2016).

Of the 69 sampled soil profiles (0-200 cm), 35 were denoted as "*allophanic* profiles", *i.e.* every sample's spectrum fell in the
280 *allophanic* cluster. The soil was composed of *allophanic* materials down to 200 cm. The samples from eight profiles (the "*halloysiti*c profiles") were exclusively of the *halloysitic* spectral class, and 26 soil profiles (the *allophanic-halloysitic* profiles) contained soil of both types (Fig. 3).

Of the 26 *allophanic-halloysitic* profiles, 20 had soils of the *allophanic* spectral class at surface, and soils of the *halloysitic* spectral class at depth. In five of the other *allophanic-halloysitic* profiles, no clear relationship between depth and spectral class was present.
285 The thickness of the *allophanic* material at sites with *allophanic-halloysitic* soil profiles was taken as being equal to the depth at which *halloysitic* material first appeared in the soil profile. Calculated in this way, the thickness of *allophanic* material varied from 0 cm in the *halloysitic* soil profiles to 200 cm in the *allophanic* soil profiles (Fig. 4a).

In this micro-watershed, allophanes predominated over halloysite because total annual rainfall was above 3000 mm (Parfitt et al. 1983) and because the study site does not experience a pronounced dry season. Therefore, a high leaching rates (Gomez-Delgado
290 et al., 2011; Benegas et al., 2014; Welsh et al., 2018) of Si probably resulted in the formation of materials in which Al groups predominate (Al:Si>> 2). A majority of soil profiles (35) were indeed rich in SRO minerals and organo-mineral (Al/Fe) complexes down to 200 cm. In 26 other profiles, contents of SRO minerals and amorphous materials were also present but decreased with depth (Fig. 3). In these *allophanic-halloysitic* soil profiles, allophane contents (10 g 100 g^{-1} soil) were intermediate between the allophane contents in *allophanic* soil profiles (> 15 g 100 g^{-1} soil) and in *halloysitic* soil profiles (< 5 g 100 g^{-1} soil) (Fig. 3). This
295 coexistence of allophane and halloysite marked a transition from the *allophanic* soils to the *halloysitic* soils. This transition could be explained by formation of halloysite at microsites within an allophane matrix (Aomine & Wada, 1962 quoted in Parfitt 2009), when periods of leaching alternate with periods of desiccation (Churchman et al., 2016). The resulting fluctuations in activity of Si in soil solutions could cause differential formation of amorphous and crystalline weathering products. A transition zone attributable to that phenomenon has been observed in pedons along the slopes of volcanoes in Ecuador at intermediate altitudes
300 (Zehetner et al., 2003). However, the variations in the soil profile types observed in the present study occurred over much shorter distances than in Zehetner's study. Only 20% of the soil profile in the watershed were of the *halloysitic* type, in which halloysite minerals predominated at all depths. Halloysite could be the result of an older soil development from allophane weathering on older tephras, or the result of silica-rich environments, such in buried soil layers that receive silica from overlying soil layer, or in zones where restricted drainage prevents removal of Si via leaching (Churchman et al. 2016). This is likely to occur in Si-rich,
305 rhyolitic tephras than in basaltic ones (Dahlgren et al., 2004). Even if in the last millennia, most of the volcanic deposits at Turrialba have been andesitic ashes that varied little in composition over time (Meijer and Buurman, 2003), recent investigations found rhyolitic materials and multistage magmas in which rhyolite is mixed with end members of basaltic andesite (Devitre et al., 2018).

As climate, paleoclimate, and composition of volcanic ejecta were identical over the studied 1-km^2 micro-watershed, we assumed that the thickness and composition of different soil materials likely resulted from an integration through time of several pedogeomorphic processes: (i) regular new ash deposits, (ii) deeper weathering of parent materials, (iii) hydrological dynamics (i.e. Si leaching or accumulation of Si), and (iv) soil erosion (Gessler et al., 2000), i.e. in unstable landscape positions, the erosion of topsoil could have removed *allophanic* soil materials and exposed the older *halloysitic* subsoils (Zehetner et al., 2003). The influence of the site's topography upon hydrology, pedogenesis, and topsoil erosion (Fig. 4a) superimposed upon the general weathering trend, and could contribute to explaining the spatial variation of the different soil profiles and *allophanic* material thickness in the micro-watershed.

3.2 SOC contents

3.2.1 Prediction of SOC contents from MIR spectra

The model built on the whole data set yielded better results (higher R^2, higher RPD and smaller RMSE) than those using two soil groups separately (Table 3). The SOC prediction have been considered accurate according to the usual criteria (Chang et al., 2001), and are comparable in performance to those developed by McDowell et al (2012) for Hawaiian soil samples. However, this model predicted negative SOC contents for 12 samples, which all belonged to the *halloysitic* spectral class. Negative predicted SOC contents have been seen previously in an Andosol dataset in Mc Dowell et al. (2012), who noted difficulties in predicting low SOC contents accurately. Those difficulties could be explained by the presence of different soil materials, with their own properties to stabilize SOC. Consequently, we preferred the SOC-content predictions made by the models that were built on soils clusters that were differentiated by their MIRS spectra and thus their type of soil material. For these models, which did not predict any negative SOC contents, the predictions of low SOC contents were better with a decreasing RMSE from 5 to 2 gC kg^{-1} soil for samples in the *halloysitic* cluster (Table 3).

3.2.2 SRO minerals and metal-organo complexes explain the variation of SOC contents

The SOC contents were explained by soil depth and by clustering based on soil MIR spectra. SOC contents decreased significantly with soil depth, and were much larger in *allophanic* than in *halloysitic* spectral classes at all depths (Table 4). As already and commonly observed, the more SRO minerals that are present, the more SOC is concentrated and preserved in soils (Fig. 5, 6; e.g. Torn et al. 1997; Chevallier et al. 2010). The concentration of SOC in the *allophanic* materials of this study were similar to those observed in the pedogenetic A horizons of Andosols from Costa Rica (Meijer and Buurman, 2003), Ecuador (Zehetner et al. 2003), Martinique (Chevallier et al. 2010), and Hawai (Torn et al. 1997).

The strongest correlations were between Fe$_o$, or Al$_p$, and SOC, suggesting both (i) an important role of organo-metal complexes in SOC stabilization, but also (ii) a stronger impact of SRO mineral with Fe on SOC contents at all depths (Fig. 5) than was expected from Kinoshita et al.'s (2016) research on the surface soil (0-5 cm) at the same site. Ferrihydrite is known to be highly involved in the stabilization of SOC in Andosol (Kleber et al. 2005; Matus et al. 2014; Filimonova et al. 2016; Parfitt et al. 1997). In our study, we did not explicitly analyse the form of Fe extractable by oxalate, whether ferrihydrite dominated Fe forms or not, but the preservation of SOC related to Fe$_o$, even in relatively small amounts (0.9-1.3 g Fe$_o$ 100g^{-1} soil), seemed to be the major factor explaining SOC-content variations at the surface and at depth (Fig. 5 and Fig. 6). In addition, the relation between SOC and Fe$_o$ appeared to have a threshold near 1.3 g Fe$_o$ 100g^{-1} soil. Beyond that threshold, all soil samples were in the *allophanic* cluster, SOC

were high and very sensitive to small variations in Fe_o content (SOC content = 82 Fe_o – 66, r^2= 0.73). In contrast, below that threshold SOC were lower and less sensitive to Fe_o content (SOC content = 24 Fe_o – 5, r^2= 0.85). However, our results did not provide analysis for determining whether the form of Fe_o present in ALL materials differed from that in *halloysitic* materials.

The organo-Al complexes, represented by Al_p, were the second most important soil variable in explaining the variation of SOC (Fig. 6). The importance of Al_p on SOC stabilization has already been noted in literature (Beare et al., 2014; Huygens et al., 2005), although the organo-Al stabilization is not fully understood (Takahashi and Dahlgren, 2016). The most given explanations were: Al toxicity for microorganisms, or electrostatic sorption between Al and SOC which limits the accessibility of SOC to microorganisms or to their enzymes. The organo-Al complexes were observed to be more effective than allophane in protecting SOC from degradation (Boudot, 1992; Powers and Schlesinger, 2002), especially in the surface soil layers (Fig. 6) where contents of organo-Al complexes and the $Al_p:Al_o$ ratios tended to be high (Fig. 3). We thus assumed that the majority of active Al groups should pertain to organo-Al complexes, which stabilized the vegetation organic inputs.

To a lesser extent, the SOC contents were explained by calculated parameters linked to soil andic properties, the ratio $(Al_o-Al_p)/Si_o$, allophane, $Al_o + 0.5 Fe_o$, and to soil depth (Fig. 5, 6; Table 4). The mineral matrix explained the SOC stabilization even better at depth than at the surface (Fig. 6c). The relationship between allophane or $Al_o + 0.5 Fe_o$, and SOC was mediated by soil depth. For a given amount of allophane, the soil was richer in SOC at the surface than at depth. In surface samples, C inputs by vegetation likely explained the higher SOC contents as well as the slightly greater variation of that content (Table 4; Matus et al., 2014). Nevertheless, the SOC contents are reported to be poorly correlated with above-ground biomass in Andosols (Noponen et al., 2013) and mainly controlled by contents of organo- Al complexes and SRO minerals even in the topsoil (Kinoshita et al., 2016). Furthermore, the decrease in SOC content with depth in *allophanic* cluster was much less noticeable than in other studies on Andosols. The SOC contents reported here decreased slightly and smoothly with depth throughout the soil profile (Fig. 3), and still remained high (40 gC kg^{-1} soil) at the 200-cm depth. The ratio $(Al_o-Al_p)/Si_o$ displayed a threshold near 2.5 (Fig.6) similarly to that was noticed for Fe_o at 1.3 g Fe_o 100g^{-1} soil. Beyond that threshold, all soil samples were in *allophanic* clusters and SOC was very sensitive to a small variation in $(Al_o-Al_p)/Si_o$ ratio.

As the volcanic ejecta at the study site contained little carbon (2 to 3 mgC g^{-1} ash, data not shown), the SOC-content distribution in the watershed was controlled mainly by vegetation inputs with pedogenic processes. The dominant pedogenic processes could be the development of an A horizon by andosolization, which consists of the stabilization of accumulations of vegetation-derived organic matter by active Al and Fe, in parallel with rapid formation of SRO minerals from volcanic ejecta. Rapid weathering of the ejecta released Si, Al, and Fe faster than crystalline minerals could form. Complexation of humic/colloidal organic substances with metal to form organo-metal (Al/Fe) complexes was especially noticeable at the soil surface. Preferential precipitations of metastable SRO minerals (allophane, imogolite, ferrihydrite) occurred in the subsoil (Ugolini and Dahlgren, 2002, Kramer et al., 2012). The permanent humid weather and the regular ash deposition favoured these mechanisms (Ugolini and Dahlgren, 2002; Mora et al., 2014) and the stabilization of SOC (Buurman et al. 2007; Chevallier et al., 2010). Regular burial of soil by volcanic ashes would explain why some locations have a very thick layer of *allophanic* material, assimilated to a young pedogenetic Andosol A horizon rich in SOC. Over time, or in specific locations where Si could accumulated, halloysite could form at the expense of SRO minerals. Formation of halloysite with lower surface area and charge density would cause loss of stabilized OC (Torn et al., 1997). Our results showed that deep-soil carbon contents as well as surface-soil carbon contents were essentially driven by the

type of soil material and contents of SRO minerals. The combination of soil development and mineralogy was a powerful factor
380   for explaining SOC content, regardless of the soil depth.

3.3 SOC stocks

3.3.1 Predictions of Bd

The prediction model for Bd (Sect. 2.5.2) gave accurate results, with a RMSE of about 0.09 and a average Bd of 0.8 g cm^{-1} (Table
5). This level of performance was close that of Cambou et al.'s (2016) predictions of Bd from near-infrared spectroscopy of
385   Luvisols. Bd is determined not only by soil constituents but also by soil structure, which has no effect upon the MIR spectral
signature of a soil sample that has been dried and sieved to <200 µm. Nevertheless, the MIR spectra of such a sample could contain
enough information to predict Bd. Here, andic properties (e.g. $Al_o + 0.5\ Fe_o$), which are well predicted by MIRS was known to
control Bd strongly (Shoji et al. 1996). Therefore, these well predicted properties were used as proxy for Bd prediction. The MIR
spectroscopy of soil samples prepared for soil analysis seemed to offer promise for predicting Bd of Andosols.

390   The Bd increased slightly with soil depth: from 0.65 to 0.75 in *allophanic* soil materials, and from 0.87 to 0.98 in *halloysitic* soil
materials. As already and commonly observed, soil horizons with SRO minerals have much lower Bd than soil with crystalline
clays (e.g. Mora et al. 2014) independently of soil depth (Table 6).

3.3.2 Variation of SOC stocks in the Aquiares watershed

Most of the ecosystem C in the watershed is in the soil (from 24 to 72 kgC m^{-2}) compared to the above-ground biomass of coffee
395   and *Erythrina poeppigiana* of 2.8 ± 0.2 kgC m^{-2} (Charbonnier et al., 2017), or the below-ground biomass of coffee down to 400
cm at 0.9 kgC m^{-2} (Defrenet et al., 2016). The SOC stocks were highly variable in both dimensions, horizontal and vertical along
the soil depth (Fig. 7). The relative distribution of SOC stocks with depth, *i.e.* the ratio of SOC stocks at 0-100 cm to SOC stocks
at 0-200 cm, also varied and was lower in the *allophanic* (0.56 ± 0.03) than in the *halloysitic* (0.75 ± 0.10) soil profile type. Our
results were consistent with most other studies (e.g. Mora et al. 2014; Batjes 2014), the *allophanic* soil materials with high contents
400   of SRO minerals had much larger SOC stocks, at every soil depth, than did *halloysitic* soil materials with crystalline clays (Fig. 7).
Thus SOC stocks in the upper 200 cm of soil varied markedly among the three soil profile types defined in Sect. 3.1.3. The range
of calculated stocks for *halloysitic* soil profiles was 24.5 ± 0.5 kgC m^{-2}, compared to 49.9 ± 1.8 kgC m^{-2} for *allophanic - halloysitic*
soil profiles and 72.4 ± 2.0 kgC m^{-2} for *allophanic* soil profiles (Fig. 7). The high SOC stocks in *allophanic* soil profiles confirmed
that large stocks can be present under plantations at moderate altitudes (1000 m a.s.l.), possibly as a result of organic-material
405   accumulation from previous land cover (pristine forest, then households) along with andosolization. Larger SOC stocks were also
measured in deep, homogeneous volcanic-ash soils of Andean ecosystems in Ecuador under upper montane forest and high altitude
paramo (87 ± 12 kgC m^{-2} within the 0-200 cm depth, Tonneijck et al., 2010), as well as in young Hydric Andosols derived from
recent volcanic ash (Poulenard et al., 2003). Nevertheless, when Al_o-Al_p/Si_o ratio > 2.5 and Fe_o > 1.3 g 100g^{-1} soil (see Sect. 3.2.2.),
the SOC stocks were high with high variations (50 and 74 kgC m^{-2}) which remained unexplained.

410   Globally, the distribution of SOC stocks down to 200 cm was based on the distribution of the soil profile type in the watershed and
could not be predicted from the SOC stocks at surface. For *halloysitic* soil profiles, we observed a linear relationship between SOC
stock in 0-20 cm and SOC stock in 0-200 cm as already reported for soils with different texture but without SRO minerals

(Andriamananjara et al. 2016). However, this relationship did not hold in *allophanic* and *allophanic - halloysitic* soil profiles (Figure 8a). The SOC stocks down to 200 cm could not be easily predicted from the SOC stock at the surface because of the discontinuity in the type of soil material in horizontal and vertical (*i.e. allophanic-halloysitic* soil profile) dimensions.

The thickness of the *allophanic* soil material layer was a better predictor of SOC stock in 0-200 cm than the SOC stocks in the topsoil (Figure 8b) because the SOC stocks varied only weakly within this layer and is especially high. Thus, the distribution of this layer in the watershed explained the distribution of SOC stocks (0-200 cm). Thick *allophanic* soil material and high SOC stocks were on gentle slopes (Fig. 4; Mora et al., 2014), while outcroppings of *halloysitic* soil material and small SOC stocks were found where slopes changed abruptly (Fig. 4). We thus hypothesized that erosion and regular ash deposition moulded the watershed's landscape, the thickness of a young andic A horizon, and the distribution of SOC within the upper two meters of that soil. Pedogeomorphic processes could explain the high variation of SOC stocks at fine-scales within the landscape.

4 Conclusion

The SOC stocks of the studied 1 km^2 watershed were much larger than the SOC stocks stored in vegetation biomass. The large variation of SOC stocks in the 0-200 cm soil profiles at fine scales (m) was related to the horizontal and vertical variations in SRO mineral contents. Knowing surface SOC stocks provided little information about deep SOC stocks, whereas SOC stocks within the 0-200 cm depth were positively correlated with thickness of a *allophanic* soil material layer, likely a young andic A horizon. In this volcanic environment, where soil minerology varies greatly, MIR spectroscopy was confirmed as a convenient and accurate tool for classifying soil materials as either andic or non-andic. As a result, MIR spectroscopy also proved useful for predicting SOC contents, bulk density, and SOC stocks for a large set of samples. The main conclusion of our study is that ignoring topography and soil pedogenesis would introduce a serious weakness in current approaches for evaluating regional C stocks, especially in young volcanic areas where mineralogy controls SOC stocks and varies strongly on a fine spatial scale. Improved knowledge of pedogenesis is thus needed for understanding the variability and distribution of SOC stocks.

5 Author contribution

O. Roupsard and R. Kinoshita designed the experiments. F. Guidat, E de Melo Viginio Filho, and P Lehner carried them out. A. Albrecht performed the lab analyses and the MIRS analyses. K. Fujisaki performed the statistics. T. Chevallier prepared the manuscript with contributions from all co-authors.

6 Competing interests

The authors declare that they have no conflict of interest.

7 Acknowledgements

Our study was part of the work of the CoffeeFlux observatory, developed by CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement) and CATIE (Centro Agronómico Tropical de Investigación y Enseñanza). This site contributes to FLUXNET (CR-AqC) and belongs to the SOERE F-ORE-T network of observatories, with support from Ecofor, Allenvi, and the French national research infrastructure ANAEE-F (http://www.anaee-france.fr/fr/). This work was also supported by AIRD (SAFSE) and the ANR, the French Research Agency through two projects: Ecosfix (ANR-2010-STRA-003-01) and MACACC (ANR-13-AGRO-0005).

The authors give warm thanks for the field support we received in Costa Rica, including that from Álvaro Barquero and his family; from Aléxis Pérez; and from the following persons at Cafetalera Aquiares: Alfonso and Diego Robelo, Luis Guillermo Navarro, Manuel Jara and Rafael Acuña Vargas.

References

Andriamananjara, A., Hewson, J., Razakamanarivo, H., Andrisoa, R. H., Ranaivoson, N., Ramboatiana, N., Razafindrakoto, M., Ramifehiarivo, N., Razafimanantsoa, M. P., Rabeharisoa, L., Ramananantoandro, T., Rasolohery, A., Rabetokotany, N., and Razafimbelo, T.: Land cover impacts on aboveground and soil carbon stocks in Malagasy rainforest, Agriculture Ecosystems and Environment, 233, 1-15, https://doi.org/10.1016/j.agee.2016.08.030, 2016.

Basile-Doelsch, I., Amundson, R., Stone, W. E. E., Masiello, C. A., Bottero, J. Y., Colin, F., Masin, F., Borschneck, D., and Meunier, J. D.: Mineralogical control of organic carbon dynamics in a volcanic ash soil on La Réunion, European Journal of Soil Science, 56, 689-703, https://doi.org/10.1111/j.1365-2389.2005.00703.x, 2005.

Batjes, N. H.: Total carbon and nitrogen in the soils of the world, European Journal of Soil Science, 65, 4-21, https://doi.org/10.1111/ejss.12115, 2014.

Beare, M. H., McNeill, S. J., Curtin, D., Parfitt, R. L., Jones, J. S., Dodd, M. B., and Sharp, J.: Estimating the organic carbon stabilisation capacity and saturation deficit of soils: a New Zealand case study, Biogeochemistry, 120, 71-87, https://doi.org/10.1007/s10533-014-9982-1, 2014.

Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., and Sommer, S.: Using Imaging Spectroscopy to study soil properties, Remote Sensing of Environment, 113, S38–S55, http://dx.doi.org/10.1016/j.rse.2008.09.019, 2009.

Benegas, L., Ilstedt, U., Roupsard, O., Jones, J., and Malmer, A.: Effects of trees on infiltrability and preferential flow in two contrasting agroecosystems in Central America, Agriculture, Ecosystems and Environment 183, 185-196, https://doi.org/10.1016/j.agee.2013.10.027, 2014.

Blakemore, L.C., Searle, P.L., Daly, B.K., 1981. Soil Bureau Laboratory Methods: Methods for chemical analysis of soils. New Zealand Soil Bureau Scientific Report, 10A, CSIRO, New Zealand.

Boudot, J. P.: Relative efficiency of complexed aluminum, noncrystalline Al hydroxide, allophane and imogolite in retarding the biodegradation of citric acid, Geoderma, 52, 29-39, https://doi.org/10.1016/0016-7061(92)90073-G, 1992.

Bounouara, Z., Chevallier, T., Balesdent, J., Toucet, J., Sbih, M., Bernoux, M., Belaissaoui, N., Bouneb, O., and Bensaid, R.: Variation in soil carbon stocks with depth along a toposequence in a sub-humid climate in North Africa (Skikda, Algeria), Journal of Arid Environments, 141, 25-33, https://doi.org/10.1016/j.jaridenv.2017.02.001, 2017.

Buurman, P., Peterse, F., and Almendros Martin, G.: Soil organic matter chemistry in allophanic soils: a pyrolysis-GC/MS study of a Costa Rican Andosol catena, European Journal of Soil Science, 58, 1330-1347, https://doi.org/10.1111/j.1365-2389.2007.00925.x, 2007.

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C., and Barthès, B. G.: Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field, Geoderma, 261, 151-159, https://doi.org/10.1016/j.geoderma.2015.07.007, 2016.

Cardinael, R., Chevallier, T., Barthès, B., Saby, N., Parent, T., Dupraz, C., Bernoux, M., and Chenu, C.: Impact of alley cropping agroforestry on stocks, forms and spatial distribution of soil organic carbon - A case study in a Mediterranean context, Geoderma, 259-260, 288-299, https://doi.org/10.1016/j.geoderma.2015.06.015, 2015.

CENIGA, 1998. Hojas Topográficas Escala 1:25000. Proyect TERRA.

Chang, C. W., Laird, D. A., Mausbach, M. J., and Hurburgh, C. R.: Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties., soil Science Society of America Journal, 65, 480-490, https://doi.org/10.2136/sssaj2001.652480x, 2001.

Charbonnier, F., Roupsard, O., le Maire, G., Guillemot, J., Casanoves, F., Lacointe, A., Vaast, P., Allinne, C., Audebert, L., Cambou, A., Clement-Vidal, A., Defrenet, E., Duursma, R. A., Jarri, L., Jourdan, C., Khac, E., Leandro, P., Medlyn, B. E., Saint-Andre, L., Thaler, P., Van den Meersche, K., Aguilar, A. B., Lehner, P., and Dreyer, E.: Increased light-use efficiency sustains net primary productivity of shaded coffee plants in agroforestry system., Plant Cell and Environment 40, 1592-1608, https://doi.org/10.1111/pce.12964, 2017.

Chevallier, T., Voltz, M., Blanchart, E., Chotte, J. L., Eschenbrenner, V., Mahieu, M., and Albrecht, A.: Spatial and temporal changes of soil C after establishment of a pasture on a long-term cultivated vertisol (Martinique), Geoderma, 94, 43-58, https://doi.org/10.1016/S0016-7061(99)00064-6, 2000.

Churchman, J., Pasbakhsh, P., Lowe, D. J., and Theng, B. K. G.: Unique but diverse: some observations on the formation, structure, and morphology of halloysite, Clays Minerals, 51, 395-416, https://doi.org/10.1180/claymin.2016.051.3.14, 2016.

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N., Bernoux, M., and Barthès, B. G.: National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy, Geoderma, 276, 41-52, https://doi.org/10.1016/j.geoderma.2016.04.021, 2016.

Costa Junior, C., Corbeels, M., Bernoux, M., Piccolo, M. C., Siqueira Neto, M., Feigl, B. J., Cerri, C. E. P., Cerri, C. C., Scopel, E., and Lal, R.: Assessing soil carbon storage rates under no-tillage: Comparing the synchronic and diachronic approaches, Soil & Tillage Research, 134, 207-212, https://doi.org/10.1016/j.still.2013.08.010, 2013.

Dahlgren, R. A., Saigusa, M., and Ugolini, F. C.: The nature, properties and management of volcanic soils, in: Advances in Agronomy, Vol 82, edited by: Sparks, D. L., Advances in Agronomy, 113-182, 2004.

Defrenet, E., Roupsard, O., Van den Meersche, K., Charbonnier, F., Pastor Pérez-Molina, J., Khac, E., Prieto, I., Stokes, A., Roumet, C., Rapidel, B., de Melo Virginio Filho, E., Vargas, V. J., Robelo, D., Barquero, A., and Jourdan, C.: Root biomass, turnover and net primary productivity of a coffee agroforestry system in Costa Rica: effects of soil depth, shade trees, distance to row and coffee age, Annals of Botany, 118, 833-851, https://doi.org/10.1093/aob/mcw153, 2016.

Devitre, C., Gazel, E., Quesada, P., Tracy, R., Lucke, O., Soto, G., and Alvarado-Induni, G.: Geochemical Evidence for Multi-Stage Chaotic Magma Mixing at Turrialba Volcano, Costa Rica, Fall meeting AGU, Washington, D.C., 10-14 dec 2018, 2018.

Don, A., Schumacher, J., Scherer-Lorenzen, M., Scholten, T., and Schulze, E. D.: Spatial and vertical variation of soil carbon at two grassland sites - Implications for measuring soil carbon stocks, Geoderma, 141, 272-282,

515   https://doi.org/10.1016/j.geoderma.2007.06.003, 2007.

Feller, C., Albrecht, A., Blanchart, E., Cabidoche, Y. M., Chevallier, T., Hartmann, C., Eschenbrenner, V., Larre-Larrouy, M. C., and Ndandou, J. F.: Soil organic carbon sequestration in tropical areas. General considerations and analysis of some edaphic determinants for Lesser Antilles soils, Nutrient Cycling in Agroecosystems, 61, 19-31, https://doi.org/10.1023/A:1013359319380, 2001.

520   Filimonova, S., Kaufhold, S., Wagner, F. E., Hausler, W., and Kogel-Knabner, I.: The role of allophane nano-structure and Fe oxide speciation for hosting soil organic matter in an allophanic Andosol, Geochimica et Cosmochimica Acta, 180, 284-302, https://doi.org/10.1016/j.gca.2016.02.033, 2016.

Gessler, P., Chadwick, O., Chamran, F., Althouse, L., and Holmes, K.: Modeling Soil-Landscape and ecosystem properties using terrain attributes, Soil Science Society of America Journal, 64, 2046-2056, https://doi.org/10.2136/sssaj2000.6462046x, 2000.

525   Gómez-Delgado, F., Roupsard, O., le Maire, G., Taugourdeau, S., Pérez, A., van Oijen, M., Vaast, P., Rapidel, B., Harmand, J. M., Voltz, M., Bonnefond, J. M., Imbach, P., and Moussa, R.: Modelling the hydrological behaviour of a coffee agroforestry basin in Costa Rica, Hydrology and Earth System Sciences, 15, 369-392, https://doi.org/10.5194/hess-15-369-2011, 2011.

Hidalgo, C., Etchevers, J. D., Martinez-Richa, A., Yee-Madeira, H., Calderon, H. A., Vera-Graziano, R., and Matus, F.: Mineralogical characterization of the fine fraction (< 2 mu m) of degraded volcanic soils and tepetates in Mexico, Applied Clay

530   Science, 49, 348-358, https://doi.org/10.1016/j.clay.2009.11.007, 2010.

Huygens, D., Boeckx, P., Van Cleemput, O., Oyarzun, C., and Godoy, R.: Aggregate and soil organic carbon dynamics in South Chilean Andisols, Biogeosciences, 2, 159-174, https://doi.org/10.5194/bg-2-159-2005, 2005.

Janik, L. J., Merry, R. H., and Skjemstad, J. O.: Can mid infrared diffuse reflectance analysis replace soil extractions ?, Australian Journal of Experimental Agriculture, 38, 681-696, https://doi.org/10.1071/EA97144, 1998.

535   Jobbagy, E. G., and Jackson, R. B.: The vertical distribution of soil organic carbon and its relation to climate and vegetation, Ecological Applications, 10, 423-436, https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2, 2000.

Kinoshita, R., Roupsard, O., Chevallier, T., Albrecht, A., Taugourdeau, S., Ahmed, Z., and Van Es, H.: Large topsoil organic carbon variability is controlled by Andisol properties and effectively assessed by VNIR spectroscopy in a coffee agroforestry system of Costa Rica, Geoderma, 262, 254-265, https://doi.org/10.1016/j.geoderma.2015.08.026, 2016.

540   Kleber, M., Mikutta, R., Torn, M., and Jahn, R.: Poorly crystalline mineral phases protect organic matter in acid subsoil horizons, European Journal of Soil Science, 56, 717-725, https://doi.org/10.1111/j.1365-2389.2005.00706.x, 2005.

Kramer, M. G., Sanderman, J., Chadwick, O. A., Chorover, J., and Vitousek, P. M.: Long-term carbon storage through retention of dissolved aromatic acids by reactive particles in soil, Global Change Biology 18, 2594-2605, https://doi.org/10.1111/j.1365-2486.2012.02681.x, 2012.

545   Lal, R.: Soil carbon sequestration impacts on global climate change and food security., Science, 304, 1623-1627, https://doi.org/10.1126/science.1097396, 2004.

Levard, C., Doelsch, E., Basile-Doelsch, I., Abidin, Z., Miche, H., Masion, A., Rose, J., Borschneck, D., and Bottero, J. Y.: Structure and distribution of allophanes, imogolite and proto-imogolite in volcanic soils, Geoderma, 183, 100-108, https://doi.org/10.1016/j.geoderma.2012.03.015, 2012.

550   Mathieu, J. A., Hatté, C., Balesdent, J., and Parent, É.: Deep soil carbon dynamics are driven more by soil type than by climate: a worldwide meta-analysis of radiocarbon profiles, Global Change Biology, 21, 4278-4292, https://doi.org/10.1111/gcb.13012, 2015.

Matus, F., Rumpel, C., Neculman, R., Panichini, M., and Mora, M. L.: Soil carbon storage and stabilisation in andic soils: A review, Catena, 120, 102-110, https://doi.org/10.1016/j.catena.2014.04.008, 2014.

555   Mayer, L. M.: Relationships between mineral surfaces and organic carbon concentrations in soils and sediments., Chemical Geology, 114, 347-363, https://doi.org/10.1016/0009-2541(94)90063-9, 1994.

Mayer, L. M., and Xing, B.: Organic matter-surface area relationships in acid soils, Soil Science Society of America Journal, 65, 250-258, https://doi.org/10.2136/sssaj2001.651250x, 2001.

Mayer, L. M., Schick, L., Hardy, K., Wagai, R., and McCarthy, J.: Organic matter in small mesopores in sediments and soils, 560   Geochimica et Cosmochimica Acta, 68, 3863-3872, https://doi.org/10.1016/j.gca.2004.03.019, 2004.

McCarthy, J. F., Ilavsky, J., Jastrow, J. D., Mayer, L. M., Perfect, E., and Zhuang, J.: Protection of organic carbon in soil microaggregates via restructuring of aggregate porosity and filling of pores with accumulating organic matter, Geochimica et Cosmochimica Acta, 72, 4725-4744, https://doi.org/10.1016/j.gca.2008.06.015, 2008.

McDowell, M. L., Bruland, G. L., Deenik, J. L., Grunwald, S., and Knox, N. M.: Soil total carbon analysis in Hawaiian soils with 565   visible, near-infrared and mid-infrared diffuse reflectance spectroscopy, Geoderma, 189-190, 312-320, https://doi.org/10.1016/j.geoderma.2012.06.009, 2012.

Meijer, E. L., and Buurman, P.: Chemical trends in a perhumid soil catena on the Turrialba volcano (Costa Rica), Geoderma, 117, 185-201, https://doi.org/10.1016/S0016-7061(03)00122-8, 2003.

Misnany, B., Tranter, G., McBratney, A. B., Brough, D. M., and Murphy, B. W.: Regional transferability of mid-infrared diffuse 570   reflectance spectroscopic prediction for soil chemical properties, Geoderma, 153, 155-162, https://doi.org/10.1016/j.geoderma.2009.07.021, 2009.

Mizota, C., and Van Reewijk, L. P.: Clay mineralogy and chemistry of soils formed in volcanic material in diverse climatic regions. Soil Monograph n°2, International Soil Reference and Information Center, Wageningen, 185 pp., 1989.

Mora, J. L., Guerra, J. A., Armas-Herrera, C. M., Arbelo, C. D., and Rodríguez-Rodríguez, A.: Storage and depth distribution of 575   organic carbon in volcanic soils as affected by environmental and pedological factors, Catena, 123, 163-175, https://doi.org/10.1016/j.catena.2014.08.004, 2014.

Mora-Chinchilla, R., 2000. Geomorfología de la Cuenca del Río Turrialba. Universidad de Costa Rica, San José.

Nocita, M., Stevens, A., Van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B. G., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A. M., Genot, V., Guerrerro, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-580   Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J. M., Sheperd, K. D., Stenbergb, B., Towett, E. K., Vargas, R., and Wetterlind, J.: Soil Spectroscopy: An alternative to wet chemistry for soil monitoring, Advances in Agronomy, 132, https://doi.org/10.1016/bs.agron.2015.02.002, 2015.

Noponen, M., Healey, J. R., Soto, G., and Haggar, J. P.: Sink or source - The potential of coffee agroforestry systems to sequester atmospheric CO2 into soil organic carbon, Agriculture, Ecosystems and Environment, 175, 60-68, 585   https://doi.org/10.1016/j.agee.2013.04.012, 2013.

Parfitt, R. L., Russell, M., and Orbell, G. E.: Weathering sequence of soils from volcanic ash involving allophane and halloysite, New Zealand, Geoderma, 29, 41-57, https://doi.org/10.1016/0016-7061(83)90029-0, 1983.

Parfitt, R. L., and Childs, C. W.: Estimation of forms of Fe and Al: a review, and analysis of constrastig soils by dissolution and Moessbauer methods, Australian Journal of Soil Research, 26, 121-144, https://doi.org/10.1071/SR9880121, 1988.

590   Parfitt, R. L., Theng, B. K. G., Whitton, J. S., and Shepherd, T. G.: Effects of clay minerals and land use on organic matter pools, Geoderma, 75, 1-12, https://doi.org/10.1016/S0016-7061(96)00079-1, 1997.

Parfitt, R. L.: Allophane and imogolite: role in soil biogeochemical processes, Clays Minerals, 44, 135-155, https://doi.org/10.1180/claymin.2009.044.1.135, 2009.

Payan, F., Jones, D. L., Beer, J., and Harmand, J. M.: Soil characteristics below Erythrina poeppigiana in organic and conventional
595   Costa Rican coffee plantations, Agroforestry Systems, 76, 81-93, https://doi.org/10.1007/s10457-008-9201-y, 2009.

Percival, H. J., Parfitt, R. L., and Scott, N. A.: Factors controlling soil carbon levels in New Zealand grasslands: Is clay content important ?, Soil Science Society of America Journal, 64, 1623-1630, https://doi.org/10.2136/sssaj2000.6451623x, 2000.

Poulenard, J., Podwojewski, P., and Herbillon, A. J.: Characteristics of non-allophanic andisols with hydric properties from the Ecuadorian paramos, Geoderma, 117, 267-281, https://doi.org/10.1016/s0016-7061(03)00128-9, 2003.

600   Powers, J. S., and Schlesinger, W. H.: Relationships among soil carbon distributions and biophysical factors at nested spatial scales in rain forests of northeastern Costa Rica, Geoderma, 109, 165-190, https://doi.org/10.1016/s0016-7061(02)00147-7, 2002.

Rasse, D. P., Mulder, J., Moni, C., and Chenu, C.: Carbon turnover kinetics with depth in a French loamy soil, Soil Science Society of America Journal, 70, 2097-2105, https://doi.org/10.2136/sssaj2006.0056, 2006.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna,
605   Austria. URL https://www.R-project.org/.

Scheel, T., Dörfler, C., and Kalbitz, K.: Precipitation of dissolved organic matter by aluminium stabilizes carbon in acidic forest soils, Soil Science Society of America Journal, 71, 64-74, https://doi.org/10.2136/sssaj2006.0111, 2007.

Shen, Q., Suarez-Abelenda, M., Camps-Arbestain, M., Calvelo Pereira, R., McNally, S. R., and Kelliher, F. M.: An investigation of organic matter quality and quantity in acid soils asinfluenced by soil type and land use, Geoderma, 328, 44-55,
610   https://doi.org/10.1016/j.geoderma.2018.05.006, 2018.

Shi, S., Zhang, W., Zhang, P., Yu, Y., and Ding, F.: A synthesis of change in deep soil organic carbon stores with afforestation of agricultural soils, Forest Ecology and Management, 296, 53-63, https://doi.org/10.1016/j.foreco.2013.01.026, 2013.

Shoji, S., Nanzyo, M., Dahlgren, R. A., and Quantin, P.: Evaluation and proposed revisions of criteria for Andosols in the world reference base for soil resources, Soil Science, 161, 604-615, https://doi.org/10.1097/00010694-199609000-00005, 1996.

615   Soriano-Disla, J. M., Janik, L. J., Rossel, R. A. V., Macdonald, L. M., and McLaughlin, M. J.: The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties, Applied Spectroscopy Reviews, 49, 139-186, https://doi.org/10.1080/05704928.2013.811081, 2014.

Takahashi, T., and Dahlgren, R. A.: Nature, properties and function of aluminum–humus complexes in volcanic soils, Geoderma, 263, 110-121, https://doi.org/10.1016/j.geoderma.2015.08.032, 2016.

620   Terra, F. S., Dematte, J. A. M., and Rossel, R. A. V.: Proximal spectral sensing in pedological assessments: vis-NIR spectra for soil classification based on weathering and pedogenesis, Geoderma, 318, 123-136, https://doi.org/10.1016/j.geoderma.2017.10.053, 2018.

Tonneijck, F. H., Jansen, B., Nierop, K. G. J., Verstraten, J. M., Sevink, J., and De Lange, L.: Towards understanding of carbon stocks and stabilization in volcanic ash soils in natural Andean ecosystems of northern Ecuador, European Journal of Soil Science,
625   61, 392-405, https://doi.org/10.1111/j.1365-2389.2010.01241.x, 2010.

Torn, M., Trumbore, S., Chadwick, O., Vitousek, P., and Hendricks, D.: Mineral control of soil organic carbon storage and turnover, Nature, 389, 170-173, https://doi.org/10.1038/38260, 1997.

Ugolini, F. C., and Dahlgren, R. A.: Soil development in volcanic ash, Global Journal of Environmental Research, 6, 69-81, 2002.

Visacarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Dematte, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthes, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Bottcher, K., Brodsky, L., Du, C. W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C. B., Knadel, M., Morras, H. J. M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. M. R., Sanborn, P., Sellitto, V. M., Sudduth, K. A., Rawlins, B. G., Walter, C., Winowiecki, L. A., Hong, S. Y., and Ji, W.: A global spectral library to characterize the world's soil, Earth-Science Reviews, 155, 198-230, https://doi.org/10.1016/j.earscirev.2016.01.012, 2016.

Wada, K. J.: Allophane and Imogolite, in: Minerals in Soil Environments, edited by: Dixon, J. B., and Weeds, S. B., 2nd ed. Soil Sc. Soc. Am., WI, 1051-1108, 1989.

Welsh, K., Boll, J., Sanchez-Murillo, R., and Roupsard, O.: Isotope hydrology of a tropical coffee agroforestry watershed: Seasonal and event-based analyses, Hydrological Processes, 32, 1965-1977, https://doi.org/10.1002/hyp.13149, 2018.

Zehetner, F., Miller, W. P., and West, L. T.: Pedogenesis of Volcanic Ash Soils in Andean Ecuador, Soil Science Society of America Journal, 67, 1797-1809, https://doi.org/10.2136/sssaj2003.1797, 2003.
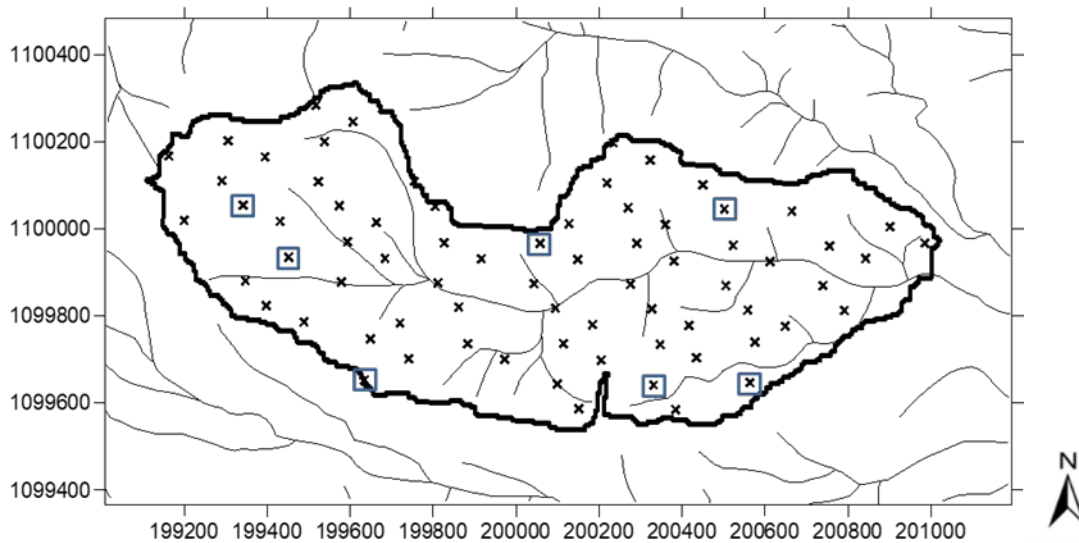
Figure 1. Map of the study site: a 1-km^2 micro-watershed. "X" symbols mark the 69 sampling locations. The seven squares show where pits were dug down to a depth of 2 m for bulk-density measurements. Thin lines indicate permanent stream channels. Numbers along the axes are UTM coordinates. See Gómez-Delgado et al. (2011) for more geographical details.
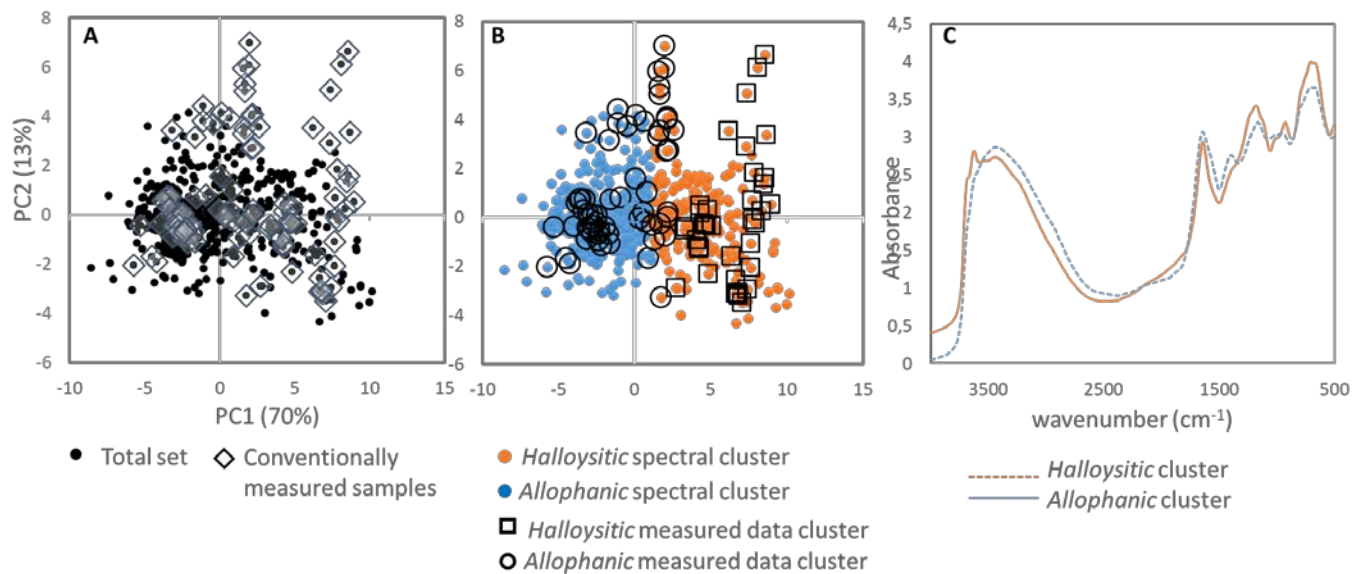
Figure 2. Mid-infrared (MIR) spectra analyses for the total set of soil samples (n=598) and for the samples (n= 98) that were chemically analysed conventionally to determine contents of Al_o, Al_p, Fe_o, and Si_o. A- Principal Component Analysis (PCA) of all spectra. Diamond-shapes identify the 98 conventionally analysed samples. B- PCA and identification of 2 clusters built either on the PCA of all the 598 MIR spectra (solid blue and red points) or on the PCA of the compositions of the 98 conventionally analysed samples (empty circles and squares). C- Average MIR spectra of the soil samples from *allophanic* and *halloysitic* spectral clusters.
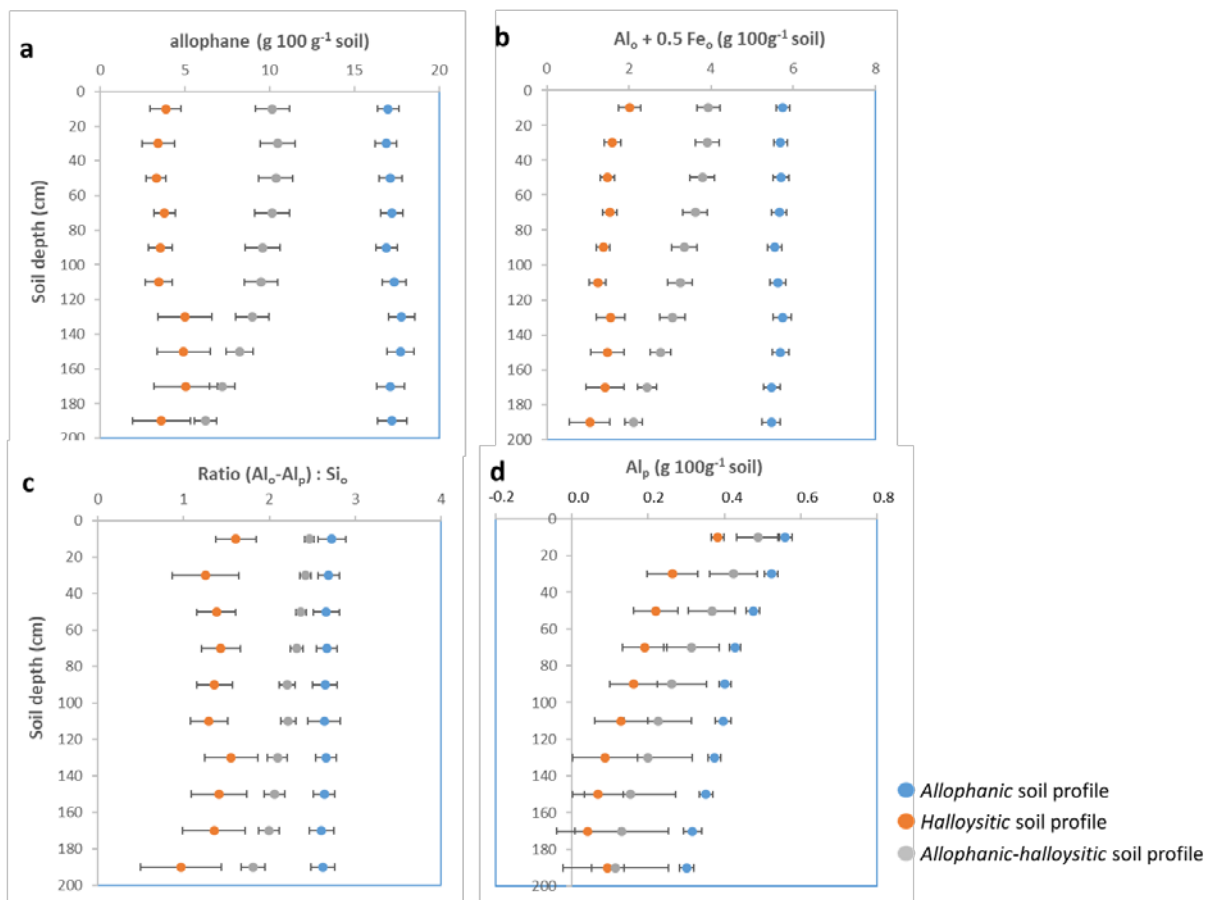
Figure 3. The three types of soil profiles, according to four characteristics of andic soils (a- allophane content, b- $Al_o + 0.5\ Fe_o$ index, c- Al:Si ratio, and d- organo-Al complexes or Al_p content). The *Allophanic* soil profile shown is the mean soil profile of 35 soil profiles; *Halloysitic* soil profile is the mean soil profile of eight soil profiles, and *Allophanic-Halloysitic* soil profile is the mean soil profile of 26 soil profiles. Error bars are error types.

640
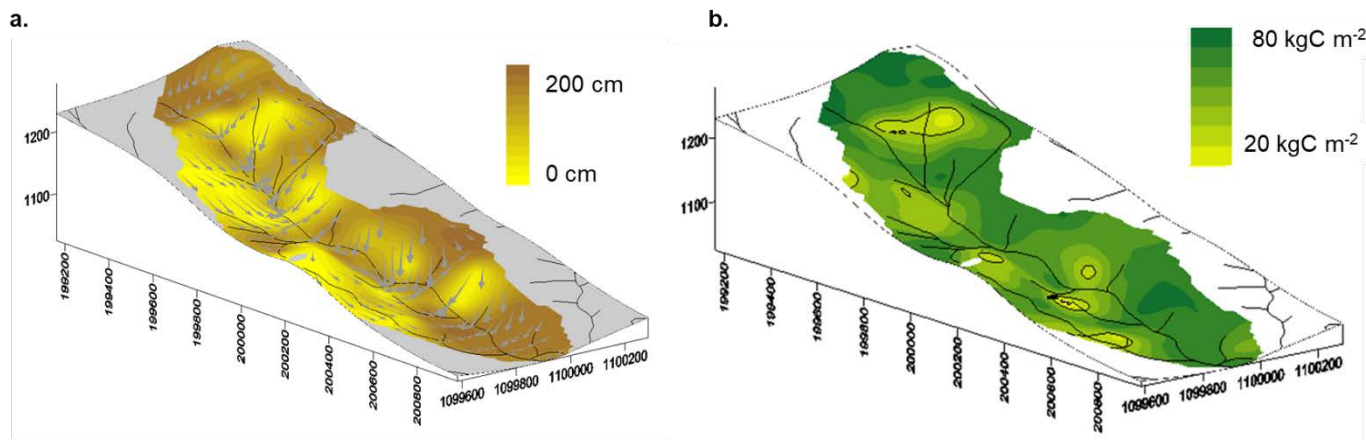
Figure 4.; a- Map of the thickness of the *allophanic* soil material layer (in cm) in the Aquiares watershed; the fine grey arrows show directions and steepness of elevation gradients: longer arrows indicate steeper slopes. b- Map of SOC stocks (kgC m⁻²) in the upper two meters of soil. Altitudes above sea level are in m; x and y axes show UTM coordinates.
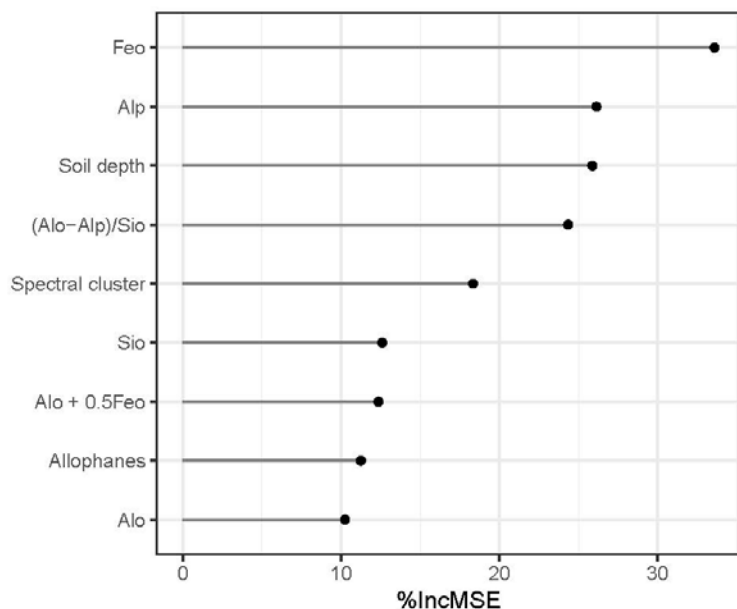
Figure 5. Relative importance (%IncMSE) of variables in the random forest model for the prediction of SOC content.
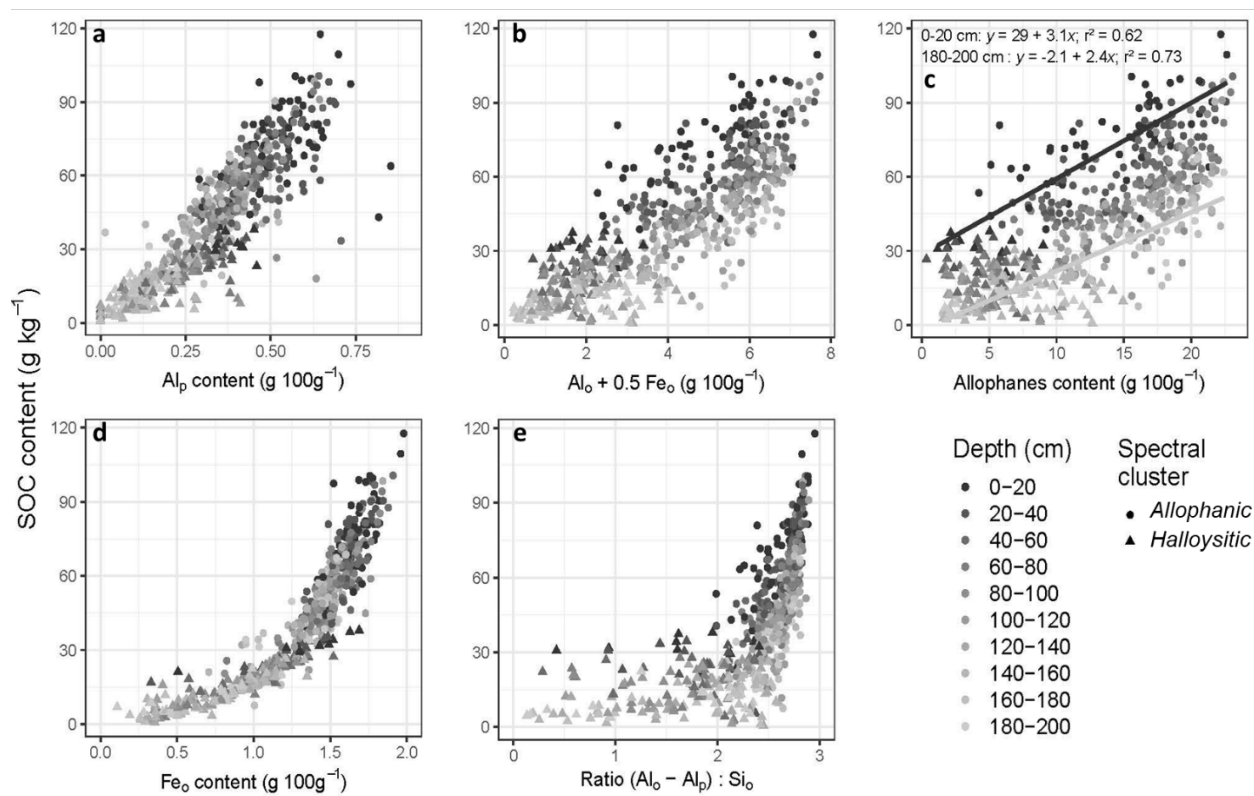
Figure 6. Relationships, according to soil-sampling depth and spectral-cluster classification, between SOC content and a- Al_p content; b- Al_o + 0.5 Fe_o index; c- allophane content (showing 2 linear regressions on results for samples from the 0-20 cm depth and the 180-200 cm); d- Fe_o content; and e- Al:Si ratio.
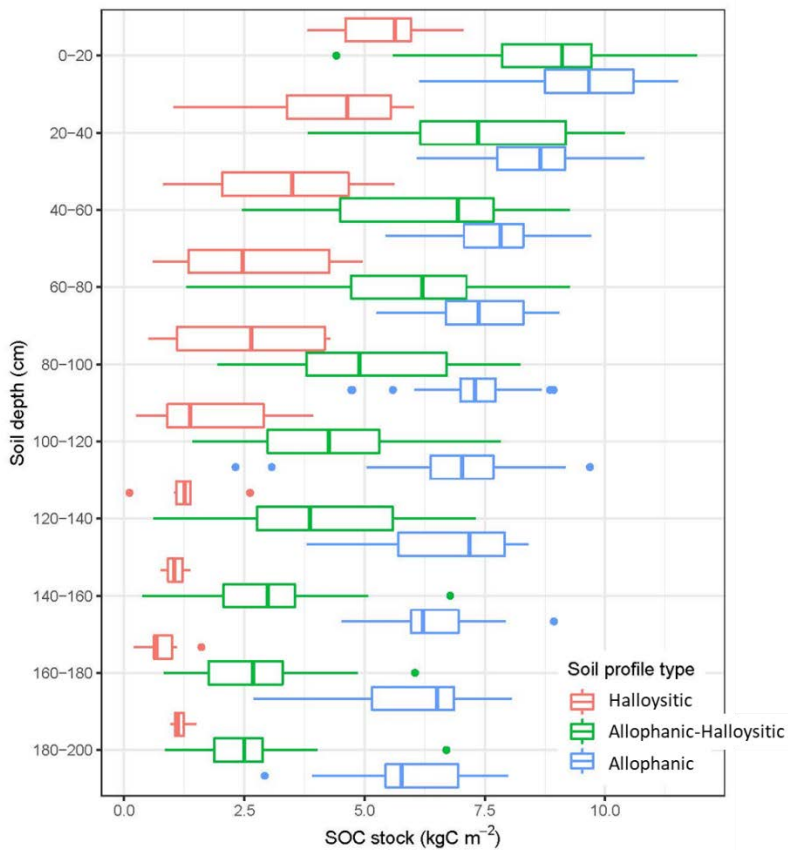
Figure 7: Mean SOC stocks in the three types of soil profiles. The *allophanic* profile is the mean profile of the 35 soil profiles composed by samples only in the *allophanic* spectral cluster; *Halloysitic* profile is the mean of the 8 soil profiles composed by samples only in the *halloysitic* spectral cluster; and *allophanic-halloysitic* profile is the mean of the 26 soil profiles composed by samples in both spectral cluster.
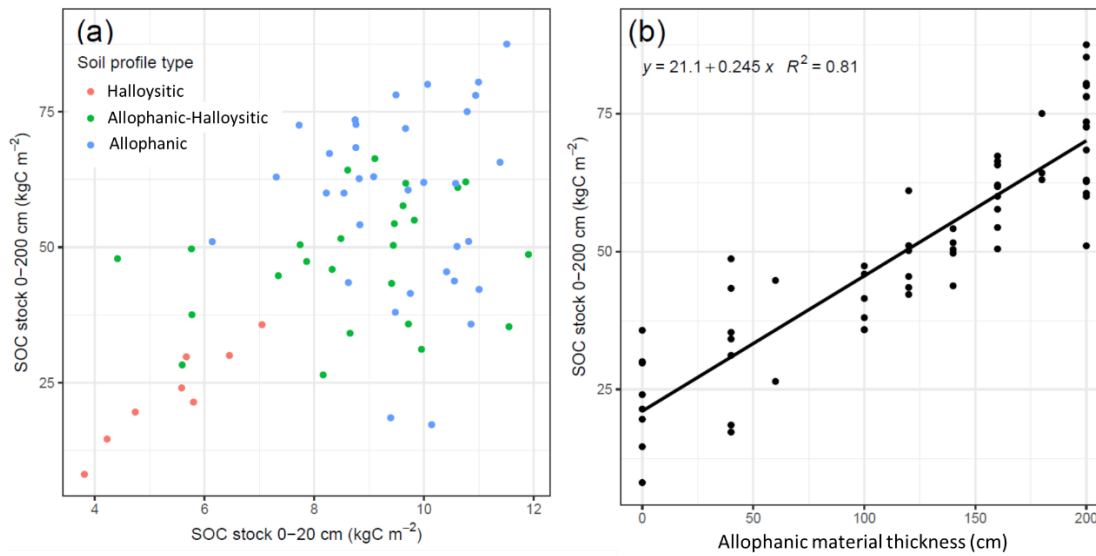
Figure 8: Relationship between SOC stock at 0-200 cm depth and a) the SOC stock in 0-20 cm depth, b) the thickness of the *allophanic* soil material.

Table 1. Validation statistics of modified partial least square (mPLS) regression for the models used to predict SOC and extractable Al, Si, and Fe contents from MIR spectra.

		n	n outliers	Mean	Sd	RMSE	R^2	RPD
Al_o	Calibration	69	0	2.18	1.72	0.4	0.94	4.9
Al_o	External validation	29		2.85	2.15	0.8	0.85	2.6
Si_o	Calibration	69	3	0.88	0.53	0.14	0.93	3.9
Si_o	External validation	29		1.06	0.64	0.19	0.91	3.4
Fe_o	Calibration	69	0	1.0	0.5	0.2	0.81	2
Fe_o	External validation	29		1.1	0.5	0.5	0.19	1
Al_p	Calibration	69	1	0.26	0.17	0.07	0.85	2.4
Al_p	External validation	29		0.24	0.12	0.06	0.72	2
SOC	Calibration	69	2	39.7	26.1	3.6	0.97	7.3
SOC	External validation	29		39.7	21.3	9.4	0.86	2.3

645   n: number of soil samples for calibration. Mean and Standard deviation (Sd) of the measured Al_o, Si_o, Fe_o, Al_p population (in g 100 g^{-1} soil) and SOC content (g kg^{-1}) used for model calibration RMSE: Root Mean Square Error, R^2: cross-validation determination coefficient, RPD : ratio of Sd to RMSE, Sd : Standard deviation.

Table 2. Organo-Al complexes and SRO minerals contents (means ± standard deviation) of the predicted data of all soil samples (All) and for those in *allophanic* and *halloysitic* spectral clusters. Letters indicate significant differences among soil depth at p < 0.05.

Depth (cm)	Al_p g 100 g⁻¹ soil All	Allophanic	Halloysitic	Fe_o g 100 g⁻¹ soil All	Allophanic	Halloysitic	Allophanes g 100 g⁻¹ soil All	Allophanic	Halloysitic	$(Al_o-Al_p)/Si_o$ All	Allophanic	Halloysitic	$Al_o + 0.5 Fe_o$ g 100 g⁻¹ soil All	Allophanic	Halloysitic	Number of samples per cluster All	Allophanic	Halloysitic
0-20	0.51±0.12 a	0.54±0.11 a	0.37±0.07 a	1.6±0.21 a	1.65±0.13 a	1.33±0.33 a	12.8±6.2 b	14.5±5.3 a	4.3±2.3 ab	2.5±0.5 a	2.6±0.2 a	1.7±0.6	4.6±1.7 ab	5.1±1.4	2.1±0.7 a	68	57	11
20-40	0.46±0.13 b	0.5±0.11 ab	0.31±0.11 ab	1.5±0.26 ab	1.6±0.12 ab	1.15±0.34 b	12.9±6.2 ab	15.3±4.5 ab	4.1±2.1 b	2.4±0.6 ab	2.6±0.2 ab	1.6±0.9	4.5±1.8 a	5.3±1.2	1.9±0.6 abc	69	54	15
40-60	0.41±0.14 c	0.46±0.1 b	0.24±0.1 bc	1.43±0.29 bc	1.56 0.13 bc	1.05±0.31 bc	12.9±6.3 ab	15.8±4.5 ab	4.6±2.0 ab	2.4±0.5 ab	2.6±0.2 bc	1.7±0.6	4.5±1.9 a	5.3±1.2	1.9±0.7 ab	66	49	17
60-80	0.36±0.12 d	0.41±0.08 c	0.21±0.1 cd	1.36±0.3 cd	1.46±0.17 cd	0.97±0.28 cd	12.8±6.3 ab	15.7±4.5 ab	4.4±1.6 ab	2.4±0.5 ab	2.6±0.2 abc	1.7±0.5	4.4±1.9 abc	5.2±1.2	1.8±0.6 abc	67	50	17
80-100	0.32±0.13 de	0.38±0.09 cd	0.19±0.1 cde	1.29±0.34 de	1.47±0.18 de	0.88±0.26 de	12.4±6.3 ab	15.9±3.8 ab	4.4±2.4 ab	2.3±0.5 ab	2.6±0.2 bc	1.6±0.5	4.2±1.9 abc	5.3±1.1	1.7±0.7 abc	67	47	20
100-120	0.3±0.15 ef	0.37±0.11 cd	0.16±0.1 def	1.21±0.38 ef	1.41±0.19 ef	0.8±0.32 e	12.6±6.5 ab	16.3±4.0 ab	4.8±2.3 ab	2.3±0.5 ab	2.6±0.2 c	1.7±0.5	4.2±2.0 abc	5.3±1.1	1.8±0.7 abc	62	42	20
120-140	0.27±0.13 fg	0.34±0.09 de	0.15±0.09 efg	1.17±0.38 ef	1.41±0.17 ef	0.79±0.31 ab	12.4±6.6 a	16.7±3.8 ab	5.4±3.1 ab	2.3±0.6 ab	2.6±0.1 bc	1.7±0.6	4.1±2.0 abc	5.4±1.1	1.9±1.0 abc	55	34	21
140-160	0.23±0.13 fgh	0.34±0.09 de	0.14±0.08 fg	1.11±0.39 fg	1.4±0.19 f	0.84±0.34 ef	11.9±6.4 a	17.3±3.9 b	7.2±3.9 a	2.2±0.6 ab	2.6±0.1 bc	1.9±0.7	3.9±2.0 abc	5.6±1.1	2.4±1.3 ab	53	25	28
160-180	0.2±0.13 gh	0.3±0.11 e	0.12±0.08 g	1.00±0.39 gh	1.31±0.19 f	0.73±0.3 ef	10.8±6.3 ab	16.6±3.7 ab	6.0±3.2 ab	2.2±0.7 ab	2.6±0.1 abc	1.8±0.7	3.5±1.9 bc	5.3±1.0	2.0±1.0 bc	48	22	26
180-200	0.19±0.12 h	0.31±0.08 e	0.1±0.06 g	0.98±0.38 g	1.33±0.18 f	0.71±0.26 f	10.4±6.5 ab	16.9±3.8 ab	5.6±3.0 a	2.1±0.7 b	2.6±0.2 abc	1.7±0.7	3.4±2.0 c	5.4±1.0	1.9±1.0 c	42	18	24

Table 3 Cross validation statistics of modified partial least square regression for the models of prediction by MIRS of soil organic carbon contents (gC kg^{-1} soil) after calibration on all of the soil samples (All) or on two clusters defined by their MIRS spectra, or by their conventionally measured data (Al$_o$, Al$_p$, Si$_o$, Fe$_o$, Al$_o$+0.5 Fe$_o$, Allophane, (Al$_o$-Al$_p$)/Si$_o$).

	n	n outliers	Mean	Sd	R^2	RMSE	RPD	n predicted values <0
All samples	98	2	33.2	23.5	0.95	5.3	4.4	12
Allophanic Spectral cluster	61	3	46.0	20.3	0.91	5.7	3.6	0
Halloysitic Spectral cluster	37	2	12.0	7.6	0.92	2.1	3.6	0
Allophanic Measured data cluster	31	1	53.9	19.7	0.88	6.9	2.8	0
Halloysitic Measured data cluster	67	3	23.6	18.4	0.93	4.6	4.0	19

n number of soil samples, Mean and Standard deviation (Sd) of the measured SOC population used for model calibration. R^2, coefficient of regression, RMSE, Root Mean Standard error, RPD, ratio of standard error of prediction and Sd, between measured and predicted SOC contents from the cross validation populations.

Table 4 Descriptive statistics for SOC contents of all soil samples with depth, and with their spectral cluster; p-value expresses results from t-test between samples from the two spectra cluster classes (Welsh, two sided alternative). For a given cluster, means followed by the same letters do not differ significantly at p=0.05.

SOC g kg^{-1}soil	All soil samples			*Allophanic* cluster (398 samples)			*Halloysitic* cluster (199 samples)			p-value
	Mean	std	n	Mean	Std	n	Mean	std	n	
0-20	68.0	22.5	68	75.1 f	16.8	57	31.2 d	5.2	11	<0.0001
20-40	58.5	21.6	69	67.3 ef	14.9	54	26.8 cd	7.8	15	<0.0001
40-60	49.8	21.3	66	59.9 de	14.1	49	20.9 bc	6.5	17	<0.0001
60-80	45.4	21	67	54.9 cd	14.7	50	17.5 ab	7.3	17	<0.0001
80-100	41.2	20.9	67	51.8 bd	14.8	47	16.2 ab	6.6	20	<0.0001
100-120	35.8	20.2	62	46.0 abc	15.7	42	14.3 ab	7.6	20	<0.0001
120-140	34	21.1	55	46.8 abc	15.9	34	13.3 a	7.5	21	<0.0001
140-160	28.6	19.6	53	45.6 abc	14.2	25	13.5 a	7.5	28	<0.0001
160-180	24.4	17.9	48	39.5 a	14.9	22	11.6 a	6.7	26	<0.0001
180-200	24.1	17.4	42	41.1 ab	12.5	18	11.3 a	5.1	24	<0.0001

Table 5 Cross validation statistics of modified partial least square regression for the models used to predict soil bulk densities (g cm^{-3}) from MIR spectra.

n	n outliers	Mean	Sd	R^2	RMSE	RPD
62	4	0.84	0.18	0.74	0.091	1.97

Table 6 Descriptive statistics for bulk densities with depth and with their spectral cluster. P-value express results from t-test between samples from the two cluster classes (Welsh, two sided alternative). For a given cluster, means followed by the same letters do not differ significantly at p=0.05.

Bd g cm^{-3}	All soil samples		Allophanic Cluster (398 soil samples)		Halloysitic Cluster (199 soil samples)		p-value
	Mean	Sd	Mean	Sd	Mean	Sd	
0-20	0.68 f	0.14	0.65 a	0.11	0.87 c	0.11	<0.0001
20-40	0.70 ef	0.15	0.65 de	0.12	0.87 bc	0.12	<0.0001
40-60	0.72 ef	0.16	0.66 de	0.12	0.88 bc	0.15	<0.0001
60-80	0.74 de	0.15	0.69 cd	0.12	0.89 bc	0.13	<0.0001
80-100	0.77 cd	0.17	0.71 bc	0.13	0.93 abc	0.14	<0.0001
100-120	0.80 abc	0.16	0.73 ab	0.11	0.94 abc	0.15	<0.0001
120-140	0.81 abc	0.16	0.73 ab	0.12	0.93 ab	0.13	<0.0001
140-160	0.82 bc	0.16	0.72 ab	0.10	0.91 abc	0.14	<0.0001
160-180	0.87 ab	0.15	0.76 a	0.11	0.96 a	0.12	<0.0001
180-200	0.88 a	0.15	0.75 ab	0.11	0.98 a	0.09	<0.0001