

## Research Article

# A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem

Anita Bayer,<sup>1</sup> Martin Bachmann,<sup>1</sup> Andreas Müller,<sup>1</sup> and Hermann Kaufmann<sup>2</sup>

<sup>1</sup> Department of Land Surface, German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Weßling, Germany

<sup>2</sup> Remote Sensing Section, Department of Geodesy and Remote Sensing, German Research Centre for Geosciences (GFZ), Telegrafenberg, 14473 Potsdam, Germany

Correspondence should be addressed to Anita Bayer, anita.bayer@dlr.de

Received 16 February 2012; Revised 20 April 2012; Accepted 21 May 2012

Academic Editor: Eyal Ben-Dor

Copyright © 2012 Anita Bayer 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.

The accurate assessment of selected soil constituents can provide valuable indicators to identify and monitor land changes coupled with degradation which are frequent phenomena in semiarid regions. Two approaches for the quantification of soil organic carbon, iron oxides, and clay content based on field and laboratory spectroscopy of natural surfaces are tested. (1) A physical approach which is based on spectral absorption feature analysis is applied. For every soil constituent, a set of diagnostic spectral features is selected and linked with chemical reference data by multiple linear regression (MLR) techniques. (2) Partial least squares regression (PLS) as an exclusively statistical multivariate method is applied for comparison. Regression models are developed based on extensive ground reference data of 163 sampled sites collected in the Thicket Biome, South Africa, where land changes are observed due to intensive overgrazing. The approaches are assessed upon their prediction performance and significance in regard to a future quantification of soil constituents over large areas using imaging spectroscopy.

## 1. Introduction

The soil as upper layer of the Earth's surface is the most important layer for energy and nutrition flows necessary for the development of vegetation and thus of key importance for landscape analysis. The characterization of an ecosystem's soil condition and its spatial and temporal changes are vital indicators for soil health and particularly in agricultural ecosystems directly linked to crop production. In semiarid regions, land cover changes coupled with degradation and soil erosion are frequent phenomena which may be the result of long-term management practices or may be linked to climate change. In particular the depletion of carbon inventories in soils is accentuated by soil degradation and erosion [1] and directly causes not only environmental but also economical problems. In the semiarid subtropical Thicket Biome as part of the Eastern Cape Province of South Africa, land changes are observed due to decades of overgrazing by goats. This has caused the unique ecosystem to change from

dense shrubland with a high rate of carbon sequestration in vegetation and peripheral soils to an open savannah-like system. Chemical and physical soil attributes can serve as tracers to assess and monitor such phenomena. However, a mapping of their spatial distribution and temporal development is limited using conventional soil analyses since this would require intensive sampling and analysis efforts.

Despite that, field and imaging spectroscopy provides a time- and cost-effective tool for a mapping of selected chemical and physical soil attributes over large areas. The soil constituents can be determined based on diagnostic absorption characteristics inherent in a soil's reflectance spectrum. In general, the direct application of diagnostic spectral features for quantification purposes is limited as many features are influenced by (1) the overlapping of spectral features and (2) additional effects of other, mostly physical, soil properties (e.g., surface roughness) which influence the spectral reflectance in a nonsignificant way.

This paper presents two approaches for a quantification of soil organic carbon, iron oxides, and clay content based on field and laboratory spectroscopy. The first is a physical approach which is based on spectral absorption feature analysis. For every soil constituent, a set of distinct spectral features are selected according to their presence in the spectra and to their documentation in respective literature. They are linked with chemical ground reference information by multiple linear regression techniques. This approach is set up to take advantage of several spectral features and characteristics and various properties describing these to gather the most information content provided by physical features. It is tested if the advantage of combining several spectral features, although each spectral feature might be influenced, will result in prediction models that reach the performance of statistical techniques. As a second method, commonly used partial least squares regression (PLS) as an exclusively statistical multivariate method is applied for comparison. The suitability of PLS techniques for the mapping of several soil constituents is well known (e.g., [2]). The two approaches are investigated for their prediction performance and significance for three spectral datasets measured in field and laboratory and for two spectral resolutions (full ASD resolution and spectral resolution of the HyMap sensor). The objective of this research is to allow for a robust quantification of carbon inventories in semiarid soils and additionally to provide indicators for ecosystem function in relation to land degradation.

In the next step, these approaches will be applied for a future large-scale quantification of soil constituents for a South African study site in the Thicket Biome. Hyperspectral imagery (approximately 320 km<sup>2</sup>) of the HyMap airborne sensor [3] was acquired of this area in 2009, coupled with an extensive collection of spectral and chemical ground reference data (163 reference plots). In the Thicket Biome, accurate spatial information can be used to quantify the consequences of unsustainable land management techniques and detect eroded or degraded areas in their primary stage.

## 2. State of the Art in Soil Spectroscopy

Many studies exist where constituents of the upper soil layer are described based on spectral data from a laboratory or field environment. Examples include soil organic carbon (e.g., [4–7]), soil iron oxides (e.g., [8–10]), and soil texture together with clay content (e.g., [7, 11, 12]).

With increasing amount in the last years, airborne spectrometer data were used for predictions of soil constituents. A direct transfer of approaches developed on laboratory or field data to an airborne scanner is limited by technical aspects depending on the sensor system such as a lower spatial and spectral resolution, increased signal-to-noise ratio, atmospheric influences, BRDF effects, and spectral mixing within pixels. A recent summary of key studies using imaging spectroscopy to study soil properties can be found in Ben-Dor et al. [13].

Viscarra Rossel et al. [2] outline the potential of different spectral domains in combination with different methods for the derivation of quantitative soil information by using soil

spectroscopy. Partial least squares regression (PLS) as multivariate calibration procedure is a well-established method for these purposes and was often used in previous studies (e.g., [14–16]). PLS applies statistically derived spectral characteristics to describe the relationship between soil chemistry and spectral information. However, the significance and transferability of PLS models built upon local observations to a regional scale is often limited because of their high degree of adaptation (see, e.g., [15]). To allow for a mapping of soil constituents for ecosystems of a regional extent, more robust algorithms are needed. In previous studies, approaches directly applying spectral absorption features or band indices are considered to be more suitable since they are physically based (e.g., [17]). These approaches appear as more robust and allow for a transfer to regions of similar environmental conditions. In the following paper, a special focus is set to studies applying spectral feature parameters for soil quantification and in particular based on airborne sensors.

The importance of soil organic carbon for soil stability and fertility together with its contribution to the global carbon cycle resulted in global initiatives to quantify ecosystem carbon stocks and their changes. Hence, there have been very intense efforts of using spectroscopy for these purposes resulting in various studies dealing with the quantification of soil organic carbon (e.g., [14, 15, 18, 19]). Beside multivariate techniques, the estimation of soil organic carbon was done in particular based on an overall decrease in reflectance in the VIS/NIR region (see [4]) and also using spectral indices (e.g., [4, 6]).

In a similar way, several studies have applied absorption feature parameters for a quantification of iron oxides content. Richter et al. [20] therefore used the band depth of the 900 nm feature and evaluated the spectral influence of variable soil textures on prediction accuracy for a laboratory setup. The enrichment of iron oxides in Israelian sand dunes (rubification) was mapped by Ben-Dor et al. [21] using the CASI airborne sensor, by application of the redness index as an iron-related band index.

For a large-scale mapping of clay minerals, Chabrilat et al. [22] applied the varying shape of the 2200 nm feature of smectite, illite, and kaolinite to detect and map their spatial distribution from airborne AVIRIS and HyMap data of Colorado, USA, while Gomez et al. [23] and also Lagacherie et al. [24] successfully applied the continuum removed band depth of the 2200 nm feature to quantify clay content.

## 3. Study Area and Data Collection

Subtropical Thicket is a unique vegetation type located in the semiarid valleys that cover around 17% of the Eastern Cape Province of South Africa [25]. Among these, more than 70% of all vegetation units are considered as “moderately” to “severely” degraded [26]. Thicket is an unusual vegetation type for a semiarid environment since an almost complete cover of dense vegetation is present in pristine conditions. Thicket vegetation comprises a high component of the succulent shrub *Portulacaria afra* [27], which accounts for large proportions of carbon in its biomass and the peripheral

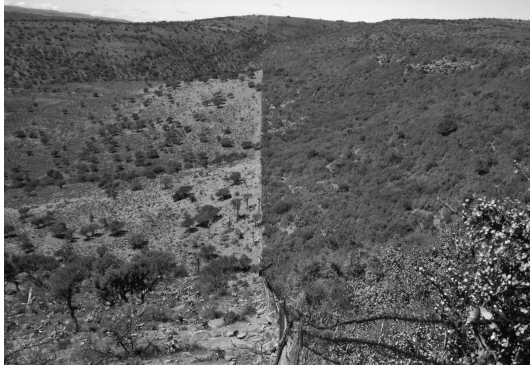


FIGURE 1: High-contrast land cover characteristics are found at fence lines between pasture and game farms in the Eastern Cape Province, South Africa. The pasture farm on the left side is highly degraded whereas the land on the right side is only slightly influenced and is nearly pristine Thicket shrub vegetation. The consequences of unsustainable land use can be detected and monitored using IS.

soils. Intensive goat farming since the early 1900s resulted in the loss of *P. afra* and the transformation of wide parts of Thicket vegetation to an open savannah-like system, which is accompanied by a severe loss of biodiversity and ecosystem carbon stocks [25]. Fence lines between farms of different utilization show pristine and transformed vegetation in direct neighborhood (Figure 1).

The study area is located near the city of Port Elizabeth (33.0°S/25.3°E; see Figure 2). It was selected within the Thicket Biome with an extent of 75 × 4 km, so that it includes the highest variance of various Thicket vegetation classes. It splits in two parts separated by a national park. Terrain elevation changes from 300 to 930 m in the northern part to 130 to 400 m in the southern section. The study area has a warm semiarid climate with monthly mean temperatures ranging from 13°C in winter months to 25°C in summer months (data by [28]). Annual precipitation averages 200 to 400 mm. The underlying mudstones and sandstones have resulted in the development of loamy and sandy soils at the surface (mostly cambisols and luvisols), (own investigations and [29]).

**3.1. Field Sampling.** During two sampling campaigns 163 natural, nonagricultural sites were sampled for ground truth, 96 in June 2009 and 67 in September/October 2009. The plots were randomly distributed within topographic units and the two sections of the study area as it is shown in Figure 2. Most of the samples were taken in areas where bare soil is the dominant land cover fraction. The amount and variability of sites sampled should result in a complete coverage of the variance of soils present. Each surveyed spot was chosen to represent the typical conditions of the surrounding area. For each plot, the general location-specific information (GPS coordinates, terrain information, land use, etc.), field spectra of the exposed soils and soil samples were collected. Additional information about vegetation coverage density and type as well as surface conditions was assessed of the center spot (1 × 1 m) and its surrounding (10 × 10 m).

Soil samples of the representative spot included the top layer to a maximum depth of around 1 cm and were limited to the crust if a physical soil crust was present at the surface. For 38 plots, field spectral measurements were not possible due to overcast conditions; thus soil field spectra are only available for 125 of the sampled sites.

Airborne HyMap imaging spectrometer data were acquired over the study area in October 2009 with a ground resolution of 3.3 m. These data are used for appropriate spectral resampling of point spectral datasets but not in the scope of this study and thus not described here in detail.

**3.2. Laboratory Chemical Analysis.** Soil samples were chemically analyzed for organic carbon and dithionite extractable iron using the Walkley-Black analysis [30], respectively, citrate-bicarbonate-dithionite extraction method (CBD, [31]). Particle size was determined in five fractions by pipette method [32]. The analysis showed the presence of sandy and loamy soils with generally low organic carbon contents. None of the histogram distributions shows a normal distribution (Figure 3). For all parameters, lower concentrations were over represented. For organic carbon, only two samples exceed 3.5% and show high contents of 5.7 and 5.9%. Both of these samples were taken at extraordinary sites, since one sample was taken in an area affected by droppings of grazing animals, while at the other sample site a partial grass layer covers the bare soil surface that might cause an increased carbon input to the soil. Iron oxides, which primarily are the results of the weathering processes of the underlying sand- and mudstones range up to 10.6%. Field surveys show that the distribution of clay and sand in overlying soils is directly related to the underlying parental material. Clay and sand concentrations increase where mudstones, respectively, sandstones, occur close to the surface. All samples were used for further analysis, even though the chemical contents of some samples differ from the distribution of the rest of the population (see Figure 3).

**3.3. Spectral Reference Data.** The measurement protocol resulted in three spectral datasets from field and laboratory environment, (1) insitu field spectra of the undisturbed soil surface, (2) bare soil field spectra, where an eventually present cover of small stones was removed, and (3) laboratory spectra.

For spectroradiometric measurements, a portable Field-Spec Pro 2 spectrometer (Analytical Spectral Devices, Inc.) was used. The measurements were taken with the bare fibre (FOV of 25°) and were directly converted to reflectance using a Spectralon panel as a reference. Measurement height was kept constant at 1.15 m, resulting in a footprint diameter of approximately 50 cm. For spots where the undisturbed bare soil fraction was very small, an alternative height of 0.5 m was applied. If small stones (mm to cm range) overlaid the surface, measurements were carried out first of the unchanged surface (insitu) and then only from the bare soil surface with the stone cover removed carefully. Stone coverage for around half of the sites was low (<3%). Despite this, particularly in regions where soils are shallow and developed

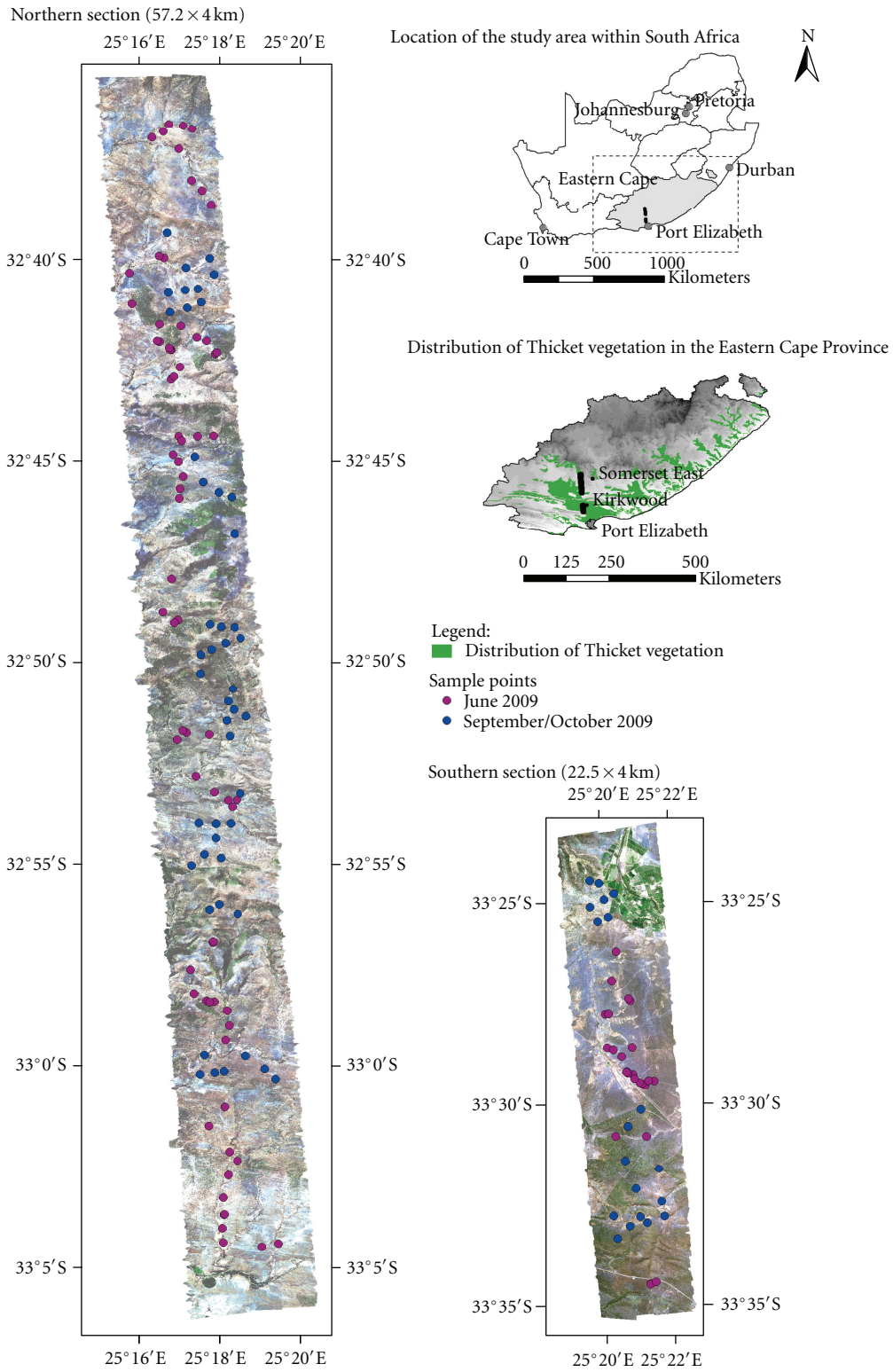


FIGURE 2: Detailed view of the HyMap imagery of the two transects west of Somerset East and in the Kirkwood vicinity in the Eastern Cape Province, South Africa, showing the distribution of the sampled sites. Each transect section comprises three HyMap flight lines.



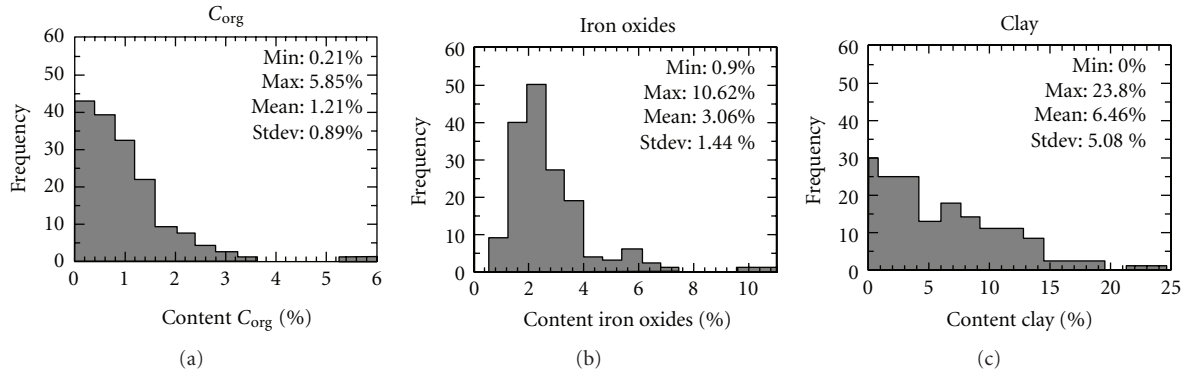


FIGURE 3: Histogram distribution with statistical parameters for 164 topsoil samples collected in the Thicket Biome, South Africa, analyzed for soil organic carbon, iron oxides, and clay content.

upon mudstones, it was significantly higher and ranged up to around 50%.

After the samples were dried, ground to dissolve aggregates, and sieved, the soil fraction smaller than 2 mm was spectrally measured in the laboratory. Constant illumination conditions were given by two Quartz halogen lamps (300 W each) at a zenith angle of 30° and a distance between spectrometer probe and sample surface of 15 cm. After five measurements were taken with the bare fibre, each sample was turned by 90° and measured again to reduce illumination effects of the soil surface. Averaged field and laboratory spectra were smoothed using a Savitzky-Golay filter [33]. Additionally, for the spectra measured in the field, the ranges of atmospheric water bands were interpolated from 1345 to 1445 nm and 1800 to 1960 nm.

Since the quantification methods are developed to be applied to hyperspectral imagery, each dataset was used in both full ASD resolution of 2087 spectral bands for the laboratory dataset and 1829 bands for field datasets, respectively (382–2468 nm, 1 nm resampling), and additionally resampled to the spectral characteristics of the calibrated HyMap imagery of 116 selected bands (456–2455 nm, spectral resolution 13–17 nm).

The different measurement setups are reflected in the statistics of the spectral libraries (Figure 4). The mean reflectance of the spectra measured in the laboratory is higher than of the two-field datasets because the pre-treatment applied beforehand to the samples (homogenizing and sieving) prevents shadow effects that may be caused by soil aggregates or extraneous material (Figure 4(a)). The controlled laboratory conditions result in very low standard deviations. Despite this, the spectral variance of the field measurements is stretched by influences that are not characteristic for specific soil constituents, such as surface condition and illumination effects (Figures 4(b) and 4(c)). In the insitu field measurements, average soil reflectance is reduced by small stones overlying the soil surface. Detailed statistics show that the presence of stones also introduced a further variability to the measurements, thus the standard deviation is higher compared to corresponding bare soil field measurements. Bare soil field spectra in particular present large extremes

in maximum and minimum spectra (Figure 4(c)). Where physical soil crusts are present, the smooth soil surface in combination with a reduced size of the particles at the soil surface increases soil reflectance. Thus the reflectance of individual samples measured in a field environment exceeded the reflectance of the corresponding laboratory spectra.

## 4. Methodology

Two methods are applied to establish a quantitative relationship between the chemical contents of the surveyed soil constituents and the spectral signal. Approach A is a physically based model, where spectral feature analysis is coupled with multiple linear regression techniques, while approach B applies partial least squares regression as an exclusively multivariate statistical method. 75% of the data (94 samples for field datasets, 123 samples for laboratory dataset) were selected by random stratified sampling based on the chemical reference of each soil constituent for training and used for model calibration, while the remaining 25% (31, resp., 40 samples) were used as test set to validate the models. The training and test sets thus are representative for the distribution of the chemical reference values as they occur in the South African study area. Even though the test samples are not used for model calibration, they cannot be considered as completely independent from the training set due to the spatial proximity of the measurements.

**4.1. Approach A: Multiple Linear Regression of Spectral Feature Parameters.** This approach applies a set of spectral features, found in previous literature to be characteristic for soil organic carbon, iron oxides, and clay content for subsequent multiple linear regression analysis. Since specific spectral properties which are related to physical absorption processes are used, this approach is seen as a physical model. It was developed to take advantage of the combination of several spectral features and characteristics. By parameterization of these spectral properties, one takes advantage of existing physically based information.

**4.1.1. Selection of Spectral Features.** The selection of diagnostic spectral characteristics is based on their presence in the

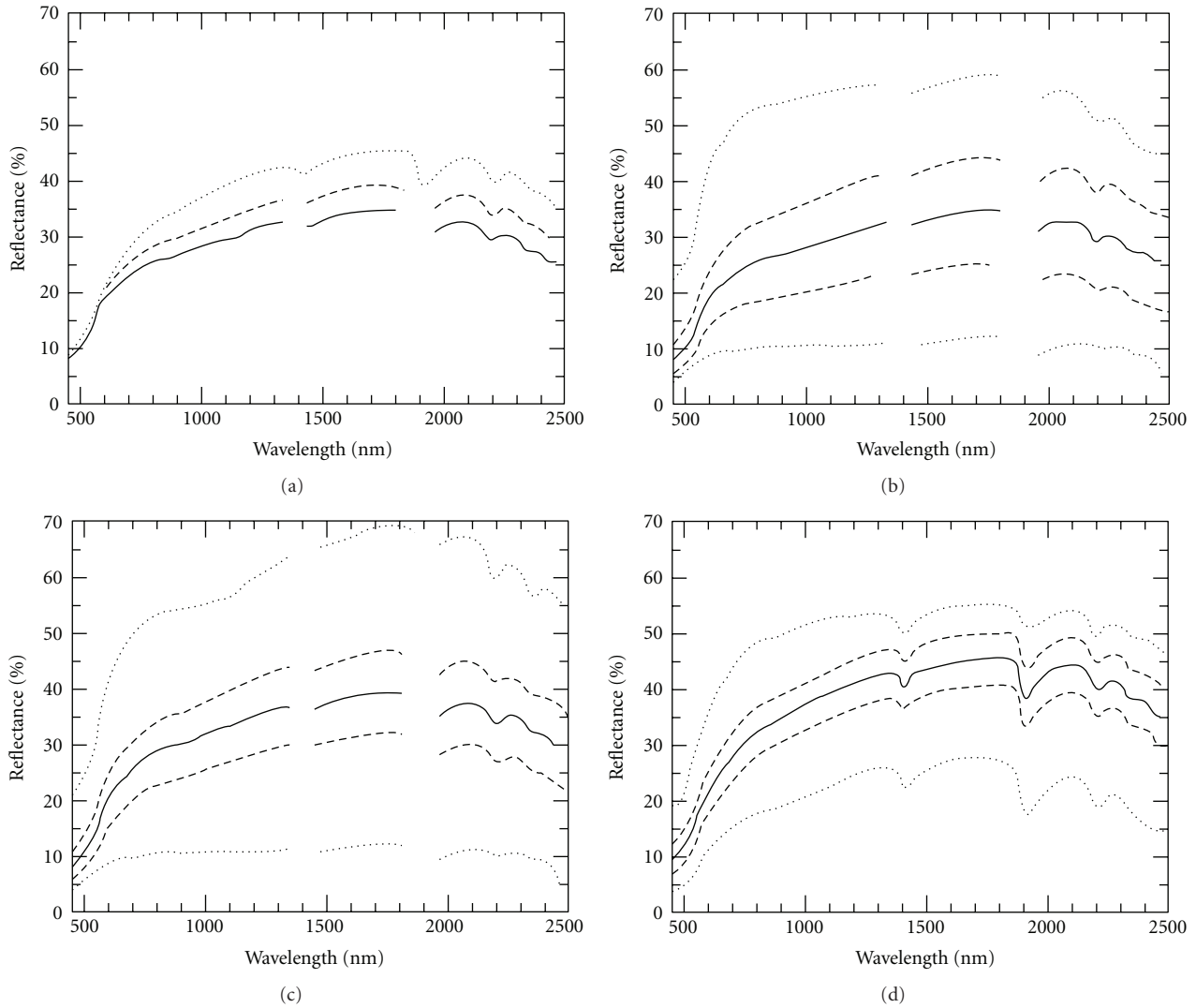


FIGURE 4: Comparison of mean spectra of the three spectral libraries (a) of insitu field (solid), field bare soil (dashed), and laboratory measurements (dotted). Statistical overview of 125 insitu field soil measurements (b), 125 field bare soil (c), and 163 laboratory spectra (d) with mean (solid), positive and negative standard deviation (dashed), minimum and maximum (dotted). All spectra are given in full ASD spectral resolution.

spectra and on well-known literature, for example, [7, 8, 34–40]. From this existing information only the most important and strong features are selected to increase the robustness of the algorithm. Features unique to one soil constituent are preferred, but not all spectral characteristics can fulfil this requirement. Table 1 lists the spectral features that are selected to be used in this study and references to a summary of previous studies where these features were described for enriched soils or nonsynthetic mineral powders.

The selected spectral characteristics are divided into three feature types: absorption features (AF), features of the spectral curve (CF), and features of its continuum as convex hull of a spectrum (HF). With these three types it is assumed that all specific spectral characteristics are covered that are described in previous studies as significant for soil organic carbon, iron oxides, and clay content.

Spectral detection and assessment of soil organic carbon is challenging since the numerous spectral features inherent to organic materials are often weak matching the complex chemistry of organic matter. Among the absorptions previous studies found to be characteristic for molecules and combinations of organic matter, the wavelength ranges around 1730 and 2330 nm are selected because they are identified as the most distinct in previous studies. Especially the absorptions around 1730 nm are unique for organic carbon since in this wavelength region no other absorptions of common minerals present in soils and rocks occur. In both wavelength regions, several second and third overtone absorptions of functional groups assigned to cellulose, lignin, starch, pectin, glucan, protein, wax, and humic acid as components of soil organic matter overlap and in combination lead to a noticeable absorption (e.g., [5, 34]). The

TABLE 1: Spectral features that are used for the delineation of soil organic carbon, iron oxides and clay content within approach A. References to studies previously describing these spectral characteristics are given.

	Feature and type	Wavelength	Assignment	Reference
$C_{org}$	(i) 1730 nm (AF) CR <sup>1</sup> : 1600–1815 nm $\lambda_{max}^2$ : not defined	1650/1669 nm	$2\nu^3$ of aromatic CH stretch	[5, 7]
		$\approx 1700$ nm	$2\nu$ of CH stretch	[34]
		1706/1754 nm	$2\nu$ of CH, alkyl doublet	[7]
		1706 nm	$4\nu$ of aliphatic CH stretch	[7]
		1726 nm	$2\nu$ of aliphatic CH stretch	[5]
		1761/1769 nm	$2\nu$ of aliphatic CH stretch	[5]
		1730–1852 nm	$4\nu$ of methyl CH	[7]
		(ii) 2330 nm (AF) CR: 2240–2410 nm $\lambda_{max}$ : 2330 nm	2275/2279 nm	$3\nu$ of CH <sub>2</sub> , CH <sub>3</sub>
	$\approx 2300$ nm		$\nu + \nu^4$ of CH stretch	[34]
	2307–2469 nm		$3\nu$ of methyl CH stretch	[7]
	2309 nm		$3\nu$ of aliphatic CH stretch	[5]
	2331 nm		$3\nu$ of CH <sub>2</sub> , COO	[5]
	2337/2386 nm		$3\nu$ of COO, CH <sub>3</sub>	[5]
	2347 nm		$3\nu$ of aliphatic CH stretch	[5]
	2381 nm		$3\nu$ of CO stretch of carbohydrates	[7]
	(iii) 450–740 nm (HF)	Decrease in reflectance in the visible range		(e.g., [4–6, 35])
(iv) 1460–1750 nm (HF)	Decrease in reflectance in the near to shortwave infrared range		(e.g., [4, 6, 35])	
Iron oxides	(i) 550 nm (AF) CR: 450–680 nm $\lambda_{max}$ : 550 nm	$\approx 490$ nm	ET <sup>5</sup> band of Fe <sup>3+</sup>	[7, 36]
		503 nm	Goethite	[37]
		$\approx 510$ nm	ET band of Fe <sup>2+</sup>	[36, 38]
		529 nm	ET band of hematite	[7]
		535 nm	Hematite	[37]
		$\approx 550$ nm	ET band of Fe <sup>2+</sup> , hematite	(e.g., [36, 38, 39])
	(ii) 700 nm (AF) CR: 580–800 nm $\lambda_{max}$ : 700 nm	650 nm	ET band of hematite and goethite	[7, 39]
		665 nm	ET band of goethite	[37]
		700 nm	$4\nu$ of O–H	[7]
		$\approx 700$ nm	ET band of Fe <sup>3+</sup>	(e.g., [35, 36, 38])
	(iii) 900 nm (AF) CR: 750–1300 nm $\lambda_{max}$ : 870 nm	$\approx 850$ –870 nm	ET band of hematite	[39, 40]
		860 nm	ET band of Fe <sup>2+</sup> and Fe <sup>3+</sup>	[34]
		868 nm	Hematite	[37]
		$\approx 870$ nm	ET band of Fe <sup>3+</sup>	[35, 36, 38]
		884 nm	ET band of hematite	[7]
		$\approx 900$ nm	Transition bands of Fe <sup>2+</sup> and Fe <sup>3+</sup>	(e.g., [34, 38])
		900–930 nm	ET band of goethite	[39]
		920 nm	ET band of goethite	[7]
		$\approx 930$ nm	$3\nu$ of O–H stretch, Goethite	[7, 37, 40]
		940 nm	$\nu + \nu$ of O–H stretch	[7]
		1000–1100 nm	ET band of Fe <sup>2+</sup>	(e.g., [35, 36])
		1025 nm	ET band of Fe <sup>3+</sup>	[8]
1075 nm	ET band of Fe <sup>3+</sup>	[8]		
(iv) 550–590 nm (CF)	Decrease in reflectance in the blue wavelength range towards the ultraviolet light		(e.g., [34, 38, 39])	
(v) 450–750 nm (HF)	Decrease in reflectance in the visible range		(e.g., [34, 38, 39])	

TABLE 1: Continued.

	Feature and type	Wavelength	Assignment	Reference
Clay	(i) 2200 nm (AF) CR: 2100–2290 nm $\lambda_{\max}$ : 2206 nm	2160/216 + 2208/2209 nm	$\delta^6$ of AlOH bend of kaolinite doublet and $\nu + \nu$ of OH stretch	[7, 37, 50]
		$\approx 2200$ nm	$\delta$ of AlOH bend and $\nu + \nu$ of OH stretch of montmorillonite and illite	(e.g., [35, 40, 50, 51])
		2200/2204 nm	Montmorillonite	[37]
		2206 nm	$\nu$ of OH stretch of montmorillonite and illite	[7]
		2208 nm	$\delta$ of AlOH bend of kaolinite doublet	[7]
		2216 nm	Illite	[37]
		2230 nm	$\delta$ of AlOH bend of smectites	[7]
	(ii) 2340 nm (AF) CR: 2270–2410 nm $\lambda_{\max}$ : 2340 nm	2308/2312 nm	Kaolinite	[37]
		2336 nm	Illite	[37]
		2340 nm	$\nu$ of OH stretch of illite	(e.g., [7, 22, 50])
		2372/2376 nm	Kaolinite	[37]
	(iii) 450–700 nm (HF)	Increase in reflectance in the visible range		(e.g., [34, 42])
	(iv) 1460–1750 nm (HF)	Increase in reflectance in the near to shortwave infrared range		(e.g., [34, 42])

<sup>1</sup>CR: range where continuum removal is performed; <sup>2</sup> $\lambda_{\max}$ : wavelength of maximal absorption predominantly found in literature; <sup>3</sup> $\nu$ : overtone absorption bands; <sup>4</sup> $\nu + \nu$ : combination bands of fundamental and overtone absorptions; <sup>5</sup>ET: electronic transition bands; <sup>6</sup> $\delta$ : fundamental absorption bands.

general decrease in reflection of organic-carbon-rich soils, for instance, in the visible due to the darkness of humic acid, is included as a feature of the spectral continuum. Several studies employ different wavelength regions in order to describe this property (e.g., [5, 6]) or tested the performance of various spectral indices to detect soil organic carbon (e.g., [4]). In this study, the two ranges in the visible region between 450 and 740 nm and in the near to shortwave infrared range between 1460 and 1750 nm are used to describe the characteristic decrease in reflectance resulting of a presence of soil organic carbon (HF).

Two selected features of organic matter are ambiguous but still can be applied since the local conditions of the study area rule out any influences. First, this applies to the absorption near 2330 nm, which is also inherent to carbonate minerals, but since there is no carbonate bedrock present in the study area and carbonate precipitation sediments occur only very locally and mostly in deeper soil layers. Second, in addition to the influence of soil organic carbon, the overall reflectance is also affected by soil moisture content and grain size distribution. However, an influence of soil moisture can be neglected, because all soils were dry at the day of and before the time of the survey. This is also shown by local weather data [41]. The effect of grain size on overall reflectance and also a feature close to the carbon feature at 2330 nm are also included as spectral characteristics for the determination of clay content, but it is not possible to resolve the proportion of each of the two constituents.

For the determination of soil iron oxides, we apply the strong absorptions occurring in the visible range around 550, 700 nm, and near 900 nm. All of them result from electronic transition bands of the bi- and trivalent iron ions

as compounds of the most common iron oxides goethite and hematite (e.g., [34, 35, 39]). Electronic transition bands usually appear as broad absorptions. The decrease in reflectance in the blue towards the ultraviolet region as result of a broad absorption band in the UV range is described by various authors (e.g., [38, 39]) and included as two features of the spectral curve: first, the change in reflectance in the blue wavelength range as slope of the spectral curve between 550 and 590 nm (CF) and second, the shape of the continuum is included over the entire VIS range (450 to 740 nm, HF) to pick up the shape of the curve with a reduced influence of the 550 nm absorption.

The spectroscopic determination of the clay content is done using the prominent AlOH absorption feature around 2200 nm that is inherent to all clay minerals, and a minor hydroxyl feature around 2340 nm. Since both features result from fundamental and overtone absorptions of functional groups inherent to clay minerals, they occur as pronounced and narrow absorptions. As illite and montmorillonite are reported as the dominant clay minerals in the study area, 2206 nm was chosen as center wavelength of the 2200 nm absorption. Further, the spectral effect of the grain size distribution was used as indirect indicator for the clay content in a soil. The effect of an increasing content of finer grain sizes leading to an increase in reflectance (e.g., [34, 42]) is analyzed as shape and mean reflectance of the spectral curve in the VIS and also in the NIR/SWIR range (HF).

**4.1.2. Feature Parameterization.** In feature parameterization, the spectral datasets are analyzed for the selected spectral characteristics of the three types introduced and are transferred to numerical parameters that describe the shape of



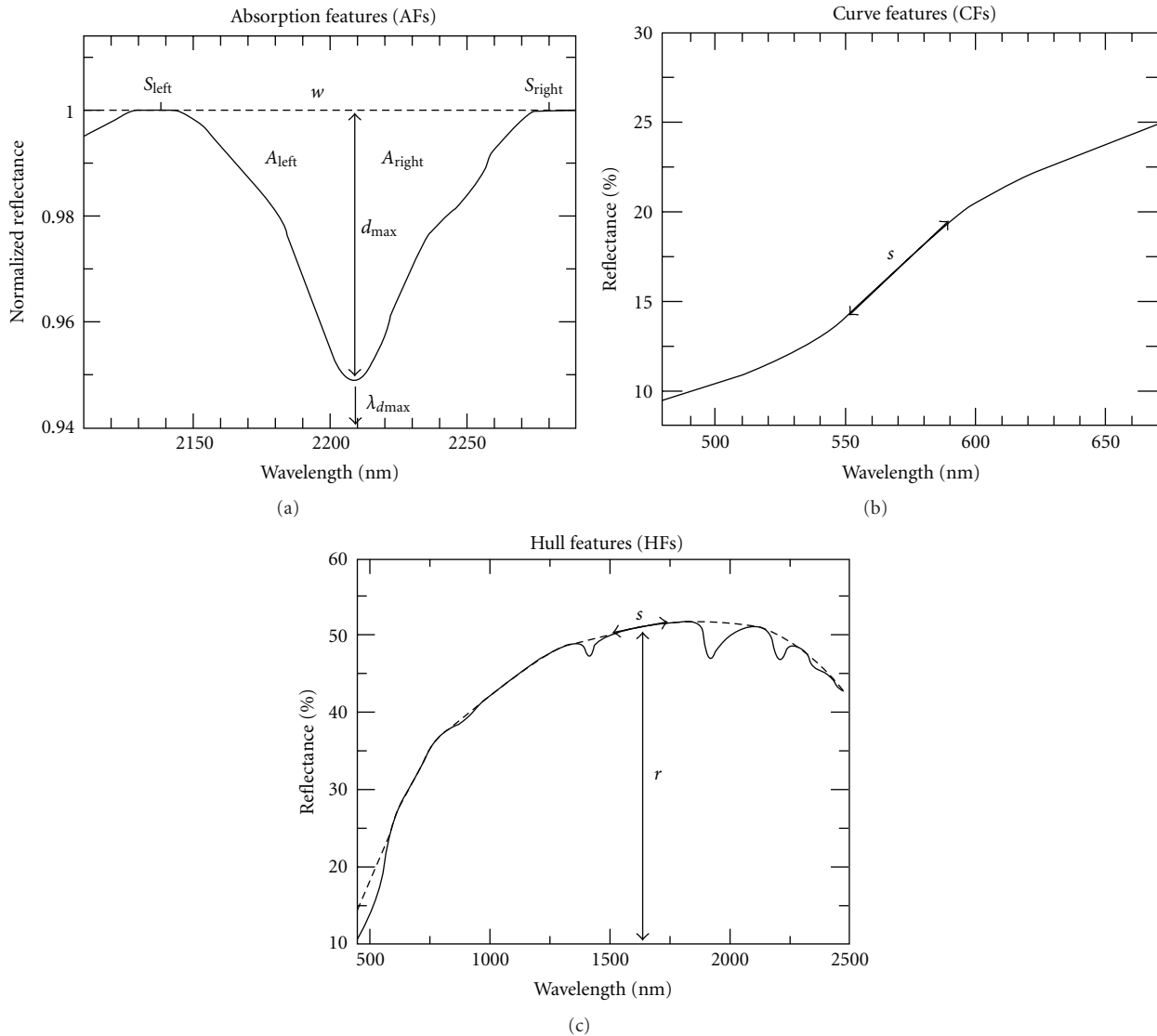


FIGURE 5: Parameterization of variables for the three types of described spectral features (AF, CF, and HF) used for the determination of soil organic carbon, iron oxides, and clay. Solid line: reflectance curve, dotted: continuum of the reflectance curve.

the spectral features (see Figure 5). These spectral feature variables are used for subsequent regression analysis. Table 1 lists parameters of the applied spectral features that are used for parameterization.

Spectral absorption features (AF) are parameterized from continuum removed reflectance spectra to isolate them from the trend of overall reflection and allow for intercomparison (see Figure 5(a)). Continuum removal (see [43]) is calculated individually for a defined interval around the feature being mapped (CR in Table 1). An AF is described by the following six variables, similar to spectral absorption feature analysis described by various authors (e.g., [37, 44, 45]): depth ( $d_{max}$ ) and wavelength ( $\lambda_{dmax}$ ) of maximal absorption, absorption depth at supposed characteristic wavelength according to respective literature ( $d_{lit}$ ), feature width ( $w$ ) as distance between the two determined feature shoulders ( $S_{left/right}$ ), the area between normalized continuum and spectral curve

( $A$ ), and asymmetry ( $AS = A_{left}/A_{right}$ ). Note that for the AF near 1730 nm, only five variables are calculated because no center wavelength ( $\lambda_{lit}$ ) is defined since this feature consists of several weaker absorptions overlapping in this wavelength region (see Tables 1 and 2). To describe the significant shape of the spectral curve in a defined wavelength range, two characteristics are introduced. (1) Curve features (CF) describe changes in reflectance occurring in a specific wavelength range (see Figure 5(b)), for example, the one triggered by the strong absorption of iron oxides in the ultraviolet. They are characterized only by the mean slope of the spectral curve ( $s$ ), calculated from a line fitted in the given wavelength range (see Table 1). (2) Features of the convex hull (HF) describe a soil constituent's effect on a broader spectral region that does not produce a distinct absorption (see Figure 5(c)). One region in the VIS (450 to 740 nm) and a second in the SWIR range (1460 to 1750 nm) were selected.

TABLE 2: Number of spectral features used and total number of calculated variables for determination of key soil constituents.

	Absorption features (6 var.)	Curve features (1 var.)	Hull features (2 var.)	Total number of variables
$C_{org}$	2 <sup>1</sup>	0	2	15
Iron	3	1	1	21
Clay	2	0	2	16

<sup>1</sup>For the AF near 1730 nm, only five variables are calculated (see text).

These wavelength regions of the two HF are used for the determination of multiple soil constituents. HF features are parameterized by mean slope ( $s$ ) and mean reflectance ( $r$ ) of the convex hull of the spectrum in a defined wavelength range so that they describe only the spectral shape without the influence of distinct local absorptions. Both variables of HF features are calculated by using a line fit.

The parameterization of  $n$  spectra and  $m$  spectral variables results in a  $m \times n$  sized matrix. Table 2 shows the number of spectral features analyzed for the determination of soil organic carbon, iron oxides, clay, and the resulting number of variables that describe the spectral shape and subsequently are used in regression analysis to model parameter contents.

**4.1.3. Model Calibration by MLR Analysis.** Multiple linear regression analysis is used to establish a functional relationship between spectral variables and chemical reference data. Figure 6 summarizes crucial steps of the work flow applied in the feature-based regression approach. MLR is conducted separately for each soil constituent using the ground reference data collected in the field campaigns. The group of spectral variables resulting from the parameterization of the selected spectral characteristics initially is checked for three parameters to ensure their suitability for MLR analysis: (1) the normalized standard deviation of each variable’s values (standard deviation in relation to mean) must be higher than 0.001 to exclude very small variables which would only cause an undesirable statistical adaptation during regression analysis. Such low standard deviations, for instance, would be reached for the depth of an absorption feature that is only weakly established or only occurs in some of the spectra. (2) Spectral variables result in redundant values which may be a result of a reduced spectral resolution (e.g.,  $d_{max}$  and  $d_{lit}$ ), and (3) the variables are analyzed for their variance to identify variables with a highly invariant part that are likely to be controlled by a superordinate factor (e.g., sensor band positions). These types of spectral variables are not suitable for regression analysis and are excluded from the analysis.

Calculated feature variables are subsequently standardized to allow for a comparison and ranking of regression coefficients. Together with chemical reference values related to the soil constituent under consideration, multiple regression analysis is performed resulting in an initial relationship without further intervention. Leave-one-out cross-validation is conducted to analyze the collection of spectral variables for significance and in order to check the sample population for outliers. Spectral variables are considered insignificant

for regression analysis if the absolute value of a regression coefficient’s mean is smaller than twice its standard deviation (see [46]) determined in leave-one-out cross-validation. Spectral variables found to be insignificant were excluded from further analysis since they provide no additional value for model development and might provide statistical adaptation. A sample  $i$  is identified as outlier if the absolute deviation of  $R_i^2$  from the mean of all  $R_i^2$  is higher than twice the standard deviation of all  $R_i^2$  determined in leave-one-out cross-validation. Samples identified as outliers may be removed from the population in this step by manual interaction if a proper reason is eminent (e.g., problems with sampling or analysis). The final multiple linear regression model is established based on the significant feature variables and including all samples. Soil parameter prediction models are calibrated for the training spectral datasets of both spectral resolutions. The models are evaluated in test set validation to assess the predictability of each model.

**4.2. Approach B: Partial Least Squares Regression.** For comparison, calibration models were built upon the same spectral datasets using partial least squares regression, as it is a well-established chemometric technique and often applied in soil spectroscopy. PLS is based on a projection of the predictor ( $x$ ) and response ( $y$ ) variables into a set of latent variables (or PLS factors) and corresponding scores, minimizing the dimensionality of the data while maximizing the covariance between  $x$  and  $y$  variables. A detailed description of the PLS technique is given by [47]. For PLS modeling the ParLes software was applied [48].

The spectral datasets were preprocessed with one or a combination of two of in total 11 manipulation methods (transformation of reflectance ( $R$ ) to  $\log(1/R)$ , 5 light scattering and baseline corrections such as multiplicative scatter correction (MSC), standard normal variate transform (SNV), wavelet detrending, and so forth, calculation of 1st and 2nd derivative, mean centering, variance scale of the data, and a combination of the latter two). The performance of each preprocessing setting was evaluated in leave-one-out cross-validation (CV). For each setting, the optimal number of latent variables to be used for modeling was examined by the variation of the root mean square error (RMS) and the Akaike information criterion (AIC) as a function of the number of latent variables. The optimal number of PLS factors was determined at local minima of RMS and AIC within a steady trend of these two factors. The one preprocessing setting in combination with its corresponding optimal number of PLS factors performing best in cross-validation was selected and subsequently used for model calibration based on the observations of the training set. Calibrated models were further applied to the test set for validation purposes (see Figure 6).

If several preprocessing settings provided similar CV accuracies, then model calibration and validation were performed for each setting. The PLS model with highest accuracies in both calibration and validation, as well as a minimal difference in between them was selected. This way the most significant, and robust PLS prediction models were retrieved and overfitting of the models was prevented.

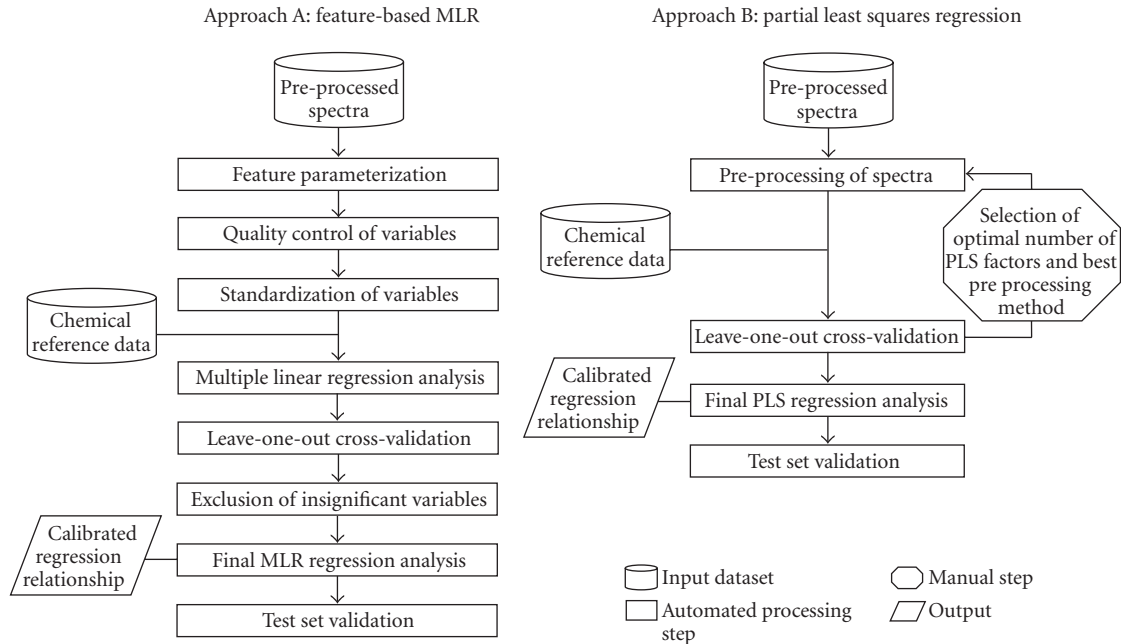


FIGURE 6: Work flow for feature-based MLR and PLS regression analysis.

## 5. Results and Discussion

For each soil constituent, a model was set up using the two modeling approaches, three spectral datasets, and two spectral resolutions, resulting in 12 models per constituent. All samples, though split into training and test sets, were applied for modeling. Model performance was assessed for each method and dataset and compared based on the model's correlation coefficient ( $R^2$ ) for predicted versus measured compositions, root mean square error (RMS), and ratio of performance to deviation (RPD). The RPD is defined as the ratio of the standard deviation of the reference samples divided by the RMS. This was done for both calibration and validation using the corresponding datasets of training data (94 samples for field and 123 for laboratory measurements) and test data (31/40 samples) (see Tables 3 and 5).

Goodness of predictions is evaluated following the qualitative classifications of Chang et al. [9]. They suggest  $R^2$  greater than 0.80 and RPD values greater than 2.0 as indicators for excellent prediction models.  $R^2$  between 0.50 and 0.80 and RPD values between 2.0 and 1.4, were considered as models of medium quality which are useful for quantitative predictions in most applications. Models with  $R^2$  lower than 0.50 and RPD lower than 1.4 are to be ranked not useable.

**5.1. Approach A: Multiple Linear Regression of Spectral Feature Parameters.** Details of the calibration models established using approach A are given in Table 3. The number of spectral variables found as insignificant and excluded from the final model development depends on its variation during cross-validation. Thus, the number of significant variables varies for all models established for a certain parameter. Table 4 gives the five spectral variables, which show the highest coefficients in the regression equation of

every feature-based regression model and thus are the most important ones for model development. Their influence is determined based on the regression coefficient and given in % of the summed absolute values of all regression coefficients. Negatively signed influences indicate negative regression coefficients. The absolute values of the regression coefficients, and thus the influences determined therefrom, mainly depend on the number and character of participating spectral variables. Thus, their absolute values cannot be expected to be identical for different models. Nonetheless, their ranges are significant and used to identify the most important spectral variables.

**Soil Organic Carbon.** All prediction models for soil organic carbon provide excellent calibration accuracies with  $R^2_{\text{Cal}}$  between 0.75 and 0.81,  $\text{RMS}_{\text{Cal}}$  around 0.45%, and all  $\text{RPD}_{\text{Cal}}$  higher than 2.0. The good correlation is also represented in the scatter plots of measured versus calculated soil organic carbon concentrations exemplarily shown for the model built upon bare soil field spectra in HyMap's spectral resolution (Figure 8). The accuracy of the test set validation is good especially for the laboratory datasets and in the same range as the calibration ( $\text{RPD}_{\text{Val}}$  close to 2.0). These models particularly showed good  $\text{RMS}_{\text{Val}}$  around 0.35%. However validation accuracy is slightly lower but reasonable for the field datasets ( $R^2_{\text{Val}}$  between 0.49 and 0.62,  $\text{RMS}_{\text{Val}}$  between 0.43 and 0.54%,  $\text{RPD}_{\text{Val}}$  between 1.26 and 1.57).

The models apply 12 to 15 of 16 calculated spectral variables. In the models developed based on field spectra, the influences concentrate on few variables, and both spectral variables describing the two hull features are of great importance (see Table 4). In all these models,  $r_{\text{HF SWIR1}}$ ,  $s_{\text{HF VIS}}$ , and  $r_{\text{HF VIS}}$  are the dominant factors in the regression

TABLE 3: Calibration and validation accuracies for approach A—Multiple linear regression of spectral parameters. The models further to be applied for a large-scale prediction of soil constituents based on HyMap imagery are highlighted.

	Spectral dataset	No. of spectral variables	Calibration (94 <sup>1</sup> /123 <sup>2</sup> samples)			Validation (31 <sup>1</sup> /40 <sup>2</sup> samples)			
			$R^2_{\text{Cal}}$	$\text{RMS}_{\text{Cal}}$	$\text{RPD}_{\text{Cal}}$	$R^2_{\text{Val}}$	$\text{RMS}_{\text{Val}}$	$\text{RPD}_{\text{Val}}$	
ASD spectral resolution	$C_{\text{org}}$	(1) In situ field	12 of 16	0.77	0.47	2.09	0.49	0.54	1.26
		(2) Bare soil field	14 of 16	0.79	0.44	2.22	0.56	0.48	1.42
		(3) Laboratory	13 of 16	0.75	0.47	2.01	0.74	0.36	1.93
	Iron oxides	(1) In situ field	19 of 21	0.63	0.80	1.64	0.24	0.90	1.07
		(2) Bare soil field	15 of 21	0.64	0.78	1.68	0.21	0.93	1.04
		(3) Laboratory	19 of 21	0.75	0.76	2.01	0.23	1.63	0.69
	Clay	(1) In situ field	14 of 16	0.31	4.12	1.20	0.01	5.13	0.88
		(2) Bare soil field	11 of 16	0.21	4.39	1.13	0.03	4.80	0.94
		(3) Laboratory	16 of 16	0.23	4.60	1.14	0.05	4.52	1.00
HyMap spectral resolution	$C_{\text{org}}$	(1) In situ field	13 of 16	0.79	0.44	2.19	0.51	0.54	1.27
		(2) Bare soil field	12 of 16	<b>0.81</b>	<b>0.43</b>	<b>2.28</b>	<b>0.62</b>	<b>0.43</b>	<b>1.57</b>
		(3) Laboratory	15 of 16	0.77	0.46	2.08	0.77	0.35	2.00
	Iron oxides	(1) In situ field	18 of 21	0.62	0.80	1.64	0.17	1.04	0.93
		(2) Bare soil field	19 of 21	<b>0.66</b>	<b>0.76</b>	<b>1.73</b>	<b>0.26</b>	<b>0.93</b>	<b>1.03</b>
		(3) Laboratory	20 of 21	0.73	0.79	1.94	0.25	1.58	0.71
	Clay	(1) In situ field	14 of 16	0.28	4.20	1.18	0.00	5.05	0.89
		(2) Bare soil field	14 of 16	<b>0.17</b>	<b>4.49</b>	<b>1.11</b>	<b>0.01</b>	<b>4.91</b>	<b>0.92</b>
		(3) Laboratory	12 of 16	0.25	4.53	1.16	0.05	4.62	0.98

<sup>1</sup>Number of samples in training and test set for field datasets. <sup>2</sup>Number of samples in training and test set for laboratory datasets.

relationships (see Figure 5 for abbreviations of spectral variables). Together they account for an influence of more than 60%. The two last ones show a negative influence as it was to expect, since increasing contents of soil organic carbon reduce reflectance in the visible, and thus have an effect on slope and mean reflectance. Although the same effect was expected for  $r_{\text{HF SWIR}1}$ , its coefficient is positive. In the models developed based on the laboratory dataset, the influences are distributed on more spectral variables and in particular the variables describing the 1730 nm absorption are predominant.  $A_{\text{AF}1730}$ ,  $r_{\text{HF SWIR}1}$ , and  $d_{\text{maxAF}1730}$  show up as important factors.

As calibration and validation accuracies show that all prediction models for soil organic carbon are of high quality according to the applied classification [9]. All models reach the criteria of excellent prediction models based on  $\text{RPD}_{\text{Cal}}$ , while their validation performance is well within medium-class models. The selection of spectral variables that account for the most influences on regression relationships is very constant and seems to be reasonable especially for field data, where the variables of the spectral continuum show up as the most influential ones. However, for the laboratory dataset, the influences are significantly different and more connected to the specific absorption around 1730 nm.

*Iron Oxides.* Calibrated iron oxides' prediction models show  $R^2_{\text{Cal}}$  between 0.62 and 0.75,  $\text{RMS}_{\text{Cal}}$  around 0.80%, and  $\text{RPD}_{\text{Cal}}$  between 1.64 and 2.01. Again with significantly better

performance of the laboratory spectra. Even if  $R^2_{\text{Cal}}$  values for all iron oxides prediction models are reasonable, lower  $\text{RPD}_{\text{Cal}}$  values already indicate reduced predictability. Validation results confirm this with very low  $R^2_{\text{Val}}$  (0.17 to 0.26), increased  $\text{RMS}_{\text{Val}}$ , and low  $\text{RPD}_{\text{Val}}$ . Especially the models built upon the laboratory soil spectra are providing significantly lower validation accuracies. Scatter plots of validation samples revealed the poor prediction of one sample which is characterized by an unusually high reflectance in the VIS and probably not representative for the spectral variation. In total feature-based iron oxides' prediction models thus do not reach the quality of the soil organic carbon prediction models with significantly poorer validation accuracies.

The predominant part of the iron oxides' prediction models are built with 18 to 20 of the 21 spectral variables describing spectral feature properties. One model applies only 15 variables. In the models developed based on field spectra,  $A_{\text{AF}900}$ ,  $A_{\text{AF}550}$ ,  $d_{\text{max AF}550}$ ,  $d_{\text{max AF}700}$ , and  $d_{\text{lit AF}900}$  are of prominent importance (see Table 4). Maximal absorption depths are usually positively signed, as expected. It is noticeable that the area of the 900 nm absorption feature ( $A_{\text{AF}900}$ ) consequently is signed negatively. For the models developed based on laboratory spectra,  $d_{\text{max AF}900}$ ,  $d_{\text{lit AF}900}$ ,  $A_{\text{AF}550}$ , and  $A_{\text{AF}900}$  appear as the most important spectral variables. Also within the models built on the laboratory spectra, negative coefficients of  $A_{\text{AF}900}$  can be observed also for the characteristic absorption at 900 nm. This may suggest



TABLE 4: Influence ( $I$ ) of the five most important spectral variables on the regression equation of each of the feature-based regression models of approach A.

Spectral dataset	Rank 1		Rank 2		Rank 3		Rank 4		Rank 5	
	$I$ [%]	Variable	$I$ [%]	Variable	$I$ [%]	Variable	$I$ [%]	Variable	$I$ [%]	Variable
$C_{org}$	36.6	$r_{HF}$ SWIR1	-24.0	$s_{HF}$ VIS	-10.9	$r_{HF}$ VIS	-6.0	$s_{HF}$ SWIR1	-5.9	$A_{AF2330}$
	30.5	$r_{HF}$ SWIR1	-21.1	$s_{HF}$ VIS	-8.4	$r_{HF}$ VIS	8.0	$d_{max}$ AFI730	-7.7	$d_{max}$ AF2330
	27.6	$A_{AF1730}$	16.9	$r_{HF}$ SWIR1	-14.1	$d_{max}$ AFI730	-11.8	$r_{HF}$ VIS	-9.0	$s_{HF}$ VIS
ASD spectral resolution	-23.7	$A_{AF550}$	21.8	$d_{max}$ AF550	-12.0	$A_{AF900}$	4.9	$d_{lit}$ AF900	4.9	$d_{max}$ AF700
	-17.6	$A_{AF900}$	10.6	$d_{lit}$ AF900	-10.4	$d_{lit}$ AF550	9.9	$d_{max}$ AF700	9.8	$d_{max}$ AF900
	29.2	$d_{max}$ AF900	-14.7	$d_{lit}$ AF900	-10.3	$A_{AF900}$	6.7	$s_{HF}$ VIS	-5.6	$d_{max}$ AF550
Clay	29.3	$A_{AF2200}$	-29.1	$d_{lit}$ AF2200	7.9	$r_{HF}$ SWIR1	5.6	$r_{HF}$ VIS	4.9	$\lambda_{dmax}$ AF2200
	-26.4	$d_{lit}$ AF2200	23.2	$A_{AF2200}$	7.8	$d_{max}$ AF2200	7.7	$r_{HF}$ SWIR1	7.5	$r_{HF}$ VIS
	-28.3	$d_{max}$ AF2200	18.8	$d_{lit}$ AF2200	-12.4	$d_{lit}$ AF2340	12.0	$A_{AF2200}$	7.5	$d_{max}$ AF2340
$C_{org}$	34.5	$r_{HF}$ SWIR1	-20.8	$s_{HF}$ VIS	-13.1	$r_{HF}$ VIS	8.1	$A_{AF1730}$	5.7	$AS_{AF2330}$
	30.4	$r_{HF}$ SWIR1	-18.6	$s_{HF}$ VIS	-12.0	$r_{HF}$ VIS	9.3	$d_{max}$ AFI730	7.4	$AS_{AF2330}$
	21.3	$A_{AF1730}$	-14.2	$d_{max}$ AFI730	10.1	$r_{HF}$ SWIR1	-7.6	$d_{max}$ AF2330	-7.4	$r_{HF}$ VIS
HyMap spectral resolution	-14.0	$A_{AF900}$	12.3	$d_{max}$ AF700	-10.4	$d_{lit}$ AF700	8.8	$d_{max}$ AF900	7.9	$A_{AF550}$
	22.0	$A_{AF550}$	-16.7	$d_{max}$ AF550	9.5	$d_{max}$ AF700	-7.8	$A_{AF900}$	-7.3	$A_{AF700}$
	18.9	$d_{max}$ AF900	-18.0	$A_{AF550}$	11.7	$d_{lit}$ AF550	-8.3	$d_{lit}$ AF900	-7.1	$A_{AF900}$
Clay	24.4	$A_{AF2200}$	-22.2	$d_{lit}$ AF2200	11.2	$r_{HF}$ SWIR1	-6.6	$s_{HF}$ VIS	-6.5	$s_{HF}$ SWIR1
	24.8	$A_{AF2200}$	-21.7	$d_{lit}$ AF2200	11.8	$\lambda_{dmax}$ AF2200	7.7	$r_{HF}$ SWIR1	-5.3	$AS_{AF2200}$
	26.3	$A_{AF2200}$	-21.2	$d_{max}$ AF2200	-14.3	$d_{lit}$ AF2340	11.6	$A_{AF2340}$	-6.2	$\lambda_{dmax}$ AF2340

Symbols: absorption features (AFs):  $A_{AF}$ : area,  $d_{max}$  AF: maximum depth,  $\lambda_{dmax}$  AF: wavelength of  $d_{max}$  AF;  $d_{lit}$  AF: depth at wavelength position given in literature,  $AS_{AF}$ : asymmetry factor; Hull features (HF):  $r_{HF}$ : mean reflectance in interval,  $s_{HF}$ : slope in interval.

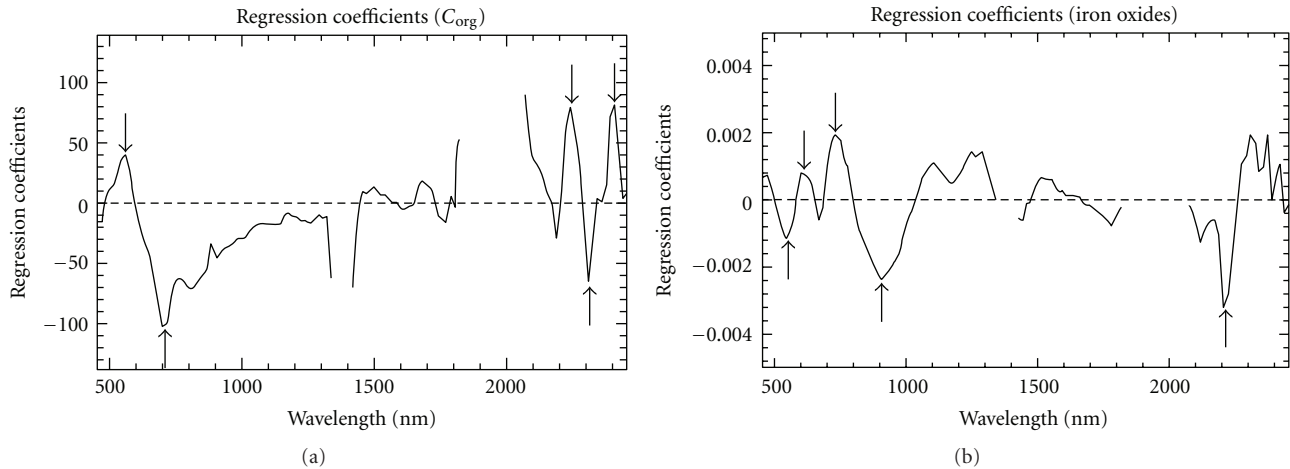


FIGURE 7: Regression coefficients of PLS models for the prediction of soil organic carbon (a) and iron oxides (b) built upon bare soil field spectra in HyMap's spectral resolution of 116 bands. Shape and range of spectral regression coefficients are related to the selected data pretreatment method (A:  $\log(1/R)$  + first derivative, B: Mean center).

that with increasing iron oxides' content more pronounced and steep features despite broad ones occur.

All together, the iron oxides' prediction model calibrations are of medium quality. Validation reveals drawbacks in prediction accuracy. The selection of important spectral variables includes attributes of the prominent iron absorptions, especially the one around 900 nm.

*Clay.* The calibration of feature-based clay prediction models leads to low correlations with  $R_{\text{Cal}}^2$  between 0.17 and 0.31 and  $\text{RMS}_{\text{Cal}}$  constantly above 4.0% for both spectral resolutions. Results of validation using the independent test set are poor and show no significant correlation ( $\text{RMS}_{\text{Val}}$  up to 5.13%,  $\text{RPD}_{\text{Cal}}$  below 1.00).

Soil clay content prediction models are built based on 11 to all 16 derived spectral variables.  $A_{\text{AF2200}}$ ,  $d_{\text{max AF2200}}$ ,  $d_{\text{Alit AF2200}}$ , and  $r_{\text{HF SWIR1}}$  frequently appear among the most important spectral variables used in the multiple linear regression equations of all six models (see Table 4). Though, considering all models, the influences observed for spectral variables are highly variant. There seems to be no connection between the selection of specific variables and the field or laboratory origin of the spectral base data. Some variables (e.g.,  $A_{\text{AF2200}}$ ,  $r_{\text{HF SWIR1}}$ ) appear consequently positively signed, whereas, for instance,  $d_{\text{Alit AF2200}}$  and  $d_{\text{max AF2200}}$  occur positively and negatively signed.

The reduced calibration accuracies, poor validation accuracies and no significant pattern within the spectral variables represented in the regression equation and their specific influence indicate that the feature-based regression approach fails to establish significant prediction models for soil clay content. Developed regression relationships indicate a high degree of statistical adaptation and finally a low significance of each model.

**5.2. Approach B: Partial Least Squares Regression.** Table 5 gives an overview of the prediction models for the selected

soil constituents using partial least squares regression techniques. The application of the different pretreatment methods allowed the derivation of consistent PLS prediction models, which apply between 4 and 9 PLS factors for regression analysis. In over 40% of the models, mean centering of the data revealed to be the best pretreatment technique, while transformation of reflectance ( $R$ ) to  $\log(1/R)$  and the first derivative were the appropriate methods for each 11% of the models. In the remaining cases, the combination of two pretreatment methods performed best. Regression coefficients of established PLS models provide information about the importance of certain wavelength for the regression model. Although they are highly variant depending on the settings of each model, some of their information can be related to physical soil properties.

Soil organic carbon prediction models for the three spectral datasets were very consistent for a number of pretreatment techniques, showing only slight differences. Calibration accuracies of the best PLS models for every dataset in terms of  $R_{\text{Cal}}^2$  ranged between 0.77 and 0.82, and  $\text{RPD}_{\text{Cal}}$  ranged between 2.11 and 2.38 for both applied spectral resolutions, both with low  $\text{RMS}_{\text{Cal}}$  slightly above 0.4%. Test set validation results of organic carbon models are also of good quality ( $R_{\text{Val}}^2$  between 0.53 and 0.79,  $\text{RMS}_{\text{Val}}$  between 0.32 and 0.52%,  $\text{RPD}_{\text{Val}}$  between 1.30 and 2.11), with highest accuracies for laboratory spectra and considerably lower for insitu field spectra. Nonetheless, organic carbon prediction models established by partial least squares techniques indicate good performance in both calibration and validation and do not indicate model overfitting. The regression coefficients of most soil organic carbon PLS models indicate a high importance of wavelengths in the VIS range, which may be related to soil colour (see Figure 7(a) for an example). Distinct features additionally occurring in the SWIR can be linked to absorptions of functional groups of soil organic carbon (see Table 1).

For the calibration of iron oxides prediction models, consequently 7 or more PLS factors were needed to set up

TABLE 5: Calibration and validation accuracies for approach B—partial least squares regression. The models further to be applied for a large-scale prediction of soil constituents based on HyMap imagery are highlighted.

	Spectral dataset	Pretreatment	No. of factors	Calibration (94 <sup>1</sup> /123 <sup>2</sup> samples)			Validation (31 <sup>1</sup> /40 <sup>2</sup> samples)			
				$R^2_{\text{Cal}}$	RMS <sub>Cal</sub>	RPD <sub>Cal</sub>	$R^2_{\text{Val}}$	RMS <sub>Val</sub>	RPD <sub>Val</sub>	
ASD spectral resolution	C <sub>org</sub>	(1) In situ field	1st derivative	6	0.79	0.45	2.16	0.53	0.52	1.30
		(2) Bare soil field	MSC + 1st derivative	6	0.82	0.41	2.38	0.65	0.42	1.61
		(3) Laboratory	Mean center + 1st derivative	6	0.82	0.40	2.36	0.69	0.45	1.53
	Iron oxides	(1) In situ field	Mean center	7	0.66	0.76	1.74	0.39	0.84	1.14
		(2) Bare soil field	Mean center	8	0.69	0.73	1.80	0.43	0.80	1.20
		(3) Laboratory	Log(1/R) + mean center	9	0.82	0.64	2.39	0.43	0.98	1.14
	Clay	(1) In situ field	Mean center	7	0.23	4.34	1.15	0.02	4.63	0.97
		(2) Bare soil field	Log(1/R)	5	0.11	4.67	1.06	0.06	4.42	1.02
		(3) Laboratory	Mean center	8	0.33	4.31	1.22	0.08	4.43	1.02
HyMap spectral resolution	C <sub>org</sub>	(1) In situ field	Log(1/R) + mean center	5	0.77	0.46	2.11	0.59	0.47	1.43
		(2) Bare soil field	<b>Log(1/R) + 1st derivative</b>	5	<b>0.82</b>	<b>0.41</b>	<b>2.38</b>	<b>0.62</b>	<b>0.45</b>	<b>1.51</b>
		(3) Laboratory	Log(1/R)	9	0.79	0.43	2.20	0.79	0.32	2.11
	Iron oxides	(1) In situ field	Mean center	7	0.67	0.75	1.75	0.43	0.76	1.26
		(2) Bare soil field	<b>Mean center</b>	7	<b>0.66</b>	<b>0.76</b>	<b>1.73</b>	<b>0.45</b>	<b>0.77</b>	<b>1.25</b>
		(3) Laboratory	Log(1/R) + mean center	9	0.81	0.67	2.28	0.49	0.88	1.27
	Clay	(1) In situ field	Mean center	4	0.14	4.58	1.09	0.01	5.06	0.89
		(2) Bare soil field	<b>Mean center</b>	5	<b>0.15</b>	<b>4.56</b>	<b>1.09</b>	<b>0.03</b>	<b>4.47</b>	<b>1.01</b>
		(3) Laboratory	1st derivative	8	0.37	4.16	1.27	0.10	4.43	1.02

<sup>1</sup>Number of samples in training and test sets for field datasets. <sup>2</sup>Number of samples in training and test sets for laboratory datasets.

significant models. The PLS models show correlations ( $R^2_{\text{Cal}}$ ) between 0.66 and 0.82 for all six combinations of spectral datasets and spectral resolutions. Resulting RMS<sub>Cal</sub> errors between 0.64 and 0.76 and RPD<sub>Cal</sub> between 1.73 and 2.39 indicate slightly lower prediction ability than for organic carbon. Test set validation of the established iron oxides' prediction models show reduced accuracies, which is most obvious in a decrease in  $R^2$  and RPD, while RMS for the most models only slightly increases from calibration to validation (RMS<sub>Val</sub> between 0.76 and 0.98%). No iron oxides model reaches an RPD<sub>Val</sub> of 1.4 indicating a medium predictive model [9]. PLS regression coefficients of iron oxides' models (example in Figure 7(b)) show a increased importance of wavelengths in the VIS range, which can be related to electronic transition bands of the iron ions primarily the one around 900 nm (Table 1). In the SWIR, the 2208 nm clay absorption shows up, representing an indirect correlation with features related to clay. In fact, a correlation of iron oxides and clay content could be identified based on chemical reference ( $R^2$  of 0.20).

The calibration accuracy of clay predictive models with  $R^2_{\text{Cal}}$  below 0.4 and RPD<sub>Cal</sub> below 1.3, respectively, is below the accuracies of medium prediction models. RMS errors are constantly above 4.0%. Among these models, the best correlations were reached for the laboratory datasets. Corresponding validation results are equally poor. Regression coefficients are highly variant. These results show that for estimations of the clay content in soils no significant prediction

model providing sufficient accuracy could be established when using partial least squares regression techniques and based on the ground reference data of the South African study site.

*5.3. Comparison of Modeling Approaches for Soil Organic Carbon and Iron Oxides.* Of all functional relationships established with the two approaches, spectra measured in a controlled laboratory led to the best models (lowest RMS<sub>Val</sub> and highest RPD<sub>Val</sub>), since they are not biased by influences present in a field environment (compare larger variance of field datasets in Figure 4). This was expected and also observed in previous studies (e.g., [24]). Generally, insitu field spectral surveys usually performed lowest. This is very likely to be a result of small stones naturally overlying the bare soil surface and disturbing the correlation between spectral signature and chemical concentrations. The resampling of the full ASD spectral resolution of the field and laboratory dataset to the reduced HyMap resolution has a low influence on model performances. Similar losses in model accuracy when reducing spectral resolution were reported by various authors (e.g., [8, 24]).

For organic carbon prediction models, the difference in performance between feature-based MLR (approach A) and PLS techniques (approach B) is low. However the predictability of the feature-based approach is slightly lower than that of PLS techniques. Average RPD<sub>Cal</sub> of the feature-based MLR models for soil organic carbon is 2.15 and 2.27 of

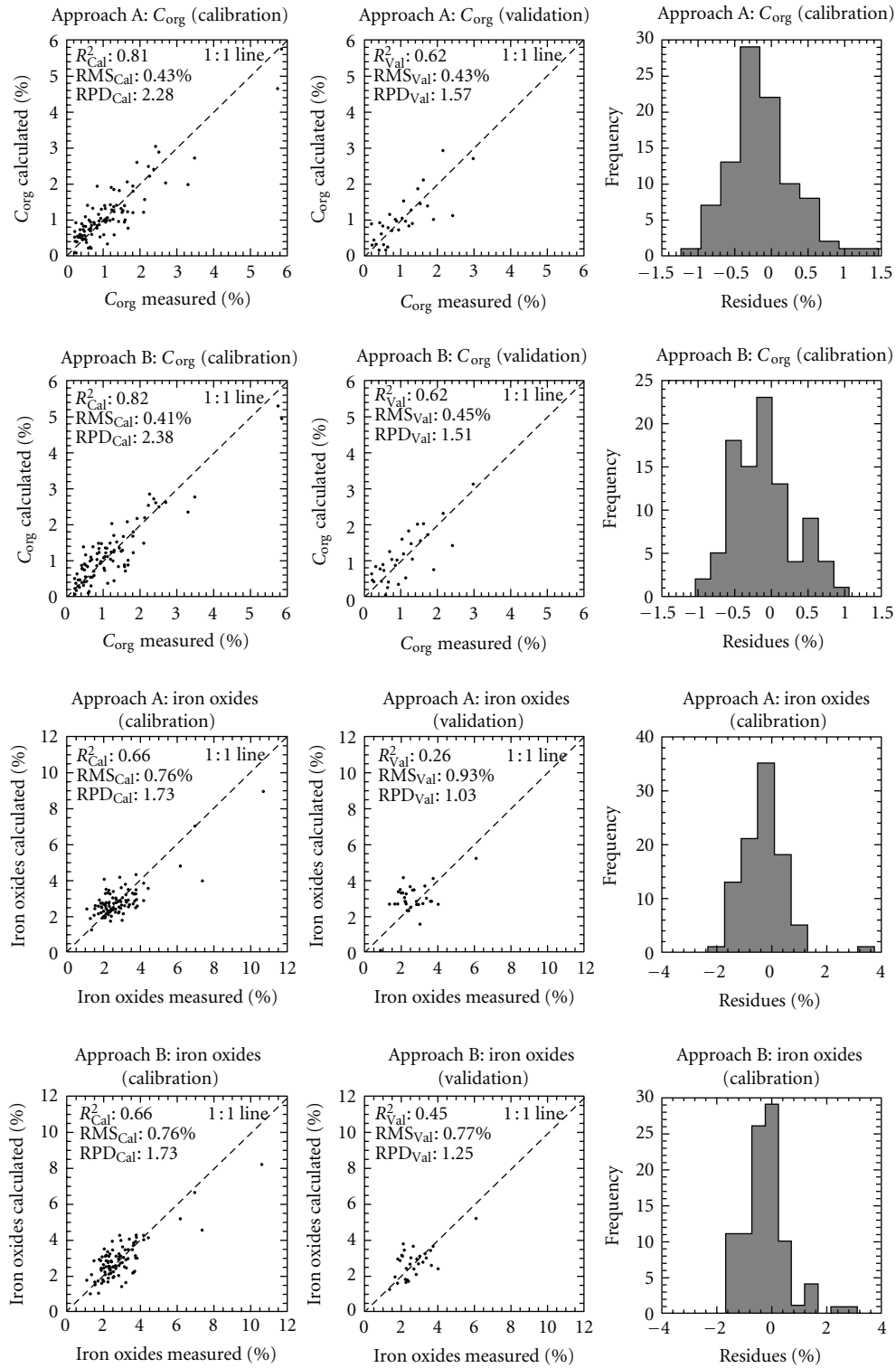


FIGURE 8: Correlation plots and histograms of calibration residues for the prediction of soil organic carbon and iron oxides using the two presented modeling approaches and developed based on bare soil field spectra in HyMap’s spectral resolution of 116 bands. Approach A: multiple linear regression of spectral parameters, approach B: partial least squares regression.



the PLS models, while validation accuracies are comparable for both approaches (average  $RMS_{Val}$  of 0.44%,  $RPD_{Val}$  of 1.58 for both approaches). For approach A the most important spectral variables in the regression models were selected very consistently and are significant.

Prediction models for iron oxides provide average calibration correlation coefficients around 0.67 for approach A and 0.72 for approach B with  $RPD_{Cal}$  of 1.77 and 1.95, respectively. Both approaches show  $RMS_{Cal}$  around 0.75%. Lower  $RPD_{Cal}$  and higher root mean square errors of the iron calibrations of both approaches generally indicate a reduced predictive power which is confirmed in validation results using the independent test set ( $R^2_{Val}$  for all models is below 0.50,  $RPD_{Val}$  of 0.91 as average of approach A models and 1.21 of approach B models). Thus, also within the iron oxides' prediction models, the ones established using PLS regression techniques perform slightly better. With these results the iron oxides' prediction models are of medium quality for both approaches, even though they cannot reach prediction accuracies of organic carbon models.

As the detailed results show, the model accuracies for soil organic carbon and iron oxides predictions, determined by a combination of spectral feature analysis and linear multiple regression techniques, are in the same range that statistical PLS approaches provide. Though, the performance of approach A is slightly lower for both soil organic carbon and iron oxides compared to the PLS approach. The difference is thus small for soil organic carbon models but more apparent for the prediction of iron oxides.

Figure 8 shows scatter plots of measured versus calculated distributions for the prediction of soil constituents using the two modeling approaches and based on bare field spectra in the resolution of the HyMap imagery. These are the models further to be applied to the HyMap imagery for a large-scale prediction of soil parameters. Their statistics are highlighted in Tables 3 and 5. The histograms of the residues are normally distributed. This clearly indicates that the established models are able to model the chemical variance of the selected parameters, even though the chemical contents of the reference samples are not normally distributed (compare Figure 3).

In the chemical reference data, 2 samples for soil organic carbon and 1 sample for iron oxides attract attention because they exceed the overall distribution of the other samples (compare Figure 3). These samples appear in the calibration scatter plots of Figure 8 for both approaches close to the 1:1 line, indicating that they might have a great influence on model calibration. But when testing regression models built on the presented spectral libraries excluding these samples, similar models and results were achieved as they are presented here.

**5.4. Prediction of Clay Content in Soils.** All regression models established and evaluated in this study do not adequately predict soil clay content. Average calibration correlations are low for both approaches ( $R^2_{Cal}$  of 0.23 as average of all models), with high  $RMS_{Cal}$  (average of 4.41%) and low  $RPD_{Cal}$  (average of 1.15). Validation fails for both approaches.

An explanation for the observed poor correlation of spectral and chemical information regarding clay content might result from the variable geology of alternating layers of sandstones and mudstones. Although in the study area developed under the same climatic conditions, the formation of soils highly depends on the source material and thus soils which developed based on sandstones very likely differ in their chemical and spectral properties from ones which developed based on mudstones. In a field environment the presence of soil physical crusts may be one additional factor. Soil physical crusts in the study area can locally be very well developed and reach over 1 cm in thickness. They form during rain drop impacts on the uncovered soil surface and cause a disintegration of soil aggregates and particle movement which results in a thin clay-rich surface layer of low permeability [49]. Soil crusts are problematic for spectral analyses, since the chemistry of the surface layer which is measured by a field spectrometer does not correspond to the chemistry of the bulk sample of the upper 1 cm analyzed in the laboratory.

## 6. Summary and Conclusions

This study presented two approaches for the quantification of relevant soil parameters from spectral data. Regression models were established for soil organic carbon, iron oxides, and clay content based on 163 samples measured in a laboratory environment and thereof 125 samples additionally measured in two field setups. The spectral datasets were investigated in original ASD spectral resolution and reduced resolution matching calibrated hyperspectral imagery of the study area obtained from the airborne HyMap sensor. The first approach is a physical model based on spectral feature analysis. A combination of selected spectral features is described by numerous variables that are used for multiple linear regression analysis. As second approach, conventional partial least squares regression analysis was chosen for comparison. The best PLS model for each spectral library was selected as combination of an appropriate preprocessing method and the optimal number of latent variables determined in leave-one-out cross-validation.

Results show that the two presented approaches provide similar capabilities to set up significant prediction models particularly for soil organic carbon and iron oxides. Good and medium-class prediction models could be established ( $RPD_{Cal}$  of 2.19 for  $C_{org}$  and 1.83 for iron oxides as average of all models of both approaches). Among these, organic carbon models for both approaches showed the best calibration accuracies, with  $RPD_{Cal}$  of 2.15 for feature-based regression and 2.27 for PLS regression with corresponding  $RMS_{Cal}$  of around 0.44% (as average of all  $C_{org}$  models of each approach). Iron oxides' models presented good and medium calibration accuracies but in particular showed a reduced validation performance. Both approaches failed to establish significant clay content prediction models ( $RPD_{Cal}$  of 1.15 and  $RMS_{Cal}$  of 4.41 as average of all models), which could be a result of the variable geology of the study area.

For the predominant part of the models built for soil organic carbon and iron oxides, the prediction performance

of the feature-based regression approach was in the same range as the statistically based PLS regressions provided. Also for the models of both approaches, the same trends were observed such as reduced validation accuracies compared to calibration for iron oxides models and generally low predictabilities of clay models. However, the feature-based approach in general performed slightly lower than the PLS approach. The difference was small for soil organic carbon predictions (comparable validation performance of MLR and PLS models), though slightly more apparent for iron oxides models. In the feature-based models, the spectral variables dominating the regression relationships were very consistent and support the significance of the established models, though different variables were predominant depending on laboratory or field data as base. The importance of specific wavelengths in PLS models was highly variant as result of the statistical modeling process. A correlation to physical features was only present in some models.

Compared to other studies working in agricultural environments [15, 19], the accuracy of prediction models for both approaches developed in this study is slightly lower. This is likely a result of the large size and the highly variant characteristics of the South African study area caused by nonagricultural environment, differences in geology and soil types. The difficulty to achieve prediction models for large areas with changing conditions (referred to as global calibrations) was addressed in previous studies (e.g., [19]). Based on AHS-160 data Stevens et al. [19] predicted soil organic carbon using among others PLS techniques for an agronomic Luxembourgian region. An  $RPD_{val}$  of 1.47 was obtained with a global calibration, which is in the same range as the prediction models for soil organic carbon established with the two approaches presented in this study. Stevens et al. could well improve this  $RPD_{val}$  to 2.76 using local calibrations based on agrogeological regions and soil types. In our case the lack of detailed spatially continuous information matching the variation of geology and soil types in the South African study area prevented the investigation of the performance of local calibrations.

A further factor-lowering prediction RPD is the small variability in measured contents of organic carbon in our study area. The ground reference sites were mainly selected to have a significant bare soil component so that these sites can also be used as validation targets for airborne hyperspectral imagery. Thus, no soil samples were taken in densely vegetated regions characterized by increased input of  $C_{org}$  and thus higher concentrations that would have increased the variability in the ground reference which is directly related to the modeling RPD. This corresponds with findings reported in, for example, [16, 19].

Results of this research show that it is possible to establish regression relationships on a physical basis that reach the predictability of PLS-derived regression models. This can be achieved by the application of a set of spectral characteristics for each of the investigated soil constituents and the inclusion of various properties of these spectral features. Statistical adaptation within regression analysis is reduced to a minimum. Because several spectral properties of each soil

constituent are further combined in the proposed approach to establish a significant model, this approach is suggested to be more robust compared to similar physical approaches that are based on the analysis of only one diagnostic spectral feature or band ratio. Additionally, the established models are rather simple and computationally unproblematic due to a limited number of variables in the regression relationship.

Physically based approaches are often considered to be more robust compared to exclusively statistical methods such as PLS, and the transferability of such approaches, is expected to be higher (e.g., [17]). For the newly developed approach this will be tested based on the hyperspectral imagery of the South African study area. For this, the developed feature-based regression relationships will be applied for a large-scale quantification of key soil parameters from hyperspectral imagery of the South African study area. The image data acquired in 2009 are to be processed with a combination of methods that allow the extraction of the soil spectral signal from mixing signatures caused by small-scale changes in land cover. First results for the large-scale prediction of topsoil organic carbon and iron oxides using calibrations developed on bare soil field spectra and using the feature-based regression approach are very promising. However, since the prediction models for soil clay content do not reach significant accuracies, they are not suited to be applied for a derivation of large-scale soil information. The examination of the HyMap imagery and the transfer of the methodology presented in this study to the image data are beyond the scope of this study.

Resulting soil parameter maps are valuable information to quantify the soil degradation status within the South African study area. In the Thicket Biome those information can be used to detect eroded and degraded areas in their primary stage in regard to direct the restoration of selective regions. The method was developed for semiarid areas in general and not adapted to specific conditions in the study area. A transfer to other regions of similar environmental conditions will be further investigated.

## Acknowledgments

This research study was funded as Ph.D. project being part of the Helmholtz EOS Network, a collaboration of German Helmholtz research centers. The PRESENCE Network (a Living Lands initiative) supported the intensive field sampling. The colleagues from DLR and GFZ are acknowledged for their valuable contribution to many parts of this work.

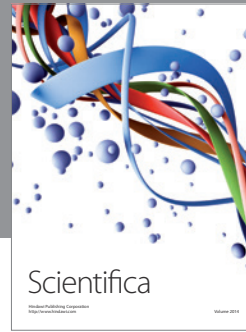
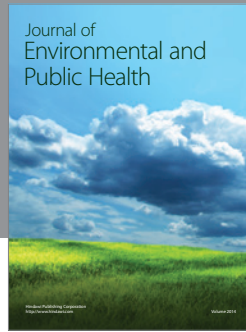
## References

- [1] R. Lal, "Soil carbon sequestration to mitigate climate change," *Geoderma*, vol. 123, no. 1-2, pp. 1-22, 2004.
- [2] 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.
- [3] T. Cocks, R. Janssen, A. Stewart, I. Wilson, and T. Shields, "The HyMap airborne hyperspectral sensor: the system, calibration

- and performance,” in *Proceedings of the 1st EARSEL Workshop on Imaging Spectroscopy*, Zurich, Switzerland, 1998.
- [4] H. M. Bartholomeus, M. E. Schaepman, L. Kooistra, A. Stevens, W. B. Hoogmoed, and O. S. P. Spaargaren, “Spectral reflectance based indices for soil organic carbon quantification,” *Geoderma*, vol. 145, no. 1-2, pp. 28–36, 2008.
  - [5] E. Ben-Dor, Y. Inbar, and Y. Chen, “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 Sensing of Environment*, vol. 61, no. 1, pp. 1–15, 1997.
  - [6] J. Hill and B. Schütt, “Mapping complex patterns of erosion and stability in dry mediterranean ecosystems,” *Remote Sensing of Environment*, vol. 74, no. 3, pp. 557–569, 2000.
  - [7] 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.
  - [8] E. Ben-Dor and A. Banin, “Visible and near-infrared (0.4–1.1 $\mu$ m) analysis of arid and semiarid soils,” *Remote Sensing of Environment*, vol. 48, no. 3, pp. 261–274, 1994.
  - [9] C.-W. Chang, D. A. Laird, M. J. Mausbach, and C. R. Hurburgh, “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.
  - [10] A. Palacios-Orueta and S. L. Ustin, “Remote sensing of soil properties in the Santa Monica Mountains I. Spectral analysis,” *Remote Sensing of Environment*, vol. 65, no. 2, pp. 170–183, 1998.
  - [11] E. Ben-Dor and A. Banin, “Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties,” *Soil Science Society of America Journal*, vol. 59, no. 2, pp. 364–372, 1995.
  - [12] T. H. Waiser, C. L. S. Morgan, D. J. Brown, and C. T. Hallmark, “In situ characterization of soil clay content with visible near-infrared diffuse reflectance spectroscopy,” *Soil Science Society of America Journal*, vol. 71, no. 2, pp. 389–396, 2007.
  - [13] E. Ben-Dor, S. Chabrillat, J. A. M. Dematté et al., “Using Imaging Spectroscopy to study soil properties,” *Remote Sensing of Environment*, vol. 113, no. 1, pp. S38–S55, 2009.
  - [14] C. Gomez, R. A. Viscarra Rossel, and A. B. McBratney, “Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: an Australian case study,” *Geoderma*, vol. 146, no. 3-4, pp. 403–411, 2008.
  - [15] A. Stevens, B. van Wesemael, H. Bartholomeus, D. Rosillon, B. Tychon, and E. Ben-Dor, “Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils,” *Geoderma*, vol. 144, no. 1-2, pp. 395–404, 2008.
  - [16] T. Udelhoven, C. Emmerling, and T. Jarmer, “Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study,” *Plant and Soil*, vol. 251, no. 2, pp. 319–329, 2003.
  - [17] V. L. Mulder, S. de Bruin, M. E. Schaepman, and T. R. Mayr, “The use of remote sensing in soil and terrain mapping-a review,” *Geoderma*, vol. 162, no. 1-2, pp. 1–19, 2011.
  - [18] T. Selige, J. Böhner, and U. Schmidhalter, “High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures,” *Geoderma*, vol. 136, no. 1-2, pp. 235–244, 2006.
  - [19] A. Stevens, T. Udelhoven, A. Denis et al., “Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy,” *Geoderma*, vol. 158, no. 1-2, pp. 32–45, 2010.
  - [20] N. Richter, T. Jarmer, S. Chabrillat, C. Oyonarte, P. Hostert, and H. Kaufmann, “Free iron oxide determination in mediterranean soils using diffuse reflectance spectroscopy,” *Soil Science Society of America Journal*, vol. 73, no. 1, pp. 72–81, 2009.
  - [21] E. Ben-Dor, N. Levin, A. Singer, A. Karnieli, O. Braun, and G. J. Kidron, “Quantitative mapping of the soil rubification process on sand dunes using an airborne hyperspectral sensor,” *Geoderma*, vol. 131, no. 1-2, pp. 1–21, 2006.
  - [22] S. Chabrillat, A. F. H. Goetz, L. Krosley, and H. W. Olsen, “Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution,” *Remote Sensing of Environment*, vol. 82, no. 2-3, pp. 431–445, 2002.
  - [23] C. Gomez, P. Lagacherie, and G. Coulouma, “Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements,” *Geoderma*, vol. 148, no. 2, pp. 141–148, 2008.
  - [24] P. Lagacherie, F. Baret, J. B. Feret, J. Madeira Netto, and J. M. Robbez-Masson, “Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements,” *Remote Sensing of Environment*, vol. 112, no. 3, pp. 825–835, 2008.
  - [25] R. G. Lechmere-Oertel, G. I. H. Kerley, and R. M. Cowling, “Patterns and implications of transformation in semi-arid succulent thicket, South Africa,” *Journal of Arid Environments*, vol. 62, no. 3, pp. 459–474, 2005.
  - [26] J. W. Lloyd, E. C. van den Berg, and A. R. Palmer, “Patterns of transformation and degradation in the Thicket Biome, South Africa,” Report 39, Terrestrial Ecology Research Unit, University of Port Elizabeth, Port Elizabeth, South Africa, 2002.
  - [27] A. J. Mills and M. V. Fey, “Transformation of thicket to savanna reduces soil quality in the Eastern Cape, South Africa,” *Plant and Soil*, vol. 265, no. 1-2, pp. 153–163, 2004.
  - [28] ARC-ISCW, *Weather Data of Three Selected Climate Stations in the Eastern Cape Province*, ARC-ISCW, Pretoria, South Africa, 2011.
  - [29] A. J. Mills and R. M. Cowling, “Rate of carbon sequestration at two thicket restoration sites in the Eastern Cape, South Africa,” *Restoration Ecology*, vol. 14, no. 1, pp. 38–49, 2006.
  - [30] A. Walkley and I. A. Black, “An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method,” *Soil Science*, vol. 37, pp. 29–37, 1934.
  - [31] O. Mehra and M. Jackson, “Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate,” *Clays and Clay Minerals*, vol. 7, pp. 317–327, 1958.
  - [32] G. W. Gee and J. W. Bauder, “Particle size analysis,” in *Methods of Soil Analysis. Part 1*, A. Klute, Ed., pp. 383–411, American Society of Agronomy, Soil Science Society, Madison, Wis, USA, 2nd edition, 1986.
  - [33] 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.
  - [34] R. Clark, “Spectroscopy of rocks and minerals, and principles of spectroscopy,” in *Remote Sensing for the Earth Sciences*, vol. 3, pp. 3–58, John Wiley & Sons, New York, NY, USA, 1999.
  - [35] M. F. Baumgardner, L. F. Silva, L. L. Biehl, and E. R. Stoner, “Reflectance properties of soils,” *Advances in agronomy*. Vol. 38, pp. 1–44, 1985.
  - [36] G. Hunt and J. Salisbury, “Visible and near-infrared spectra of minerals and rocks: 1. Silicate minerals,” *Modern Geology*, vol. 1, pp. 283–300, 1970.

- [37] C. Grove, S. Hook, and E. Paylor II, *Laboratory Reflectance Spectra of 160 Minerals, 0.4 to 2.5 Micrometers*, Jet Propulsion Laboratory, National Aeronautics and Space Administration, JPL publication 92-2, Pasadena, Calif, USA, 1992.
- [38] G. Hunt, J. Salisbury, and C. Lenhoff, "Visible and near-infrared spectra of minerals and rocks: III. Oxides and hydroxides," *Modern Geology*, vol. 2, pp. 195–205, 1971.
- [39] R. V. Morris, H. V. Lauer, C. A. Lawson, E. K. Gibson Jr, G. A. Nace, and C. Stewart, "Spectral and other physico-chemical properties of submicron powders of hematite ( $\alpha$ -Fe<sub>2</sub>O<sub>3</sub>), maghemite ( $\gamma$ -Fe<sub>2</sub>O<sub>3</sub>), magnetite (Fe<sub>3</sub>O<sub>4</sub>), goethite ( $\alpha$ -FeOOH) and lepidocrocite ( $\gamma$ -FeOOH)," *Journal of Geophysical Research*, vol. 90, no. 4, pp. 3126–3144, 1985.
- [40] G. R. Hunt, "Spectroscopic properties of rocks and minerals," in *Handbook of Physical Properties of Rocks*, R. S. Carmichael, Ed., vol. 1, pp. 295–385, CRC Press, Boca Raton, Fla, USA, 1982.
- [41] CSIR, *Weather Data of a Climate Station in the Thicket Biome, South Africa*, CSIR, Pretoria, South Africa, 2009.
- [42] E. R. Stoner and M. F. Baumgardner, "Characteristic variations in reflectance of surface soils," *Soil Science Society of America Journal*, vol. 45, no. 6, pp. 1161–1165, 1981.
- [43] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research*, vol. 89, no. 7, pp. 6329–6340, 1984.
- [44] R. N. Clark, G. A. Swayze, K. E. Livo et al., "Imaging spectroscopy: earth and planetary remote sensing with the USGS Tetracorder and expert systems," *Journal of Geophysical Research E*, vol. 108, no. 12, pp. 5–1, 2003.
- [45] F. van der Meer, "Analysis of spectral absorption features in hyperspectral imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 5, no. 1, pp. 55–68, 2004.
- [46] W. Kessler, *Multivariate Datenanalyse*, Wiley-VCH Verlag, Weinheim, Germany, 2007.
- [47] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [48] R. A. V. Viscarra Rossel, "ParLeS: software for chemometric analysis of spectroscopic data," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, pp. 72–83, 2008.
- [49] D. S. McIntyre, "Permeability measurements of soil crusts formed by raindrop impact," *Soil Science*, vol. 85, pp. 185–189, 1958.
- [50] R. N. Clark, T. V. V. King, M. Klejwa, G. A. Swayze, and N. Vergo, "High spectral resolution reflectance spectroscopy of minerals," *Journal of Geophysical Research*, vol. 95, no. 8, pp. 12–680, 1990.
- [51] G. R. Hunt, "Spectral signatures of particulate minerals in the visible and near infrared," *Geophysics*, vol. 42, no. 3, pp. 501–513, 1977.





**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

