land

MDPI

Article

Insights into the Effects of Study Area Size and Soil Sampling Density in the Prediction of Soil Organic Carbon by Vis-NIR Diffuse Reflectance Spectroscopy in Two Forest Areas

Massimo Conforti [1] and Gabriele Buttafuoco [2,*]

[1] National Research Council of Italy, Research Institute for Geo-Hydrological Protection (CNR-IRPI), Via Cavour 4/6, 87036 Rende, Italy
[2] National Research Council of Italy, Institute for Agricultural and Forest Systems in the Mediterranean (CNR-ISAFOM), Via Cavour 4/6, 87036 Rende, Italy
* Correspondence: gabriele.buttafuoco@cnr.it; Tel.: +39-0984841466

Abstract: Sustainable forest land management requires measuring and monitoring soil organic carbon. Visible and near-infrared diffuse reflectance spectroscopy (Vis-NIR, 350–2500 nm), although it has become an important method for predicting soil organic carbon (SOC), requires further studies and methods of analysis to realize its full potential. This study aimed to determine if the size of the study area and soil sampling density may affect the performance of Vis-NIR diffuse reflectance spectroscopy in the prediction of soil organic carbon. Two forest sites in the Calabria region (southern Italy), which differ in terms of area and soil sampling density, were used. The first one was Bonis catchment area (139 ha) with a cover consisting mainly of Calabrian pine, while the second was Mongiana forest area (33.2 ha) within the "Marchesale" Biogenetic Nature Reserve, which is covered by beech. The two study areas are relatively homogeneous regarding parent material and soil type, while they have very different soil sampling density. In particular, Bonis catchment has a lower sampling density (135 samples out of 139 ha) than Mongiana area (231 samples out of 33.2 ha). Three multivariate calibration methods (principal component regression (PCR), partial least square regression (PLSR), and support vector machine regression (SVMR)) were combined with different pretreatment techniques of diffuse reflectance spectra (absorbance, ABS, standard normal variate, SNV, and Savitzky–Golay filtering with first derivative (SG 1st D). All soil samples (0–20 cm) were analyzed in the laboratory for SOC concentration and for measurements of diffuse reflectance spectra in the Vis-NIR region. The set of samples from each study area was randomly divided into a calibration set (70%) and a validation set (30%). The assessment of the goodness for the different calibration models and the following SOC predictions using the validation sets was based on three parameters: the coefficient of determination (R^2), the root mean square error (*RMSE*), and the interquartile range (*RPIQ*). The results showed that for the two study areas, different levels of goodness of the prediction models depended both on the type of pretreatment and the multivariate method used. Overall, the prediction models obtained with PLSR and SVMR performed better than those of PCR. The best performance was obtained with the SVMR method combined with ABS + SNV + SG 1st D pretreatment ($R^2 \geq 0.77$ and RPIQ > 2.30). However, there is no result that can absolutely provide definitive indications of either the effects of the study area size and soil sampling density in the prediction of SOC by vis-NIR spectroscopy, but this study fostered the need for future investigations in areas and datasets of different sizes from those in this study and including also different soil landscapes.

Keywords: soil organic carbon; forest soil; Vis-NIR reflectance spectroscopy; spectral pretreatment

Citation: Conforti, M.; Buttafuoco, G. Insights into the Effects of Study Area Size and Soil Sampling Density in the Prediction of Soil Organic Carbon by Vis-NIR Diffuse Reflectance Spectroscopy in Two Forest Areas. *Land* **2023**, *12*, 44. https://doi.org/10.3390/land12010044

Academic Editor: Yiyun Chen

Received: 21 November 2022
Revised: 18 December 2022
Accepted: 21 December 2022
Published: 23 December 2022

1. Introduction

The European vision for soil by 2050 is anchored in the EU biodiversity strategy for 2030 [1] and in the Climate Adaption Strategy [2] with the purpose of contributing significantly to several objectives of the European Green Deal [3] and Sustainable Development Goal 15.3 of the United Nations. This vision is aimed at achieving that all soil ecosystems are in a healthy condition and hence more resilient to global change and able to provide as many soil-related ecosystem services as possible. In this context, soil organic carbon (SOC) plays a key role in the terrestrial carbon (C) cycle and, particularly, in forest ecosystems, which together with other wooded land, cover over 43.5% of European territory [4]. Forest soils are the largest C sinks and about 80 percent of terrestrial organic carbon stocks involved in the active C cycle are soil-related and only about 20 percent in vegetation [5,6]. Consequently, the increase of C sequestration in form of soil organic carbon (SOC) has a critical role in the mitigation of global warming [7–9]. Moreover, soil organic carbon has profound effects on soil and environmental quality because it strongly affects soil properties (physical, chemical, and biological) and soil processes; it is an important sink and source of main plant and microbial nutrients [10,11].

Measuring and monitoring soil organic carbon is fundamental for the management of forest ecosystems and there is also an increasing demand for SOC data to support environmental monitoring and spatial modeling. In this context, visible and near-infrared reflectance spectroscopy (Vis-NIR, 350–2500 nm) has become an important method for measuring and monitoring soil organic carbon [12–18]. Vis-NIR diffuse reflectance spectroscopy is effective and less time-consuming, and cost-effective than conventional analytical methods, especially when large numbers of samples need to be measured [13,16,18–20]. However, direct quantitative prediction of soil constituents from soil diffuse reflectance spectra is almost impossible because individual absorptions are difficult to identify since a given spectrum results from the interactions among these soil constituents and their absorption of electromagnetic energy. In addition, the absorptions in the Vis-NIR spectral region are largely non-specific, weak, and broad because of the overlapping absorptions of soil constituents and their usually small concentrations [13,21]. Vis-NIR diffuse reflectance spectroscopy is based on the assumption that the soil diffuse reflectance in the 350–2500 nm spectral region is a linear combination of the spectral signatures from its various constituents, weighted by their abundance [18,22,23]. Accordingly, changes in different soil constituents (physical, chemical, mineralogical, and biological) result in the different spectral features which can be detected by diffuse reflectance spectroscopy [14,17,23,24]. Therefore, the analysis of soil diffuse reflectance spectra data requires complex mathematical treatments to extract useful information and relate the traditional measures of the soil constituents (which from now on will be referred to as soil properties and are the calibration data) to the spectral data. The calibration data are the dependent predicted (**Y**) data, whereas the spectra are the independent (**X**) data [21,25,26]. This mathematical approach is known as chemometrics and is applied using specific wavelengths selected by methods of stepwise regression or can embody full high-resolution spectra using multiple linear regression (MLR), principal components regression (PCR), partial least-squares regression (PLSR), and machine learning (ML) algorithms such as random forest (RF), support vector machine regression (SVMR), and artificial neural network (ANN), and many others [20, 27–31].

The calibration process allows a model to be defined that relates the measures of the soil property(-ies) (**Y**, calibration data) to the diffuse reflectance spectral data (**X**) and is one of the biggest crucial issues that must be addressed for the quality of the prediction of the soil property of interest. The developed calibration models will be used to predict chemical or physical properties in unmeasured soil samples.

However, the results of the calibration process are affected by the mathematical model and its required assumptions (e.g., linearity, no multicollinearity, number of observations, homoscedasticy, etc.), by the pretreatment of the spectral calibration data to reduce the impact of different effects (nonlinearities, additive and multiplicative effects,

etc.) maintaining a sufficient amount of useful information and fulfill the assumptions required by the chosen model and improving calibration [15,21,25,27,32].

There is no best multivariate calibration method, and each one has advantages and disadvantages that in some way affect their choice [21,33–35]. Methods such as PCR or PLSR, compared to MLR, have the advantage of handling multicollinearity of data, but implicitly assume a linear relationship between diffuse reflectance spectra and soil properties that, if not satisfied, mathematical treatments of the spectra are required to establish it [25]. Instead, other methods, such as ANN and SVMR, account for nonlinearity in the relationship between soil diffuse reflectance and soil properties [21].

Furthermore, the effectiveness and accuracy of the predictions are highly dependent on the calibration set, whichever calibration method is chosen. Therefore, the calibration set to ensure robustness and reliability of predictions, must be representative of the whole set and selected according to the spectral characteristics [36] or analytical properties [37] or both [38]. Other studies have demonstrated the usefulness of using different algorithms to select calibration subsets that are representative of the total dataset [39,40] or optimizing the number of soil samples to be included in the calibration dataset [41]. However, not much consideration has been given to the fact that the accuracy and effectiveness of the calibration set may vary depending on the size of the study area and, therefore, the variability of the soil property(ies) of interest or on the soil sampling density.

The soil property chosen for this study is soil organic carbon (SOC), which belongs to the primary group of soil constituents that causes spectral absorptions in the Vis-NIR region and diffuse reflectance spectroscopy has been successfully applied [13,19,26,42–48].

The main objective of this study was to determine if the size of the study area and soil sampling density may affect the performance of Vis-NIR diffuse reflectance spectroscopy in the estimation of soil organic carbon. To this purpose, two forest areas in the Calabria region (southern Italy), which differ in land cover, size, sampling density, content, and range of variation of SOC, were selected.

2. Materials and Methods

2.1. Description of Study Areas

The forest areas are located in the Sila and Serre Vibonesi massifs (Calabria region, southern Italy) (Figure 1a). The first study site is the Bonis catchment (Figure 1b), located in the upland landscape of the Sila Massif (northern Calabria), which is a westward-draining tributary of the Cino river. The catchment area is approximately 139 hectares with an altitude ranging from 979 to 1300 m above sea level and mainly covered by Calabrian pine (*Pinus laricio* Poiret) [49]. From a lithological perspective, the Bonis catchment is predominantly characterized by Paleozoic granitoid rocks, which are often deeply fractured and weathered. For most of the catchment area, the bedrock is covered with thick layers of saprolite [50]. In the upper reaches of the catchment, the morphology is characterized by a relatively flat to gentle-inclined crests, which consists of fragments of the old planation surfaces (paleosurfaces), shaped during the Late Pliocene–Early Pleistocene [51,52]. Steep slopes, often cut by deep incisions, surround these crests. Slope gradient varies from 0° to about 50° with a mean value of 21°. According to the Keys to Soil Taxonom [53], the most common soils in the study area are mainly Typic Xerumbrepts and Ultic Haploxeralf [54]. The prevailing soil textural classes are sandy loam and sandy clay loam.

The second study area, known as Mongiana, is part of the "Marchesale" Biogenetic Nature Reserve (Serre Massif, southern Calabria) (Figure 1c) with an extension of about 33 ha and elevation ranging from 1112 to 1236 m above sea level. Beech (*Fagus sylvatica*) characterizes the plant cover [55], whereas the outcrop rocks consist of Paleozoic granitic, frequently weathered and covered by a thickness of several meters of regolith, some of which are colluvial deposits [56,57]. The morphology of the landscape is characterized by paleosurfaces that are the remnants of flat or gently sloping highlands, which are often

sharply separated by steep slopes and V-shaped valleys [56,58]. Slope gradient ranges between 0° to about 45° with a mean slope of 10°. Soils of the study area can be classified as Humic Dystrudept and Lithic Udipsamments [58]. The prevailing soil textural classes are sandy loam, loam, and silt loam [59].

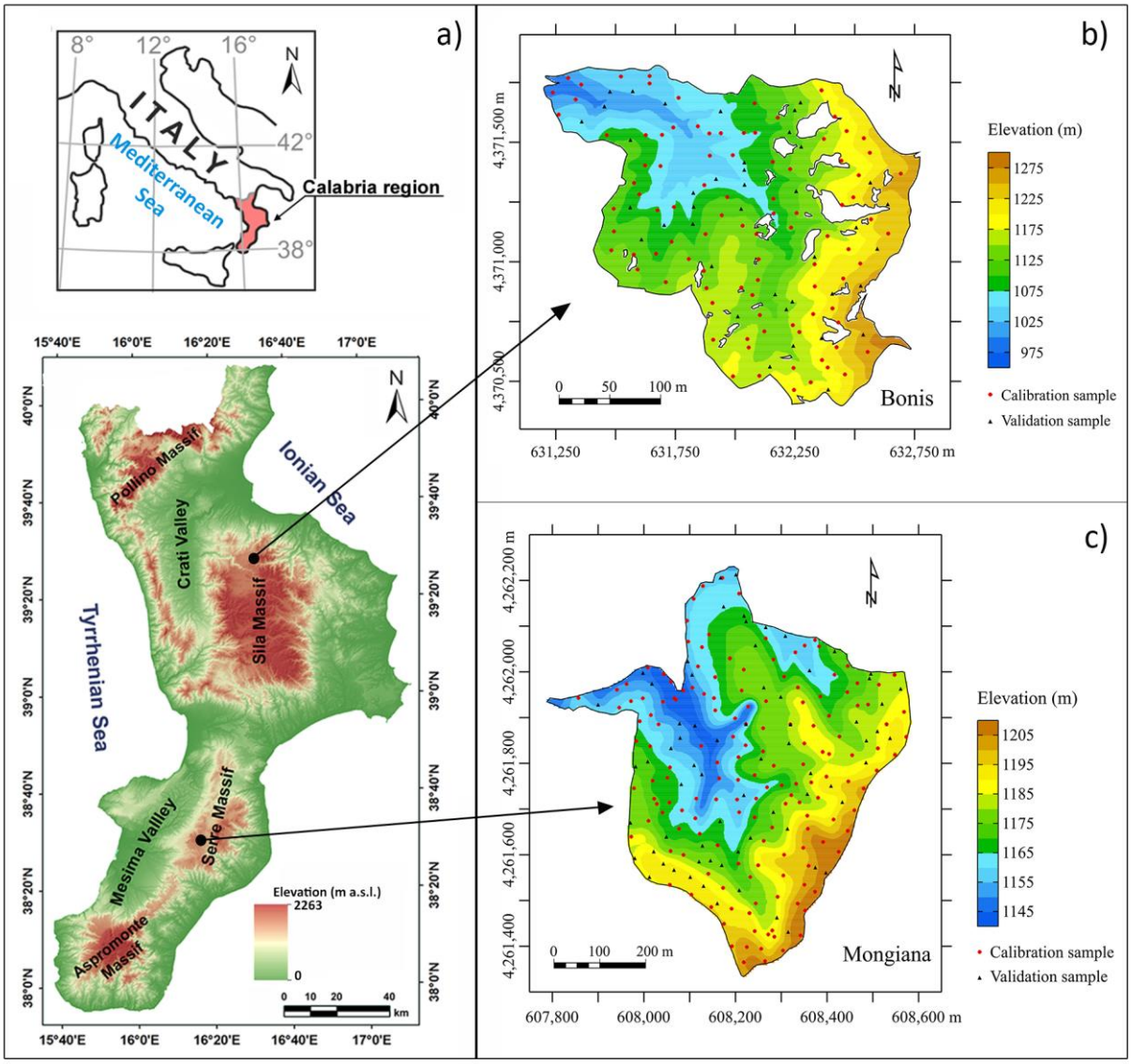

Figure 1. (**a**) Study area locations; (**b**) Bonis catchment; (**c**) Mongiana site. Soil sample locations used for calibration (red dots) and validation (black dots) are also reported.

2.2. Soil Sample Collection and Preparation

Soil samples in both study areas were collected according to the ISO international standard [60]. In the Mongiana area, approximately 75% of the soil samples collected were in areas with soils classified as Humic Dystrudepts, while the remaining 25% were collected in areas with soils classified as Lithic Udipsamments. In the Bonis catchment area, about 85% of the topsoil was collected in areas characterized by Typic Xerumbrepts and 15% in soils classified as Ultic Haploxeralf. Topsoil (0–20 cm) samples were collected in 2012 in Mongiana and in 2016 in the Bonis catchment area. At each soil sampling location, surface litter was removed and soil was sampled using a metallic core cylinder with a diameter of 7.5 cm and a height of 20 cm (883.1 cm^3). The position of the soil sampling sites was recorded using a hand-held GNSS device (Figure 1b,c) and 135 topsoil samples (density sampling = 0.97 samples/ha) were collected within the Bonis catchment,

whereas 231 topsoils were sampled in the Mongiana area (density sampling = 7 samples/ha).

The soil samples were brought to the laboratory, oven dried at 40 °C for 48 h, then were gently crushed in an agate mortar to break up larger aggregates and remove visible roots; next, each sample was sieved at 2 mm. These procedures homogenized soil samples moisture and roughness, minimizing their effect on the spectra measurements.

2.3. SOC and Spectroscopic Analyses

Soil samples were finely ground, then were passed through a 0.25 mm sieve, and analyzed for SOC concentration by dry combustion using a Shimadzu TOC-L analyzer with an SSM-5000A solid sample module (Shimadzu Corporation, Kyoto, Japan). In addition, each SOC measurement was replicated three times.

The diffuse reflectance spectra were measured in laboratory using an ASD FieldSpec IV spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) with a range of wavelength between 350 and 2500 nm and a spectral sampling interval of 1 nm. The soil spectra were measured in a dark room to reduce the effect of external light and with the spectroradiometer held in the nadir position at a distance of 10 cm from the soil sample. The latter was placed inside a petri dish (9 cm in diameter and 1 cm deep) and levelled with a spatula to obtain a smooth surface. For the artificial illumination, a 50-watt halogen lamp with a zenith angle of 30° was used, placed at a distance of approximately 25 cm from the soil sample. Before the first scan and after every set of five samples, under the same lighting conditions, the diffuse reflectance spectrum of a Spectralon panel (20 cm × 20 cm, Labsphere Inc., North Sutton, NH, USA) was measured as a white reference [59].

To reduce the noise level, the average of 50 spectra for each soil sample was calculated. Finally, the diffuse reflectance spectra were resampled every 10 nm wavelength to reduce the total number of raw reflectance bands and avoid overfitting [14].

2.4. Spectral Pretreatments

Before developing calibration models that relate SOC measurements to spectral data, several pretreatments were applied to the raw diffuse reflectance spectra to minimize noise and optimize calibration accuracy. A soil spectrum contains irrelevant information, such as electrical noise, sample background, and stray light. Therefore, it is necessary to eliminate this irrelevant information to improve the prediction models by using chemometric methods [33,61,62]. The commonly used spectral pretreatments techniques include apparent absorbance, mean centering, auto-scaling, normalization, smoothing, derivatives, standard normal variate transformation, and multiplicative scatter correction [33,62,63].

In this study, three different pretreatments were applied to each spectrum of both datasets. In the first pretreatment, the diffuse reflectance (R) spectra were transformed into absorbance (ABS) spectra by ABS = Log (1/R) to improve the identification of characteristic absorption bands. Then, the standard normal variate (SNV) was applied. SNV is a row-oriented transformation, which centers and scales individual spectra and removes the scatter effect from spectral data. A result of the transformation is that on the vertical scale, each spectrum is centered on zero and varies roughly from −2 to +2 [27,32]. Finally, the Savitzky–Golay (SG) filtering method [64] and first derivative (1st D) [61] were applied: the Savitzky–Golay (SG) filtering method reduces the random noise of the measurements [65], whereas the first derivative (1st D) enhances the spectral resolution and removes the additive constant background effects [61].

2.5. Calibration Model Construction

The spectral data and corresponding SOC measurements are used to develop calibration models, which form functional models between the two sets of measurements (spectral data and SOC concentrations) and find out if these functional models are good

(accurate) and, finally, to predict the SOC concentrations of new and unmeasured soil samples from spectral data. The accuracy of a calibration model is a measure of its systematic error, which is measured by the difference between the predicted SOC value and the true one. The assessment of the calibration model accuracy is made using an independent validation set [26,65]. Therefore, it is essential to split both datasets (spectral and SOC) into a calibration set and a validation set. For this purpose, both datasets were randomly partitioned into calibration set (70% of total samples) and validation set (remaining 30% of total samples).

To construct the calibration models for the two study areas and evaluate the effect of the different pretreatments, three multivariate calibration methods were used: principal components regression (PCR) [66], partial least squares regression (PLSR) [67], and support vector machine regression (SVMR) [68]. Each method was combined with the three pretreatments techniques mentioned above and a total of 18 prediction models were developed. All prediction models were generated by using the Unscrambler 11 (Aspen Technology, Inc., Bedford, MA, USA).

In the PCR method, principal component analysis (PCA) and multiple linear regression (MLR) are combined to reduce the number of spectral bands, which are potentially correlated, and to solve multicollinearity problems [69]. The PCA decomposes the matrix of the spectral data (**X**) into orthogonal Principal Components (PCs). The first PCs explaining most of the raw data variance will be used as predictors instead of spectral bands against the soil property (**Y**) using MLR.

The PLSR method is widely used to construct predictive models when there are many predictor variables that are highly collinear [16,70]. The PLSR aims to find a few linear combinations which explain most of the variation in both predictors (**X**, spectra) and response variable (**Y**, soil property) [16,66]. A set of orthogonal factors called latent variables, which maximize the covariance between **X** and response variable **Y**, is extracted from the spectra. In practice, **X** and **Y** are decomposed into factor scores and factor loadings in such a way that the most relevant part of the **X**-variation is used for regression and the noise ignored. The choice of the number of factors to be retained is made for each model by leave-one-out cross-validation.

The SVMR is based on statistical learning theory and can efficiently handle large input spaces and works well with sparse data [71]. The SVMR algorithm transfers the raw data into a feature space of higher dimensionality and transforms the original predictors using kernel functions. The objective is identifying an interpolation function by fitting the training data to a tube with radius ε using boundary samples (support vectors) close to the margin defined by ε. A cost function aiming to simultaneously minimize the coefficients' size and the prediction errors allows to obtain the best regression model. The method requires selecting the kernel function, the parameterization of a cost parameter C, and a kernel parameter γ. Here, a linear kernel was used, whereas by cross validation the optimization of the parameters C and ε was solved [41]. More details about the implementation of SVMR for soil spectroscopic modeling can be found in Viscarra Rossel and Behrens [21].

2.6. Performance of Prediction Models

As is well known, the basis of the criteria for assessing the prediction performance of models are the errors obtained from estimating the measured values of SOC (in this case) using them. For this purpose, the classical coefficient of determination (R^2) and the root mean square error of prediction (*RMSE*) were used. The coefficient of determination (R^2) measures how strong the relationship is between measured and estimated SOC values. The higher R^2 is, the better the prediction performance of the model. The RMSE is the parameter commonly used to describe the prediction ability of a model [29] and is calculated as root mean squared error:

$$RMSE = \sqrt{\frac{1}{n}(y_n - \hat{y}_n)^2} \quad (1)$$

where $\hat{y}$ is the predicted value and y the true value, and n the number of samples in the calibration or validation set. Since *RMSE* is a measure of an average error, the smaller it is, the better.

Furthermore, the ratio of performance to interquartile distance (*RPIQ*) was calculated, in order to better represent the data spread and remove any range effect, regardless of the distribution [65]. The *RPIQ* is defined as:

$$RPIQ = \frac{(Q_3 - Q_1)}{RMSE} \quad (2)$$

where Q_3 is the upper quartile and Q_1 is the lower quartile. Comparing models, the best performance is achieved by the model with the largest *RPIQ*.

3. Results

3.1. Descriptive Statistics of SOC and Vis-NIR Spectra Features.

The soil organic carbon concentration data have greater values at the Mongiana area than at Bonis catchment (Table 1). Indeed, the mean (6.26%) at Mongiana as well as the maximum value (13.20%) are higher than the mean (2.66) and maximum value (11.02%) at the Bonis catchment area. Furthermore, the SOC data at Mongiana have a fairly symmetrical distribution with a skewness of just 0.15, while the SOC data at the Bonis catchment have a highly asymmetric distribution with a positive skewness of 2.65. The large asymmetry of the SOC data at the Bonis catchment area is well reflected by the large difference between the maximum SOC value and the upper quartile (Q_3) (Table 1).

In addition, the degree of variation in SOC values is greater in the Bonis catchment (CV = 0.49) than in the Mongiana area (CV = 0.31) (Table 1).

The basic statistics for the calibration and validation sets of SOC at both study areas are enough similar to those of the whole SOC data set (Table 1) indicating their sufficient statistical homogeneity.

Table 1. Summary statistics of soil organic carbon (SOC) at the two study areas of the whole sets of data and calibration and validation sets.

Parameter	Bonis Catchment			Mongiana Area		
	Whole Set	**Calibration Set**	**Validation Set**	**Whole Set**	**Calibration Set**	**Validation Set**
N	135	95	40	231	162	69
Mean (%)	2.66	2.65	2.70	6.26	6.15	6.53
Minimum (%)	0.67	1.07	0.67	1.01	1.01	2.57
Q_1	1.87	1.90	1.77	5.11	4.90	5.34
Median (%)	2.38	2.38	2.40	6.31	6.30	6.45
Q_3	3.21	3.12	3.40	7.44	7.35	7.61
Maximum (%)	11.02	11.02	7.64	13.2	11.80	13.2
St. Dev. (%)	1.30	1.29	1.35	1.89	1.91	1.80
Skewness (-)	2.65	3.25	1.34	0.15	-0.01	0.65
Kurtosis (-)	12.35	17.11	2.40	0.69	0.21	1.61
CV (-)	0.49	0.49	50.00	0.31	0.31	0.28

N = number of observations; St. Dev. = standard deviation; Q_1 = lower quartile; Q_3 = upper quartile; CV = coefficient of variation.

The diffuse reflectance spectra (Figure 2) show similar shape and typical absorption bands near 1400, 1900, and 2200 nm, but a high variation in reflectance intensities. Since the shape and reflectance intensities of the spectral curves are strongly influenced by

different soil constituents (e.g., organic carbon, iron oxides, or soil texture) [29,72,73], the visual inspection of diffuse reflectance spectra reveals that the reflectance response is influenced by SOC concentration, specifically reflectance has a tendency to decrease with increasing SOC and vice versa.

In some soil diffuse reflectance spectra, the absorption peaks related to the bending and stretching of the hydroxyl (OH) features of hygroscopic or free water (close to 1400 and 1900 nm) are small while in others they are large (Figure 2). The same behavior can be observed for absorption at the bands around 2200 nm that are related to the OH of clay minerals lattice (Figure 2).

Clear differences between the diffuse reflectance spectra for the two study areas can be observed and are likely to be mainly due to differences in soil type, mineralogy, and soils physico-chemical composition. The average diffuse reflectance values of soil spectra of the two study areas are significantly different, mainly because of the difference in the mean values of SOC concentration (Figure 2).

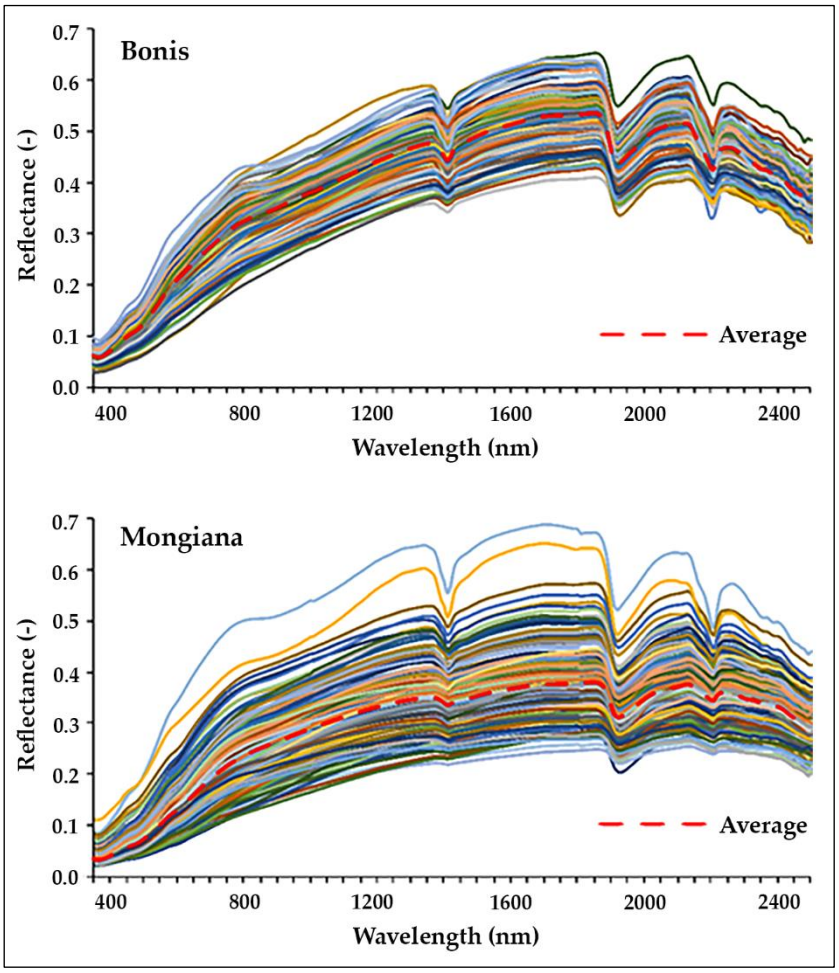

Figure 2. Diffuse reflectance spectra for all topsoil samples at both study areas. For each study area, the average reflectance spectrum (dashed red line) is also shown.

3.2. Performance of Pretreatment Methods on SOC Prediction Models

Eighteen calibration models were obtained using PCR, PLSR, and SVMR coupled with different pretreatment methods. The performances of the 18 models are reported in Table 2. In general, the prediction models obtained different levels of accuracy depending on the pretreatment(s) and multivariate calibration method used. The values of R^2 in the calibration models vary between 0.53 and 0.86 for SOC in the Bonis catchment and between 0.72 and 0.85 in the Mongiana area (Table 2). The *RMSE* values vary between 0.50

and 0.88 for SOC in the Bonis catchment, and between 0.75 and 1.01 in the Mongiana area. Finally, *RPIQ* values range between 1.35 and 2.40 for SOC at the Bonis catchment and between 2.38 to 3.23 at the Mongiana area.

The best calibration model for SOC prediction at the Bonis catchment was obtained by using the SVMR coupled with absorbance spectra pretreated with standard normal variate (SNV), Savitzky–Golay (SG) filtering and the first derivative (1st D). It obtained the highest value of R^2 equal to 0.86 (Table 2), but the lowest RMSE (0.50%) and the highest RIPQ (2.40) were obtained using PLSR with only the transformation of reflectance spectra to absorbance (Table 2). The worst performance resulted for the calibration model obtained using PCR coupled with the spectra of diffuse reflectance pretreated with ABS + SNV + SG 1st D (Table 2).

The best calibration model in the Mongiana area resulted by applying the SVMR method coupled with absorbance spectra pretreated with standard normal variate (SNV), Savitzky–Golay (SG) filtering and the first derivative (1st D) (Table 2). It obtained the highest value of R^2 and RPIQ and lowest RMSE. Finally, the lowest accuracy resulted for the calibration model obtained from the SVMR coupled with the reflectance spectra transformed into absorbance. (Table 2).

Table 2. Performance of the calibration models obtained by PCR, PLSR, and SVMR, coupled with different spectroscopic pretreatment techniques for SOC prediction at the two study areas. The best results of the models are highlighted in bold.

Dataset	Pretreatment Method	R^2 (-)			RMSE (%)			RPIQ (-)		
		PCR	PLSR	SVMR	PCR	PLSR	SVMR	PCR	PLSR	SVMR
Bonis (n = 90)	ABS	0.80	**0.85**	0.75	0.57	**0.50**	0.75	2.09	**2.40**	1.58
	ABS + SNV	0.79	0.81	**0.83**	0.59	**0.56**	0.59	2.03	**2.11**	2.01
	ABS + SNV + SG 1st D	0.53	0.84	**0.86**	0.88	**0.51**	0.55	1.35	**2.32**	2.15
Mongiana (n = 162)	ABS	0.81	**0.82**	0.72	0.83	**0.81**	1.01	2.89	**2.97**	2.38
	ABS + SNV	0.74	0.73	**0.78**	0.98	0.99	**0.90**	2.46	2.43	**2.68**
	ABS + SNV + SG 1st D	0.77	0.78	**0.85**	0.92	0.90	**0.75**	2.62	2.67	**3.23**

The next step was to evaluate the performance of all calibration models obtained using an independent SOC dataset (validation set) for both study areas. The results of this evaluation are summarized in Table 3. The values of R^2 vary from 0.52 to 0.82, whereas *RPIQ* from 1.67 to 2.84 (Table 3). Generally, these results are in accordance with those obtained for the calibration models. The best validation performance for SOC in the Bonis catchment was obtained using PLSR coupled with the soil reflectance spectra transformed into ABS (Table 3). In contrast, for SOC prediction in the Mongiana area, limited predictive abilities were found using the ABS + SNV and ABS + SNV + SG 1st D pretreatment techniques coupled with PCR (Table 3).

Table 3. Validation results of the models obtained by PCR, PLSR, and SVMR, coupled with different spectroscopic pretreatment techniques for SOC prediction at the two study areas. The best results of the models are highlighted in bold.

Dataset	Pretreatment Method	R^2 (-)			RMSE (%)			RPIQ (-)		
		PCR	PLSR	SVMR	PCR	PLSR	SVMR	PCR	PLSR	SVMR
Bonis (n = 45)	ABS	0.79	**0.82**	0.79	0.61	**0.57**	0.68	2.64	**2.84**	2.37
	ABS + SNV	0.70	**0.72**	0.72	0.74	**0.71**	0.76	2.17	**2.25**	2.12
	ABS + SNV + SG 1st D	0.75	0.73	**0.82**	0.67	0.70	**0.63**	2.38	2.30	**2.54**
Mongiana (n = 69)	ABS	0.73	**0.73**	0.52	1.03	**1.01**	1.36	2.21	**2.24**	1.67
	ABS + SNV	0.62	0.67	**0.71**	1.20	1.13	**1.08**	1.89	2.01	**2.11**
	ABS + SNV + SG 1st D	0.62	0.65	**0.77**	1.20	1.16	**0.97**	1.89	1.96	**2.34**

Overall, the validation results show that calibration models built using PLSR and SVMR as methods perform better than those obtained by PCR (Table 3). Therefore, the average R^2 value for the models built with PLSR and SVMR is 0.72, while that for the models built with PCR is 0.70. Based on the R^2 and *RPIQ* values of the validation, the SOC prediction models for the Bonis catchment area performed better than those obtained for the Mongiana area (Figure 3). That might be due to the lower degree of variation of the SOC in terms of coefficient of variation (Table 1) in the Mongiana area compared to those in the Bonis catchment.

In order to summarize the results obtained, Figure 3 shows the scatter plots of the best performance of the SOC prediction models in the two study areas obtained by PLSR and SVMR coupled with each of the three combinations of diffuse reflectance spectra pretreatments.

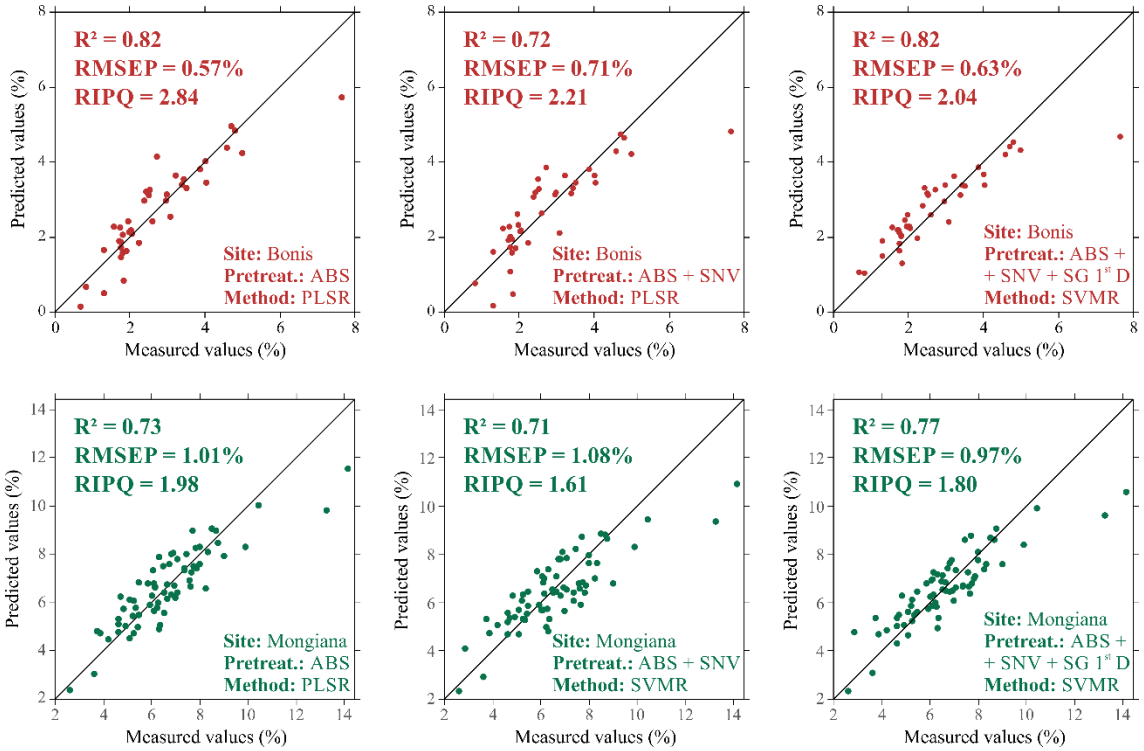

Figure 3. Scatter plots of the best performances of the SOC prediction models obtained by PLSR, and SVMR coupled with each of the three different spectroscopic pretreatment techniques. Above and in red are shown the results for the Bonis catchment area, and below and in green those for the Mongiana area. ABS = Absorbance; SNV = Standard Normal Variate; SG 1st D = Savitzky–Golay 1st order derivative.

4. Discussion

The different degree of spatial variation of the SOC concentration in the two study areas (Table 1) likely affected the spectral response of the soil samples (Figure 2) as also reported in the literature [21,74-76].

In the two study areas, SOC concentrations ranged from 0.67%, in the Bonis basin, to 13.2%, in the Mongiana area (Table 2). The lowest SOC concentrations were observed mainly in the Lithic Udipsamments and secondarily in the Typic Xerumbrepts; whereas the highest SOC values were observed in the Humic Dystrudepts and the Ultic Haploxeralf. Furthermore, the different distribution pattern of SOC can probably be related to the morphological characteristics of the two study areas [54,57,58]. In particular, high and very high concentrations of SOC were found in topsoils collected on the summit palaeosurface, gentle slopes and valley floors. In contrast, low to moderate SOC concentrations (<3%) were measured on convex slopes with slopes generally greater than 20°, which are often affected by water erosion processes [54,57].

The visual inspection of the soil spectra (Figure 2) showed that the Vis-NIR diffuse reflectance spectra of soil samples were similar in appearance with respect to spectral

shape, but there were evident differences in terms of reflectance intensity because variation in SOC content.

Also confirmed in the literature is the apparent similarity in the shape of the diffuse reflectance spectra coupled with the different intensity of the reflectance values (Figure 2) probably related to variations in SOC concentration [24,41,63,77]. The spectral signatures with lower reflectance generally refer to soil samples with higher content of SOC whereas in soil spectra with high reflectance, SOC is relatively low [24,63,76]. Spectra with low diffuse reflectance values have a concave or linear shape, which results in a decrease in the slope of the curve between the 500 and 800 nm domains (Figure 2) that is typical of soils with high SOC concentrations [78,79]. Moreover, the soil reflectance spectra showed important absorption bands that could be related to some soil constituents (Figure 2). Particularly, the absorption bands around 1400 and 1900 nm mainly due to OH and metal-OH bonds of soil clay minerals, and the absorption bands at 2200 nm due to the interaction of near-infrared radiation with lattice minerals [16,78].

The results suggest that coupling Vis-NIR spectroscopy with the three multivariate calibration methods (PCR, PLSR, SVMR) allows to discriminate the spatial variability of SOC content, which represent the key pedogenetic features for evaluate soil quality, as well as soil development or degradation [79]. Moreover, the pretreatments of spectral reflectance data can help to fulfil the assumptions of linear multivariate calibration and improve the calibration models [25]. That would explain the high performance of PLSR. These pretreatment techniques have their own specific purposes as well as effects on the diffuse reflectance spectra and can allow to highlight absorption bands or portion of reflectance spectra more sensitive to SOC or remove noise caused by particle size and light scattering [62,63]. Thus, in a context of great variability of SOC content as in these two study areas, which are characterized by different land cover, soil types, and soil samples size, spectroscopy pretreatments methods may be useful to increase the accuracy of prediction models [33,76,80–82]. For the SOC in the two study areas, the capability of the PCR, PLSR, SVMR methods coupled with three pretreatments (ABS, ABS + SNV, ABS + SNV + SG 1st D) showed different performances, which may be related to the different SOC spatial variation between the two study areas. However, the results also showed that each of the pretreatments used in this study had different effects on the SOC calibration models, either improving or worsening the performance of the prediction models (Table 2). This finding is confirmed by several studies in which spectra pretreatment has been reported to improve the accuracy of SOC prediction models [72,81,82], while other studies show acceptable results without the application of pretreatment to spectral data [72,83,84].

Overall, the lowest variations in performance of the prediction models were obtained for the PLSR method, which consequently provides more stable models for predicting SOC content than PCR and SVMR methods. Additionally, when the raw spectra were transformed into ABS, the PLSR models presented the best performances for both sites (Tables 2 and 3). These results are in agreement with the findings of Carvalho et al. (2022) [63] in which different pretreatments and multivariate calibration methods for predicting SOC by Vis-NIR spectroscopy in southern Brazil were compared. Nevertheless, for both sites studied, coupling SVMR and ABS + SNV + SG 1st D was the best combination of multivariate calibration method and pretreatment techniques (Tables 2).

The performance of the calibration models in the validation fully agreed with that of the calibration models. Consequently, confirming the combination of multivariate calibration methods and spectra preprocessing techniques used in model calibration (Tables 2 and 3). The best accuracy in the SOC predictions of the validation set data was obtained always by SVMR combined with and ABS + SNV + SG 1st D pretreatments.

However, the results of the validation for both areas showed that all the prediction models underestimate the higher SOC concentrations, especially for the PCR models. That may be due to the low number of samples having high SOC concentrations in the calibration set.

The findings highlighted that the different size of soil dataset, as well as the soil sampling density, for the two areas, do not have a direct effect on SOC prediction models. Therefore, the performance of the local prediction models can be mainly related to the SOC range of variation and to the combination of different multivariate methods and the spectral pretreatments used [41,72]. Furthermore, according to Gholizadeh et al. [85] and Vašát et al. [86], applying different pretreatment techniques could improve the results of prediction; the pretreatments might remove physical phenomena in the reflectance spectra and increase the performance of the prediction models or exploratory analysis [33,62]. That suggests the need to test different multivariate calibration methods and investigating the effect of different spectral pretreatments [72,87].

In addition, large soil property datasets generally have larger variances due to the great heterogeneity of soils and the variability of soil morphology and land use/cover; therefore, it is more difficult to obtain high performance prediction models. In these cases, therefore, it may be necessary to use soil and spectral databases with an adequate number of soil samples for different environments and soil types. Conversely, small rather than large spectral databases for estimating SOC at a local scale could provide accurate prediction models [88].

5. Conclusions

This study was carried out in two forested areas of the Calabria region (south Italy) to determine if the performance of calibration models obtained combining Vis-NIR diffuse reflectance spectra and some multivariate calibration methods for predicting SOC content in forest soils might depend on the size of the study area and on the soil sampling density.

It was found that, generally, the performance of the multivariate calibration methods in the prediction of SOC depended on the spectral pretreatments both in calibration and validation procedures. In the model validation, support vector machine regression (SVMR) outperformed with the increase of pretreatments.

Regarding the study areas, in the prediction of SOC at the Bonis catchment (139 ha and sampling density = 0.97 samples/ha), PLSR outperformed with all pretreatments in calibration procedure, whereas in validation PLSR was the best method 2 out 3 pretreatments. The performance of SVMR increased with the number of pretreatments, but PLSR performed slightly better. However, in validation, the absolute best model was obtained by PLSR using absorbance spectra.

In the prediction of SOC at the Mongiana area (33 ha and sampling density = 7 samples/ha), using absorbance spectra, PCR outperformed in the calibration models, whereas PLSR outperformed in validation. Using the other two pretreatments, the best method was SVMR both in calibration and validation. In validation, the absolute best model was obtained by SVMR and absorbance spectra with SNV and SG 1st order derivative.

The results of this study highlighted various aspects of the use of Vis-NIR spectroscopy in the prediction of SOC under two conditions of study area size and sampling density using different calibration methods in combination with a variety of spectra pretreatments. However, there was no result that can absolutely provide definitive indications of either the effects of the study area size and soil sampling density. In fact, a detailed analysis of the results of the calibration and validation models would still allow further discussion on the best performance achieved, but no clear conclusions about the effects of the study area size and soil sampling density in predicting SOC by Vis-NIR spectroscopy. What apparently seems to be a failure of the study actually fosters the need to future investigations in areas and datasets of different sizes from those in this study and including different soil landscapes, which may allow generalizing the results obtained.

Author Contributions: Conceptualization, M.C. and G.B.; soil sampling and laboratory analysis, M.C. and G.B.; formal analysis and data curation, M.C. and G.B.; writing—original draft, M.C. and G.B.; writing—review and editing, M.C. and G.B. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support for this work comes from the project "ALForLab" PON03PE_00024_1 co-funded by the National Operational Programme for Research and Competitiveness (PON R&C) 2007–2013, through the European Regional Development Fund (ERDF) and national resource (Revolving Fund—Cohesion Action Plan (CAP) MIUR) and the project LIFE09 ENV/IT/078 Managing forests for multiple purposes: carbon, biodiversity, and socio-economic wellbeing (ManFor C.BD.).

Data Availability Statement: The data presented in this study are available on request.

Acknowledgments: The authors thank the anonymous reviewers of this paper for providing constructive comments, which have contributed to the improvement of the published version.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Commission. *EU Biodiversity Strategy for 2030. Bringing Nature Back into Our Lives*; Brussels, Belgium, 2020.
2. European Commission. *Forging a Climate-Resilient Europe—The New EU Strategy on Adaptation to Climate Change*; Brussels, Belgium, 2021.
3. European Commission. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, The European Green Deal, COM(2019) 640 Final*; Brussels, Belgium, 2019.
4. European Commission. *New EU Forest Strategy for 2030*; Brussel, Belgium, 2021.
5. Lorenz, K.; Lal, R. *Carbon Sequestration in Forest Ecosystems*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2010; ISBN 9789048132652.
6. Blume, H.-P.; Brümmer, G.W.; Fleige, H.; Horn, R.; Kandeler, E.; Kögel-Knabner, I.; Kretzschmar, R.; Stahr, K.; Wilke, B.-M. Soil Organic Matter. In *Scheffer/SchachtschabelSoil Science*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 55–86; ISBN 978-3-642-30942-7.
7. Minasny, B.; Malone, B.P.; McBratney, A.B.; Angers, D.A.; Arrouays, D.; Chambers, A.; Chaplot, V.; Chen, Z.-S.; Cheng, K.; Das, B.S.; et al. Soil Carbon 4 per Mille. *Geoderma* **2017**, *292*, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002.
8. von Lützow, M.; Kögel-Knabner, I.; Ekschmitt, K.; Matzner, E.; Guggenberger, G.; Marschner, B.; Flessa, H. Stabilization of Organic Matter in Temperate Soils: Mechanisms and Their Relevance under Different Soil Conditions—A Review. *Eur. J. Soil Sci.* **2006**, *57*, 426–445. https://doi.org/10.1111/j.1365-2389.2006.00809.x.
9. Smith, P.; Martino, D.; Cai, Z.; Gwary, D.; Janzen, H.; Kumar, P.; McCarl, B.; Ogle, S.; O'Mara, F.; Rice, C.; et al. Greenhouse Gas Mitigation in Agriculture. *Philos. Trans. R. Soc. B: Biol. Sci.* **2008**, *363*, 789–813. https://doi.org/10.1098/rstb.2007.2184.
10. Nieder, R.; Benbi, D.K. *Carbon and Nitrogen in the Terrestrial Environment*; Springer: Dordrecht, The Netherlands, 2008; Volume 9781402084; ISBN 9781402084331.
11. Murphy, B.W. Impact of Soil Organic Matter on Soil Properties—A Review with Emphasis on Australian Soils. *Soil Res.* **2015**, *53*, 605. https://doi.org/10.1071/sr14246.
12. Ben-Dor, E.; Banin, A. Banin Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. *Soil Sci. Soc. Am. J.* **1995**, *59*, 364. https://doi.org/10.2136/sssaj1995.03615995005900020014x.
13. Ge, Y.; Morgan, C.L.S.; Wijewardane, N.K. Visible and Near-infrared Reflectance Spectroscopy Analysis of Soils. *Soil Sci. Soc. Am. J.* **2020**, *84*, 1495–1502. https://doi.org/10.1002/saj2.20158.
14. Shepherd, K.D.; Walsh, M.G. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988. https://doi.org/10.2136/sssaj2002.9880.
15. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Dor, E.B.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Adv. Agron.* **2015**, *132*, 139–159. https://doi.org/10.1016/bs.agron.2015.02.002.
16. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* **2006**, *131*, 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007.
17. Nanni, M.R.; Demattê, J.A.M. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *Soil Sci. Soc. Am. J.* **2006**, *70*, 393–407. https://doi.org/10.2136/sssaj2003.0285.
18. Demattê, J.; Morgan, C.; Chabrillat, S.; Rizzo, R.; Franceschini, M.; da Silva Terra, F.; Vasques, G.; Wetterlind, J. Spectral Sensing from Ground to Space in Soil Science: State of the Art, Applications, Potential, and Perspectives. In *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*; Thenkabail, P.S., Ed.; CRC Press: Boca Raton, FL, USA, 2015; pp. 661–732; ISBN 9780429089442.

19. Viscarra Rossel, R.A.; Hicks, W.S. Soil Organic Carbon and Its Fractions Estimated by Visible-near Infrared Transfer Functions. *Eur. J. Soil Sci.* **2015**, *66*, 438–450. https://doi.org/10.1111/ejss.12237.
20. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. In *Advances in Agronomy*; Academic Press: Cambridge, MA, USA, 2010; Volume 107, pp. 163–215; ISBN 9780123810335.
21. Viscarra Rossel, R.A.; Behrens, T. Using Data Mining to Model and Interpret Soil Diffuse Reflectance Spectra. *Geoderma* **2010**, *158*, 46–54. https://doi.org/10.1016/J.GEODERMA.2009.12.025.
22. Stevens, A.; van Wesemael, B.; Bartholomeus, H.; Rosillon, D.; Tychon, B.; Ben-Dor, E. Laboratory, Field and Airborne Spectroscopy for Monitoring Organic Carbon Content in Agricultural Soils. *Geoderma* **2008**, *144*, 395–404. https://doi.org/10.1016/j.geoderma.2007.12.009.
23. Ben-Dor, E.; Chabrillat, S.; Demattê, J.A.M.; Taylor, G.R.; Hill, J.; Whiting, M.L.; Sommer, S. Using Imaging Spectroscopy to Study Soil Properties. *Remote Sens. Environ.* **2009**, *113*, S38–S55. https://doi.org/10.1016/j.rse.2008.09.019.
24. Conforti, M.; Matteucci, G.; Buttafuoco, G. Using Laboratory Vis-NIR Spectroscopy for Monitoring Some Forest Soil Properties. *J. Soils Sediments* **2018**, *18*, 1009–1019. https://doi.org/10.1007/s11368-017-1766-5.
25. Gobrecht, A.; Roger, J.-M.; Bellon-Maurel, V. Chapter Four—Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation: A Review. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Cambridge, MA, USA, 2014; Volume 123, pp. 145–175.
26. Reeves, J.B.; McCarty, G.W.; Calderon, F.; Hively, W.D. Chapter 20—Advances in Spectroscopic Methods for Quantifying Soil Carbon. In *Managing Agricultural Greenhouse Gases*; Liebig, M.A., Franzluebbers, A.J., Follett, R.F., Eds.; Academic Press: San Diego, CA, USA, 2012; pp. 345–366; ISBN 978-0-12-386897-8.
27. Næs, T.; Isakson, T.; Fearn, T.; Davies, T.; Ziegel, E.R. *A User-Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Charlton, Chichester, UK, 2004; Volume 46; ISBN 1906715254.
28. Farifteh, J.; van der Meer, F.; Atzberger, C.; Carranza, E.J.M. Quantitative Analysis of Salt-Affected Soil Reflectance Spectra: A Comparison of Two Adaptive Methods (PLSR and ANN). *Remote Sens. Environ.* **2007**, *110*, 59–78. https://doi.org/10.1016/j.rse.2007.02.005.
29. Varmuza, K.; Filzmoser, Peter. *Introduction to Multivariate Statistical Analysis in Chemometrics*; CRC Press: Boca Raton, FL, USA, 2009; ISBN 978-1-4200-5947-2.
30. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2021; ISBN 978-1-0716-1417-4.
31. Westerhaus, M.; Workman Jr., J.; Reeves III, J.B.; Mark, H. Quantitative Analysis. In *Near-Infrared Spectroscopy in Agriculture*; John Wiley & Sons, Ltd.: *Madison, Wisconsin, USA*, 2004; pp. 133–174; ISBN 9780891182368.
32. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard Normal Variate Transformation and De-Trending of near Infrared Diffuse Reflectance Spectra 106. *Appl. Spectrosc.* **1989**, *43*, 772–777. https://doi.org/10.1366/0003702894202201.
33. Vasques, G.M.; Grunwald, S.; Sickman, J.O. Comparison of Multivariate Methods for Inferential Modeling of Soil Carbon Using Visible/near-Infrared Spectra. *Geoderma* **2008**, *146*, 14–25. https://doi.org/10.1016/j.geoderma.2008.04.007.
34. Mouazen, A.M.; Kuang, B.; de Baerdemaeker, J.; Ramon, H. Comparison among Principal Component, Partial Least Squares and Back Propagation Neural Network Analyses for Accuracy of Measurement of Selected Soil Properties with Visible and near Infrared Spectroscopy. *Geoderma* **2010**, *158*, 23–31. https://doi.org/10.1016/j.geoderma.2010.03.001.
35. Stevens, A.; Udelhoven, T.; Denis, A.; Tychon, B.; Lioy, R.; Hoffmann, L.; van Wesemael, B. Measuring Soil Organic Carbon in Croplands at Regional Scale Using Airborne Imaging Spectroscopy. *Geoderma* **2010**, *158*, 32–45. https://doi.org/10.1016/J.GEODERMA.2009.11.032.
36. Guerrero, C.; Zornoza, R.; Gómez, I.; Mataix-Beneyto, J. Spiking of NIR Regional Models Using Samples from Target Sites: Effect of Model Size on Prediction Accuracy. *Geoderma* **2010**, *158*, 66–77. https://doi.org/10.1016/j.geoderma.2009.12.021.
37. Aïchi, H.; Fouad, Y.; Walter, C.; Viscarra Rossel, R.A.; Lili Chabaane, Z.; Sanaa, M. Regional Predictions of Soil Organic Carbon Content from Spectral Reflectance Measurements. *Biosyst. Eng.* **2009**, *104*, 442–446. https://doi.org/10.1016/j.biosystemseng.2009.08.002.
38. Williams, P. C. (2001). Implementation of near-infrared technology. In: Williams P.C. and Norris K. (editors) Near-Infrared Technology in the Agricultural and Food Industries, 145–169. American Association of Cereal Chemists: St. Paul, Minnesota, USA. ISBN: 1891127241.
39. Ramirez-Lopez, L.; Schmidt, K.; Behrens, T.; van Wesemael, B.; Demattê, J.A.M.; Scholten, T. Sampling Optimal Calibration Sets in Soil Infrared Spectroscopy. *Geoderma* **2014**, *226–227*, 140–150. https://doi.org/10.1016/j.geoderma.2014.02.002.
40. Minasny, B.; McBratney, A.B. Conditioned Latin Hypercube Sampling for Calibrating Soil Sensor Data to Soil Properties. In *Proximal Soil Sensing*; Springer: Dordrecht, The Netherlands, 2010; pp. 111–119; ISBN 978-90-481-8859-8.
41. Lucà, F.; Conforti, M.; Castrignanò, A.; Matteucci, G.; Buttafuoco, G. Effect of Calibration Set Size on Prediction at Local Scale of Soil Carbon by Vis-NIR Spectroscopy. *Geoderma* **2017**, *288*, 175–183. https://doi.org/10.1016/j.geoderma.2016.11.015.
42. Zimmermann, M.; Leifeld, J.; Fuhrer, J. Quantifying Soil Organic Carbon Fractions by Infrared-Spectroscopy. *Soil Biol. Biochem.* **2007**, *39*, 224–231. https://doi.org/10.1016/j.soilbio.2006.07.010.
43. Vasques, G.M.; Grunwald, S.; Sickman, J.O. Modeling of Soil Organic Carbon Fractions Using Visible–Near-Infrared Spectroscopy. *Soil Sci. Soc. Am. J.* **2009**, *73*, 176. https://doi.org/10.2136/sssaj2008.0015.

44. Clairotte, M.; Grinand, C.; Kouakoua, E.; Thébault, A.; Saby, N.P.A.; Bernoux, M.; Barthès, B.G. National Calibration of Soil Organic Carbon Concentration Using Diffuse Infrared Reflectance Spectroscopy. *Geoderma* **2016**, *276*, 41–52. https://doi.org/10.1016/j.geoderma.2016.04.021.
45. McCarty, G.W.; Reeves, J.B.; Reeves, V.B.; Follett, R.F.; Kimble, J.M. Mid-Infrared and Near-Infrared Diffuse Reflectance Spectroscopy for Soil Carbon Measurement. *Soil Sci. Soc. Am. J.* **2002**, *66*, 640–646. https://doi.org/10.2136/sssaj2002.0640.
46. Colombo, C.; Palumbo, G.; di Iorio, E.; Sellitto, V.M.; Comolli, R.; Stellacci, A.M.; Castrignanò, A. Soil Organic Carbon Variation in Alpine Landscape (Northern Italy) as Evaluated by Diffuse Reflectance Spectroscopy. *Soil Sci. Soc. Am. J.* **2014**, *78*, 794. https://doi.org/10.2136/sssaj2013.11.0488.
47. Cozzolino, D.; Morón, A. Potential of Near-Infrared Reflectance Spectroscopy and Chemometrics to Predict Soil Organic Carbon Fractions. *Soil Tillage Res.* **2006**, *85*, 78–85. https://doi.org/10.1016/j.still.2004.12.006.
48. Ben-Dor, E.; Inbar, Y.; Chen, Y. The Reflectance Spectra of Organic Matter in the Visible Near-Infrared and Short Wave Infrared Region (400–2500 Nm) during a Controlled Decomposition Process. *Remote Sens. Environ.* **1997**, *61*, 1–15. https://doi.org/10.1016/S0034-4257(96)00120-4.
49. Buttafuoco, G.; Castrignanò, A. Study of the Spatio-Temporal Variation of Soil Moisture under Forest Using Intrinsic Random Functions of Order k. *Geoderma* **2005**, *128*, 208–220. https://doi.org/10.1016/j.geoderma.2005.04.004.
50. le Pera, E.; Sorriso-Valvo, M. Weathering and Morphogenesis in a Mediterranean Climate, Calabria, Italy. *Geomorphology* **2000**, *34*, 251–270. https://doi.org/10.1016/S0169-555X(00)00012-X.
51. Molin, P.; Fubelli, G.; Dramis, F. Evidence of Tectonic Influence on Drainage Evolution in an Uplifting Area: The Case of Northern Sila (Calabria, Italy). *Geogr. Fis. E Din. Quat.* **2012**, *35*, 49–60. https://doi.org/10.4461/GFDQ.2012.35.5.
52. Luca, F.; Robustelli, G.; Conforti, M.; Fabbricatore, D. Geomorphological Map of the Crotone Province (Calabria, South Italy). *J. Maps* **2011**, *7*, 375–390. https://doi.org/10.4113/jom.2011.1190.
53. Soil Survey Staff. *Keys to Soil Taxonomy*, 12th ed.; USDA-Natural Resources Conservation Service: Washington, DC, USA, 2014.
54. ARSSA. Carta Dei Suoli Della Regione Calabria—Scala 1:250000. In *Monografia Divulgativa; Servizio Agropedologia.* Agenzia Regionale per Lo Sviluppo e per i Servizi in Agricoltura: Soveria Mannelli, Italy, 2003.
55. Conforti, M.; Matteucci, G.; Buttafuoco, G. Organic Carbon and Total Nitrogen Topsoil Stocks, Biogenetic Natural Reserve 'Marchesale' (Calabria Region, Southern Italy). *J. Maps* **2017**, *13*, 91–99. https://doi.org/10.1080/17445647.2016.1262795.
56. Calcaterra, D.; Parise, M.; Dattola, L. Caratteristiche Dell'alterazione e Franosita Di Rocce Granitoidi Nel Bacino Del Torrente Alaco (Massiccio Delle Serre, Calabria). *Boll. Della Soc. Geol. Ital.* **1996**, *115*, 3–28.
57. Conforti, M.; Longobucco, T.; Scarciglia, F.; Niceforo, G.; Matteucci, G.; Buttafuoco, G. Interplay between Soil Formation and Geomorphic Processes along a Soil Catena in a Mediterranean Mountain Landscape: An Integrated Pedological and Geophysical Approach. *Environ. Earth Sci.* **2020**, *79*, 59. https://doi.org/10.1007/s12665-019-8802-2.
58. Conforti, M.; Lucà, F.; Scarciglia, F.; Matteucci, G.; Buttafuoco, G. Soil Carbon Stock in Relation to Soil Properties and Landscape Position in a Forest Ecosystem of Southern Italy (Calabria Region). *Catena* **2016**, *144*, 23–33. https://doi.org/10.1016/j.catena.2016.04.023.
59. Conforti, M.; Froio, R.; Matteucci, G.; Buttafuoco, G. Visible and near Infrared Spectroscopy for Predicting Texture in Forest Soil: An Application in Southern Italy. *IForest* **2015**, *8*, 339–347. https://doi.org/10.3832/ifor1221-007.
60. The International Organization for Standardization. ISO 18400-101:2017—Soil Quality—Sampling—Part 101: Framework for the Preparation and Application of a Sampling Plan 2017; pp. 1–15.
61. Viscarra Rossel, R.A. ParLeS: Software for Chemometric Analysis of Spectroscopic Data. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 72–83. https://doi.org/10.1016/j.chemolab.2007.06.006.
62. Rinnan, Å.; Van Den Berg, F.; Engelsen, S.B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.
63. Carvalho, J.K.; Moura-Bueno, J.M.; Ramon, R.; Almeida, T.F.; Naibo, G.; Martins, A.P.; Santos, L.S.; Gianello, C.; Tiecher, T. Combining Different Pre-Processing and Multivariate Methods for Prediction of Soil Organic Matter by near Infrared Spectroscopy (NIRS) in Southern Brazil. *Geoderma Reg.* **2022**, *29*, e00530. https://doi.org/10.1016/j.geodrs.2022.e00530.
64. Savitzky, Abraham.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. https://doi.org/10.1021/ac60214a047.
65. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.M.; McBratney, A.B. Critical Review of Chemometric Indicators Commonly Used for Assessing the Quality of the Prediction of Soil Attributes by NIR Spectroscopy. *TrAC Trends. Anal. Chem.* **2010**, *29*, 1073–1081.
66. Martens, H.; Næs, T. *Multivariate Calibration*; John Wiley & Sons Inc: Chichester, UK, 1991; ISBN 978-0-471-93047-1.
67. Geladi, P.; Kowalski, B.R. Partial Least-Squares Regression: A Tutorial. *Anal Chim Acta* **1986**, *185*, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9.
68. Schölkopf, B.; Smola, A.J. *Learning with Kernels*; The MIT Press: London, UK, 2018; ISBN 9780262536578.
69. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.
70. Cozzolino, D.; Morón, A. The Potential of Near-Infrared Reflectance Spectroscopy to Analyse Soil Chemical and Physical Characteristics. *J. Agric. Sci.* **2003**, *140*, 65–71. https://doi.org/10.1017/S0021859602002836.

71. Vohland, M.; Besold, J.; Hill, J.; Fründ, H.-C. Comparing Different Multivariate Calibration Methods for the Determination of Soil Organic Carbon Pools with Visible to near Infrared Spectroscopy. *Geoderma* **2011**, *166*, 198–205. https://doi.org/10.1016/j.geoderma.2011.08.001.
72. Moura-Bueno, J.M.; Dalmolin, R.S.D.; Horst-Heinen, T.Z.; ten Caten, A.; Vasques, G.M.; Dotto, A.C.; Grunwald, S. When Does Stratification of a Subtropical Soil Spectral Library Improve Predictions of Soil Organic Carbon Content? *Sci. Total Environ.* **2020**, *737*, 139895. https://doi.org/10.1016/j.scitotenv.2020.139895.
73. Viscarra Rossel, R.A.; Bui, E.N.; de Caritat, P.; McKenzie, N.J. Mapping Iron Oxides and the Color of Australian Soil Using Visible-near-Infrared Reflectance Spectra. *J. Geophys. Res. Earth Surf.* **2010**, *115*, *1-13*. https://doi.org/10.1029/2009JF001645.
74. Gomez, C.; Lagacherie, P.; Coulouma, G. Continuum Removal versus PLSR Method for Clay and Calcium Carbonate Content Estimation from Laboratory and Airborne Hyperspectral Measurements. *Geoderma* **2008**, *148*, 141–148. https://doi.org/10.1016/j.geoderma.2008.09.016.
75. Clark, R. N., & Wiley. (1999). Spectroscopy of rocks and minerals and principles of spectroscopy. In A. N. Rencz (Ed.), Manual of remote sensing, Volume 3, Remote sensing for the earth sciences (pp. 3–58). John Wiley & Sons Inc: New York
76. Moura-Bueno, J.M.; Dalmolin, R.S.D.; ten Caten, A.; Dotto, A.C.; Demattê, J.A.M. Stratification of a Local VIS-NIR-SWIR Spectral Library by Homogeneity Criteria Yields More Accurate Soil Organic Carbon Predictions. *Geoderma* **2019**, *337*, 565–581. https://doi.org/10.1016/j.geoderma.2018.10.015.
77. Song, Y.; Li, F.; Yang, Z.; Ayoko, G.A.; Frost, R.L.; Ji, J. Diffuse Reflectance Spectroscopy for Monitoring Potentially Toxic Elements in the Agricultural Soils of Changjiang River Delta, China. *Appl. Clay Sci.* **2012**, *64*, 75–83. https://doi.org/10.1016/j.clay.2011.09.010.
78. Riefolo, C.; Castrignanò, A.; Colombo, C.; Conforti, M.; Ruggieri, S.; Vitti, C.; Buttafuoco, G. Investigation of Soil Surface Organic and Inorganic Carbon Contents in a Low-Intensity Farming System Using Laboratory Visible and near-Infrared Spectroscopy. *Arch. Agron. Soil Sci.* **2020**, *66*, 1436–1448. https://doi.org/10.1080/03650340.2019.1674446.
79. Conforti, M.; Buttafuoco, G.; Leone, A.P.A.P.; Aucelli, P.P.P.C.C.; Robustelli, G.; Scarciglia, F. Studying the Relationship between Water-Induced Soil Erosion and Soil Organic Matter Using Vis-NIR Spectroscopy and Geomorphological Analysis: A Case Study in Southern Italy. *Catena* **2013**, *110*, 44–58. https://doi.org/10.1016/j.catena.2013.06.013.
80. Knox, N.M.; Grunwald, S.; McDowell, M.L.; Bruland, G.L.; Myers, D.B.; Harris, W.G. Modelling Soil Carbon Fractions with Visible Near-Infrared (VNIR) and Mid-Infrared (MIR) Spectroscopy. *Geoderma* **2015**, *239*, 229–239. https://doi.org/10.1016/j.geoderma.2014.10.019.
81. Pinheiro, É.; Ceddia, M.; Clingensmith, C.; Grunwald, S.; Vasques, G. Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon. *Remote Sensor* **2017**, *9*, 293. https://doi.org/10.3390/rs9040293.
82. Dotto, A.C.; Dalmolin, R.S.D.; ten Caten, A.; Grunwald, S. A Systematic Study on the Application of Scatter-Corrective and Spectral-Derivative Preprocessing for Multivariate Prediction of Soil Organic Carbon by Vis-NIR Spectra. *Geoderma* **2018**, *314*, 262–274. https://doi.org/10.1016/j.geoderma.2017.11.006.
83. Araújo, S.R.; Wetterlind, J.; Demattê, J.A.M.; Stenberg, B. Improving the Prediction Performance of a Large Tropical Vis-NIR Spectroscopic Soil Library from Brazil by Clustering into Smaller Subsets or Use of Data Mining Calibration Techniques. *Eur. J. Soil Sci.* **2014**, *65*, 718–729. https://doi.org/10.1111/ejss.12165.
84. Wijewardane, N.K.; Ge, Y.; Wills, S.; Loecke, T. Prediction of Soil Carbon in the Conterminous United States: Visible and Near Infrared Reflectance Spectroscopy Analysis of the Rapid Carbon Assessment Project. *Soil Sci. Soc. Am. J.* **2016**, *80*, 973–982. https://doi.org/10.2136/sssaj2016.02.0052.
85. Gholizadeh, A.; Borůvka, L.; Saberioon, M.M.; Kozák, J.; Vašát, R.; Němeček, K. Comparing Different Data Preprocessing Methods for Monitoring Soil Heavy Metals Based on Soil Spectral Features. *Soil Water Res.* **2016**, *10*, 218–227. https://doi.org/10.17221/113/2015-SWR.
86. Vašát, R.; Kodešová, R.; Klement, A.; Borůvka, L. Simple but Efficient Signal Pre-Processing in Soil Organic Carbon Spectroscopic Estimation. *Geoderma* **2017**, *298*, 46–53. https://doi.org/10.1016/j.geoderma.2017.03.012.
87. Angelopoulou, T.; Balafoutis, A.; Zalidis, G.; Bochtis, D. From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review. *Sustainability* **2020**, *12*, 443. https://doi.org/10.3390/su12020443.
88. Guerrero, C.; Wetterlind, J.; Stenberg, B.; Mouazen, A.M.; Gabarrón-Galeote, M.A.; Ruiz-Sinoga, J.D.; Zornoza, R.; Viscarra Rossel, R.A. Do We Really Need Large Spectral Libraries for Local Scale SOC Assessment with NIR Spectroscopy? *Soil Tillage Res.* **2016**, *155*, 501–509. https://doi.org/10.1016/j.still.2015.07.008.