Article

Soil Order-Land Use Index Using Field-Satellite Spectroradiometry in the Ecuadorian Andean Territory for Modeling Soil Quality

Susana Arciniegas-Ortega [1,2,*], Iñigo Molina [1,*] and Cesar Garcia-Aranda [1,*]

1   Department of Land Surveying and Cartography Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

2   Department of Environmental Engineering, Universidad Central del Ecuador, Quito 170129, Ecuador

*   Correspondence: srarciniegas@uce.edu.ec (S.A.-O.); inigo.molina@upm.es (I.M.); cesar.garciaa@upm.es (C.G.-A.)

Abstract: Land use conversion is the main cause for soil degradation, influencing the sustainability of agricultural activities in the Ecuadorian Andean region. The possibility to identify the quality based on the spectral properties allows remote sensing methods to offer an alternative form of monitoring the environment. This study used laboratory spectroscopy and multi-spectral images (Sentinel 2) with environmental covariates (physicochemical parameters) to find an affordable method that can be used to present spatial prediction models as a tool for the evaluation of the quality of Andean soils. The models were developed using statistical techniques of logistic regression and linear discriminant analysis to generate an index based on soil order and three indexes based on the combination of soil order and land use. This combined approach offers an effective method, relative to traditional laboratory methods, to derive estimates of the content and composition of soil constituents, such as electrical conductivity (CE), organic matter (OM), pH, and soil moisture (HU). For Mollisol index.3 with Páramo land use, a value of organic matter (OM) $\geq$8.6% was obtained, whereas for Mollisol index.4 with Shrub land use, OM was $\geq$6.1%. These results reveal good predictive (estimation) capabilities for these soil order–land use groups. This provides a new way to monitor soil quality using remote sensing techniques, opening promising prospects for operational applications in land use planning.

Keywords: remote sensing; soil quality; soil properties; indices; Andean region

1. Introduction

Soil is the main natural resource for food and energy production [1]. It controls the movement of water in the landscape and functions as a biological filter for the possible leaching of pollutants into environmental spheres [2]. However, soil can be degraded by chemical and physical processes, which reduces its ability to function as a base for the development of a healthy layer for vegetation. Therefore, acknowledging soil conditions by the effects on vegetation can represent site conditions [3,4].

Gholizadeh and Kopačková (2019) [1] considered that conventional methods of soil health evaluation in large areas involve several expensive and time-consuming variables such as collection of field data, chemical analysis in a laboratory, and geostatistical interpolation. Alternately, several studies have shown the possibility of characterizing soils and identifying their quality by correlating both physicochemical and spectral parameters [5]. Therefore, the use of remote sensing spectrometry products in environmental evaluation studies offers a complementary alternative to in situ monitoring procedures to aid in research, control, and monitoring of the soil component. The application field for these tools in the soil component is extensive [6] through the study of soil characteristics such

as reflectance, degradation, and possible polluting agents with the processing of satellite images permitting the inspection and monitoring of large areas in a fixed time and place [7].

In Ecuador, the properties and pedogenetic processes of soil have been studied in terms of rock type, geomorphology, taxonomic classification, and soil order. An increasing breach between the available information on the main soils and their quality [8] leads to understanding the condition of the soil to allow for the planification of healthy and sustainable territories, as determined by goals 12 and 15 of the Sustainable Development Goals (SDG) [9]. These goals correspond to responsible consumption and production, and life on land. Therefore, it is necessary to determine the quality of soil to develop fast, feasible, and affordable estimation methods for monitoring and assessing areas.

In the study area, the predominate soils are Andisol and Mollisol, which originate from weathering of volcanic material (ash) [10]. These relatively young soils can convey high agricultural potential [11]. Andisols, also known as *páramos*, are clay loam soils capable of retaining enormous amounts of water; on the contrary, Mollisols are fertile soils with a high organic matter content that cover approximately 70% of the Cayambe canton, a political–administrative unit of Ecuador, where the research's basin is located. However, the lack of land use and occupation policies has caused the expansion of agricultural activity boundaries [12], causing the loss of the *páramo*. The main goal is to understand whether there is any relationship between the spectra measured in the samples collected in field with the corresponding bands measured by satellite, combining physicochemical analysis to quantify and model the quality of Andean soils caused by agricultural activity. This will be achieved by (i) compiling and analyzing the physicochemical parameters of soil based on quality standards for agricultural activities, obtaining indices that classify the soils based on their order (Andisol and Mollisol) and use, and (ii) determining whether the soil is associated with some of the physicochemical qualities considered, validating land use and order models based on field reflectance data, satellite reflectance, and physicochemical qualities.

There are several studies based on national models capable of predicting spectra limited in an infrared laboratory with statistical algorithm analysis [13,14]. In this research, with the use of these spectroscopic methods for the evaluation of soil quality, models were developed for estimating indicators based on the combination of soil order, land use, and physicochemical characteristics, using logistic regression analysis, linear discriminant analysis, and regression trees. The approach offers a method that derives the estimates using the ratio of laboratory/satellite spectra when the soil is well represented by the calibration samples used to build the predictive models [13]. Therefore, the performance of these local models can be used in other geographic spaces by incorporating the spectra into a dataset for that area [15].

Consequently, the combination of laboratory spectroscopy and multispectral images with environmental covariates is an adequate methodological alternative to obtain models that are adjusted for the prediction of the quality of Andean soils, independently of other methodological approaches that have been used [16–19].

2. Materials and Methods

2.1. Study Area

This study was done at Río Blanco basin, located in the Cayambe canton, Pichincha province, Ecuador (Figure 1), where agricultural activities are mainly related to the dairy–floricultural corridor via the cultivation of exportation products, such as roses and summer flowers (17.32 km^2), livestock for milk production (399.201 km^2), and, in smaller quantities, agricultural activity, especially high Andean crops (10.984 km^2) [12]. Unfortunately, in Río Blanco basin, no studies have been aimed towards soil quality. The few studies available are related to land use and occupation, risks associated with the presence of the Cayambe volcano, agriculture [20], soil rehabilitation with Cangahua [21], and watershed management [22].
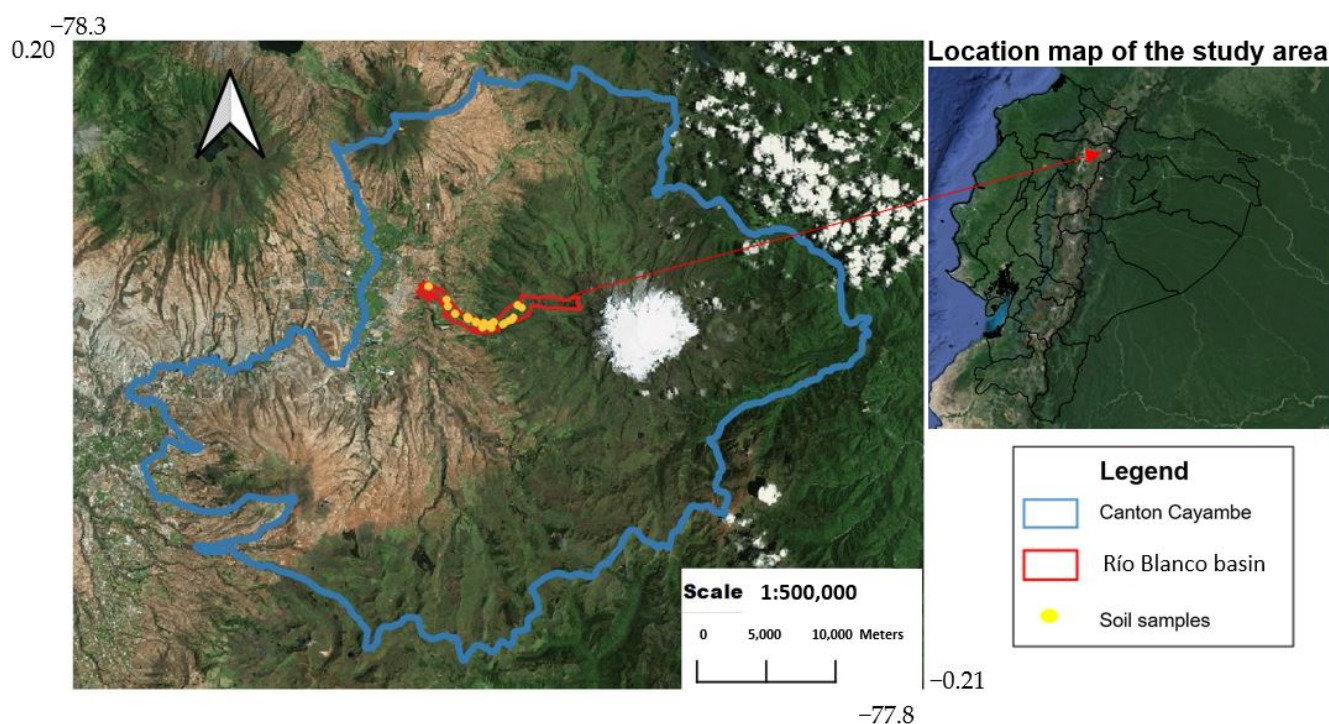
Figure 1. Geographical location of the study area and location of the soil sampling points (Google Maps).

The existing black and brown soils in the basin can retain high amounts of water and organic matter content, known as Mollisol and Andisol [10]. The soils have been influenced by the volcanic activity of the Cayambe during its genesis, causing slopes ranging from gentle to steep [7]. Vegetation coverage is characterized by the presence of cultivated grass (27%), herbaceous and shrubby moorland (45.6%), flowers and short-cycle crops (16.5%), and eucalyptus trees (8.1%) [10].

2.2. Sentinel-2 Satellite Images

The satellite imagery used for this research was acquired by the Sentinel-2 (S2) satellite constellation (2A and 2B Earth Observing Missions) launched in 2015 and 2016, which has been used extensively for monitoring land cover and vegetation [23,24]. The S2 satellites are identical and operate in a sun-synchronous orbit at a mean altitude of 786 km. The main S2 payload is a multi-spectral instrument (MSI), which is a push-broom sensor that registers the radiation reflected from the Earth passing through the atmosphere in 13 spectral bands distributed in four bands at 10 m, six bands at 20 m, and three bands at 60 m spatial resolution. Figure 2 depicts the range and spectral response functions of the S2A/S2B MSI instruments for these bands [25]. The S2 satellites have a swath width of 290 km. The visible bands (VIS) B1, B2, B3, B4 at 10 m resolution; near-infrared bands (NIR) B5, B6, B7, B8A at 20 m resolution and B8 at 10 m; and shortwave infrared (SWIR) B11 and B12 are most useful for retrieving geophysical surface parameters [26].

Meanwhile, the 60 m resolution bands are used for atmospheric corrections, which are of crucial importance for most EOS applications and permit the development and evaluation of robust atmospheric correction algorithms such as Sen2Cor. In the study area, located inside a swath overlap, the revisit frequency of each satellite is four to five days in an 11° forward-looking view angle; the presence of two identical satellites allows a geometric revisit time between two and three days, supporting near-continuous monitoring of vegetation and land surface processes [26].

The imagery used in this research was acquired on 16 July 2018 by the Sentinel 2B platform (see Supplementary Materials). The Level-2A product was used after the

processing carried out in the Level-1C product available in the Open Access Hub of the European Space Agency (ESA) DataHUB server [27]. The identifier of the product is referenced as S2B_MSIL1C_20180716T153619_N0206_R068_T17NRA_20180716T202613. Furthermore, in the text, it will be referred to as S2BL2A.
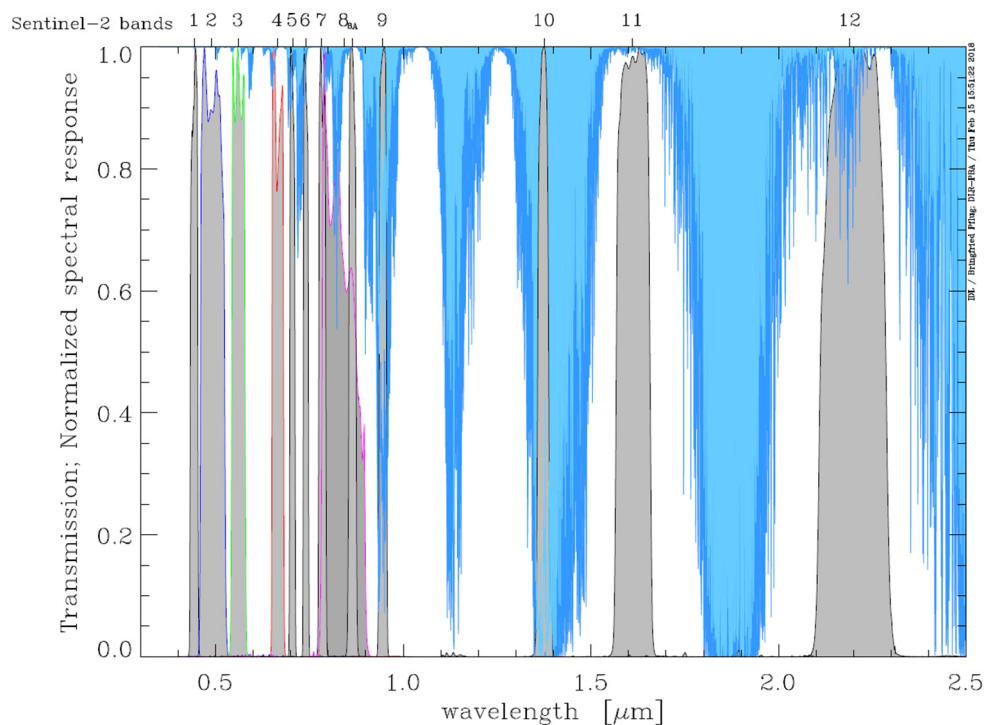

Figure 2. S2A and S2B-MSI spectral response functions (SRF) [28].

2.3. Soil Sampling

Soil samples were collected during four field trips during the months of June, July, and August 2018. The samples were placed in hermetically sealed sleeves for physicochemical analysis (Figure 3a,b), and in two bulk density cylinders (Figure 3c) for spectroradiometry analysis. The cylinder lids were covered with geomembranes to keep the samples unaltered, as explained by Yánez and Arciniegas (2019) [7], to comply with the provisions of Section 2.5.
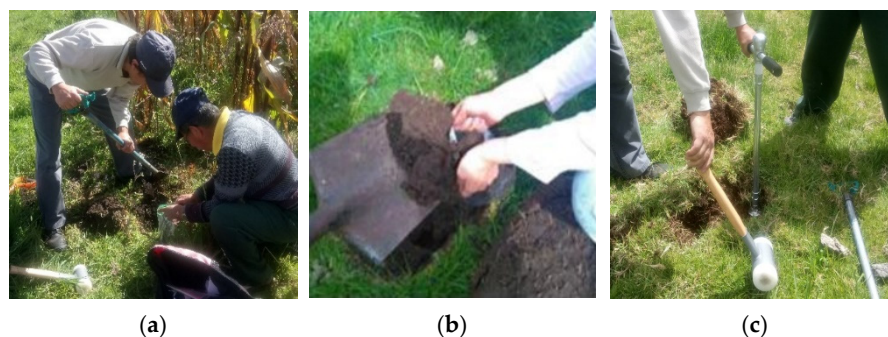

(a)   (b)   (c)

Figure 3. Soil sample collected in the field. (**a**) Shows site preparation. (**b**) Shows sample put in sealed sleeves for physicochemical analysis (**c**) Shows bulk density cylinders for spectroradiometry analysis.

A total of 36 surface soil samples (0–10 cm) were collected from the study site using core drilling, cylinders, a stainless-steel shovel, and a GARMIN GPSMAP 62sc handheld navigator (accurate to within ±4 m). After removing vegetation from the soil surface in a quadrant of approximately 30 × 30 cm, 1 kg of soil sample was collected from each

sampling point. The number of samples was calculated by the type of composite sample according to Ecuadorian environmental standards (Book VI, Annex2) [29], establishing two samples per homogeneous zone, and one sample for zones ID02 and ID08, which presented collection problems.

2.4. Physicochemical Parameter Measurements

To determine the quality of soil, the physical, chemical, and biological components of the soil and their interactions must be considered; despite the different measurements possible, not all parameters are relevant for the soil in a particular scenario [30]. In this study, the physicochemical components were selected considering two criteria: The first was based on the reflectivity of the soils that are conditioned to organic matter, that interfere with the spectral curves [31], and the second according to Friedman et al. 2001 [32], who set a minimum number of indicators for agricultural activities, due to human use and management. The parameters measured were soil moisture, pH, electrical conductivity (CE), and organic matter (OM). Other parameters, such as heavy metals, could not be measured for economic reasons.

For the 36 soil samples, the analysis was carried out as follows: For soil moisture, pH, and conductivity, the methods established in NOM-021-RECNAT-2000 (Mexican official standard that establishes the specifications of fertility, salinity, and soil classification) were used [33]. The AS-05 gravimetric method was applied to soil moisture, the AS-02 electrometric method was applied to pH, and the AS-20 method was applied to electrical measure conductivity.

For the determination of OM, the soil samples were sent to the certified soil, foliar, and water laboratory of the Agency for the Regulation and Control of Phytosanitary and Zoosanitary (Agrocalidad) of Ecuador, where the volumetric PEE/SFA 09 method (Walkley Black's analytical method consisting of wet oxidation of the soil sample) was applied [34].

2.5. Spectral Measurements in Laboratory and Satellite Image Sentinel-2

Spectral analysis of the soil samples was carried out in the Ecuadorian Space Institute (IEE) laboratory, which facilitated the use of the ASD FieldSpec4 spectroradiometer (Analytical Spectral Devices). The equipment has a spectral resolution of 3 nm in the range of 350–1000 nm and 10 nm in the range of 1001–2500 nm. In this regard, the spectral bands of the MSI sensor onboard satellite S2 are within the spectral range of the ASD instrument.

In the laboratory, the soil spectra were collected using a small spectralon placed in the HiBrite MudLight device with an optical fiber connected to generate an artificial light source (Figure 4). The ASD FieldSpec4 spectroradiometer was used to obtain soil spectral data in the laboratory. The measurement protocol followed the methodology described in Figure 5 [35].

Figure 4. Location of cylinder with respect to HiBrite MudLight device on support.
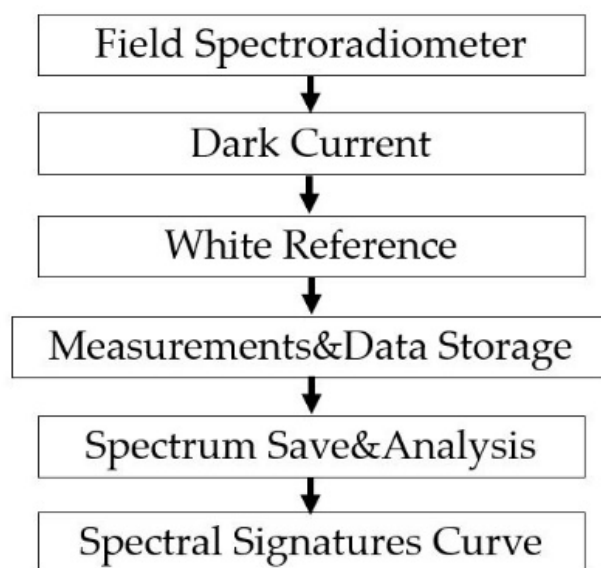
Figure 5. Methodology for the generation of spectral signatures [35,36].

After the common configuration and control settings, i.e., output folders, connecting optical fiber, etc., the appropriate integration time was set given the lighting conditions to optimize the measurements. Subsequently, the dark current was also recorded. Then, a reference target or white reference (spectralon) was measured until a horizontal line with a reflectance value of 1 was presented. Laboratory soil spectra measurements followed these pre-operation phases, and the resulting measured spectra were processed with the corresponding software (ViewSpec Pro). An important issue that also needed to be addressed was that of the signal-to-noise ratio, or SNR, during measurements, which is related to the signal component. The spectra measurement procedure was performed using the principle of a continuous fiber optic cable, as specified in [36]. This technique has the advantage of significantly degrading the SNR, as it avoids interactions with other media between the recording device and the sample. Soil measurements were carried out by controlling the direction of the optical fiber to always point towards the target in order to avoid anisotropy effects. Subsequently, an average of all measurements was made to obtain the spectrum. Finally, the displays of the spectra that were selected were shown.

The S2BL2A product specified in 2.2 was used for the satellite image, with bottom-of-atmosphere (BOA) reflectance, in which only the bands matching the S2 product were considered. The centroid of each pixel was determined to obtain the spectral values per band on a 20 × 20 m grid.

These laboratory spectroradiometry measurements and satellite images were used to establish indices (the table in Section 2.8.1) that compare the physicochemical parameters of the soil with the spectral bands of the S2-MSI sensor. Since it was agreed to only use the equipment supplied in the Laboratory of the Ecuadorian Space Institute, in-field spectroradiometry measurements were not possible. The measurements of the spectra were carried out only in the laboratory, as in [37–39], generating a different model to those already known to evaluate soil quality [40,41].

2.6. Land Use/Soil Order Dataset

Before the processing specified in Section 2.8, a dataset was built based on the in-field soil samples (spectra and physicochemical parameters) and reflectance per homogeneous area in the S2BL2A satellite image. These data consisted 345,408 observations as a product of the combination of two datasets and constituted the geographic population of the area under study. The data were then treated based on the combination of the variables "land order" and "land use", whose total set of observations was as follows (Table 1). The

combinations shown in Table 1 allowed different models to be established based on soil order and land use.

Table 1. Soil samples classified according to the order of the soil and by land use.

| Land Use | Order of the Soil | | Total |
	Andisol	Mollisol	
Agricultural	0	4284	4284
Shrub	968	13,496	14,464
Forest	40	18,476	18,516
Páramo *	188,760	38,072	226,832
Pasture	9892	71,420	81,312
Total	199,660	145,748	345,408

* A *páramo* is a fragile neotropical high mountain ecosystem. In Ecuador it has an average height of 3300 m above sea level.

Model 1 was established from the dataset by applying a simple random method of 5% of the population to compare what was provided by each sample until one dataset was left as a result of the information provided between one model and another being the same. The data were divided into training and testing groups. The training dataset consisted of 70% of the total number of observations in the sample, and the test dataset consisted of 30% of the total number of observations in the sample. With the dataset the model was elaborated and later executed.

Model 2 was developed from the dataset by selecting those corresponding to the Andisol soil order (Table 1), with 199,660 observations. Three categories were selected: Shrub, Páramo, and Pasture, leaving Forest out of the analysis due to its low frequency. Composite samples of 70% of each land use category were randomly selected, thus leaving a composite sample of 139,734 random observations, which were part of the training dataset. Therefore, the test dataset contained the remaining 30% of the 59,886 observations.

For the Mollisol order, considering the data's trend, based on the geographical distribution of the different land uses (Figure 6), it was observed that some of the soils had a particular use owing to their nature and population growth, among other characteristics of the area.
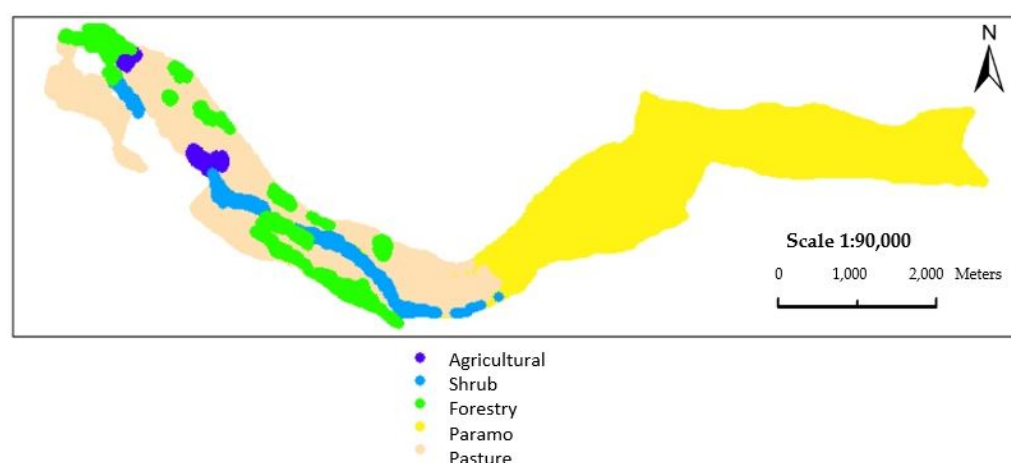

Figure 6. Geographical distribution of land uses.

The dataset corresponding to the Mollisol soil order (Table 1), with 145,748 observations, was considered to form Model 3 from a stratified random sample of 70% of the total. This allowed us to consider two subsets based on land use, resulting in two more models. The first subset (Model 3), called "Mollisol 1", was made up of Forest, Páramo, and Pasture; the second subset (Model 4), "Mollisol 2", was composed of Agriculture and Shrub. Each

training dataset represented a 70% stratified random sample based on the categories of Mollisol 1 (Table 2) and Mollisol 2 (Table 3). Consequently, the test dataset included the remaining 30% of the data. For Mollisol 1 it was 38,391, and for Mollisol 2, 5335.

Table 2. Training sample to obtain the discriminant function in the classification of land use of the Mollisol 1 order.

Land Use	Forest	Páramo	Pasture	Total
Samples	12,933	26,650	49,994	89,577
Ratio	0.1444	0.2975	0.5581	1.00

Table 3. Training sample to obtain the discriminant function in the classification of land use of the Mollisol 2 order.

Land Use	Agriculture	Shrub	Total
Samples	2998	9447	12,445
Ratio	0.2409	0.7591	1.00

2.7. Homogeneous Zones

The area of the Rio Blanco basin was extracted from the hydrological database of the Cayambe canton. Land use was extracted from the national productive systems database [10], and slope data and soil order data were obtained from a geopedology base map [10]. The multi-criteria technique was applied to create homogeneous zones [7] considering the following criteria: soil order and land use. This made it possible to establish a spatial analysis unit for correlation evaluation tests, determining the number of samples to be taken in the field. In total, within the Rio Blanco basin, 31 zones with similar characteristics were identified in previous work [7], of which 19 were analyzed for easy access in the study area (Figure 7).
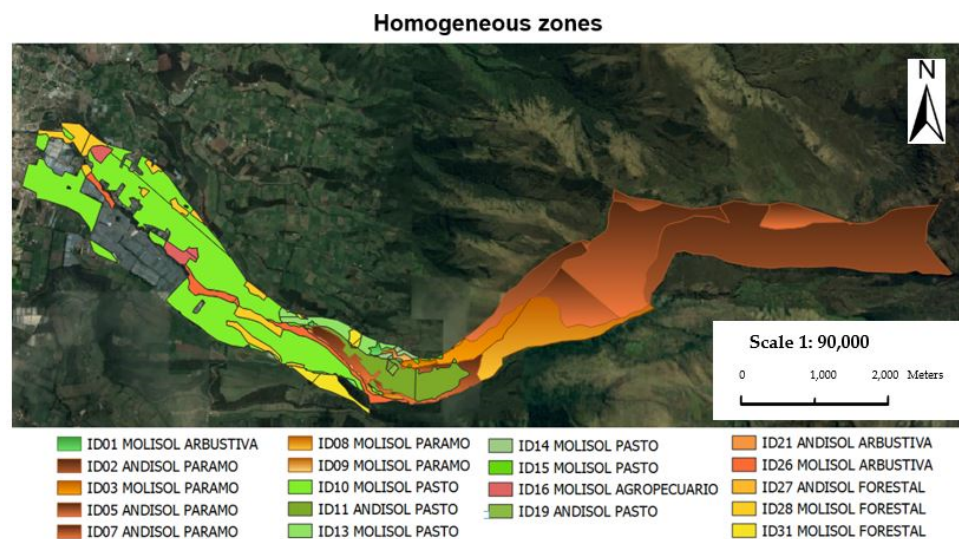

Figure 7. Homogeneous areas analyzed in the study.

Using the physicochemical laboratory analysis results, the Thiessen polygons were calculated in each homogeneous zone from the soil-sampling points. Likewise, the topology of spatial relations (intersection and inside) was applied [42] with the spectral values per satellite band to identify the spatial relationship of the satellite spectral values concerning the physicochemical characteristics of each homogeneous zone [43].

The Mexican (NOM-021-RECNAT-2000) [33] and Ecuadorian (Book VI, Annex 2) environmental standards [29] were also considered to establish homogeneous areas as a

unit of spatial analysis. This allowed sampling points to be defined based on the criteria of slope, soil order, and land use, resulting in 19 homogeneous areas (Figure 7).

2.8. Regression and Spatial Analysis

To identify the relationship between the spectral behavior of soil in the laboratory and satellite images associated with the physicochemical qualities of soils, we proceeded to apply regression techniques based on the theory of machine learning, which makes them useful for predicting outcomes, identifying patterns, and making decisions with minimal human intervention [44]. Subsequently, discriminant analysis techniques, geostatistics, and non-parametric models (regression trees) were applied.

This study started by creating a logistic regression model (Figure 8) with the dependent variable "soil order" (Andisol, Mollisol), and as explanatory variables, the reflectance levels of the spectral behavior of soil in laboratory and satellite images, called Model 1, to later select those variables that were statistically significant, from which this model generated the soil index required (Table 4). The basic form of the logistic regression model is as follows (Equation (1)) [45]:

$$\text{logit}(yy) = \beta o + \beta 1 x 1 + \beta 2 x 2 + \cdots + \beta i x i \tag{1}$$

where

y: soil order dependent variables Andisol and Mollisol;
xi: independent variables (predictors): wavelengths of the spectroradiometer convoluted by the spectral response of soils in the laboratory (B04c, B05c, B06c, B07c, B08c, B08Ac) and bands corresponding to the MSI sensor of satellite S2 (B04s, B05s, B06s, B07s, B08s, B08As);
βo: constant (intercept);
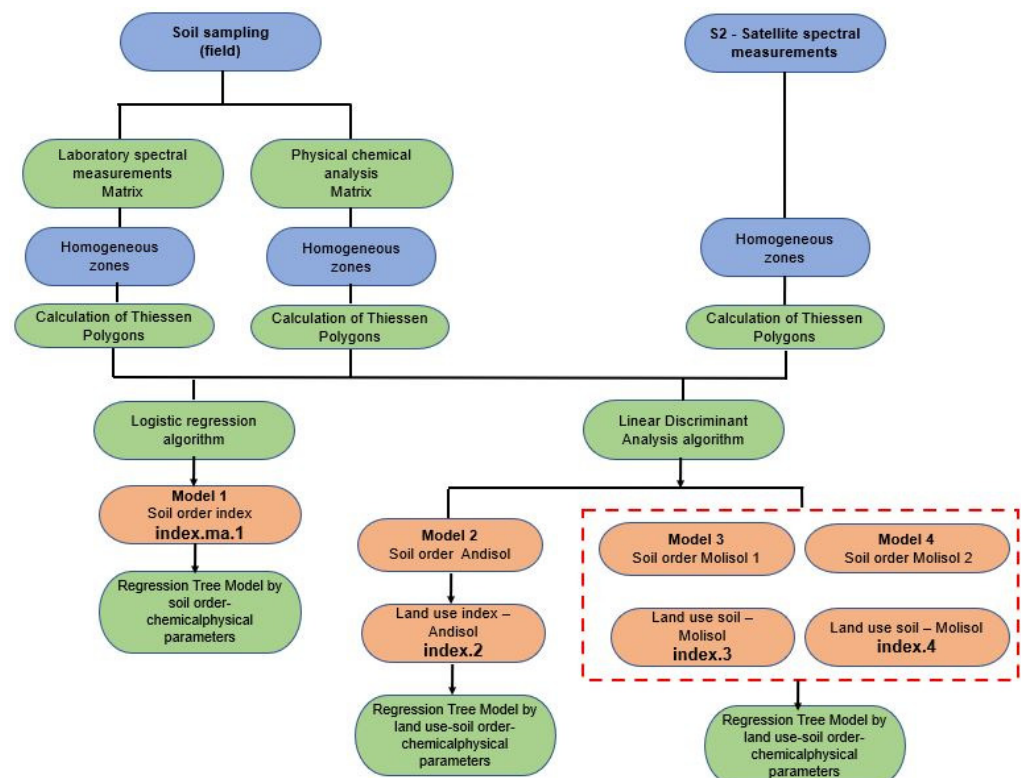βi: coefficients of predictor variables.

Figure 8. Flowchart indicating the methodological steps implemented to create the model, indexes, and predictions of physicochemical parameters.

Table 4. Models and indices by soil order and soil order–land use.

Model	Index	Description	Dependent Variable	Independent Variable
Model 1	index.ma.1	Separates soil order: Mollisol, Andisol	Mollisol Andisol	Field bands, Satellite bands,
Model 2	index.2	Separates land use of the Andisol order	Andisol	Field bands, Satellite bands, Shrub, Páramo, Pasture
Model 3	index.3	Separates land use of the Mollisol order	Mollisol (Mollisol 1 group) for land use: Forest, Páramo, Pasture	Field bands, Satellite bands, Forest, Páramo, Pasture
Model 4	index.4	Separates land use of the Mollisol order	Mollisol (Mollisol 2 group) for land use: Agriculture, Shrub	Field bands, Satellite bands, Agriculture, Shrub

Several logistic regression trials with the dependent variable of soil order were tested. The statistically significant variables ($p \leq 0.05$) were verified until the most highly significant set of variables was obtained. The statistically significant model consisted of a combination of independent variables related to the reflectance levels of the soil spectra in the laboratory and satellite image.

The land use variable was related to the taxonomic classification of soil, so when considering only this variable, there was no way to separate the different land uses and generate an index that allowed us to estimate the physicochemical characteristics based on the land use variable of soil. To resolve this difficulty, the soil order variable was considered, and within each of these categories—Andisol and Mollisol—the land use variable that yielded the best classification levels by soil order was analyzed, as discussed in Section 2.6. Considering land use as the dependent variable, a statistical methodology known as linear discriminant analysis was applied (Figure 8) [46], which allowed a linear combination of variables to be identified that could be used to determine the group to which each individual belonged. In this case, the individuals were identified as homogeneous areas in which different soil samples were collected to develop Model 2, Model 3, and Model 4 (Table 4).

2.8.1. Obtaining Index

The standardization of the coefficients from the logistic regression model (Model 1) followed a pattern where each coefficient was divided by the sum of its coefficients, in such a way that the sum of the coefficients of the index was equal to 1 (Equation (2)).

$$index.ma.1 = k + cso[i] * xi + \ldots \ldots \tag{2}$$

where

k: constant;
cso[i] = standardized soil order coefficients;
xi: independent variables (predictors): wavelengths of the spectroradiometer convoluted by the spectral response of soils in the laboratory (B04c, B05c, B06c, B07c, B08c, B08Ac) and bands corresponding to the MSI sensor of satellite S2 (B04s, B05s, B06s, B07s, B08s, B08As).

From the linear discriminant function, the coefficients of the models were standardized by land use by soil order of the models: Model 2, Model 3, and Model 4, where each coefficient of the model was divided by the sum of its coefficients in such a way that the sum of the coefficients of the indices (Index 2, Index 3, Index 4) was equal to 1 Equation (3).

$$index = cso[1] * x1 + \ldots . + cso[i] * xi \tag{3}$$

where

cso[i] = standardized coefficients of land use by soil order;

xi: independent variables (predictors): wavelengths of the spectroradiometer convoluted by the spectral response of soils in the laboratory (B02c, B03c, B04c, B05c, B06c, B07c, B08c, B08Ac, B011c, B012c) and bands corresponding to the MSI sensor of satellite S2 (B02s, B03s, B04s, B05s, B06s, B07s, B08s, B08As, B011c, B012c).

In this case, the data were treated based on the combination of soil order and land use, and the entire set of observations is shown in Table 1.

An association analysis was applied between each physicochemical parameter and the indices, which determined no linear trend between the different pairs of variables that were compared. The geostatistical surface was created using the inverse distance weighted (IDW) method, which assumes that closer objects are similar to those far apart. Therefore, any unknown location will probably have an equal value to the nearest known locations [47]. It was possible to determine how the order of soil and land use were spatially distributed in terms of probability and to establish predictions of each physicochemical parameter through a set of non-parametric models known as decision trees [48], where the dependent variable can be categorical or numeric. The regression tree model was applied since the dependent variable was each of the calculated indices and was numerical, and the explanatory variables (covariate) were the order of the soil and the order–land use, together with each physicochemical parameter, considered as independent variables, as shown examples in Table 5, which are explained in Section 3.5. The space defined by the regression tree models, as part of a non-parametric analysis, consisted of dividing the predictor space into boxes (regions) [49]. For example, the areas were a function of Index 1 and the order of soil to estimate the physicochemical organic matter (OM) parameter to make the prediction. Consequently, a regression tree model was generated to describe the association between each index and each physicochemical parameter, as shown in Figure 8.

Table 5. Regression trees.

Regression Trees	Dependent Variable	Covariates (Soil Order and/or Land Use)	Physicochemical Parameter Covariate (Examples)
ARMAH	index.ma.1	Soil order: Andisol—Mollisol	Soil moisture
ARUSAMO	index.2	Soil order: Andisol Land use: Shrub, Páramo, Pasture	OM
ARUSM3MO	index.3	Soil order: Mollisol 1 Land use: Forest, Páramo, Pasture	OM
ARUSM4MO	index.4	Soil order: Mollisol 2 Land use: Shrub, Agriculture	OM

For a better understanding, the general methodological framework was divided into three parts, as shown in Figure 8. All statistical and graphical calculations were performed using RStudio software [50].

2.8.2. Validation

The validity of Model 1 was tested by calculating the confusion matrix [49] to determine the classification error of the samples and the accuracy, together with the Kappa statistic to indicate the degree of agreement between the measured data and the predicted value by the model, concerning their order in Andisol or Mollisol, as well as the sensitivity and specificity, which indicate the probability of correctly classifying the soil samples from the model. The Kappa coefficient must be equal to zero. When the Kappa coefficient differs from 0, it means that the data obtained from the validation model, as predicted data, agree with the measured data used to generate the model. Sensitivity refers to the ability of a model to identify the order of soils. In contrast, specificity indicates the ability of the model to identify soil samples that do not correspond to the order of the soil to be classified.

In Model 2, among the different land uses of the order Andisol with each of the two sets of test data selected randomly, the classification errors were evaluated using the confusion matrix and the accuracy metric.

For Model 3, among land uses such as Forest, Páramo, and Pasture of soil order Mollisol, those that did not participate in the elaboration of the model needed to be classified based on Model 3 and have their corresponding confusion matrices calculated in order to determine the classification error and accuracy metric.

The same criteria applied for Model 4, between the agricultural and shrubland uses of the soil order Mollisol.

Once each model was validated, the effects of soil order and land use on each physicochemical parameter were determined; one-way analysis of variance (ANOVA) was used [49], with a dependent variable for each physicochemical parameter and an independent variable for the order and land use of the soil. Several ANOVA models were tested for soil order and other land use models within each order. For each case, the null hypothesis was that the mean of each physicochemical parameter is the same in each order of soil, or the mean of each physicochemical parameter is the same for each land use within each order, versus the alternate hypotheses indicating that at least one pair of mean values is different. The statistical decision criteria are based on a significance level of 5% ($\alpha = 0.05$); thus, if the *p*-value is less than or equal to 0.05 the null hypothesis is rejected, and it can be concluded that the effect of the independent variable is significant in relation to the mean of the dependent variable. Otherwise, if the *p*-value is greater than 0.05, then the available data do not yield enough information to conclude that the independent variable is significant in relation to some variation of the dependent variable.

3. Results

3.1. Physicochemical Analysis

The different soil samples were analyzed using standardized physicochemical methods. The soil moisture results were in a range from 12.23% to 74.99%. The areas that predominated with the highest percentage of soil moisture were moors, and those with the lowest soil moisture were forest and shrub areas. The lowest water content was observed in Mollisol.

Regarding pH, the most acidic soils corresponded to undisturbed moors, whereas the least acidic soils corresponded to cultivated pastures. The range was 4.55 to 5.76, which, according to Mexican regulations, ranges from moderately acidic to strongly acidic, and according to Ecuadorian regulations, it would be out of range.

The electrical conductivity of the soils was within the limits established by both the Ecuadorian and Mexican regulations, <200 μS/cm and <1 dS/m, respectively. The areas with the lowest electrical conductivity were moors, whereas the highest electrical conductivity was observed in grasses.

For OM, the values ranged from 2.78% to 16.06%. According to Mexican standards, the zones range from very low to very high levels for soils of volcanic origin. The zone with the lowest OM percentage was forest, followed by passage areas, and the zone with the highest OM content was *páramo* areas.

3.2. Analysis of Spectral Signatures

In this section the behavior of each band was determined with respect to the intensity of reflectance of the soil samples. For each cylinder, two spectral measurements were made in opposite sections of the same tube, and for each soil sample 10 spectral measurements per section were averaged for a single representative spectrum per homogeneous area, which resulted in graphs (Figure 9a,b) as a function of reflectance and wavelength per sample in each homogeneous area.
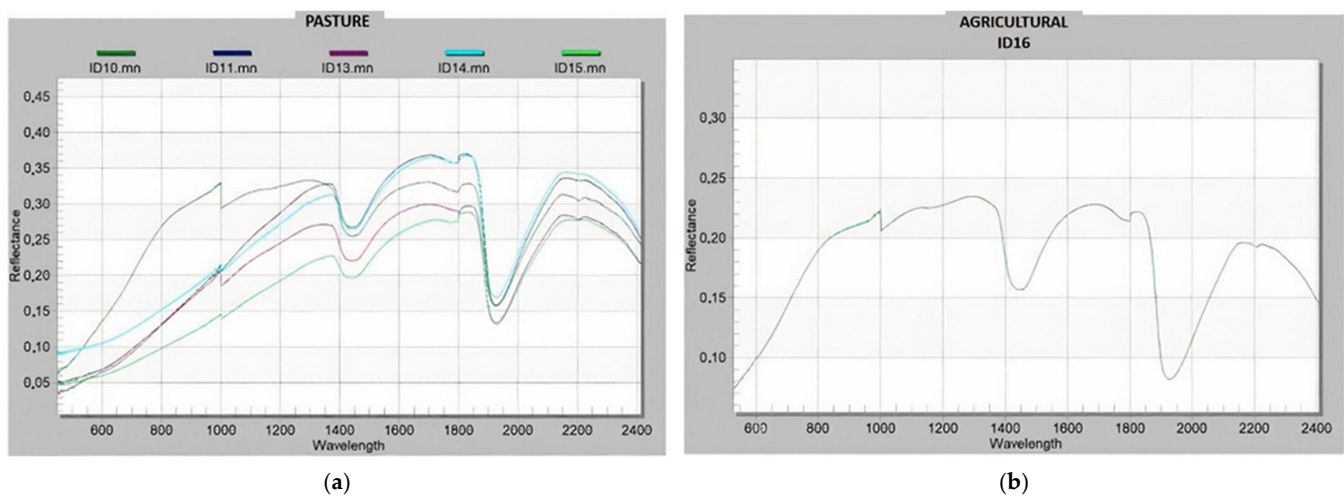
(**a**)   (**b**)

Figure 9. Spectral measurements in the laboratory according to land use. (**a**) Shows the curves by homogeneous zones ID10, ID11, ID13, ID1, ID15. (**b**) Shows the curve by homogeneous zone ID16.

The reflectance of the spectra was graphically analyzed in the laboratory to determine the behavior of the soils related to their spectral signatures of the Andisol and Mollisol orders. The spectral signatures obtained in the laboratory presented a pattern related to the typical spectral signature of soils, ranging from the visible range (VNIR) to near-infrared (NIR) to short-wave infrared (SWIR).

The graph in Figure 9a shows the pasture curves, where sample ID11 of the Andisol soil order presented the same intensity of reflectance as sample ID14 of the Mollisol order, which were the highest compared to the other samples. Sample ID10 of the Mollisol order had a medium intensity of reflectance, unlike sample ID15 of the Mollisol order and sample ID13 of the Mollisol order with lower values of reflectance intensity. This variation in the curves is related to the properties and state of these soils [51], considering the variation of each of the land uses. Thus, in the graph in Figure 9b, sample ID16 of the Mollisol order, with agricultural land use, may indicate changes in the characteristics and status of agricultural use in the months of June, July, and August.

It can be said that the graphs made a difference in the behavior of the soil order based on the associated land use.

This could be related to the reflectance records of the Sentinel-2 satellite images (Figure 10a,b) to improve spectral differences by calculating soil order indices based on land use and physicochemical parameters, as explained in the next section.
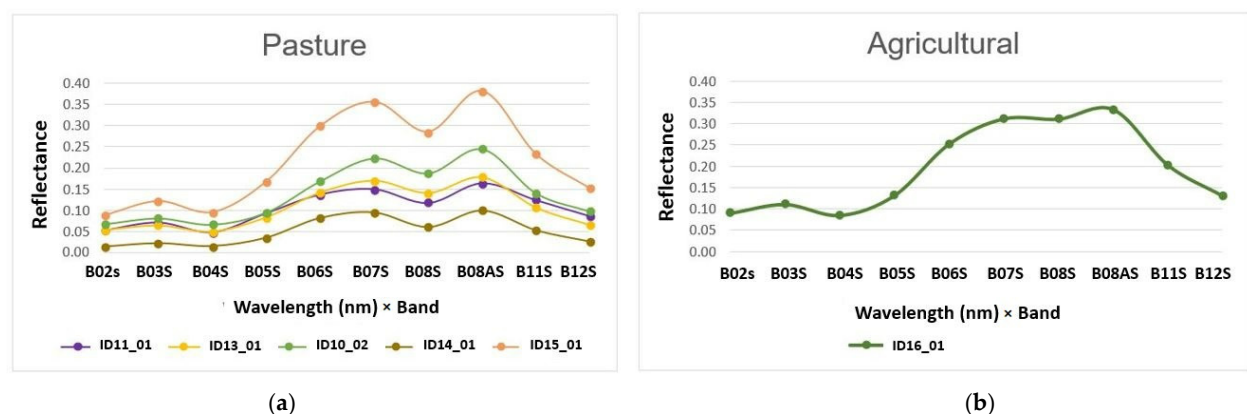

(**a**)   (**b**)

Figure 10. Spectral signatures of Sentinel 2B. (**a**) Shows the spectral signatures of Pasture by homogeneous zones. (**b**) Shows the average spectral signature of Agriculture obtained in the homogeneous zone 16 of Mollisol soil.

3.3. Development and Validation of Models Based on Soil Reflectance Levels in Laboratory and Satellite Image

3.3.1. Model 1, by the Orders of Andisol and Mollisol Soils

From the logistic regression calculation, Model 1 was obtained, whose structure is shown in Table 6.

Table 6. Logistic regression of Model 1 based on the reflectance levels of the soil spectra in the laboratory and satellite image.

Variables	Estimated Coefficients	Error Std	Value z	Pr(Z > |z|)	Decision
(Intercept)	-6.125×10^0	8.011×10^2	-76.459	$<2.0 \times 10^{-16}$	$p < 0.0001$
B04c	-7.596×10^3	5.784×10^1	-131.317	$<2.0 \times 10^{-16}$	$p < 0.0001$
B05c	9.382×10^3	1.080×10^2	86.872	$<2.0 \times 10^{-16}$	$p < 0.0001$
B06c	9.149×10^3	1.284×10^2	71.263	$<2.0 \times 10^{-16}$	$p < 0.0001$
B07c	-1.395×10^4	1.105×10^2	-126.278	$<2.0 \times 10^{-16}$	$p < 0.0001$
B08c	-3.630×10^0	4.775×10^0	-0.760	0.447	$p > 0.1000$
B08Ac	3.021×10^3	2.114×10^1	142.912	$<2.0 \times 10^{-16}$	$p < 0.0001$
B04s	5.655×10^0	1.151×10^0	4.913	8.96×10^{-7}	$p < 0.0001$
B05s	4.412×10^1	1.417×10^0	31.131	$<2.0 \times 10^{-16}$	$p < 0.0001$
B06s	-5.892×10^1	1.968×10^0	-29.941	$<2.0 \times 10^{-16}$	$p < 0.0001$
B07s	-5.889×10^1	2.164×10^0	-27.216	$<2.0 \times 10^{-16}$	$p < 0.0001$
B08s	8.076×10^0	5.844×10^1	13.819	$<2.0 \times 10^{-16}$	$p < 0.0001$
B08As	8.902×10^1	1.933×10^0	46.059	$<2.0 \times 10^{-16}$	$p < 0.0001$

p-value of the model: $p < 0.0001$.

The coefficients of Model 1 were both positive and negative. This model was composed of explanatory variables, consisting of a combination of the spectral behavior of the soil in the laboratory and satellite image, with the particularity that there are reflectance levels related to the characteristics of red, red border, and near-infrared. Classical vegetation indices were composed, but in this case, the objective was to classify the order of the soil in Andisol and Mollisol. Furthermore, one of the independent variables was not significant (B08c), which did not influence the global significance of this logistic regression model ($p < 0.0001$).

Based on the training dataset, we obtained a confusion matrix (Table 7).

Table 7. Training dataset from Model 1 based on reflectance levels of soil spectra in laboratory and satellite image.

Predicted Data	Data Real		Total
	Andisol	**Mollisol**	
Andisol	137,147	6006	143,153
Mollisol	2615	96,018	98,633
Total	139,762	102,024	241,786

The confusion matrix indicated a training error of 3.5%, which means that the model was good for classifying soils in relation to their order in Andisol or Mollisol based on the spectral behavior of the soil in laboratory and satellite image, and satellite related to the red reflectance of any modality. The false-positive and false-negative coefficients were relatively low (Table 7), at 1.9% (2615/139,762) and 5.89% (6006/102,024), respectively. In other words, 2615 soil samples of the Andisol order were classified as Mollisol, and 6006 Mollisol soil samples were classified as Andisol. We then evaluated the model using a test dataset to describe the validation process.

Model 1 Validation

The diagnostic evaluation of Model 1, from the diagnostic statistics using the test dataset, was generally good because the accuracy, sensitivity, and specificity were above 95%. On the other hand, the *p*-value of the Kappa statistic (Table 8) was more significant than 0.05 ($p > 0.05$), indicating that the null hypothesis that the measurements obtained through Model 1 were equivalent to the real data is not rejected.

Table 8. Diagnostic statistics for the validation of Model 1.

Statistical	Value	Statistical	Value	Statistical	Value
Success	0.9644	Kappa	0.9268	Sensitivity	0.9811
CI 95%	0.9633–0.9656			Specificity	0.9416

3.3.2. Model 2 for Variable Land Use of the Andisol Order

As explained in Section 2.8.1, to classify land use according to the Andisol soil order, linear discriminant analysis was applied (Figure 8). The following results were obtained (Table 9).

Table 9. Median reflectance levels of soil spectra in laboratory and satellite image, according to the group defined by land use for the Andisol order.

	Laboratory (B0ic) Satellite Image (B0is)						
Land Use	B02c	B03c	B04c	B05c	B06c	B07c	B08c
Shrub	0.0679	0.0765	0.0980	0.1108	0.1229	0.1381	0.1644
Páramo	0.0733	0.0855	0.1183	0.1363	0.1514	0.1693	0.1995
Pasture	0.0554	0.0650	0.0878	0.1012	0.1137	0.1290	0.1461
Land Use	B08Ac	B11c	B12c	B02s	B03s	B04s	B05s
Shrub	0.1680	0.3096	0.2812	0.0090	0.0363	0.0228	0.0698
Páramo	0.2006	0.3537	0.2843	0.0147	0.0400	0.0400	0.0785
Pasture	0.1600	0.3409	0.3161	0.0166	0.0481	0.0354	0.0848
Land Use	B06s	B07s	B08s	B08As	B11s	B12s	
Shrub	0.2134	0.2502	0.2791	0.2710	0.1296	0.0624	
Páramo	0.1591	0.1844	0.2083	0.2065	0.1594	0.0902	
Pasture	0.2413	0.2940	0.3272	0.3208	0.1791	0.1006	

The spectral values of the soil in the laboratory had greater weight in the classification of the different land uses than in the satellite image as a function of spectrometry in the laboratory and satellite image. Regardless of the sign, the coefficients of the soil spectral values in the laboratory were greater than the coefficients of the values in the satellite image. Consequently, the first component of this linear discriminant function explains that 96.5% of the total variability of the three different land uses had lower coefficients; even though the reflectance values in the satellite image were lower, these variables were important for the classification of land use as a function of the Andisol soil order. The first and second discriminant components were the linear combinations of the variables that best discriminate between the three land uses of the Andisol order, which in this case corresponded to the entire spectrum of soil in the laboratory and satellite image, respectively. Figure 11 shows the results of the soil classification based on the linear discriminant function model (Model 2).

The numbers 1, 2, and 3 represent the mean of each dataset. The means were quite separate, which implies a good classification of the land use of the Andisol order. In addition, based on the first linear discriminator, better discrimination was observed between the soils of Pasture and Páramo or between soils of Shrub and Páramo use than between the soils of Shrub and Pasture use. This situation could be because these land uses, in some cases, have relatively small neighboring units. Based on the training dataset, a confusion matrix was obtained (Table 10). In Figure 11 and Table 10, a good classification of the land uses of the Andisol order was observed, with a classification error of 0.51%.
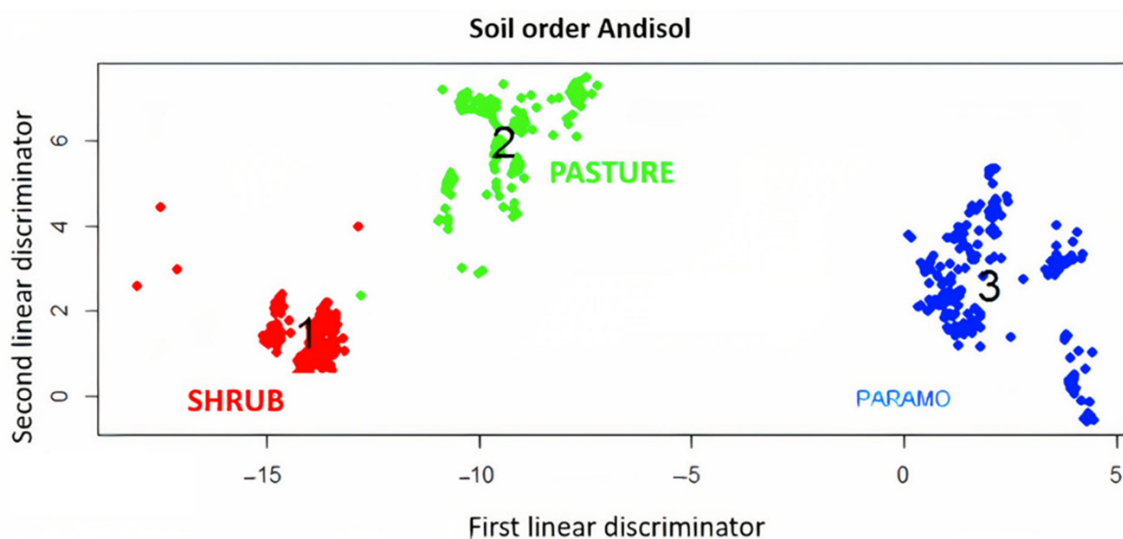
Figure 11. Classification of land uses of the Andisol order.

Table 10. Classification of land uses of the Andisol order based on Model 2 for the training set.

Real Land Use	Land Use Predicted by Model 2			Total
	Shrub	Páramo	Pasture	
Shrub	671	0	0	671
Páramo	0	132,132	0	132,132
Pasture	693	12	6220	6925
Total	1364	132,144	6220	139,728

Model 2 Validation

From the first group data for Shrub, Páramo, and Pasture land uses of the Andisol order, corresponding to the 30% that were not part of the model calculation, a confusion matrix was obtained (Table 11) from which a good classification of the land uses of the Andisol order was obtained, whose classification error was only 0.50% and accuracy 99.5%. Similar results were obtained for the second set of randomly selected data.

Table 11. Classification of land uses of the Andisol order based on Model 2 for a first test set.

Real Land Use	Land Use Predicted by Model 2			Total
	Shrub	Páramo	Pasture	
Shrub	289	0	2	291
Páramo	0	56,628	0	56,628
Pasture	295	5	2667	2967
Total	584	56,633	2669	59,886

3.3.3. Model for Variable Land Use of the Mollisol Order

(a) Model 3 for the variable of land use of the Mollisol order 1 from all wavelengths of the soil spectrum in laboratory and satellite image

The results were obtained from the application of LDA (Table 12).

Table 12. Coefficients of the linear discriminant function with dependent variable of the land use of the Mollisol 1 order and independent variables of reflectance levels of soil spectral behavior in laboratory and satellite image.

Field Bands	Coefficients LD1	Satellite Bands	Coefficients LD1	Field Bands	Coefficients LD2	Satellite Bands	Coefficients LD2
B02c	−708.68	B02s	−5.48	B02c	−410.48	B02s	13.59
B03c	2125.64	B03s	20.24	B03c	466.47	B03s	−33.49
B04c	−3693.53	B04s	−13.93	B04c	−883.87	B04s	17.64
B05c	887.38	B05s	−5.09	B05c	451.14	B05s	9.81
B06c	2377.94	B06s	3.01	B06c	105.54	B06s	−12.59
B07c	−805.64	B07s	3.55	B07c	327.71	B07s	−7.35
B08c	180.12	B08s	−0.92	B08c	−280.28	B08s	3.01
B08Ac	−388.23	B08As	−5.63	B08Ac	468.29	B08As	16.89
B11c	−32.67	B11s	4.67	B11c	−25.013	B11s	−4.61
B12c	50.53	B12s	1.16	B12c	6.73	B12s	−1.11

For the Mollisol 1 order, observing the coefficients of the linear discriminant function (Table 12), it resulted that the spectral behavior of soils measured in the laboratory exhibited a higher contribution to discriminate land uses Forest, Páramo, and Pasture for the Mollisol 1 order, compared to the coefficients derived from the satellite image. Consequently, the first component of this linear discriminant function (LD1) explained 70.9% of the total variability of the three different land uses, implying that although the reflectance values in the satellite image had lower coefficients, these variables were important for the classification of land use from the Mollisol 1 soil order. The first and second discriminant components were the linear combinations of the variables that best discriminated between the three Mollisol 1 land use types. Figure 12 shows a representation of the linear discriminant function for this particular case of Mollisol 1 land use, with a minimum overlap between Páramo and Pasture, with a classification error of only 0.47% and an accuracy of 99.52%.
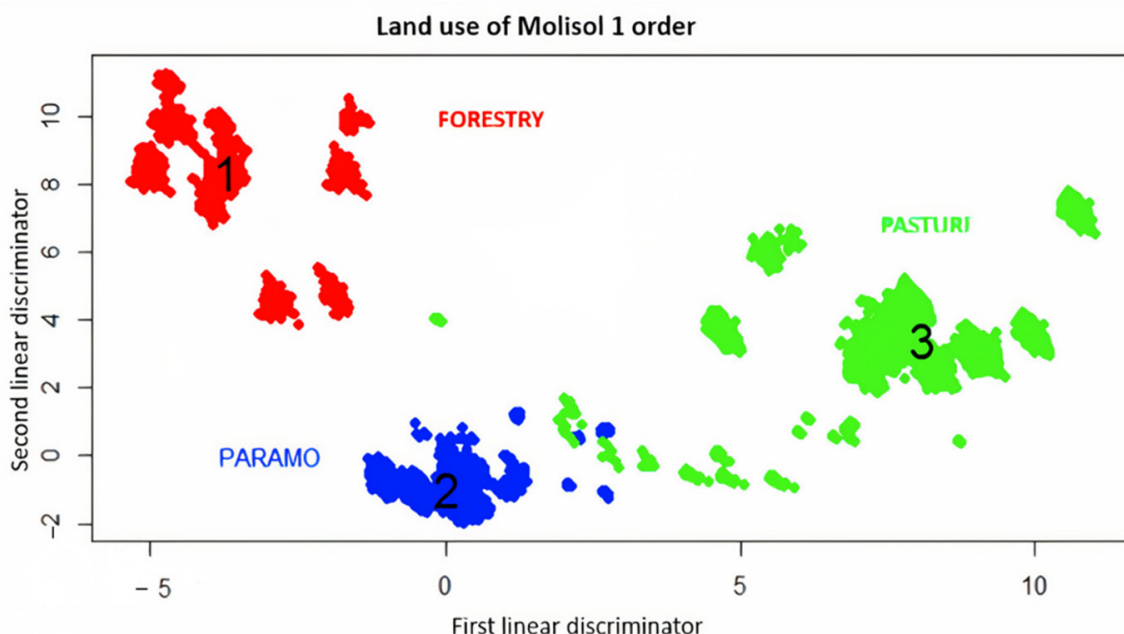

Figure 12. Representation of the linear discriminant function: land use—Mollisol 1.

(b) Model 4 for the variable of land use of the Mollisol 2 order from all wavelengths of the soil spectra in laboratory and satellite image

The following results were obtained from the application of LDA (Table 13):

Table 13. Coefficients of the linear discriminant function with dependent variable of the use of soil of the Mollisol 2 order and independent variables of the reflectance levels of the soil spectra in laboratory and satellite image.

Ground Bands	Ground Coefficients LD	Satellite Bands	Satellite Coefficients LD
B02c	2796.26	B02s	−2.27
B03c	−3969.19	B03s	−1.87
B04c	−6701.17	B04s	4.90
B05c	18,157.90	B05s	2.94
B06c	−10,820.44	B06s	3.71
B07c	−311.00	B07s	−2.37
B08c	−1444.39	B08s	0.21
B08Ac	2194.74	B08As	−1.85
B11c	35.04	B11s	1.42
B12c	−36.25	B12s	−8.34

In relation to the coefficients of the linear discriminant function (Table 13), it was found that the soil spectral values measured in the laboratory had a higher contribution for the classification of the considered land uses compared to those of the satellite image. Regardless of the sign, the coefficients of the soil spectrum in the laboratory were higher than the coefficients of the spectrum in the satellite image. Consequently, the only component of this linear discriminant function explained 100% of the total variability of the two different land uses, which implies that even though the spectral values of the satellite image had lower coefficients, these variables were important for the classification of these land uses as a function of the Mollisol 2 soil order. Figure 13 represents a good classification of the uses of soils order Mollisol 2.
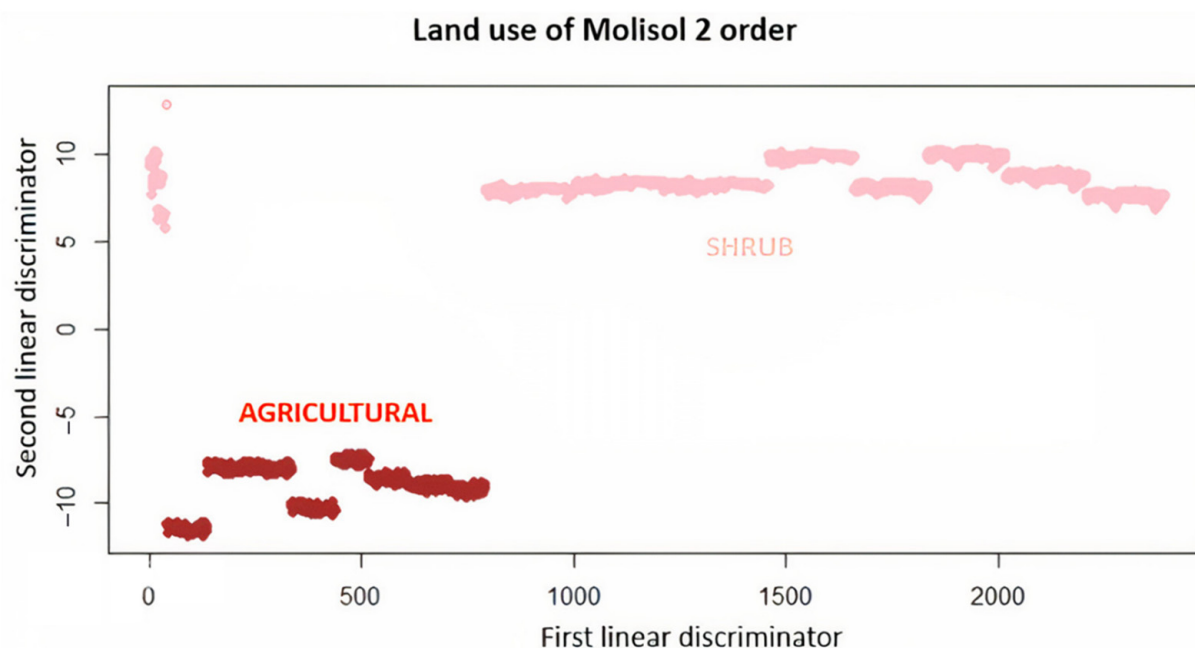

Figure 13. Representation of the linear discriminant function: land use—Mollisol 2.

Model 4 Validation

For the validation of Model 4, we tested 30% of the remaining data, called the test dataset, to classify the soils based on Model 4, and obtained the following confusion matrix, shown in Table 14.

Table 14. Classification of land use of the Mollisol 2 order based on Model 4 for the test dataset.

| Real Land Use | Land Use Predicted by Model 4 | | Total |
	Agriculture	Shrub	
Agriculture	1286	0	1286
Shrub	0	4049	4049
Total	1286	4049	5335

In Table 14, for the remaining 30%, a good classification of the land use was observed in Agriculture and Shrub for the Mollisol 2 order, considering the same behavior indicated in the training data.

3.4. Index Development

3.4.1. Index for Andisol and Mollisol Soil Orders from Model 1

According to the methodological process indicated in Section 2.8.1, the index.ma.1 (Mollisol Andisol Index) was obtained (Equation (4)):

$$\text{index.ma.1} = -\,0.21 - 272.64\text{B04c} + 336.76\text{B05c} + 328.37\text{B06c}$$

$$-\,500.64\text{B08c} - 0.13\text{B08c} + 108.45\text{B08Ac} + 0.20\text{B04s} \tag{4}$$

$$+\,1.58\text{B05s} - 2.11\text{B06S} - 2.11\text{B07s} + 0.29\text{B08s} + 3.20\text{B08As}$$

The index.ma.1 separates the soils according to their order into Andisol and Mollisol. If the index values are positive, they correspond to soils of the Andisol order; if index.ma.1 takes negative values, they correspond to soils of the Mollisol order. The descriptive statistics of index.ma.1 are shown in Table 15.

Table 15. Descriptive statistics of the index.ma.1 according to the order of the soil.

| Orden | Descriptive Statistics | | | | | | |
	Minimum	Q1	Median	Mean	S	Q3	Maximum
Andisol	−0.6407	0.0936	0.1908	0.2377	0.1792	0.3310	1.3658
Mollisol	−3.6134	−2.1854	−1.9723	−1.4961	0.5688	−0.1383	0.4836
Global	−3.6134	−0.3100	0.0669	−0.1340	0.5868	0.2290	1.3658

For the soils of the Andisol order, the mean level of the index was 0.2377 with a relatively low level of variability, equal to 0.1792. For soils of the Mollisol order, the mean level of the index was lower, −1.4961, presenting a higher level of variability equal to 0.5688, which can also be classified as a high level of variability. In this way, index.ma.1 classifies soils according to their order.

3.4.2. Indices Depending on the Variable of Land Use of the Andisol and Mollisol Orders

(a) Index for the Land Use of the Andisol Order from Model 2

The index obtained from the discriminant function of Model 2 was expressed as follows (Equation (5)):

$$\text{index.2} = -\,22.24\text{B02c} - 67.95\text{B03c} - 149.75\text{B04c} + 402.09\text{B05c}$$

$$+\,438.95\text{B06c} - 705.75\text{B07c} + 108.45\text{B08Ac} + 0.20\text{B04s}$$

$$+\,29.83\text{B11x} - 23.95\text{B12c} - 2.67\text{B02s} - 2.49\text{B03s} \tag{5}$$

$$+\,2.90\text{B04s} + 1.13\text{B05s} + 1.56\text{B06} - 0.99\text{B07s}$$

$$+\,0.18\text{B08s} - 0.89\text{B08As} + 0.67\text{B11s} - 1.55\text{B12s}$$

The descriptive statistics of Index 2 are displayed in Table 16.

Table 16. Descriptive statistics of Index 2 for land use of the Andisol order.

Land Use	Descriptive Statistics						
	Minimum	Q1	Median	Mean	S	Q3	Maximum
Shrub	−1.98	−1.29	−1.14	−1.04	0.33	−1.04	−0.43
Páramo	0.25	0.49	0.53	0.54	0.06	0.58	1.14
Pasture	−1.46	−1.30	−1.18	−1.12	0.22	−0.99	−0.28
General	−1.98	0.48	0.52	0.45	0.38	0.58	1.14

For land use of the Andisol order, the mean level of Index 2 was higher in Páramo (0.54), with a level of variability equal to 0.06 (the table in Section 3.5.2). For the land use of the Pasture type of the Andisol order, the average level of Index 2 was −1.12, with a level of variability equal to 0.22 (the table in Section 3.5.2). In the use of bushland of the Andisol order, the mean level of Index 2 was −1.04, with a level of variability of 0.33. The level of variability of the groups defined according to the land use of the Andisol order was very different, representing the natural behavior of these variables.

(b) Index for the Land Use of the Mollisol 1 Order from Model 3.

The index obtained from the discriminant function of Model 3 is expressed as follows (Equation (6)).

$$
\begin{aligned}
\text{index.3} = {}& 127.77B02c - 383.23B03c + 665.89B04c - 159.98B05c \\
& - 428.71B06c + 145.25B07c - 32.47B08c + 69.99B08Ac \\
& + 5.89B11c - 9.11B12c + 0.99B02s - 3.65B03s \\
& + 2.51B04s + 0.92B05s - 0.54B06s - 0.64B07s \\
& + 0.16B08s + 1.014B08As - 0.84B11s - 0.21B12s
\end{aligned}
\tag{6}
$$

For land uses of the Mollisol 1 order, the mean level of Index 3 was higher in forest land use, at 0.79, with the highest level of variability, equal to 0.22 (the table in Section 3.5.3). For the use of páramo land of the Mollisol 1 order, the mean level of Index 5 was −0.06, with the lowest level of variability, equal to 0.12 (Table 17). The level of variability of the groups defined as a function of the land use of the Mollisol order was different, representing the natural behavior of these variables.

Table 17. Descriptive statistics of Index 3 for land use of the Mollisol 1 order.

Land Use	Descriptive Statistics						
	Mín	Q1	Median	Mean	S	Q3	Max
Forest	0.32	0.56	0.85	0.79	0.22	0.98	1.15
Páramo	−0.11	−0.08	−0.07	−0.06	0.12	0.03	0.19
Pasture	−2.19	−1.36	−1.23	−1.25	0.19	−1.18	0.082

(c) Index for the land Use of the Mollisol 2 Order from Model 4

The fourth index obtained from the discriminant function (Model 4) was obtained by standardizing the coefficients of this model. Each coefficient of Model 4 was divided by the sum of its coefficients in such a way that the sum of the coefficients of Index 4 was equal to 1, obtaining (Equation (7)):

$$
\begin{aligned}
index.4 = &- 27.40B02c + 38.90B03c + 65.66B04c - 177.93B05c \\
&+ 106.03B06c + 3.05B07c + 14.15B08c - 21.51B08Ac \\
&- 0.34B11c + 0.35B12c + 0.02B02s + 0.02B03s \\
&- 0.05B04s - 0.03B05s - 0.04B06s + 0.02B07s \\
&- 0.002B08s + 0.02B08As - 0.01B11s + 0.08B12s
\end{aligned}
\tag{7}
$$

For land uses of the Mollisol 2 order, the mean Index 4 was higher in agricultural land use, at 0.12, with the highest level of variability, equal to 0.02 (Table 18). For the use of shrub soil of the Mollisol 2 order, the mean level of Index 6 was −0.07, with the lowest level of variability, equal to 0.01 (Table 18). The level of variability of the groups defined as a function of the land use of the Mollisol order was different, representing the natural behavior of these variables.

Table 18. Descriptive statistics of Index 4 for the land use of the Mollisol 2 order.

Land Use	Descriptive Statistics						
	Mín	**Q1**	**Median**	**Mean**	**S**	**Q3**	**Max**
Agriculture	0.09	0.10	0.11	0.12	0.02	0.12	0.15
Shrub	−0.11	−0.08	−0.07	−0.07	0.01	−0.07	−0.05

3.5. Regression Tree Models to Define Association between Indices and Physicochemical Parameters

3.5.1. Regression Tree Models of Physicochemical Parameters as a Function of Soil Order through Model 1 (index.ma.1)

The first model predicted the values of index.ma.1. as a function of the covariates, soil order, and physicochemical parameters. An example with soil moisture is presented, where for soils of the Andisol order, the mean of the index.ma.1 (Figure 14) is equal to −0.134. If the soil moisture (HU) is greater than or equal to 36.7 for soils of the Andisol order, the predicted value of the index.ma.1 is on average equal to 0.109 (i = 0.109) for n = 74,000 soil samples. However, if the soil moisture is less than 36.7, the predicted value of index.ma.1 is on average equal to 0.382 (i = 0.382), for n = 65,700 soil samples. The same interpretation for soils of Mollisol order.

Likewise, if the value of the index is positive, it corresponds to an Andisol soil order and negative to a Mollisol soil order. The predicted moisture value is at least 36.7.

3.5.2. Regression Tree Models of Physicochemical Parameters as a Function of Land Use of the Andisol Order through Model 2 (index.2)

As shown in Table 19, it was possible to obtain the effect of land use of the Andisol order on the physicochemical parameters.

Table 19. Application of one-way ANOVA to determine the effect of land use of the Andisol order on PFQ.

PFQ	F-Value	*p*-Value	Decision
OM	21,289	$<2 \times 10^{-16}$	Significant
CE	82,125	$<2 \times 10^{-16}$	Significant
pH	19,808	$<2 \times 10^{-16}$	Significant
HU	5213	$<2 \times 10^{-16}$	Significant

An example of the non-parametric regression tree model is presented below (Figure 15), with the dependent variable as index.2 and the independent variables as soil use and organic matter (OM) for soils of the Andisol order (Table 20) (ARUSAMO).
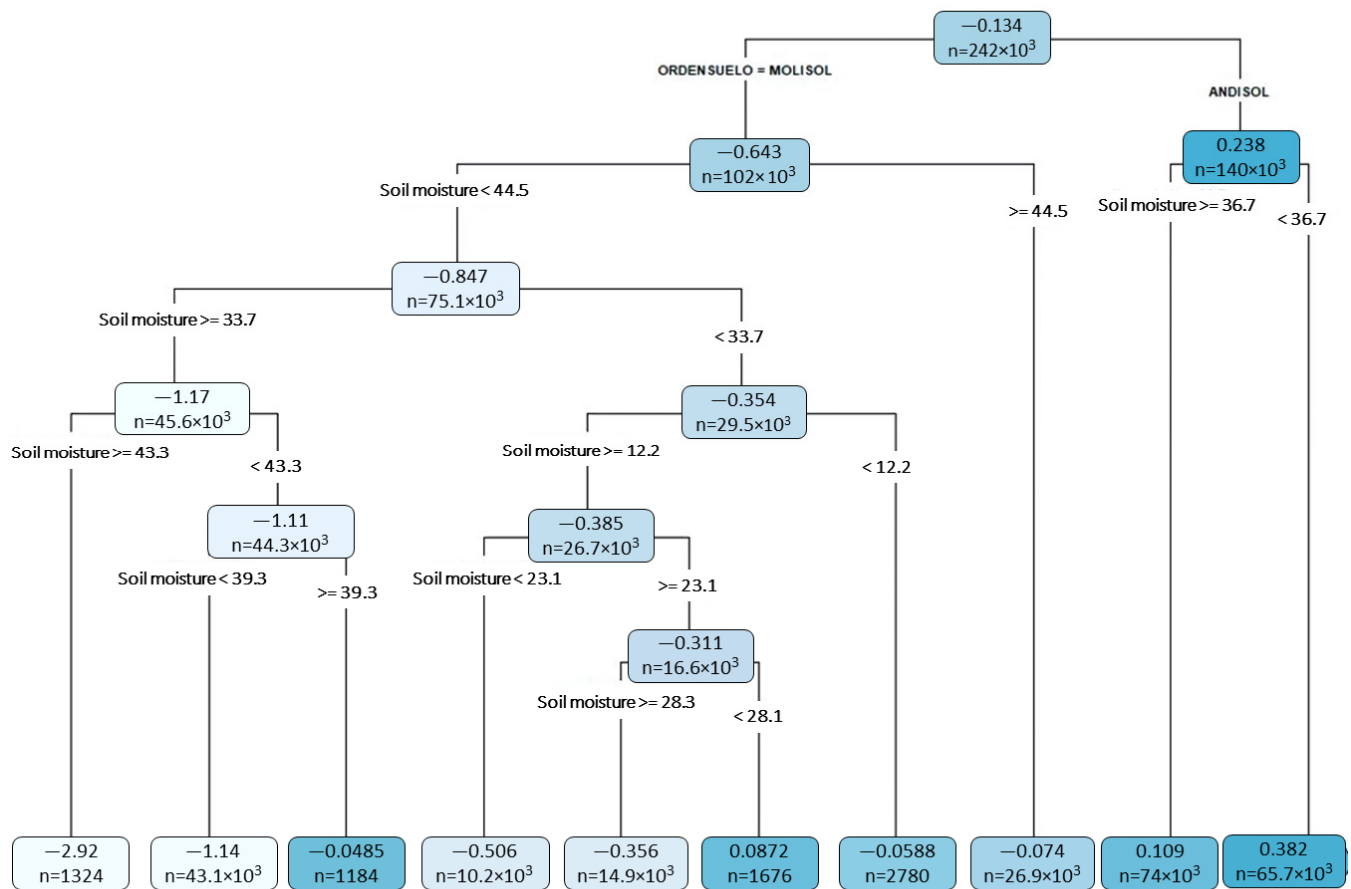
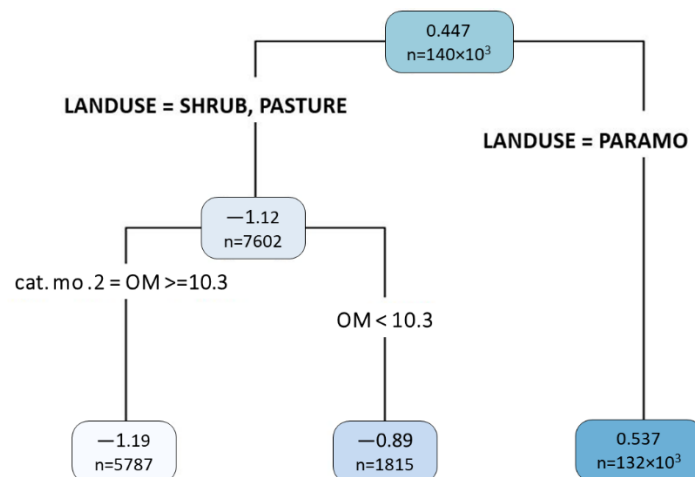**Figure 14.** Moisture Regression Tree Model for Soil Order Andisol-Mollisol (ARMAH).

Figure 15. Regression tree for Index 2, Andisol soil, and OM.

Table 20. Classification of land uses of the Andisol order according to the mean of organic matter.

Land Use	Mean of OM		Total
	OM < 10.3%	OM ≥ 10.3%	
Shrub	677	0	677
Páramo	122,254	9878	132,132
Pasture	1138	5787	6925
Total	**124,069**	**15,665**	**139,734**

3.5.3. Regression Tree Models of Physicochemical Parameters as a Function of Land Use of the Mollisol Order through Model 3 (index.3)

In Figure 16, the regression tree model of index.3 can be observed as an example in the function of land use of soil order Mollisol 1 and organic matter, which are related based on Table 21 (ARUSM3MO).
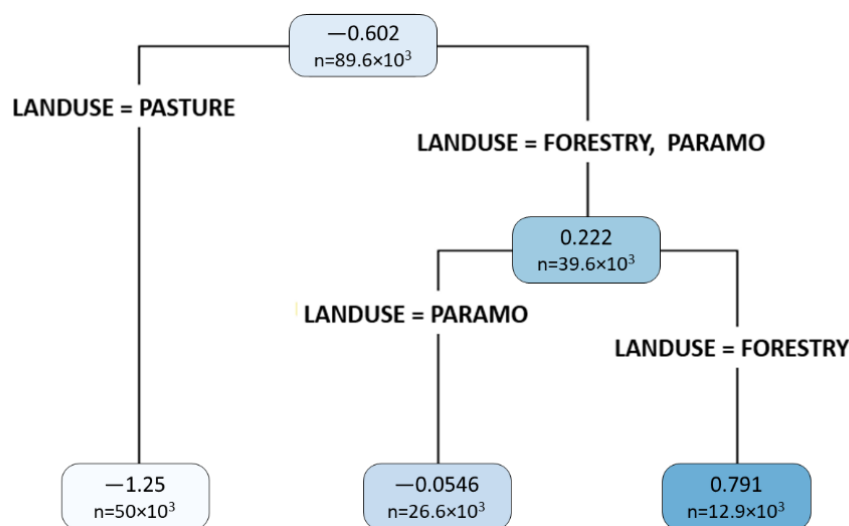

Figure 16. Regression tree for index 3: land use of Mollisol 1 and OM.

Table 21. Classification of the soil samples of the Mollisol 1 order according to the use of the soil by the mean of OM.

Land Use	Mean of OM		Total
	OM < 8.6%	OM ≥ 8.6%	
Forest	10,119	2814	12,933
Páramo	0	26,650	26,650
Pasture	41,775	8219	49,994
Total	51,894	37,683	89,577

OM was greater in Páramo (≥8.6%) than Forest and Pasture, with a misclassification of 22% for Forest and 17% for Pasture (Table 21).

3.5.4. Regression Tree Models of Physicochemical Parameters as a Function of Land Use of the Mollisol Order through Model 4 (index.4)

The land use regression tree model for soil order Mollisol 2 and OM, which are related based on Table 22 (ARUSM4MO), are shown in Figure 17.

Table 22. Classification of the soil samples of the Mollisol 2 order according to the use of the soil by the mean of OM.

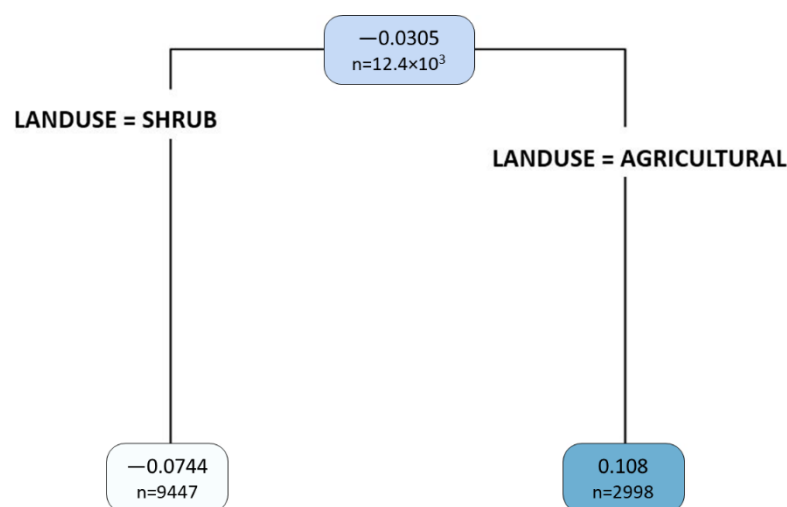| Land Use	Mean of OM		Total
	OM < 6.1%	OM ≥ 6.1%	
Agriculture	2998	0	2998
Shrub	94	9353	9447
Total	2979	9466	12,445

Figure 17. Regression tree for Index.4: land use of Mollisol 2 and OM.

OM was greater in Shrub ($\geq$6.1%) than Agriculture (<6.1%), with a misclassification of 1% for Shrub (Table 22).

4. Discussion

The results of this study allow for a description of the correlation between the physicochemical parameters with index.2 (Andisol), index.3 (Mollisol 1), index.4 (Mollisol 2), according to the soil order–land use homogeneous zones defined in Section 2.7 and based on the criteria of Zebrowski 1997 [52]. In the case of the Mollisol order soil, the prediction values for Páramo were $\geq$8.6% (Table 21), which maintained the characteristic behavior of the Ecuadorian Andean zone, as cited by Podwojewski (1999) [53]. On the contrary, for Forest and Pasture, the prediction models presented a behavior with a lower value in organic matter (<8.6%) (Table 21), a less acidic pH and lower soil moisture percentage, and a higher electrical conductivity [10]. This behavior shows the effects of the impact of human activity, with a lesser value of OM in the Agriculture land use (<6.1%) (Table 22).

The results presented in this study differ from other studies that compared different classification techniques using Sentinel-2 images [18,54], or considered the capacity of satellite observations to monitor and determine the state of the vegetation due to environmental stress factors by evaluating vegetation and chlorophyll indices [1].

Unlike other methodological approaches [17,55–57], this study demonstrates that the combination of laboratory spectroscopy and multispectral images with environmental covariates is an adequate alternative to establish spatial analysis models to predict the quality of Andean soils in terms of physicochemical variables such as CE, OM, pH, and HU. For this purpose, performing soil order–land use associations was revealed to be an important possible tool for assessing the accomplished predictive models.

(1) Performance of the Models

Equation (4) shows the distributions of the logistic regression coefficients in the R-NIR spectral range for soil order, with low false-positive (1.9%) and false-negative (5.89%) coefficients. For the Andisol soil order, the mean level of the index (index.ma.1) was 0.2377, with a level of variability equal to 0.1792 (Table 15). For the Mollisol soil order, the mean level of the index (index.ma.1) was −1.4961, with a level of variability equal to 0.5688 (Table 15). The index values ranged from approximately −6 to 2, with some outliers below −4 and a very low frequency of occurrence. This corroborates the potential of promoting soil studies based on laboratory spectral data and remote sensors, such as Ali Aldabaa et al. (2014) [19], who evaluated the feasibility of the methods for the prediction of soil surface salinity by visible near-infrared diffuse reflectance spectroscopy (VisNIR) and remote sensing (RS). Equations (5)–(7) show the distributions of the coefficients of

the linear discriminant function in the VIS-NIR-SWIR spectral range for the order–land use both in laboratory and S2 configurations. For land uses of the Andisol order, the level of variability of the defined groups was very different, which represents the natural behavior of these variables (Shrub, Páramo, Pasture), which, unlike previous research on the VIS-NIR, presented greater sturdiness considering SWIR.

(2)   Predictions of Physicochemical Parameters

The prediction performance of the R-NIR model, based on the Student *t*-test with $p < 0.0001$ for OM, CE, pH, and soil moisture, shows that the mean of each parameter in the Andisol and Mollisol soil order were different, concluding that the mean of each one was lower for the Mollisol soil order, unlike CE, where its mean was higher in Mollisol.

Regarding the results obtained from the VIS-NIR-SWIR or full spectrum model, using non-parametric regression tree models, excellent results were obtained for OM, pH, CE, and soil moisture as explanatory variables of order–land use [57]. For Mollisol 1, the 95% confidence intervals for the difference in means for the set of physicochemical parameters (CE, OM, pH, HU) were negative for Pasture and Páramo, and for Forest. This means that on average the given set of parameters had higher values in Forest. For the Mollisol 2 soil type, the 95% confidence intervals for the difference in means for the considered set of physicochemical parameters were negative for Shrub and positive for Agriculture. For Andisol-type soils, the 95% confidence intervals for the difference in means for OM in Páramo are higher than the average OM in Shrub. Similarly, but in an opposite direction, when comparing the mean OM in Pasture and Páramo land uses ($p = 0.0000 < 0.05$), the 95% confidence interval for the mean difference was negative, which implies that the average OM in soils used for pasture was less than the average OM in soils used for Páramo. Very similar results were obtained in relation to the pH physicochemical parameter, and in relation to CE and HU in all pairs of established comparisons there were statistically significant differences.

In the methodological process, the nonparametric regression tree method was successfully applied to predict the values of the model covariates by soil order or land use order (Figure 8). This statistical analysis methodology differed from those applied to date, like Adeline (2017) [41], Bao (2017) [40], Soriano-Disla (2014) [17], and Ali Aldabaa (2014) [19], where it was established that soil properties were derived from reflectance spectra that can be applied from various sources of spectral measurements, such as measurements in the laboratory, in the field, or from remote sensing systems.

These regression tree models were more flexible than those presented by Hill (2011) [55], because they did not consider non-compliance with statistical assumptions such as normality or collinearity problems between predictor variables. The regression tree models allowed for approximate estimates supported by 95% confidence intervals as a measure of the variation range of each physicochemical parameter, allowing for a reading of this from the top to the final nodes and vice versa, which was not possible in other applied models (Figures 14–17; Tables 21 and 22) [19,55,56].

Soil quality and soil degradation are crucial to develop sustainable agricultural activities [58]. The usual methods for environmental soil monitoring are very labor intensive and costly to cover large areas of land [1,13]. Satellite data in this field open up new research opportunities with great applications, as large areas of land can be analyzed and soil quality can be assessed in areas that are difficult to access [6,51].

Finally, this study is very valuable for the Ecuadorian Andean region for soil sustainability. Additionally, the results obtained in this study could be adapted in future research to other geographical regions after reviewing the soil order and land use that allow the relationships observed in the proposed model indices to be confirmed.

5. Conclusions

Soil quality is an important factor in sustainable land management. Its evaluation allows for the development and implementation of sustainable agriculture management techniques. Thus, in this study, an alternative method for the prediction of the parameters

OM, CE, pH, and soil moisture based on the R-NIR and VIS-NIR-SWIR models is presented to demonstrate its applicability in the Ecuadorian Andean region. For this purpose, logistic regression analysis and linear discriminant function analysis were used. This required the establishment of homogeneous zones defined by soil order and land use combinations to design and implement soil-sampling strategies and field–satellite spectral measurements. The findings of this study suggest that soil + RS spectroscopy is a useful technique to predict soil properties, presenting good potential as an impetus towards future soil studies.

According to the results of this study:

(1) The logistic regression function made it possible to predict the values as a soil order function and each of the physicochemical parameters described above.

(2) The linear discriminant function made it possible to treat the data based on the linear combination of the Andisol soil order variables by land use (Shrub, Páramo, Pasture), Mollisol soil order by land use (group 1: Forest, Páramo and Pasture; group 2: Agriculture and Shrub).

(3) Non-parametric models had the advantage of predicting the values of the independent variables OM, CE, pH, and soil moisture (soil properties).

Therefore, because of the achieved results, the proposed methodology might be applied to other regions and adapted to predict soil properties as a function of the site-specific soil order and land use properties. Future research should explore the variability of soil quality parameters geographically with the aim of building regional models.

References

1. Gholizadeh, A.; Kopačková, V. Detecting vegetation stress as a soil contamination proxy: A review of optical proximal and remote sensing techniques. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 2511–2524. [CrossRef]

2. Stenberg, B.; Viscarra Rossel, R.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. *Adv. Agron.* **2010**, *107*, 163–215. [CrossRef]

3. Chaerle, L.; van Der Straeten, D. Seeing is believing: Imaging techniques to monitor plant health. *Biochim. Biophys. Acta-Gene Struct. Expr.* **2001**, *1519*, 153–166. [CrossRef]

4. Fedotov, Y.; Bullo, O.; Belov, M.; Gorodnichev, V. Experimental research of reliability of plant stress state detection by laser-induced fluorescence method. *Int. J. Opt.* **2016**, *2016*, 4543094. [CrossRef]

5. Correa, R.; de Souza, C.; Álves, L. Espectroscopia de reflectancia aplicada a la caracterización espectral de suelos contaminados. In Proceedings of the Latinoamerican Remote Sensing Week—LARS, Santiago, Chile, 23–25 October 2013; pp. 1–7. (In Spanish)

6. Pérez González, E.; García Rodríguez, P. Aplicaciones de la teledetección en degradación de suelos. *Bol. Asoc. Geógr. Esp.* **2013**, *61*, 285–308. (In Spanish) [CrossRef]

7. Yánez, G.; Arciniegas, S. Caracterización Físico-Química y Espectral de Suelos con Actividad Agropecuaria en la Microcuenca del Río Blanco—Cayambe. Bachelor's Thesis, Universidad Central del Ecuador, Quito, Ecuador, 2019. (In Spanish)

8. Cruzatty, L.C.G.; Vollmann, J.E.S. Caracterización de suelos a lo largo de un gradiente altitudinal en Ecuador. *Rev. Bras. Cienc. Agrar.* **2012**, *7*, 456–464. (In Spanish) [CrossRef]

9. UNDP Ecuador. Available online: https://www.ec.undp.org/content/ecuador/es/home/sustainable-development-goals.html (accessed on 29 December 2021).

10. Instituto Espacial Ecuatoriano. *Memoria Técnica: Cantón Cayambe. Generación de Geoinformación para la Gestión del Territorio a Nivel Nacional Escala 1:25,000*; Ministerio de Defensa Nacional, SENPLADES, IEE, GADPP: Quito, Ecuador, 2013. (In Spanish)

11. Franco, W.; Peñafiel, M.; Cerón, C.; Freire, E. Biodiversidad productiva y asociada en el valle interandino norte del Ecuador. *Bioagro* **2016**, *28*, 181–192. (In Spanish)

12. GADIP Cayambe. *Actualización del Plan de Desarrollo y Ordenamiento Territorial del Cantón Cayambe 2015–2025*; GADIP Cayambe: Cayambe, Ecuador, 2015. (In Spanish)

13. Baldock, J.; Hawke, B.; Sanderman, J.; MacDonald, L. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. *Soil Res.* **2013**, *51*, 577–595. [CrossRef]

14. Stevens, A.; Nocita, M.; Tóth, G.; Montanarella, L.; van Wesemael, B. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS ONE* **2013**, *8*, e66409. [CrossRef] [PubMed]

15. Bayat, B.; van der Tol, C.; Verhoef, W. Remote sensing of grass response to drought stress using spectroscopic techniques and canopy reflectance model inversion. *Remote Sens.* **2016**, *8*, 557. [CrossRef]

16. Bellon-Maurel, V.; McBratney, A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils—Critical review and research perspectives. *Soil Biol. Biochem.* **2011**, *43*, 1398–1410. [CrossRef]

17. Soriano-Disla, J.M.; Janik, L.J.; Viscarra Rossel, R.A.; Macdonald, L.M.; McLaughlin, M.J. 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]

18. Borras, J.; Delegido, J.; Pezzola, A.; Pereira, M.; Morassi, G.; Camps-Valls, G. Clasificación de usos del suelo a partir de imágenes Sentinel-2. *Teledetección* **2017**, *48*, 55–66. (In Spanish) [CrossRef]

19. Ali Aldabaa, A.; Weindorf, D.; Chakraborty, S.; Sharma, A.; Li, B. Combination of proximal and remote sensing methods for rapid soil salinity quantification. *Geoderma* **2014**, *239–240*, 34–46. [CrossRef]

20. Calvache, M.; Ballesteros, H.; Andino, J. Caracterizacion del potencial natural de los suelos dedicados a la ganaderia lechera en la sierra ecuatoriana: Caso Cayambe. In Proceedings of the VIII Congreso Ecuatoriano de la Ciencia del Suelo, Manabí, Ecuador, 26 September 2002. (In Spanish)

21. Prat, C.; Moreno, J.; Hidrobo, J.; Trujillo, G.; Ortega, C.; Etcheveres, J.; Hidalgo, C.; Baéz, A.; Gallardo, J. *Understanding Mountain Soils: A Contribution from Mountain Areas to the International Year of Soils*; FAO: Rome, Italy, 2015; pp. 111–115.

22. Tucci, C. *Plan de Manejo Integrado de los Recursos Hídricos en la Cuenca Alta del Río Guayllabamba*; BID Banco Interamericano de Desarrollo Económico, FONAG Fondo Para La Protección Del Agua: Quito, Ecuador, 2009. (In Spanish)

23. Clevers, J.; Kooistra, L.; van den Brande, M. Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sens.* **2017**, *9*, 405. [CrossRef]

24. Immitzer, M.; Vuolo, F.; Atzberger, C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]

25. European Space Agency. *Sentinel 2 Document Library*; ESA Technical Document, ref. COPE-GSEG-EOPG-TN-15-0007; ESA: Paris, France, 2017.

26. ESA. Available online: https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-2/instrument-payload (accessed on 28 January 2021).

27. ESA Copernicus. Available online: https://scihub.copernicus.eu/dhus/#/home (accessed on 16 July 2018).

28. ESA Sentinel Online. Available online: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument (accessed on 28 January 2021).

29. Ministry of the Environment and Water. *Book VI, Annex 2, Norma de Calidad Ambiental del Recurso Suelo y Criterios de Remediación para Suelos Contaminados*; Registro Oficial del Ecuador: Quito, Ecuador, 2015.

30. USDA. *Guía para la Evaluación de la Calidad y Salud del Suelo*; USDA: Washington, DC, USA, 1999. (In Spanish)

31. Baumgardner, M.F.; Silva, L.R.F.; Biehl, L.L.; Stoner, E.R. Reflectance properties of soils. In *Advances in Agronomy*; Academic Press: Cambridge, MA, USA, 1986; Volume 38, pp. 1–44. [CrossRef]

32. Friedman, D.; Hubbs, M.; Tugel, A.; Seybold, C.; Sucik, M. *Guidelines for Soil Quality Assessment in Conservation Planning*; USDA: Washington, DC, USA, 2001.

33. Secretaría de Medio Ambiente y Recursos Naturales de México. *Norma Oficial Mexicana NOM-021-SERMANAT-2000 que Establece las Especificaciones de Fertilidad, Salinidad y Clasificación de Suelos, Estudio, Muestreo y Análisis*; Diario Oficial de la Federación: Mexico City, Mexico, 2002. (In Spanish)

34. Hidrobo, J.R.; Mendez, E.G. Validación del Método Analítico Walkley y Black de Materia Orgánica en Suelos Arcillosos, Francos y Arenosos del Ecuador. Bachelor's Thesis, Universidad Central del Ecuador, Quito, Ecuador, 2016. (In Spanish)

35. Borana, S.; Yadav, S.; Parihar, S. Hyperspectral Data Analysis for Desertic Vegetation of Jodhpur Area. In Proceedings of the 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), Jaipur, India, 7–8 February 2020; pp. 171–175. [CrossRef]

36. Field Spectroradiometer Signal-to-Noise Ratio—The ASD Advantage. Available online: https://www.materials-talks.com/field-spectroradiometer-signal-to-noise-ratio-the-asd-advantage/ (accessed on 29 May 2022).

37. Orjuela, I.; Camacho, J.; Rubiano, Y. Uso de espectrómetros de infrarrojo medio (MIRS) para la determinación de carbono del suelo en oxisoles. In Proceedings of the XII Congreso Latinoamericano y Del Caribe de Ingeniería Agrícola, Bogotá, Colombia, 23 May 2016. (In Spanish)

38. Estrella, S.; Jácome, N. Cuantificación del Contenido de Arcilla en los Suelos del Estado de Santa Catarina, Brasil a Través de los Datos Espectrales Obtenidos Emdiante Fieldspec Pro 3. Bachelor's Thesis, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, 2016. (In Spanish)

39. Bonett, J.; Camacho, J.; Ramirez, L. Análisis de respuestas espectrales en el infrarrojo medio de suelos de Colombia. In Proceedings of the XX Congreso Latinoamericano y XVI Congreso Peruano de la Ciencia del Suelo, Cusco, Peru, 9–15 November 2014. (In Spanish) [CrossRef]

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

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

42. Mark, D.; Egenhofer, M. Topology of Prototypical Spatial Relations between Lines and Regions in English and Spanish. In Proceedings of the Third International Symposium on Spatial Data Handling, Sydney, Australia, 17–19 August 1988.

43. Poveda, K.; Arciniegas-Ortega, S. Análisis del Comportamiento Espectral del Suelo a Partir de sus Características Físico-químicas en la Microcuenca del río Blanco, Cayambe. Bachelor's Thesis, Universidad Central del Ecuador, Quito, Ecuador, 2020. (In Spanish)

44. Amat, J. Estadística con R. Git Hub. 2020. Available online: https://rpubs.com/Joaquin_AR/229736 (accessed on 20 May 2020). (In Spanish)

45. Dalgaard, P. *Introductory Statistics with R*; Springer: New York, NY, USA, 2008; pp. 227–248. [CrossRef]

46. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer: New York, NY, USA, 2013; pp. 130–163, 303–330.

47. Witold, F.; Heather, S. *Interpolate Temperatures Using the Geostatistical Wizard*; Esri: Redlands, CA, USA, 2019.

48. Rojo, J. *Árboles de Clasificación y Regresión*; Consej Superior de Investigaciones Científicas Laboratorio de Estadística: Madrid, Spain, 2006. (In Spanish)

49. Gareth, J.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2021. [CrossRef]

50. The R Project for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 18 April 2020).

51. Chuvieco, E. *Teledetección Ambiental: La Observación de la Tierra Desde el Espacio*, 3rd ed.; Ariel Ciencias: Barcelona, Spain, 2008. (In Spanish)

52. Zebrowski, C. Los suelos con cangahua en el Ecuador. In *Suelos Volcánicos Endurecidos*; IRD: Quito, Ecuador, 1997; pp. 128–137. (In Spanish)

53. Podwojewski, P. *Los Suelos de las Altas Tierras Andinas: Los Páramos del Ecuador*; Sociedad Ecuatoriana de la Ciencia del Suelo, IRD: Quito, Ecuador, 1999. (In Spanish)

54. Peng, Y.; Kheir, R.B.; Adhikari, K.; Malinowski, R.; Greve, M.B.; Knadel, M.; Greve, M.H. Digital Mapping of Toxic Metals in Qatari Soils Using Remote Sensing and Ancillary Data. *Remote Sens.* **2016**, *8*, 1003. [CrossRef]

55. Hill, J.; Udelhoven, T.; Vohland, M. The Use of Laboratory Spectroscopy and Optical Remote Sensing Soil Properties. In *Precision Crop Protection—The Challenge and Use of Heterogenety*; Springer: Dordrecht, The Netherlands, 2010; pp. 67–85. [CrossRef]

56. Yufeng, G.; Thomasson, A.; Ruixiu, S. Remote sensing of soil properties in precision agriculture. *Front. Earth Sci.* **2011**, *5*, 229–238. [CrossRef]

57. Rosero-Vlasova, O.; Pérez-Cabello, F.; Montorio, R.; Vlassova, L. Assessment of Laboratory VIS-NIR-SWIR Setups with Different Spectroscopy Accessories for Characterisation of Soils from Wildfire Burns. *Biosyst. Eng.* **2016**, *152*, 51–67. [CrossRef]

58. Baroudy, A.; Ali, A.; Mohamed, F.; Shokr, M.; Savin, I.; Poddubsky, A.; Ding, Z.; Kheir, A.; Aldosari, A.; Elfadaly, A.; et al. Modeling Land Suitability for Rice Crop Using Remote Sensing and Soil Quality. *Sustainability* **2020**, *12*, 9653. [CrossRef]