

## Research Article

# Predicting Soil Organic Carbon Content Using Hyperspectral Remote Sensing in a Degraded Mountain Landscape in Lesotho

**Freddy Bangelesa,<sup>1</sup> Elhadi Adam ,<sup>1</sup> Jasper Knight,<sup>1</sup> Inos Dhau,<sup>2</sup> Marubini Ramudzuli,<sup>2</sup> and Thabiso M. Mokotjomela<sup>1</sup>**

<sup>1</sup>*School of Geography Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa*

<sup>2</sup>*Department of Geography and Environmental Studies, University of Limpopo, Sovenga 0727, South Africa*

Correspondence should be addressed to Elhadi Adam; [elhadi.adam@wits.ac.za](mailto:elhadi.adam@wits.ac.za)

Received 4 September 2019; Accepted 11 March 2020; Published 12 April 2020

Academic Editor: Teodoro M. Miano

Copyright © 2020 Freddy Bangelesa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Soil organic carbon constitutes an important indicator of soil fertility. The purpose of this study was to predict soil organic carbon content in the mountainous terrain of eastern Lesotho, southern Africa, which is an area of high endemic biodiversity as well as an area extensively used for small-scale agriculture. An integrated field and laboratory approach was undertaken, through measurements of reflectance spectra of soil using an Analytical Spectral Device (ASD) FieldSpec® 4 optical sensor. Soil spectra were collected on the land surface under field conditions and then on soil in the laboratory, in order to assess the accuracy of field spectroscopy-based models. The predictive performance of two different statistical models (random forest and partial least square regression) was compared. Results show that random forest regression can most accurately predict the soil organic carbon contents on an independent dataset using the field spectroscopy data. In contrast, the partial least square regression model overfits the calibration dataset. Important wavelengths to predict soil organic contents were localised around the visible range (400–700 nm). This study shows that soil organic carbon can be most accurately estimated using derivative field spectroscopy measurements and random forest regression.

## 1. Introduction

Soil organic carbon (SOC) is an important property related to soil biological, physical, and chemical characteristics and constitutes a major component of the global carbon cycle [1]. SOC is classified as the third most important global carbon sink (>2500 Pg) and contains almost double the carbon that is found in the atmosphere (750 Pg) and terrestrial living biomass (560 Pg) combined [2–6]. In agricultural landscapes, SOC depletion as a result of accelerated soil erosion can lead to reduced crop yields, lowered moisture retention capacity, and reduced nutrient status [7–10]. SOC also contributes to stabilization of the soil and formation of aggregates, which can promote resistance to erosion [11, 12]. In mountain landscapes, however, soil physical properties are highly variable spatially, where changes in slope steepness, depth of weathering products, slope processes, and

microclimate give rise to variations in vegetation types and soil properties and high rates of soil erosion [13].

In the mountainous highlands of Lesotho, southern Africa, soil erosion is a significant problem due to a combination of geologic, climatic, ecological, and human factors [14–16]. Weathering of the underlying Jurassic basalts has produced a low-strength mixture of silica and expansive clay minerals [17]. Plagioclase within the basalts has been affected by zeolitization and chloritization, and olivine in particular has been replaced by iron oxides, serpentine, and clays (mainly montmorillonite) [18]. These weathering products make the resulting soil susceptible to erosion by surface sheet flow, subsurface clay expansion, slaking and soil piping, and landslide/debris flow activity caused by subsurface waterlogging and failure (e.g., [19]). Although many studies have been concerned with calculating soil volume loss by erosion through gullies (locally known as dongas)

and evaluating change in gully morphology over time (e.g., [20, 21], few studies have considered organic carbon export as part of the sediment yield [22, 23]. This is important, however, because SOC export depletes the remaining nutrient base of soils within the catchment and can lead to lower agricultural productivity as well as environmental degradation, and SOC can change downstream trophic status and may lead to eutrophication and other negative consequences within aquatic ecosystems.

The first step in quantifying potential SOC export from a catchment is to map spatial patterns of SOC storage in surface soils. This is because (1) SOC content is highest in the soil A horizon, nearest the land surface, and (2) overland or sheet flow leading to enhanced soil erosion first affects this A horizon, which is therefore preferentially lost by erosion. Accurate measurement of SOC at different spatial scales is challenging [24] because it has been most commonly based on grid sampling of soil in the field. However, these techniques are expensive, time-consuming, and not spatially continuous [25], and a close sampling interval is required to capture SOC spatial patterns effectively. In mountain environments, high-density sampling resolution is particularly difficult to achieve because of poor accessibility and high relief and also because of the high variability of soils, ecosystems, and SOC found in these environments. Therefore, there is a growing demand for effective methods in quantifying SOC in mountain environments.

Remote sensing techniques offer a cost-effective, reproducible, and rapid method of quantifying spatially distributed data on SOC [26]. This is possible through the correlation between soil reflectance and soil organic content. Previous investigations show that increasing values of SOC are inversely proportional to an overall decrease in reflectance in the visible (Vis, 400–700 nm), near-infrared (NIR, 700–1400 nm), and shortwave infrared (SWIR, 1400–2500 nm) regions of the electromagnetic spectrum [27]. McMorrow et al. [28] observed absorption from 677 nm to 1108 nm which is associated with SOC and iron oxides. Other absorption features are present within the range of NIR and SWIR, related to lignin and cellulose at 1120 nm and 2100 nm, respectively. Airborne and satellite platforms have largely contributed to the assessment of SOC by examining these different spectral signatures, but one of the drawbacks of these technologies is that they cannot discriminate between carbon from surface vegetation and soils, and they generate a mixed pixel signal. It is also difficult to estimate SOC using satellite and airborne remote sensing methods when SOC concentrations within the soil are small because it results in a very weak signal [29].

Laboratory and field spectroscopy methods in the visible and near-infrared bands (Vis-NIR, 400–2500 nm) are therefore advantageous in measuring and modelling SOC content [30]. These approaches are both rapid and nondestructive [31]. Under controlled laboratory conditions, SOC content can be estimated with high precision and accuracy [32], whereas field-based measurements may be affected by atmospheric conditions, soil moisture, texture, and shadow effects [33]. Some researchers have reported satisfactory

results when using Vis-NIR spectra to predict SOC in the field [34, 35]; however, results vary according to soil type and moisture content and thus cannot be applied everywhere. Viscarra Rossel and Behrens [29] argued that soil absorption spectral signatures are likely to overlap and vary spatially and temporally. This is a motivation to develop SOC models for different regions or ecological/geomorphic contexts. There are few studies addressing SOC modelling using either laboratory or field measurements in southern Africa [35], especially in mountainous regions where SOC is vulnerable to land use change mainly by overgrazing, soil erosion, and climate change.

This study therefore sought to (i) estimate SOC content using spectroscopy measurements under field and laboratory conditions, (ii) compare the performance of two different modelling approaches (partial least squares regression (PLSR) and nonlinear random forest (RF)) in predicting SOC content, and (iii) evaluate the role of spectral derivatives in determining model outputs. The purpose of this analysis is to identify better modelling approaches for SOC using remotely sensed data, which will enable a quicker and more accurate evaluation of SOC, in particular in poorly known areas such as mountains.

## 2. Materials and Methods

*2.1. Study Area.* The study area is in eastern Lesotho, southern Africa (Figure 1). In this region, mountains of the Drakensberg-Maluti range rise to 3482 m a.s.l. at Thabana Ntlenyana, with river valleys incised into Jurassic basalts, and low-nutrient xerophytic grasslands present on flat-topped mountain summit [36, 37]. The mean annual rainfall is 775 mm with 85% falling in the summer season [38]. The mean winter monthly temperatures range from  $-6.3^{\circ}\text{C}$  to  $5.1^{\circ}\text{C}$ , and the mean summer maximum temperatures range from  $16.5^{\circ}\text{C}$  at high altitudes to  $29^{\circ}\text{C}$  in the lowlands [39]. Lesotho has a highly degraded mountain landscape [40, 41].

Most of the population (86%) depends on subsistence agriculture [42], and increased population and climate change have meant that cattle and sheep grazing has expanded from river valleys to high elevation pastures, leading to overgrazing and a loss of ecological integrity in these areas [43, 44]. Consequently, agricultural productivity has been declining in Lesotho [45].

*2.2. Soil Sampling and Spectral Measurements.* Fieldwork was conducted along the Mokhotlong River in eastern Lesotho (Figure 1) in October 2015 (austral spring/summer). This river flows east to west and has incised through Jurassic basalts, giving rise to a highly meandering river pattern with bedrock spurs and steep valley sides. The river floodplain is very narrow with small strip agricultural fields located adjacent to the river channel and, where slopes sediments are available, terraced fields are present along lower valley slopes (Figure 2). Higher elevation areas at the tops of valley sides and on plateau summits ( $\sim 2600$ – $2900$  m a.s.l.) are not enclosed and are characterised by tussocky grasslands [37].

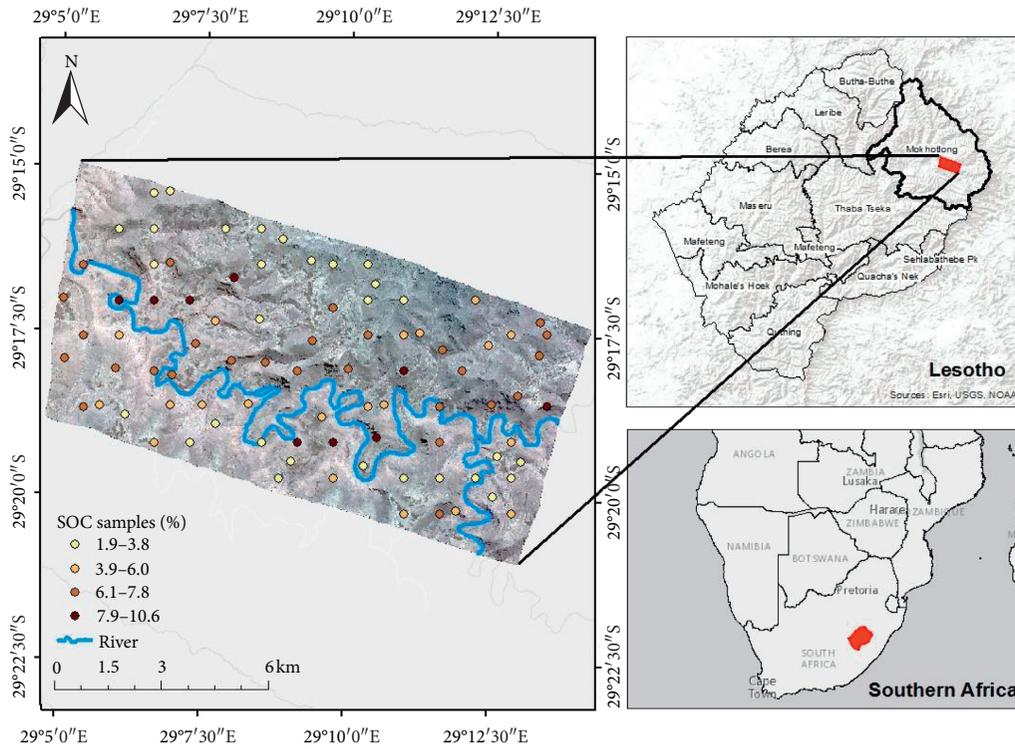


FIGURE 1: Location of the study area in eastern Lesotho, showing the positions of soil samples/ASD data points, and soil organic carbon values ( $\text{g } 100 \text{ g}^{-1}$ ).

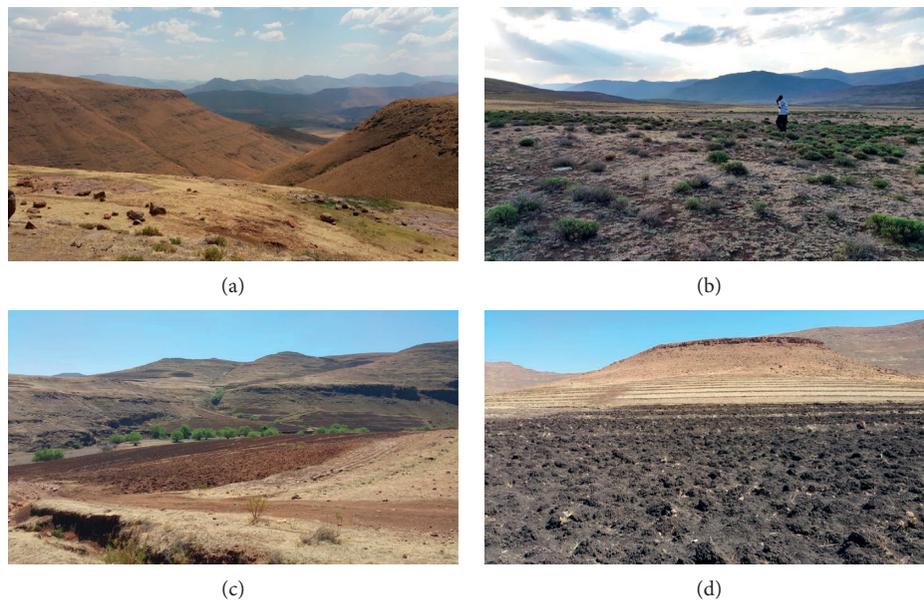


FIGURE 2: Field photos from eastern Lesotho, showing (a) flat basalt plateau surfaces with deeply incised river valleys, (b) xerophytic and low-nutrient shrubby grassland with a degraded land surface, (c) view of agricultural fields on lower valley slopes; note the presence of trees along river margins and the absence of significant vegetation on upper valley slopes, and (d) terraced fields along lower valley slopes, replaced by grassland on steeper upper slopes. A basalt depositional surface is preserved around the hill summit.

2.3. *Spectral Measurements in the Field.* Soil spectral reflectance was measured in the field using an Analytical Spectral Device (ASD) FieldSpec® 4 optical sensor (Analytical Spectral Devices, Inc., Boulder, CO, USA). This instrument measures wavelengths from 350 to 2500 nm and

with 3–10 nm spectral resolution. Spectra were recorded for the region 350–1000 nm with a sampling interval of 1.4 nm, and 2 nm for the region 1000–2500 nm. ASD sampling points in the study area ( $n = 109$ ) were generated randomly using Hawth’s Analysis Tool within ArcMap. All points were

converted to latitude and longitude, and a handheld GPS was used to navigate to these locations in the field. Once the sampling point was located, a plot of 10 m by 10 m was drawn where the coordinates of the sampling point were considered as the centroid of the plot. Within each plot, 3 subplots of 2 m by 2 m dimensions were randomly selected in order to take into consideration any variability within the plot. Five spectral measurements from the nadir at about 1 m height and with 5° field of view above the soil surface were scanned in each subplot and averaged, giving a total of 15 spectral measurements for each plot. For every measurement, the white reference panel was used to calibrate atmosphere conditions and irradiance of the sun. A surface (top 5 cm) soil sample of ~400 g was taken from each of these three subplots for subsequent laboratory analysis.

**2.4. Spectral Measurements in the Laboratory.** Spectral measurements on the 109 soil samples collected in the field were made in the laboratory. All measurements were done on samples that had been air-dried and lightly crushed in a pestle and mortar to a uniform fine sediment. Measurements were made on a black background plate. The soils were scanned using the ASD with the white reference panel for comparison and with exactly the same methodology as in the field.

**2.5. SOC Analysis.** In the laboratory, soil samples were first dried and lightly crushed. On this mixed sample, the loss on ignition (LOI) method was used to quantify SOC concentration [46]. A subsample (~20 g) was combusted in a muffle furnace for 8 hours at 430°C. LOI was measured as the difference between the oven-dry soil mass and the soil mass after combustion, divided by the oven-dry soil mass [47]. LOI values were converted to SOC with a factor of 0.55 [48].

**2.6. Spectra Preprocessing and Transformation.** Before modelling, the noisy ends of the spectra were removed in order to correct for low-intensity radiation appearing at the spectra edge. The spectrum below 400 nm was removed. The water vapour absorption features ranging between 1350–1460, 1790–1960, and 2350–2500 nm which can affect the model were also removed [49, 50]. The same operation was done with laboratory spectral data for standardization. The laboratory and field spectral first derivative transformation and Savitzky–Golay smoothing were applied in order to enhance the spectral signal [51]. Savitzky–Golay smoothing was first applied for reducing the noise effect before derivative transformation because derivative spectra are sensitive to noise [52] (Figure 3). The same methodology was undertaken for both field and laboratory spectral measurements.

**2.7. Statistical Analysis.** All statistical analysis was implemented in R program v3.1.3 [53]. Before modelling, 70% of the data (training dataset) were randomly selected for training the models for both laboratory and field data and the rest (30%) as the testing dataset. The normality of SOC values within the datasets was checked using the

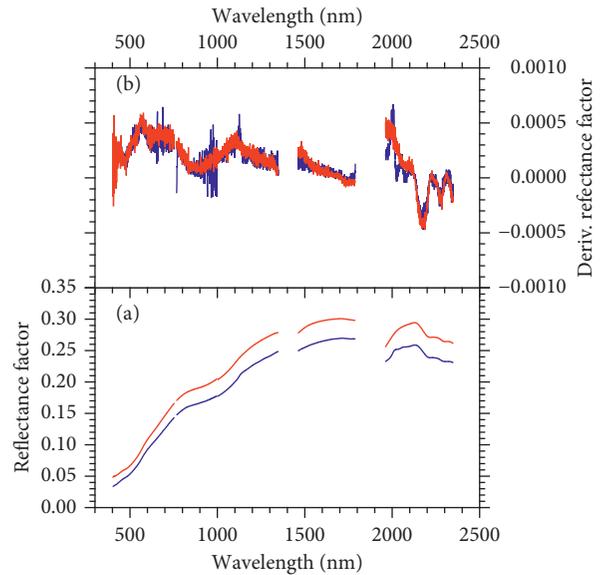


FIGURE 3: Spectral data with removed noisy regions (300–399, 1350–1640, 1790–1960, and 2350–2500 nm): (a) untransformed and (b) first derivative transformation. Blue lines denote field spectral data and red lines denote laboratory spectral data.

Kolmogorov–Smirnov goodness-of-fit test in order to select the most suitable statistical test. We used the Kruskal–Wallis test [54] to compare the training, testing, and whole datasets because of the lack of normality in the testing dataset. Student’s  $t$ -test was then used to compare the averaged field spectra with the averaged laboratory spectral data, which were both normally distributed. In order to identify outliers, Hotelling’s  $T^2$  distribution for multivariate analysis was performed with the field spectral data [55].

**2.7.1. Partial Least Squares Regression (PLSR).** The PLSR method was implemented in order to construct predictive models when the independent variables are many, noisy, and highly collinear such as hyperspectral reflectance data [56]. The method uses orthogonal factors or components, called latent variables, as new independent variables of the dependent variable. These latent variables are simply linear combinations of the original independent variables, but in contrast to principal component regression, they are extracted such that they explain as much of the covariance between the dependent and independent variables as possible. PLSR is discussed in detail by Martens and Næs [57]. The optimum number of factors through the leave-one-out cross-validation method was used in order to minimize overfitting [58]. The root mean square error of cross-validation (RMSECV) was used in order to evaluate the optimal number of components that minimizes the RMSECV.

PLSR uses many algorithms for feature selection. In this research, the variable importance in projection (VIP) was used in order to identify key wavelengths. The logic of VIP is to accumulate the importance of each variable being reflected by the weight from each component. VIP provides a list of ranked variables and a threshold of between 0.83 and 1.21 which is used to select key wavelengths [59]. In this

study, wavelengths where peak maxima were above a VIP threshold of 1 were selected as key wavelengths [60].

**2.7.2. Random Forest (RF) Regression.** The RF regression is a machine learning algorithm based on a classification and regression tree [61, 62]. The model uses recursive partitioning to split the data (spectra) into different homogeneous groups, named regression trees (ntree). Each tree is individually grown to its optimum size based on a bootstrap sample from the training dataset (70%) without any pruning (a continuous selection of input variables at every node). In RF regression, a random subset of variables (mtry) is selected to determine the split at each node [61]. The model uses a deterministic algorithm to select the number of random samples and variables from the training dataset. In each tree, the data that are not included in the tree (the out-of-bag (OOB) data, 30% of the total dataset) are predicted and the OOB error is produced in terms of mean square errors through the difference between OOB data and data used to grow the regression trees [61, 63]. The OOB error provides an estimation of the important variables by calculating how much the OOB error is increased when a variable is changed, while all others remain unchanged [64]. This attribute enables the operator to select or train the RF model to focus on certain features. In this study, the RF algorithm was implemented for both field and laboratory spectral datasets. Because of the large number of variables, model optimization which is computationally intense was not implemented. The default setting for mtry (1/3 of the total number of wavelengths) and ntree (500) was used.

The recursive feature selection [65] was performed to determine the least number of wavelengths that predict SOC concentration with greatest accuracy. The recursive feature elimination algorithm is a wrapper feature selection method that uses all features (variables) as a starting point [66]. Models with low accuracy are removed from the current subset. The procedure ends when the given numbers of variables are dropped [66]. The combination of the recursive feature selection with the important variables ranked according to the percent decrease in mean squared error helped us to identify key wavelengths.

**2.8. Model Validation.** Many parameters can be used to assess the performance of models in spectroscopy. Spectra models are generally assessed in terms of their coefficient of determination ( $R^2$ ) and root mean square error (RMSE). Other parameters such as the Akaike information criterion (AIC) and the ratio prediction to deviation (RPD) can also be used. In this study, all of these parameters were used for completeness. The root mean square error of calibration (RMSEC) and validation (RMSEP) are calculated as follows:

$$\text{RMSEC} = \sqrt{\frac{\sum (y_m - y_p)^2}{N}}, \quad (1)$$

$$\text{RMSEP} = \sqrt{\frac{\sum (y_m - y_v)^2}{N}},$$

where  $y_m$  are measured values from the laboratory measurement,  $y_p$  are predicted values derived from spectral data using either PLSR or RF,  $y_v$  are predicted values obtained using the validation set, and  $N$  is the number of samples. The model with a lowest coefficient RMSE has the best performance. The AIC is a compromise between model accuracy and model parsimony [67] and is calculated as follows:

$$\text{AIC} = n \ln \text{RMSE} + 2p, \quad (2)$$

where  $n$  is the number of samples and  $p$  the number of features used in the prediction. The model with the smallest AIC has the best performance. The RPD was also used to compare the performance of models of different datasets, as well as their practicability. This metric is a way of normalizing the RMSEs of prediction in order to compare calibration models where the measured variables have different variances [68], and is calculated as follows:

$$\text{RPD} = \frac{\text{STDEV}(y)}{\text{RMSEP}}, \quad (3)$$

where  $\text{STDEV}(y)$  refers to the standard deviation of the reference data (calibration dataset) and  $\text{RMSEP}$  refers to the root mean square error of prediction.

The six categories of interpretation as suggested by Viscarra Rossel et al. [69] were adopted as follows: an RPD value of  $>2.5$  indicates excellent models/predictions;  $2.0 < \text{RPD} < 2.5$  indicates very good quantitative models/predictions;  $1.8 < \text{RPD} < 2.0$  indicates good models/predictions, where quantitative predictions are possible;  $1.4 < \text{RPD} < 1.8$  indicates fair models/predictions which may be used for assessment and correlation;  $1.0 < \text{RPD} < 1.4$  indicates poor models/predictions, where only high and low values are distinguishable; and  $\text{RPD} < 1.0$  indicated very poor models/predictions, and their uses are not distinguishable.

### 3. Results

**3.1. SOC Sample Analysis.** Of the 109 soil samples collected, 94 were analysed for SOC content, excluding outliers. Hotelling's test detected 4 outliers with a horizontal cut off limit of 6.32845 and a vertical cut off limit of 3.98695. The statistical description of SOC of the calibration dataset, validation dataset, and the whole dataset is presented in Table 1. SOC values ranged from  $1.93 \text{ g } 100 \text{ g}^{-1}$  to  $10.6 \text{ g } 100 \text{ g}^{-1}$  with a mean value of  $5.04 \text{ g } 100 \text{ g}^{-1}$  and a standard deviation of 2.11. The Kolmogorov–Smirnov test shows that all datasets were not normally distributed ( $p < 0.05$  values of 0.0126, 0.0020, and 0.2100 for the whole dataset, calibration dataset, and validation dataset, respectively). All subsets have a skewed distribution. The Kruskal–Wallis test for skewed distributions indicates that there are no significant differences among the three datasets at 5% significant level ( $p$  value = 0.64). Thus, both calibration and validation datasets statistically represent the whole dataset.

**3.2. Comparison of Field and Laboratory Spectra.** The mean reflectance values of field and laboratory measurements were computed for all 94 samples. In general, the reflectance

TABLE 1: Descriptive statistics of SOC (in  $\text{g } 100 \text{ g}^{-1}$ ) values within the whole, calibration, and validation datasets.

Dataset	$N$	Min	Max	Mean	Median	Std	Skewness	Kurtosis
Whole dataset	94	1.93	10.66	5.14	4.97	2.13	0.47	2.27
Calibration dataset	65	2.17	9.57	4.97	4.87	1.99	0.42	2.04
Validation dataset	29	1.94	10.67	5.55	5.55	2.42	0.39	2.19

values of the SOC measured in the laboratory are higher than those measured in the field, confirmed by a one-tailed Student's  $t$ -test (significance at 5% level,  $p$  value = 0.024). Pearson's correlation test reveals that both spectra measurement are strongly correlated at 5% significant level ( $R = 0.99$ ,  $p$  value  $< 2.2e^{-16}$ ).

### 3.3. Key Wavelength Selection

**3.3.1. Number of Variables Selected by VIP.** VIP algorithms were computed with PLSR for both field and laboratory spectra. The algorithm selected 80, 720, 57, and 881 key wavelengths for the first derivative laboratory spectra, the raw laboratory spectra, the first derivative field spectra, and raw field spectra, respectively. The location of 57 selected wavelengths of the first derivative field spectra is presented in Figure 4.

**3.3.2. Number of Variables Selected by the Recursive Feature Selection.** The recursive feature selection was computed for both field and laboratory spectra. The smallest number of wavelengths for each model that would offer the best prediction using the random forest method was identified. The lowest number of key wavelengths (49) was obtained with the first derivative field spectra with a RMSE of  $0.61 \text{ g } 100 \text{ g}^{-1}$ ; the location of these wavelengths is presented in Figure 4. The use of 105 wavelengths produces the lowest RMSE ( $0.62 \text{ g } 100 \text{ g}^{-1}$ ) for the first derivative laboratory spectral data. For the laboratory raw spectral data, the use of 105 variables produced the lowest RMSE ( $0.61 \text{ g } 100 \text{ g}^{-1}$ ). For the field raw spectral data, the use of 789 variables produced the lowest cross-validation error (1.059).

**3.3.3. The Position of Key Wavelengths.** Table 2 shows the position of key wavelengths selected by both the VIP algorithm and the recursive feature selection method for all combinations of models. For interpretation purposes, the functional group and vibration mode of wavelength are also presented. The recursive feature selection algorithm selects around half of the key wavelengths in all datasets between the range 400 and 700 nm. The VIP algorithm implemented under the first derivative spectral data (laboratory and field) also selects most of the key wavelengths in the same range. However, the VIP algorithm implemented under the raw spectral data (field and laboratory) did not select key wavelengths within the visible spectrum, and most wavelengths selected are around 2000–2200, 1400–1500, and 1350–1450 nm.

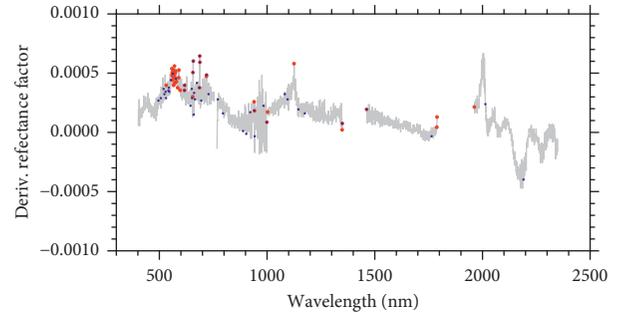


FIGURE 4: Field spectra first derivative and location of key wavelength selected by VIP (red dots) and RF recursive feature selection (blue dots).

**3.4. Model Development.** In total, 16 models have been developed, 8 with the field spectral data and 8 with the laboratory spectral data. For each spectral data (field or laboratory), 4 models were developed with RF and 4 others for PLSR using different transformations: raw data, the first derivative (FD), key wavelengths (K), and the combination of first derivative and key (FD-K) wavelengths. RPD was used as the most important criterion in ranking those models.

**3.4.1. Models from Laboratory Spectral Data.** Results from the two different modelling approaches used to predict SOC (PLSR and RF) using the laboratory spectral data are presented in Table 3. The effects of wavelength selection and the spectral first-derivative on the model are also presented. The best prediction model is selected according to its performance with respect to fitting to the validation dataset.

The results show that PLSR overfits the calibration dataset with  $R_c^2$  between 0.99 and 0.89, but with relatively low accuracy in predicting a new dataset. PLSR models are very good models/prediction ( $2.0 < \text{RPD} < 2.5$ ), whereas RF models are considered as excellent models/predictions ( $\text{RPD} > 2.5$ ). The best predictive model for the laboratory data was obtained with L-FD-RF-K ( $\text{RPD} = 3.77$ ,  $R_p^2 = 0.87$ , and  $\text{RMSEP} = 0.64 \text{ g } 100 \text{ g}^{-1}$ ).

**3.4.2. Models from Field Spectral Data.** Table 4 shows the 8 different models developed using the field datasets. The results indicate that PLSR models are likely to overfit the calibration dataset. All RF models show excellent models/predictions ( $\text{RPD} > 2.5$ ). The best prediction model for the field spectra data was achieved with F-FD-RF-K ( $\text{RPD} = 3.77$ ,  $R_p^2 = 0.88$ , and  $\text{RMSEP} = 0.64 \text{ g } 100 \text{ g}^{-1}$ ), followed by F-FD-RF ( $\text{RPD} = 3.03$ ,  $R_p^2 = 0.89$ , and  $\text{RMSEP} = 0.79 \text{ g } 100 \text{ g}^{-1}$ ). As for RF algorithms, first

TABLE 2: Position of key wavelengths selected by both the VIP algorithm and the recursive feature selection method for all combinations of models.

Wavelength (nm)	Possible assignment	VIP (PLSR)				RF			
		Lab		Field		Lab		Field	
		Raw	FD	Raw	FD	Raw	FD	Raw	FD
2200–2450	Comb C-H stret	—	—	—	—	+	—	—	—
2000–2200	Comb N-H stret, comb O-stret	++	—	—	—	+	+	—	+
1790–1960	Water	+	+	+	+	+	—	+	—
1650–1780	1st overt C-H stret	+	—	++	—	—	—	+	—
1400–1500	1st overt N-H stret and O-H stret	++	—	++	—	—	—	++	—
1300–1420	Comb C-H stret	+	—	—	—	+	—	+	—
1350–1460	Water	+	—	++	+	—	—	—	+
1100–1225	2nd overt C-H stret	+	—	+	—	+	++	+	+
950–1100	2nd overt N-H stret and O-H stret	—	—	—	+	+	+	+	+
850–950	3rd overt C-H stret	+	—	+	+	++	+	++	+
775–850	3rd overt N-H stret	+	—	+	+	++	+	+	+
400–700	Mineral (Fe oxides)	—	+++	+	+++	+++	+++	+++	+++

TABLE 3: Performance of all RF regression and PLSR models in the calibration and validation laboratory datasets (RMSEC and RMSEP in  $g\ 100\ g^{-1}$ ).

Models		Calibration set				Validation set		
	Model	Pretreatment	RMSEC	$R_c^2$	AIC	RMSEP	$R_p^2$	RPD
Lab spectra	PLSR	None	0.60	0.90	126	0.87	0.86	2.27
		FD	0.46	0.94	90.98	0.99	0.82	2.00
		K	0.12	0.99	82	0.95	0.84	2.10
		FD-K	0.63	0.89	131.8	0.99	0.82	2.00
	RF	None	0.65	0.89	—	0.79	0.82	3.03
		K	0.80	0.84	—	0.79	0.86	3.03
		FD	0.64	0.89	—	0.79	0.84	3.03
		FD-K	0.59	0.90	—	0.64	0.87	3.77

TABLE 4: Performance of all RF regression and PLSR models in the calibration and validation field datasets (RMSEC and RMSEP in  $g\ 100\ g^{-1}$ ).

Models		Calibration set				Validation set		
	Model	Pretreatment	RMSEC	$R_c^2$	AIC	RMSEP	$R_p^2$	RPD
Field spectra	PLSR	None	0.89	0.99	-123	1.05	0.80	1.88
		FD	0.550	0.99	-186	0.88	0.86	2.26
		K	0.67	0.86	63	0.75	0.86	2.97
		FD-K	0.64	0.89	132	1.04	0.80	1.90
	RF	None	0.9	0.79	-	0.79	0.76	3.03
		FD	0.74	0.85	-	0.79	0.89	3.03
		K	0.87	0.80	-	0.87	0.77	3.03
		FD-K	0.60	0.90	-	0.64	0.88	3.77

derivative and key wavelength selection improve the performance of F-PLSR. Compared to F-PLSR (RPD = 1.88,  $R_p^2 = 0.80$ , and RMSEP =  $1.05\ g\ 100\ g^{-1}$ ), both F-FD-PLSR (RPD = 2.26,  $R_p^2 = 0.86$ , and RMSEP =  $0.88\ g\ 100\ g^{-1}$ ) and F-PLSR-K (RPD = 2.27,  $R_p^2 = 0.86$ , and RMSEP =  $0.75\ g\ 100\ g^{-1}$ ) predict SOC content in an independent dataset with higher accuracy.

#### 4. Discussion

This study highlights the complexity of relationships between field and laboratory measurements of the spectral signatures of soils and their relationships to SOC. The

remote sensing approach adopted here is suitable for inaccessible and geomorphically variable mountain landscapes. The complexity of the study area is depicted by high variability of soil properties including SOC values (Table 1). This can be explained by the geomorphological context of the study area which is characterised by steep bedrock-controlled slopes with stepped profiles, underfit river valleys [70], and nutrient-poor soils with limited bioavailability of nitrogen [37]. These soils accumulate on flat basalt weathered surfaces and in association with river floodplains and terraces. Subsistence farmers use these latter locations for small-scale agriculture (Figure 2).

High soil variability leads to high  $R^2$ , RPD, and RMSEP values [71, 72]. This can also explain why results from our RF models are more accurate than the results that Viscarra Rossel and Behrens [29] found using 72 wavebands under laboratory conditions and slightly better than what Nawar and Mouazen [73] have recently found under laboratory conditions. Their best predictive RF model had a  $R^2$  of 0.84, RMSEP of  $0.14 \text{ g } 100 \text{ g}^{-1}$ , and RPD = 2.55, with a relatively low range of SOC in the whole dataset. However, the results we have obtained with the PLSR model are within the range that Viscarra Rossel and Behrens [29] and Li et al. [74] found under similar laboratory conditions, and Stevens et al. [75] and Viscarra Rossel and Behrens [29] found under similar field conditions, using the same regression approaches.

Comparing RF to PLSR, our results found that RF outperforms PLSR in predicting SOC values in an independent dataset. The predictive performance of RF can be explained by the fact that machine learning algorithms are less sensitive to nonlinear and complex data [29]. The weakness of PLSR to accurately predict SOC values in a new dataset has also been reported in some studies (e.g., [75, 76]).

It is difficult to compare the performance of laboratory and field models because they both achieve very similar results with slightly different combinations of input variables. However, Li et al. [74] found that the PLSR model achieved better accuracy within their laboratory spectral data than in their field data. The results also revealed that the reflectance of laboratory measurements is visually higher than the field reflectance. This is because of the presence of soil moisture in the field [77] which increases the forward scattering of light and enhances absorption at all wavelengths [78]. However, Viscarra Rossel et al. [79] found no significant difference between field and laboratory measurements because they removed the spectral bands for water absorption. This may be a suitable method for accounting for variations in soil moisture in field samples.

The effect of spectral derivative and key wavelength selection on improving the raw model was evident in this study. This has also been reported by other researchers (e.g., [80, 81]). Peng et al. [81] assessed the impact of 8 different preprocessing methods and showed that including the first derivative achieved the best result. Li et al. [74] demonstrated the superiority of the first derivative with SG smoothing to improve PLSR. Vasques et al. [80] also used SG smoothing to improve SOC model results. Viscarra Rossel and Behrens [29] improved the RF method by using a discrete wavelet transform algorithm as a feature selection method. These previous studies show the importance of key wavelength selection. Our results show that wavelengths between 400 and 700 nm are most important to predict SOC content (Figure 4), which is also in accordance with some previous studies [29, 82].

According to Viscarra Rossel and Hicks [83], the visible portion of the soil spectrum is mostly related to Fe oxide, either goethite or haematite. Different soil chromophores including chlorophyll, tannis, humic substances, iron oxide, and clay minerals may also explain the high correlation between spectral data and SOC in the visible region [34]. Viscarra Rossel et al. [79] found correlations with wavelengths around 410, 570, and 660 nm in the visible part of the

spectrum. Under laboratory conditions, Wang et al. [72] reported 440, 560, 625, 740, and 1336 nm as the principal spectral bands to predict SOC. Nocita et al. [84] suggested that the spectral region between 580 and 680 nm was sufficient to predict SOC. Our results show some other important wavelengths at around 2000–2200 and 1400–1500 nm. Molecular vibration and rotation of organic functional groups are the main factors leading to absorption in the NIR region. Wavelengths around 2000–2200 nm, for example, may be attributed to the effects of carbonyl C=O/CH stretch vibration [83] or clay minerals [85] (Table 2). Spectral peaks around 1455 nm can be attributed to the presence of water but with the phenomenon of spectral overlapping, as discussed by Viscarra Rossel and Behrens [29]. Nevertheless, Stuart [86] attributed the spectral portion between 1400 and 1500 nm to the first overtone N-H stretching and first overtone O-H stretching.

## 5. Conclusions and Wider Implications

This study shows that SOC values in degrading mountainous landscapes can be predicted using different modelling approaches based on field and laboratory spectral measurements. The best model was shown to be the RF because the PLSR model was more likely to overfit the calibration dataset. There were also some slight differences when these models were applied to the field and laboratory data: the PLSR model was slightly better with the laboratory compared to the field data, whereas there was no difference with the RF model. The best model results were obtained with transformed spectral data, with the key wavelengths to predict SOC values mostly localised around the visible range (400–700 nm). These results are significant because they show that different models can produce results of varying accuracy, even when based on the same datasets. This has implications for operator choice when it comes to dataset analysis for the purpose of accurately predicting SOC values.

The results of this study also have wider implications for the accurate prediction of SOC content in degraded mountain landscapes such as Lesotho. The spatial variability in SOC values (Figure 1) shows that a field-based approach alone is unlikely to be able to accurately capture this variability, even with high-resolution sampling. The application of different spectral analytical techniques means that better predictive models for SOC content can be established, using satellite and not just ground-based remote sensing data, as described in this study. This will allow for a better understanding of spatial patterns of SOC in inaccessible mountain landscapes. Furthermore, quantifying carbon storage within soils, and coupling this to land surface models of soil erosion (e.g. [13]), can mean that carbon export from mountains can be calculated. This is critical for carbon budgeting as well as evaluation of soil nutrient status.

## Data Availability

The soil analysis and the spectral data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by NRF grant 90526 (to JK).

## References

- [1] T. J. Battin, S. Luyssaert, L. A. Kaplan, A. K. Aufdenkampe, A. Richter, and L. J. Tranvik, "The boundless carbon cycle," *Nature Geoscience*, vol. 2, no. 9, pp. 598–600, 2009.
- [2] C. Tarnocai, J. G. Canadell, E. A. G. Schuur, P. Kuhry, G. Mazhitova, and S. Zimov, "Soil organic carbon pools in the northern circumpolar permafrost region," *Global Biogeochemical Cycles*, vol. 23, no. 2, 2009.
- [3] L. Condron, C. Stark, M. O'Callaghan, P. Clinton, and Z. Huang, "The role of microbial communities in the formation and decomposition of soil organic matter," in *Soil Microbiology and Sustainable Crop Production*, G. R. Dixon and E. L. Tilston, Eds., pp. 81–118, Springer Science, Berlin, Germany, 2010.
- [4] J. A. J. Dungait, D. W. Hopkins, A. S. Gregory, and A. P. Whitmore, "Soil organic matter turnover is governed by accessibility not recalcitrance," *Global Change Biology*, vol. 18, no. 6, pp. 1781–1796, 2012.
- [5] P. Smith, "Soils and climate change," *Current Opinion in Environmental Sustainability*, vol. 4, no. 5, pp. 539–544, 2012.
- [6] D. Wang, S. Chakraborty, D. C. Weindorf et al., "Synthesized use of VisNIR DRS and PXRF for soil characterization: total carbon and total nitrogen," *Geoderma*, vol. 243–244, pp. 157–167, 2015.
- [7] A. J. Mills and M. V. Fey, "Declining soil quality in South Africa: effects of land use on soil organic matter and surface crusting," *South African Journal of Plant and Soil*, vol. 21, no. 5, pp. 388–398, 2004.
- [8] R. Lal, "Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands," *Land Degradation & Development*, vol. 17, no. 2, pp. 197–209, 2006.
- [9] D. Hillel and C. Rosenzweig, *Handbook of Climate Change and Agroecosystems*, Imperial College Press, London, UK, 2010.
- [10] D. S. Powlson, P. J. Gregory, W. R. Whalley et al., "Soil management in relation to sustainable agriculture and ecosystem services," *Food Policy*, vol. 36, pp. S72–S87, 2011.
- [11] L. Ferreras, E. Gomez, S. Toresani, I. Firpo, and R. Rotondo, "Effect of organic amendments on some physical, chemical and biological properties in a horticultural soil," *Bioresource Technology*, vol. 97, no. 4, pp. 635–640, 2006.
- [12] O. A. Deraghme, J. R. Jensen, and C. T. Petersen, "Near-saturated hydraulic properties in the surface layer of a sandy loam soil under conventional and reduced tillage," *Soil Science Society of America Journal*, vol. 72, pp. 1728–1737, 2008.
- [13] H. J. Smith, A. J. van Zyl, A. S. Claassens, J. L. Schoeman, and M. C. Laker, "Soil loss modelling in the Lesotho highlands water project catchment areas," *South African Geographical Journal*, vol. 82, no. 2, pp. 64–69, 2000.
- [14] S. Grab and M. Nüsser, "Towards an integrated research approach for the Drakensberg and Lesotho mountain environments: a case study from the Sani plateau region," *South African Geographical Journal*, vol. 83, no. 1, pp. 64–68, 2001.
- [15] J. N. Mbata, "Land use practices in Lesotho: implications for sustainability in agricultural production," *Journal of Sustainable Agriculture*, vol. 18, no. 2–3, pp. 5–24, 2001.
- [16] M. E. Meadows and M. T. Hoffman, "The nature, extent and causes of land degradation in South Africa: legacy of the past, lessons for the future?" *Area*, vol. 34, no. 4, pp. 428–437, 2002.
- [17] F. G. Bell and D. R. Haskins, "A geotechnical overview of Katse Dam and transfer tunnel, Lesotho, with a note on basalt durability," *Engineering Geology*, vol. 46, no. 2, pp. 175–198, 1997.
- [18] E. Garzanti, M. Padoan, M. Setti, A. López-Galindo, and I. M. Villa, "Provenance versus weathering control on the composition of tropical river mud (southern Africa)," *Chemical Geology*, vol. 366, pp. 61–74, 2014.
- [19] R. J. Edwards, W. N. Ellery, and J. Dunlevey, "The role of the in situ weathering of dolerite on the formation of a peatland: the origin and evolution of Dartmoor Vlei in the KwaZulu-Natal Midlands, South Africa," *CATENA*, vol. 143, pp. 232–243, 2016.
- [20] M. P. W. Sonneveld, T. M. Everson, and A. Veldkamp, "Multi-scale analysis of soil erosion dynamics in KwaZulu-Natal, South Africa," *Land Degradation & Development*, vol. 16, no. 3, pp. 287–301, 2005.
- [21] J. Keay-Bright and J. Boardman, "Evidence from field-based studies of rates of soil erosion on degraded land in the central Karoo, South Africa," *Geomorphology*, vol. 103, no. 3, pp. 455–465, 2009.
- [22] C. Mchunu and V. Chaplot, "Land degradation impact on soil carbon losses through water erosion and CO<sub>2</sub> emissions," *Geoderma*, vol. 177–178, pp. 72–79, 2012.
- [23] D. Müller-Nedebock, P. Chivenge, and V. Chaplot, "Selective organic carbon losses from soils by sheet erosion and main controls," *Earth Surface Processes and Landforms*, vol. 41, no. 10, pp. 1399–1408, 2016.
- [24] N. J. Kuhn, T. Hoffmann, W. Schwanghart, and M. Dotterweich, "Agricultural soil erosion and global carbon cycle: controversy over?" *Earth Surface Processes and Landforms*, vol. 34, 2009.
- [25] E. Goidts and B. van Wesemael, "Regional assessment of soil organic carbon changes under agriculture in Southern Belgium (1955–2005)," *Geoderma*, vol. 141, no. 3–4, pp. 341–354, 2007.
- [26] R. J. Gehl and C. W. Rice, "Emerging technologies for *in situ* measurement of soil carbon," *Climatic Change*, vol. 80, no. 1–2, pp. 43–54, 2007.
- [27] A. H. Al-Abbas, P. H. Swain, and M. F. Baumgardner, "Relating organic matter and clay content to the multispectral radiance of soils," *Soil Science*, vol. 114, no. 6, pp. 477–485, 1972.
- [28] J. M. McMorrow, M. E. J. Cutler, M. G. Evans, and A. Al-Roichdi, "Hyperspectral indices for characterizing upland peat composition," *International Journal of Remote Sensing*, vol. 25, no. 2, pp. 313–325, 2004.
- [29] R. A. Viscarra Rossel and T. Behrens, "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma*, vol. 158, no. 1–2, pp. 46–54, 2010.
- [30] J. Wetterlind and B. Stenberg, "Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples," *European Journal of Soil Science*, vol. 61, no. 6, pp. 823–843, 2010.
- [31] R. Guénon, M. Vennetier, N. Dupuy, S. Roussos, A. Pailler, and R. Gros, "Trends in recovery of Mediterranean soil chemical properties and microbial activities after infrequent

- and frequent wildfires,” *Land Degradation & Development*, vol. 24, no. 2, pp. 115–128, 2013.
- [32] M. Cohen, R. S. Mylavarapu, I. Bogrekci, W. S. Lee, and M. W. Clark, “Reflectance spectroscopy for routine agronomic soil analyses,” *Soil Science*, vol. 172, no. 6, pp. 469–485, 2007.
- [33] J. B. Reeves, “Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done?” *Geoderma*, vol. 158, no. 1-2, pp. 3–14, 2010.
- [34] A. M. Mouazen, M. R. Maleki, J. De Baerdemaeker, and H. Ramon, “On-line measurement of some selected soil properties using a VIS-NIR sensor,” *Soil and Tillage Research*, vol. 93, no. 1, pp. 13–27, 2007.
- [35] M. Nocita, L. Kooistra, M. Bachmann, A. Müller, M. Powell, and S. Weel, “Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern Cape Province of South Africa,” *Geoderma*, vol. 167-168, pp. 295–302, 2011.
- [36] Bureau of Statistics and Planning, *Lesotho Agricultural Situation Report (1982–2004)*, Bureau of Statistics and Planning, Maseru, Lesotho, 2007.
- [37] J. Knight, S. W. Grab, and C. Carbutt, “Influence of mountain geomorphology on alpine ecosystems in the Drakensberg alpine centre, southern Africa,” *Geografiska Annaler: Series A, Physical Geography*, vol. 100, no. 2, pp. 140–162, 2018.
- [38] T. K. Saha, “Impact of climate change on agricultural production in Lesotho: a case study,” *African Crop Science Conference Proceedings*, vol. 10, pp. 273–277, 2011.
- [39] S. J. Mason and M. R. Jury, “Climatic variability and change over southern Africa: a reflection on underlying processes,” *Progress in Physical Geography: Earth and Environment*, vol. 21, no. 1, pp. 23–50, 1997.
- [40] S. W. Grab and C. Deschamps, “Geomorphological and geoecological controls and processes following gully development in alpine mires, Lesotho,” *Arctic, Antarctic and Alpine Research*, vol. 36, pp. 49–58, 2004.
- [41] K. B. Showers, *Imperial Gullies: Soil Erosion and Conservation in Lesotho*, Ohio University Press, Athens OH, USA, 2005.
- [42] N. S. Eash, D. M. Lambert, M. V. Marake, C. Thierfelder, F. R. Walker, and M. D. Wilcox, *Small-holder Adoption of Conservation Agriculture in Lesotho and Mozambique*, Huazhong Agricultural University, Wuhan, China, 2013.
- [43] M. Nüsser, “Pastoral utilization and land cover change: a case study from the Sanqebethu Valley, eastern Lesotho,” *Erdkunde*, vol. 56, no. 2, pp. 207–221, 2002.
- [44] M. Nüsser and S. W. Grab, “Land degradation and soil erosion in the eastern Highlands of Lesotho, southern Africa,” *Die Erde*, vol. 133, pp. 291–311, 2002.
- [45] L. Sicili, *Conservation Agriculture and Sustainable Crop Intensification in Lesotho*, Vol. 10, FAO, Rome, Italy, 2010.
- [46] M. E. Konen, P. M. Jacobs, C. L. Burras, B. J. Talaga, and J. A. Mason, “Equations for predicting soil organic carbon using loss-on-ignition for north central U.S. soils,” *Soil Science Society of America Journal*, vol. 66, no. 6, pp. 1878–1881, 2002.
- [47] E. E. Schulte and B. G. Hopkins, “Estimation of organic matter by weight loss-on-ignition,” in *Soil Organic Matter: Analysis and Interpretation*, F. R. Magdoff, M. A. Tabatabai, and E. A. Hanlon, Eds., SSSA, Madison, WI, USA, pp. 21–31, 1996.
- [48] M. J. J. Hoogsteen, E. A. Lantinga, E. J. Bakker, J. C. J. Groot, and P. A. Tittonell, “Estimating soil organic carbon through loss on ignition: effects of ignition conditions and structural water loss,” *European Journal of Soil Science*, vol. 66, no. 2, pp. 320–328, 2015.
- [49] M.-L. Smith, M. E. Martin, L. Plourde, and S. V. Ollinger, “Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: comparison between an airborne (aviris) and a spaceborne (hyperion) sensor,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1332–1337, 2003.
- [50] P. S. Thekabail, E. A. Enclona, M. S. Ashton, C. Legg, and M. J. De Dieu, “Hyperion, IKONOS, ALI and ETM+ sensors in the study of African rainforests,” *Remote Sensing of Environment*, vol. 90, no. 1, pp. 23–43, 2004.
- [51] A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least squares procedures,” *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [52] F. Tsai and W. Philpot, “Derivative analysis of hyperspectral data,” *Remote Sensing of Environment*, vol. 66, no. 1, pp. 41–51, 1998.
- [53] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [54] W. H. Kruskal and W. A. Wallis, “Use of ranks in one-criterion variance analysis,” *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.
- [55] J. E. Jackson, *A User’s Guide to Principal Components*, p. 592pp, John Wiley & Sons, New York, NY, USA, 1991.
- [56] S. Wold, “PLS for multivariate linear modeling,” in *Chemometric Methods in Molecular Design*, H. van de Waterbeemd, Ed., pp. 195–218, VCH, Weinheim, Germany, 1995.
- [57] H. Martens and T. Naes, “Assessment, validation and choice of calibration method,” in *Multivariate Calibration*, H. Martens, Ed., pp. 237–266, John Wiley & Sons, New York, NY, USA, 1989.
- [58] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, NY, USA, 1993.
- [59] T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, “A review of variable selection methods in Partial Least Squares Regression,” *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 62–69, 2012.
- [60] I.-G. Chong and C.-H. Jun, “Performance of some variable selection methods when multicollinearity is present,” *Chemometrics and Intelligent Laboratory Systems*, vol. 78, no. 1-2, pp. 103–112, 2005.
- [61] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [62] A. Liaw and M. Weiner, “Classification and regression by random forest,” *R News*, vol. 2, pp. 18–22, 2002.
- [63] J. Maindonald and W. J. Braun, *Data Analysis and Graphics Using R: An Example-Based Approach*, Cambridge University Press, Cambridge, UK, 2nd edition, 2006.
- [64] K. J. Archer and R. V. Kimes, “Empirical characterization of random forest variable importance measures,” *Computational Statistics & Data Analysis*, vol. 52, no. 4, pp. 2249–2260, 2008.
- [65] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [66] J. Kittler, “Feature set search algorithms,” in *Pattern Recognition and Signal Processing*, C. H. Chen, Ed., pp. 41–60, Elsevier, Amsterdam, Netherlands, 1978.
- [67] H. Akaike, “Maximum likelihood identification of Gaussian autoregressive moving average models,” *Biometrika*, vol. 60, no. 2, pp. 255–265, 1973.
- [68] A. B. McBratney, B. Minasny, and G. Tranter, “Necessary meta-data for pedotransfer functions,” *Geoderma*, vol. 160, no. 3-4, pp. 627–629, 2011.
- [69] R. A. Viscarra Rossel, D. J. J. Walvoort, A. B. McBratney, L. J. Janik, and J. O. Skjemstad, “Visible, near infrared, mid

- infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties,” *Geoderma*, vol. 131, no. 1-2, pp. 59–75, 2006.
- [70] J. Knight and S. W. Grab, “Drainage network morphometry and evolution in the eastern Lesotho highlands, southern Africa,” *Quaternary International*, vol. 470, pp. 4–17, 2018.
- [71] B. Stenberg, R. A. Viscarra Rossel, A. M. Mouazen, and J. Wetterlind, “Visible and near infrared spectroscopy in soil science,” *Advances in Agronomy*, vol. 107, pp. 163–215, 2010.
- [72] J. Wang, T. He, C. Lv, Y. Chen, and W. Jian, “Mapping soil organic matter based on land degradation spectral response units using Hyperion images,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 12, pp. S171–S180, 2010.
- [73] S. Nawar and A. M. Mouazen, “On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning,” *Soil and Tillage Research*, vol. 190, pp. 120–127, 2019.
- [74] S. Li, Z. Shi, S. Chen et al., “In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai-tibet plateau,” *Environmental Science & Technology*, vol. 49, no. 8, pp. 4980–4987, 2015.
- [75] A. Stevens, B. van Wesemael, G. Vandenschrack, S. Touré, and B. Tychon, “Detection of carbon stock change in agricultural soils using spectroscopic techniques,” *Soil Science Society of America Journal*, vol. 70, no. 3, pp. 844–850, 2006.
- [76] C.-W. Chang, D. A. Laird, M. J. Mausbach, and C. R. Hurburgh Jr, “Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties,” *Soil Science Society of America Journal*, vol. 65, no. 2, pp. 480–490, 2001.
- [77] B. Kuang and A. M. Mouazen, “Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms,” *European Journal of Soil Science*, vol. 62, no. 4, pp. 629–636, 2011.
- [78] D. B. Lobell and G. P. Asner, “Moisture effects on soil reflectance,” *Soil Science Society of America Journal*, vol. 66, no. 3, pp. 722–727, 2002.
- [79] R. A. Viscarra Rossel, S. R. Cattle, A. Ortega, and Y. Fouad, “In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy,” *Geoderma*, vol. 150, no. 3-4, pp. 253–266, 2009.
- [80] G. M. Vasques, S. Grunwald, and J. O. Sickman, “Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra,” *Geoderma*, vol. 146, no. 1-2, pp. 14–25, 2008.
- [81] X. Peng, T. Shi, A. Song, Y. Chen, and W. Gao, “Estimating soil organic carbon using VIS/NIR spectroscopy with SVMR and SPA methods,” *Remote Sensing*, vol. 6, no. 4, pp. 2699–2717, 2014.
- [82] M. Vohland, J. Besold, J. Hill, and H.-C. Fründ, “Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy,” *Geoderma*, vol. 166, no. 1, pp. 198–205, 2011.
- [83] R. A. Viscarra Rossel and W. S. Hicks, “Soil organic carbon and its fractions estimated by visible-near infrared transfer functions,” *European Journal of Soil Science*, vol. 66, no. 3, pp. 438–450, 2015.
- [84] M. Nocita, A. Stevens, G. Toth, P. Panagos, B. van Wesemael, and L. Montanarella, “Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach,” *Soil Biology and Biochemistry*, vol. 68, pp. 337–347, 2014.
- [85] P. S. Nayak and B. K. Singh, “Instrumental characterization of clay by XRF, XRD and FTIR,” *Bulletin of Materials Science*, vol. 30, no. 3, pp. 235–238, 2007.
- [86] B. H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications*, John Wiley & Sons, Chichester, UK, 2004.