



Article

Implementing Cloud Computing for the Digital Mapping of Agricultural Soil Properties from High Resolution UAV Multispectral Imagery

Samuel Pizarro ^{1,2}, Narcisa G. Pricope ³, Deyanira Figueroa ¹, Carlos Carbajal ⁴, Miriam Quispe ⁵, Jesús Vera ⁵, Lidiana Alejandro ⁵, Lino Achallma ⁵, Izamar Gonzalez ⁵, Wilian Salazar ⁶, Hildo Loayza ^{7,8}, Juancarlos Cruz ⁴ and Carlos I. Arbizu ^{6,*}

- ¹ Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Carretera Saños Grande-Hualahoyo Km 8 Santa Ana, Huancayo, Junin 12002, Peru; spizarro@unpc.edu.pe (S.P.); deyanirafigueroa66@gmail.com (D.F.)
 - ² Facultad de Zootecnia, Andean Ecosystem Research Group, Universidad Nacional del Centro del Perú, Av. Mariscal Castilla 3089, Huancayo, Junin 12002, Peru
 - ³ Department of Earth and Ocean Sciences, University of North Carolina Wilmington, 601 S College Rd., Wilmington, NC 28403, USA; pricopen@uncw.edu
 - ⁴ Dirección de Supervisión y Monitoreo en las Estaciones Experimentales Agrarias, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; cmcarbajal@gmail.com (C.C.); jcruz@inia.gob.pe (J.C.)
 - ⁵ Dirección de Supervisión y Monitoreo en las Estaciones Experimentales Agrarias, Instituto Nacional de Innovación Agraria (INIA), Carretera Saños Grande-Hualahoyo Km 8 Santa Ana, Huancayo, Junin 12002, Peru; miriamrocioq@gmail.com (M.Q.); jvera@lamolina.edu.pe (J.V.); lidiana.alejandro@gmail.com (L.A.); linoachamen@gmail.com (L.A.); izamargonzaleztoivar@gmail.com (I.G.)
 - ⁶ Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; r_cambioclimatico@inia.gob.pe
 - ⁷ International Potato Center (CIP), Headquarters P.O. Box 1558, Lima 15024, Peru; h.loayza@cgiar.org
 - ⁸ Programa Académico de Ingeniería Ambiental, Facultad de Ingeniería, Universidad de Huánuco, Huánuco 10001, Peru
- * Correspondence: carlos.arbizu@untrm.edu.pe
† Current address: Facultad de Ingeniería y Ciencias Agrarias, Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Cl. Higos Urco 342, Chachapoyas, Amazonas 01001, Peru.



Citation: Pizarro, S.; Pricope, N.G.; Figueroa, D.; Carbajal, C.; Quispe, M.; Vera, J.; Alejandro, L.; Achallma, L.; Gonzalez, I.; Salazar, W.; et al.

Implementing Cloud Computing for the Digital Mapping of Agricultural Soil Properties from High Resolution UAV Multispectral Imagery. *Remote Sens.* **2023**, *15*, 3203. <https://doi.org/10.3390/rs15123203>

Academic Editor: Thomas Cudahy

Received: 11 May 2023

Revised: 13 June 2023

Accepted: 14 June 2023

Published: 20 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The spatial heterogeneity of soil properties has a significant impact on crop growth, making it difficult to adopt site-specific crop management practices. Traditional laboratory-based analyses are costly, and data extrapolation for mapping soil properties using high-resolution imagery becomes a computationally expensive procedure, taking days or weeks to obtain accurate results using a desktop workstation. To overcome these challenges, cloud-based solutions such as Google Earth Engine (GEE) have been used to analyze complex data with machine learning algorithms. In this study, we explored the feasibility of designing and implementing a digital soil mapping approach in the GEE platform using high-resolution reflectance imagery derived from a thermal infrared and multispectral camera Altum (MicaSense, Seattle, WA, USA). We compared a suite of multispectral-derived soil and vegetation indices with in situ measurements of physical-chemical soil properties in agricultural lands in the Peruvian Mantaro Valley. The prediction ability of several machine learning algorithms (CART, XGBoost, and Random Forest) was evaluated using R^2 , to select the best predicted maps ($R^2 > 0.80$), for ten soil properties, including Lime, Clay, Sand, N, P, K, OM, Al, EC, and pH, using multispectral imagery and derived products such as spectral indices and a digital surface model (DSM). Our results indicate that the predictions based on spectral indices, most notably, SRI, GNDWI, NDWI, and ExG, in combination with CART and RF algorithms are superior to those based on individual spectral bands. Additionally, the DSM improves the model prediction accuracy, especially for K and Al. We demonstrate that high-resolution multispectral imagery processed in the GEE platform has the potential to develop soil properties prediction models essential in establishing adaptive soil monitoring programs for agricultural regions.

Keywords: soil mapping; UAV; Google Earth Engine; machine learning; cloud computing

1. Introduction

It is estimated that by the year 2050, the demand for food for human consumption will increase by up to 70% and, in the absence of available land for agricultural expansion, agricultural intensification predicated on optimal water, soil, and crop management will become increasingly necessary in order to maintain the productive capacity of agricultural lands [1]. Soils, as a heterogeneous system and dominant factor of agricultural production, are characterized by distinct physical and chemical properties that affect the health of crops and determine, to a large extent, their development and sustenance throughout the duration of the growing cycle [2]. Therefore, the efficient management of agricultural soils requires the adoption of site-specific management practices that account for the existing variability of soils and subsequently crops. As such, detailed spatial information is required to delineate oftentimes homogeneous management units, with similar physical and chemical properties [1,3]. Typically, the physical and chemical properties of soils can be determined with laboratory-based analysis, but for large fields and at the scale of agricultural management systems, these methods prove to be costly and have multiple drawbacks [4]. Recently, to assuage some of these limitations in assessing the productive potential of different soil and agricultural systems, a wide range of remote sensing methods have been applied to determine soil properties and assess overall agricultural productive potential, with demonstrably good results, faster and at comparatively large spatial scales [5]. Specifically, on the ground soil mapping technologies depend heavily on the use of geographic information systems (GIS) and global positioning system (GPS) and are increasingly being supplemented with remote sensing technologies for integrated digital soil mapping approaches or soil information systems that use sophisticated data analysis workflows to predict soil properties based on environmental predictors [6].

The cutting edge in the development of integrated soil information systems is therefore at the intersection of GIS, GPS, and remote sensing. Remote sensing technology in particular can provide valuable information on crop health, growth, and yields without the need for physical intervention. Techniques such as aerial photography, multispectral, and hyperspectral imaging and most recently Unoccupied Aerial Vehicles (UAVs) can be used to collect data that can be analyzed to optimize irrigation management, identify pest infestations, and detect disease outbreaks in crops, as well as soil properties [7]. Deery et al. [8] and Prashar and Jones [9] found that close-range remote sensing technology can accurately measure crop growth and identify crop stress caused by factors such as water or nutrient deficiencies. Jindo et al. [10] used remote sensing to identify pest infestations in potato crops, while Luo et al. [11] found that remote sensing can be used to identify disease outbreaks in crops, allowing for timely interventions to prevent yield losses. Furthermore, Cheng et al. [12] demonstrated that remote sensing can be used to map crop water productivity, leading to significant water savings while maintaining or increasing crop yields. These studies highlight the potential of remote sensing technology to assist in crop and soil management, allowing users and farmers to make informed decisions, optimizing their operations and scale production through means of agricultural intensification.

As an important emerging component of remote sensing technologies, UAVs can be used to support agricultural intensification and optimize water, soil, and crop management to meet the global increasing demand for food. Generally, UAVs are being utilized for crop monitoring, precision agriculture [13], harvest monitoring, and, importantly, soil mapping. UAVs lead the cutting edge in digital soil mapping given their ability to contribute to deriving high-resolution maps of soil properties and characteristics [14,15]. UAVs equipped with various multispectral and hyperspectral sensors; those images allow for the generation of multiple related indices from reflectance data supplemented with in situ information through algorithms and with indicators of detailed vegetative development and soil prop-

erties [16], including carbon (C), nitrogen (N), water content, and soil texture [17]. Similarly, widely utilized spectral indices such as NDVI (Normalized Difference Vegetation Index), in combination with more specific vegetation indices (VIs) derived from visible and near infrared (NIR) data have been shown to be related to soil organic carbon [13], infiltration rates [18], and soil moisture and evapotranspiration metrics [19]. However, there are relatively few instances in the literature on the determination of specific soil physical-chemical properties from remotely sensed multispectral imagery. For instance, multispectral imagery collected by an unmanned aerial vehicle (UAV) was used to determine soil properties such as organic matter content [20], clay content [21], soil moisture content [22] or map soil properties such as sand, silt, clay, cation exchange capacity (CEC), soil organic carbon (SOC) and nitrogen [23].

With increasingly high temporal repeat and spatially denser datasets available from both satellite and UAV platforms, machine learning models are leveraged more in order to develop soil predictive models and formalized digital soil mapping frameworks, because they improve the prediction accuracy and eliminate most of the statistical restrictions that regression, kriging, and their variations demanded [24]. In order to implement these complex models with massive datasets though, conventional desktop workstations became insufficient to generate prediction maps in near real-time. An alternative solution to this challenge is the use of cloud computing, and specifically Google Earth Engine (GEE) [25], which has become a free geospatial data analysis platform, capable of storing and analyzing high-resolution imagery as raster data using its computing infrastructure where machine learning algorithms are designed to run in multiple processors simultaneously, reducing the time of processing and resources with accurate results.

The aim of this work is to explore the feasibility to design and implement a digital soil mapping approach using the Google Earth Engine (GEE) platform utilizing high-resolution reflectance imagery derived from a thermal infrared and multispectral camera (Altum model; MicaSense Inc.) flown aboard a UAV, compared with in situ measurements of soil physical-chemical properties (lime (%), clay (%), sand (%), electrical conductivity (EC) (mS/m), nitrogen (N) (ppm), phosphorus (P) (ppm), potassium (K) (ppm), organic matter (OM) (%), aluminum (Al) (ppm) and pH). Accordingly, we had the following hypotheses: (i) it is feasible to implement UAV imagery and ML to predict soil properties in the cloud using Google Earth Engine, (ii) the use of spectral indices and DSM improve the accuracy of predicted values of soil properties, (iii) ML models are efficient in manage multiple datasets of predictors to perform spatially consistent soil properties maps.

First, we present the material and methods section, where we describe the study area, how soil samples were collected, the imagery acquisition and software processing, statistical and spatial analysis, and validation. Then, we show the results, the soil parameters determined by the laboratory, the correlation analysis between soil parameters and predictors, and the evaluation of machine learning models. Finally, we discuss the results and present the conclusions.

2. Materials and Methods

2.1. Study Area

The Peruvian central zone, especially the Mantaro Valley, is purely agricultural and simultaneously the largest agricultural area in the highlands of Peru. It is estimated that between 40,000 and 70,000 ha are cultivated in the lowlands. The soil data collection was carried out at the Santa Ana Agricultural Research Station (Santa Ana for the rest of the text) of the Instituto Nacional de Innovación Agraria (INIA) (75°13'17.60''W, 12°0'42.36''S) (Figure 1), which is located in the El Tambo district, Huancayo province and department of Junin (Peru). Santa Ana is located in the southeast of Peruvian Mantaro Valley at the base of an alluvial fan landscape. This inter-andean valley is located in the Peruvian central highlands at a mean altitude of 3250 m.a.s.l. with a length of 53 km and a width ranging from 4 to 21 km in places. Approximately 20.7% of this important inter-andean valley can be used for agriculture.

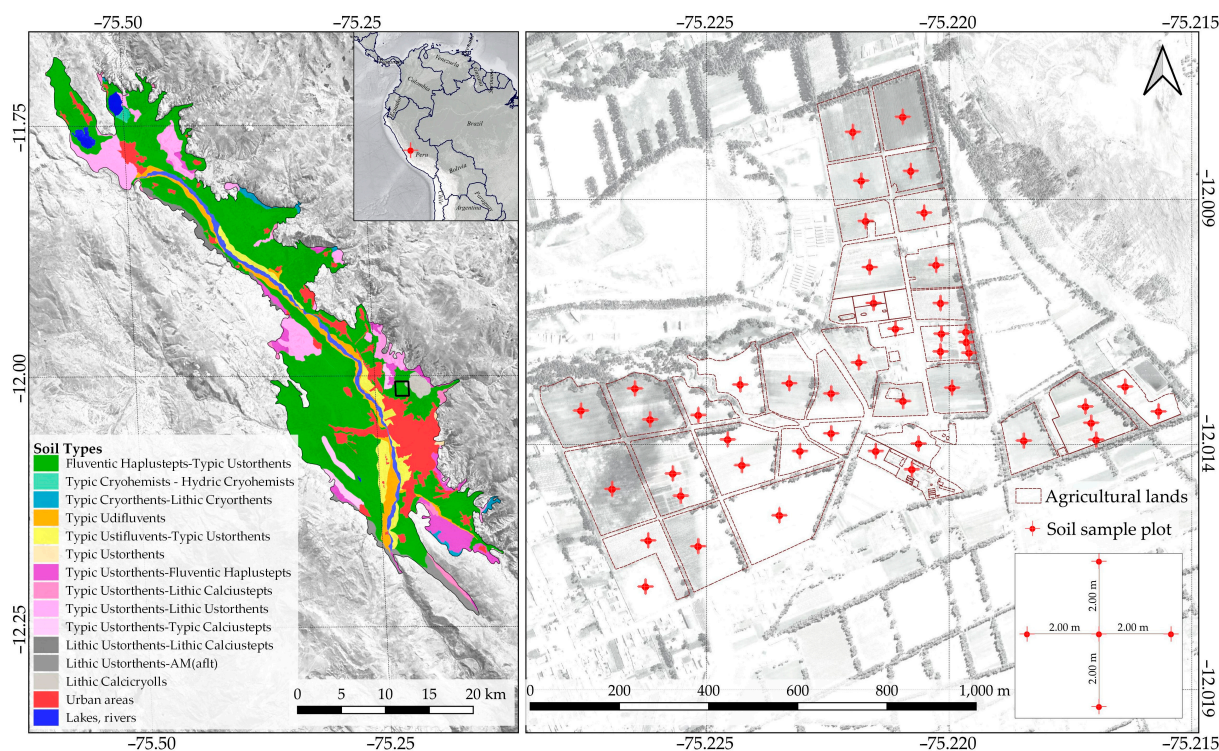


Figure 1. Location of the study area, Santa Ana, Junin (Peru).

Santa Ana has an altitudinal gradient range from 3303 to 3325 m.a.s.l. The physiography is dominated by the plain landscape of the mountain valley. The climate is characterized by periods of rain between November and March, a transition period from April to October and a dry season between May and August with a total amount of 477 mm/year. The average temperature ranges from 3.90 to 20.2 °C, with the lowest temperatures between May and August, and frost events between July and August [26]. The agricultural fields cover 49.83 ha from 67.08 ha distributed in 42 parcels, with flood irrigation canals. The sowing period occurs between October to May.

2.2. Methodological Framework

The methodological framework employed in this study is presented in Figure 2, and described in more detail in the following five methods subsections:

2.2.1. Field Sampling of Chemical and Physical Soil Parameters

A total of 46 soil samples were collected at 30 cm depth in one of the widest stretches of the Mantaro valley at the Santa Ana experimental station; the sample plots were located around the central point of each parcel and georeferenced using a D-RTK 2-DJI GNSS GPS. This approach is simple to implement because the estimation of the quality measures and their precision is straightforward and gives relatively precise estimates, with no assumptions needed in quantifying the standard error of the estimated quality measures [27].

The physicochemical analyses to determine lime (%), clay (%), sand (%), electrical conductivity (EC) (mS/m), nitrogen (N) (ppm), phosphorus (P) (ppm), potassium (K) (ppm), organic matter (OM) (%), aluminum (Al) (ppm) and pH of the soil were carried out at the Laboratorio de Suelos, Aguas y Foliare (LABSAF) of Santa Ana. The samples were dried at room temperature (15–30 °C), disaggregated, homogenized and sieved (2 mm). Soil pH was determined according to the US EPA 9045 D method [28], electrical conductivity (E.C) in (mS/m) was determined according to the ISO 11265:1994/Cor 1 method [29],

organic matter (%), total nitrogen (5% of M.O.); available phosphorus (ppm), available potassium (ppm), Al (ppm) and texture (%) according to the Mexican Official Standard [30].

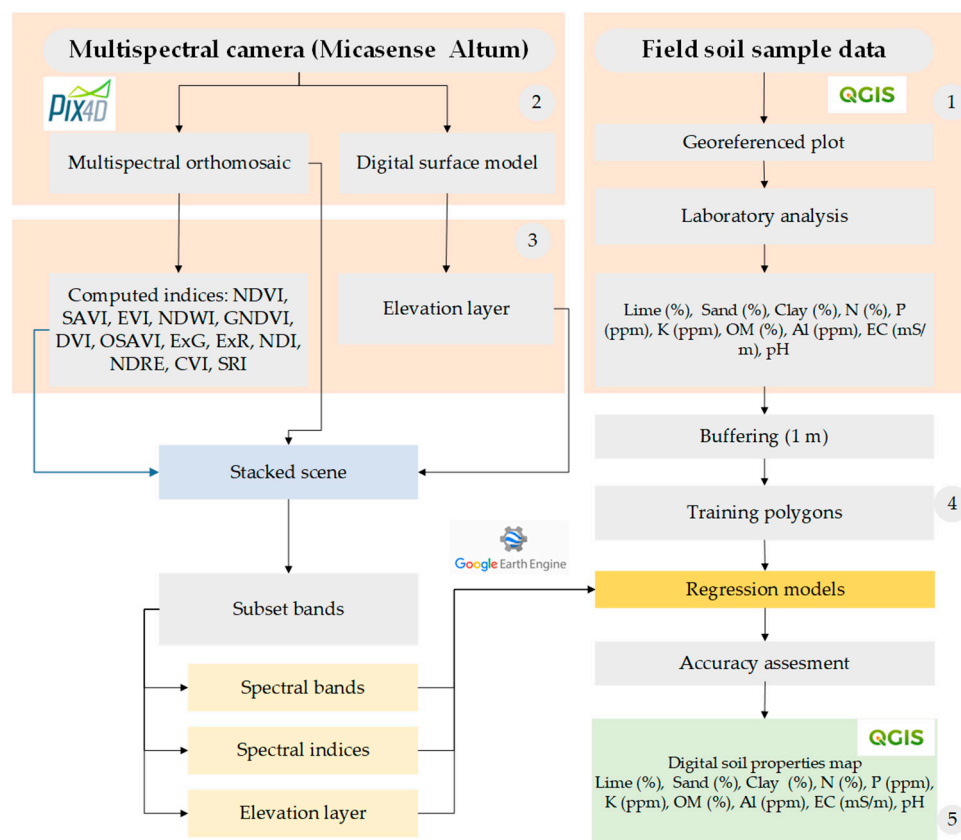


Figure 2. Representation of the methodological framework used in this study.

2.2.2. Acquisition and Processing of Multispectral Imagery

A thermal and multispectral Altum camera (MicaSense, Inc., Seattle, WA, USA), on board a DJI Matrice 300 RTK UAV (DJI, Shenzhen, China) were used to take 16-bit multispectral photos, with 5 spectral bands (blue (475 ± 20 nm), green (560 ± 20 nm), red (668 ± 10 nm), NIR (840 ± 40 nm), and RE (717 ± 10 nm)) at 3.2 megapixels' resolution (2064×1544 pixels) and LWIR thermal band (160×120 pixels) 0.01 megapixels' resolution. Detailed characteristics of the UAV, camera, and flight plan used are shown in Figure 3.

The flight plan was executed roughly at noon local time on 8 August 2022, at a height above the ground of 150 m. The photos were taken every 2.0 s with 75% front and side overlap. Finally, these photos were stored in 16-bit .tiff format.

The photogrammetric processing was carried out in the Pix4D Pro Mapper software (Prilly, Switzerland). The relative differences between the initial and optimized internal parameters are minimal (0.48%), indicating that initial parameters are reliable for the construction of the orthomosaic. We collected 8 ground control points with a D-RTK 2-DJI GNSS GPS (Horizontal: 1 cm + 1 ppm(RMS); Vertical: 2 cm + 1 ppm(RMS)) and introduced them into the processing flow to improve the topographic precision of the point cloud and orthomosaic reflectance bands, with a final ground surface distance (GSD) of 15.42 cm. Based on the point cloud, a digital surface model (DSM) was generated, at the same resolution as the orthomosaic and exported in .tiff format.

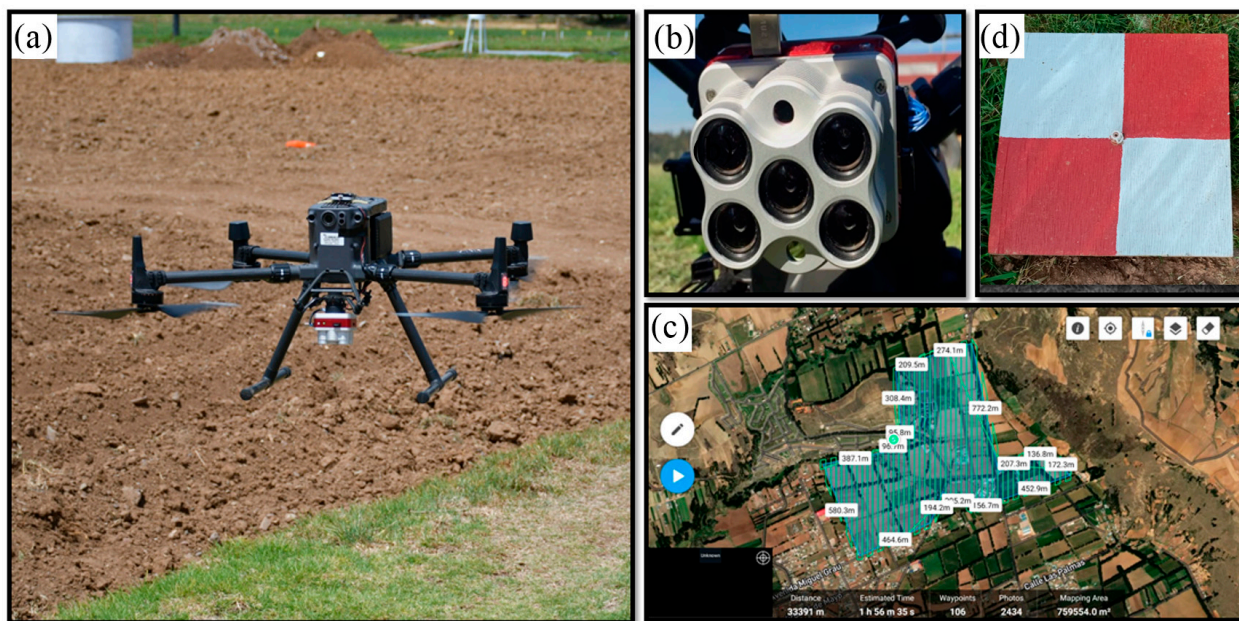


Figure 3. (a) Matrice 300 UAV integrated with the Altum sensor serving as the imaging platform used in this study, (b) Altum camera, (c) flight plan for the study image, and (d) GCP.

2.2.3. Model Development and Statistical Analysis

Variable Extraction

To develop the spatial soil parameters distribution models, we utilized the multispectral 14 spectral indices commonly used in vegetation and soil analysis, shown in Table 1. These indices include vegetation, soil and water indices. A circular buffer of 0.5 m in diameter was used for each sampled point, where the reflectance values of each pixel are converted into an observation replica that contrasts with the concentration value of the soil parameter of interest. Vegetation indices were calculated through the different combinations of reflectance, and compiled as predictors together with the pure spectral bands.

Table 1. Spectral indices extracted from the Micasense Altum imagery.

Bands	Wavelength (nm)
Normalized Difference Vegetation Index (NDVI) [31]	$\frac{(NIR - RED)}{(NIR + RED)}$
Enhanced Vegetation Index (EVI) [32]	$G \times \frac{NIR - RED}{NIR + C1 \times RED + C2 \times BLUE + L}$
Normalized Difference Water Index (NDWI) [33]	$NDWI = \frac{(GREEN - NIR)}{(GREEN + NIR)}$
Soil Adjusted Vegetation Index (SAVI) [32]	$\left(\frac{(NIR - RED)}{(NIR + RED + 1)} \right) (1 + L)$
Green Normalized Difference Vegetation Index (GNDVI) [34]	$\frac{(NIR - GREEN)}{(NIR + GREEN)}$
Difference Vegetation Index (DVI) [35]	$(NIR - RED)$
Optimized Soil Adjusted Vegetation Index (OSAVI) [36]	$(1 + 0.16) \left(\frac{(NIR - RED)}{(NIR + RED + 0.16)} \right)$
Excess Green index (ExG) [37]	$2 \times GREEN - RED - BLUE$
Excess Red index (ExR) [38]	$2 \times RED - GREEN$
ExG – ExR [39]	$ExG - ExR$
Normalized Difference Index (NDI) [40]	$\frac{(GREEN - RE)}{(GREEN + RE)}$
Red-edge Normalized Difference Vegetation Index (NDRE) [41]	$\frac{(NIR - RED)}{(NIR + RED)}$
Chlorophyll vegetation index (CVI) [42]	$NIR \times \left(\frac{RED}{GREEN^2} \right)$
Simple Ratio Red/Blue Iron Oxide (SRI) [43]	$RED / BLUE$

MicaSense Altum multispectral central wavelengths: B, G, R, RE and NIR: 474, 560, 668, 717 and 840 nm.

2.2.4. Spatial Analysis

The soil sample data were randomly split into training (70%) and validation data (30%). Using a compiled set of spectral bands (1), spectral indices (2) and the DSM (3), multiple models were developed using logic based algorithms available in the GEE platform [25] and applied iteratively to four dataset stacks (spectral bands, spectral bands + DSM, spectral indices, spectral indices + DSM). We used logic-based machine learning regression methods to map soil properties, such as, decision tree (i.e., CART), Gradient boosting (XGBoost), and random forest (RF) (Figure 4).

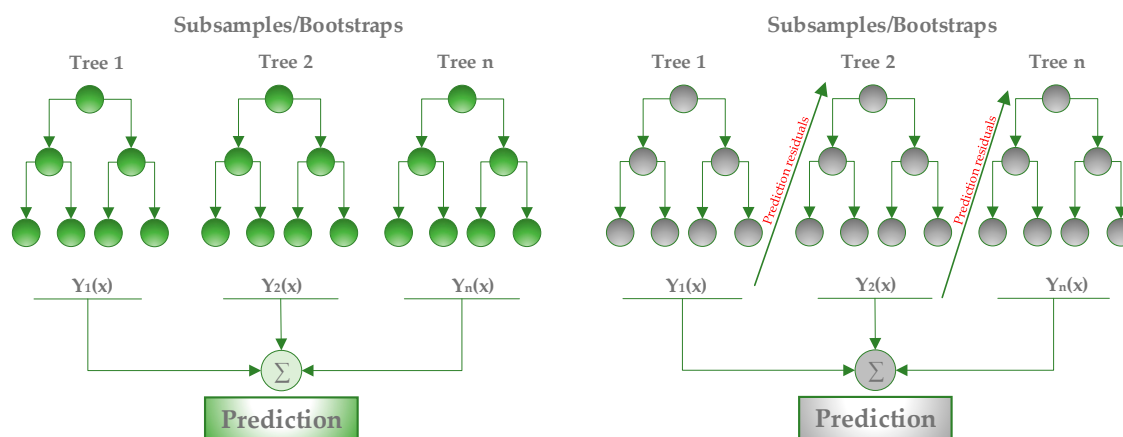


Figure 4. Bagging/random forest example (left), boosting/XG-boost example (right).

CART is a non-parametric algorithm used for classification or regression analysis, with the ability to suppress data noise, using a non-parametric regression method that adds a set of decision trees in a binomial partition [44]. Additionally, regression trees replace the missing data and manage the abnormal data, the hierarchical structure of classification also allows model interactions between predictor variables [45]. However, it is not a stable model in the sense that small changes in the input space can generate a completely different tree. For this reason, CART is used as a base learner to construct more complex models such as RF and XGBoost [46].

Random Forest [47] constructs multiple decision trees that are sampled independently during training, typically improving classification by voting results compared to a single decision tree model. The algorithm makes no assumptions about the data distribution; and can handle scores and continuous variables simultaneously and has good nonlinear data mining capabilities and generalization capabilities [48].

XGBoost is an improvement of the gradient boosting algorithm and has been widely used in classification and regression analysis [49], with generally good accuracy. The decision trees in XGBoost are trained sequentially with adjustments made from the error of the previous tree, while in RF they are built in parallel and independently [46]. In addition, it selects random subsets to fit individual predictors iteratively, in order to obtain the minimized loss function and introduces the stochastic gradient boosting procedure, which can reduce the risk of overfitting and improve the generalization of models with regularization.

For random forest and XGBoost, we defined the number of generated decision trees of 100 leaving the other parameters by default; and for the other classifiers we used the default configuration. A total of 120 models were built between the combination of predictors and input data.

2.2.5. Model Validation and Accuracy Assessment

In order to evaluate the performance of the models developed, an accuracy assessment was conducted to evaluate the performance of regression. The coefficient of determination (R^2), the Root Means Square Error (RMSE), and Mean Absolute Error (MAE) were used to

compare the accuracy of different models. More specifically, the R^2 was used to measure the variation between the measured and predicted soil parameters evaluated; the RMSE was used to assess the magnitude of error between the measurements and the predicted soil parameter. MAE and RMSE express the average prediction error in units of the variable of interest. Regarding validation metrics, the closer R^2 is to 1, and the closer RMSE and MAE are to 0, the better the model fit is considered. To select the best model, we used the higher estimation accuracy, and the smaller error by soil parameter modeled.

$$R^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

where n is the number of samples (individual plot) in the data set, y_i is the measured soil property, \hat{y}_i is the predicted soil property based on the UAS imagery, and \bar{y}_i indicates the average of the measured soil property.

Finally, we used variable importance metrics that consider that every time a split of a node is made on a variable, the impurity criterion for the two descendent nodes is less than the parent node and adding up the decreases for each individual variable over all trees in the forest gives a fast variable importance that is often very consistent with the permutation importance measure. That provided us with an additional means of assessing how each predictor variable enabled accuracy improvements in the optimized soil parameter prediction model, in terms of a normalized percentage contribution.

3. Results

3.1. Descriptive Statistics

The descriptive statistics of the soil properties analyzed are shown in Table 2. Lime, Clay and Sand values ranged from 10.83 to 58.71%, with a balanced texture. The standard deviation (SD) range from 4.45 to 8.46%, and the coefficient of variation (CV) range between 16.72 to 22.04% which indicate moderate variability according to the classification proposed by Wilding and Drees [50].

Table 2. Descriptive statistics of soil Properties.

Soil Property	Minimum	Maximum	Mean	Median	SD	CV (%)
Lime (%)	27.42	58.71	38.08	37.35	6.24	16.72
Clay (%)	10.83	34.78	20.28	20.04	4.42	22.04
Sand (%)	22.58	56.46	41.64	42.30	8.46	20.01
EC (mS/m)	1.58	9.37	3.75	3.57	1.48	41.53
N (ppm)	0.07	0.23	0.12	0.11	0.02	21.06
P (ppm)	7.47	57.88	29.78	30.80	11.11	36.07
K (ppm)	57.88	335.42	107.16	97.80	45.29	46.31
OM (%)	1.48	4.57	2.31	2.29	0.48	21.06
Al (ppm)	0.27	9.46	4.27	4.32	2.56	59.23
pH	5.25	6.88	6.09	6.06	0.35	5.83

EC values in the entire study area varied greatly, ranging from 1.58 to 9.37 mS/m, and the SD and CV were 1.48 mS/m and 41.53%, respectively. Al ranged from 0.27 to 9.46 ppm, and the SD and CV were 2.56% and 59.23%, respectively. K ranged from 57.88 to 335.42 ppm, and the SD and CV were 45.29 ppm and 59.23%, respectively. P ranged from 7.47 to 57.88 ppm, and the SD and CV were 11.11 ppm and 36.07%, respectively. These

samples indicated high variability (CV > 35%) which may be attributed to random factors such as environmental factors and measurement errors [51].

N values ranged from 0.07 to 0.23 ppm and the SD and CV were 0.02 ppm and 21.06%, respectively. OM values ranged from 1.48 to 4.57% and CV were 0.48 ppm and 21.06%, respectively; for both variables indicate a moderate variability.

Soil pH values in the entire study area varied from 5.25 to 6.88, with the mean and median values of 6.09 and 6.06%, respectively, which indicate a low variability. The soil properties in the entire study area were classified as acid.

3.2. Correlation Analysis between Predictors and Soil Properties

Figure 5 shows Pearson’s correlation coefficients (r), between soil properties and predictors composed of spectral bands, spectral indices and the DSM obtained from the Altum imagery calculated with the corrplot library [52], in R environment [53]. Overall, lime shows very low correlations (r = 0–0.19) with most predictors, low correlations (r = 0.20–0.39) with NIR, LWIR; and moderate correlation (r = 0.4–0.59) with DSM. Clay has negative moderated correlations with blue, red, LWIR bands, and NDWI, ExR and NRE; and moderate positive correlations with most spectral indices. Sand shows moderate correlation with LWIR band and high negative correlation (r = 0.6–0.79) with DSM. In summary, textural soil properties expressed by contents of clay, sand and lime, show better correlations with spectral indices and DSM than spectral bands.

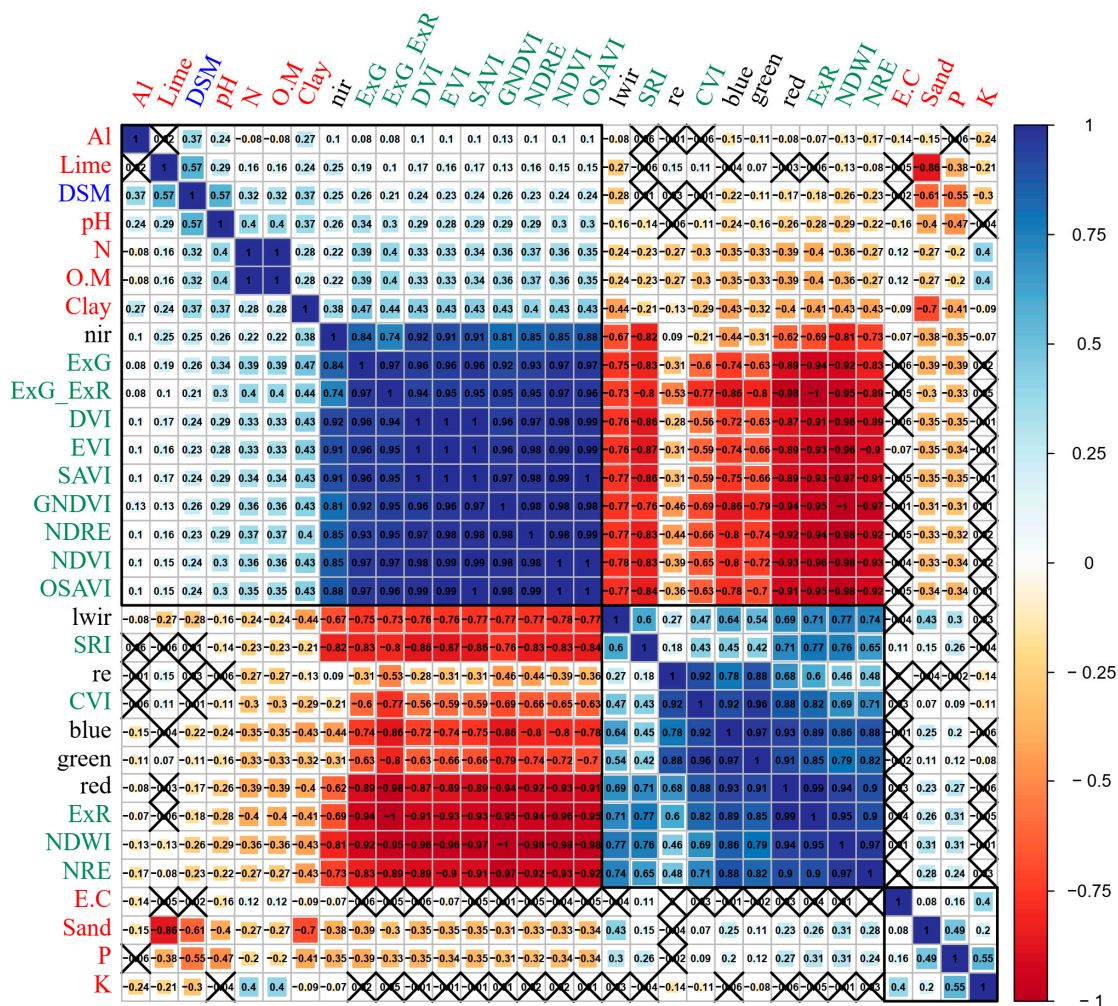


Figure 5. Correlation coefficients between measured soil physical chemical properties and predictors. r—Pearson’s correlation coefficient; Significant at 5% probability; X—non-significant.

N shows low correlation with most spectral and indices bands, and moderate correlation with ExR and ExG_ExR spectral indices. P presents very low and low correlations with spectral bands and indices and moderate correlation with DSM, the spectral indices slightly improve the correlation. K shows non-significant correlation with most predictors except a low correlation with DSM. OM has low correlation for most spectral bands, spectral indices and DSM, and a moderate correlation for ExR index. Al presents a very low correlation with most predictors and a low correlation with DSM. EC shows non-significant correlation with most predictors except a very low correlation with NIR, EVI and SRI index. pH shows very low and low correlations with spectral bands and indices and moderate correlation with DSM, the spectral indices slightly improve the correlation. Pearson's correlation coefficients reveal that in general, the predictors selected gave a poor relationship with soil properties in the study area, spectral indices, and DSM used were better correlated with soil properties than spectral bands.

3.3. Analysis of Modeling Results

The evaluation of the machine learning regression models in the training and validation datasets, response to a combination of predictors bands, spectral indices, and the DSM, the results of accuracy are shown in Table 3.

Soil properties like clay, sand and phosphorus (P) present the highest accuracy (R^2 : 0.89 to 0.91) and smallest errors in RMSE (1.39 to 3.71) and MAE (0.29 to 1.20) when RF and spectral indices (SI) were used together. Lime, nitrogen (N), organic matter (OM), electrical conductivity (EC) and pH present high accuracy (0.81 to 0.92) and smaller errors, RMSE (0.01 to 0.53) and MAE (0.03 to 6.18) when CART and SI are used together. The Al and K content show better prediction accuracy if the DSM predictor is combined with SI (R^2 : 0.89, 0.88), minimizing the error RMSE (0.82, 18.36) and MAE (1.18, 0.001) through the CART model. In general, the selected models had a satisfactory predictive capacity for all the soil properties tested, with slight superiority of CART models combined with SI for Lime, N, OM, EC and pH, and adding DSM is better for K and Al. RF combined with SI shows better performance for clay, sand and P. XGBoost presents the worst performance for all the soil properties evaluated.

3.4. Prediction Results and Relative Importance of the Predictors

Based on the previous results, we selected the best regression models and created maps of the spatial quantitative distribution of each soil property in Figures 6–8. The maps generated by the ten selected models show a gradient concentration of soil properties from the north to south, and in general, the CART and RF models show similar spatial distribution. The relative importance of the predictors (note that the importance value has been converted to percentage) for the selected models with the highest accuracy and small errors, are based on spectral indices, which revealed the similarities in the main predictors in the for CART and RF models evaluated.

Table 3. Evaluation of the prediction effects of the different models in predicting soil properties.

Algorithm	Predictors	Training										Validation									
		Lime	Clay	Sand	N	P	K	OM	Al	EC	pH	Lime	Clay	Sand	N	P	K	OM	Al	EC	pH
R-square																					
CART	SB	0.60	0.52	0.51	0.23	0.45	−0.04	0.16	0.36	−0.14	0.26	0.46	0.49	0.61	0.54	0.46	0.01	0.41	0.28	0.01	0.39
	SB.DSM	0.37	0.33	0.44	0.41	0.28	−0.23	0.43	0.24	0.03	0.19	0.25	0.35	0.40	0.50	0.29	0.01	0.46	0.12	−0.09	0.34
	SI	0.92	0.88	0.89	0.82	0.83	0.55	0.81	0.91	0.89	0.90	0.89	0.86	0.76	0.84	0.89	0.72	0.84	0.88	0.86	0.87
	SI.DSM	0.90	0.72	0.84	0.66	0.87	0.81	0.66	0.91	0.79	0.80	0.89	0.85	0.82	0.85	0.83	0.88	0.85	0.89	0.72	0.92
XGBoost	SB	0.40	0.42	0.43	0.34	0.37	0.20	0.34	0.35	0.21	0.33	0.32	0.41	0.42	0.26	0.37	0.15	0.26	0.35	0.20	0.34
	SB.DSM	0.34	0.39	0.39	0.31	0.35	0.27	0.31	0.34	0.24	0.35	0.30	0.38	0.37	0.26	0.32	0.15	0.26	0.30	0.26	0.32
	SI	0.54	0.53	0.54	0.41	0.53	0.39	0.41	0.54	0.53	0.50	0.50	0.51	0.50	0.30	0.52	0.29	0.30	0.52	0.53	0.50
	SI.DSM	0.51	0.53	0.49	0.40	0.51	0.40	0.40	0.50	0.48	0.48	0.48	0.51	0.46	0.31	0.50	0.34	0.31	0.51	0.47	0.49
RF	SB	0.70	0.72	0.71	0.63	0.66	0.46	0.63	0.66	0.37	0.64	0.59	0.77	0.71	0.60	0.68	0.44	0.60	0.65	0.33	0.65
	SB.DSM	0.66	0.68	0.72	0.60	0.63	0.52	0.60	0.63	0.43	0.64	0.56	0.74	0.67	0.59	0.61	0.45	0.59	0.55	0.42	0.65
	SI	0.89	0.89	0.89	0.82	0.89	0.71	0.82	0.89	0.87	0.89	0.88	0.91	0.89	0.78	0.89	0.64	0.78	0.88	0.86	0.90
	SI.DSM	0.85	0.85	0.87	0.75	0.83	0.71	0.75	0.83	0.76	0.84	0.82	0.87	0.84	0.74	0.82	0.65	0.74	0.81	0.77	0.86
RMSE																					
CART	SB	3.98	2.99	5.93	0.02	8.15	42.18	0.41	2.07	1.61	0.31	4.49	3.32	5.34	0.02	8.29	53.32	0.43	2.12	1.41	0.27
	SB.DSM	4.98	3.54	6.32	0.02	9.37	45.84	0.34	2.25	1.49	0.32	5.28	3.74	6.60	0.02	9.52	53.79	0.41	2.35	1.48	0.28
	SI	1.77	1.47	2.84	0.01	4.51	27.70	0.20	0.77	0.51	0.11	2.00	1.71	4.21	0.01	3.76	28.49	0.22	0.88	0.53	0.12
	SI.DSM	2.03	2.29	3.34	0.01	3.99	18.02	0.26	0.76	0.69	0.16	2.05	1.77	3.66	0.01	4.58	18.36	0.22	0.82	0.75	0.10
XGBoost	SB	4.87	3.29	6.36	0.02	8.74	37.00	0.36	2.08	1.34	0.29	5.01	3.57	6.48	0.02	8.95	49.37	0.48	2.02	1.27	0.28
	SB.DSM	5.11	3.38	6.58	0.02	8.87	35.40	0.37	2.10	1.31	0.29	5.09	3.64	6.77	0.02	9.28	49.39	0.48	2.09	1.21	0.28
	SI	4.27	2.96	5.73	0.02	7.53	32.37	0.34	1.75	1.04	0.25	4.29	3.25	6.03	0.02	7.83	45.03	0.47	1.73	0.97	0.24
	SI.DSM	4.40	2.97	6.01	0.02	7.69	32.03	0.35	1.82	1.09	0.26	4.40	3.23	6.26	0.02	7.97	43.55	0.46	1.75	1.03	0.24
RF	SB	3.44	2.30	4.56	0.01	6.42	30.34	0.27	1.50	1.20	0.22	3.92	2.22	4.62	0.02	6.42	40.19	0.35	1.48	1.16	0.20
	SB.DSM	3.66	2.44	4.44	0.01	6.69	28.55	0.28	1.57	1.14	0.21	4.06	2.38	4.87	0.02	7.01	39.94	0.36	1.69	1.08	0.20
	SI	2.07	1.41	2.78	0.01	3.68	22.19	0.19	0.87	0.55	0.12	2.13	1.39	2.82	0.01	3.71	32.12	0.26	0.88	0.53	0.11
	SI.DSM	2.41	1.68	3.07	0.01	4.50	22.35	0.22	1.05	0.74	0.14	2.61	1.67	3.38	0.01	4.80	31.74	0.28	1.08	0.68	0.13

Table 3. Cont.

Algorithm	Predictors	Training										Validation									
		Lime	Clay	Sand	N	P	K	OM	Al	EC	pH	Lime	Clay	Sand	N	P	K	OM	Al	EC	pH
MAE																					
CART	SB	1.04	1.50	0.71	17.92	1.65	0.01	0.19	3.89	0.16	2.90	1.07	1.65	0.61	23.21	2.08	0.01	0.20	4.10	0.14	2.56
	SB.DSM	1.13	1.90	0.68	19.94	2.46	0.01	0.16	4.95	0.17	3.38	1.15	1.96	0.76	24.20	2.69	0.01	0.19	5.25	0.15	3.67
	SI	0.18	0.39	0.13	5.36	0.40	0.001	0.05	1.30	0.03	0.73	0.21	0.47	0.15	6.18	0.40	0.00	0.04	0.88	0.03	1.09
	SI.DSM	0.21	0.87	0.22	3.51	0.46	0.001	0.09	1.15	0.05	0.89	0.20	0.70	0.22	3.32	0.44	0.00	0.08	1.18	0.03	0.90
XGBoost	SB	1.60	2.55	0.88	21.95	3.65	0.01	0.24	6.55	0.22	5.27	1.53	2.82	0.83	26.68	3.66	0.01	0.29	6.79	0.21	5.40
	SB.DSM	1.60	2.58	0.87	21.37	3.78	0.01	0.25	6.71	0.22	5.41	1.59	2.87	0.82	26.89	3.78	0.01	0.29	7.14	0.21	5.63
	SI	1.34	2.21	0.70	18.79	3.19	0.01	0.22	5.65	0.19	4.67	1.31	2.49	0.65	23.14	3.19	0.01	0.27	5.92	0.18	5.01
	SI.DSM	1.37	2.23	0.72	18.69	3.22	0.01	0.22	5.85	0.19	4.83	1.32	2.47	0.67	23.36	3.20	0.01	0.27	6.14	0.18	5.08
RF	SB	1.01	1.54	0.68	16.94	2.29	0.01	0.16	4.13	0.14	2.98	0.98	1.55	0.67	21.04	2.51	0.01	0.19	4.29	0.14	2.91
	SB.DSM	1.05	1.64	0.67	17.14	2.39	0.01	0.17	4.48	0.14	2.95	1.08	1.68	0.65	21.38	2.58	0.01	0.20	4.83	0.14	3.24
	SI	0.53	0.86	0.31	10.06	1.20	0.00	0.10	2.15	0.07	1.58	0.53	0.89	0.29	12.84	1.17	0.01	0.12	2.26	0.07	1.60
	SI.DSM	0.69	1.10	0.42	11.92	1.56	0.01	0.13	2.90	0.10	1.84	0.70	1.14	0.40	15.40	1.67	0.01	0.15	3.15	0.09	2.10

SB: Spectral bands (blue, green, red, red edge, nir, lwir), SI: Spectral indices (NDVI, EVI, NDWI, SAVI, GNDVI, DVI, OSAVI, ExG, ExR, ExG-ExR, NDI, NDRE, CVI, SRI), DSM: Digital surface model.

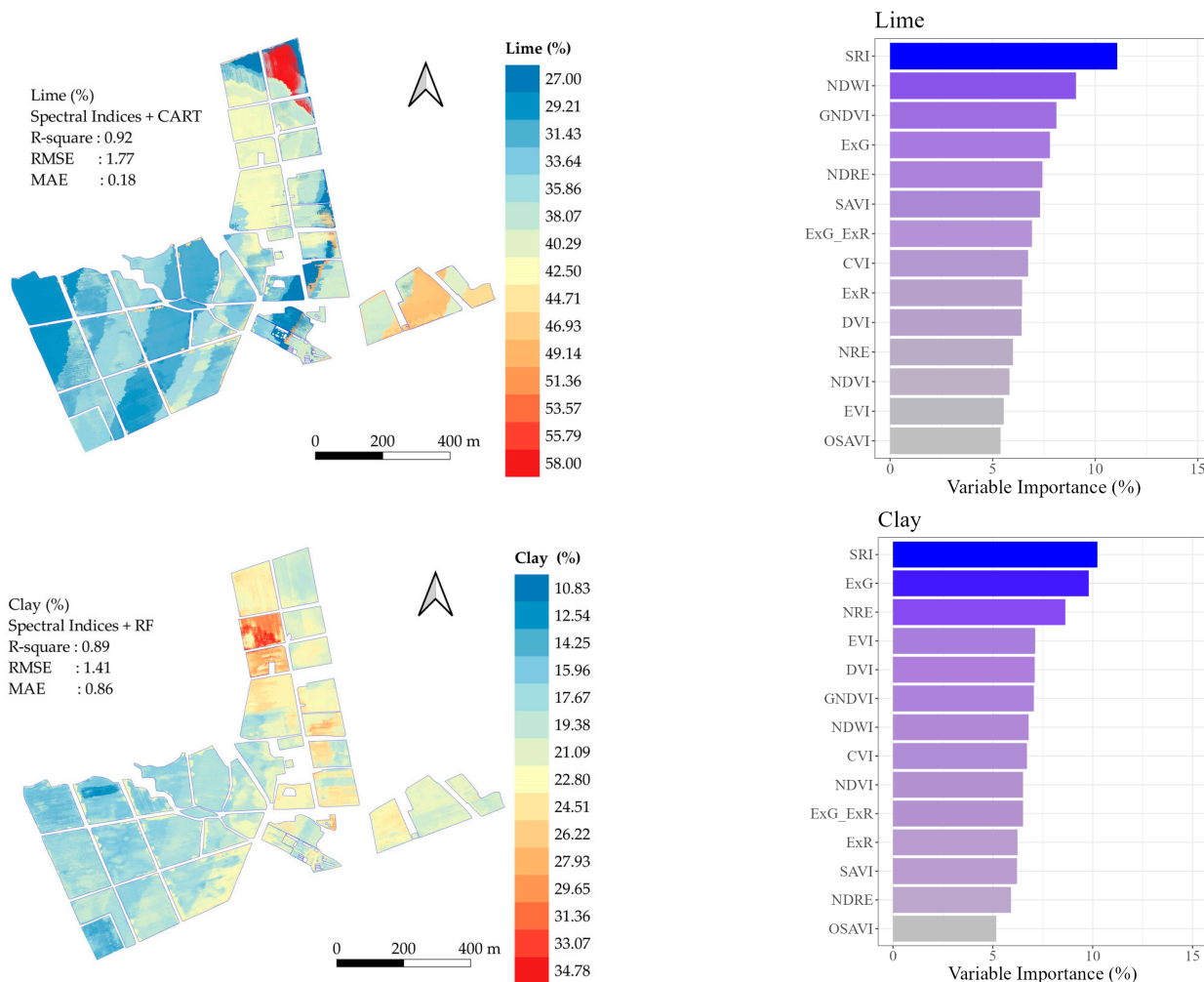


Figure 6. Prediction maps for Lime and Clay (left) and relative importance of predictors (right).

The SRI index was the main explanatory predictor for most models (already 10% of the total relative importance), except for K and Al; for which the DSM is the main explanatory predictor. For lime, clay and sand the predictors ExG, GNDVI, NDWI, NRE and EVI, showed different hierarchical characteristics in the first four explanatory predictors, and the maps for clay and sand show an opposite density distribution, due to the location of the study area in an alluvial fan. For N and P, the spectral indices are the most important variables, but not for K and Al, for which the second variable importance is still SRI. OM has GNDVI and EXG_ExR and NDWI, as the first four important variables. These results are similar for EC and pH. The major importance of predictors for all properties was not consistent with the Pearson correlation results.

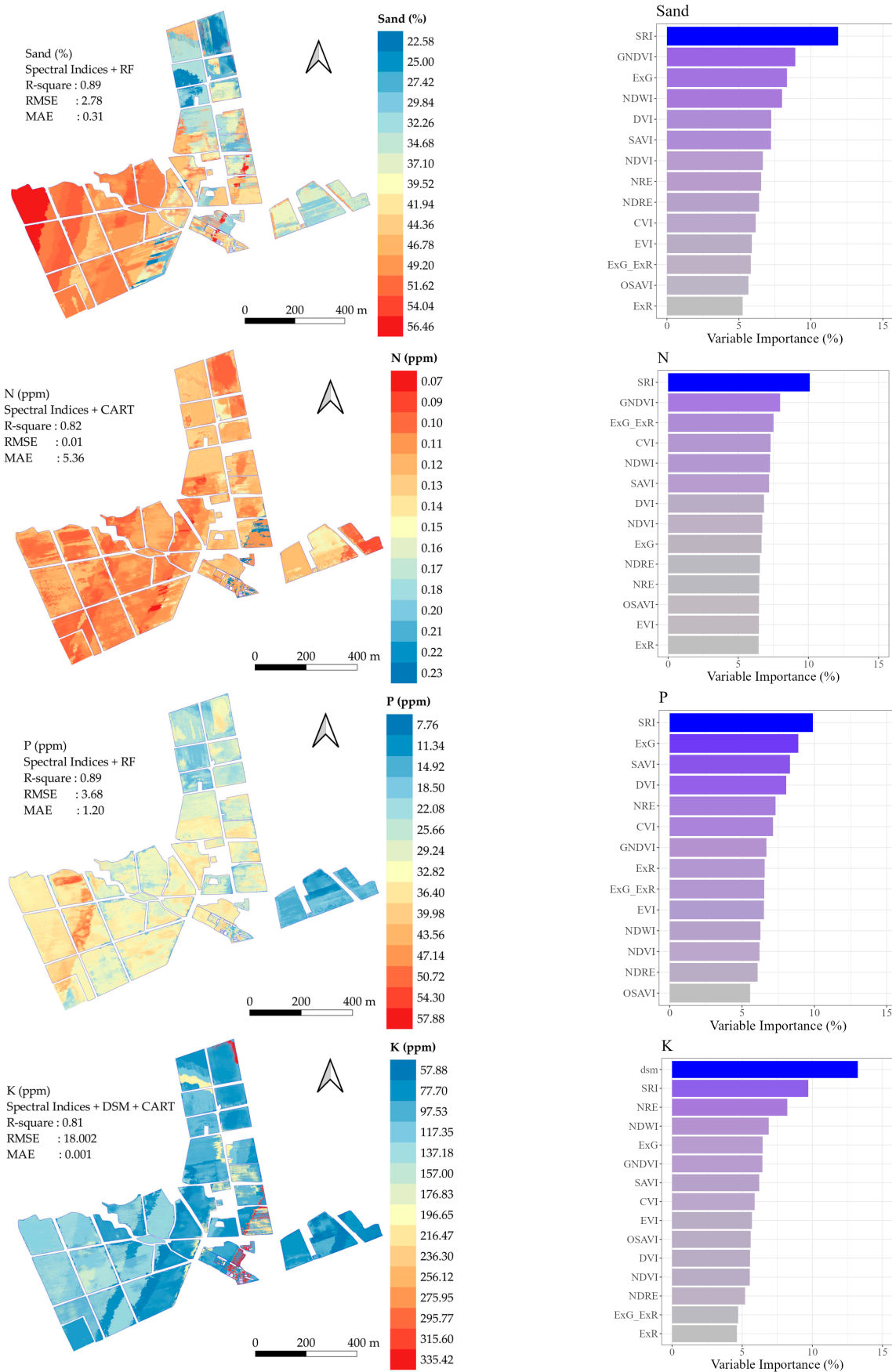


Figure 7. Prediction maps for sand, N, P and K (left) and relative importance of predictors (right).

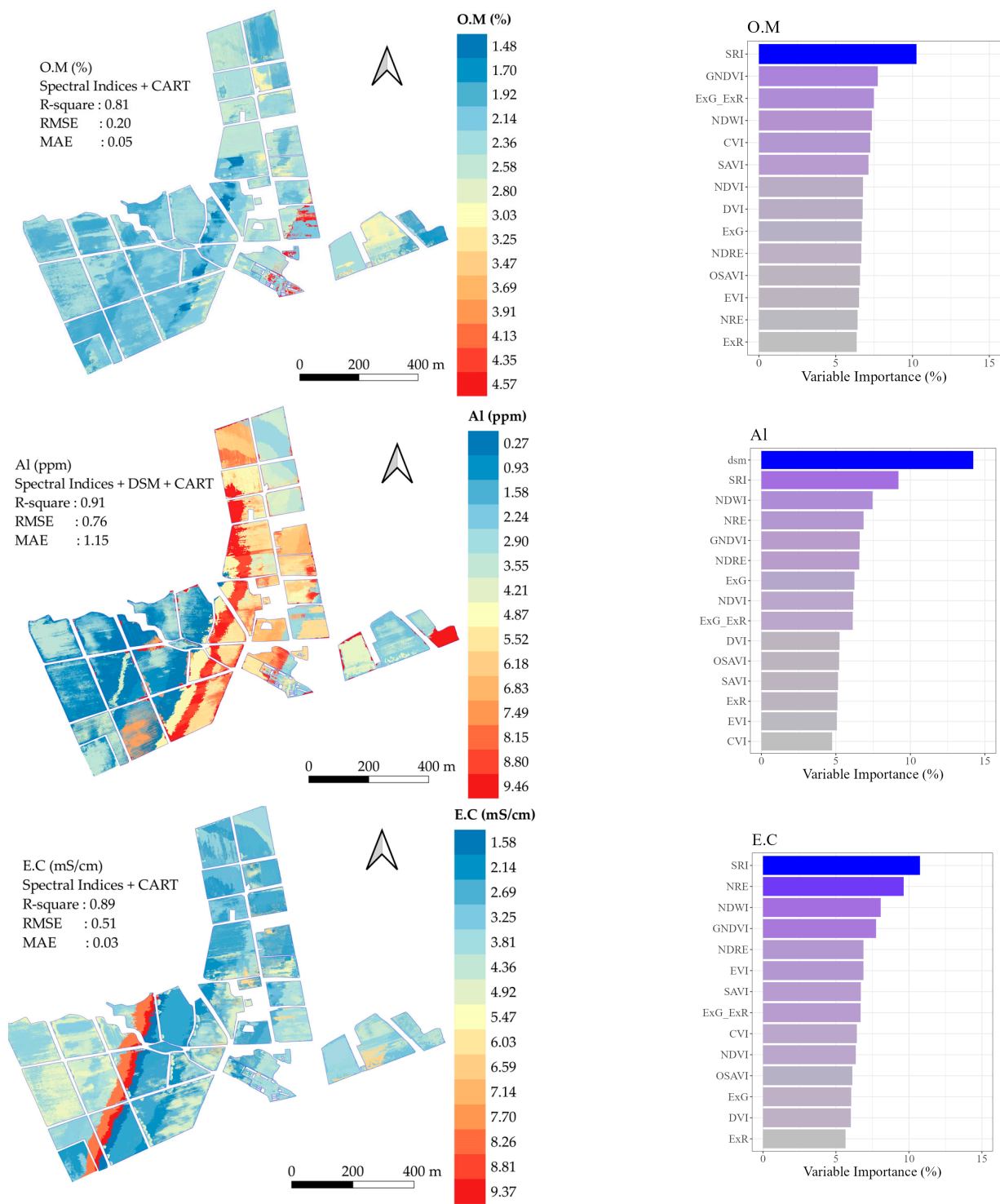


Figure 8. Cont.

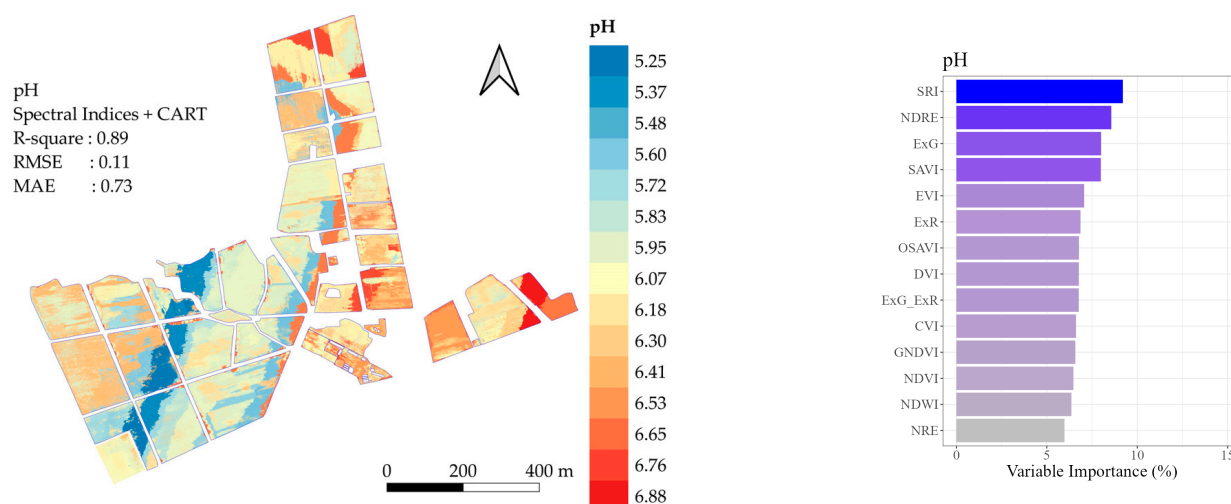


Figure 8. Prediction maps for OM, Al, EC and pH (left) and relative importance of predictors (right).

4. Discussion

The sustainable intensification of agriculture will be necessary to meet the growing global demand for food and while conventional agricultural intensification can lead to negative environmental impacts, sustainable intensification, which includes optimal water, soil, and crop management, as well as integrated crop-livestock management, can increase yields while reducing environmental impacts [54,55]. Although most studies using UAVs have focused on monitoring crops based on vegetation parameters and phenology [56], our results showed that it is feasible to use UAV-based multispectral imagery in an effective way to predict soil properties for agricultural lands, using open-access GEE platform through regression models using machine learning as Random Forests, CART and XGBoost without having to purchase or download a software [20,57]. Although the use of GEE is time-consuming in data preparation [58], the training phase and model construction is faster than traditional computing methods for UAV-based imagery [59]. Raster datasets are easily processed in parallel by subdividing an area into tiles [25], additionally the physical resources needed are smaller.

Considering the low temporal variation of soil properties, we designed the study in the dry season in order to avoid any potential pitfalls caused by the weather and potential confusion introduced by soil moisture.

We contribute to prior work on UAV-derived measurements of soil physical and chemical characteristics, yet we compute significantly more metrics than prior work has accomplished. For instance, prior studies have extracted variables such as sand, silt, clay, cation exchange capacity (CEC), soil organic carbon (SOC) and nitrogen [23], organic matter content [20], clay content [21], and soil moisture content [22]. However, none of these prior studies attempted to conduct a full-spectrum assessment of the spatial characteristics of both physical and chemical soil properties as conducted in this work.

Considering the correlograms and performance of the extracted bands and calculated spectral indices (Figure 4), in some cases it is common to find a negative R-square when the model performs worse than the mean of the observations as predictors [60,61], however the use of spectral indices as predictors improve the correlations with soil properties. These results can be attributed mainly to the presence of typical absorption features for soil organic matter in the VIS and NIR spectral regions, respectively [17]. Bogrekci and Lee [62], reported monitoring systems to detect phosphorus in soil using diffuse reflectance spectroscopy in the ultraviolet (UV), VIS, and NIR regions. Similarly, available phosphorus was detected in the range of 300–700 nm [63]. Soil organic matter is better correlated with spectral indices, likely given that the spectral regions of vegetation indices were more sensitive to changes in soil organic carbon and clay than the other indices [44]. The addition

of three-dimensional data (such as the DSM) improves correlations for soil properties, likely because this predictor is related to soil variability and depends on the morphology of the landscape, regularly used in soil mapping because of the critical role that landscape morphology plays in soil formation [64].

Since machine learning models do not require an assumption of normality [65], data transformations were not performed for the model used in this work, and RF is considered a method that reduces the uncertainty of CART, because RF averages a group of fully grown trees, and work for a large collection of de-correlated, noisy, approximately unbiased trees are built, the average of the trees reduce the model variance and the uncertainty, and was used on mapping soils [58,66,67].

However, some studies showed that CART has better performance for pH, compared with RF [60,68]. When the interpretability of the resulting model is important for the user, logical-based machine learning models are more appropriate, as they do not function as “black boxes” [69], and the main advantage is that the former provides an estimate of the relative importance of the predictors in the model, and avoids the elimination of predictive covariates that may be relevant for soil, even if there are correlations between them [70]. Furthermore, it is necessary to consider that differences in soil condition, particularly moisture, have a significant effect on the composition and amount of reflected and emitted energy from a soil surface, which reduces the reflectance over the entire spectrum [71]. Moreover, several factors including soil roughness, crop residues, and tillage can generate variability in soil structure and further complicate reflectance values collected from close-range remote sensing [72].

In addition, it is also important to map the spatial error at the pixel level, more than just show a statistical metric as RMSE [73], but is mandatory a systematic sampling that can capture the spatial variability at a small scale, but this detailed soil survey should be expensive in time and cost [74].

Although our results show that the combination of multiple spectral indices, and topography can effectively predict soil properties, further improvements are still necessary. Specifically, given that field-sampled values of soil fertility properties such as N, P and K can change significantly both within and between every crop season as a function of variability in soil treatments. In order to reduce the uncertainty and improve the prediction accuracy of similar prediction modeling, it would be necessary to make continuous field soil evaluations to capture the spatial variability in these soil properties over short time periods.

5. Conclusions

Our work contributes to the current cutting-edge science highlighting the benefits of using high-resolution imagery collected using multispectral sensors onboard UAVs for precision agriculture and highly detailed soil information systems. We demonstrated how machine learning algorithms computed in the cloud using Google Earth Engine can be a solution to make processing more accessible without the use of physical servers, for predicting soil properties at a detailed level with satisfactory results. We found that UAV-derived multispectral indices can improve soil properties prediction when combined with digital surface models constructed from UAV imagery and that the most significant predictors are SRI, GNDWI, NDWI, and ExG. Lastly, by comparing three machine learning techniques (CART, RF, and XGBoost), we demonstrated that CART models perform better and are more spatially consistent than RF and XGBoost models for most of the soil properties we investigated. These results suggest that the application of machine learning algorithms with ground-truth data augmentation is effective in the spatial estimation of soil properties using UAV-based multispectral imagery and can contribute to more efficient and effective crop and agricultural field management.

Author Contributions: S.P. and C.C. designed the methodology; J.V., M.Q. and H.L. provided and validated field data, J.V. and L.A. (Lino Achallma) collected soil samples and prepare test reports, L.A. (Lidiana Alejandro) and I.G. performed laboratory analyses for all physicochemical parameters; S.P., C.C. and D.F. performed the data processing; S.P. and C.C. analyzed the data; W.S., J.C. and C.I.A. funding acquisition, S.P., N.G.P., D.F., C.C. and M.Q. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the following two projects “Mejoramiento de los servicios de investigación y transferencia tecnológica en el manejo y recuperación de suelos agrícolas degradados y aguas para riego en la pequeña y mediana agricultura en los departamentos de Lima, Áncash, San Martín, Cajamarca, Lambayeque, Junín, Ayacucho, Arequipa, Puno y Ucayali” CUI 2487112 and “Creación del servicio de agricultura de precisión en los Departamentos de Lambayeque, Huancavelica, Ucayali y San Martín 4 Departamentos” CUI 2449640 of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government.

Data Availability Statement: The data presented in this study are openly available in Dataverse at <https://doi.org/10.21223/PKPXQF>.

Acknowledgments: Santa Ana’s LABSAF and AGPRES teams for providing infrastructure and equipment for the soil data collection and laboratory analysis. We thank STC project “Precision agriculture: determination of aerial biomass and yield of corn (*Zea mays*) and wheat (*Triticum aestivum*) crop using machine learning applied to unmanned aerial vehicle images”. C.I.A. thanks Vicerrectorado de Investigación of UNTRM.

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

References

1. Sona, G.; Passoni, D.; Pinto, L.; Pagliari, D.; Masseroni, D.; Ortuani, B.; Facchi, A. UAV Multispectral Survey to Map Soil and Crop for Precision Farming Applications. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.-ISPRS Arch.* **2016**, *2016*, 1023–1029. [[CrossRef](#)]
2. Porta, J.; López, M.; Roquero, C. *Edafología Para La Agricultura y El Medio Ambiente*; Ediciones Mundi-Prensa: Madrid, Spain, 2003.
3. Corwin, D.L.; Lesch, S.M.; Shouse, P.J.; Soppe, R.; Ayars, J.E. Identifying Soil Properties That Influence Cotton Yield Using Soil Sampling Directed. *Agron. J.* **2003**, *95*, 352–364. [[CrossRef](#)]
4. Srinet, R.; Nandy, S.; Padalia, H.; Ghosh, S.; Watham, T.; Patel, N.R.; Chauhan, P. Mapping Plant Functional Types in Northwest Himalayan Foothills of India Using Random Forest Algorithm in Google Earth Engine. *Int. J. Remote Sens.* **2020**, *41*, 7296–7309. [[CrossRef](#)]
5. Das, B.S.; Sarathjith, M.C.; Santra, P.; Sahoo, R.N.; Srivastava, R.; Routray, A.; Ray, S.S. Hyperspectral Remote Sensing: Opportunities, Status and Challenges for Rapid Soil Assessment in India. *Curr. Sci.* **2015**, *108*, 860–868.
6. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
7. Wang, D.; Wan, B.; Liu, J.; Su, Y.; Guo, Q.; Qiu, P.; Wu, X. Estimating Aboveground Biomass of the Mangrove Forests on Northeast Hainan Island in China Using an Upscaling Method from Field Plots, UAV-LiDAR Data and Sentinel-2 Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *85*, 101986. [[CrossRef](#)]
8. Deery, D.M.; Rebetzke, G.J.; Jimenez-Berni, J.A.; James, R.A.; Condon, A.G.; Bovill, W.D.; Hutchinson, P.; Scarrow, J.; Davy, R.; Furbank, R.T. Methodology for High-Throughput Field Phenotyping of Canopy Temperature Using Airborne Thermography. *Front. Plant Sci.* **2016**, *7*, 1808. [[CrossRef](#)]
9. Prashar, A.; Jones, H.G. Infra-Red Thermography as a High-Throughput Tool for Field Phenotyping. *Agronomy* **2014**, *4*, 397–417. [[CrossRef](#)]
10. Jindo, K.; Teklu, M.G.; van Boheeman, K.; Njehia, N.S.; Narabu, T.; Kempenaar, C.; Molendijk, L.P.G.; Schepel, E.; Been, T.H. Unmanned Aerial Vehicle (UAV) for Detection and Prediction of Damage Caused by Potato Cyst Nematode *G. Pallida* on Selected Potato Cultivars. *Remote Sens.* **2023**, *15*, 1429. [[CrossRef](#)]
11. Luo, L.; Chang, Q.; Wang, Q.; Huang, Y. Identification and Severity Monitoring of Maize Dwarf Mosaic Virus Infection Based on Hyperspectral Measurements. *Remote Sens.* **2021**, *13*, 4560. [[CrossRef](#)]
12. Cheng, M.; Jiao, X.; Shi, L.; Penuelas, J.; Kumar, L.; Nie, C.; Wu, T.; Liu, K.; Wu, W.; Jin, X. High-Resolution Crop Yield and Water Productivity Dataset Generated Using Random Forest and Remote Sensing. *Sci. Data* **2022**, *9*, 641. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, W.; Wang, K.; Chen, H.; He, X.; Zhang, J. Ancillary Information Improves Kriging on Soil Organic Carbon Data for a Typical Karst Peak Cluster Depression Landscape. *J. Sci. Food Agric.* **2012**, *92*, 1094–1102. [[CrossRef](#)]
14. Zhang, Y.; Han, W.; Zhang, H.; Niu, X.; Shao, G. Evaluating Soil Moisture Content under Maize Coverage Using UAV Multimodal Data by Machine Learning Algorithms. *J. Hydrol.* **2023**, *617*, 129086. [[CrossRef](#)]
15. Heil, J.; Jörges, C.; Stumpe, B. Fine-Scale Mapping of Soil Organic Matter in Agricultural Soils Using UAVs and Machine Learning. *Remote Sens.* **2022**, *14*, 3349. [[CrossRef](#)]

16. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sens.* **2017**, *9*, 1110. [[CrossRef](#)]
17. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* **2006**, *131*, 59–75. [[CrossRef](#)]
18. Francos, N.; Romano, N.; Nasta, P.; Zeng, Y.; Szabó, B.; Manfreda, S.; Ciraolo, G.; Mészáros, J.; Zhuang, R.; Su, B.; et al. Mapping Water Infiltration Rate Using Ground and Uav Hyperspectral Data: A Case Study of Alento, Italy. *Remote Sens.* **2021**, *13*, 2606. [[CrossRef](#)]
19. Hassan-Esfahani, L. *High Resolution Multi-Spectral Imagery and Learning Machines in Precision Irrigation Water Management*; Utah State University: Logan, UT, USA, 2015; p. 153.
20. Zhou, J.; Xu, Y.; Gu, X.; Chen, T.; Sun, Q.; Zhang, S.; Pan, Y. High-Precision Mapping of Soil Organic Matter Based on UAV Imagery Using Machine Learning Algorithms. *Drones* **2023**, *7*, 290. [[CrossRef](#)]
21. Shabou, M.; Mougnot, B.; Chabaane, Z.L.; Walter, C.; Boulet, G.; Aissa, N.B.; Zribi, M. Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi-Arid Lands. *Remote Sens.* **2015**, *7*, 6059–6078. [[CrossRef](#)]
22. Matese, A.; Toscano, P.; Di Gennaro, S.; Genesio, L.; Vaccari, F.; Primicerio, J.; Belli, C.; Zaldei, A.; Bianconi, R.; Gioli, B. Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. *Remote Sens.* **2015**, *7*, 2971–2990. [[CrossRef](#)]
23. Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLoS ONE* **2017**, *12*, e0170478. [[CrossRef](#)]
24. Keskin, H.; Grunwald, S. Regression Kriging as a Workhorse in the Digital Soil Mapper’s Toolbox. *Geoderma* **2018**, *326*, 22–41. [[CrossRef](#)]
25. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
26. Instituto Geofísico del Perú. *Atlas Climático de Precipitación y Temperatura Del Aire En La Cuenca Del Río Mantaro*; Fondo Editorial del Consejo Nacional del Ambiente—CONAM, Ed.; Instituto Geofísico del Perú: Lima, Perú, 2005.
27. Brus, D.J.; Kempen, B.; Heuvelink, G.B.M. Sampling for Validation of Digital Soil Maps. *Eur. J. Soil Sci.* **2011**, *62*, 394–407. [[CrossRef](#)]
28. US Environmental Protection Agency Method 9045D Soil and Waste PH. 2004.
29. International Standard Organisation (ISO). Soil Quality: Determination of the Specific Electrical Conductivity. 1996. Available online: <https://www.iso.org/standard/19243.html> (accessed on 10 May 2023).
30. Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). Norma Oficial Mexicana NOM-021-RECNAT-2000. 2002. Available online: <http://www.ordenjuridico.gob.mx/Documentos/Federal/wo69255.pdf> (accessed on 10 May 2023).
31. Rouse, J.; Haas, R.; Schell, J.; Deering, D. Monitoring Vegetation Systems in the Great Plains with ERTS. In Proceedings of the Third Earth Resources Technology Satellite Symposium, Washington, DC, USA, 10–14 December 1974; Volume 351, p. 309.
32. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [[CrossRef](#)]
33. McFeeters, S.K. The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432. [[CrossRef](#)]
34. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS- MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
35. Richardson, A.J.; Everitt, J.H. Using Spectral Vegetation Indices to Estimate Rangeland Productivity. *Geocarto Int.* **1992**, *7*, 63–69. [[CrossRef](#)]
36. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
37. Woebbecke, D.M.; Meyer, G.E.; Von Bargen, K.; Mortensen, D.A. Color Indices for Weed Identification under Various Soil, Residue, and Lighting Conditions. *Trans. Am. Soc. Agric. Eng.* **1995**, *38*, 259–269. [[CrossRef](#)]
38. Hindman, T.; Meyer, G.E. Machine Vision Detection Parameters for Plant Species Identification. *Syst. Eng.* **1998**, *3543*, 327–335.
39. Meyer, G.E.; Neto, J.C. Verification of Color Vegetation Indices for Automated Crop Imaging Applications. *Comput. Electron. Agric.* **2008**, *63*, 282–293. [[CrossRef](#)]
40. Bannari, A.; Morin, D.; Bonn, F.; Huete, A.R. A Review of Vegetation Indices. *Remote Sens. Rev.* **1995**, *13*, 95–120. [[CrossRef](#)]
41. Gitelson, A.; Merzlyak, M.N. Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus Hippocastanum* L. and *Acer Platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *J. Plant Physiol.* **1994**, *143*, 286–292. [[CrossRef](#)]
42. Vincini, M.; Frazzi, E.; D’Alessio, P. A Broad-Band Leaf Chlorophyll Vegetation Index at the Canopy Scale. *Precis. Agric.* **2008**, *9*, 303–319. [[CrossRef](#)]
43. Hewson, R.D.; Cudahy, T.J.; Huntington, J.F. Geologic and Alteration Mapping at Mt Fitton, South Australia, Using ASTER Satellite-Borne Data. *Int. Geosci. Remote Sens. Symp.* **2001**, *2*, 724–726. [[CrossRef](#)]

44. Jin, X.; Du, J.; Liu, H.; Wang, Z.; Song, K. Remote Estimation of Soil Organic Matter Content in the Sanjiang Plain, Northeast China: The Optimal Band Algorithm versus the GRA-ANN Model. *Agric. For. Meteorol.* **2016**, *218–219*, 250–260. [CrossRef]
45. Schuler, U.; Herrmann, L.; Ingwersen, J.; Erbe, P.; Stahr, K. Comparing Mapping Approaches at Subcatchment Scale in Northern Thailand with Emphasis on the Maximum Likelihood Approach. *Catena* **2010**, *81*, 137–171. [CrossRef]
46. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A Review of Machine Learning Applications in Wildfire Science and Management. *Environ. Rev.* **2020**, *28*, 478–505. [CrossRef]
47. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
48. Gholizadeh, A.; Žižala, D.; Saberioon, M.; Borůvka, L. Soil Organic Carbon and Texture Retrieving and Mapping Using Proximal, Airborne and Sentinel-2 Spectral Imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [CrossRef]
49. Mayr, A.; Binder, H.; Gefeller, O.; Schmid, M. The Evolution of Boosting Algorithms. *Methods Inf. Med.* **2014**, *53*, 419–427. [CrossRef] [PubMed]
50. Wilding, L.P.; Drees, L.R. Spatial Variability and Pedology. *Dev. Soil Sci.* **1983**, *11*, 83–116.
51. Reza, S.K.; Nayak, D.C.; Chattopadhyay, T.; Mukhopadhyay, S.; Singh, S.K.; Srinivasan, R. Spatial Distribution of Soil Physical Properties of Alluvial Soils: A Geostatistical Approach. *Arch. Agron. Soil Sci.* **2016**, *62*, 972–981. [CrossRef]
52. Wei, T.; Simko, V. Corrplot: Visualization of a Correlation Matrix (Version 0.84) 2017, 18. Available online: <https://github.com/taiyun/corrplot> (accessed on 5 February 2023).
53. R Core Team R: A Language and Environment for Statistical Computing 2021. Available online: <https://www.R-project.org> (accessed on 5 February 2023).
54. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global Food Demand and the Sustainable Intensification of Agriculture. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20260–20264. [CrossRef] [PubMed]
55. Herrero, M.; Thornton, P.K.; Notenbaert, A.; Msangi, S.; Wood, S.; Kruska, R.; Dixon, J.; Bossio, D.; Steeg, J.; van de Freeman, H.A.; et al. *Drivers of Change in Crop–Livestock Systems and Their Potential Impacts on Agro-Ecosystems Services and Human Wellbeing to 2030*; ILRI: Nairobi, Kenya, 2012.
56. van der Merwe, D.; Burchfield, D.R.; Witt, T.D.; Price, K.P.; Sharda, A. *Drones in Agriculture*, 1st ed.; Elsevier Inc.: Amsterdam, The Netherlands, 2020; Volume 162, ISBN 9780128207673.
57. Padarian, J.; Minasny, B.; McBratney, A.B. Using Google’s Cloud-Based Platform for Digital Soil Mapping. *Comput. Geosci.* **2015**, *83*, 80–88. [CrossRef]
58. Ganerød, A.J.; Bakkestuen, V.; Calovi, M.; Fredin, O.; Rød, J.K. Where Are the Outcrops? Automatic Delineation of Bedrock from Sediments Using Deep-Learning Techniques. *Appl. Comput. Geosci.* **2023**, *18*, 100119. [CrossRef]
59. Bennett, M.K.; Younes, N.; Joyce, K. Automating Drone Image Processing to Map Coral Reef Substrates Using Google Earth Engine. *Drones* **2020**, *4*, 50. [CrossRef]
60. Hengl, T.; Heuvelink, G.B.M.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; MacMillan, R.A.; De Jesus, J.M.; Tamene, L.; et al. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLoS ONE* **2015**, *10*, e0125814. [CrossRef]
61. Keshavarzi, A.; del Árbol, M.Á.S.; Kaya, F.; Gyasi-Agyei, Y.; Rodrigo-Comino, J. Digital Mapping of Soil Texture Classes for Efficient Land Management in the Piedmont Plain of Iran. *Soil Use Manag.* **2022**, *38*, 1705–1735. [CrossRef]
62. Bogrekcı, I.; Lee, W.S. Spectral Soil Signatures and Sensing Phosphorus. *Biosyst. Eng.* **2005**, *92*, 527–533. [CrossRef]
63. Maleki, M.R.; Mouazen, A.M.; De Ketelaere, B.; Ramon, H.; De Baerdemaeker, J. On-the-Go Variable-Rate Phosphorus Fertilisation Based on a Visible and near-Infrared Soil Sensor. *Biosyst. Eng.* **2008**, *99*, 35–46. [CrossRef]
64. Cavazzi, S.; Corstanje, R.; Mayr, T.; Hannam, J.; Fealy, R. Are Fine Resolution Digital Elevation Models Always the Best Choice in Digital Soil Mapping? *Geoderma* **2013**, *195–196*, 111–121. [CrossRef]
65. Hengl, T.; Macmillan, R.A. *Predictive Soil Mapping with R*; Lulu.Com: Morrisville, NC, USA, 2019; ISBN 978-0-359-30635-0.
66. Ma, G.; Ding, J.; Han, L.; Zhang, Z.; Ran, S. Digital Mapping of Soil Salinization Based on Sentinel-1 and Sentinel-2 Data Combined with Machine Learning Algorithms. *Reg. Sustain.* **2021**, *2*, 177–188. [CrossRef]
67. Nussbaum, M.; Spiess, K.; Baltensweiler, A.; Grob, U.; Keller, A.; Greiner, L.; Schaepman, M.E.; Papritz, A. Evaluation of Digital Soil Mapping Approaches with Large Sets of Environmental Covariates. *Soil* **2018**, *4*, 1–22. [CrossRef]
68. Egelberg, J.; Pena, N.; Rivera, R.; Andruk, C. Assessing the Geographic Specificity of PH Prediction by Classification and Regression Trees. *PLoS ONE* **2021**, *16*, e0255119. [CrossRef]
69. Khaledian, Y.; Miller, B.A. Selecting Appropriate Machine Learning Methods for Digital Soil Mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. [CrossRef]
70. Akpa, S.I.C.; Odeh, I.O.A.; Bishop, T.F.A.; Hartemink, A.E. Digital Mapping of Soil Particle-Size Fractions for Nigeria. *Soil Sci. Soc. Am. J.* **2014**, *78*, 1953–1966. [CrossRef]
71. Nocita, M.; Stevens, A.; Noon, C.; Van Wesemael, B. Prediction of Soil Organic Carbon for Different Levels of Soil Moisture Using Vis-NIR Spectroscopy. *Geoderma* **2013**, *199*, 37–42. [CrossRef]
72. Gelaw, A.M.; Singh, B.R.; Lal, R. Organic Carbon and Nitrogen Associated with Soil Aggregates and Particle Sizes Under Different Land Uses in Tigray, Northern Ethiopia. *L. Degrad. Dev.* **2015**, *26*, 690–700. [CrossRef]

73. Zhang, M.; Zhang, M.; Yang, H.; Jin, Y.; Zhang, X.; Liu, H. Mapping Regional Soil Organic Matter Based on Sentinel-2a and Modis Imagery Using Machine Learning Algorithms and Google Earth Engine. *Remote Sens.* **2021**, *13*, 2934. [[CrossRef](#)]
74. Zhao, Z.; Ashraf, M.I.; Meng, F.R. Model Prediction of Soil Drainage Classes over a Large Area Using a Limited Number of Field Samples: A Case Study in the Province of Nova Scotia, Canada. *Can. J. Soil Sci.* **2013**, *93*, 73–78. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.