Article

Determination of Mehlich 3 Extractable Elements with Visible and Near Infrared Spectroscopy in a Mountainous Agricultural Land, the Caucasus Mountains

Elton Mammadov [1,*] , Michael Denk [2] , Frank Riedel [2], Cezary Kaźmierowski [3] , Karolina Lewinska [3] , Remigiusz Łukowiak [4], Witold Grzebisz [4], Amrakh I. Mamedov [5] and Cornelia Glaesser [2]

1   Institute of Soil Science and Agrochemistry, Azerbaijan National Academy of Sciences, 5 M. Rahim, Baku 1073, Azerbaijan
2   Institute of Geosciences and Geography, Martin Luther University Halle-Wittenberg, Von-Seckendorff Platz 4, 06120 Halle (Saale), Germany; michael.denk@geo.uni-halle.de (M.D.); frank.riedel@geo.uni-halle.de (F.R.); cornelia.glaesser@geo.uni-halle.de (C.G.)
3   Department of Environmental Remote Sensing and Soil Science, Adam Mickiewicz University in Poznan, Krygowskiego 10, 61-680 Poznan, Poland; cezark@amu.edu.pl (C.K.); karolina.lewinska@amu.edu.pl (K.L.)
4   Department of Agricultural Chemistry and Environmental Biogeochemistry, Nature University of Poznan, Wojska Polskiego 38/42, 60-625 Poznan, Poland; lukowiak@up.poznan.pl (R.Ł.); witegr@up.poznan.pl (W.G.)
5   Faculty of Agriculture, ALRC, Tottori University, 1390 Hamasaka, Tottori 680-0001, Japan; amrakh03@yahoo.com
*   Correspondence: elton.eldaroglu@gmail.com

Abstract: Soil spectroscopy is a promising alternative to evaluate and monitor soil and water quality, particularly in mountainous agricultural lands characterized by intense degradation and limited soil tests reports; a few studies have evaluated the feasibility of VIS-NIR spectroscopy to predict Mehlich 3 (M3) extractable nutrients. This study aimed to (i) examine the potential of VIS-NIR spectroscopy in combination with partial least squares regression to predict M3-extractable elements (Ca, K, Mg, P, Fe, Cd, Cu, Mn, Pb, and Zn) and basic soil properties (clay, silt, sand, $CaCO_3$, pH, and soil organic carbon-SOC), (ii) find optimal pre-processing techniques, and (iii) determine primary prediction mechanisms for spectrally featureless soil properties. Topsoil samples were collected from a representative area (114 samples from 525 ha) located in the mountainous region of NW Azerbaijan. A series of pre-processing steps and transformations were applied to the spectral data, and the models were calibrated and evaluated based on the coefficient of determination (R^2), root mean square error (RMSE), and the residual prediction deviation (RPD). The leave-one-out cross-validated predictions showed that the first derivative spectra produce higher prediction accuracies (R^2 = 0.51–0.91; RPD = 1.20–2.29) for most soil properties. The evaluation of the model performance with optimal pre-processing techniques revealed that both calibration and validation models produce considerable differences in RPD values associated with sample size and the random partition of the calibration or validation subsets. The prediction models were excellent or very good (RPD > 2.0) for $CaCO_3$, SOC, sand, silt, Ca, and Pb, good or fair (1.4 < RPD < 2.0) for clay, K, Cd, pH, Fe, Mn, and Cu, and poor (1.0 < RPD < 1.4) for Mg, P, and Zn. Principal component and correlation, stepwise regression analysis, and variable importance in projection procedures allowed to elucidate the underlying prediction mechanisms. Unlike the previous studies, the spectral estimations of pH, Ca, Mg, P, Fe, Pb, and Cd concentrations were linked to their correlation with $CaCO_3$ rather than soil organic matter, whereas Mg and P concentrations were also connected to Fe-oxides. Soil particle sizes contributed to predicting K concentration but confounded the prediction of P and Zn concentration. The weaker correlations of Mn, Cu or Zn with $CaCO_3$, particle sizes, SOC, Fe, and spectral data yielded to their lower prediction accuracy. The major prediction mechanisms for M3-extractable elements relied on their relations with $CaCO_3$, pH, clay content and mineralogy, and exchangeable cations in the context of their association with land use. The results can be used in mountain lands to evaluate and control the effect of management on soil quality indices and land degradation neutrality. Further studies are needed to develop most advantageous sampling schemes and modeling.

1. Introduction

Land-use intensification associated with land-use change and conventional agriculture significantly affects soil properties, quality, and ecosystem services, and hence comprises important aspects for future sustainable land management [1]. Therefore, monitoring soil properties in response to agricultural practices, land use, mining, and climate change is critical, especially in mountainous regions, where the soils are prone to severe land degradation processes, such as on-site and off-site impact, erosion, leaching clay particles and nutrients, and deterioration of water quality by chemicals [2]. Recent developments of site-specific crop management and precision agriculture have significantly heightened the need for more time- and cost-efficient techniques and methods of soil analyses [3]. Spectroscopic techniques are considered fast, cost-effective, and non-hazardous, and multiple soil characteristics can be estimated from the same spectra [4]. The number of studies examining the potential of applications of reflectance spectroscopy in agriculture is steadily increasing all over the world [4]. Many of these investigations have evaluated individual wavelength ranges separately, while others have combined several ranges of wavelength to estimate soil properties and quality indices [1,5–7]. Among these techniques, laboratory-based visible and near infrared spectroscopy (VIS-NIR or NIR) has been reported as an alternative to traditional soil analytical methods, which are often time consuming and costly, providing limited information on the soil spatial variability [4,8–10].

Partial least squares regression (PLSR) is one of the most commonly used methods to establish relations between soil properties and spectral data, as reported by review studies [11,12]. The robustness of the prediction models depends not only on data pre-processing and proper calibration methods, but also on soil properties, relationships between them, and the number of samples used for the calibration and validation [13–15]. Many studies have focused on the potential of VIS-NIR spectroscopy to predict a wide range of soil physical, chemical, and biological properties in different climatic regions of the world [4,16,17]. Along with basic soil properties, different forms of nutrients and heavy metals in soils have been predicted with VIS-NIR spectra in varying scales [6,7,18,19]. Compared to basic soil properties (pH, texture, soil organic carbon (SOC), exchangeable cations, electrical conductivity (EC), available N and P, microbial biomass, and respiration), fewer studies have focused on the prediction of Mehlich 3 (M3) extractable chemicals, e.g., microelements and metals using VIS-NIR reflectance spectra.

The M3-extraction procedure [20] is a simple, rapid, and universal soil chemical analysis method used worldwide [21,22]. With its extraction power on the border between weak (e.g., $CaCl_2$) and strong (e.g., *aqua regia*) ranges, the M3 method enables the extraction of a wide spectrum of elements, such as macro (e.g., Ca, Mg, K, and P), micro (e.g., Cu, Fe, Mn, Ni, and Zn), and toxic elements (e.g., As, Cd, Cr, and Pb) [20]. Moreover, M3 extraction, which is related to the available or total macro- and micronutrients, heavy metal contents, and soil exchangeable cations, could have great potential to be used for massive analysis of soils and water samples from mountain agricultural areas with a heterogenic landscape, site-specific farming, and specific soil systems and properties for application of advanced approaches across various scales (e.g., GIS, remote sensing, precision agriculture) to control soil quality and land degradation neutrality [23].

As a reference method, M3-extractable elements have been calibrated with VIS-NIR reflectance spectra in a few studies [18,24–26], whereas other works have focused on predicting M3-extractable elements using NIR spectroscopy [24,27,28]. Abdi et al. [24] applied modified PLSR and 40 spectral pretreatments, both separately and combined, to improve VIS and NIR calibration models for predicting M3-extractable elements as well as other basic soil properties. They concluded that the prediction accuracy for P, Fe, and Ca with

NIR was better than that based on VIS-NIR spectra, whereas other soil parameters were predicted more accurately with the whole VIS-NIR spectrum [24]. Hively et al. [25] attempted to predict different soil properties including M3-extractable elements in flat and well-drained local cultivated fields using airborne imaging spectrometer data (400–2450 nm) combined with PLSR. Despite the fact that the spectral data were acquired via an air-borne manner, 15 of 19 soil properties were predicted with an $R^2 > 0.50$ [25]. Vašát et al. [26] focused on enhancing continuum-removed spectra of VIS-NIR reflectance data by using all spectral features and their parameters and then relating these parameters to M3-extractable nutrients by employing a multiple linear regression technique. The authors compared their results with PLSR outcomes that were achieved without using pre-processing techniques and found notable differences between them. Therefore, the capability of PLSR combined with un-processed spectra was not thoroughly apparent [26]. Luce et al. [18] assessed the potential of VIS-NIR spectroscopy to predict total, M3- and DTPA-extractable, and water-soluble heavy metal concentrations in different agricultural soils treated with paper mill biosolids and byproduct lime. The study concluded that VIS-NIR spectroscopy has the capacity to predict total M3 and DTPA nutrients and, to a lesser extent, water-soluble soil heavy metals, especially when site-specific samples were used. The validation results were good to excellent ($R^2 > 0.70$, RPD > 1.70) for M3-extracted Cu, Cd, and Ni and weaker ($R^2 < 0.70$, RPD < 1.70) for Zn [18].

Besides the VIS-NIR prediction capabilities, several studies have focused on the prediction mechanisms of a series of underlying soil properties including plant-available macro and micronutrients [18,19,29]. In this regards, principal component analysis (PCA) is an effective approach to elucidate complex interrelationships among soil properties and, hereafter, the relations between soil properties and spectra [19,30]. Cheng et al. [19] successfully employed PCA biplots, as well as correlation and partial correlation analysis, to clarify mechanisms behind the estimation of heavy metal concentrations from VIS-NIR spectra. Thus, particular emphasis needs to be given to spectrally featureless soil properties, especially M3-extractable elements, since much less is known about their prediction mechanisms. Some basic soil properties are contributors, and others are confounders for the prediction of spectrally featureless soil characteristics. In this context, we propose VIS-NIR spectroscopy to be a supplementary tool for estimating M3-extractable elements and the assessment of soil properties in the mountainous agricultural lands, including the Lesser Caucasus Mountains, where high soil and land-use variability and degradation is typical.

Generally, more attention has been paid to arable soils, especially in lowland areas, whereas soils in mountainous regions have received less consideration in view of spectroscopy approaches [12]. More than 50% of Azerbaijan territory is mountainous, which necessitates the agricultural importance of the mountainous lands, including the Lesser Caucasus Mountains. In recent decades, this fragile mountain ecosystem experienced serious land degradation, soil quality reduction (e.g., soil macro- and micro-nutrients, SOC, clay particles, pH, and Ca), and damaged water resources by sediment and nutrients or chemicals through flooding, excessive runoff, and water erosion. These environmental problems are associated with both natural and anthropogenic factors such as changing climate conditions, heterogeneous topography, and improper land management without conservation measures, including intensive cultivation and land-use change from grassland to cropland, uncontrolled livestock grazing, and deficiency of ground cover. Improper cultivation is considered one of the main causes of land degradation in the Caucasus Mountains, and small-scale farming (0.2 to 3 ha) due to land tenure ownership is challenged with limited methodologies and knowledge of soil surveying and controlling soil quality via sustainable soil management and conservation, i.e., increasing SOC, minimizing soil disturbance, using crop rotation, and cover crop. The recently initiated implementation of sustainable land management practices for maintaining soil fertility and reducing land degradation neutrality within the Global Soil Partnership (i) targets landscape restoration and reverse land degradation via monitoring and evaluation, and (ii) requires updated land use and soil quality information, which is solely based on very limited, outdated

data collected in the 1980s [31]. Therefore, monitoring the dynamics of land productivity, soil quality, and fertility associated with agricultural management practices and land-use change in the instable system of the Lesser Caucasus Mountains requires the application of both soil survey and spectroscopic prediction and hence future upscaling land degradation neutrality.

Only a very limited number of reports are available from Azerbaijan [32–34] and other mountainous regions of the world on the status of soil fertility, including the spectroscopic prediction of M3 nutrients for soil quality monitoring purposes; in this sense, this study is a pioneering study. Therefore, the objectives of this study are (i) to evaluate the feasibility of VIS-NIR spectroscopy and PLSR modelling to predict M3-extractable elements (Ca, Mg, Cd, Cu, Fe, K, P, Pb, Mn, and Zn) in a representative test area of the Caucasus Mountains, (ii) to analyze the input of different pre-processing techniques, and (iii) to study the prediction mechanism of underlying soil characteristics. Thus, the spectrally active basic soil properties, namely $CaCO_3$, particle sizes, SOC, and pH, that affect nutrients solubility and mobility were involved in the VIS-NIR estimations. Unlike the past reports, this study focused on the VIS-NIR spectroscopic determination of M3 elements and basic soil properties in a comparatively large mountainous agricultural area (525 ha) where small-scale farming and different land-use types are typical.

2. Materials and Methods

2.1. Study Area and Soil Sampling

The study was performed in the foothill belt in the Lesser Caucasus Mountains in Azerbaijan (Figure 1). The study area (525 hectares) is located between 40.8401° N and 40.8701° N latitudes, 45.6201° E and 45.6401° E longitudes, and its elevation ranges from 745 to 1100 m. The typical climate is temperate (Mediterranean climate) with mild–dry winters. The annual precipitation and average air temperature are 570 mm and 10.2 °C, respectively. According to the World Reference Base for Soil Classification [35], the accommodating soil group is Kastanozem, and the soil moisture regime is xeric. The study area is complex in geological setting with a high diversity of rock types.
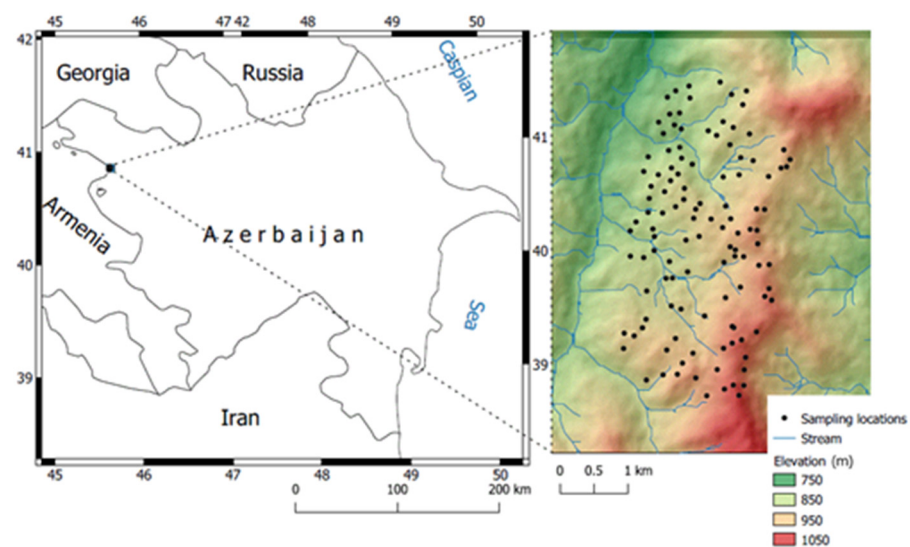
Figure 1. Geographical location of the study area (Digital Elevation Model data (USGS SRTM 1-ArcSecond Global DEM) was downloaded from www.earthexplorer.usgs.gov, accessed on 11 July 2018).

The morphology is typically characterized by gentle slopes (~15°) with tonalites and quartz-diorites of the upper Jurassic lower Cretaceous. The typical parent material is sedimentary rocks and carbonate-rich deluvial sediments, especially in the footslopes, where slopes are gentle. Soil erosion is a typical phenomenon in the study area associated with natural conditions, slope inclination, and land-use intensity. Therefore, severely

eroded soils and outcrops of bedrocks occur in the south-, south-east-faced, and steep slopes [31,33]. Agricultural land-use types for the study area are pastures, hayfields, cereals, and potato planting. Although local shrubby lands are typical, those areas are often under agricultural use as permanent pastures. The soil properties are highly variable and related to variations in topography, geological settings, and land-use history [36].

A soil survey (114 samples: ~1 sample per 5 hectare) was conducted with consideration of the terrain heterogeneity of the investigated foothill area to collect samples from areas representing the different land-use types [21]. The sampling locations were randomly (irregularly) designated and cover variations in land use, topography, geological substrate, and the erosive state of the soil system. A Garmin GPS Map 62s was used to determine geographical coordinates of the sampling locations. The soil samples (each of ~2 kg) were collected from the topsoil horizon (0–15 cm) and air-dried, milled, and sieved through a 2 mm sieve in the laboratory. The samples were divided into two parts for spectral measurements and chemical analyses.

2.2. Chemical Analyses and Descriptive Statistics

The particle sizes (sand, silt, and clay) were determined by the hydrometer method [37], calcium carbonate ($CaCO_3$) content using a pressure-calcimeter [38], and soil organic carbon (SOC) content using the Walkley–Black method [39]; soil pH was measured in deionized water with a soil solution ratio of 1:1. Since the laboratory determination of clay mineralogy is time-consuming and expensive, the cation exchange capacity (CEC) to clay ratio (clay activity in soil taxonomy) of the soils was used to characterize the clay mineralogy and the CEC/Clay ratio was also well-predicted ($R^2 = 0.86$) from mid-infrared diffuse reflectance spectra [40]. The elements Ca, Cd, Cu, Fe, K, Mg, Mn, P, Pb, and Zn were digested using the Mehlich 3 procedure [20] and then measured by Atomic Absorption Spectroscopy (Thermo Scientific ICE 3000 Series, Pittsburgh, PA, USA). The elements K, Mg, and Ca were extracted by Mehlich III and Ammonium Acetate and measured with Inductively Coupled Plasma emission spectroscopy for validation purposes and CEC calculation as a sum of cations.

The detection limits were calculated as three times the standard deviation of 20 replicate measurements of the blank. The used standard solution concentrations were (i) 0, 5, 10, 25, 50, and 100 mg dm^{-3} for K, Ca, Mg, and P, (ii) 0.0, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, and 10.0 mg dm^{-3} for Zn, Cu, Mn, Pb, and Cd, and (iii) 0.0, 0.5, 1.0, 2.5, 5.0, 10.0, 25.0, 50.0, and 100.0 mg dm^{-3} for Fe. The used standard solutions (K, Mg, and Ca: AVS—Titrinorm 1000 mg dm^{-3} $\pm$ 0.2% VWR Chemicals; $PO_4{}^3$: 10 g dm^{-3} Fixanal, Fluka analytical; Zn, Cu, and Mn: 1000 mg dm^{-3} $\pm$ 4 mg dm^{-3} Fixanal, Fluka analytical; Fe: 999 mg/dm^3 $\pm$ 4 mg dm^{-3} Fixanal, Fluka analytical; Cd and Pb: 1 g dm^{-3} $\pm$ 5 mg dm^{-3} Fixanal, Regional Office of Measures and Control, department of analytical chemistry) and the solution concentration range were related to the routine method used by the soil testing laboratories.

The results of the chemical analyses were statistically analyzed to characterize the data nature and spatial distribution and variability of the soil parameters. The descriptive statistics included the calculation of the mean, maximum, minimum, median, standard deviation (SD), and the coefficient of variation (CV). The CV values were considered [41] small (<20%), moderate (20–50%), or large (>50%). Since no reference data showed a normal distribution, Spearman's Rho correlation coefficients (*r*) were used to characterize the relationships between the studied soil properties. The relationships were considered very weak, weak, moderate, strong, and very strong when the *r* values varied between 0.00 ± 0.19, 0.20 ± 0.39, 0.40 ± 0.59, 0.60 ± 0.79, and 0.80 ± 1.00, respectively [42]. The statistical analysis was computed in an *R* software environment [43]. The analysis of variance (ANOVA) was conducted to assess the effects of land use on soil properties, where mean comparisons were performed using the Tukey–Kramer HSD test at $p < 0.05$. A stepwise regression analysis was used to examine the effects of basic soil properties on M3 elements.

2.3. VIS-NIR Reflectance Measurements and Pre-Processing of Spectroscopic Data

For the spectral reflectance measurement, samples were placed inside circular transparent Petri dishes (10 cm in diameter and 1.5 cm in depth), and the sample surface was flattened to create a smooth surface. An ASD FieldSpecPro FR (Analytical Spectral Devices Inc., Boulder, CO, USA) spectroradiometer was used to measure soil reflectance in a dark room. Its spectral resolution was 3 nm in the 350–1000 nm range and 10 nm in the 1000–2500 nm range. The sampling interval was 1.4 nm in the 350–1000 nm range and 2 nm in the 1000–2500 nm range. The sample was illuminated using an ASD Pro Lamp equipped with a 50 W halogen lamp under a zenith angle of 45° located at 31 cm from the soil sample. The bare fiberoptic of the spectroradiometer was placed in nadir position, 11 cm above the sample surface, allowing the circular measurement area of approximately 4 cm in diameter. The measured raw spectral data were divided by the reflectance values of a white reference panel (Spectralon®, 99% reflectivity) measured under the same illumination conditions to obtain relative reflectance spectra of the samples. Each sample was measured from four positions, by rotating the Petri dish 90° each time. This approach served to minimize possible spectral variations due to the surface geometry of the samples. Each single measurement was based on internally averaged 50 spectra. Afterwards, the 12 spectra (3 spectra per 90° rotation) of each sample were exported from binary to ASCII format using ViewSpecPro (Analytical Spectral Devices Inc., Boulder, CO, USA). The acquired 12 spectra were separately involved in splice-correction and then averaged. Finally, the obtained spectrum for each sample was involved in spectral pre-processing techniques.

Pre-processing methods are used to remove signal noise, accentuate features, and to improve mineral identification and the accuracy of the prediction [44]. For this reason, different methods, such as Savitzky–Golay smoothing [45], first and second derivatives [46], continuum removal [46], multiplicative scatter correction (MSC) [47], standard normal variate [48], and de-trending and absorbance (Abs) transformations [48] were applied to the averaged reflectance spectra. The selected pre-processing techniques were based on our previous study [33] that tested the feasibility of smoothing and derivatives with different moving window sizes (mainly 7, 9 and 11), as well as derivatives with different gaps and segment sizes. The pre-processing techniques were performed using the "prospectr" package [49] in the R software environment [43].

2.4. PLSR Model Calibration and Validation

Relations between the spectral data and the tested soil properties were studied using PLSR modelling [50,51]. The PLSR technique is a standard bilinear calibration method that uses data compression by reducing the large number of measured collinear spectral variables to fewer non-correlated latent variables or factors. The compressed latent variables result from the method similar to principal components analysis of measured spectral data. These factors represent the relevant structural information of the data and are utilized to predict the dependent variables. The PLSR algorithm is not described in this study, whereas further information can be found in numerous studies [50].

For the calibration of regressions models, a two-staged approach was used. First, optimal pre-processing methods for each soil property were determined. For this reason, PLSR models were calibrated for all pre-processing methods using all samples. Leave-one-out cross-validation [10] was applied based on the highest coefficient of determination (R^2) and lowest root mean square error (RMSE) of the cross-validated predictions. The number of factors (NF) to be used in the models was selected based on the lowest RMSE [52] and the Akaike Information Criterion (AIC) [53]. In the second stage, PLSR models were calibrated using 88 samples (75%) and the optimal pre-processing methods found in the first stage and were validated with the remaining independent 26 sample (25%) subset. The procedure served to test the performance of the models with a set of independent validation samples. The validation and calibration subsets were randomly selected in the R software environment [43].

The evaluation of the predictive performance of the models was based on R^2, RMSE, and the residual prediction deviation (RPD) [54]. A six level interpretations of RPD [10], values were adopted: RPD < 1.0 indicates very poor models/predictions, and their use is not recommended; 1.0 < RPD < 1.4: poor models/predictions, where only high and low values are distinguishable; 1.4 < RPD < 1.8: fair models/predictions, which may be used for assessment and correlation; 1.8 < RPD < 2.0: good models/predictions, where quantitative predictions are possible; 2.0 < RPD < 2.5: very good, quantitative models/predictions; and RPD > 2.5: excellent models/predictions.

2.5. Determination of Important Wavelengths and Prediction Mechanisms

Statistical analysis is a commonly applied approach to evaluate the prediction mechanism of spectrally featureless soil properties from VIS-NIR spectra. This approach involves internal correlation between spectrally active and passive soil properties to analyze the underlying mechanisms [29,30]. Another widely used approach is the analysis of important wavelengths through the PLSR model. In this study, both methods were employed. To determine the important wavelengths used in the model calibrations, the Variable Importance in the Projection (VIP) approach was used [51,55,56]. For an observed variable y, VIP was calculated by:

$$VIPk(a) = K \sum_a W^2 ak \left(\frac{SSYa}{SSYt} \right)$$

where VIP is the importance of the Kth predictor variable based on a model with a factors, W_{ak} is the corresponding loading weight of the Kth variable in the ath PLSR factor, SSY_a is the explained sum of squares of Y by a PLSR model with a factors, SSY_t is the total sum of squares of Y, and K is the total number of predictor variables. The wavelength is considered important [51] if the VIP > 1, which is used as a threshold level [55,57].

The correlations between soil properties and reflectance values of the reflectance spectra were computed for each wavelength and plotted across the whole wavelength range. Furthermore, PCA and partial correlation analysis were used to evaluate (i) the relations between soil properties and the spectral data, and (ii) the contribution of basic soil properties to the prediction of M3-extractable elements according to Cheng et al. [19]. The first two principal components of the PCA (PC1 and PC2) were selected as the factors representing the spectral data. Afterwards, the PC1 and PC2 were combined with the soil properties, and PCA was used to plot soil properties and spectral data against principal factors (F1 and F2) in the two-dimensional space. Then, Spearman's Rho correlation coefficients were computed between soil properties and spectral data (PC1) by controlling each basic soil property separately [19]. The statistical analyses, as well PLSR modeling and VIP scores, were conducted using the "pls" [58] and "plsVarSel" [59] packages in R software environment [43].

## 3. Results

3.1. Reference Data

The descriptive statistics of the soil properties are presented in Table 1. The soil reaction (pH in H_2O) varies from moderately acid to slightly alkaline. The soils are predominantly silty clay and silty clay loam, and partially clay loam in texture [21]. The sand content is largely variable (~60%) as opposed to the silt fraction, the predominating particle size in respect to its mean. The clay and SOC content vary considerably (~35%) with a mean value of 21% and 3.41%, respectively. The $CaCO_3$ content is largely variable (~84%), with a mean value of 2.91%. Such variability of SOC, clay, and $CaCO_3$ is attributed to the distribution of parent material depending on the topography, elevation (e.g., change in temperature, precipitation, pH by elevation), and dissolution and leaching processes [2,60]. Among the M3 elements, Ca, K, and P were largely variable (~50%), and others varied moderately (20–40%). The mean concentration of M3 nutrients (mg kg^{-1}) followed the sequence of Ca > Mg > K > Mn > Fe > P > Zn > Pb > Cu > Cd (Table 1).

Table 1. Descriptive statistics of the tested soil properties.

Soil Property	Min	Max	Mean	Median	SD	CV (%)
Ca (mg kg^{-1})	133.07	1674.77	726.14	658.89	388.95	53.7
Cd (mg kg^{-1})	0.191	0.465	0.32	0.334	0.064	20.0
Cu (mg kg^{-1})	0.67	3.08	1.141	1.015	0.405	35.5
Fe (mg kg^{-1})	19.8	116.16	58.85	55.72	23.45	39.8
K (mg kg^{-1})	74.2	806.27	262.66	225.29	127.73	47.1
Mg (mg kg^{-1})	190.82	615.66	446.6	438.46	92.92	20.8
Mn (mg kg^{-1})	24.69	145.22	109.97	115.87	23.72	21.6
P (mg kg^{-1})	2.41	20.34	5.74	4.8	2.98	51.9
Pb (mg kg^{-1})	1.32	4.59	3.17	3.4	0.86	27.1
Zn (mg kg^{-1})	1.98	7.74	4.59	4.62	0.9	20.0
Sand (%)	9	78	29	25	17	59.6
Silt (%)	21	77	50	54	11	22.0
Clay (%)	2	38	21	23	8	34.8
SOC (%)	1.34	6.85	3.41	3.08	1.15	34.8
CaCO$_3$ (%)	0.00	8.81	2.91	2.16	2.41	83.6
pH	5.75	7.82	7.25	7.47	0.46	6.3

Min—minimum, Max—maximum, SD—standard deviation, CV—coefficient of variation.

The Spearman's Rho correlation coefficients among the soil properties were computed and sorted according to the first principal components in the correlation matrix for better visualization purpose (Figure 2). Significant relations between the basic soil properties, and basic soil properties and M3-extractable elements were noted.
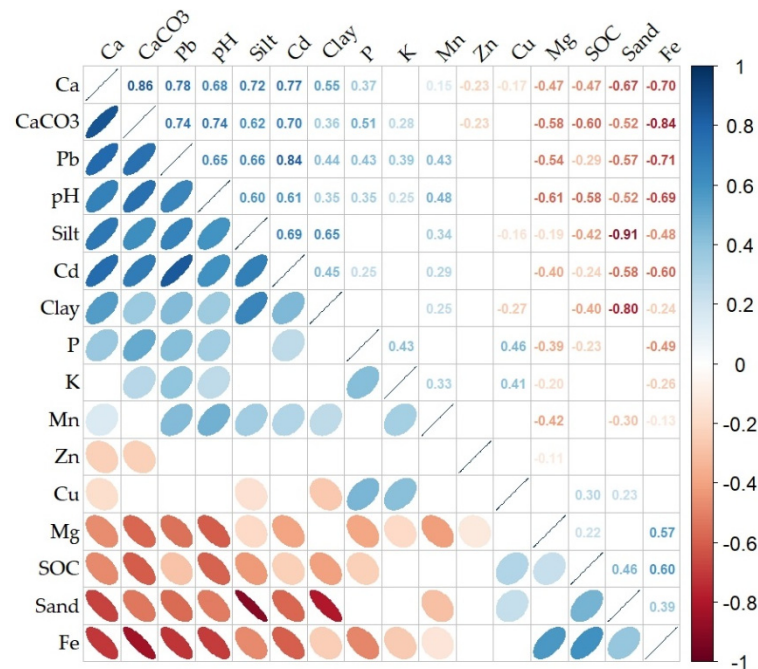
Figure 2. Spearman's Rho correlation coefficients between tested soil constituents (blank cells stand for insignificant correlations at the level of $p \geq 0.05$).

The clay content was negatively correlated ($r = 0.80$ *) with the sand and positively correlated ($r = 0.65$ *) with the silt content. A weak positive correlation was found between the clay content and CaCO$_3$ or pH, whereas the correlation between the silt content and CaCO$_3$ or pH was strong. However, both clay and silt content were negatively related with SOC. Unlike those, the sand content was positively correlated with SOC and negatively correlated with CaCO$_3$. These relations are attributed to the characteristic distribution of quartz-rich bedrock and carbonate-rich parent material (e.g., clay mineralogy) at higher

and low elevations, respectively. Both the SOC and sand contents increased with elevation, whereas the $CaCO_3$ content decreased. It appears that the SOC content is controlled by soil texture and clay mineralogy and associated soil characteristics (e.g., pH, $CaCO_3$ and Fe) with regard to the land use and elevation, and soil types. Accordingly, the $CaCO_3$ content was found positively correlated ($r = 0.84$ *) with Ca (the main source of Ca) and pH ($r = 0.74$ *), and negatively correlated with SOC and Fe ($r > 0.60$ *).

The M3-extractable elements, such as Ca, Cd, Fe, Mg, and Pb, showed higher and significant correlations with the basic soil properties, whereas others (Cu, K, P, Mn, and Zn) mainly showed no or weaker, yet significant correlations (Figure 2). Cd and Pb showed a strong positive, whereas Fe and Mg showed a negative reliance on pH and $CaCO_3$. As is obvious from the correlation coefficient between them ($r = 0.84$ *), Cd and Pb occur very closely in the soil matrix. Their relations to $CaCO_3$, clay, and pH probably confirm that the underlying calcareous parent material plays an important role as their main source, and that the availability of these elements depends on pH. Furthermore, the inter-correlations (e.g., between Ca and Cd, Ca and Pb, Cd and Pb, and Fe and Mg) show that those elements are similarly bound in the soil matrix.

The results suggested that the M3 elements are related to the variation in soil type or texture and basic soil properties (mostly $CaCO_3$ and pH), along with coexisting metals that can reliably assist as predictive factors [61]. Moreover, the role of $CaCO_3$ (e.g., Ca and electrolyte sources, and pH modifier) on soil biogeochemistry is still overlooked in mountain soils and needs greater attention for the monitoring of land degradation and quality [2,60]. Thus, for the current study, the tested soil properties can be formally categorized into two groups. The first group comprises $CaCO_3$, clay, silt, pH, and their reliant elements (Ca, Cd, Pb, K, P, and Mn), and the second group comprises sand, SOC, Fe, and Mg, whereas Zn and Cu have complex relations due to their weak relations with others.

3.2. Land-Use Effects on Soil Properties

In the last three to four decades, the fragile mountain ecosystems in the Caucasus region experienced serious land degradation without conservation. This has never been evaluated at the level of basic soil properties as well as regarding the plant available macro- and micro-nutrients. We assumed significant effects of different land uses on soil properties causing their regional heterogeneity. Typical land uses include arable land (potato and cereals), pastures (grazing from March to December), hayfields (preserved from March to July, then used as pasture), and locally spread shrubbery areas, which are less exposed to grazing. The land-use effects on soil properties are related to soil type (clay mineralogy) and management intensity, because basic properties and M3-extractable elements are sensitive to soil management and natural disturbances (Table 2), resulting in a strong variation of soil properties between and within land use (Supplementary Figure S1).

Soil pH, $CaCO_3$, silt, clay, K, Ca, P, and CEC showed similar trends by land use (arable > hay > pasture ≥ shrub land). Similar trends were noted for Pb and Cd. For sand, CEC/clay, Mg, and Fe, the trends were opposite. The SOC content is related to the land disturbance, e.g., carbon sequestration potential for different land uses (arable < pasture < hay < shrub land); the trends for Zn and Cu were similar to SOC (Table 2, Supplementary Figure S1). The elevation (pasture > shrub > hay > arable land) is associated with the change in parent material (e.g., CEC/clay), increasing precipitation and decreasing temperature, pH and salinity, and $CaCO_3$ and exchangeable Ca (Ca and electrolyte resources), while promoting the leaching of Ca, clay, and slit fractions and increased clay and silt content in the lower part of the study area. Though soil organic matter and clay can act as buffers to resist pH variations, the leaching of free cations, oxides (e.g., Fe), and $CaCO_3$, as well as the removal of cations (e.g., Ca, Mg and K) by plant uptake or harvesting, and the continuous formation of carbonic acid and application of acidic fertilizers can decrease soil pH, particularly in the higher elevated areas. This affects the soil buffering capacity and CEC, as well as the solubility and transport of M3 elements such as Fe, Mn, Cu, Zn, and P [2].

Table 2. Soil properties as affected by land use or elevation. Within each row, means labelled with the same letter are not significantly different at $p < 0.05$.

Soil Properties	Land Use			
	Arable	Hayfield	Pasture	Shrub
Elevation (m)	872.2 [c]	910.9 [b]	971.5 [a]	953.4 [a]
SOC (%)	2.2 [c]	3.5 [b]	3.4 [b]	4.3 [a]
pH	7.6 [a]	7.4 [a]	7.1 [b]	7.0 [b]
Sand (%)	16.4 [b]	18.6 [b]	38.4 [a]	39.3 [a]
Silt (%)	57.8 [a]	57.1 [a]	43.5 [b]	44.6 [b]
Clay (%)	26.4 [a]	25.3 [a]	17.8 [b]	15.2 [b]
$CaCO_3$ (%)	5.0 [a]	3.7 [b]	1.8 [c]	1.2 [c]
K (mg kg^{-1})	317.8 [a]	281.3 [ab]	259.1 [ab]	214.9 [b]
Mg (mg kg^{-1})	421.4 [a]	439.9 [a]	462.0 [a]	461.0 [a]
Ca (mg kg^{-1})	1058.0 [a]	879.8 [b]	544.3 [c]	426.4 [c]
CEC (meq/100 g)	9.6 [a]	8.8 [a]	7.2 [b]	6.5 [b]
CEC/Clay	0.4 [b]	0.4 [b]	0.5 [ab]	0.5 [a]
Zn (mg kg^{-1})	4.4 [b]	4.9 [ab]	4.4 [b]	5.0 [a]
Cu (mg kg^{-1})	1.0 [a]	1.2 [a]	1.1 [a]	1.2 [a]
Pb (mg kg^{-1})	3.8 [a]	3.6 [a]	2.7 [b]	2.7 [b]
Cd (mg kg^{-1})	0.4 [a]	0.4 [a]	0.3 [b]	0.3 [b]
Mn (mg kg^{-1})	114.2 [ab]	117.3 [a]	105.1 [b]	103.9 [b]
Fe (mg kg^{-1})	39.8 [b]	58.3 [a]	67.4 [a]	69.3 [a]
P (mg kg^{-1})	7.6 [a]	5.5 [b]	5.0 [b]	4.7 [b]

CEC—cation exchange capacity, CEC/clay—clay mineralogy index.

The variation in the land-use trend was also associated with the CEC/clay ratio of the samples (0.35–0.55 (63 samples) > 0.20–0.35 (40 samples) > 0.55–75 (12 samples)), revealing that mixed clay mineralogy (CEC/clay = 0.35–0.55) was the dominant parameter following illite + kaolinite mineralogy (CEC/clay = 0.20–0.35); samples with >0.55 (e.g., mixed clay mineralogy) were related to undisturbed areas with higher SOC or steep slope areas under shrub with the lowest clay content (<6%) as a result of clay loss or leaching. This affected the uncertainty of the relations between basic soil properties and their effect on the variations of M3 elements. The contribution of basic soil properties to M3-extractable elements by stepwise analysis showed various orders of parameters: for most of the M3 elements (Mg, Ca, Mn, Fe, Pb, and Cd), a larger variation (73–95%) could be explained by the association of the SOC, $CaCO_3$, pH, CEC, silt, clay, and CEC/clay, whereas for the M3 elements K, Zn, and P, this variation was ≤30%. The effect of land use was significant in all soil properties with the exception of Cu and Mg, revealing that whereas the intensity of land-use change, agricultural management, and soil conservation was not controlled by the community, specific trends in the soil properties by land use could be observed (Supplementary Table S1).

3.3. Effects of Spectral Pre-Processing Methods on the PLSR Calibration

The results of cross-validated predictions showed that the used pre-processing techniques considerably improved the prediction accuracies compared to the raw spectra. The AIC and RMSE values of the optimal pre-processing methods were consistent in all cases. The optimal pre-processing technique found for each soil property was selected, and their performances based on cross-validation are summarized in Table 3. The first derivative was the optimal pre-processing method for all the tested soil properties (except clay and SOC). For clay and SOC content, multiplicative scatter correction (MSC) and absorbance were found as the optimal pre-processing methods, respectively. Overall, the second derivatives and CR produced the poorest predictions.

Table 3. Statistics of cross-validation, calibration, and validation prediction for each soil constituent based on the optimal pre-processing methods.

Soil Property	Optimal Pre-Processing Method	NF	Cross-Validation (*n* = 114)			Calibration (*n* = 88)			Validation (*n* = 26)		
			RMSE	R^2	RPD	RMSE	R^2	RPD	RMSE	R^2	RPD
$CaCO_3$	D_1-Gap-S_{10}	11	0.81	0.96	2.97	0.83	0.96	3.02	0.99	0.89	2.16
pH	D_1-Gap-S_{10}	4	0.36	0.69	1.44	0.35	0.66	1.47	0.37	0.65	1.49
Clay	MSC	11	3.78	0.84	1.85	4.02	0.85	1.75	3.77	0.83	2.20
Silt	D_1-Gap-S_{10}	4	5.32	0.82	2.07	5.00	0.85	2.19	5.75	0.81	2.02
Sand	D_1-Gap-S_{10}	4	8.18	0.81	2.08	8.21	0.82	2.02	8.35	0.81	2.19
SOC	Absorbance	16	0.45	0.93	2.53	0.54	0.94	2.19	0.43	0.90	2.16
Ca	D_1-Gap-S_{10}	10	180.60	0.91	2.15	179.80	0.94	2.21	165.10	0.86	2.39
Cd	D_1-Gap-S_{10}	6	0.035	0.80	1.81	0.036	0.82	1.79	0.037	0.74	1.88
Cu	D_1-Gap-S_{10}	10	0.275	0.80	1.47	0.327	0.80	1.31	0.180	0.76	1.74
Fe	D_1-Gap-S_{10}	9	14.66	0.82	1.60	14.79	0.83	1.49	16.31	0.76	1.72
K	D_1-Gap-S_{10}	10	72.21	0.85	1.85	78.93	0.85	1.55	65.32	0.83	2.62
Mg	D_1-Gap-S_{10}	8	67.14	0.73	1.39	74.30	0.75	1.27	66.00	0.62	1.36
Mn	D_1-Gap-S_{10}	10	13.72	0.85	1.73	14.77	0.86	1.57	14.25	0.71	1.81
P	D_1-Gap-S_{10}	8	2.19	0.73	1.36	2.26	0.72	1.25	2.41	0.51	1.43
Pb	D_1-Gap-S_{10}	10	0.375	0.91	2.29	0.380	0.93	2.17	0.396	0.84	2.48
Zn	D_1-Gap-S_{10}	6	0.750	0.56	1.20	0.890	0.54	0.99	0.670	0.57	1.43

D_1-Gap-S_{10}—1st derivative with gap segment size of 10 bands; MSC—multiplicative scatter correction; NF—the number of factors.

Based on the cross-validation results, the predictions for $CaCO_3$ and SOC showed excellent models (RPD > 2.5), whereas the prediction of the sand, silt, clay, and pH were characterized with very good (2.0 < RPD < 2.5), good (1.8 < RPD < 2.0), and fair model (1.4 < RPD < 1.8) performance, respectively (Table 3). Among the M3-extractable elements, the Ca and Pb contents were predicted with the highest accuracy, whereas the predictions for Cd and K produced good model results. However, fair predictions were obtained for Cu, Fe, and Mn. As expected from the correlations with the basic soil properties, Mg, P, and Zn were poorly predicted (1.0 < RPD < 1.4). The number of factors (NF) varied depending on the soil property, ranging from 4 (for sand, silt, and pH) to 16 (for SOC).

3.4. PLSR Model Calibration and Validation

PLSR models based on all samples allowed us to determine the optimal pre-processing methods for each soil property. These models were again calibrated and validated with subsets of 75% and 25% of all samples, respectively. In view of RMSE, both calibration and validation models were appropriately similar to those of the cross-validation results. Nevertheless, the RPD values revealed unstable models for some of the predicted soil properties (Table 3). For the calibration subset, excellent predictions were solely obtained for $CaCO_3$; the model validation showed very good predictions (R^2 = 0.89, RPD = 2.16). For SOC, both calibration and validation subsets revealed very good models (R^2 > 0.90, RPD > 2.0). The performances of the calibration and validation subsets for sand and pH were similar to those of the cross-validation setup. The validation subset produced very good results for sand and fair results for pH. Although the validation subset revealed a very good model for clay (RPD = 2.20), the calibration model was fair (RPD = 1.75).

With regard to the models for M3 elements, the calibration and validation models of Ca, Cd, Fe, Mg, and Pb performed similarly. The validation showed very good predictions for Ca and Pb and good predictions for Cd (Table 3). Although Fe and Mg are significantly correlated with basic soil properties (e.g., $CaCO_3$, pH, SOC), the PLSR models of the validation subset produced fair predictions for them. The validation for the M3-extractable elements Cu, K, Mn, P, and Zn showed better performances than the corresponding calibration subset. The calibration models for Cu, P, and Zn were poor, whereas the validation subset produced fair prediction models for the corresponding elements. Likewise, the

calibration models were fair for K and Mn, whereas the validation subset revealed excellent and good models, respectively.

3.5. Prediction Mechanisms and Important Wavelengths

3.5.1. Correlations between Soil Properties and Spectral Reflectance Data

The Pearson correlation coefficients between soil constituents and reflectance values of the smoothed spectra for each wavelength were computed and plotted to visualize the relations between them (Figure 3). Generally, the correlation coefficients varied nearly from -0.80 to 0.90, likely depending on the soil properties across the wavelength range. However, the interpretation of correlations is complicated due to the complexity of overlapping spectral features of soil parameters. According to the correlation pattern, the soil properties could be categorized into two groups: the first group ($CaCO_3$, clay, pH, Ca, Cd, K, Mn, P, and Pb) had positive (Figure 3a), and the second group (sand, SOC, Fe, Cu, Mg, and Zn) had negative correlation coefficients (Figure 3b). The soil properties in each group could be further divided into subgroups, e.g., based on the similarity of the correlation patterns: clay, pH, Mn, Pb, and Cd; $CaCO_3$ and Ca; K and P; SOC and Fe; and sand, Mg, and Zn.
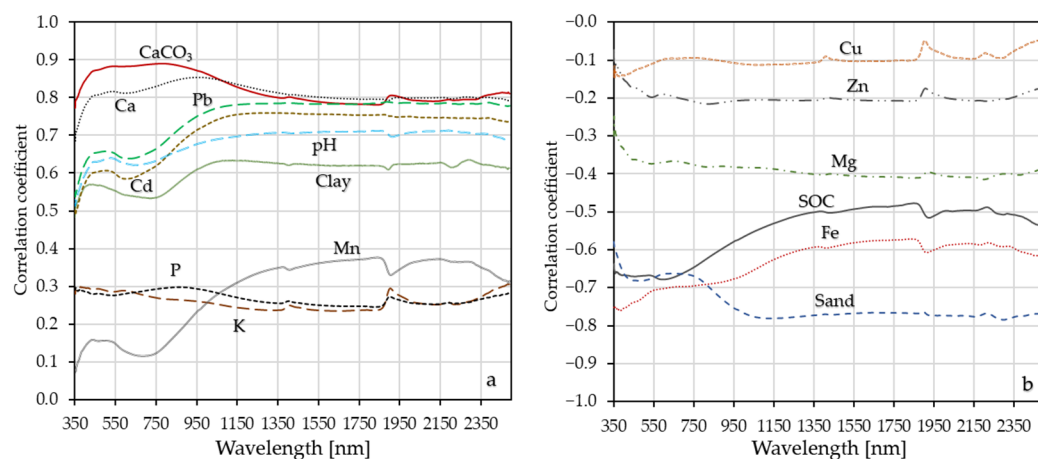
Figure 3. Correlations between soil properties and spectral reflectance across the 350–2500 nm wavelength region ((**a**)—positively and (**b**)—negatively correlated soil properties).

The relations between $CaCO_3$ and the reflectance spectra formed the highest positive correlation coefficients. Its maximum value ($r = 0.89$ *) was found in VIS region at 790 nm. As expected from the correlation between $CaCO_3$ and Ca ($r = 0.86$ *), extractable Ca also showed a high correlation to the VIS region. Cd and Pb showed a very similar correlation pattern to that of pH or clay with lower correlations in the VIS ($r = 0.50$–0.65), and higher correlations ($r = 0.60$–0.75 *) in the NIR range. Mn, K, and P also showed a similar shape, but with much lower correlation coefficients ($r < 0.40$ *). Among the negatively correlated soil constituents, sand, Fe, and SOC revealed higher correlation coefficients ($r = 0.50$–0.78 *) across the whole wavelength region (Figure 3b). The sand content apparently showed higher correlations in the NIR and SWIR range. In this range, the correlation coefficients were >0.70, and its maximum value ($r = 0.77$ *) was found at 2308 nm. However, the well-correlated Fe and SOC ($r = 0.60$ *) showed strong relations in the VIS range, with a similar/parallel trend in the NIR and SWIR range (Figure 3b). Higher correlative relations between SOC and reflectance spectra were observed in the VIS region (especially from 540 to 690 nm), with the highest correlation ($r = 0.69$ *) observed at the 605 nm waveband. The relation between the reflectance spectra and Cu, Zn, and Mg was generally weak across VIS-NIR region.

The M3-extractable elements that were (i) positively correlated with the reflectance spectra showed similar spectral correlation patterns to primary soil properties such as $CaCO_3$, pH, and clay, and those that (ii) were negatively correlated with the reflectance spectra showed less similarity with SOC and sand. Despite the soil properties were

positively or negatively correlated, absorption features were distinctly reflected in the correlation spectra. Likewise, for $CaCO_3$, Ca, K, and P, the absorption feature at 1900 nm appeared as a correlation peak, whereas an inverse trend was visible for pH, clay, and Mn (Figure 3). Generally, soil reflectance increased with an increase in $CaCO_3$ and clay content and decreased with increase in sand and SOC contents. As $CaCO_3$ and SOC significantly affect the soil color, these soil constituents showed the highest impacts in the VIS part of the spectrum. Fe showed even stronger correlations than SOC in the VIS region.

3.5.2. Principal Component and Partial Correlation Analysis

The PCA biplot was employed to visualize the relationships among soil properties and spectral data (Figure 4). The PCA of the spectral data showed that PC1 and PC2 explained 94.59% and 4.75% (in total 99.35%) of the variance of the original spectral data, respectively. Therefore, PC1 and PC2 were used to investigate the spectral information and its relations to the soil properties. The biplot represented a highly discriminating pattern of the basic soil properties, M3 elements, and the first two principal components of the spectral data. About 60% of the total variability was explained by F1 and F2 (Figure 4).
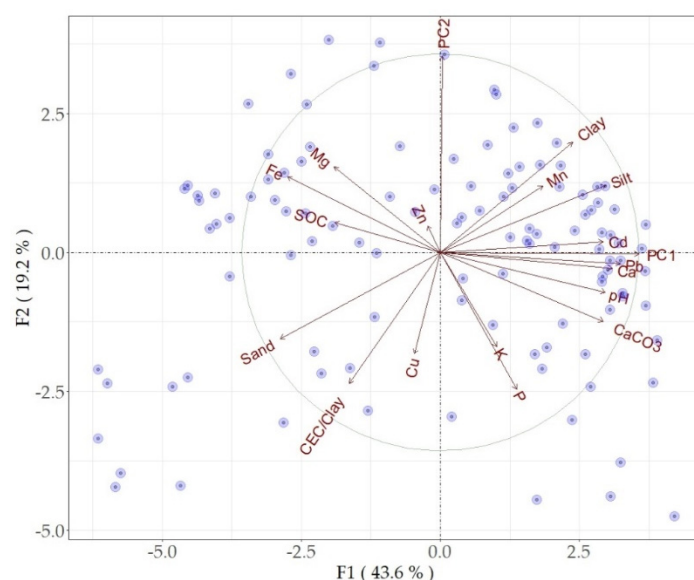
Figure 4. Biplot of PCA factors 1 and 2, based on soil properties and the first two principal components (PC1 and PC2) of the spectral reflectance data.

The PC1 was positively correlated with a group of soil properties ($CaCO_3$, silt, clay, pH, Ca, Cd, Pb, and Mn) and negatively correlated with Fe, Mg, and SOC. The PC2 specifically showed orthogonality with the majority of parameters, whereas positive correlation occurred with Zn (containing less information) and negative correlation with sand, CEC/Clay, Cu, P, and K. Most of the soil properties except Zn, Cu, K, and Mn were approximately proportional to the variability included in the two components. The relations among particle sizes and CEC/Clay showed the expected typical pattern due to apparent collinearity. The sample distribution pattern far from the biplot origin confirmed considerable variability and better performance (Figure 4).

For partial correlation analysis, we selected the variable PC1 to represent and investigate the spectral information and its relationships to the soil properties, since significant correlations (r = 0.37–0.88 *) were found between PC1 and all soil properties except Cu, Mn, and Zn (Table 4). The correlation coefficients showed that the contribution of $CaCO_3$ was the highest compared to other parameters. The correlation coefficients between PC1 and SOC, pH, Fe, Ca, Cd, Mg, P, Pb, and Zn decreased when $CaCO_3$ was controlled. The particle sizes had a varying effect on the M3 elements. For example, when the particle sizes were controlled, the correlation between PC1 and K became insignificant, whereas the correlation

between PC1 and P and Zn remained significant. At the same time, slight effects of particle sizes on SOC, pH, Mn, Cu, and Cd were observed. Unlike other controlled parameters, SOC had a small effect, yet PC1 was still significantly correlated with the soil properties except for Cu, Mn, and Zn. The Fe content was found as the main controlling factor for Mg and P, and the correlation between these elements and PC1 became insignificant when Fe was controlled. Removing $CaCO_3$, clay, and Fe simultaneously resulted in incisively decreasing the correlation coefficients between PC1 and the M3 elements. Though the correlations remained significant, they were weaker for Ca, Cd, and Pb and decreased from 0.88, 0.75, and 0.76 to 0.37, 0.32, and 0.28, respectively.

Table 4. Spearman's Rho correlation coefficients between the first principal component (PC1) of soil reflectance spectra and soil properties (1st row), and the partial correlation coefficients with controlling basic soil properties.

Category Controlled	CaCO₃	Clay	Silt	Sand	SOC	pH	Fe	Ca	Cd	Cu	K	Mg	Mn	P	Pb	Zn
							Spearman's Rho correlation coefficients									
	0.84 *	0.62 *	0.84 *	−0.78 *	−0.59 *	0.70 *	−0.68 *	0.88 *	0.75 *	−0.10	0.37 *	−0.40 *	0.17	0.39 *	0.76 *	−0.17
							Partial correlation coefficients									
CaCO₃		0.63 *	0.74 *	−0.73 *	−0.18	0.20	0.06	0.56 *	0.42 *	−0.09	0.26 *	0.20	0.23	−0.08	0.37 *	0.05
Clay	0.84 *		0.73 *	−0.60 *	−0.47 *	0.65 *	−0.70 *	0.82 *	0.67 *	0.09	0.23	−0.51 *	0.05	0.53 *	0.69 *	−0.26 *
Silt	0.74 *	0.18		−0.08	−0.47 *	0.45 *	−0.59 *	0.72 *	0.43 *	0.07	0.17	−0.43 *	−0.18	0.46 *	0.50 *	−0.31 *
Sand	0.80 *	0.00	0.49 *		−0.41 *	0.55 *	−0.66 *	0.76 *	0.59 *	0.14	0.18	−0.51 *	−0.06	0.47 *	0.61 *	−0.24 *
SOC	0.75 *	0.52 *	0.80 *	−0.70 *		0.54 *	−0.51 *	0.84 *	0.77 *	0.10	0.33 *	−0.34 *	0.23	0.31 *	0.76 *	−0.14
pH	0.67 *	0.56 *	0.73 *	−0.68 *	−0.31 *		−0.39 *	0.77 *	0.57 *	−0.04	0.28 *	0.04	−0.22	0.21	0.56 *	−0.27 *
Fe	0.68 *	0.64 *	0.79 *	−0.76 *	−0.30 *	0.43 *		0.76 *	0.58 *	−0.06	0.28 *	−0.01	0.14	0.08	0.53 *	−0.16
CaCO₃ + Clay + Fe			0.59 *	−0.49 *	−0.03	0.15		0.37 *	0.32 *	0.11	0.14	0.03	0.11	0.12	0.28 *	−0.07

* stands for significant correlations at the level of $p < 0.05$.

3.5.3. Important Wavelengths

To determine important wavelengths (variables) used in predictions, the VIP scores of the PLSR models using the optimal pre-processing techniques for each soil constituent were computed and plotted across the 350–2500 nm wavelength region (Figure 5). Generally, a large similarity was observed for the VIP score pattern of all the tested soil constituents except for clay and SOC. Unlike all others, the VIP graph was characterized by a simple pattern for SOC, whereas multiple VIP peaks were observed for clay. The VIP pattern of SOC showed a comparatively wider wavelength range (350–735 nm) in the VIS region, with a spectral feature centered at 1892 nm that was also found important for SOC and other soil properties. For the clay prediction, the VIP scores showed a different pattern, particularly in the VIS region. The two wavelength ranges centered at 574 and 781 nm contain soil color information and contributed to the clay prediction. Further VIP peaks centered at 1415, 1892, 1920, 2210, and 2302 nm also contributed to the clay content estimation. A VIP peak negligibly exceeding the threshold centered at 1024 nm was noticed as an important predictor for clay as well.

For predicting $CaCO_3$, a wide spectral range from 350 to 583 nm, and some more narrow ranges in the NIR and SWIR region (1383, 1414, 1880, 1913, 2178, 2311, 2415, 2341, and 2460 nm) were found as important predictors (Figure 5). As expected from the relation between $CaCO_3$ and pH, their VIP patterns and important wavelengths were apparently similar. The wavelength centered at 380 nm was found important for the prediction of all tested soil constituents. Despite the fact that correlations between the spectra and the sand content in the NIR region were higher than in the VIS region (Figure 3), the VIS range apparently contributed to its prediction, with a wavelength range from 360 to 570 nm. At the same time, the VIP peaks centered at 860, 1380, 1415, 1915, 2178, and 2246 nm were also found as essential predictors.
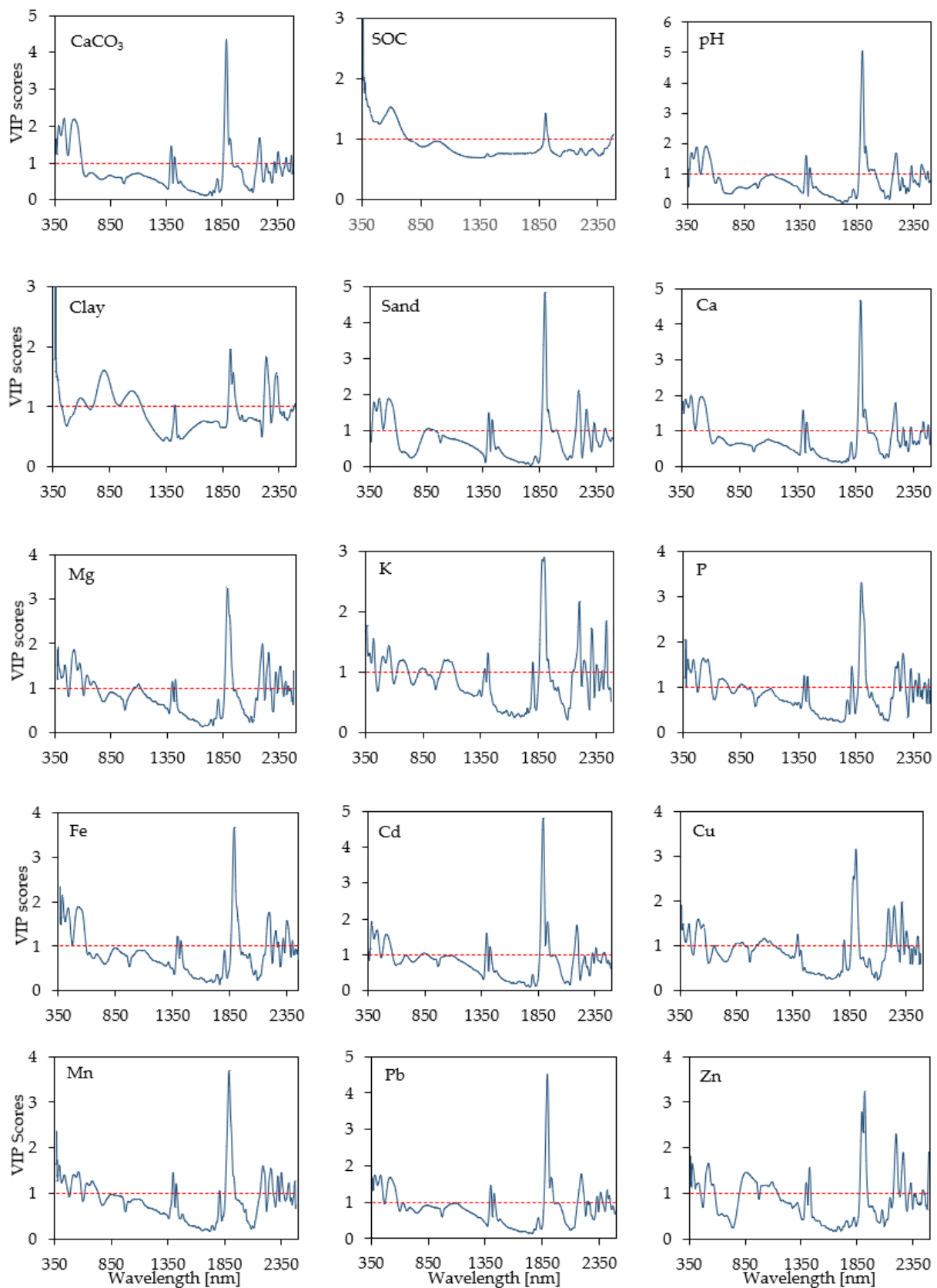
Figure 5. Variable importance in projection (VIP) scores associated with the PLSR models for the soil properties. The wavelength is considered as important if the VIP scores exceed the threshold value of 1 (dashed red line).

As expected from the similarity of the correlation spectra (Figure 3), the VIP patterns of Ca, Cd, and Pb were similar to that of CaCO$_3$, and pH indicated the same important wavelengths (Figure 5). Unlike Ca and Pb, a VIP peak slightly exceeding the threshold limit centered at ~850 nm was a significant predictor for the Cd content. Likewise, the same VIP peak was also typical for the predictions of K, P, and Zn (Figure 3). Since Cu was one of the poorly predicted elements, the interpretation of important wavelengths requires caution, considering its weak relationship with the reflectance spectra. Multiple narrow wavelength ranges (centered at 380, 434, 530, and 654 nm) contributed to its prediction (Figure 5). Further peaks were observed at 1805, 1885, 2200, 2249, 2315, 2358, and 2460 nm. For the Fe prediction, a wider range of wavelengths with peaks centered at 384, 435, and 530 nm, and at 1400, 1426, 2200, and 2280 nm was found important. The VIP graph of K and P showed a similar pattern, e.g., with distinct peaks centered at 1400, 1800, 1900, and 2200 nm. In the case of Mg and Cu, the VIP scores indicated a similar pattern, i.e., four VIP peaks (435, 520, 567, and 635 nm) in the VIS region and several peaks in the NIR and SWIR region (1095, 1395, 1424, 2205, 2261, and 2326 nm). For the prediction of Mn, the VIP scores identified three consecutive wavelength ranges (375–468, 485–586, and 613–735 nm) comprising double peaks similar to cases of K, P, and Mg. Although the Zn content was poorly predicted, the VIP scores revealed a series of important wavelengths (380, 430, 530, 860, 1095, 1395, 1425, 1892, and 2260 nm).

4. Discussion

4.1. Accuracy Analysis of the PLSR Models

Both calibration and validation models produced differences in RPD values compared to the cross-validation setup (Table 3). The calibration and validation models performed differently for CaCO$_3$ and clay. An excellent and very good calibration model for CaCO$_3$ and clay, respectively, was substituted with a very good and fair validation model. With regard to the model performances for the M3 elements, the calibration and validation models for predicting Ca, Cd, Fe, Mg, and Pb performed similarly. However, the performances of the validation model for the elements Cu, K, Mn, and Zn were better than the corresponding calibration subset. The calibration models were fair for K and Mn, whereas the validation subset revealed excellent and good models, respectively. Thus, all the soil constituents under consideration were estimated based on VIS-NIR spectral data, and the R^2 values of the validation models ranged from 0.51 to 0.90. The soil properties could be ranked as following according to the cross-validation model performances: CaCO$_3$ > SOC > Pb > Ca > sand > silt > clay > K > Cd > Mn > Fe > Cu > pH > Mg > P > Zn (Table 3).

Such differences in the model performances were found in previous studies. Similar to our results, it was reported that calibration, validation, and prediction processes can be complicated and dependent on soil types and its physiochemical properties, the soil management history, chemical characteristics, and the analytical method, along with the sample size and distribution, as well as the prediction approaches [4,7,13,14,27–29,54,61]. Based on 2250 soil samples that were collected in a 5-year period from four soil groups varying in clay mineralogy (Alfisols, Inceptisols, Entisols, and Vertisols) and a series of robust regression analyses, Golia and Diakoloukas [61] concluded that several factors (pH, electric conductivity, SOC, soil texture, exchangeable cations, and metal oxides along with coexisting metals) affect the total concentrations of Fe and Cd in soils. For instance, total Cd was reliably predicted with Cu and Zn in Entisols; clay, Pb, and CEC in Alfisols; and clay, Pb, and CEC in Vertisols [61]. These results are in line with our findings (Supplementary Table S1, Figure 4).

Models may perform better when variations in the whole sample set are well-reflected in the calibration and validation data set. The observed dissimilarities in our study can be explained by the descriptive statistics of the calibration and validation subsets used to predict the M3 elements. Furthermore, the prediction accuracies of the soil properties are associated with the statistical distribution and the high variability of the measured parameters [18,29]. Additionally, it is reasonable that the samples on the boundaries of the

entire set are included in the calibration subset [62]. In accordance with recent studies, K is one of the complex elements to estimate with VIS-NIR spectra regardless of the extraction procedure. For instance, a good model was obtained by Rodríguez-Pérez et al. [7], whereas Bilgili et al. [13] found a poor model (R^2 = 0.25–0.32) for the K content. Chang et al. [27] obtained a more accurate model (R^2 = 0.64) for K extracted in M3 compared to K extracted in NH4OAc (R^2 = 0.55). In our case, a higher performance of the validation subset for K was probably not relied on for the robustness of the PLSR prediction, because the change in RPD was based on the standard deviation of the corresponding subsets. The standard deviation (not shown) of the calibration subset (122 mg/kg) was considerably narrower than the SD of the validation subset (171 mg/kg).

Few studies have evaluated the potential of VIS-NIR spectra to predict M3-extractable soil heavy metals. Predicting multiple soil properties and M3-extractable elements with NIR spectra and principal component regression by using a large amount of soil samples (collected from a national inventory) allowed us to obtain successful models for Ca (R^2 > 0.80), fair models for Fe, K, Mg, and Mn (R^2 = 0.50–0.80), and poor models (R^2 < 0.50) for Cu, P, and Zn [27]. Reliable calibrations for Ca, Cu, and Mg (R^2 > 0.70, RPD > 1.75), and less-reliable calibrations for Al, Fe, K, Mn, P, and Zn (R^2 > 0.70, RPD < 1.75) were found in another study [28]. The models were consistent for only Ca, Mg, and Mn when independent samples were used for the validation. The study presumed that the pH of the M3-extractant may affect the solubility of most of these nutrients, regardless of the soil texture and, consequently, the potential of NIR spectra to predict them. Higher correlations of the nutrients with clay content improved their predictability by NIR spectra [28].

The results of NIR spectral prediction of M3 elements and soil characteristics based on 448 samples collected over a 7-year period from an experimental site revealed that the predictions were excellent for Ca, Mn, pH, and Mg, whereas moderately successful results were obtained for Al, total C, Fe, and Zn, and poor results for P and Cu [24]. Luce et al. [18] assessed the potential of VIS-NIR spectroscopy to predict total, M3- and DTPA-extractable, and water-soluble heavy metal concentrations in different agricultural soils repeatedly treated with biosolid and liming by-product. They concluded that VIS-NIR spectroscopy has the lesser capacity to predict water-soluble heavy metal concentrations, especially when site-specific samples were used. In contrast, the validation results were excellent for M3-extracted Cu and Cd, good for Ni, and poor for Zn [18]. These results were similar to our study for Cd and Zn, though inverse for Cu, i.e., the study obtained poor models for predicting Cu from site-specific test areas.

The prediction accuracy of the models was concordant with their correlation bands (between soil constituent and reflectance spectrum). The prediction accuracy of the M3 elements was consistent with their relations to basic soil properties (Figures 2 and 4), confirming that the major spectral prediction mechanisms for the M3-extractable elements rely on the correlation of those elements with $CaCO_3$, pH, sand, clay, and SOC. Moreover, the inter-correlations between secondary soil properties and spectrally active soil components such as Fe oxides, organic matter, clay, $CaCO_3$, and pH were also related to the vital predictive mechanisms [24,26,28], which was consistent with the current study.

4.2. Effect of Pre-Processing Techniques on PLSR Predictions

One of the challenges in soil spectroscopy is to develop pre-processing techniques to extract more information from the spectra [11]. Abdi et al. [24] employed 40 pre-processing techniques (different gap sizes and smoothing points) to predict total C, N, P, pH, and M3-extractable elements using 448 soil samples. Best estimates were obtained from smoothed spectra for Cu, Zn, and total and available P, first derivative spectra for K, Ca, and Al, and second derivative spectra for total C, N, pH, Fe, Mg, and Mn. Nduwamungu et al. [28] applied 12 pre-processing techniques for predicting M3-extractable elements. In this study, first and second derivative spectra produced less reliable predictions for Fe, K, Al, P, Mn, and Zn.

The feasibility of a series of pre-processing techniques (11 filtering techniques, 8 normalizations, and 7 linearization/scaling techniques; in total, 26) was tested to predict basic soil properties [33]. Absorbance and MSC spectra combined with filtering techniques were used in PLSR modelling as pre-processing techniques. The first derivative spectra with a gap segment size of 10 bands allowed us to remove particle size effects, reduce the drift of the baseline, and highlight certain parts of the spectral information. This pre-processing approach outperformed other techniques for all soil properties. MSC removes undesired scatter or texture effects and thereby improved the prediction of the clay content [63]. The MSC was also effective in diminishing the baseline drift caused by the differences in grinding and optical setups [64]. Absorbance spectra provided the most reliable model for predicting SOC content due to spectral features of organic matter occurring in the VIS and NIR region [10,65]. Thus, the optimal pre-processing methods found in this study are in agreement with findings in previous studies [13,15,24].

4.3. Analysis of Prediction Mechanisms

4.3.1. Correlation, PCA and Partial Correlation Analysis

Spectral characteristics of a large variety of soil properties and M3 elements were described in numerous studies [13,26,66,67]. Correlation patterns of the M3-extractable elements (i) Ca, Cd, K, Mn, P, and Pb were similar to those of $CaCO_3$, silt, pH, and clay, (ii) Mg and Zn were alike to sand, and (iii) Cu was resembled to SOC and Fe (Figure 3). The higher correlative relations between SOC and the reflectance spectra in the VIS region reached a maximum (r = 0.69 *) at 605 nm. A similar correlation pattern was described in recent studies [12,68] as an absorption peak centered at 600 nm, contributing to the strong estimation of SOC. SOC has a broad absorption in the VIS region and shows a pattern in spectra as linear, concave, or an upward shift shape that also depends on soil mineralogy [69]. Fe showed even stronger correlations than SOC that are most likely related to features of Fe oxides linked with soil types [61]. Equally, the concentrations of spectrally featureless heavy metals in soils were estimated via the correlations among Fe oxides [30], organic matter [69], and clay minerals [70]. The weak relation between the reflectance spectra and Cu, Zn, and Mg was in agreement with other studies [26].

The results of PCA and consequent partial correlation analysis showed significant correlations between PC1 of the spectral data and the soil properties except Cu, Mn, and Zn (Table 4). $CaCO_3$ was the most influential factor followed by Fe, clay, and SOC for predicting spectrally featureless soil properties. The small influence of SOC on M3-extracted metals was not fully consistent with previous studies. Correlations with Pb, Zn, and Cd or As, Cr, and Cu are governed by soil organic matter or Fe (oxides) along with coexisting metals and soil texture that can reliably assist as predictive factors [18,19,29,61]. Therefore, based on our analyses, M3-extractable elements are indirectly bound to iron oxides, clay minerals, or organic matter and can be assessed using the spectral features of these soil constituents, which is in line with other studies [4,24,26].

4.3.2. Analysis of Important Wavelengths

The VIP score patterns of the tested soil constituents conceded a large similarity (Figure 5). Most soil properties (silt, pH, Ca, Cd, Pb, Mn) showed a similar VIP pattern to that of $CaCO_3$. Our results for predicting the clay content and $CaCO_3$ were comparable to various other studies [8,71–76]. As expected from the relation between $CaCO_3$ and pH, their VIP patterns and important wavelengths were similar. The wavelength centered at 380 nm was found important for all tested soil properties and is presumably related to clay mineralogy [4,75]. In another study, the VIP peak centered at 380 nm was related to free Fe (382 nm) content associated with influence of Fe oxides and hydroxides [5,75]. The VIP graph was characterized by simple pattern for SOC and comparatively wider wavelength range in the VIS region (350–735 nm). The spectral feature centered at 1892 nm was found important not only for SOC, but also for all other soil properties (Figure 5). A relatively narrow range in the VIS region (360–570 nm) apparently contributed to the

sand content prediction. It could be linked to the spectral features of iron oxides and SOC, whereas a weak VIP peak at 860 nm may match to the absorption feature of hematite [72]. These relations were confirmed by the significant correlation ($r \geq 0.40$ *) between sand and SOC or Fe (Figures 2 and 5). Further wavelengths (1380, 1415, 1915, 2178, and 2246 nm) are presumably related to water (1380 and 1915 nm) and clay minerals (1415, 2160, and 2230 nm) [4,59]. The rest of the VIP peaks (1880, 2315, and 2336 nm) were probably related to carbonates, also found as notable predictors, as indicated by a moderate correlation ($r = -0.52$ *) between $CaCO_3$ and sand [8,72]. A strong correlation between sand and clay specifies the substantial contribution of clay mineralogy to sand prediction. The incomplete incorporation of the clay minerals and spectral features of hematite may be one of the reasons for the relatively low prediction accuracy of the sand content compared to the clay content (Table 3).

As anticipated from the similarity of correlation spectra and VIP patterns (Figures 3 and 5), the important wavelengths for Ca, Cd, and Pb were the same as that of $CaCO_3$ and pH (Figure 2). A VIP peak centered at ~850 nm was a significant predictor for the Cd content. Likewise, the same VIP peak was also important for predicting K, P, and Zn. The band assignment for this case was difficult since the wavelength corresponded to the absorption feature of alkyl [71]. Several VIP peaks (380, 434, 530, and 654 nm) corresponding to the absorption features of Fe oxides contributed to the prediction of Cu (Figure 5). The peak at 654 nm was likely related to electronic transition processes of goethite at 650 nm [71]. However, no correlation between Fe and Cu, and lower ($r = 0.30$ *), yet significant correlation between Cu and SOC (Figure 2), showed that the contribution of organic matter in the VIS region was obvious. The VIP peaks centered at 948 and 1905 nm are probably related to water absorption bands, while the peaks at 880, 1097, 1148, and 1394 nm were related to aromatics (1100 nm), alkyl (880, 1138 nm), and hydroxyl (1400 nm) groups, respectively [71]. The peaks observed at 2200, 2249, and 2460 nm corresponded to the absorption features of clay minerals, as confirmed by the significant correlation with clay content. The further three VIP peaks observed at 1805, 2315, and 2358 nm fell within the methyl absorption feature ranges [71].

For the Fe prediction, a wider range in the VIS region (Figure 5) three VIP peaks (384, 435 and 530 nm) was important, which demonstrated the close relation to the absorption features of free Fe, goethite, and hematite [5]. The VIP peaks at 1400, 1426, 2200 and 2280 nm were presumably related to hydroxyl (1400 nm), clay minerals (1415 nm and 2206 nm) and organic matter (2275 nm). The contribution of organic matter to the Fe prediction is confirmed through the significant correlation ($r = 0.60$ *) between Fe and SOC (Figure 2). Similar to the Cd prediction, other important wavelengths (2326 and 2360 nm) might be associated with the absorption feature of methyl [71]. However, the wavelength of 2326 nm should be treated with care, since an absorption feature of $CaCO_3$ occurs at 2336 nm. In addition, a strong correlation ($r = -0.84$ *) between Fe and $CaCO_3$ (Figure 2) was considered in partial correlation analysis to understand the contribution of the basic soil properties (e.g., controlling factor $CaCO_3$, Fe, clay) for the prediction of M3-extractable elements (Table 4) [19].

A large similarity between the correlation spectra and the VIP patterns (Figures 3 and 5) for K and P seized the moderate correlation between them ($r = 0.43$ *). Four VIP peaks observed at 435, 520, 567, and 635 nm corresponded to the absorption features of Fe oxides [71], which confirmed the contribution of Fe to the prediction of P (Figure 2, Table 4). In accordance with partial correlation analysis, K and P were not controlled by SOC, despite the moderate correlation between Fe and SOC (Figure 2). It should be noted that the variation in Fe content is explained by $CaCO_3$, pH, and SOC (73%), whereas the variation in P is related to $CaCO_3$, clay, and pH (32%) (Supplementary Table S1). Increasing the $CaCO_3$ content increases the pH value, which, in turn, affects the solubility of nutrients (Fe, K, P, etc.). Other peaks can be assigned to different soil constituents, such as aromatics (1095 nm), kaolinite (1395 and 1424 nm), illite or smectitte (2205 nm), aliphatics (2261 nm), and carbonates (2326 nm) [4,71]. Similar patterns and overlapping of multiple VIP peaks

were found for Mg and Cu (Figure 5). As is obvious from the significant correlations between Mg and the spectrally active CaCO$_3$ or Fe ($r = -0.60$ *), and SOC ($r = -0.22$ *), spectral features of the mineral composition could play an important role in predicting Mg. Analogue to Mg, K, and P, three consecutive wavelength ranges (375–468, 485–586, and 613–735 nm) comprising double peaks were the important wavelengths for the prediction of Mn. The significant correlation between Mn and soil texture reveals that the clay mineralogy plays an important role in its prediction. Zn is weakly correlated with only CaCO$_3$ and Ca, and thus, the interpretation of the VIP sores was challenging. Clearly visible peaks at 380, 430, and 530 nm in the VIS region corresponded to absorption features of Fe oxides [5,71], whereas others (860, 1095, 1395, 1425, 1892, and 2260 nm) were probably related to the absorption bands of organic matter (853 and 1100 nm), water, and clay minerals [72].

The correlation spectra and VIP score patterns of the soil constituents revealed that the improved prediction of M3 elements relies on spectrally active soil constituents. Moreover, relations between basic soil properties and heavy metal concentrations could be inferred from the similarity of their important wavelengths [29]. The overlapping of the important wavelengths of soil properties also implied their close associations.

5. Conclusions

We examined the potential of VIS-NIR reflectance spectroscopy and PLSR modelling to predict M3-extractable elements and basic soil properties, while focusing on the effect of different pre-processing techniques and prediction mechanisms for M3-extractable elements. The M3 nutrients are much less investigated, particularly in mountainous land with high variability of soil properties due to the climatic factors, land-use history, and degradation. The pre-processing techniques considerably improved prediction accuracies, and the first derivative of the spectra was found optimal for all the tested soil properties except SOC and the clay content, for which both absorbance and MSC spectra were the most reliable techniques. The models' predictions were excellent for CaCO$_3$ and SOC, very good for sand, silt, Ca, and Pb, good for clay, K, and Cd, fair for pH, Fe, Mn, and Cu, and poor for Mg, P, and Zn.

The PCA biplot, partial correlation, stepwise regression analysis, and VIP procedures allowed us to elucidate the underlying prediction mechanisms for M3-extractable nutrients. Unlike the past studies, the spectral estimation of M3-extractable elements was mostly related to their correlation with CaCO$_3$ rather than other soil properties and soil organic matter; the latter showed a negligible effect in this case. The spectral estimation of pH, Ca, Mg, P, Fe, Pb, and Cd concentrations were based on their relations with CaCO$_3$, whereas Mg and P concentrations were associated with Fe oxides. Soil texture displayed a varying effect on M3 elements: clay, silt, and sand were found to contribute to predicting K concentration, whereas they appeared as confounding factors for predicting P and Zn concentration. The nutrients Mn, Cu, and Zn, which had a weak correlation with CaCO$_3$, particle sizes, SOC, Fe, and spectral reflectance, were predicted with lower accuracy. The soil constituents that correlated closely with reflectance were reliably predicted and vice versa. The underlying prediction mechanisms for the soil properties that have no spectral response (pH, Ca, K, Mg, P, Cd, Cu, Mn, Pb, and Zn) relied on their co-variation with spectrally active soil properties. The results of this study can be used in mountain lands to evaluate and control effects of land use, agricultural management, and conservation on land degradation neutrality and soil quality indices. The results contribute to building a pioneering spectral database for the soils of the Caucasus Mountains, as well as for global spectral soil libraries, and assist with the monitoring of soil and water quality (e.g., land degradation and rehabilitation) in relevant mountainous regions. Further focus needs to be given to (i) applying different calibration strategies, (ii) improving the sampling approach and selecting process of the calibration and validation samples, and (iii) extending the spectral methodology for soil profile assessment.

Author Contributions: Conceptualization, E.M., C.K., M.D. and C.G.; methodology, E.M., C.K., F.R., M.D., K.L., R.Ł., W.G. and A.I.M.; spectral measurements: E.M., M.D. and F.R.; software, E.M., C.K. and F.R.; validation, E.M., C.K., F.R. and C.G.; formal analysis, F.R., M.D., C.G., W.G. and A.I.M.; investigation, E.M., C.K. and W.G.; resources, M.D., F.R. and C.G.; data curation, E.M. and C.K.; writing—original draft preparation, E.M.; writing—review and editing, E.M., C.K., A.I.M., M.D. and C.G.; visualization, E.M.; supervision, C.G. and C.K.; project administration, M.D. and C.K.; funding acquisition, C.K. All authors have read and agreed to the published version of the manuscript.

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

References

1. Paz-Kagan, T.; Shacha, M.; Zaady, E.; Karnieli, A. A spectral soil quality index (SSQI) for characterizing soil function in areas of changed land use. *Geoderma* **2014**, *230*, 171–184. [CrossRef]
2. Mamedov, A.I.; Tsunekawa, A.; Tsubo, M.; Fujimaki, H.; Kawai, T.; Kebede, B.; Mulualem, T.; Abebe, G.; Wubet, A.; Levy, G.J. Soil structure stability under different land uses in association with polyacrylamide effects. *Sustainability* **2021**, *13*, 1407. [CrossRef]
3. Stoorvogel, J.; Kooistra, L.; Bouma, J. Managing Soil Variability at Different Spatial Scales as a Basis for Precision Agriculture. In *Soil-Specific Farming Precision Agriculture*; Advances in Soil Science; Rattan, L., Stewart, B., Eds.; CRC Press: New York, NY, USA, 2016; pp. 37–72. [CrossRef]
4. Soriano-Disla, J.; Janik, L.; Viscarra Rossel, R.; MacDonald, L.; McLaughlinad, M. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.* **2014**, *49*, 139–186. [CrossRef]
5. Islam, K.; Singh, B.; McBratney, A. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust. J. Soil Res.* **2003**, *41*, 1101–1114. [CrossRef]
6. Riedel, F.; Denk, M.; Müller, I.; Barth, N.; Gläßer, C. 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* **2018**, *315*, 188–198. [CrossRef]
7. Rodríguez-Pérez, J.R.; Marcelo, V.; Pereira-Obaya, D.; García-Fernández, M.; Sanz-Ablanedo, E. Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. *Agronomy* **2021**, *11*, 1895. [CrossRef]
8. Mouazen, A.; Kuang, B.; De Baerdemaeker, J.; Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* **2010**, *158*, 23–31. [CrossRef]
9. Pinheiro, É.F.; Ceddia, M.; Clingensmith, C.; Grunwald, S.; Vasques, G. Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the Central Amazon. *Remote Sens.* **2017**, *9*, 293. [CrossRef]
10. Viscarra Rossel, R.; Walvoort, D.; McBratney, A.; Janik, L.; Skjemstad, J. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]
11. Mehmood, T.; Liland, K.; Snipen, L.; Sæbø, S. A review of variable selection methods in partial least squares regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69. [CrossRef]
12. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.; Clairotte, M.; Csorba, A. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Adv. Agron.* **2015**, *132*, 139–159. [CrossRef]
13. Bilgili, V.; van Es, M.; Akbas, F.; Durak, A.; Hively, D. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semiarid area of Turkey. *J. Arid Environ.* **2010**, *74*, 229–238. [CrossRef]
14. Nawar, S.; Buddenbaum, H.; Hill, J.; Kozak, J.; Mouazen, A. Estimating the soil clay content and organic matter by means of different calibration methods of Vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res.* **2016**, *155*, 510–522. [CrossRef]

15. Vasques, G.; Grunwald, S.; Sickman, J. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* **2008**, *146*, 14–25. [CrossRef]

16. Fang, Q.; Hong, H.; Zhao, L.; Kukolich, S.; Yin, K.; Wang, C. Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy. *Hindawi J. Spectrosc.* **2018**, *2018*, 3168974. [CrossRef]

17. Recena, R.; Fernández-Cabanás, V.; Delgado, A. Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. *Geoderma* **2019**, *337*, 368–374. [CrossRef]

18. Luce, M.; Ziadi, N.; Gagnon, B.; Karam, A. Visible near infrared reflectance spectroscopy prediction of soil heavy metal concentrations in paper mill biosolid- and liming by-product-amended agricultural soils. *Geoderma* **2017**, *288*, 23–36. [CrossRef]

19. Cheng, H.; Shen, R.; Chen, Y.; Wan, Q.; Shi, T.; Wang, J.; Wan, Y.; Hong, Y.; Li, X. Estimating heavy metal concentrations in suburban soils with reflectance spectroscopy. *Geoderma* **2019**, *336*, 59–67. [CrossRef]

20. Mehlich, A. Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Commun. Soil Sci. Plant Anal.* **2008**, *15*, 1409–1416. [CrossRef]

21. Burt, R.; Soil Survey Staff (Eds.) *Soil Survey Staff, Soil Survey Field and Laboratory Methods Manual, Soil Survey Investigations Report No. 51*; Version 2.0.; U.S. Department of Agriculture, Natural Resources Conservation Service: Redmond, OR, USA, 2014; 487p.

22. Ziadi, N.; Tran, T. Mehlich 3 extractable elements. In *Soil Sampling and Methods of Analysis*; Carter, M., Gregorich, E., Eds.; CRC Press: Boca Raton, FL, USA; Taylor & Francis: Boca Raton, FL, USA, 2007; pp. 81–88.

23. Zbíral, J. Determination of plant-available micronutrients by the Mehlich 3 soil extractant—a proposal of critical values. *Plant Soil Environ.* **2016**, *62*, 527–531. [CrossRef]

24. Abdi, D.; Tremblay, G.; Ziadi, N.; Bélanger, G.; Parent, L. Predicting soil phosphorus-related properties using near-infrared reflectance spectroscopy. *Soil Sci. Soc. Am. J.* **2012**, *76*, 2318–2326. [CrossRef]

25. Hively, W.; McCarty, G.; Reeves, J.; Lang, M.; Oesterling, R.; Delwiche, S. Use of airborne hyperspectral imagery to map soil properties in tilled agricultural fields. *Appl. Environ. Soil. Sci.* **2011**, *2011*, 358193. [CrossRef]

26. Vašát, R.; Kodešová, R.; Borůvka, L.; Klement, A.; Jakšík, O.; Gholizadeh, A. 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* **2014**, *232–234*, 208–218. [CrossRef]

27. Chang, C.; Laird, D.; Mausbach, M.; Maurice, J.; Hurburgh, J. Near-Infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480–490. [CrossRef]

28. Nduwamungu, C.; Ziadi, N.; Parent, L.E.; Tremblay, G. Mehlich 3 extractable nutrients as determined by near-infrared reflectance spectroscopy. *Can. J. Soil Sci.* **2009**, *89*, 579–587. [CrossRef]

29. Wang, J.J.; Cui, L.J.; Gao, W.X.; Shi, T.Z.; Chen, Y.Y.; Gao, Y. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma* **2014**, *216*, 1–9. [CrossRef]

30. Wu, Y.; Chen, J.; Ji, J.; Gong, P.; Liao, Q.; Tian, Q.; Ma, H. A mechanism study of reflectance spectroscopy for investigating heavy metals in soils. *Soil Sci. Soc. Am. J.* **2007**, *71*, 918–926. [CrossRef]

31. Ismayilov, A. Soil resources of Azerbaijan. In *Soil Resources of Mediterranean and Caucasus Countries Extension of the European Soil Database*; Yigini, Y., Panagos, P., Montanarella, L., Eds.; Publications Office of the European Union, European Commission Joint Research Centre Institute for Environment and Sustainability: Luxembourg, 2013; Chapter II; pp. 16–36. [CrossRef]

32. Feyziyev, F.; Babayev, M.; Priori, S.; L'Abate, G. Using visible-near infrared spectroscopy to predict soil properties of Mugan Plain, Azerbaijan. *J. Soil. Sci.* **2016**, *6*, 52–58. [CrossRef]

33. Mammadov, E.; Denk, M.; Riedel, F.; Lewinska, K.; Kaźmierowski, C.; Glaesser, C. Visible and Near-Infrared Reflectance Spectroscopy for Assessment of Soil Properties in the Caucasus Mountains, Azerbaijan. *Commun. Soil Sci. Plant Anal.* **2020**, *51*, 2111–2136. [CrossRef]

34. Ismayilov, A.; Feyziyev, F.; Mammadov, E.; Babayev, M. Soil organic carbon prediction by Vis-NIR Spectroscopy: Case Study the Kur-Aras Plain, Azerbaijan. *Commun. Soil Sci. Plant Anal.* **2020**, *51*, 726–734. [CrossRef]

35. *IUSS Working Group WRB, World Reference Base for Soil Resources 2014, Update 2015, International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; World Soil Resources Reports No. 106; FAO: Rome, Italy, 2015.

36. Mammadov, E.; Nowosad, J.; Glaesser, C. Estimation and mapping of surface soil properties in the Caucasus Mountains, Azerbaijan using high-resolution remote sensing data. *Geoderma Reg.* **2021**, *26*, e00411. [CrossRef]

37. Gee, G.; Bauder, J. Physical and Mineralogical Methods: Particle-Size Analysis. In *Methods of Soil Analysis*, 2nd ed.; Agronomy Monograph, Part 1; Klute, A., Ed.; American Society of Agronomy/Soil Science Society of America: Madison, WI, USA, 1986; pp. 383–411.

38. Sparks, D.L.; Fendorf, S.E.; Toner, C.V.; Carski, T.H. Kinetic Methods and Measurements. In *Methods of Soil Analysis*; Part 3—Chemical Methods; Sparks, D.L., Ed.; SSSA Book Series: Madison, WI, USA, 1996; pp. 1275–1307.

39. Nelson, D.; Sommers, L. Total Carbon, Organic Carbon, and Organic Matter. In *Methods of Soil Analysis*, 2nd ed.; Part 2; Sparks, D.L., Ed.; American Society of Agronomy/Soil Science Society of America: Madison, WI, USA, 1996; pp. 961–1010. [CrossRef]

40. Bloesch, P.M. Prediction of the CEC to clay ratio using mid-infrared spectroscopy. *Soil Res.* **2012**, *50*, 1–6. [CrossRef]

41. Ameyan, O. Surface soil variability of a map unit on Niger river alluvium. *Soil Sci. Soc. Am. J.* **1984**, *50*, 1289–1293. [CrossRef]

42. Evans, J. *Straightforward Statistics for the Behavioral Sciences*; Brooks/Cole Publishing: Pacific Grove, CA, USA, 1996; 600p.

43. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019; Available online: https://www.cran.r-project.org (accessed on 1 November 2021).

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

45. Savitzky, A.; Golay, M. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

46. Clark, R.; Roush, T. Reflectance spectroscopy quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.* **1984**, *89*, 6329–6340. [CrossRef]

47. Geladi, P.; McDougall, D.; Martens, H. Linearization and scatter correction for near infrared reflectance spectra of meat. *Appl. Spectrosc.* **1985**, *39*, 491–500. [CrossRef]

48. Barnes, R.; Dhanoa, M.; Lister, S. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777. [CrossRef]

49. Stevens, A.; Ramirez-Lopez, L. An Introduction to the *Prospectr* Package in R. Available online: https://CRAN.R-project.org/package=prospectr/vignettes/prospectr.html (accessed on 9 June 2014).

50. Geladi, P.; Kowalski, B. Partial least-squares regression: A tutorial. *Anal. Chimica. Acta.* **1986**, *185*, 1–17. [CrossRef]

51. Wold, S.; Eriksson, L.; Trygg, J.; Kettaneh, N. *The PLS Method—Partial Least Squares Projections to Latent Structures—And Its Applications in Industrial RDP (Research, Development, and Production)*; Technical Report for PLS in Industrial RPD (Research, Development, and Production); Umea University: Umeå, Sweden, 2004; pp. 1–44.

52. Efron, B.; Tibshir, R. *An Introduction to the Bootstrap*; Chapman Hall/CRC Press: Boca Raton, FL, USA, 1994; 430p.

53. Akaike, H. Akaike's Information Criterion. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 25–29.

54. Williams, P. *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P., Norris, K., Eds.; American Association of Cereal Chemists: Saint Paul, MN, USA, 2001; pp. 35–55.

55. Chong, I.; Jun, C. Performance of some variable selection methods when multi-collinearity is present. *Chemom. Intell. Lab. Syst.* **2005**, *78*, 103–112. [CrossRef]

56. Wold, S.; Trygg, J.; Berglund, A.; Antti, H. Some recent developments in PLS modelling. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 131–150. [CrossRef]

57. Tenenhaus, M. *La Régression PLS*; TECHNIP: Paris, France, 1998; 264p. (In French)

58. Mevik, B.; Wehrens, R. The *pls* package: Principal component and partial least squares regression in R. *Jour. Stat. Softw.* **2007**, *18*, 1–24. [CrossRef]

59. Liland, K.; Mehmood, T.; Sæbø, S. The plsVarSel package: Variable selection in partial least squares regression in R. *Jour. Stat. Softw.* **2017**, *0.9.4*, 1–23.

60. Rowley, M.C.; Grand, S.; Verrecchia, E.P. Calcium—Mediated stabilization of soil organic carbon. *Biogeochemistry* **2018**, *137*, 27–49. [CrossRef]

61. Golia, E.E.; Diakoloukas, V. Soil parameters affecting the levels of potentially harmful metals in Thessaly area, Greece: A robust quadratic regression approach of soil pollution prediction. *Environ. Sci. Pollut. Res.* **2021**, 1–9. [CrossRef]

62. Sun, W.C.; Zhang, X.; Sun, X.J.; Sun, Y.L.; Cen, Y. Predicting nickel concentration in soil using reflectance spectroscopy associated with organic matter and clay minerals. *Geoderma* **2018**, *327*, 25–35. [CrossRef]

63. Isaksson, T.; Næs, T. The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl. Spectrosc.* **1988**, *42*, 1273–1284. [CrossRef]

64. Shi, T.Z.; Liu, H.Z.; Chen, Y.Y.; Fei, T.; Wang, J.J.; Wu, G.F. Spectroscopic diagnosis of arsenic contamination in agricultural soils. *Sensors* **2017**, *17*, 1036. [CrossRef] [PubMed]

65. Shepherd, K.; Walsh, M. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988–998. [CrossRef]

66. Udelhoven, T.; Emmerling, C.; Jarmer, T. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least square regression: A feasibility study. *Plant Soil* **2003**, *251*, 319–329. [CrossRef]

67. Leone, A.; Viscarra Rossel, R.; Amenta, P.; Buondonno, A. Prediction of soil properties with PLSR and vis-NIR spectroscopy: Application to Mediterranean Soils from Southern Italy. *Curr. Anal. Chem.* **2012**, *8*, 283–299. [CrossRef]

68. Bartholomeus, H.; Schaepman, E.; Kooistra, L.; Stevens, A.; Hoogmoed, B.; Spaargaren, O. Spectral reflectance based indices for soil organic carbon quantification. *Geoderma* **2008**, *145*, 28–36. [CrossRef]

69. Ben Dor, E.; Inbar, Y.; Chen, Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sens. Environ.* **1997**, *61*, 1–15. [CrossRef]

70. Rathod, P.H.; Rossiter, D.G.; Noomen, M.F.; van der Meer, F.D. Proximal spectral sensing to monitor phytoremediation of metal-contaminated soils. *Int. J. Phytorem.* **2013**, *15*, 405–426. [CrossRef]

71. Viscarra Rossel, R.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]

72. Clark, R. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In *Manual of Remote Sensing*; Rencz, A., Ed.; Wiley & Sons: New York, NY, USA, 1999; Volume 3, pp. 3–58.

73. Ben Dor, E.; Irons, J.; Epema, G. Soil Reflectance. In *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*; Rencz, A., Ed.; Wiley & Sons: New York, NY, USA, 1999; Volume 3, pp. 111–188.

74. Gaffey, S. Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 microns): Calcite, aragonite, and dolomite. *Am. Mineral.* **1986**, *71*, 151–162.

75. Strens, R.; Wood, B. Diffuse reflectance spectra and optical properties of some iron and titanium oxides and hydroxides. *Mineral. Mag.* **1977**, *43*, 347–357. [CrossRef]

76. Goetz, A.; Chabrillat, S.; Lu, Z. Field reflectance spectrometry for detection of swelling clays at construction sites. *Field Anal. Chem. Tech.* **2001**, *5*, 143–155. [CrossRef]