Article

Mapping Land Cover Types for Highland Andean Ecosystems in Peru Using Google Earth Engine

Samuel Edwin Pizarro [1,*], Narcisa Gabriela Pricope [2], Daniella Vargas-Machuca [3], Olwer Huanca [4] and Javier Ñaupari [1]

[1] Rangeland Ecology and Utilization Laboratory, Department of Animal Production, Universidad Nacional Agraria La Molina, Av. La Molina SN, Lima 15000, Peru; jnaupariv@lamolina.edu.pe
[2] Department of Earth and Ocean Sciences, University of North Carolina Wilmington, 601 S College Rd., Wilmington, NC 28403, USA; pricopen@uncw.edu
[3] Instituto de Montaña, Vargas Machuca 408, Miraflores, Lima 15000, Peru; dvargas@mountain.org
[4] Nor Yauyos Cochas Landscape Reserve, Servicio Nacional de Áreas Naturales Protegidas, Av. Huancavelica 3113—Covica, Huancayo 12001, Peru; ohuanca@sernanp.gob.pe
* Correspondence: 20171457@lamolina.edu.pe; Tel.: +51-963952025

Abstract: Highland Andean ecosystems sustain high levels of floral and faunal biodiversity in areas with diverse topography and provide varied ecosystem services, including the supply of water to cities and downstream agricultural valleys. Google (™) has developed a product specifically designed for mapping purposes (Earth Engine), which enables users to harness the computing power of a cloud-based solution in near-real time for land cover change mapping and monitoring. We explore the feasibility of using this platform for mapping land cover types in topographically complex terrain with highly mixed vegetation types (Nor Yauyos Cochas Landscape Reserve located in the central Andes of Peru) using classification machine learning (ML) algorithms in combination with different sets of remote sensing data. The algorithms were trained using 3601 sampling pixels of (a) normalized spectral bands between the visible and near infrared spectrum of the Landsat 8 OLI sensor for the 2018 period, (b) spectral indices of vegetation, soil, water, snow, burned areas and bare ground and (c) topographic-derived indices (elevation, slope and aspect). Six ML algorithms were tested, including CART, random forest, gradient tree boosting, minimum distance, naïve Bayes and support vector machine. The results reveal that ML algorithms produce accurate classifications when spectral bands are used in conjunction with topographic indices, resulting in better discrimination among classes with similar spectral signatures such as *pajonal* (tussock grass-dominated cover) and short grasses or rocky groups, and moraines, agricultural and forested areas. The model with the highest explanatory power was obtained from the combination of spectral bands and topographic indices using the random forest algorithm (Kappa = 0.81). Our study presents a first approach of its kind in topographically complex Cordilleran terrain and we show that GEE is particularly useful in large-scale land cover mapping and monitoring in mountainous ecosystems subject to rapid changes and conversions, with replicability and scalability to other areas with similar characteristics.

Keywords: land cover mapping and monitoring; Landsat 8 OLI; Google Earth Engine; highland landscapes; machine learning algorithms

1. Introduction

The interaction between conservation and the use of natural resources in and around protected areas characterized by multiple conflicting land uses arises from multiple interacting causes and requires continuous monitoring of land cover associated with land use changes and conversions [1]. In order to understand and quantify the resulting landscape dynamics and patterns, it is necessary to have large-scale and repeat information on the integrity of ecosystems for the implementation and evaluation of management policies of natural resources [2]. In addition, land use and land cover (LULC) mapping of mountain

ecosystems is very important for researchers and governments, given that these are global biodiversity hotspots that support the livelihoods of billions of people worldwide, and the sustainable management of these ecosystems is listed on the United Nations 2030 Agenda for sustainable development under the Sustainable Development Goal (SDG) 15.4 [3]. However, given the topographic, abiotic and biotic complexity and heterogeneity of mountain ecosystems globally, it remains challenging to quantify the dynamic nature of LULC [4], which leads to structural limitations in achieving SDG 15.4 and creating sustainable protection strategies [5]. As such, in-depth assessments of mountain ecosystem integrity and stability at regional scales continue to provide value towards achieving a global framework that can be applied towards monitoring progress towards SDG 15.4.

High-elevation landscapes in Peru and across the Andean Cordillera present complex topography, altitudinal and environmental gradients that contain a variety of ecosystems and vegetation communities; more broadly characterized by multiple conflicting land uses (protected areas, grazing, agriculture, mining), with implications for water provision and the maintenance of other ecosystem services needed for agriculture and water supply for both local and coastal populations [6]. Across these high-value, sensitive ecosystems, it is becoming ever more important that stakeholders and decision-makers have the ability to monitor LULCC in near-real time and take advantage of the computational advances in geospatial technologies [7].

Optical remote sensing data represent a feasible and economical option for the continuous monitoring of land cover changes, conversions and transitions [8,9] given its ability to collect data remotely, at consistent temporal scales. When working with mid-resolution satellite imagery, most of the pixels contain a mixture of the reflected radiance of different land cover types [10], and it is very difficult to find fully homogeneous land cover pixels in semi-arid regions, especially for rare and degraded vegetation types growing in areas that are hard to access [11]. In addition, a diverse topography reduces the accuracy of land cover classification in complex terrain because it produces variations in surface illumination between shaded areas and those receiving direct sunlight, as result the reflectance values of land cover vary greatly within classes [12]. One approach to overcome these issues is to use advanced classification methods based on machine learning (ML) that can process high-dimensional data to avoid overfitting. ML has been used to map vegetation in mountains because they have good potential for accurately classifying natural land cover types [13,14]. Supervised classification is one of the most essential and classical machine learning techniques in remote sensing with the goal of optimizing the performance criteria of prediction using data samples and learning rules from a training set, that then get generalized to the total data set even on high-dimensional, complex data [15].

Furthermore, the increasing accessibility of spatial imagery and the development of cloud-based spatial analysis computational frameworks such as GEE enable the development and implementation of spatial analysis, such as land cover classification even in contexts where processing such large volumes of data had been previously prohibitive [16]. GEE's programming interface allows users to define, create and run custom algorithms based on their analysis needs [17]. GEE is a platform for geospatial and multitemporal analysis on a planetary scale, which by incorporating the computational capabilities of Google, represents a free access tool for the analysis of a wide variety of socio-environmental problems with great usability especially in less-developed country contexts [18]. GEE's catalog includes the complete archive of Landsat satellite imagery, Sentinel-1, 2 and 3 imagery, Moderate Resolution Imaging Spectrometer (MODIS), National Oceanographic and Atmospheric Administration Advanced Very High-Resolution Radiometer (NOAA AVHRR) imagery, as well as climate, topographic and land cover data, among others [18,19]. The existence of these freely available resources is especially critical for researchers in developing countries where the computational infrastructure used to be and oftentimes continues to be a limiting factor. GEE has been shown to provide reliable results when investigating ecological restoration policies [2], the conservation status of mountain ecosystems [5],

high-alpine rangeland ecosystem change trajectories [20], plant functional types [13] and for mapping vegetation and land use types [21].

Since 1972, Landsat satellite missions have continuously provided a historical record of imagery that forms the basis for extracting land use and land cover data at global, regional and local scales [22]. The characteristics of Landsat images such as their spatial resolution of 30 m and their temporal resolution of 16 days [22], as well as their vast historical record have established the Landsat archive as a leader in mapping and monitoring of land cover and the ease of retrieval of this deep imagery data set via GEE is now revolutionizing how we approach mapping and monitoring LULC distribution [21,23].

The current study aimed to design and implement a repeatable and transferable GEE approach for mapping land cover classes in the Nor Yauyos Cochas Landscape Reserve (RPNYC acronym in Spanish), using remote sensing data derived from Landsat 8 OLI imagery and topographic-derived indices from ALOS PALSAR DEM. In this context, we demonstrate a first attempt at using remote sensing technologies that rely on optical sensors for mapping vegetation cover in Andean highland ecosystems employing cloud computing implemented in the Google Earth Engine (GEE) and machine learning (ML).

2. Materials and Methods

2.1. Study Area

RPNYC is located in the central Andes of Peru, in the regions of Lima and Junín (Figure 1). The RPNYC has an area of 221,268.48 ha and is surrounded by a buffer zone of 109,503.20 ha, a significant area to monitor in the absence of remotely sensed data. In total, 62.1% of the surface of the reserve is located in the Lima region, Yauyos Province, while 37.9% is located in the Junín region [24].
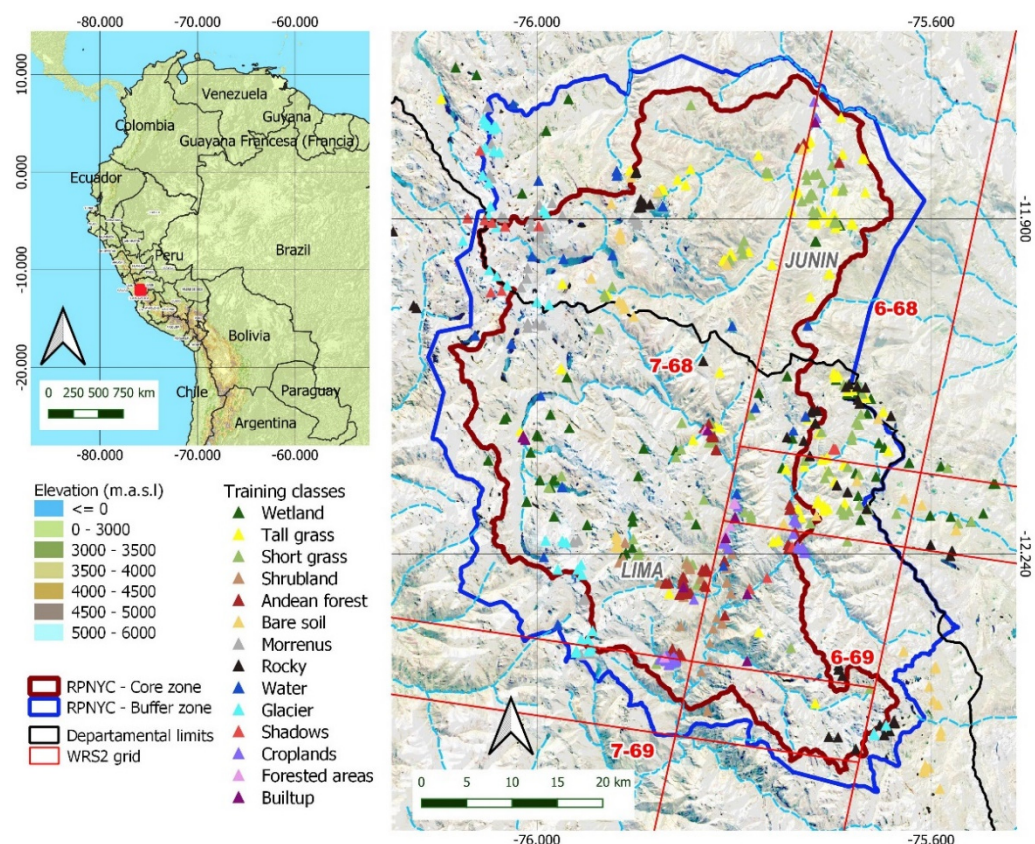

Figure 1. Location of RPNYC district in Peru (**left**) and map of RPNYC boundaries (**right**) with location of the training points by land cover type.

RPNYC was established by the National Natural Protected Areas Department (SER-NANP) in 2001 with the dual aim of preserving its ecological and cultural value, and promoting a harmonious and economically sustainable relationship with the rural communities surrounding it [25]. The reserve supports great biodiversity of flora and fauna, including environmentally sensitive grasslands, wetlands and high Andean forests, used by the local population for various agricultural activities [25], under regulation and supervision by SERNANP [26]. Water supply to the cities and agricultural valleys downstream represents one of the key ecosystem services supplied by RPNYC. Approximately 11 million people, including residents of the capital Lima, depend on the water originating in this reserve [26].

The RPNYC has an altitudinal gradient that ranges from 2630 to 5660 m above sea level (m.a.s.l). The physiography is dominated by landscapes of high mountains whose slopes present a very rugged topography, strongly divided by channels and deep valley bottoms. RPNYC covers six distinct biomes, with moisture regimes ranging from arid to rainforest, home to ten types of vegetation covers and a total of 330 plant species, such as trees, shrubs, grasses, succulents and epiphytes. The fauna consists of 75 species of birds, 15 species of mammals, 4 reptile species, 1 amphibian and 2 native fish species [24].

Extensive breeding and grazing of domestic livestock such as sheep, camelids, cattle and horses constitutes the main agricultural land use in the reserve. In the surrounding buffer communal lands, the sustainable management of vicuñas (*Vicugna vicugna*) is carried out, a wild species of South American camelid, of important cultural value for local populations, cataloged by the International Union for Conservation of Nature (IUCN) as Least Concern. Subsistence agricultural activities are carried out in the valley bottoms and gentler mountain slopes by means of terraced crops, similar to densely populated urban centers downstream [24].

The climate varies with altitude, generating multiple microclimates, with periods of rain between November and March (minimum of 2.05 mm / day in November and maximum of 4.48 mm/day in February), a transition period from April to October (1.68 mm/day in April and 1.90 mm/day in October) and a dry season between May and August (0.46 mm/day in August and 0.23 mm/day in July). The average temperature ranges from 3 to 7.5 °C, with lowest temperatures between May and August, and frost events between July and August [27].

Data from vegetation censuses and in-country assessments identify five native and endemic vegetation types of high biodiversity value: Bofedales, tall grass (tussock grass-dominated cover), Puna mat grass, bushes and relict forests [24,28]. High-altitude areas lacking vegetation cover were classified as: bare soil, moraines, rocky areas, lagoons and glaciers and finally, areas of agricultural and urban land use were divided into reforested, agricultural and urban areas.

2.2. Methodological Framework

The methodological framework employed in this study is presented in Figure 2, and described in the following six methods subsections: Landsat image preprocessing (Section 2.3), training and validation sample collection (Section 2.4), classification feature input (Section 2.5), separability analysis (Section 2.6), classifiers (Section 2.7), accuracy assessment (Section 2.8) and calculation of land cover extents (Section 2.9).

2.3. Landsat Image Preprocessing

Due to the number of spectral bands, Landsat 8 Operational Land imager (OLI) surface reflectance images were used as the prime source of data for the study. A search for all USGS Surface Reflectance (SR) images from the Landsat 8 OLI/TIRS sensor covering RPNYC reserve (located within Worldwide Reference System 2, path 6, row 68, path 6, row 69, path 7, row 68, path 7, row 69) from January to June 2018 according to the field data collection dates was performed within the GEE's data pool. Pre-processing of Landsat 8 OLI/TIRS images consisted of application of quality bands to mask cloud and cloud

shadows contamination, using a CFmask function [29] and verified with a final visual test. The composite imagery used in the ML classifications was assembled from 31 scenes (see Supplemental Materials, Table S1), through a median reducer based on GEE tools, to use the more frequent pixel value in the stack.
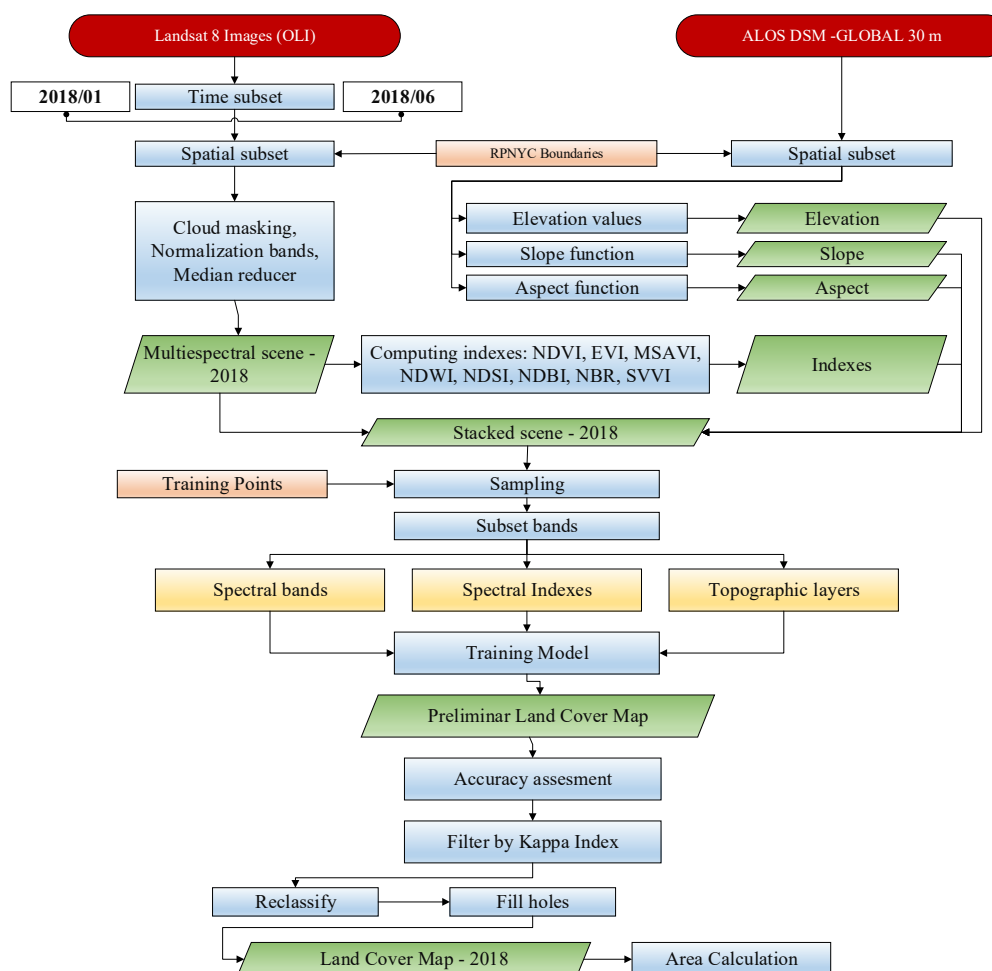

Figure 2. Representation of the methodological framework used in this study.

2.4. Training and Validation Sample Collection

Ground reference data on vegetation composition within accessible regions of the RPNYC were collected during and after the rainy season 2018 (Jan–June). Relatively homogeneous areas of 90×90 m^2 were chosen as survey plots, on the ground, based on accessibility and visual inspection of patches connected by ranger trails to ensure sampling within the fourteen land cover classes shown in Figure 1.

At every plot, the dominant vegetation land cover class (Table 1) was determined based on dominant species cover through a rapid assessment process developed by the Rangeland Ecology ad Utilization Laboratory at the National Agrarian University La Molina, Peru (LEUP-UNALM acronym in Spanish). All the survey plot locations were recorded with a global navigation satellite system receiver (Garmin 62S) and subsequently checked for accuracy.

In order to complete training data sets for land cover classes, we used other reference sources such as the 2015 national vegetation map [30] (rocky areas, glaciers and bare soil classes), Google Earth imagery (pan-sharpened QuickBird, GeoEye and WorldView-2 imagery) and visual interpretation.

Training samples were generated from a total of 3601 pixels available (75% from field data, and 25% from Google Earth imagery and the 2015 national vegetation map) and

the samples were split into a validation data set, defined by a stratified random sampling method for each cover class, to at least 30% of the total points. This procedure ensured independence between the training and validation data [31].

Table 1. Vegetation land cover classes identified in RPNYC [24,25].

Land Cover	Description
Wetlands	Areas with excess of water from high rates of orographic precipitation, dominated by cushion plants, distributed in low-slope areas and above 3800 m.a.s.l.
Tall grass	Areas composed chiefly of perennial grasses such as Festuca, Poa, Stipa and Calamagrostis species, distributed along the slopes and tops of hilly and mountains landscapes, between an altitude range of 3900 to 4900 m.a.s.l.
Short grass	Areas dominated by dwarf herbaceous forbs and cushion plants growing in areas of moderate water content. Erect shrubs, mosses and lichens are of minor importance. Spatial distribution similar to tall grass land cover.
Shrublands	Communities conformed by species of *Baccharis* and *Parastrephia*, distributed between an altitude range of 2750 to 3800 m.a.s.l.
Andean forest	Areas generally relegated to rocky slopes or ravines dominated by shrubs of the genus *Polylepis* associated with the genera: *Buddleja, Clethra, Gynoxys, Podocarpus* or *Prumnopitys*, distributed between a wide altitude range of 600 to 4100 m.a.s.l.

2.5. Classification Feature Input

Due to the extreme elevation range and steep slopes and its interaction with the vegetation cover in the RPNYC reserve, a reflectance normalization process was applied to Landsat 8 images by dividing each reflectance band by the reflectance sum of all bands [21,32], using band variables from Band 2 to Band 7.

Eight spectral indices were derived from the reflectance normalized Landsat spectral bands as shown in Table 2.

Table 2. Extracted spectral indices from Landsat 8 imagery.

Bands	Wavelength (μm)
Normalized Difference Vegetation Index (NDVI)	$\text{NDVI} = \frac{(\text{NIR} - \text{RED})}{(\text{NIR} + \text{RED})}$
Enhanced Vegetation Index (EVI)	$\text{EVI} = \text{G} \times \frac{\text{NIR} - \text{RED}}{(\text{NIR} + \text{C1} \times \text{RED} + \text{C2} \times \text{BLUE} + \text{L})}$
Modified Soil Adjusted Vegetation Index (MSAVI)	$\text{MSAVI} = \frac{2 \times \text{NIR} + 1 - \sqrt{(2 \times \text{NIR})^2 - 8(\text{NIR} - \text{RED})}}{2}$
Normalized Difference Water Index (NDWI)	$\text{NDWI} = \frac{(\text{GREEN} - \text{RED})}{(\text{GREEN} + \text{RED})}$
Normalized Difference Snow Index (NDSI)	$\text{NDSI} = \frac{(\text{GREEN} - \text{SWIR1})}{(\text{GREEN} + \text{SWIR1})}$
Normalized Burn Ratio (NBR)	$\text{NBR} = \frac{(\text{NIR} - \text{SWIR1})}{(\text{NIR} + \text{SWIR1})}$
Normalized Difference Built-Up Index (NDBI)	$\text{NDBI} = \frac{(\text{SWIR1} - \text{NIR})}{(\text{SWIR1} + \text{NIR})}$
Spectral Variability Vegetation Index (SVVI)	$\text{SVVI} = \text{SD}(\text{BLUE, GREEN, RED, NIR, SWIR1, SWIR2}) - \text{SD}(\text{NIR, SWIR1, SWIR2})$

NDVI is one of the most widely used vegetation indices thanks to its ability to discriminate vegetation/non-vegetation covers [33], while EVI can be useful in areas of high leaf area index (LAI) where NDVI may saturate [34]. MSAVI minimizes the effect of bare soil in the computation of the Soil-Adjusted Vegetation Index (SAVI) [35], whereas NWDI is useful for discriminating water from land due to its sensitivity to open water. [36] NDSI is related to the presence of snow in a pixel while ignoring cloud cover [37], NBR is designed to highlight burned areas and estimate fire severity [38], while NDBI highlights urban areas where there is typically a higher reflectance in the shortwave-infrared part of the

spectrum [39]. Finally, SVVI is used to discriminate forest, secondary forest, savanna and agriculture classes that would otherwise be difficult to differentiate spectrally [40].

Given the importance of topographic variables in classifying vegetation cover classes, we utilized the ALOS PALSAR DEM data (30 m) [41] to derive slope (degrees, ranging from 0 to 90°), aspect (0 to 360°), transformed by taking the trigonometric sine values of aspect to avoid circular data values [42]. Sine values of aspect represent the degree of east-facing slopes, as the values range from 1 (i.e., east-facing) to -1 (west-facing).

From the Landsat data and ALOS PALSAR topographic derivatives, we defined six sets of stacked imagery for predictor inputs, in the following order: spectral bands; spectral indices; spectral bands + spectral indices; spectral bands + topographic-derived indices; spectral indexes + topographic-derived indices; and spectral bands + spectral indices + topographic-derived indices that were then used to implement classifiers on.

2.6. Separability Analysis

The spectral separability analysis was performed for all classification data set inputs, and the range of values generated by this method is 0.0 to 2.0, where a value of 0.0 means that the pair of classes have a perfect resemblance and the value of 2.0 indicates that the class pairs are completely different. Ideally, the inter-class separability value in the classification scheme used is more than 1.9. It is calculated using the followings formulas [43].

$$J_{xy} = 2\left(1 - e^{-B}\right) \tag{1}$$

$$B = \frac{1}{8}(x-y)^t\left(\frac{\sum x + \sum y}{2}\right)^{-1}(x-y) + \frac{1}{2}\ln\left(\frac{\left|\frac{\sum x + \sum y}{2}\right|}{|\sum x|^{\frac{1}{2}}|\sum y|^{\frac{1}{2}}}\right) \tag{2}$$

where x is first spectral signature vector; y is second spectral signature vector; Σx is covariance matrix of sample x; and Σy is covariance matrix of sample y. The Jeffries–Matusita distance is asymptotic to 2 when signatures are completely different, and tends to 0 when signatures are identical.

2.7. Classifiers

A variety of machine learning image classification methods have been used to map vegetation type on large mountain areas, such as, support vector machines (SVM) [44], decision tree (i.e., CART), naïve Bayes and random forest (RF) classifiers [21,45].

SVM is used for classification to determine a hyperplane that can divide training data into a predetermined number of categories, maximizing the distance from the data points of classes to an optimal separation vector of a hyperplane created from the variables [46].

Decision tree classifiers use variables subsequently to find split (partitioning) points and try to split subsets further with other variables, etc. Groups of pixels can be further divided based on tree growing and pruning parameters, until optimal classification is achieved [14].

Random Forest [47] classifiers construct multiple decision trees that are sampled independently during training, typically improving classification by vote results compared to a single decision tree model.

To implement classification, we used GEE's platform—a web-based IDE—through JavaScript Code Editor [18], which offers a variety of geospatial analysis tools and visualization interface. In these environments, we use six supervised machine learning classification algorithms available [48] shown in Table 3, and applied iteratively to all six data set stacks as indicated in Section 2.4.

For random forest and gradient tree boosting, 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 36 models were built between the combination of classifiers and input data.

Table 3. Supervised machine learning for GEE in 2021.

Group	Algorithm
Logic-based algorithms	CART (Classification and Regression Tree) [49]
	Random Forest [47]
	Gradient Tree Boosting [50]
Statistical learning algorithms	Minimum Distance [51]
	Naive Bayes classifiers [52]
Support vector machines	Voting SVM (Support Vector Machines) [53]

Note: Examples in bold represent machine learning algorithms used in GEE.

2.8. Accuracy Assessment

A confusion matrix and the accuracy assessment of each resulting classification output for the six computed algorithms on the six sets were computed using the 'caret' [54] and 'diffeR' [55] packages in R. We calculated the overall accuracy (OA), Kappa coefficient (K), producer accuracy (PA) and user accuracy (UA), complemented with a quantity disagreement (Q) and allocation disagreement (A) assessment [56]. We used these metrics combined to assess the quality of the resulting land cover maps. The OA determines the overall efficiency of an algorithm and is measured by dividing the total number of correctly classified samples by the total number of testing samples.

The K indicates the degree of agreement between the ground truth data and the predicted values, while the PA measures how well a pixel has been classified and includes the error of omission (the proportion of observed features on the ground that are erroneously excluded from a class). The UA measures the reliability of the map, informing how well the map represents what is on the ground and it includes the error of commission which refers to pixels erroneously included in a class [57].

Q is the amount of difference between the reference data and a comparison map that is due to the less than perfect match in the proportions of the categories, and A is the amount of difference between the reference data and a comparison map that is due to the less than optimal match in the spatial allocation of the categories, given the proportions of the categories in the reference and comparison maps (e.g., classifying built-up in a location where built-up development is not observed) [56].

Finally, we used variable importance which considers 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 an additional means of assessing how each predictor variable enabled accuracy improvements in the optimized land cover classification model, in terms of a normalized percentage contribution.

2.9. Calculation of Land Cover Areas/Extents

During the final step, we created a mask of areas that were excluded from the analysis based on CFmask filter [29], which glacier pixels were misclassified as clouds, reclassifying them as glacier pixels. This process was performed using the "raster" package [58] in R 3.6 software. Then, we tabulated pixels by land cover class and estimated the cover area adjusted to the total area for the core and buffer zone of the RPNYC. The area results were compared with a published map from the environment minister in 2015 expressed in ha for all the RPNYC.

3. Results

3.1. Separability Analysis

Considering the different input data sets, there are a small improvement on JM distance when topographic indices are added for all the land cover classes (Figure 3), showing the better separability when all the variables are included.
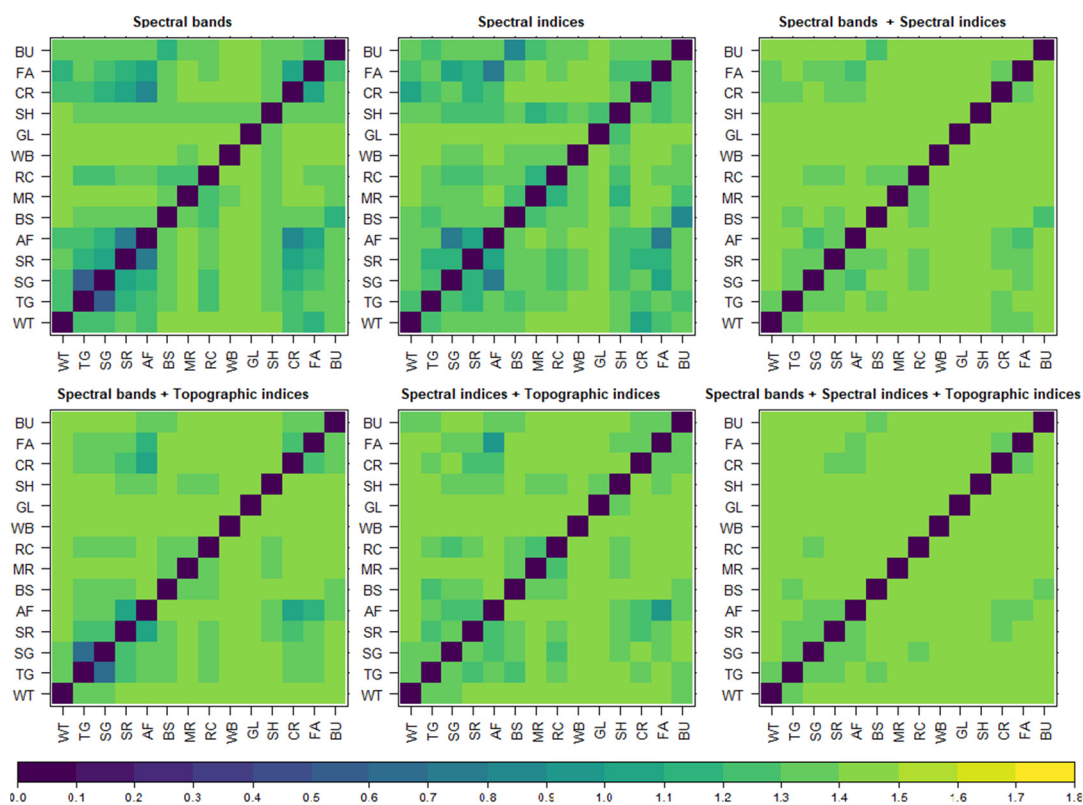
Figure 3. Jefferies–Matusita distance separability matrix for training classes. Notes: WT—Wetlands; TG—Tall grass; SG—Short grass; SR—Shrublands; AF—Andean forest; BS—Bare soil; MR—Moraine; RC—Rocky; WB—Water bodies; GL—Glacier; SH—Shadows, FA—Forested reas; CR—Croplands; BU—Built-up.

Due to the landscape heterogeneity, we consider that the JM distance has acceptable levels for the separability of the training areas, the next step is to conduct the classification process. Overall, class separability is adequate and would provide a fairly accurate classification.

3.2. Accuracy Assessment Results

Table 4 provides the accuracy estimates of the iterations between the ML classifiers available in GEE and the six sets of stacked imagery for predictor inputs of the study area. The results showed that for all the sets of iterations, the random forest classifier yielded the better performance (0.81), followed by gradient tree boosting; additionally, the accuracy increases dramatically for all the models when topographic-derived indices are added as predictors. There are no marked observed differences between predictors based on spectral bands and spectral indices separately. The lowest performance were the naïve Bayes and minimum distance classifiers.

Table 5 depicts the confusion matrix obtained from the analysis of the validation data set for the best model identified (RF: spectral bands + topographic indices). Overall, the classification between vegetation land cover classes and non-vegetation is similar. In particular, the comparison between wetlands, short grasses and tall grasses presents more confusion errors. In contrast, the classification accuracies of water bodies and built-up areas are higher (both user's and producer's accuracies) than the other classes.

The integration of topographic-derived indices resulted in improvements in the Kappa coefficient for all models (0.08–0.61) and was the most important variable in predicting land cover types (elevation and slope), as shown in Figure 4, for the best three classification models (random forest).

Table 4. Accuracy assessment of land cover classification using machine learning algorithms from Google Earth Engine (GEE).

| Combination | Metric | Supervised Classification Algorithm | | | | | |
		CART	Random Forest	Support Vector Machine	Minimum Distance	Gradient Tree Boosting	Naïve Bayes
Spectral bands	K	0.57	0.66	0.58	0.38	0.65	0.43
	Q (%)	6.31	5.99	9.79	17.46	5.93	14.56
	A (%)	32.80	24.36	27.84	38.92	25.13	36.92
Spectral indices	K	0.55	0.64	0.51	0.19	0.63	0.07
	Q (%)	5.80	6.19	21.71	33.18	5.93	84.34
	A (%)	34.60	26.03	21.97	42.14	27.38	3.25
Spectral bands + Spectral indices	K	0.57	0.66	0.59	0.38	0.65	0.37
	Q (%)	6.38	6.19	9.79	17.91	5.86	22.23
	A (%)	32.60	24.16	27.32	38.47	25.64	35.70
Spectral bands + Topographic indices	K	0.74	**0.81**	0.66	0.40	0.80	0.50
	Q (%)	4.64	**4.06**	7.41	25.71	3.67	20.43
	A (%)	18.69	**12.89**	23.52	29.25	13.92	25.52
Spectral indices + Topographic indices	K	0.73	0.79	0.60	0.30	**0.81**	0.21
	Q (%)	4.90	4.90	11.28	27.71	**4.19**	48.20
	A (%)	19.07	13.79	24.87	36.79	**12.89**	26.22
Spectral bands + Spectral indices + Topographic indices	K	0.72	0.78	0.65	0.40	**0.81**	0.48
	Q (%)	3.87	5.41	8.18	25.39	**3.80**	19.14
	A (%)	21.52	14.24	23.45	29.51	**13.60**	27.90

Note: K: Kappa coefficient, Q: Quantity disagreement, A: Allocation disagreement.

Table 5. Confusion matrix of best classification.

Class	WT	TG	SG	SR	AF	BS	MR	RC	WB	GL	SH	CR	FA	BU	Total	UA	PA
WT	140	3	4	1	0	0	0	0	0	0	0	0	0	1	149	93.96	86.96
TG	8	191	31	0	8	2	0	1	0	0	3	0	0	0	244	78.28	81.28
SG	13	28	200	4	1	4	0	4	0	0	0	0	0	0	254	78.74	78.74
SR	0	1	0	51	2	0	0	1	0	0	1	0	0	0	56	91.07	85.00
AF	0	6	7	1	108	0	0	0	0	0	0	19	6	0	147	73.47	75.52
BS	0	2	1	1	1	108	0	5	0	0	0	0	0	0	118	91.53	93.91
MR	0	0	0	0	0	0	64	4	0	0	0	0	0	0	68	94.12	64.00
RC	0	2	11	2	0	1	4	101	0	0	0	0	0	0	121	83.47	85.59
WB	0	0	0	0	0	0	0	0	58	0	0	0	0	0	58	100.00	100.00
GL	0	0	0	0	0	0	31	2	0	114	6	0	0	0	153	74.51	100.00
SH	0	0	0	0	1	0	1	0	0	0	30	0	0	0	32	93.75	75.00
CR	0	2	0	0	19	0	0	0	0	0	0	81	2	0	104	77.88	81.00
FA	0	0	0	0	3	0	0	0	0	0	0	0	10	0	13	76.92	50.00
BU	0	0	0	0	0	0	0	0	0	0	0	0	2	33	35	94.29	97.06
Total	161	235	254	60	143	115	100	118	58	114	40	100	20	34		OA: 83.05%	
											K: 0.81		Q: 4.06		A: 12.89		

Notes: Diagonal elements in this matrix indicate the number of samples for which predictions match reference data. Off-diagonal elements in rows and columns, respectively, capture errors of commission and omission. UA: user's accuracy; PA: producer's accuracy; OA: overall accuracy; K: Kappa coefficient; Q: Quantity disagreement; A: Allocation disagreement. WT—Wetlands; TG—Tall grass; SG—Short grass; SR—Shrublands; AF—Andean forest; BS—Bare soil; MR—Moraine; RC—Rocky; WB—Water bodies; GL—Glacier; SH—Shadows, FA—Forested areas; CR—Croplands; BU—Built-up.

3.3. Classification Result

Based on the previous results, we selected the best classification map and the spatial distribution of each land cover class is shown in Figure 5. We can recognize the spatial distribution of different land covers related to elevation, slope and aspect gradients, agriculture areas are distributed in the north and south side of the RPNYC in relative lowland

areas, grazing lands dominates the high elevation around glaciers and water bodies in the basins head. Pixels classified as shadows were reclassified as NoData and not displayed on the final map.
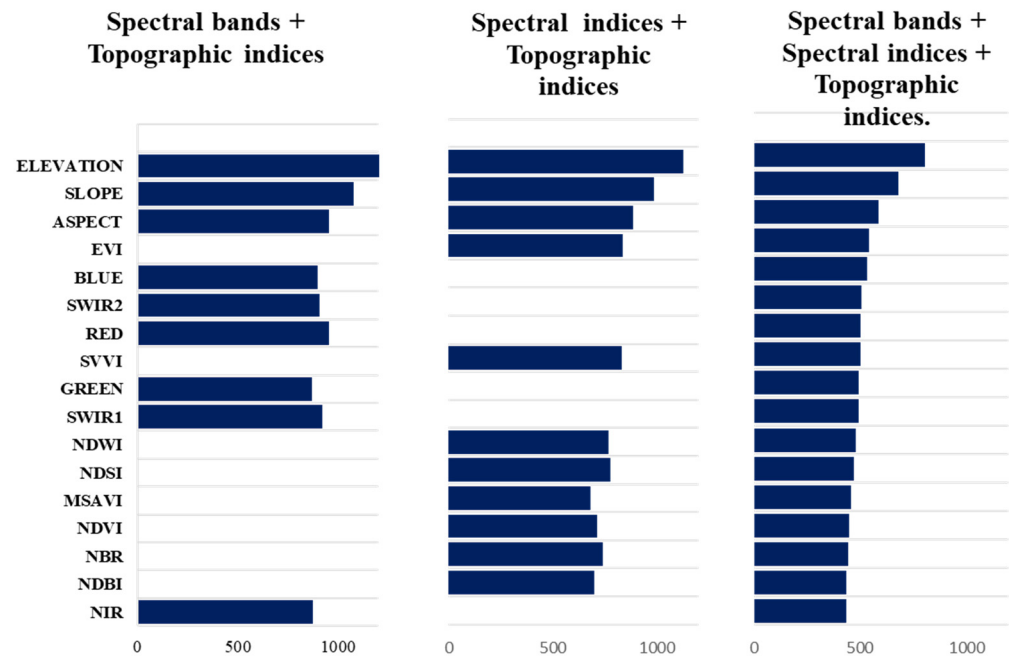

Figure 4. Variable importance analyses for random forest classifiers.

Regarding the distribution of the land covers for the two best maps for 2018 (Table 6), the reserve is dominated by a combination of tall grasses and short grasses along the core and buffer zone (56.26% and 55.06%, respectively), followed by a combination of no vegetation covers such as bare soil, moraine and rocky (28.88% and 31.76 %, respectively). Wetlands, glaciers and water bodies extend more in the buffer zone than in the core area, in contrast to Andean forests and shrublands which are more characteristic in the core zone (6.38%) than in the buffer zone (2.80%). With respect to forested and agricultural areas (0.93%), both are concentrated around built-up areas and along rivers into the core and buffer zones and represent 0.14% of the total surface of RPNYC.

Table 6. Distribution of the surface area of each land cover class for the core and buffer zone of RPNYC in percentages.

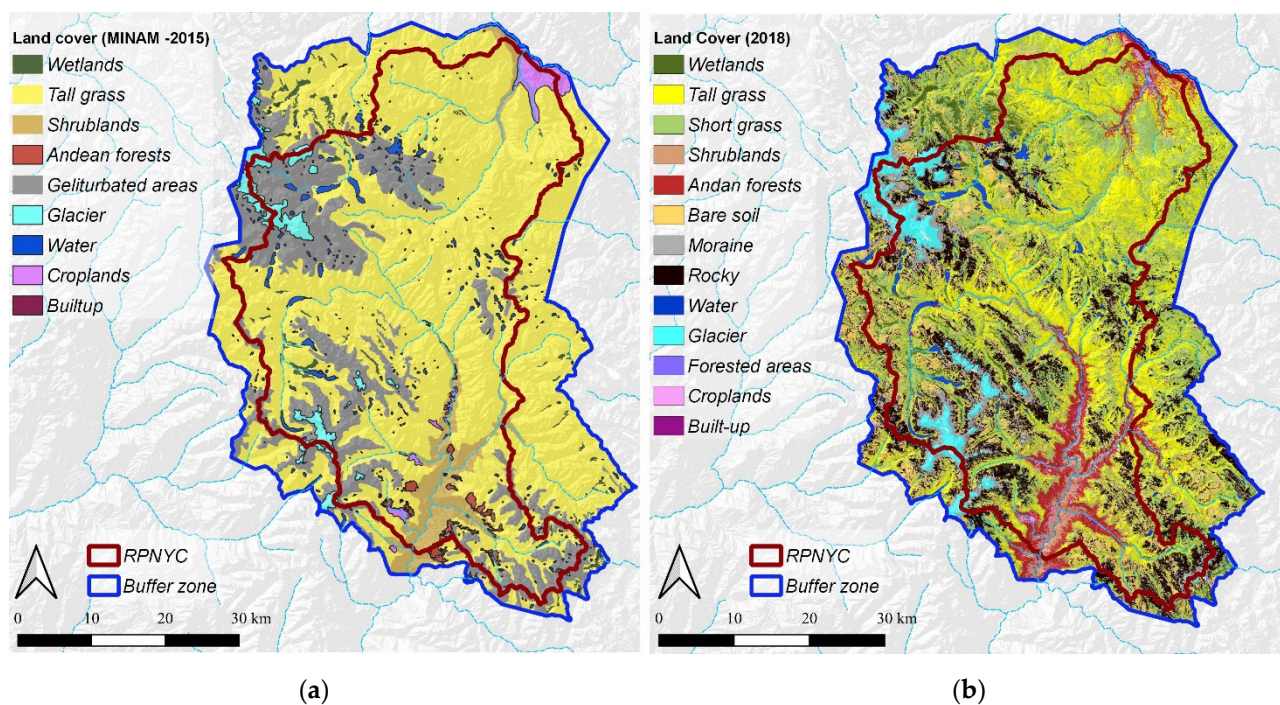| Land Cover Class	RF. (Spectral Bands + TOPOGRAPHIC Indices)		MINAM—2015	
Zone	Core (%)	Buffer (%)	Core (%)	Buffer (%)
Wetlands	3.95	6.04	0.65	2.30
Tall grass	24.26	20.96	66.60	72.57
Short grass	28.35	31.47		
Shrublands	2.84	1.48	5.76	2.51
Andean forest	6.32	2.99	0.79	0.39
Bare soil	9.11	12.15	21.07	17.15
Moraine	4.76	4.81		
Rocky	13.81	14.35		
Water bodies	1.79	0.94	1.90	1.25
Glacier	3.26	3.87	1.90	2.75
Agriculture	0.94	0.40	1.32	1.07
Forested Areas	0.08	0.04		
Built-up	0.53	0.51		0.01
Total	100.00	100.00	100.00	100.00

(**a**)

(**b**)

Figure 5. Spatial representation of the classification output with the highest overall classification accuracy: random forest + spectral bands + spectral indices + topographic layers (**a**) and the National Land cover map assessment 2015 (**b**).

4. Discussion

The use of cloud computing platforms for geospatial big data analysis, such as Google Earth Engine, provide multitemporal data sets of satellite imagery and preprocessed spatial models and allow us to build specific workflows, in order to obtain detailed maps in short timeframes relative to traditional computing approaches, such as [59] who used >650,000 Landsat scenes and >1million central processing unit (CPU) hours to produce annual maps of tree cover gain and loss, performing these computations in just several days.

Our results show that is feasible to use the open-access GEE platform to implement land cover mapping in complex Andean ecosystems, based on composite multitemporal mosaics and field data information using machine learning classifiers, without having to purchase or download data and software [2,5,13,20,21].

These characteristics confer significant advantages that reduce the learning gap to process spatial data for technicians and resource managers for the task of spatially monitoring protected areas in an effective way. Another advantage is the feasibility to share the GEE code and apply them to another areas, with minimum modifications, considering a growing community of user's support, and discussion forums, with the only critical requirement being a stable internet connection.

For the more general purpose of updating land cover maps using visual interpretation of satellite imagery or semi-automatic updating through GIS analysis, it is feasible if the study area is not extensive [31]; however, natural protected areas are generally large areas. Then, GEE is particularly useful for these purposes.

In this study, we show that the inclusion of topographic indices improved the discrimination of the various cover classes by minimizing the terrain shadow effects and greatly decreasing the misclassification between water bodies and shadows. Thus, the overall classification accuracy of the land cover spatial models was improved when topographic indices were combined with spectral bands or spectral indices separately. However, no improvement was observed in the classification accuracy when spectral bands and indices are combined together, except the effect of topographic indices that can improve classification accuracy in general [2,60].

The best classifiers tested in this study were able to consistently minimize the effects of clouds and the derived no data pixels, and the accuracy metrics for the select model outperformed coincides with other studies that used machine learning classifiers, especially RF in distinguishing several vegetation types for heterogeneous and large regions such as RPNYC [21,61]. RF can cope with high-dimensional problems very easily thanks to the pruning strategy and is very beneficial by alleviating the often-reported overfitting problem of simple decision trees [62].

As shown in Table 5, confusion errors occurred among all classes, except water bodies. Notably, the highest confusion was found between tall and short grasses. This is particularly severe since both have very similar ecological and visual characteristics and are dominated by ecologically similar vegetation types of grasses [63]. In the field, we noted that the wetland class was found adjacent to tall grass and short grass areas, forming transition zones, which adds errors to the classification. Confusion was also common among moraines and glaciers, which is attributable to the proximity of both classes and the loss of glacial coverage [64].

Another consideration when interpreting the classification accuracies is the availability of the training samples for the supervised classification. As shown in Table 5, for example, wetlands have a larger number of training polygons compared to the forested areas class or croplands. This is because the RPNYC, since its foundation in 2001, has not allowed the expansion of the agricultural frontier and reforestation with exotic species [28]. On the other hand, the shadow class is usually found in physically smaller and inaccessible area relative to other classes, for example, in canyons.

Given that persistent cloud cover is an issue for large, high-altitude areas, we show that the issue can effectively be minimized using the multi-temporal composite and cloud masking approach such as CFmask function [29].

We find differences between the land cover selected map and an official land cover map [30], which was built based on 2011 Landsat 4–5 TM imagery. The principal differences are related to the information classes and proportion of areas, for example tall and short grasses are considered as one class, and the area of these two classes in our selected maps represents 75% of the official map. Similarly, we detected an area five times larger for wetlands. However, we found some similarities with respect to the water bodies and glacier extents.

The main limitations of this work mostly pertain to uncertainties between the Landsat-derived land cover maps and the reference data. The reference data samples that were utilized during the classifier training and testing phases were limited in quantity due to the remote accessibility. Another major challenge we encountered was the mismatch in time between the available cloud-free Landsat images and the reference data. The MINAM reference map was produced three years earlier than the study period (2018) and used 657 verification points and 415 control points for the entire country, so the comparison with the maps generated from our work is tenuous. The field survey and the high spatial resolution reference images retrieved from Google Earth (2017) are minimally a year apart from the study period. This created difficulty in analyzing the classification products in conjunction with the available reference data set.

5. Conclusions

Our results demonstrate that the use of cloud-based computing and machine learning feasibly allows us to build an open-access, accurate and relatively gap-free approach for land cover mapping in a heterogeneous and large Andean ecosystem using freely available Landsat imagery. This approach can be utilized to advance natural resource planning and management in less-developed country contexts and is also useful in informing long-term monitoring of vegetation in remote protected areas by establishing key land cover classes for any reference year. Furthermore, establishing the extent and distribution of land cover classes along with their spectral and topographic characteristics may allow us to extrapolate these classifications to non-protected areas in similar ecosystems. Our

work highlights the importance of the inclusion of topographic-derived indices, in the differentiation of complex land covers distributed in diverse and highly heterogeneous topography to improve classification accuracies unlike the inclusion of spectral indices. Finally, we established an approach for generating a reliable land cover map in a protected reserve most representative of Andean ecosystems, thus establishing a baseline for future change analyses and environmental planning. This work has implications for achieving SDG 15.4 focused on monitoring and assessing changes in mountain ecosystems and for the long-term monitoring and assessment of environmental changes and conservation efforts in mountains.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs14071562/s1, Table S1: List of utilized Landsat 8 images, Code S1: Google Earth engine code.

References

1. De Jong, L.; De Bruin, S.; Knoop, J.; van Vliet, J. Understanding Land-Use Change Conflict: A Systematic Review of Case Studies. *J. Land Use Sci.* **2021**, *16*, 223–239. [CrossRef]
2. Yan, Y.; Zhuang, Q.; Zan, C.; Ren, J.; Yang, L.; Wen, Y.; Zeng, S.; Zhang, Q.; Kong, L. Using the Google Earth Engine to Rapidly Monitor Impacts of Geohazards on Ecological Quality in Highly Susceptible Areas. *Ecol. Indic.* **2021**, *132*, 108258. [CrossRef]
3. UNCCD. SDG 15: Life on Land- Facts and Figures, Targets, Why It Matters. Available online: https://knowledge.unccd.int/publications/sdg-15-life-land-facts-and-figures-targets-why-it-matters (accessed on 29 January 2022).
4. Lambin, E.F.; Geist, H.J.; Lepers, E. Dynamics of Land-Use and Land-Cover Change in Tropical Regions. *Annu. Rev. Environ. Resour.* **2003**, *28*, 205–241. [CrossRef]
5. Bian, J.; Li, A.; Lei, G.; Zhang, Z.; Nan, X. Global High-Resolution Mountain Green Cover Index Mapping Based on Landsat Images and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 63–76. [CrossRef]
6. Flores, E.R. Climate Change: High Andean Rangelands and Food Security. *Rev. Glaciares Y Ecosistemas Montaña* **2016**, *1*, 73–80.
7. Herrick, J.E.; Beh, A.; Barrios, E.; Bouvier, I.; Coetzee, M.; Dent, D.; Elias, E.; Hengl, T.; Karl, J.W.; Liniger, H.; et al. The Land-potential Knowledge System (Landpks): Mobile Apps and Collaboration for Optimizing Climate Change Investments. *Ecosyst. Health Sustain.* **2016**, *2*, e01209. [CrossRef]
8. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global Land Cover Mapping at 30 m Resolution: A POK-Based Operational Approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [CrossRef]
9. Joshi, N.; Baumann, M.; Ehammer, A.; Fensholt, R.; Grogan, K.; Hostert, P.; Jepsen, M.R.; Kuemmerle, T.; Meyfroidt, P.; Mitchard, E.T.A.; et al. A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sens.* **2016**, *8*, 70. [CrossRef]
10. Townshend, J.R.G.; Huang, C.; Kalluri, S.N.V.; Defries, R.S.; Liang, S.; Yang, K. Beware of Per-Pixel Characterization of Land Cover. *Int. J. Remote Sens.* **2000**, *21*, 839–843. [CrossRef]
11. Roberts, D.A.; Gardner, M.; Church, R.; Ustin, S.; Scheer, G.; Green, R.O. Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models. *Remote Sens. Environ.* **1998**, *65*, 267–279. [CrossRef]
12. Pimple, U.; Sitthi, A.; Simonetti, D.; Pungkul, S.; Leadprathom, K.; Chidthaisong, A. Topographic Correction of Landsat TM-5 and Landsat OLI-8 Imagery to Improve the Performance of Forest Classification in the Mountainous Terrain of Northeast Thailand. *Sustainability* **2017**, *9*, 258. [CrossRef]
13. 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]

14. Sluiter, R.; Pebesma, E.J. Comparing Techniques for Vegetation Classification Using Multi- and Hyperspectral Images and Ancillary Environmental Data. *Int. J. Remote Sens.* **2010**, *31*, 6143–6161. [CrossRef]
15. Alpaydın, E. *Introduction to Machine Learning*, 3rd ed.; MIT Press: London, UK, 2014; ISBN 978-0-262-02818-9.
16. Mananze, S.; Pôças, I.; Cunha, M. Mapping and Assessing the Dynamics of Shifting Agricultural Landscapes Using Google Earth Engine Cloud Computing, a Case Study in Mozambique. *Remote Sens.* **2020**, *12*, 1279. [CrossRef]
17. Bauddh, K.; Kumar, S.; Singh, R.P.; Korstad, J. *Ecological and Practical Applications for Sustainable Agriculture*; Springer: Singapore, 2020; ISBN 9789811533723.
18. 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]
19. Kumar, L.; Mutanga, O. Google Earth Engine Applications; Remote Sensing. *Remote. Sens.* **2019**, *11*, 591.
20. Zhou, B.; Okin, G.S.; Zhang, J. Leveraging Google Earth Engine (GEE) and Machine Learning Algorithms to Incorporate in Situ Measurement from Different Times for Rangelands Monitoring. *Remote Sens. Environ.* **2020**, *236*, 111521. [CrossRef]
21. Tsai, Y.H.; Stow, D.; Chen, H.L.; Lewison, R.; An, L.; Shi, L. Mapping Vegetation and Land Use Types in Fanjingshan National Nature Reserve Using Google Earth Engine. *Remote Sens.* **2018**, *10*, 927. [CrossRef]
22. Schroeder, T.A.; Cohen, W.B.; Song, C.; Canty, M.J.; Yang, Z. Radiometric Correction of Multi-Temporal Landsat Data for Characterization of Early Successional Forest Patterns in Western Oregon. *Remote Sens. Environ.* **2006**, *103*, 16–26. [CrossRef]
23. Xie, Y.; Sha, Z.; Yu, M. Remote Sensing Imagery in Vegetation Mapping: A Review. *J. Plant Ecol.* **2008**, *1*, 9–23. [CrossRef]
24. MINAM. *Inventario y Evaluación Del Patrimonio Natural En La Reserva Paisajística Nor Yauyos Cochas*; MINAM: Lima, Peru, 2011; p. 264.
25. INRENA. *Reserva Paisajística Nor Yauyos Cochas—Plan Maestro 2006—2011*; INRENA: Lima, Peru, 2006; p. 263.
26. Dourojeanni, P.; Fernandez-Baca, E.; Giada, S.; Leslie, J.; Podvin, K.; Zapata, F. Vulnerability assessments for ecosystem-based adaptation: Lessons from the Nor Yauyos Cochas Landscape Reserve in Peru. In *Climate Change Adaptation Strategies–An Upstream-downstream Perspective*; Springer: Cham, Switzerland, 2016; pp. 141–160. [CrossRef]
27. FDA. Estudio de La Vulnerabilidad e Impacto Del Cambio Climático Sobre La Reserva Paisajística Nor Yauyos Cochas. In *Escenarios Climáticos Futuros y Distribución Futura de Especies*; EbA Montaña: Lima, Peru, 2013.
28. SERNANP. *Reserva Paisajística Nor Yauyos-Cocha—Plan Maestro 2016–2020*; MINAM: Lima, Peru, 2016; p. 107.
29. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Joseph Hughes, M.; Laue, B. Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]
30. Ministerio del Ambiente (MINAM). Mapa Nacional de Cobertura Vegetal—Memoria Descriptiva. Available online: https://www.gob.pe/institucion/minam/informes-publicaciones/2674-mapa-nacional-de-cobertura-vegetal-memoria-descriptiva (accessed on 15 January 2022).
31. Chen, W.; Li, X.; He, H.; Wang, L. Assessing Different Feature Sets' Effects on Land Cover Classification in Complex Surface-Mined Landscapes by ZiYuan-3 Satellite Imagery. *Remote Sens.* **2018**, *10*, 23. [CrossRef]
32. Wu, C. Normalized Spectral Mixture Analysis for Monitoring Urban Composition Using ETM + Imagery. *Remote Sens. Environ.* **2004**, *93*, 480–492. [CrossRef]
33. 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, Remote Sensingcenter, Texas A&M hivemity, Colfegp Station, Texas, Washington, DC, USA, 10–14 December 1974; Volume 351, p. 309.
34. Huete, A.R.; Didan, K.; Miura, T.; Rodriguez, E.; Gao, X.; Ferreira, L. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]
35. 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]
36. 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]
37. Hall, D.K.; Riggs, G.A. Mapping Global Snow Cover Using Moderate Resolution Imaging Spectroradiometer (MODIS) Data. *Glaciol. Data* **1995**, *33*, 13–17.
38. Key, C.H.; Benson, N. *Measuring and Remote Sensing of Burn Severity: The CBI and NBR*; U.S. Geological survey Open-File Report; U.S. Geological Survey Wildland Fire Workshop: Los Alamos, NM, USA, 2000; pp. 2–11.
39. Zha, Y.; Gao, J.; Ni, S. Use of Normalized Difference Built-up Index in Automatically Mapping Urban Areas from TM Imagery. *Int. J. Remote Sens.* **2003**, *24*, 583–594. [CrossRef]
40. Coulter, L.L.; Stow, D.A.; Tsai, Y.H.; Ibanez, N.; Shih, H.C.; Kerr, A.; Benza, M.; Weeks, J.R.; Mensah, F. Classification and Assessment of Land Cover and Land Use Change in Southern Ghana Using Dense Stacks of Landsat 7 ETM + Imagery. *Remote Sens. Environ.* **2016**, *184*, 396–409. [CrossRef]
41. Rosenqvist, A.; Shimada, M.; Ito, N.; Watanabe, M. ALOS PALSAR: A Pathfinder Mission for Global-Scale Monitoring of the Environment. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3307–3316. [CrossRef]
42. Jiliang, X.; Xiaohui, Z.; Zhengwang, Z.; Guangmei, Z.; Xiangfeng, R. Multi- Scale Analysis on Wintering Habitat Selection of Reeves' s Pheasant (Syrmaticus Reevesii) in Dongzhai National Nature Reserve, Henan Province, China. *Acta Ecol. Sin.* **2006**, *26*, 2061–2067.

43. Richards, J.A. *Supervised Classification Techniques*, 5th ed.; Springer: Berlin/Heidelberg, Germany, 2013; ISBN 9783642300615.
44. Noi, P.T.; Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors* **2018**, *18*, 18. [CrossRef]
45. Vega Isuhuaylas, L.A.; Hirata, Y.; Santos, L.C.V.; Torobeo, N.S. Natural Forest Mapping in the Andes (Peru): A Comparison of the Performance of Machine-Learning Algorithms. *Remote Sens.* **2018**, *10*, 782. [CrossRef]
46. Petropoulos, G.P.; Kalaitzidis, C.; Prasad Vadrevu, K. Support Vector Machines and Object-Based Classification for Obtaining Land-Use/Cover Cartography from Hyperion Hyperspectral Imagery. *Comput. Geosci.* **2012**, *41*, 99–107. [CrossRef]
47. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
48. Kotsiantis, S. Supervised Machine Learning: A Review of Classification Techniques. *Informatica* **2007**, *31*, 249–268.
49. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Springer: Singapore, 2006; ISBN 9780387310732.
50. Mayr, A.; Binder, H.; Gefeller, O.; Schmid, M. The Evolution of Boosting Algorithms: From Machine Learning to Statistical Modelling. *Methods Inf. Med.* **2014**, *53*, 419–427. [CrossRef]
51. Wacker, A.G.; Landgrebe, D.A. *Minimum Distance Classification in Remote Sensing*; LARS Technical Reports; Purdue University: Lafayette, Indiana, 1972; p. 25.
52. Caruana, R.; Niculescu-Mizil, A. An Empirical Comparison of Supervised Learning Algorithms. In Proceedings of the 23rd International Conference on Machine Learning—ICML'06, Carnegie Mellon University, Pittsburgh, Pennsylvania, 25–29 June 2006; pp. 161–168.
53. Vapnik, V. *The Nature of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 1999; ISBN 0387987800.
54. Max, K.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Ziem, A.; Scrucca, L.; et al. *Caret: Classification and Regression Training*; R Package Version 6.0-86; Astrophysics Source Code Library: Cambridge, MA, USA, 2020.
55. Pontius, R.; Ali, S. DiffeR: Metrics of Difference for Comparing Pairs of Maps or Pairs of Variables. 2019, p. 19. Available online: https://cran.r-project.org/web/packages/diffeR/diffeR.pdf (accessed on 28 January 2022).
56. Pontius, R.; Millones, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *Int. J. Remote Sens.* **2011**, 37–41. [CrossRef]
57. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]
58. Hijmans, R.J.; Van Etten, J.; Sumner, M.; Cheng, J.; Baston, D.; Bevan, A.; Bivand, R.; Busetto, L.; Canty, M.; Fasoli, B.; et al. Raster: Geographic Data Analysis and Modeling. 2020, p. 249. Available online: https://cran.r-project.org/web/packages/raster/raster.pdf (accessed on 28 January 2022).
59. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef]
60. Räsänen, A.; Virtanen, T. Data and Resolution Requirements in Mapping Vegetation in Spatially Heterogeneous Landscapes. *Remote Sens. Environ.* **2019**, *230*, 111207. [CrossRef]
61. Azzari, G.; Lobell, D.B. Landsat-Based Classification in the Cloud: An Opportunity for a Paradigm Shift in Land Cover Monitoring. *Remote Sens. Environ.* **2017**, *202*, 64–74. [CrossRef]
62. Campos-Taberner, M.; Moreno-Martínez, Á.; García-Haro, F.J.; Camps-Valls, G.; Robinson, N.P.; Kattge, J.; Running, S.W. Global Estimation of Biophysical Variables from Google Earth Engine Platform. *Remote Sens.* **2018**, *10*, 1167. [CrossRef]
63. Yaranga, R.; Custodio, M.; Chanamé, F.; Pantoja, R. Floristic Diversity in Grasslands According to Plant Formation in the Shullcas River Sub-Basin, Junin, Peru. *Sci. Agropecu.* **2018**, *9*, 511–517. [CrossRef]
64. Weng, C.; Bush, M.B.; Curtis, J.H.; Kolata, A.L.; Dillehay, T.D.; Binford, M.W. Deglaciation and Holocene Climate Change in the Western Peruvian Andes. *Quat. Res.* **2006**, *66*, 87–96. [CrossRef]