Review Article

Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review

Qian Fang,[1,2] Hanlie Hong [ID],[1] Lulu Zhao,[1,3] Stephanie Kukolich,[3] Ke Yin,[1] and Chaowen Wang[4]

[1]*School of Earth Sciences, China University of Geosciences, Wuhan 430074, China*
[2]*Department of Soil, Water, and Environmental Science, University of Arizona, Tucson, AZ 85721, USA*
[3]*Department of Geosciences, University of Arizona, Tucson, AZ 85721, USA*
[4]*Gemological Institute, China University of Geosciences, Wuhan 430074, China*

Correspondence should be addressed to Hanlie Hong; honghl8311@aliyun.com

Clay minerals are the most reactive and important inorganic components in soils, but soil mineralogy classifies as a minor topic in soil sciences. Revisiting soil mineralogy has been gradually required. Clay minerals in soils are more complex and less well crystallized than those in sedimentary rocks, and thus, they display more complicated X-ray diffraction (XRD) patterns. Traditional characterization methods such as XRD are usually expensive and time-consuming, and they are therefore inappropriate for large datasets, whereas visible and near-infrared reflectance spectroscopy (VNIR) is a quick, cost-efficient, and nondestructive technique for analyzing soil mineralogic properties of large datasets. The main objectives of this review are to bring readers up to date with information and understanding of VNIR as it relates to soil mineralogy and attracts more attention from a wide variety of readers to revisit soil mineralogy. We begin our review with a description of fundamentals of VNIR. We then review common methods to process soil VNIR spectra and summary spectral features of soil minerals with particular attention to those <2 μm fractions. We further critically review applications of chemometric methods and related model building in spectroscopic soil mineral studies. We then compare spectral measurement with multivariate calibration methods, and we suggest that they both produce excellent results depending on the situation. Finally, we suggest a few avenues of future research, including the development of theoretical calibrations of VNIR more suitable for various soil samples worldwide, better elucidation of clay mineral-soil organic carbon (SOC) interactions, and building the concept of integrated soil mapping through combined information (e.g., mineral composition, soil organic matter-SOM, SOC, pH, and moisture).

1. Introduction

Soils are open, complex, and dynamic systems as well as fundamental natural environments for animals, plants, microorganisms, and human interaction [1]. Mineral composition is the most fundamental property of a soil, and soil minerals account generally for half the soil volume [2]. According to Churchman [3], clay minerals in the soil context are "secondary inorganic compounds of <2 μm size" including Fe, Al, and Mn oxides (hydroxides and oxyhydroxides), as well as noncrystalline phases. Importantly, they are the most reactive and important inorganic components in soils, and they occur commonly in close association with the most reactive

organic matter [4, 5]. Clays influence soil function through both their bulk properties and their associations with their huge outer/inner surfaces (e.g., cation exchange capacity [6]). The effort involved in comprehensive understanding of the nature of soil minerals is of particular importance as they may help us explain and predict how different soil types function [7].

However, soil mineralogy (mainly clay mineralogy) is still a minor topic in soil sciences. This may be due partly to the unjustified assumption that a given soil mineral will have the same properties as those of its better-crystallized counterpart that formed in a more "geologic" context (e.g., sedimentary kaolinite will have the same properties as pedogenic kaolinite) [4]. Revisiting soil mineralogy has

been gradually important, for instance, in terms of the manner by which soil minerals are defined and investigated [8].

The most commonly used method to characterize soil minerals is XRD, which is fundamentally qualitative. Since soil clay minerals are generally more complex and less well crystallized than those of geological environments [9–11], they display more complicated XRD patterns [12, 13]. Despite quantitative improvements of XRD [14], mineral characterization is usually expensive and time-consuming [2]. Some chemical extraction procedures can be useful in the analysis of Fe oxides. However, this is expensive, time-consuming, and can complicate our scientific interpretation of the soil by changing the chemical equilibrium between soil solution and solid phases in soil specimens [15, 16]. Thus, these conventional analyses are not appropriate for larger scale soil studies, and we must use an alternative method to target and characterize soil minerals.

Visible and near-infrared reflectance spectroscopy (VNIR, 350–2500 nm), that is, the study of light of the visible and near-infrared reflected from material surfaces, is a quick, cost-efficient, and nondestructive technique in soil sciences [17, 18]. This technique has been greatly developed in soil sciences in the past several decades and has seen apparent exponential growth over the past 20 years [19]. VNIR has been of increasing interest for the analyses of soil parameters including soil organic carbon, pH, bulk texture, elemental concentration, and cation exchange capacity [20, 21]. In soil mineralogy, VNIR can be used to characterize various soil mineralogic properties such as clay mineral composition, clay content, and mineral weathering/alteration degree, although quartz and feldspar have weak/nonexistent absorption in the VNIR range [22–24]. In this paper, we aim to bring readers up to date with VNIR as it relates to soil mineralogy and we seek to attract more attention from readers to revisit soil mineralogy.

2. Fundamentals of VNIR

The VNIR part of the electromagnetic spectrum includes both the visible (350–780 nm) and near-infrared (780–2500 nm) ranges, which overlaps with the optical radiation range (100–1000 nm; Figure 1). Sometimes, the 350–1000 wavelength range is referred as VNIR (visible-near-infrared), and the 1000–2500 range is referred as the SWIR (short-wave infrared) in remote sensing literature [25]. The human eyes and brain can process spectral information from the visible region and see color, while modern spectroscopy can observe precise details over a much broader wavelength range.

2.1. Absorption, Scattering, and Emission. When photons enter a solid, liquid, or gaseous material, they will either be absorbed, reflected from its surface, or pass through it [26]. The reflective process is defined as scattering, and the scattered photons can be detected and measured. Photons can also be detected when they are emitted from a surface with a temperature above absolute zero [25]. Three general physical processes (i.e., electronic transitions, vibrational
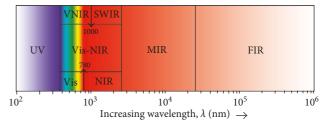

FIGURE 1: Optical spectrum between 10^2 to 10^6 nm.

transitions, and rotational transitions) result in the absorption bands in the spectra of materials. The absorption bands in the VNIR range are derived from both the electronic and vibrational transitions [27, 28].

2.2. Causes of Absorptions in the VNIR Region

2.2.1. Electronic Transitions. Discrete ions and atoms have independent energy states. A photon is emitted from an atom when one of its electrons moves to a lower energy state. When an atom absorbs a photon of a given wavelength, its electrons move from a relatively low electron state to a higher one [25]. These electron processes occur because of their high energy and mobility. The electronic processes are mainly caused by (1) crystal-field effects. Since iron is a very common transition element in minerals, a common electronic process revealed in the visible region is due to unfilled d-orbitals of Fe-oxide minerals [24, 29]. Electron energy levels are influenced by many factors, including the valence state of the atom (e.g., Fe^{2+} and Fe^{3+}), the type of ligands, the asymmetry of the location it occupies, the distance between the metal ion and the ligand, and the deformation degree of the site [28]. (2) Charge transfer: it is dominated by mineralogy, and it is hundred times more powerful than the crystal-field effects. It is the main reason of the red color of hydroxides and Fe oxides. Moreover, the conduction bands and color centers can also be causes of the electronic transitions in some minerals [25].

2.2.2. Vibrational Transitions. The bonds in a crystal lattice or molecule vibrate like springs. The molecule's mass and the strength of each molecular bond dominate their vibration frequency [25]. The absorption bands in the VNIR range are observed as a consequence of molecular vibrations [30]. Soil minerals (e.g., phyllosilicate and carbonate minerals), in particular, have unique absorption features in the VNIR region due to overtones and vibrational combinations related to the stretching and bending of the molecular bonds such as O-H, C-H, C-C, and N-H [31].

3. Spectroscopic Measurements

3.1. Spectral Preprocessing. The raw spectra are usually preprocessed through various approaches to accentuate features and remove signal noise [32]. The processed soil spectra facilitate mineral identification, and the accuracy of soil mineral prediction is greatly improved through the use of various preprocessing methods [33]. The following
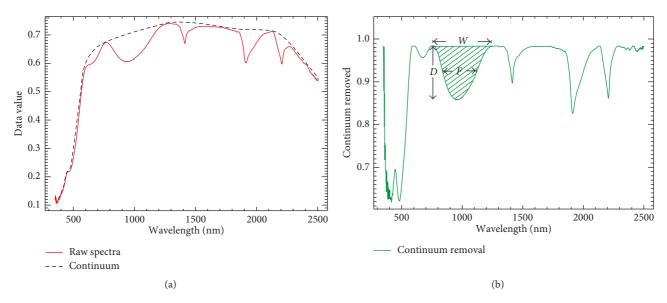
FIGURE 2: (a) Continuum removal of the spectrum corresponding to a red earth soil sample [24]; (b) a continuum-removed spectrum and some spectral parameters (i.e., *W*, *D*, and *F*).

preprocessing methods for spectra have been used in previous soil mineralogic studies.

3.1.1. Continuum Removal Approaches.
The continuum removal approach aims to remove background noise and isolate particular absorption features for identification and analysis [34]. The continuum is usually determined using local maxima to generate a hull of boundary points (Figure 2(a)) [22]. All the boundary points are fitted by straight-line segments, and then, the continuum removal is calculated by removing the original reflectance intensities from corresponding intensities of the continuum (Figure 2(b)) [23]. Continuum removal analysis is a particularly robust tool for detecting and predicting iron oxides and phyllosilicate minerals. Thus, it is feasible to substitute a statistical method to apply to soil mineralogy studies [10, 20, 22, 24].

Absorption bands in the VNIR region can be described by geometrical parameters derived from the continuum removal curve (Figure 2(b)). Four parameters are directly displayed in Figure 2(b), which include position (*P*), width (*W*), depth (*D*), and full width at half maximum (FWHM, abbreviated to "F"). The parameter asymmetry (AS) is calculated as follows:

$$AS = \frac{F_{left}}{F_{right}}, \tag{1}$$

where F_{left} represents the left width at half maximum, and F_{right} represents the right width at half maximum [20].

3.1.2. Smoothing Techniques.
Smoothing techniques are used to extract the maximum amount of information from each spectrum possibly by minimizing the influence of background noise [32]. Commonly used smoothing techniques include the Savitzky–Golay transform (SG [35]), Norris smoothing filter (NG [36]), and averaging spectra [37]. SG smoothing

eliminates the influences of ground interference noise and baseline float, thus enhancing the signal-to-noise ratio. NG smoothing removes the effects of particle-size variation when the soil samples vary in texture, moisture, and grain size [32].

3.1.3. Derivative Algorithms.
Derivative algorithms can rapidly identify characteristic positions of spectral minimum, maximum, and inflection point values [32]. Additionally, the effect of variation in optical setup and sample grinding is eliminated after derivative transformation [38]. Because the spectral noise tends to amplify with derivative transform, a smoothing technique is often used before the derivative algorithm [37]. The spectral curve after the first derivative, for example, is better at discriminating goethite and hematite and estimating their abundance, with two peaks at 435 and 535 nm for goethite and a single absorption at ∼570 nm for hematite (Figure 3) [39].

3.2. Spectral Features of Soil Minerals

3.2.1. Fe-Oxide Minerals.
Fe-oxide minerals are known to be pedogenic indicators for investigating soil temperature and moisture regimes, which are directly related to pedogenic climate evolution [24, 40]. Fe-oxide minerals are the main active components in the VNIR region (350–1000 nm) since most electron transitions are caused by various kinds of iron oxides [41, 42]. The most common Fe-oxide minerals in soils are goethite (α-FeOOH) and hematite (α-Fe$_2$O$_3$), which can track climate change [43, 44]. Goethite and hematite exhibit diagnostic spectral features in the VNIR region, and the absorption bands are generally broad and smooth (Figure 3). A strong absorption band near 920 nm indicates the presence of goethite (Figure 3(a)), and four absorption bands at 420, 480, 600, and 1700 nm can be used to map its distribution [39]. Hematite is dominated by three absorption bands at 520, 650, and 880 nm [45]. Both goethite and
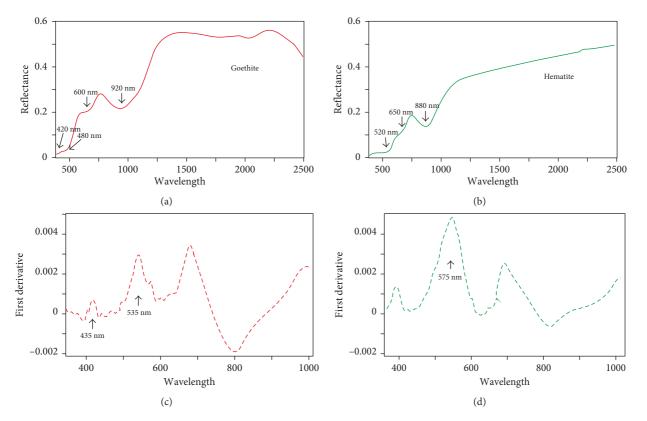
FIGURE 3: Raw spectra for goethite (a) and hematite (b) and the first-derivative spectra of goethite (c) and hematite (d) (modified from Zheng et al. [39]).

hematite have an absorption band at around 500 nm (480 for goethite and 520 for hematite, resp.); the band for goethite (at 480 nm) is narrow with intense reflectance, while the band for hematite (at 520 nm) is wide with low reflectance (Figure 3(a)). The absorptions in the VNIR region cause the vivid colors of Fe oxides, for example, yellow goethite and red hematite [37]. For a spectral curve representing a sample soil mixture, the width of the absorption band at ~870 nm (W_{870}) is higher when the soil sample contains more Fe oxides [46]. The concave shape of the 800–1000 nm range indicates the crystallinity of the Fe-oxide minerals. When a soil sample is composed of well-crystallized minerals, the corresponding spectrum reveals a symmetric and deeper feature in this range [47].

3.2.2. Clay Minerals. Clay minerals are frequently used as climatic indicators since their nature is directly influenced by the temperature and amount of precipitation at the site during pedogenesis [9, 48]. As climate conditions shift from cool/dry to warm/moist, the dominant clay minerals go from chlorite/illite → vermiculite → montmorillonite → kaolinite [24, 49]. The dominant clays in soils show diagnostic absorptions in the SWIR domain [39]. These absorption bands are caused by vibrational transitions and commonly display sharp and narrow features (Figure 4). The diagnostic bands are mainly focused on ~1400 nm (overtones caused by OH), ~1900 nm (overtones caused by molecular water), and ~2200 nm (combination tones caused by Al-OH [50, 51]).
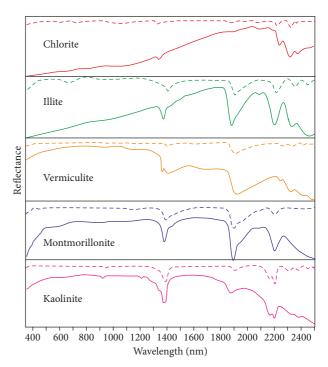

FIGURE 4: Raw and continuum-removed spectra of five common clay minerals.

Additionally, some weak absorption bands in the 2300–2500 nm region are related to the presence of Fe-OH and/or Mg-OH in the clay minerals [24].

TABLE 1: Spectral features of some predominant soil minerals.

Absorption band	Origin		Diagnostic feature	Description	Assigned soil minerals
~465 nm	Electronic transition	$^6A_{1g} \to {}^4T_{1g,4Eg}$	Width	Narrow	Goethite
				Wide	Hematite
~650 nm		$^6A_{1g} \to {}^4T_{2g}$	Position	Left → right	Goethite → hematite
~900 nm		$^6A_{1g} \to {}^4T_{1g}$	Position	Left → right	Hematite → goethite
~1400 nm	OH			Double absorption band (1395 nm and 1415 nm)	Kaolinite
			Asymmetry	>1	Kaolinite
				<1	Montmorillonite
			Position	<1910 nm	Montmorillonite/vermiculite
				>1910 nm	Kaolinite/illite
~1900 nm	H$_2$O		Depth	Intense → shallow	Montmorillonite → illite → kaolinite
	Molecular vibration		Width	Narrow	Montmorillonite
				Wide	Vermiculite/illite
				Double absorption band (2163 nm and 2208 nm)	Kaolinite
~2200 nm	Al-OH		Depth	Intense	Kaolinite
				Shallow	Vermiculite/montmorillonite/illite
			Asymmetry	>1	Kaolinite
				<1	Montmorillonite
>2300 nm	Fe-OH and Mg-OH			Additional absorptions (2340 nm and 2445 nm)	Illite
				Triple absorption features between 2330 and 2360 nm	Chlorite

The spectral characteristics of some clay minerals are showed in Figure 4 and Table 1. Chlorites are a group of clay minerals containing specific octahedral cations such as Fe, Mg, and Al [52]. Their reflectance spectra exhibit a weak absorption band at approximately 1400 nm and triple absorption features near 2300 nm. The bands at 2250 and 2350 nm are related to Fe-OH and Mg-OH, respectively [53]. Illite is characterized by three prominent absorptions at ~1400, ~1900, and ~2200 nm. Two secondary diagnostic Al-OH absorption peaks close to 2344 and 2445 nm are modified by Fe and Mg tschermak cation exchange [24, 31]. Vermiculite has two broad absorptions at 1400 and 1900 nm and two weak absorptions near 2200 and 2300 nm [39]. Montmorillonite has three strong and sharp absorption bands at ~1400, ~1900, and ~2200 nm, which are similar to but generally stronger than illite.

Additionally, the combination bands produced by the vibrations of absorbed water cause weak shoulders near 1468 nm and 1970 nm for montmorillonite spectra [37]. Kaolinite is featured by two spectral doublets: one is near 1400 nm (1390 and 1410 nm), and the other is near 2200 nm (2160 and 2210 nm).

3.2.3. Carbonates.

In soils, carbonates are leached from the surface with time and accumulate in the subsoil at a certain depth [54]. The presence of carbonate is widely used as a basic soil characteristic to describe soil types and quantify soil erosion [22]. Carbonates are characterized by several absorptions in the VNIR domain, caused by overtones and combinations of fundamental vibrations of the CO_3^{2-} ion (Figure 5) [31, 37]. A strong absorption band at ~2350 nm

FIGURE 5: Absorptions for calcite in the VNIR domain, caused by overtones and combinations of fundamental vibrations of the CO_3^{2-} ion.

and three weak absorption bands at ~1900, ~2000, and ~2160 nm were reported by Hunt and Salisbury [55] for carbonates in the NIR region, with the ~2350 nm absorption showing obvious double-band structures (Figure 5).

3.3. Prediction from the Continuum Removal Spectra.

As discussed in Section 3.1.1, several geometrical features of the absorption bands can be extracted through the continuum removal method. Those parameters (e.g., P, D, and AS) from the continuum removal spectra are key to characterizing and predicting mineral compositions in soils. Viscarra Rossel et al. [23] quantitatively estimated the mineral composition by using the continuum removal method. Compositions of soil minerals such as kaolinite, illite, Al-smectite, goethite, and hematite are considered in this study, and the parameter

D is selected for prediction. The spectroscopic predictions are generally in consistence with those interpreted by XRD analysis. According to Dufrechou et al. [20], the parameter D at ~1400, ~1900, and ~2200 nm was strongly affected by the amounts of kaolinite, illite, and montmorillonite in soil mixtures. Additionally, the estimation of montmorillonite abundance shows reliability when compared with XRD results. Five parameters (P, D, W, F, and AS) were used in the work by Zhao et al. [24] for assessing the utility of the continuum removal method. We compared these parameters with the results from both XRD and DRS analyses and found that some of the parameters are good at mineral content prediction. Furthermore, some parameters (e.g., AS at ~2200 nm) are confirmed as reliable proxies for soil weathering and paleoclimate reconstruction.

4. Chemometric Methods

VNIR spectra of soil mixtures are commonly weak and nonspecific due to (1) low concentration of particular soil minerals, (2) scatter effects caused by soil structure, (3) overlapping absorptions of soil attributes, and (4) influences of specific constituents such as quartz [37]. All of these factors pose a challenge for VNIR analyses. Therefore, useful information needs to be mathematically extracted from the spectra and correlated with soil attributes [45]. The development of VNIR in soil studies would have been impossible without the parallel application of chemometric methods [56].

Building a predictive soil mineral abundance model (i.e., multivariate calibration) is an important first step in chemometric analysis. Overall, we should understand the data and the objective of the modeling prior to building a model. Then, the spectral dataset is preprocessed and subdivided. Finally, we can proceed to build, evaluate, and select models [57].

4.1. Prior to Model Building. The first step in any model building process for the study of spectral pedology is to understand the characteristics of the dataset. We need to consider three main concepts in understanding the dataset process [57]: (1) *understanding the distribution of the responses (i.e., outcomes)*: the responses are either numerical or categorical. In the model building process for soil mineral analysis, the outcomes (e.g., contents of clay/Fe-oxide minerals) are described numerically. Understanding the characteristics of responses provides better ways for partitioning the data into calibration and validation sets; (2) *understanding the nature of the predictors*: the predictors in the spectral dataset are numerical, since they are usually the spectral signals between 350 and 2500 nm. In fact, these predictors are highly related, leading to numerically redundant information. Different predictors are suitable for different kinds of models. For example, partial least squares can be used for correlated predictors, while recursive partitioning can manage missing predictor information [58]; (3) *the relationship between the amount of the predictor set (P) and the sample set (N)*: when building a model for a soil

mineral study, the dataset commonly has far fewer samples ($N < 200$) than predictors ($P > 2000$). Therefore, a model that can handle dataset where $N < P$ is preferred.

After understanding the dataset, a preprocessing procedure is often used for improving the performance of the model [47, 59]. For a model used in a soil mineral study, the data transformations for multiple predictors contain the following methods: (1) *data reduction*: principal component analysis (PCA [60]) is a commonly used data reduction technique. In this technique, the number of datasets is largely reduced by seeking principle components (PC)—linear combinations of the predictors that capture the greatest possible variance; (2) *removing predictors*: in some cases, removing predictors prior to modeling has potential advantages. For example, Adeline et al. [59] showed that performances of the predictive models were globally stable and accurate when the spectral resolution decreased from 3 nm to 60 nm. Additionally, for a model based on a spectral signal dataset, the spectra were transformed to apparent absorbance: $A = \log(1/R)$ prior to developing a regression model, and the spectral preprocessing methods discussed in Section 3 also have potential for model performance improvement [18, 23].

4.2. Candidate Models. Once we fully understand the dataset, the next step is to setup several candidate models. The most commonly used type of model in soil mineral analysis is a regression model, which is defined as a model that predicts numerical outcomes [57]. Establishing a regression model related to the soil VNIR spectral data is the basic role of chemometric analysis [61]. The regression models are subdivided into linear and nonlinear regressions. Linear regressions are the dominant calibration methods for spectral pedology and include partial least squares regression (PLSR [62]) and principal component regression (PCR [63]). The nonlinear data are managed by data mining techniques, namely, multivariate adaptive regression splines (MARS [38]), neural networks (NN [64]), and regression tree analysis (RTA [65]).

4.2.1. Linear Regression Models. Both PLSR and PCR can deal with predictors that are highly collinear and are effective in situations where the number of predictors is far beyond the number of available samples [37]. Furthermore, PLSR and PCR are closely related and share similar prediction errors in most situations [61]. Regardless, the PLSR algorithm is usually preferred in spectral pedology analysis because (1) it maximizes covariance between response variables and predictors so that the model is more interpretable, and (2) it is a faster algorithm [45].

PLSR has been widely and successfully used in predicting the mineralogic compositions of weathering levels of soils. Viscarra Rossel et al. [2] accurately predicted the concentrations of kaolinite, illite, and smectite ($R^2 = 0.94$, 0.96, and 0.92, resp.) in mineral mixtures, although the prediction for Fe oxides was biased against measurement. Summers et al. [66] and Ostovari et al. [67] showed that the PLSR method is good at predicting $CaCO_3$ content with R^2 values of 0.69 and 0.71 for soil

samples from Australia and Iran, respectively. The total clay content and free iron in soils were also proven to be predictable attributes by the PLSR model [68, 69].

4.2.2. Nonlinear Regression Models.

The use of models that are inherently nonlinear in nature (i.e., data mining techniques) has gained increasing attention in recent years [37, 61]. A more detailed description of the nonlinear models is available in Kuhn and Johnson [57]. Previous studies have suggested that nonlinear regression models or the combination of nonlinear and linear models may provide better predictions for soil properties. Mouazen et al. [70] showed that a combined PLSR-NN model was better at predicting soil properties than a PLSR model. Viscarra Rossel and Behrens [45] proposed that the combined FS_{VIP}-ANN and FS_{MARS}-ANN models were the best models for predicting clay content, pH, and soil organic carbon (SOC) when both the parsimony and accuracy of the model were taken into consideration. Mulder et al. [71] determined the mineral composition of a soil by coupling an RTA model with exponential Gaussian optimization results. The abundances of kaolinite and calcite were predicted with acceptable RMSE values (<0.1) in both laboratory and field samples.

4.3. Model Evaluation.

Two techniques are commonly used to test the prediction performance of the model [33, 56]. In the first, the soil spectral configuration database is randomly divided into a calibration dataset and a validation dataset [72]. The calibration dataset (usually ~2/3 of the complete database) is used to derive the model, while the validation dataset (commonly contains 1/3 of the complete database) is set aside to exclusively validate the derived model. This process is used to obtain realistic estimates of prediction accuracy. The second is a procedure called cross validation. It uses the "leave-group-out" method (namely, repeated random subsampling validation method [73]) and was adopted to verify the predictive capability for the calibration dataset. A calibration dataset containing X samples is built from the total database N ($N \geq X + 1$). The soil property value of the other $N - X$ samples for validation is predicted. The prediction of relative soil mineral abundance is obtained by repeating the cross-validation process [74].

ParLeS version 3.0 is usually used for multivariate calibration performance [75]. The bias and accuracy of the prediction models are assessed by adjusting the coefficient of determination (R^2) between observed and predicted values, the mean error (ME)

$$ME = \sum_{i=1}^{N} \left(Y_i' - Y_i\right), \tag{2}$$

and the root mean-square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{n} \left(Y_i' - Y_i\right)^2}, \tag{3}$$

where N is the number of the dataset, Y_i is the observed value, and Y_i' represents the predicted value [18, 23].

We compromise between model parsimony and model accuracy to find the most satisfactory model [76]. The Akaike information criterion (AIC) is suggested for best-performing algorithm selection [45]:

$$AIC = 2p + n \times (\ln RMSE), \tag{4}$$

where p is the number of factors, and n is the number of samples used in the prediction. The best model will have the minimum AIC value.

4.4. Feature Selection.

Feature selection is mainly applied to remove redundant and/or noninformative predictors from the model [57] and may improve model accuracy. Some models such as PLSR, MARS, and RT will provide a feature selection procedure by default.

The variable importance of the projection (VIP) and b-coefficient scores obtained by the PLSR model help us measure the statistical significance of predictors and select the most important ones [77]. The VIP score of the kth predictor is calculated as follows:

$$VIPk = K \cdot \sum_{a=1}^{a_{opt}} w_{ak}^2 \cdot R_a^2, \tag{5}$$

where K represents the total number of the predictor variables, a_{opt} is an optimal number of latent variables selected by the PLSR model, w_{ak} is the loading weight for the kth latent variable, and R_a^2 represents the adjust coefficient of determination of the ath latent variable in the PLSR model [78]. A predictor (such as wavelength) is selected and considered to be very important if (1) the VIP exceeds the threshold value of one (Chong and Jun [77]) and (2) its b-coefficient is higher than the b-coefficient based on all spectral bands [23].

According to Gomez et al. [22], the important spectral bands selected by the PLSR model are related to the presence of clay minerals such as kaolinite and illite. Additionally, surrogate spectral features selected by VIP and b-coefficient approach contain enough information to satisfactorily estimate the studied soil attributes. According to Viscarra Rossel and Behrens [45], a combined NN and feature selection model (FS_{vip}-ANN) is the best method to predict clay content and pH and produce smaller RSME and AIC values.

5. Comparison between Spectral Measurement and Multivariate Calibration

The relative abundance of minerals in a soil sample can be predicted either by spectral analyses (e.g., continuum removal) or chemometric methods (e.g., PLSR and NN) [20, 24, 45, 47]. Although both types of methods correlate the spectral signal with information about the soil minerals, they differ in many ways, including their focused spectral bands, complexity, and how they are applied.

5.1. Focused Spectral Bands. Spectral analyses focus on specific absorption bands representative of the corresponding soil minerals, while the multivariable regression algorithms commonly use the signals from the whole 350–2500 nm region. In some cases, the 350–400 nm and 2450–2500 nm ranges with low instrumental signal-to-noise ratios are removed [59, 79]. Therefore, a multivariable regression model deals with over 1000 spectral bands—many more than the number of focused bands in a continuum removal study. Moreover, several geometric parameters can be extracted from each band in a spectral measurement, including *P*, *W*, *D*, *F*, and AS, whereas only the information of depth for each band can be gleaned from a chemometric study. Note that some algorithms intrinsically provide a feature selection method (e.g., SMLR and PLSR), and it has been shown that the most important features selected by a regression model are the ones that we should pay the most attention to in a spectral measurement study [22].

5.2. Complexity. Theoretically, multivariate calibration is very complicated because it involves a larger number of algorithms and because different algorithms have the potential to be combined into better predictive models, depending on the situation [45, 70, 79]. However, in practice, multivariate modeling and prediction is not that complicated. Thanks to the development of executable and fast running software such as ParLeS and Unscrambler [75, 80], the difficult calculation process can be done much more easily. On the other hand, spectral measurement studies cost more time because we must (1) identify a soil mineral based on the spectral features, (2) extract parameters from the bands, and (3) relate those parameters with the information about the soil mineral.

5.3. Application Preference. The geometric features of the spectra are more suitable for monitoring the molecular structural changes of soil minerals, since the variations of the absorption bands are caused by electron transitions (e.g., Fe^{2+} to Fe^{3+}) and molecular vibration (e.g., Al-OH versus H_2O). Thus, spectral measurement is widely and successfully applied to (1) measure mineral physicochemistry that is sensitive to changes in metamorphic grade [53, 81], (2) map and monitor mineral erosion, deposition, and weathering of minerals [24, 51], and (3) explore water and potential life on extraterrestrial objects [10, 21]. Chemometric methods are more often used in monitoring overall soil properties, since almost all of the signals in the VNIR domain are involved in the modeling process. Several soil attributes are successfully determined by an appropriate multivariate calibration technique, including soil clay [23, 69], organic matter [32, 67], and nitrogen content [82, 83].

Table 2 is a review of some soil mineralogic attributes predicted by VNIR spectroscopy using either chemometric analysis or spectral-based measurement. In this summary, most of the studies used soil samples for analysis, and many of them are among diverse soil types (Table 2). The predictions of the soil properties are still good when there is great range of soil types (e.g., 22, 45, 84, and 85). A single

mineral (e.g., kaolinite and goethite) is more precisely predicted when mineral mixtures are used in the measurements [2, 86]. The studies in Table 2 include both data collected in the laboratory and data based on field soil sensing. In the lab, the sample pretreatment and illumination conditions can be controlled to eliminate the influences of the moisture and the grain size of the soil sample [18]. While in the field, the VNIR spectroscopy may be affected by many potential problems such as variable distances between the sensor and the soil, the smearing of soil surfaces, the size of the soil aggregates, and the amount of moisture [87]. These potential problems may reduce the prediction accuracy of field-based analysis [22, 85]. However, the field-based VNIR spectroscopy is more attractive because it (1) enables the potential analysis of soil properties with promising results in previous studies [87] and (2) reduces the cost of the measurement by simplifying the sample preparation. Based on the results of the studies, PLSR is proved to be the most robust soil mineralogic analysis method amongst all of the multivariate calibrations (Table 2; $0.43 < R^2 < 0.96$). The CR-based model is good at predicting clay mineral concentration (e.g., 20, 47, and 88; $R^2 > 0.79$). In some cases, the nonlinear models (NN and MARS) exhibit better estimation in predicting soil mineralogy than the PLSR model (e.g., 45 and 89). In general, when a soil mineral is investigated by spectroscopy, the PLSR and the CR-based models are the most promising methods to provide estimates of mineral abundance.

6. Conclusions and Future Research Directions

Clay minerals in soils are more complex and less well crystallized than those in sedimentary rocks. Traditional characterization methods such as XRD are usually expensive and time-consuming, whereas VNIR is a quick, cost-efficient, and nondestructive technique for analyzing the soil mineralogic properties of large datasets. The major strength of soil mineralogy studies is that there is a direct relationship between soil minerals and their spectra, since the diagnostic absorption bands of soil minerals lie within the VNIR region. Therefore, the nature of soil mineralogy can be approached through both spectral measurement and multivariate calibration. The spectral measurement is focused on geometric information extracted from several bands (e.g., 350–400, ~1900, ~2200, and 2450–2500 nm) that relate to soil minerals. The parameters derived from the continuum removal method are mainly used for mineral identification and prediction. In a multivariate calibration analysis, the dataset contains the entire VNIR domain. The most robust model for soil mineral estimation is selected after understanding the data, data preprocessing, candidate model building, and performance assessment.

Firstly, VNIR has been greatly developed in soil sciences over the past several decades. However, no definitive results on theoretical calculations have yet been found because most soil studies occur on a regional scale so their results are only regionally representative. Thus, it is essential to further develop the theoretical calibrations of VNIR that are more suitable for soil samples worldwide, despite difficulties due to high soil variability across the globe.

TABLE 2: A review of VNIR spectroscopy for the prediction of soil attributes using feature-based and chemometric methods in previous studies.

Soil mineralogic attribute[a]	Soil type[b]	Location	Sample[c]	n_{cal}/n_{val}[d]	Method[e]	R^2[f]	RMSE[g]	Source
Clay$_T$	Alfisols and entisols	Romania	Lab/sieved (2)-air-dried	210/90	SVR	0.27	84.4	[90]
Clay$_T$	Not mentioned	France	Lab/sieved (2)-air-dried	99/49	PLSR	0.76	33.6	[59]
Clay$_T$	Diverse	Romania	Lab/sieved (2)-air-dried	210/90	SVR	0.46	85.3	[91]
Clay$_T$	Vertisols and alfisols	India	Lab/sieved (2)-air-dried	175/58	PLSR	0.80	2.2	[92]
Clay$_T$	Diverse	Denmark	Lab/sieved (2)-air-dried	480$_{CV}$	PLSR	0.94	17	[84]
Clay$_T$	Entisol and inceptisols	Italy	Lab/sieved (2)-oven-dried	175/60	PLSR	0.80	0.17	[93]
Clay$_T$	Diverse	Australia	Lab/sieved (2)-air-dried	1104$_{CV}$	DWT-ANN	0.88	64.2	[45]
Clay$_T$	Diverse	Australia	Lab/sieved (2)-air-dried	1104$_{CV}$	PLSR	0.83	77.7	[45]
Clay$_T$	Entisol	Turkey	Lab/sieved (2)-air-dried	502$_{CV}$	PLSR	0.84	38.20	[89]
Clay$_T$	Entisol	Turkey	Lab/sieved (2)-air-dried	502$_{CV}$	MARS	0.86	35.10	[89]
Clay$_T$	Diverse	France	Lab/sieved (2)-oven-dried	52$_{CV}$	PLSR	0.85	31.2	[22]
Clay$_T$	Diverse	France	Lab/sieved (2)-oven-dried	52$_{CV}$	CR	0.73	44	[22]
Clay$_T$	Diverse	France	Field/fresh	52$_{CV}$	PLSR	0.64	49.6	[22]
Clay$_T$	Diverse	France	Field/fresh	52$_{CV}$	CR	0.58	82	[22]
Clay$_T$	Diverse	USA	Lab/sieved (2)-air-dried	72$_{CV}$	PLSR	0.92	41	[85]
Clay$_T$	Diverse	USA	Field/fresh	72$_{CV}$	PLSR	0.83	61	[85]
Clay$_K$	Diverse	Australia	Lab/sieved (2)-air-dried	102$_{CV}$	PLSR	0.95	111.5	[94]
Clay$_K$	Not mentioned	China	Lab/sieved (2)-air-dried	20$_{CV}$	CR-MLR	0.80	10.41	[88]
Clay$_K$	Diverse topsoil	Australia	Lab/sieved (2)-air-dried	4606$_{CV}$	Model trees	0.52	0.8	[95]
Clay$_K$	Diverse subsoil	Australia	Lab/sieved (2)-air-dried	2492$_{CV}$	Model trees	0.46	1.1	[95]
Clay$_K$	Mineral-organic mixes	Australia	Lab/sieved (0.2)-oven-dried	8$_{CV}$	PLSR	0.94	0.36	[2]
Clay$_I$	Diverse	Australia	Lab/sieved (2)-air-dried	90$_{CV}$	PLSR	0.96	102.1	[94]
Clay$_I$	Not mentioned	China	Lab/sieved (2)-air-dried	20$_{CV}$	CR-MLR	0.79	28.14	[88]
Clay$_I$	Diverse topsoil	Australia	Lab/sieved (2)-air-dried	4606$_{CV}$	Model trees	0.41	1.2	[95]
Clay$_I$	Diverse subsoil	Australia	Lab/sieved (2)-air-dried	2492$_{CV}$	Model trees	0.40	1.5	[95]
Clay$_I$	Mineral-organic mixes	Australia	Lab/sieved (0.2)-oven-dried	8$_{CV}$	PLSR	0.96	0.34	[2]
Clay$_S$	Diverse	Australia	Lab/sieved (2)-air-dried	98$_{CV}$	PLSR	0.94	118.7	[94]
Clay$_S$	Not mentioned	China	Lab/sieved (2)-air-dried	20$_{CV}$	CR-MLR	0.84	21.35	[88]
Clay$_S$	Diverse	France	Lab/sieved (2)-air-dried	63$_{CV}$	CR	0.83	158	[20]
Clay$_S$	Diverse topsoil	Australia	Lab/sieved (2)-air-dried	4606$_{CV}$	Model trees	0.61	0.8	[95]
Clay$_S$	Diverse subsoil	Australia	Lab/sieved (2)-air-dried	2492$_{CV}$	Model trees	0.44	1.2	[95]
Clay$_S$	Mineral-organic mixes	Australia	Lab/sieved (0.2)-oven-dried	8$_{CV}$	PLSR	0.92	0.34	[2]
Clay$_S$	Not mentioned	USA	Lab/sieved (2)-air-dried	178$_{CV}$	PLSR	0.83	1	[96]
Clay$_{2:1}$	Argiudoll, hapludolls, and eutrudox	Brazil	Lab/sieved (2)-oven-dried	29$_{CV}$	CR	0.80	—	[47]
Fe$_T$	Not mentioned	France	Lab/sieved (2)-air-dried	99/49	PLSR	0.84	24[h]	[59]
Fe$_T$	Vertisols and alfisols	India	Lab/sieved (2)-air-dried	175/58	PLSR	0.78	0.15[h]	[92]
Fe$_T$	Diverse	Moravia	Lab/sieved (2)-air-dried	97$_{CV}$	MLR	0.37	12.76[h]	[97]
Fe$_T$	Cambisols and luvisols	South Africa	Lab/sieved (2)-air-dried	123/40	CR-MLR	0.21	9.3	[98]
Fe$_T$	Cambisols and luvisols	South Africa	Field/fresh	94/31	CR-MLR	0.23	16.3	[98]

TABLE 2: Continued.

Soil mineralogic attribute[a]	Soil type[b]	Location	Sample[c]	n_{cal}/n_{val}[d]	Method[e]	R^2[f]	RMSE[g]	Source
Fe$_T$	Cambisols and luvisols	South Africa	Lab/sieved (2)-air-dried	123/40	PLSR	0.43	9.8	[98]
Fe$_T$	Cambisols and luvisols	South Africa	Field/fresh	94/31	PLSR	0.43	8.0	[98]
Fe$_H$	Typic hapludox and typic eutrudox	Brazil	Lab/sieved (2)-air-dried	206$_{CV}$	PLSR	0.77	10.7	[99]
Fe$_G$	Fe-oxide mixes	USA	Lab/sieved (2)-oven-dried	6$_{CV}$	CR	0.95	25.3	[86]
C	Entisols, inceptisols, and aridisols	Iran	Lab/sieved (2)-air-dried	30/10	PLSR	0.71	52.4	[67]
C	Not mentioned	France	Lab/sieved (2)-air-dried	99/49	PLSR	0.89	42.1	[59]
C	Diverse topsoil	Iran	Lab/sieved (2)-air-dried	251$_{CV}$	PLSR	0.58	117	[100]
C	Diverse topsoil	Iran	Lab/sieved (2)-air-dried	251$_{CV}$	CR	0.35	149	[100]
C	Entisol	Turkey	Lab/sieved (2)-air-dried	502$_{CV}$	PLSR	0.65	6.80	[89]
C	Entisol	Turkey	Lab/sieved (2)-air-dried	502$_{CV}$	MARS	0.71	6.20	[89]
C	Diverse	France	Lab/sieved (2)-oven-dried	52$_{CV}$	PLSR	0.94	38.34	[22]
C	Diverse	France	Lab/sieved (2)-oven-dried	52$_{CV}$	CR	0.92	52	[22]
C	Diverse	France	Field/fresh	52$_{CV}$	PLSR	0.77	76.67	[22]
C	Diverse	France	Field/fresh	52$_{CV}$	CR	0.47	132	[22]

[a]Soil attributes: clay$_T$, total clay content; clay$_{2:1}$,2:1 clay mineral content; Clay$_K$, kaolinite content; Clay$_I$, illite content; Clay$_S$, smectite content; Fe$_T$, total iron content; Fe$_H$, hematite content; Fe$_G$, goethite content; C, carbonate content. [b]soil type: soil type is not given in the study (not mentioned); the soil samples in the study include more than three soil types (diverse). [c]sample treatment: oven-dried, sieved (<2 mm), and measured under laboratory conditions (lab/sieved (2)-oven-dried); oven-dried, sieved (<0.2 mm), and measured under laboratory conditions (lab/sieved (0.2)-oven-dried); air-dried, sieved (<2 mm), and measured under laboratory conditions (lab/sieved (2)-air-dried); fresh (wet and unprocessed) and measured under field conditions (field/fresh). [d]n_{cal}/n_{val} shows the number of samples used in the spectral calibration and validation, respectively. CV indicates that the validation was conducted independently using a statistical cross-validation technique. [e]partial least squares regression (PLSR); multivariate adaptive regression splines (MARS); support vector regression (SVR); the discrete wavelet transform-artificial neural networks (DWT-ANN); continuum removal (CR); feature-based multiple linear regression (CR-MLR). [f]coefficient of determination. [g]root mean-square error (g·kg^{-1}). —, RMSE is not given in the study. [h]mg·kg^{-1}.

Secondly, more field analyses are required for obtaining full potential of VNIR. The in situ data collection in the field is one of the advantages compared with conventional techniques. The heterogeneity of the technical and environmental factors (e.g., soil moisture, soil surface condition, and biological residue) will directly influence the characteristics of the absorption bands, causing increased uncertainty of the spectral measurements. Nevertheless, multivariate calibration models for field data show good or even better mineral prediction than laboratory data. There has been a lack of more systematic studies on the various effects of field sample data and variations in mineralogy, moisture, organic matter, and their interactions. Therefore, future work should focus on these types of studies rather than laboratory spectra.

Thirdly, VNIR may have the potential to help us investigate interactions between soil clay minerals and SOC. Mechanisms of SOC stabilization have attracted increasing interest due to their potential to influence the global carbon cycle. It is widely suggested that soil clay minerals play a central role in capturing and permanently sequestering atmospheric CO_2. Both clay content and clay mineral type exert important influences on the carbon sequestration. Because VNIR is capable of characterizing most of carbon- and hydroxyl-related properties, it should allow us to study clay-SOC interactions when combined with the other common or state-of-the-art techniques.

Finally, integrated soil mapping is needed in future large-scale soil analysis. The VNIR spectrum contains integrative information (e.g., mineral composition, SOM, SOC, pH, and moisture) of the soil attributes that reflect the nature of a soil system. Thus, we could use VNIR to map soils. More collaborative and strategic spectral studies are needed to better understand the complete nature of soil [101, 102]. Some global or national spectral libraries [103, 104] have been established to build collaborative networks for soil spectroscopy, but more spectral libraries will facilitate the wider use of VNIR and make global-scale soil monitoring possible.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] F. Riedel, M. Denk, I. Müller, N. Barth, and C. Gläßer, "Prediction of soil parameters using the spectral range between 350 and 15,000 nm: a case study based on the

permanent soil monitoring program in Saxony, Germany," *Geoderma*, vol. 315, pp. 188–198, 2018.

[2] R. A. Viscarra Rossel, R. N. McGlynnand, and A. B. McBratney, "Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy," *Geoderma*, vol. 137, no. 1, pp. 70–82, 2006.

[3] G. J. Churchman, "Is the geological concept of clay minerals appropriate for soil science? A literature-based and philosophical analysis," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 35, no. 15–18, pp. 927–940, 2010.

[4] G. J. Churchman and D. J. Lowe, "Alteration, formation, and occurrence of minerals in soils," in *Handbook of Soil Sciences*, P. M. Huang, Y. Li, and M. E. Sumner, Eds., Vol. 1, CRC Press, Boca Raton, FL, USA, 2nd edition, 2012.

[5] M. Singh, B. Sarkar, S. Sarkar et al., "Stabilization of soil organic carbon as influenced by clay mineralogy," in *Advances in Agronomy*, pp. 33–84, Elsevier, New York, NY, USA, 2018.

[6] E. Arthur, "Rapid estimation of cation exchange capacity from soil water content," *European Journal of Soil Science*, vol. 68, no. 3, pp. 365–373, 2017.

[7] B. Abdelkader, A. Ahmed, B. Mosterfa, and S. Isam, "Laboratory study of geotextiles performance on reinforced sandy soil," *Journal of Earth Science*, vol. 27, no. 6, pp. 1060–1070, 2017.

[8] G. J. Churchman, "Game changer in soil science. Functional role of clay minerals in soil," *Journal of Plant Nutrition and Soil Science*, vol. 181, no. 1, pp. 99–103, 2018.

[9] H. Hong, F. Cheng, K. Yin, G. J. Churchman, and C. Wang, "Three-component mixed-layer illite/smectite/kaolinite (I/S/K) minerals in hydromorphic soils, south China," *American Mineralogist*, vol. 100, no. 8-9, pp. 1883–1891, 2015.

[10] Q. Fang, G. J. Churchman, H. Hong et al., "New insights into microbial smectite illitization in the Permo-Triassic boundary K-bentonites, South China," *Applied Clay Science*, vol. 140, pp. 96–111, 2017.

[11] Q. Fang, H. Hong, L. Zhao et al., "Tectonic uplift-influenced monsoonal changes promoted hominin occupation of the Luonan Basin: Insights from a loess-paleosol sequence, eastern Qinling Mountains, central China," *Quaternary Science Reviews*, vol. 169, pp. 312–329, 2017.

[12] H. Hong, G. J. Churchman, K. Yin, R. Li, and Z. Li, "Randomly interstratified illite–vermiculite from weathering of illite in red earth sediments in Xuancheng, southeastern China," *Geoderma*, vol. 214-215, pp. 42–49, 2014.

[13] K. Yin, H. Hong, G. J. Churchman, Z. Li, and Q. Fang, "Mixed-layer illite-vermiculite as a paleoclimatic indicator in the Pleistocene red soil sediments in Jiujiang, southern China," *Palaeogeography, Palaeoclimatology, Palaeoecology*, 2018, In press.

[14] J. Srodon, V. A. Drits, D. K. McCarty, J. C. Hsieh, and D. D. Eberl, "Quantitative X-ray diffraction analysis of clay-bearing rocks from random preparations," *Clay and Clay Mineral*, vol. 49, no. 6, pp. 514–528, 2001.

[15] X. Huang, H. Jiang, Y. Li et al., "The role of poorly crystalline iron oxides in the stability of soil aggregate-associated organic carbon in a rice–wheat cropping system," *Geoderma*, vol. 279, pp. 1–10, 2016.

[16] E. K. Coward, A. T. Thompson, and A. F. Plante, "Iron-mediated mineralogical control of organic matter accumulation in tropical soils," *Geoderma*, vol. 306, pp. 206–216, 2017.

[17] D. J. Brown, K. D. Shepherd, M. G. Walsh, M. Dewayne Mays, and T. G. Reinsch, "Global soil characterization with VNIR diffuse reflectance spectroscopy," *Geoderma*, vol. 132, no. 3-4, pp. 273–290, 2006.

[18] M. Nocita, A. Stevens, C. Noon, and B. van Wesemael, "Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy," *Geoderma*, vol. 199, pp. 37–42, 2013.

[19] V. Bellon-Maurel and A. McBratney, "Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils–critical review and research perspectives," *Soil Biology and Biochemistry*, vol. 43, no. 7, pp. 1398–1410, 2011.

[20] G. Dufrechou, G. Grandjean, and A. Bourguignon, "Geometrical analysis of laboratory soil spectra in the short-wave infrared domain: clay composition and estimation of the swelling potential," *Geoderma*, vol. 243-244, pp. 92–107, 2015.

[21] T. A. Goudge, J. M. Russell, J. F. Mustard, J. W. Head, and S. Bijaksana, "A 40,000 yr record of clay mineralogy at Lake Towuti, Indonesia: paleoclimate reconstruction from reflectance spectroscopy and perspectives on paleolakes on Mars," *Geological Society of America Bulletin*, vol. 129, no. 7-8, pp. 806–819, 2017.

[22] 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.

[23] 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, pp. 253–266, 2009.

[24] L. Zhao, H. Hong, J. Liu et al., "Assessing the utility of visible-to-shortwave infrared reflectance spectroscopy for analysis of soil weathering intensity and paleoclimate reconstruction," *Palaeogeography, Palaeoclimatology, Palaeoecology*, 2018, In press.

[25] R. N. Clark, "Spectroscopy of rocks and minerals, and principles of spectroscopy," in *Manual of Remote Sensing*, Vol. 3, pp. 3–58, John Wiley and Sons, New York, NY, USA, 1999.

[26] M. Zude, "Optical monitoring of fresh and processed agricultural crops," in *Contemporary Food Engineering*, CRC Press, Boca Raton, FL, USA, 2008.

[27] V. C. Farmer, "Infra-red spectroscopy in mineral chemistry," in *Physicochemical Methods of Mineral Analysis*, pp. 357–388, Springer, Berlin, Germany, 1975.

[28] R. G. Burns, *Mineralogical Applications of Crystal Field Theory*, Cambridge University Press, Cambridge, UK, 2nd edition, 1993.

[29] M. Tappert, B. Rivard, D. Giles, R. Tappert, and A. Mauger, "Automated drill core logging using visible and near-infrared reflectance spectroscopy," *Economic Geology*, vol. 106, no. 2, pp. 289–296, 2011.

[30] H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise, *Near-Infrared Spectroscopy: Principles, Instruments, Applications*, John Wiley and Sons, Weinheim, Germany, 1st edition, 2008.

[31] R. N. Clark, T. King, M. Klejwa, G. A. Swayze, and N. Vergo, "High spectral resolution reflectance spectroscopy of minerals," *Journal of Geophisical Research-Solid Earth and Plants*, vol. 95, no. B8, pp. 12653–12680, 1990.

[32] Y. Tian, J. Zhang, X. Yao, W. Cao, and Y. Zhu, "Laboratory assessment of three quantitative methods for estimating the organic matter content of soils in China based on visible/near-infrared reflectance spectra," *Geoderma*, vol. 202-203, pp. 161–170, 2013.

[33] A. H. Cambule, D. G. Rossiter, J. J. Stoorvogel, and E. M. A. Smaling, "Building a near infrared spectral library

for soil organic carbon estimation in the Limpopo national park, Mozambique," *Geoderma*, vol. 183-184, pp. 41–48, 2012.

[34] R. N. Clark and T. L. Roush, "Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications," *Journal of Geophysical Research: Solid Earth*, vol. 89, no. B7, pp. 6329–6340, 1984.

[35] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[36] K. H. Norris and P. C. Williams, "Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. influence of particle size," *Cereal Chemistry*, vol. 61, no. 2, pp. 158–165, 1984.

[37] S. Bo, R. A. V. Rossel, A. M. Mouazen, and J. Wetterlind, "Chapter five–visible and near infrared spectroscopy in soil science," *Advances in Agronomy*, vol. 107, no. 107, pp. 163–215, 2010.

[38] K. D. Shepherd and M. G. Walsh, "Development of reflectance spectral libraries for characterization of soil properties," *Soil Science Society of America Journal*, vol. 66, no. 3, pp. 988–998, 2002.

[39] G. Zheng, C. Jiao, S. Zhou, and G. Shang, "Analysis of soil chronosequence studies using reflectance spectroscopy," *International Journal of Remote Sensing*, vol. 37, no. 8, pp. 1881–1901, 2016.

[40] P. Hu, Z. Jiang, Q. Liu et al., "Estimating the concentration of aluminum-substituted hematite and goethite using diffuse reflectance spectrometry and rock magnetism: feasibility and limitations," *Journal of Geophysical Research-Solid Earth*, vol. 121, no. 6, pp. 4180–4194, 2016.

[41] R. M. Cornell and U. Schwertmann, *The Iron Oxides: Structure, Proprieties, Reactions, Occurences and Uses*, Wiley-VCH Verlag, Weinheim, Germany, 2nd edition, 2003.

[42] R. J. Murphy and S. T. Monteiro, "Mapping the distribution of ferric iron minerals on a vertical mine face using derivative analysis of hyperspectral imagery (430-970 nm)," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 75, pp. 29–39, 2013.

[43] J. Torrent, Q. Liu, J. Bloemendal, and V. Barron, "Magnetic enhancement and iron oxides in the upper luochuan Loess-paleosol sequence, Chinese Loess plateau," *Soil Science Society of America Journal*, vol. 71, no. 5, pp. 1570–1578, 2007.

[44] L. Zhao, H. Hong, Q. Fang et al., "Monsoonal climate evolution in southern China since 1.2 Ma: new constraints from Fe-oxide records in red earth sediments from the Shengli section, Chengdu Basin," *Palaeogeography, Palaeoclimatology, Palaeoecology*, vol. 473, pp. 1–15, 2017.

[45] R. A. Viscarra Rossel and T. Behrens, "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma*, vol. 158, no. 1, pp. 46–54, 2010.

[46] 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, 1986.

[47] N. D. Sheldon and N. J. Tabor, "Quantitative paleoenvironmental and paleoclimatic reconstruction using paleosols," *Earth-Science Reviews*, vol. 95, no. 1-2, pp. 1–52, 2009.

[48] J. A. M. Demattê and F. Da Silva Terra, "Spectral pedology: a new perspective on evaluation of soils along pedogenetic alterations," *Geoderma*, vol. 217-218, pp. 190–200, 2014.

[49] H. W. Nesbitt, C. M. Fedo, and G. M. Young, "Quartz and feldspar stability, steady and non-steady-state weathering, and petrogenesis of siliciclastic sands and muds," *Journal of Geology*, vol. 105, no. 2, pp. 173–192, 1997.

[50] R. S. Da Cruz, C. M. Dias Fernandes, R. N. Nobre Villas et al., "A study of the hydrothermal alteration in Paleoproterozoic volcanic centers, Sao Felix do Xingu region, Amazonian Craton, Brazil, using short-wave infrared spectroscopy," *Journal of Volcanology and Geothermal Research*, vol. 304, pp. 324–335, 2015.

[51] T. Cudahy, M. Caccetta, M. Thomas et al., "Satellite-derived mineral mapping and monitoring of weathering, deposition and erosion," *Scientific Reports*, vol. 6, no. 23702, pp. 1–12, 2016.

[52] M. S. Abweny, F. J. A. van Ruitenbeek, B. de Smeth et al., "Short-Wavelength Infrared (SWIR) spectroscopy of low-grade metamorphic volcanic rocks of the Pilbara Craton," *Journal of African Earth Sciences*, vol. 117, pp. 124–134, 2016.

[53] M. P. Doublier, T. Roache, and S. Potel, "Short-wavelength infrared spectroscopy: a new petrological tool in low-grade to very low-grade pelites," *Geology*, vol. 38, no. 11, pp. 1031–1034, 2010.

[54] D. Badía, C. Martí, J. Casanova et al., "A quaternary soil chronosequence study on the terraces of the Alcanadre river (semiarid Ebro Basin, NE Spain)," *Geoderma*, vol. 241-242, pp. 158–167, 2015.

[55] G. R. Hunt and J. W. Salisbury, "Visible and near infrared spectra of minerals and rocks. II. carbonates," *Modern Geology*, vol. 2, pp. 23–30, 1971.

[56] H. M. Heiseand and R. Winzen, "Chemometrics in near-infrared spectroscopy," in *Near-Infrared Spectroscopy: Principles, Instruments, Applications*, H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise, Eds., pp. 125–162, Wiley-VCH, Weinheim, Germany, 2002.

[57] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, New York, NY, USA, 2013.

[58] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: a conditional inference framework," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 651–674, 2006.

[59] K. R. M. Adeline, C. Gomez, N. Gorretta, and J. M. Roger, "Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data," *Geoderma*, vol. 288, pp. 143–153, 2017.

[60] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[61] X. Shao, X. Bian, J. Liu, M. Zhang, and W. Cai, "Multivariate calibration methods in near infrared spectroscopic analysis," *Analytical Methods*, vol. 2, no. 11, pp. 1662–1666, 2010.

[62] S. Wold, H. Martens, and H. Wold, "The multivariate calibration problem in chemistry solved by the PLS method," in *Lecture Notes in Mathematics*, A. Ruhe and B. Kagstrom, Eds., Springer Verlag, Heidelberg, Germany, 1982.

[63] C. 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.

[64] P. H. Fidencio, R. J. Poppi, and J. C. de Andrade, "Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy," *Analytica Chimica Acta*, vol. 453, no. 1, pp. 125–134, 2002.

[65] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees*, CRC Press, Belmont, CA, USA, 1984.

[66] D. Summers, M. Lewis, B. Ostendorf, and D. Chittleborough, "Visible near-infrared reflectance spectroscopy as

a predictive indicator of soil properties," *Ecological Indicators*, vol. 11, no. 1, pp. 123–131, 2011.

[67] Y. Ostovari, S. Ghorbani-Dashtaki, H. Bahrami et al., "Towards prediction of soil erodibility, SOM and CaCO₃ using laboratory Vis-NIR spectra: a case study in a semi-arid region of Iran," *Geoderma*, vol. 314, pp. 102–112, 2018.

[68] C. Gomez, P. Lagacherie, and G. Coulouma, "Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data," *Geoderma*, vol. 189-190, pp. 176–185, 2012.

[69] M. Nouri, C. Gomez, N. Gorretta, and J. M. Roger, "Clay content mapping from airborne hyperspectral Vis-NIR data by transferring a laboratory regression model," *Geoderma*, vol. 298, pp. 54–66, 2017.

[70] A. M. Mouazen, B. Kuang, J. De Baerdemaeker, and H. Ramon, "Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy," *Geoderma*, vol. 158, no. 1-2, pp. 23–31, 2010.

[71] V. L. Mulder, M. Plötze, S. de Bruin et al., "Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1–2.4 µm) and regression tree analysis," *Geoderma*, vol. 207-208, pp. 279–290, 2013.

[72] G. M. Vasques, J. A. M. Demattê, R. A. Viscarra Rossel, L. Ramírez-López, and F. S. Terra, "Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths," *Geoderma*, vol. 223–225, pp. 73–78, 2014.

[73] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, USA, 2013.

[74] C. Gomez, Y. Le Bissonnais, M. Annabi, H. Bahri, and D. Raclot, "Laboratory vis–NIR spectroscopy as an alternative method for estimating the soil aggregate stability indexes of Mediterranean soils," *Geoderma*, vol. 209-210, pp. 86–97, 2013.

[75] R. A. Viscarra Rossel, "ParLeS: software for chemometric analysis of spectroscopic data," *Chemometrics and Intelligent Laboratory Systems*, vol. 90, no. 1, pp. 72–83, 2008.

[76] H. Akaike, "Information theory as an extension of the maximum likelihood Principle," in *Proceedings of the Second International Symposium on Information Theory*, Akadémia Kiadó, Budapest, Hungary, 1973.

[77] I. Chong and C. 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.

[78] 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.

[79] Z. Shi, W. Ji, R. A. Viscarra Rossel, S. Chen, and Y. Zhou, "Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library," *European Journal of Soil Science*, vol. 66, no. 4, pp. 679–687, 2015.

[80] N. Bao, L. Wu, B. Ye, K. Yang, and W. Zhou, "Assessing soil organic matter of reclaimed soil from a large surface coal mine using a field spectroradiometer in laboratory," *Geoderma*, vol. 288, pp. 47–55, 2017.

[81] T. A. Carrino, A. P. Crosta, C. L. Bemfica Toledo, and A. M. Silva, "Unveiling the hydrothermal mineralogy of the Chapi Chiara gold prospect, Peru, through reflectance spectroscopy, geochemical and petrographic data," *Ore Geology Reviews*, vol. 67, pp. 299–315, 2015.

[82] A. Morellos, X. Pantazi, D. Moshou et al., "Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy," *Biosystems Engineering*, vol. 152, pp. 104–116, 2016.

[83] W. Ji, S. Zhou, H. Jingyi, and L. Shuo, "In situ measurement of some soil properties in paddy soil using visible and near-infrared spectroscopy," *PLoS One*, vol. 9, no. 8, article e0159785, 2014.

[84] C. Hermansen, M. Knadel, P. Moldrup, M. H. Greve, R. Gislum, and L. W. de Jonge, "Visible–near-infrared spectroscopy can predict the clay/organic carbon and mineral fines/organic carbon ratios," *Soil Science Society of America Journal*, vol. 80, no. 6, p. 1486, 2016.

[85] 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.

[86] R. J. Murphy, S. Schneider, and S. T. Monteiro, "Consistency of measurements of wavelength position from hyperspectral imagery: use of the ferric iron crystal field absorption at ∼900 nm as an indicator of mineralogy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2843–2857, 2014.

[87] S. Xu, Y. Zhao, M. Wang, and X. Shi, "Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy," *Geoderma*, vol. 310, pp. 29–43, 2018.

[88] J. Chen, J. Chen, Q. Wang, Y. Zhang, H. Ding, and Z. Huang, "Retrieval of soil dispersion using hyperspectral remote sensing," *Journal of the Indian Society of Remote Sensing*, vol. 44, no. 4, pp. 563–572, 2016.

[89] A. Volkan Bilgili, H. M. van Es, F. Akbas, A. Durak, and W. D. Hively, "Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey," *Journal of Arid Environments*, vol. 74, no. 2, pp. 229–238, 2010.

[90] A. Raj, S. Chakraborty, B. M. Duda et al., "Soil mapping via diffuse reflectance spectroscopy based on variable indicators: an ordered predictor selection approach," *Geoderma*, vol. 314, pp. 146–159, 2018.

[91] B. M. Duda, D. C. Weindorf, S. Chakraborty et al., "Soil characterization across catenas via advanced proximal sensors," *Geoderma*, vol. 298, pp. 78–91, 2017.

[92] M. C. Sarathjith, B. S. Das, S. P. Wani, and K. L. Sahrawat, "Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils," *Geoderma*, vol. 267, pp. 1–9, 2016.

[93] M. Conforti, R. Froio, G. Matteucci, and G. Buttafuoco, "Visible and near infrared spectroscopy for predicting texture in forest soil: an application in southern Italy," *iForest-Biogeosciences and Forestry*, vol. 8, no. 3, pp. 339–347, 2015.

[94] N. J. Robinson and M. Kitching, "The 3D distribution of phyllosilicate clay minerals in western Victoria," *Geoderma*, vol. 284, pp. 152–177, 2016.

[95] R. A. Viscarra Rossel, "Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra," *Journal of Geophysical Research*, vol. 116, no. F4, 2011.

[96] A. Goetz, S. Chabrillat, and Z. Lu, "Field reflectance spectrometry for detection of swelling clays at construction sites," *Field Analytical Chemistry and Technology*, vol. 5, no. 3, pp. 143–155, 2001.

[97] R. Vašát, R. Kodešová, L. Borůvka, A. Klement, O. Jakšík, and A. Gholizadeh, "Consideration of peak parameters

derived from continuum-removed spectra to predict extractable nutrients in soils with visible and near-infrared diffuse reflectance spectroscopy (VNIR-DRS)," *Geoderma*, vol. 232-234, pp. 208–218, 2014.

[98] A. Bayer, M. Bachmann, A. Müller, and H. Kaufmann, "A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African ecosystem," *Applied and Environmental Soil Science*, vol. 2012, Article ID 971252, 20 pages, 2012.

[99] L. A. Camargo, J. Marques Junior, V. Barron, L. R. Ferracciu Alleoni, R. S. Barbosa, and G. T. Pereira, "Mapping of clay, iron oxide and adsorbed phosphate in Oxisols using diffuse reflectance spectroscopy," *Geoderma*, vol. 251, pp. 124–132, 2015.

[100] F. Khayamim, J. Wetterlind, H. Khademi, A. H. J. Robertson, A. F. Cano, and B. Stenberg, "Using visible and near infrared spectroscopy to estimate carbonates and gypsum in soils in Arid and Subhumid regions of Isfahan, Iran," *Journal of Near Infrared Spectroscopy*, vol. 23, no. 3, pp. 155–165, 2015.

[101] Q. Fang, H. Hong, L. Zhao, F. Cheng, K. Yin, and C. Wang, "Climatic implication of authigenic minerals formed during pedogenic weathering processes," *Earth Science*, vol. 43, no. 3, pp. 753–769, 2018.

[102] H. Hong, Q. Fang, C. Wang, L. Zhao, and N. Gong, "Constraints of parent magma on altered clay minerals: a case study on the ashes near the Permian-Triassic Boundary in Xinmin section, Guizhou Province," *Earth Science*, vol. 42, no. 2, pp. 161–172, 2017.

[103] R. Viscarra Rossel, "The Soil Spectroscopy Group and the development of a global soil spectral library," *NIR News*, vol. 20, no. 4, pp. 14-15, 2009.

[104] Z. Shi, Q. Wang, J. Peng et al., "Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations," *Science China Earth 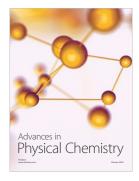Sciences*, vol. 57, no. 7, pp. 1671–1680, 2014.