

Article

# Using Google Earth Engine to Map Complex Shade-Grown Coffee Landscapes in Northern Nicaragua

Lisa C. Kelley <sup>1,2,\*</sup>, Lincoln Pitcher <sup>3</sup> and Chris Bacon <sup>2</sup>

<sup>1</sup> Department of Geography and Environment, University of Hawai'i-Mānoa, Honolulu, HI 96822, USA

<sup>2</sup> Department of Environmental Studies and Sciences, Santa Clara University, Santa Clara, CA 95053, USA; cbacon@scu.edu

<sup>3</sup> Department of Geography, University of California-Los Angeles, Los Angeles, CA 90095, USA; lincolnpitcher@ucla.edu

\* Correspondence: lckelley@hawaii.edu; Tel.: +1-808-956-8465

Received: 2 May 2018; Accepted: 11 June 2018; Published: 14 June 2018



**Abstract:** Shade-grown coffee (shade coffee) is an important component of the forested tropics, and is essential to the conservation of forest-dependent biodiversity. Despite its importance, shade coffee is challenging to map using remotely sensed data given its spectral similarity to forested land. This paper addresses this challenge in three districts of northern Nicaragua, here leveraging cloud-based computing techniques within Google Earth Engine (GEE) to integrate multi-seasonal Landsat 8 satellite imagery (30 m), and physiographic variables (temperature, topography, and precipitation). Applying a random forest machine learning algorithm using reference data from two field surveys produced a 90.5% accuracy across ten classes of land cover, with an 82.1% and 80.0% user's and producer's accuracy respectively for shade-grown coffee. Comparing classification accuracies obtained from five datasets exploring different combinations of non-seasonal and seasonal spectral data as well as physiographic data also revealed a trend of increasing accuracy when seasonal data were included in the model and a significant improvement (7.8–20.1%) when topographical data were integrated with spectral data. These results are significant in piloting an open-access and user-friendly approach to mapping heterogeneous shade coffee landscapes with high overall accuracy, even in locations with persistent cloud cover.

**Keywords:** shade-grown coffee; dry tropics; Nicaragua; land-cover classification; multi-temporal data; Landsat 8; random forest algorithm; Google Earth Engine

## 1. Introduction

Shade-grown coffee (shade coffee) production is an important component of the forested tropics [1–4]. This is particularly true of smallholder shade coffee systems characterized by complex and structurally diverse shade management practices [5–7]. In contrast to sun-grown coffee (sun coffee) or coffee grown under sparse shade, complex rustic shade coffee agroforests are associated with a diversity of provisioning ecosystem services [6] and a high diversity of associated fauna, including mammals, arthropods, ants, butterflies, and bees [7,8]. Given their structural and biological similarity to forested lands, rustic shade coffee lands also provide habitat connectivity for forest-dependent species navigating fragmented landscapes [9]. Over the past two decades, there have thus been a number of studies that have aimed to produce spatially explicit maps of shade coffee production using remote sensing technologies [10–13].

Accurately detecting shade-grown coffee using satellite imagery, however, has long proven challenging. Analysts have reported particularly low accuracy in distinguishing between shade

coffee and young and mature woodland or forest, limiting an understanding of coffee production in relation to other forested and agricultural land covers (e.g., 37.5–58.7% in an early effort by the authors of [10]; see also References [11–15]). Most shade coffee is also produced in plots of land smaller than the resolution of openly accessible satellite data (e.g., MODIS) under a canopy of forest trees [12]. Resulting rustic shade coffee systems thus closely resemble adjacent patches of forested land in visible/near-infrared wavelengths [15–17]. Topographical effects (e.g., shadows) and chronic cloud cover compound mapping challenges in the mountainous regions where most coffee is grown [11,13].

A number of recent studies have explored how refined classification algorithms and high-resolution passive spaceborne and airborne data might be used to improve smallholder coffee classification. In one study, QuickBird imagery (0.6–2.4 m spatial resolution) was combined with a neural network classifier to predict the distribution of shade coffee trees with 96.9% accuracy in New Caledonia [12]. Another study integrated aerial orthoimagery (0.25 m spatial resolution) with existing maps of forest cover, elevation data, and expert knowledge in a Bayesian predictive model to classify coffee with 87% overall accuracy in Rwanda [18]. Analysts have also explored classification using open-access Landsat data (30 m) [19], often integrating Landsat data with physiographic variables. The authors of [11], for example, showed how data on land surface temperature from Landsat ETM+ thermal imagery can help detect subtle differences in leaf and canopy density, morphology, biomass, species composition, and canopy water status in shade coffee lands (see also Reference [20]). By integrating these data with visible and infrared bands and a post-classification stratification model using elevation and precipitation data, these authors achieved a producer and user accuracy of 91.8% and 61.1% respectively for shade coffee in Costa Rica's highlands. Elsewhere, the role of multi-temporal Landsat data in accurately mapping shade coffee has been explored [13,17]. By integrating three Landsat ETM+ scenes from various seasons, for example, one study from El Salvador differentiated between coffee plantations and non-mangrove forest with 81.6% accuracy [13].

Outside the coffee sector, multi-temporal or seasonal spectral features or indices are now commonly used to identify otherwise difficult-to-classify agricultural and forested lands [21–25]. Such approaches are capable of exploiting information on intra-annual variation in dominant crops or vegetative covers [25–27]. Multi-temporal spectral datasets have been particularly effective in achieving high classification accuracy when satellite imagery is integrated with physiographic or textural variables [22,28,29]. The authors of [28], for example, obtained an 86% overall accuracy across 14 land-cover classes in a heterogeneous landscape in southern Spain by combining seasonal imagery with digital terrain measures and textural data. The authors of [22] mapped larch plantations in China with 91.9% accuracy by combining multi-seasonal Landsat 8 imagery and bivariate textural features in a random forest (RF) classification model.

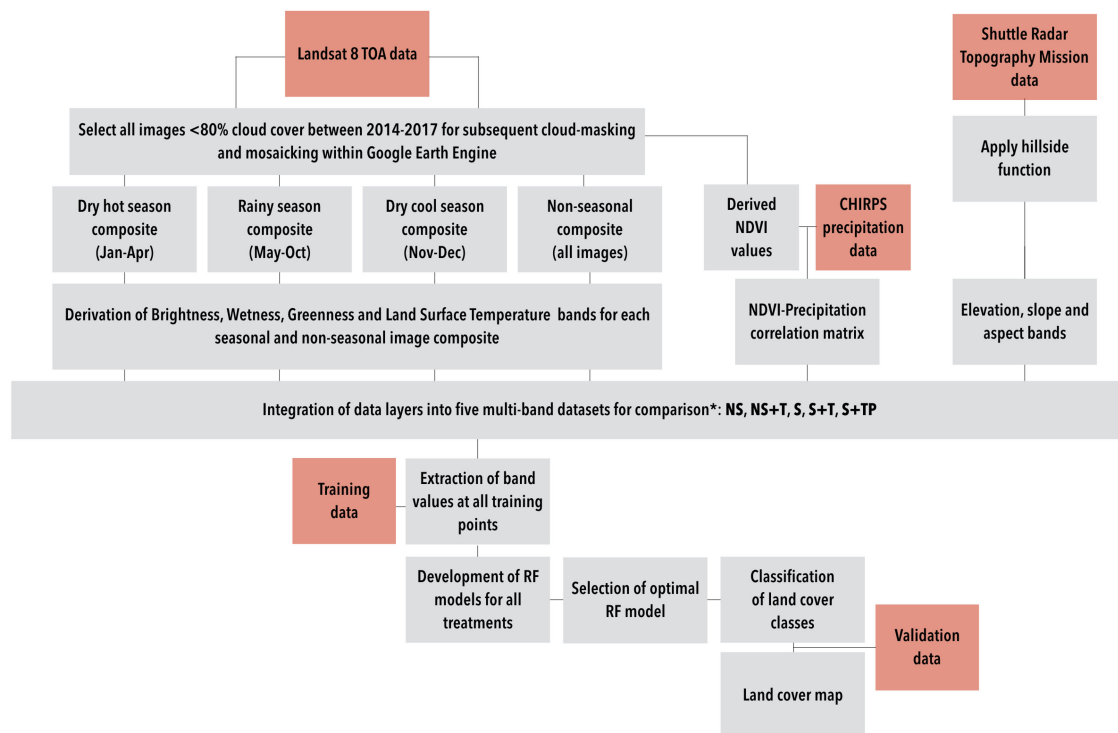
Historically, the cost of satellite imagery precluded such approaches, particularly discouraging their application across broad spatial scales and/or in regions marked by high cloud coverage [30,31]. Since 2008, however, open-access to the entire Landsat imagery archive has resolved this challenge, in effect creating the world's largest openly accessible library of geospatial information at 30-m resolution [30]. The development of Google Earth Engine (GEE) and other cloud-based computing techniques, in turn, has allowed analysts to process and analyze dense stacks of satellite imagery on the fly. GEE, for example, hosts the entire Landsat archive, provides tools and an application program interface for summoning, processing, and mosaicking this imagery, and runs all analyses in parallel across many machines in Google's cloud-based processing platform [32]. As prior work has illustrated, these developments enable analyses at high computational efficiency even when relying on multi-temporal image mosaics [23,33,34]. The authors of [33], for example, used >650,000 Landsat scenes and >1 million central processing unit (CPU) hours to produce annual maps of tree cover gain and loss, performing these computations in just several days. New cloud-based computational capacities also enable the efficient use of "ensemble" classifiers such as RF which average across hundreds of individual decision trees to ensure unbiased land-cover classifications [22,35].

This study draws on these advancements to develop an open-access and multi-temporal approach to mapping not only shade coffee but also overall land cover with high accuracy, here piloting this approach in three districts of northern Nicaragua. While there are multispectral spaceborne sensors with higher temporal (e.g., planet CubeSats) and spatial (e.g., WorldView-2/3/4) resolutions that might be used to enhance coffee and overall land-cover classifications, much of these data remain proprietary and/or have insufficient accuracies for time-series analysis, precluding their extension to other production contexts. For example, the authors of [36] used WorldView-2 to classify coffee fields in Hawai'i with 81% overall accuracy; these data, however, are not freely available and therefore cannot be used for our study in Nicaragua. The high temporal and spatial resolution of Planet CubeSat data (coupled with non-commercial access for academic research) makes this an appealing solution for land-cover classification applications. However, the absence of a cloud mask, large geolocation errors, and poor cross-platform radiometric calibrations render these data difficult to use in time-series-based spectral analyses [37]. Open-access approaches to mapping shade coffee based on Landsat data, in contrast, have often been limited by issues of cloud cover or by poor classifier performance across non-coffee land-cover classes (e.g., a 30% and 50% user's and producer's accuracy for non-woody land covers in Reference [11]), inhibiting an understanding of the landscape "matrix" [38] within which shade coffee and associated forest-dependent biodiversity are situated. For example, in the northern Nicaragua study area that we focused on, smallholder shade-coffee plots are typically interspersed with pine stands, pine-oak forests, semi-humid broadleaf tropical forest, and significant quantities of mixed-use pasture, integrated corn and bean production (or, "milpa"), and vegetable production [39]. Informed by the above developments, and to address these challenges, we pursued two primary objectives:

- (1) Developing an open-access and multi-temporal approach to classifying both shade coffee and overall land cover with high accuracy in northern Nicaragua.
- (2) Assessing the relative value of various spectral and ancillary data combinations in enabling shade coffee and overall land-cover classification accuracy in this context.

## 2. Materials and Methods

Figure 1 illustrates the full approach we adopted, while Table 1 details the open-access datasets on which we relied. We began by using dense stacks of Landsat 8 imagery within GEE to develop relatively gap-free (i.e., <5%) image mosaics for three distinct seasons in the northern Nicaragua study area: dry hot season months (January–April), rainy season months (May–October), and dry cool season months (November–December). We also developed a non-seasonal composite produced from all imagery spanning these months with the goal of more explicitly assessing how separate seasonal spectral variables may improve classification accuracy. Next, we applied a Kauth–Thomas (KT) linear transformation (or a tasseled cap-like transformation) to each image mosaic [40–42], reducing spectral data to three spectral features with known values in capturing seasonal differences in vegetative growth and soil moisture: brightness, wetness, and greenness [28,40]. Finally, we used an RF algorithm [36,43,44] to build classification models for five distinct combinations of spectral and ancillary data, training and validating each RF model with reference data collected during two field surveys. We describe each of these methods and research treatments in greater detail below.



**Figure 1.** Overview of the classification approach. Red shading indicates primary datasets used in this analysis. \*NS refers to non-seasonal data (brightness, wetness, greenness, and temperature); S refers to multi-seasonal data (rainy/dry hot/dry cool brightness, wetness, greenness, and temperature); T refers to elevation, slope, and aspect data; and P refers to the correlation matrix between precipitation and normalized difference vegetation index (NDVI). Treatments are described in greater detail in Section 2.8.

**Table 1.** Datasets Used in Land-Cover Classification.

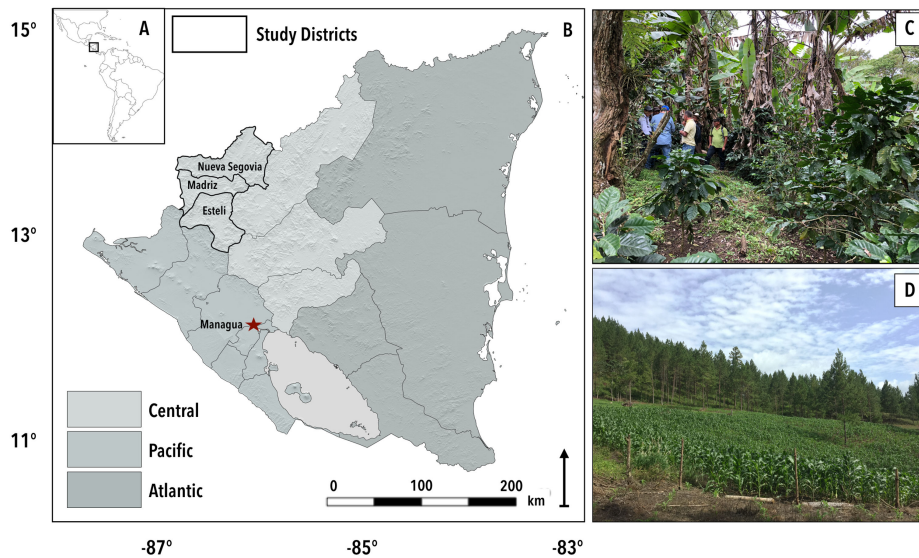
Dataset	Use	Resolution	Source
<i>Landsat 8 TOA Reflectance</i>	Spectral indices; land surface temperature data; NDVI trajectories	<i>Spatial:</i> 30 m <i>Date range:</i> 2014–2017	USGS/NASA
<i>CHIRPS Precipitation Data</i>	Construction of NDVI-precipitation correlation matrix	<i>Spatial:</i> 0.05° <i>Date range:</i> 2014–2017	Funk et al. (2014)
<i>Shuttle Radar Topography Mission</i>	Digital Elevation Raster	<i>Spatial:</i> 30 m	STRM, NASA
<i>Tree Canopy Coverage</i>	Data validation	<i>Spatial:</i> 30 m <i>Date range:</i> 2010	Sexton et al. (2013)

Notes: Acronyms are as follows: TOA, Top of Atmosphere; CHIRPS, Climate Hazards Group InfraRed Precipitation with Station data; NDVI, Normalized Difference Vegetation Index; USGS, United States Geological Survey; NASA, National Aeronautics and Space Administration; STRM, Shuttle Radar Topography Mission.

## 2.1. Study Area

This study focused on a 7080 km<sup>2</sup> area spanning the administrative districts of Estelí, Madriz, and Nueva Segovia in the central highlands of Nicaragua [45] (Figure 2). These districts have three mountain ranges running east to west with considerable topographical variation, ranging from plateaus at ~800 m above sea level (masl), and mountain peaks along the border with Honduras reaching 1600 masl, to valleys outside the urban areas of Condega and Estelí at ~700 masl [46]. The central highland region we focused on also typically experiences a drier and cooler climate than the hot, dry Pacific coast, or the humid, rainy, and mostly forested Atlantic coast [47]. The region receives an average of 1357 mm of rain annually, more than 70% of which falls during the rainy season (May–October) or the dry cool season months (November–December). Dry hot season months

(January–April) are characterized by almost no rainfall with average temperatures of 2–5 °C warmer, which triggers leaf loss among obligatory deciduous tree species [46].



**Figure 2.** Overview of the study area in northwest Nicaragua. (A) Nicaragua in a Central American regional context. (B) Primary climate zones (grey) with study areas noted using thick district boundary lines. The specific study districts included Esteli, Madriz, and Nueva Segovia. (C) Field photo of discussion about participatory mapping approaches in a structurally complex shade-coffee plot. (D) Field photo of “milpa” croplands adjacent to pine stands.

## 2.2. Land-Cover Classification Categories

Based on existing classifications of land covers and forest ecosystems in the study area [39,45], and participatory dialogue and initial data sharing with local primary cooperatives, we selected ten land-cover classes for analysis (Table 2). These 10 classes exhibited different patterns of intra-annual changes in vegetation, based on the seasonality of smallholder livelihoods, differences in the vegetative composition of these land covers, and differences in their elevation gradients (see Table 2). We exploited these distinctions in our classification approach (see Figure 1). The three cultivated coffee species (*Coffea arabica*, *Coffea canephora*, and *Coffea liberica*) grown globally are typically found in one of five different formations: rustic polyculture, traditional polyculture, commercial polyculture, shaded monoculture, and sunned monoculture [2,3]. In the northern Nicaragua study districts, nearly all smallholder coffee corresponds to “rustic coffee systems”, wherein coffee trees are planted under a thin layer of planted or naturally occurring forest and evergreen shade trees.

**Table 2.** Land-Cover Classes.

Class	Description
“Milpa” (1)	Open area seasonally covered by rotational corn and bean crops (first cycle: May–August, second cycle: September–November)
Multi-Use Pasture (2)	Open area predominantly covered by grass species though some woody tree species, including <i>Pinus</i> species, are occasionally present (<30% cover)
Pine Stands (3)	Forested area predominantly covered by <i>Pinus</i> species (>30% cover)
Pine-Oak Forests (4)	Forested area predominantly covered by mixed <i>Pinus</i> species and <i>Quercus</i> species (>30% cover)
Semi-Humid Broadleaf Forests (5)	Forested area (>30% cover) typically found at higher elevations; can consist of as much as 75% facultative deciduous species

Table 2. Cont.

Class	Description
<i>Deciduous Dry Forest</i> (6)	Forested area (>30% cover) typically found at lower elevations; consists of 75–100% deciduous species
<i>Shade Coffee</i> (7)	Agroforested area (>30% tree cover) predominantly covered by coffee production, and diverse deciduous and evergreen shade trees
<i>Built</i> (8)	Urban area, roads, other constructed settlement areas
<i>Water</i> (9)	Water bodies
<i>Wet Rice</i> (10)	Irrigated rice fields

### 2.3. Landsat Imagery

This study used open-access Landsat 8 Top of Atmosphere (TOA) reflectance data hosted within GEE to construct satellite image mosaics (see Table 1). Many available Landsat images within GEE are already processed to a relatively high level of geometric and terrain accuracy. Landsat 8 TOA reflectance data from the Collection 1 Tier 1 collection represent the highest possible quality imagery available in the Landsat 8 collection [48]. Scenes in the Tier 1 collection are consistently geo-registered, meaning that all images are corrected for displacement using ground control points and digital elevation model data. Within the Tier 1 collection, all images are registered with a root-mean-square error (RMSE)  $\leq 12$  m [48]. This geometric registration ensures the pixel-to-pixel correspondence necessary for the integration of multi-temporal imagery [23].

In addition to geometrically registered datasets, radiometric normalization is necessary when using multi-temporal imagery; this normalization ensures that spectral-radiometric properties are consistent across observations taken on different days or by different sensors [49]. Radiometric calibration involves converting raw and unprocessed digital numbers (DN) for each spectral band contained within a given Landsat scene into at-sensor radiance values that account for the specificities of the sensor that acquired the imagery, including measurement changes, mechanical malfunctions, or deterioration in sensor quality [49,50]. We elected to use the Tier 1 collection because all images in this collection were already radiometrically calibrated using well-established methods to ensure their consistency with other scenes [48].

After radiometric normalization, at-sensor reflectance also needs to be converted to a planetary reflectance value. Imagery can be converted into either Top of Atmosphere (TOA) or Surface Reflectance (SR) values [51]. We selected the TOA collection rather than SR reflectance data because initial tests suggested that a Kauth–Thomas (KT) linear transformation, described below, greatly improved classification accuracy. Currently, KT coefficients for Landsat 8 data are best established for TOA data [52–54]. TOA data have also consistently been used to produce multi-temporal image mosaics that resulted in high-accuracy land-cover classifications, particularly when spectral indices or transformations have been used to enhance spectral signals [23,25,31,55,56]. TOA reflectance values for the Tier 1 collection were computed using well-established calibration coefficients from Reference [49].

### 2.4. Development of Multi-Seasonal and Non-Seasonal Image Composites

Cloud cover is a persistent challenge in mapping land-cover types with Landsat data, particularly in mountainous regions of the tropics. Persistent clouds or shadows over certain pixels can also distort multi-temporal image mosaics, reducing accuracy. We addressed this challenge by first filtering the entire Landsat 8 TOA Tier 1 collection by date and cloud cover, selecting only those images spanning the years 2014–2017 with <80% cloud cover. Through this process, we identified a total of 52, 64, and 27 Landsat scenes for dry hot, rainy, and dry cool season months, respectively (143 scenes in total, see Supplemental Materials, Table S1). These Landsat scenes were then pre-processed using cloud- and shadow-masking algorithms made openly-accessible through the GEE user community. These approaches rely on a simpleCloudScore() function within the GEE, which produces a cloud mask based

on several indicators of cloudiness, as well as a shadow-masking function that identifies dark outliers by comparing single-date observations with pixel-based statistics based on time-series data [57].

Following the application of cloud- and shadow-masking techniques, a median pixel value was taken across all observations spanning the three-year study period. Elsewhere, analysts used this approach to produce relatively gap-free image mosaics, even in areas of persistent cloud coverage [34,58,59]. Though analysts previously used shorter durations of time to construct image mosaics (given challenges linked to inter-annual land use and cover changes, e.g., References [58,59]), a three-year period was necessary to achieve relatively gap-free image mosaics given particularly high cloud coverage in the study area. Note that a median pixel value was extracted not only from all pre-processed imagery specific to each season (dry hot, rainy, and dry cool), but also from imagery spanning all seasons to produce a non-seasonal composite. We produced this latter composite for the explicit sake of assessing the relative value of isolating seasonal spectral features.

### 2.5. Kauth–Thomas Linear Transformation

Following image processing, a Kauth–Thomas (KT) linear transformation was applied to the three seasonal and one non-seasonal image mosaics. The KT transformation was originally developed using Landsat data for agricultural applications [40], and converts original Landsat bands into a new set of bands which disaggregate pixels into brightness, greenness, and wetness components, reducing the number of spectral features for classification [28,29,60]. The resulting brightness band measures soil brightness as a weighted sum of all bands and is particularly well suited to capturing urban areas. The greenness band captures the contrast between the near-infrared bands and visible bands and is typically highest in areas of high vegetative biomass and photosynthetically active vegetation. The wetness band contrasts visible and near-infrared bands with longer-infrared bands and provides an index of either soil or surface moisture (alternatively, the amount of dead/dried vegetation) [28,41,61].

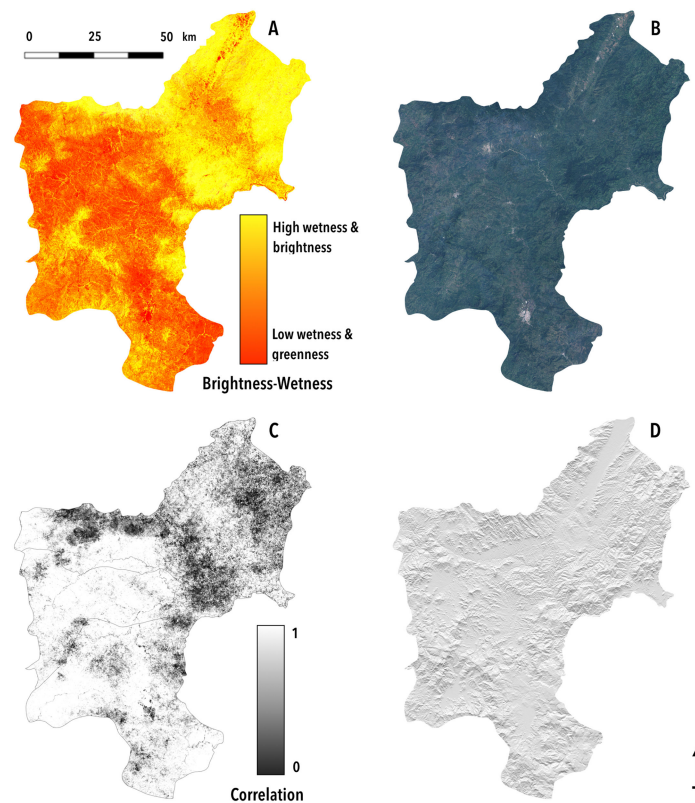
We applied a KT linear transformation, given its successful application in other complex and heterogeneous systems, and because it is well established as a means of compressing spectral data and reducing the correlation between spectral features with minimal information loss [52,61]. The KT linear transformation is also considered a particularly relevant tool in capturing phenological differences throughout the year given its capacity to represent physical features of the landscape [62]. We considered this helpful to our analysis given our emphasis on exploiting potential distinctions in the seasonality of modeled land-cover types. KT linear transformations of Landsat imagery are reliant on sensor- and data-specific (e.g., radiance and reflectance) coefficients [40,41]. We used coefficient values tailored to Landsat 8 TOA reflectance data provided by Reference [52].

### 2.6. Land Surface Temperature, Topography, and Precipitation

Non-seasonal and multi-seasonal brightness, greenness, and wetness bands were then merged with gridded land surface temperature, topography, and precipitation data, given the established value of these inputs in earlier approaches to mapping coffee [11,13,18] (Figure 3). Median values of land surface temperature available from the initial pre-processed Landsat 8 image composites were used for land surface temperature data. All topographical data layers (elevation, aspect, and slope) were then derived from the Shuttle Radar Topography Mission (STRM) Digital Elevation Data product provided by NASA/USGS/JPL-Caltech available within GEE at a 30-m resolution [63] (see Table 1). This dataset contains one band for elevation. By applying a hillshade function to this band made available through GEE and NASA SRTM values for aspect and slope were also derived.

In prior approaches to mapping shade coffee, data on precipitation were used as part of a post-classification stratification model that ensured classified coffee lands were only mapped within seemingly appropriate environmental contexts [11]. In this analysis, we explored the value of integrating precipitation data into the classification model, here assessing how information on the correlation between precipitation and changes in vegetation on a 30-day time lag could contribute to classification accuracy [64]. We included these data at the pixel scale using openly accessible

Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) pentad precipitation data [65]. The goal of modeling the relationship between precipitation and Normalized Difference Vegetation Index (NDVI) was to explore if the differentiated responses to rainfall (e.g., impervious roads vs. rain-fed croplands) also improved classification accuracy. Vegetation greenness was modeled using derived NDVI values for all cloud-masked Landsat 8 TOA images. NDVI is a common measure of vegetation productivity, and takes advantage of contrasting effects of vegetation on red and near-infrared reflectance [66,67].



**Figure 3.** (A–D) Datasets used for land-cover classification. (A) Seasonal image composite spanning dry hot months (January–April), visualized using derived brightness, greenness, and wetness bands. (B) Seasonal image composite spanning dry cool months (November–December), visualized in natural color composite using red, green, and blue spectral bands. (C) Correlation matrix of precipitation and NDVI on a 30-day lag. (D) Hillshade derived from Shuttle Radar Topography Mission (SRTM) data. Data are presented for the three study districts. Coordinates from west to east:  $-86.8^{\circ}$  to  $-85.8^{\circ}$ . Coordinates from south to north:  $12.8^{\circ}$  to  $14.1^{\circ}$

### 2.7. Training and Validation Data

Training data are an important input into supervised classification approaches, with the structure of training data particularly important to the RF classifier we drew on here [36,68,69]. One study, for example, showed that using proportionally weighted training samples for each class reduces the likelihood of the RF model over-classifying under-represented classes, while using balanced training samples reduces the likelihood of under-represented classes being omitted from the model [69]. This study relied on a balanced sample of 200 training points for all ten land-cover classes collected during two in situ surveys (10–21 July 2017; 9–19 December 2017). We collected 200 points for each class, reflecting established understandings that data points between  $10p$  and  $30p$  were needed for each class, where  $p$  = the number of spectral bands included in the input dataset [70]. We used a balanced sample



given our emphasis on achieving not only high shade-coffee mapping accuracy but also high overall classification accuracy.

In our case, it was impossible to fully randomize the collection of training data given limitations in field time and the participatory nature of the research project. We relied on a systematic random sample based on two primary sources of data: (i) a community-based mapping exercise where specific land parcels were tied to particular land covers, and (ii) six roughly two-hour transects that were walked alongside community leaders. Data from community-based maps were subsequently used within GEE to pull a random sample of points. Within walking transects, data on land covers were also collected at 20 m intervals using a consumer grade handheld global positioning system (GPS). Transects spanned diverse land-cover types, including known sites of shade coffee, semi-humid broadleaf forest, “milpa”, multi-use pasture, and deciduous dry forest. Visual interpretation of non-seasonal Landsat image composites was also used to collect training data for water and built areas.

All validated field data were then randomly shuffled and re-ordered within R software. Classes with more than 200 training points were under-sampled to create a balanced training set for classification, and values for all predictor variables were extracted from the multi-band composite at each data point using the `.sampleRegions()` algorithm within GEE. A balanced 20% subset of this dataset ( $N = 400$ ) was reserved for use in validating all RF models.

### 2.8. Classification with Random Forest

To address our second research objective, we next developed five datasets for comparison and classification (Table 3). As noted above, we included the non-seasonal image mosaic as one of these treatments to better evaluate the specific role of isolated seasonal spectral variables in enabling accurate classifications. We began by constructing an RF model for all multi-seasonal brightness, greenness, and wetness variables, and physiographic variables (**S+TP**), using the RF statistical package within the R software [71]. We then constructed RF models for non-seasonal predictor variables, and for multi-seasonal and non-seasonal brightness, greenness, and wetness bands, with and without topographical information. We compared the accuracy obtained by the full **S+TP** model with the accuracy obtained from four other models: (i) non-seasonal brightness, greenness, wetness, and temperature data (**NS**); (ii) non-seasonal brightness, greenness, wetness, and temperature data + topographical information (**NS+T**); (iii) multi-seasonal brightness, greenness, wetness, and temperature data (**S**); and (iv) multi-seasonal brightness, greenness, wetness, and temperature data + topographical information (**S+T**).

**Table 3.** Datasets for Comparison and Classification.

Dataset	Bands	Data Layers
<b>NS</b>	4	Non-seasonal brightness, greenness, wetness, and land surface temperature data
<b>NS+T</b>	7	<b>NS</b> dataset; elevation, slope, and aspect data
<b>S</b>	12	All seasonal (dry hot, rainy, and dry cool) brightness, greenness, wetness, and land surface temperature data
<b>S+T</b>	15	<b>S</b> dataset; elevation, slope, and aspect data
<b>S+TP</b>	16	<b>S+T</b> dataset; precipitation-NDVI correlation matrix

Classification models currently available within GEE include maximum entropy models, support vector machines with linear, polynomial, and Gaussian kernels, classification and regression trees, and naïve Bayes [72]. We elected to use an RF classification model, given initial classification tests and a growing body of evidence highlighting the value of RF and other “ensemble” models in enabling fine-scale classifications on the basis of marginal spectral differences between land-cover classes [28,29,73–75]. Ensemble classifiers, such as RF, were shown to effectively distinguish between

spectrally similar agricultural and forested land-cover types by building multiple trees from provided input and training data [35].

RF algorithms work by growing a user-specified number of random trees (N) to full depth, using only a randomly selected subset of predictor variables (*mtry*) to search for the best split at each node [35]. RF algorithms also apply a bagging approach (i.e., randomly resampling subsets of the training data and predictor variables with no deletion of data), allowing for the creation of numerous relatively unbiased models which can later be averaged [35,76–78]. Our approach to determining the optimal *mtry* value for each subset of predictor variables involved creating a large number of RF models for each dataset, and selecting that which produced the highest accuracy for a given dataset [78]. Supplemental Materials Table S2 provides an illustration of this approach for the full RF model. All RF models were constructed by setting N = 500, as error rates for all models stabilized before this point.

### 2.9. Accuracy and Variable Importance Assessments

The accuracy of all constructed RF models was assessed in a number of ways. We relied primarily on error matrices (Supplemental Materials Table S3) displaying RF model predictions relative to validated reference data. The diagonal elements in such matrices indicated the number of samples for which predictions matched reference data, while off-diagonal elements in rows and columns captured errors of commission and omission (pixels mistakenly committed to or omitted from a class). These matrices were used to estimate user's accuracy (the accuracy of classification despite errors of commission) and producer's accuracy (the accuracy of classification despite errors of omission). For each RF model below, we also provided kappa error (K) matrix statistics, which accounted for the possibility of chance agreement between predicted values and reference data [13]. Values of K ranged from 0 (no agreement) to 1 (perfect agreement), with K values >0.80 indicating strong agreement, and values of  $0.40 < K < 0.80$  indicating moderate agreement [76].

We relied on two further measures of validation. Firstly, we used a visual assessment of the land-cover map relative to openly accessible data on percentage tree canopy coverage, for the reference year 2010, available at 30-m resolution [77]. Secondly, we used variable importance analyses that provided us an additional means of assessing how specific multi-spectral seasonal and ancillary predictor variables enabled accuracy improvements in the optimized land-cover classification model. Below, we present the importance of these variables in terms of percentage decrease in overall accuracy when that variable was included in the analysis. Within the R software, these measures were developed using internal RF estimates of prediction errors for all training samples ( $x_i$ ) against all trees that were not constructed using training sample  $x_i$ , or an "out-of-bag" (OOB) error. Variable importance analyses assess how much OOB error increases when a particular predictor variable is removed [71,78].

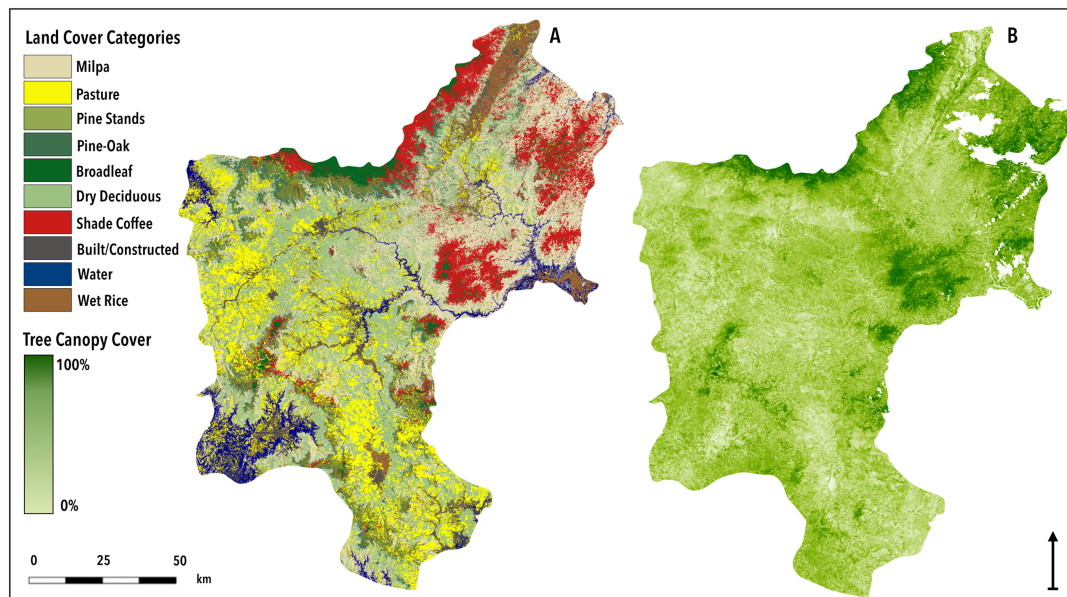
## 3. Results

### *The Distribution of Shade-Grown Coffee in Northern Nicaragua*

Running the RF model (N = 500) on the full set of multi-seasonal brightness, greenness, wetness, and ancillary physiographic bands (**S+TP**, *mtry* = 3) produced an overall accuracy of 90.5% across all classes (kappa index = 0.89), the highest accuracy achieved by any model. This model suggested that shade-grown coffee is typically grown in mid-range elevations, often adjacent to semi-humid broadleaf forests, pine-oak forests, and pine stands, while dry deciduous forest, "milpa", and pasture lands tend to dominate lowland regions (Figure 4).

Table 4 provides the accuracy estimates for the full model, depicting field-validated reference data (columns) against RF model predictions (rows). This matrix revealed relatively high user's and producer's accuracy for most woody and non-woody land covers, including "milpa" croplands, shade coffee, semi-humid broadleaf forests, pasture, pine stands, and pine-oak forests. It also revealed that confusion between shade coffee and semi-humid broadleaf forest may be the core factor limiting higher accuracy estimates across these classes. Ten percent of known shade coffee points were falsely

identified as semi-humid broadleaf forest, while 7.5% of all points classified as shade coffee were actually semi-humid broadleaf forest. Dry deciduous forest, “milpa” croplands, multi-use pastures, and built areas were also confused by the final RF model.



**Figure 4.** Land-cover classifications as compared to data on tree canopy coverage. (A) Land cover classification produced by the full multi-seasonal and topographical data model (S+TP,  $mtry = 3$ ,  $N = 500$ ) with shade coffee depicted in bright red. (B) Estimated tree canopy coverage for the reference year 2010, available at 30-m resolution [77]. Coordinates from west to east:  $-86.8^\circ$  to  $-85.8^\circ$ . Coordinates from south to north:  $12.8^\circ$  to  $14.1^\circ$

**Table 4.** Confusion Matrix of Best Classification Achieved for the 16-Band Full Dataset.

S+TP													
ID	1	2	3	4	5	6	7	8	9	10	Total	UA	PA
1	34	2	0	0	0	1	2	2	0	0	41	82.9	85.0
2	0	33	0	1	0	1	0	2	0	0	37	89.1	82.5
3	1	2	40	0	0	0	1	0	0	0	44	90.9	100
4	0	0	0	38	0	0	1	0	0	0	39	97.4	95.0
5	0	1	0	0	37	0	4	0	0	0	42	88.1	92.5
6	3	2	0	0	0	38	0	2	0	0	45	84.4	95.0
7	1	0	0	1	3	0	32	2	0	0	39	82.1	80.0
8	0	0	0	0	0	0	0	31	1	0	32	96.9	77.5
9	1	0	0	0	0	0	0	0	38	0	39	97.4	97.4
10	0	0	0	0	0	0	0	1	0	40	41	97.6	100
<b>Total</b>	40	40	40	40	40	40	40	40	39	40	361/399	<b>OA: 90.5%</b>	

**Kappa: 0.89**

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. ID: identifier; UA: user’s accuracy; PA: producer’s accuracy; and OA: overall accuracy.

Table 5 also presents the results of the full model relative to other tested subsets of predictor variables. These data depicted a trend of increasing overall land-cover classification accuracy when multi-seasonal spectral data and ancillary data layers were included in the classification model. They also revealed a significant improvement (7.8–20.1%) when topographical data were included in the model. These data showed that RF models typically produced the highest accuracy when relatively fewer predictor variables were used to make each split (i.e., <50% of predictors, see also Supplemental

Materials Table S2). Underlying error matrices for all other tested subsets of predictor variables are provided in Supplemental Materials Table S3.

**Table 5.** Increases in classification accuracy achieved with multi-seasonal spectral variables, topographical information, and precipitation data.

Dataset	NS	NS+T	S	S+T	S+TP
Number of variables	4	7	12	15	16
<i>mtry</i>	3	3	7	3	3
Overall Accuracy (%)	65.6%	85.7%	81.7%	89.5%	90.5%
Kappa Index	0.62	0.84	0.80	0.88	0.89

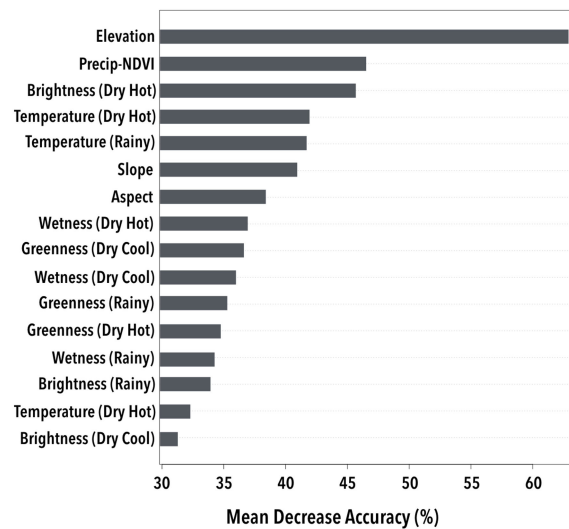
Notes: Classifications were conducted using random-forest classifiers (N = 500) with optimized numbers of split variables (*mtry*) for each subset of predictor variables tests. **NS** refers to the non-seasonal data (brightness, wetness, greenness, and temperature); **S** refers to multi-seasonal data (rainy/dry hot/dry cool brightness, wetness, greenness, and temperature); **T** refers to elevation, slope, and aspect data; and **P** refers to the correlation between precipitation and changes in NDVI.

Derived user's and producer's accuracies for each class and all models (Table 6) demonstrated the role of multi-seasonal spectral data in modestly improving the user's and producer's accuracy of shade coffee (5.3% and 5.0%, respectively, when integrated with topographic data). They also suggested the value of multi-seasonal spectral data in significantly improving the user's and producer's accuracy of "milpa", semi-humid broadleaf forest, dry deciduous forest, and built and constructed areas. While data on the correlation between precipitation and NDVI drove a modest improvement in overall land-cover classification accuracy relative to the multi-seasonal spectral and topographical model (1%, see Table 5), the inclusion of this data layer decreased user's accuracy of shade coffee and "milpa" croplands by 4.4% and 9.2%, respectively. These decreases were offset by increases in the user's accuracy of multi-use pasture and built/constructed areas by more than 10% (Table 6). Variable importance analyses illustrated the importance of elevation data, and dry hot season brightness and temperature data in enabling high overall classification accuracy in the full model (Figure 5).

**Table 6.** Changes in Per-Class Sensitivity and Specificity with Multi-Seasonal Spectral Variables, Topographical Information, and Precipitation Data.

ID	Class	NS	NS+T	S	S+T	S+TP
1	"Milpa"	34.1 (37.5)	81.6 (77.5)	59.5 (62.5)	92.1 (87.5)	82.9 (85.0)
2	Multi-Use Pasture	68.6 (60.0)	75.6 (77.5)	79.4 (67.5)	75.6 (77.5)	89.1 (82.5)
3	Pine Stands	76.0 (95.0)	95.2 (100)	88.9 (100)	95.2 (100)	90.9 (100)
4	Pine-Oak Forests	65.6 (52.5)	94.7 (90.0)	86.8 (82.5)	100 (95.0)	97.4 (95.0)
5	Broadleaf Forest	75.0 (82.5)	81.8 (90.0)	83.3 (87.5)	86.4 (95.0)	88.1 (92.5)
6	Deciduous Forest	52.8 (70.0)	80.9 (95.0)	80.4 (92.5)	83.0 (97.5)	84.4 (95.0)
7	Shade Coffee	77.1 (67.5)	81.2 (75.0)	73.7 (70.0)	86.5 (80.0)	82.1 (80.0)
8	Built/Constructed	68.8 (55.0)	78.1 (62.5)	78.8 (65.0)	84.4 (67.5)	96.9 (77.5)
9	Wet Rice	94.6 (89.7)	97.4 (94.5)	97.4 (94.9)	97.4 (94.9)	97.4 (97.4)
10	Water	51.4 (47.5)	90.5 (95.0)	88.4 (95.0)	95.2 (100)	97.6 (100)

Notes: Values reflect user's accuracy (true positive; listed first) and producer's accuracy (listed second).



**Figure 5.** Variable importance analyses for the full model (% change in mean accuracy, **S+TP**,  $mtry = 3$ ,  $N = 500$ ).

#### 4. Discussion

Accurately mapping complex and heterogeneous shade coffee landscapes is a first step in understanding the dynamics of land use and cover change in coffee-producing regions, including those linked to changes in smallholder livelihoods, market volatility, hydro-climatic change and variability, and Coffee Leaf Rust. Results from this study are significant in highlighting the value of multi-seasonal and open-access approaches in mapping complex and heterogeneous shade coffee landscapes with high overall accuracy, even in sites of persistent cloud cover. Results are also significant in showing how such approaches can be used to enable regional-scale (and larger) spatial analysis when integrated with image mosaicking techniques within GEE, demonstrated here for three districts in northwest Nicaragua.

The most significant finding from this study is the high shade coffee and overall land-cover classification accuracy achieved by the full model. Though the producer's accuracy for shade coffee achieved by this model is comparable with prior studies reliant on Landsat data, the user's accuracy for shade coffee is significantly higher (i.e., 80% vs. the 61% reported in Reference [11]). The 90.5% overall accuracy reported in this study across all land covers is also considerably higher than the 65% reported by Reference [11] across all classes, and the 76.7% reported by Reference [13] across woody land-cover classes. These results also compare favorably with prior analyses of coffee suitability in the region [78] and with available maps of forest and tree canopy coverage in the area [79]. The exception to this is the over-classification of water in the southwest and east corners of the land-cover classification. This distortion was potentially produced by the sensitivity of the RF classifier to spatial autocorrelation in the training sample, and may reflect the challenges of accurately mapping seasonally inundated water sources using multi-temporal datasets [68,80].

In this case, high classification accuracy was apparently enabled by combining multi-seasonal spectral data with topographical information, and information on precipitation-linked changes in vegetation. The full model produced an improvement in overall land cover accuracy of more than 20.4% relative to the non-seasonal image composite on its own. Despite persistent confusion between some classes, multi-seasonal spectral variables also appeared to contribute strongly to overall classification accuracy by exploiting distinctions in the floristic composition, structure, and intra-annual phenological variation across common woody and non-woody land-cover classes. In this case, multi-seasonal data was particularly valuable in driving improvements in the mapping accuracy of "milpa" croplands, dry deciduous forest, and built/constructed areas. The importance of dry hot seasonal information on brightness and land surface temperature is logical given the pronounced phenological variegation

during this point of the year (e.g., leaf senescence in dry deciduous forests, and seasonal fallowing of “milpa” fields). The importance of elevation information also corresponds with the strong shift (+37%) in coffee-mapping accuracy previously noted by Reference [18] when elevation data were included in the classification model.

In this study, including information on the correlation between precipitation and vegetation change decreased, rather than enhanced, mapping accuracy across shade coffee and semi-humid broadleaf forest classes. This finding is potentially explained by the coarser spatial resolution of the CHIRPS precipitation data used, which are available at  $0.05^\circ$  (or ~5-km resolution) in the study area vs. the 30-m resolution of Landsat 8 data. This is particularly true given that shade coffee is often proximal to semi-humid forest. Nonetheless, including these data appeared to modestly contribute to overall classification accuracy, here improving accuracy by 1% relative to seasonal and topographic variables on their own. Variable importance analyses and per-class accuracy estimates indicate that precipitation data played a strong role in driving accuracy in the full model, and played a particularly strong role in improving “milpa” and multi-use pasture classification accuracy. This is logical given that seasonal precipitation regimes closely inform the seasonality of smallholder “milpa” plantings. Future work could explore the possibility of fitting tailored precipitation models to each land-cover class, as some classes are unlikely to exhibit a simple correlative response to precipitation.

Higher accuracy of shade-grown coffee in this study was limited by apparent trade-offs between shade-coffee mapping accuracy and overall mapping accuracy. For example, for shade coffee, using just multi-seasonal and topographical predictors produced an 86.5% and 80.0% user’s and producer’s accuracy for shade coffee, respectively; these values decreased to 82.1% and 80.0% when data on precipitation and vegetation change were included in the model. Higher shade coffee mapping accuracy was also limited by classifier confusion between semi-humid broadleaf forest and shade coffee. While overall mapping accuracy was enhanced by the inclusion of multi-temporal imagery, it is possible that patterns of intra-annual phenological changes characteristic of shade coffee do not differ significantly enough from those characteristic of semi-humid broadleaf forests, particularly in borderline cases (e.g., where semi-humid broadleaf forest was thinned for timber products, or where shade-coffee lands exclusively contained shade trees also found in semi-humid broadleaf forests). One explanation for this is the high intra-class variability across both shade coffee and semi-humid broadleaf forest in northern Nicaragua, and the common occurrence of broadleaf forest species in smallholder shade-coffee plots, many of which also have home consumptive value (field observations).

A final notable facet of our analysis was our reliance on image mosaicking techniques and an RF ensemble classifier within GEE. It is possible that the use of imagery over a three-year period increased errors in sites of high inter-annual change, and it is also possible that using image mosaicking muted a signal of seasonality (i.e., by constructing mosaics from images in adjacent months). This potentially shaped the limited improvements in accuracy observed across shade coffee lands when using multi-seasonal spectral data, despite prior evidence of distinct patterns of seasonality in shade coffee plots [13,17]. Nonetheless, this technique allowed us to work around issues of frequent cloud cover across the study region, even when prioritizing seasonal imagery. Depending on study goals and desired coverage, the use of single-date images that maximize temporal distance (and differences in vegetation cover) between time periods may be the best approach. The overall high accuracy of this approach, however, suggests the value of such techniques across broad spatial scales. Our analysis also corroborates prior work illustrating the value of RF models in distinguishing between several vegetation types (e.g., “milpa” vs. multi-use pasture, pine stands vs. pine oak forests, and semi-humid broadleaf forests vs. shade coffee). Mirroring prior findings, RF models were optimized with small *mtry* values, which potentially helped avoid overly correlated individual decision trees [28].

There were at least three limitations to our study. Firstly, the value of these approaches in other parts of the tropics with shade coffee remains unexplored. The central region of northern Nicaragua we focused on is shaped by stronger seasonality when compared with other parts of the country and the tropics, and the study years corresponded with a severe drought [79]. During this

time of drought, leaf senescence was potentially more accentuated than at other times, accentuating the value of multi-seasonal spectral variables. Secondly, while this approach enabled a relatively accurate classification of shade coffee, data were too coarse to be analyzed for information on floristic composition [81], the number of vegetation strata [82,83], or the diversity of forest canopy structures [12], which are insights important to a full assessment of biological diversity. Thirdly, these data do not reveal the socio-political, economic, and cultural dynamics producing this heterogeneous and complex landscape. These factors not only mediate the distribution of biological diversity. They also shape the apparent biophysical distinctions in seasonality exploited here (e.g., through the planting of diverse evergreen shade trees for household fruit consumption in shade-coffee plots, and seasonal cropping patterns in “milpa” fields). Understanding historical and contemporary land use dynamics is an important next step in engaging this approach more reflexively.

## 5. Conclusions

The results presented here are significant in piloting an open-access, accurate, and relatively gap-free approach for mapping shade-grown coffee and adjacent landscape heterogeneity at a regional scale. The development of this approach is significant given limitations in land-cover classification accuracy, and challenges with cloud cover that characterized previous attempts to map shade-grown coffee landscapes using Landsat data. Accurately mapping shade-grown coffee and associated landscape contexts provides initial traction in developing understandings of how these land covers, and corresponding land uses, shift in relation to one another, both historically and at present. Future work can extend this study by exploring the value of this approach in other parts of the coffee-producing tropics, and in regions characterized by a greater diversity of coffee production strategies. Future work can also build on this study by exploring how a more diverse range of multi-temporal datasets can be leveraged to improve classification approaches in complex forested and agroforested landscapes.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-4292/10/6/952/s1>, Table S1: Metadata for all utilized Landsat 8 images, Table S2: Kappa Index Values for the Multi-Seasonal Spectral Model (S+TP) with Different Specifications of *mtry* and N, and Table S3: Error Matrices for All Remaining RF Model Specifications.

**Author Contributions:** Conceptualization, L.C.K.; Formal analysis, L.C.K.; Methodology, L.C.K., C.B., and L.P.; Writing—original draft, L.C.K.; Writing—review & editing, L.P. and C.B.

**Acknowledgments:** This material is based upon work supported by the National Science Foundation under Grant Number (BCS 1539795). The authors thank Matt Jones and Sharon Gourdji for access to data that facilitated this analysis, and two anonymous reviewers for their generous feedback. Access to codebases made openly accessible through Google Earth Engine and SERVIR-MEKONG greatly facilitated this research.

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

## References

1. Perfecto, I.; Rice, R.A.; Greenberg, R.; Van der Voort, M.E. Shade coffee: A disappearing refuge for biodiversity. *BioScience* **1996**, *46*, 598–608. [[CrossRef](#)]
2. Moguel, P.; Toledo, V.M. Biodiversity conservation in traditional coffee systems of Mexico. *Conserv. Biol.* **1999**, *13*, 11–21. [[CrossRef](#)]
3. Jha, S.; Bacon, C.M.; Philpott, S.M.; Ernesto Méndez, V.; Läderach, P.; Rice, R.A. Shade coffee: Update on a disappearing refuge for biodiversity. *BioScience* **2014**, *64*, 416–428. [[CrossRef](#)]
4. Getachew, T.; Zavaleta, E.; Shennan, C. Coffee landscapes as refugia for native woody biodiversity as forest loss continues in southwest Ethiopia. *Biol. Conserv.* **2014**, *169*, 384–391.
5. Cerdán, C.R.; Rebolledo, M.C.; Soto, G.; Rapidel, B.; Sinclair, F.L. Local knowledge of impacts of tree cover on ecosystem services in smallholder coffee production systems. *Agric. Syst.* **2012**, *110*, 119–130. [[CrossRef](#)]
6. Tschardtke, T.; Clough, Y.; Bhagwat, S.A.; Buchori, D.; Faust, H.; Hertel, D.; Hölscher, D.; Jührbandt, J.; Kessler, M.; Perfecto, I.; et al. Multifunctional shade-tree management in tropical agroforestry landscapes: A review. *J. Appl. Ecol.* **2011**, *48*, 619–629. [[CrossRef](#)]

7. Philpott, S.M.; Arendt, W.J.; Armbrrecht, I.; Bichier, P.; Diestch, T.V.; Gordon, C.; Greenberg, R.; Perfecto, I.; Reynoso-Santos, R.; Soto-Pinto, L.; et al. Biodiversity loss in Latin American coffee landscapes: Review of the evidence on ants, birds, and trees. *Conserv. Biol.* **2008**, *22*, 1093–1105. [[CrossRef](#)] [[PubMed](#)]
8. Imbach, P.; Fung, E.; Hannah, L.; Navarro-Racines, C.E.; Roubik, D.W.; Ricketts, T.H.; Harvey, C.A.; Donatti, C.I.; Läderach, P.; Locatelli, B.; et al. Coupling of pollination services and coffee suitability under climate change. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 10438–10442. [[CrossRef](#)] [[PubMed](#)]
9. Perfecto, I.; Vandermeer, J.; Mas, A.; Pinto, L.S. Biodiversity, yield, and shade coffee certification. *Ecol. Econ.* **2005**, *54*, 435–446. [[CrossRef](#)]
10. Langford, M.; Bell, W. Land cover mapping in a tropical hillsides environment: A case study in the Cauca region of Colombia. *Int. J. Remote Sens.* **1997**, *18*, 1289–1306. [[CrossRef](#)]
11. Cordero-Sancho, S.; Sader, S.A. Spectral analysis and classification accuracy of coffee crops using Landsat and a topographic-environmental model. *Int. J. Remote Sens.* **2007**, *28*, 1577–1593. [[CrossRef](#)]
12. Gomez, C.; Mangeas, M.; Petit, M.; Corbane, C.; Hamon, P.; Hamon, S.; De Kochko, A.; Le Pierres, D.; Poncet, V.; Despinoy, M. Use of high-resolution satellite imagery in an integrated model to predict the distribution of shade coffee tree hybrid zones. *Remote Sens. Environ.* **2010**, *114*, 2731–2744. [[CrossRef](#)]
13. Ortega-Huerta, M.A.; Komar, O.; Price, K.P.; Ventura, H.J. Mapping coffee plantations with Landsat imagery: An example from El Salvador. *Int. J. Remote Sens.* **2012**, *33*, 220–242. [[CrossRef](#)]
14. Carvalho, L.M.T.; Clevers, J.G.P.W.; Skidmore, A.K.; de Jong, S.M. Selection of imagery data and classifiers for mapping Brazilian semi-deciduous Atlantic forests. *Int. J. Appl. Earth Obs. Geoinf.* **2004**, *5*, 173–186. [[CrossRef](#)]
15. Moreira, M.A.; Adami, M.; Rudorff, B.F.T. Análise espectral e temporal da cultura do café em imagens Landsat. *Pesquisa Agropecuária Brasileira* **2004**, *39*, 223–231. [[CrossRef](#)]
16. Moreira, M.A.; Rudorff, B.F.T.; Barros, M.A.; De Faria, V.G.C.; Adami, M. Geotechnologies to map coffee fields in the states of Minas Gerais and Sao Paulo. *Eng. Agric.* **2010**, *30*, 1123–1135.
17. Chemura, A.; Mutanga, O.; Dube, T. Integrating age in the detection and mapping of incongruous patches in coffee (*Coffea arabica*) plantations using multi-temporal Landsat 8 NDVI anomalies. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 1–13. [[CrossRef](#)]
18. Mukashema, A.; Veldkamp, A.; Vrieling, A. Automated high resolution mapping of coffee in Rwanda using an expert Bayesian network. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *33*, 331–340. [[CrossRef](#)]
19. Hansen, M.C.; Loveland, T.R. A review of large area monitoring of land cover change using Landsat data. *Remote Sens. Environ.* **2012**, *122*, 66–74. [[CrossRef](#)]
20. Luvall, J.C.; Lieberman, D.; Lieberman, M.; Hartshorn, G.S.; Peralta, R. Estimation of tropical forest canopy temperatures, thermal response numbers, and evapotranspiration using aircraft-based thermal sensor. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 1393–1401.
21. Zhu, X.; Liu, D. Accurate mapping of forest types using dense seasonal Landsat time series. *ISPRS J. Photogramm. Remote Sens.* **2014**, *96*, 1–11. [[CrossRef](#)]
22. Gao, T.; Zhu, J.; Zheng, X.; Shang, G.; Huang, L.; Wu, S. Mapping spatial distribution of larch plantations from multi-seasonal Landsat-8 OLI imagery and multi-scale textures using random forests. *Remote Sens.* **2015**, *7*, 1702–1720. [[CrossRef](#)]
23. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine. *Remote Sens.* **2016**, *9*, 1065. [[CrossRef](#)]
24. Chrysafis, I.; Mallinis, G.; Gitas, I.; Tsakiri-Strati, M. Estimating Mediterranean forest parameters using multi seasonal Landsat 8 OLI imagery and an ensemble learning method. *Remote Sens. Environ.* **2017**, *199*, 154–166. [[CrossRef](#)]
25. Aguilar, R.; Zurita-Milla, R.; Izquierdo-Verdiguier, E.; de By, R.A. A cloud-based multi-temporal ensemble classifier to map smallholder farming systems. *Remote Sens.* **2018**, *10*, 729.
26. Misra, G.; Kumar, A.; Patel, N.R.; Zurita-Milla, R. Mapping a Specific Crop-A Temporal Approach for Sugarcane Ratoon. *J. Indian Soc. Remote Sens.* **2014**, *42*, 325–334. [[CrossRef](#)]
27. Hu, Q.; Wu, W.; Song, Q.; Lu, M.; Chen, D.; Yu, Q.; Tang, H. How do temporal and spectral features matter in crop classification in Heilongjiang Province, China? *J. Integr. Agric.* **2017**, *16*, 324–336. [[CrossRef](#)]



28. Rodriguez-Galiano, V.; Chica-Olmo, M. Land cover change analysis of a Mediterranean area in Spain using different sources of data: Multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. *Appl. Geogr.* **2012**, *35*, 208–218. [[CrossRef](#)]
29. Rodriguez-Galiano, V.R.; Chica-Olmo, M.; Abarca-Hernandez, F.; Atkinson, P.M.; Jeganathan, C. Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-season texture. *Remote Sens. Environ.* **2012**, *121*, 93–107. [[CrossRef](#)]
30. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [[CrossRef](#)]
31. Crawford, C.J.; Manson, S.M.; Bauer, M.E.; Hall, D.K. Multitemporal snow cover mapping in mountainous terrain for Landsat climate data record development. *Remote Sens. Environ.* **2013**, *135*, 224–233. [[CrossRef](#)]
32. 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](#)]
33. 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](#)] [[PubMed](#)]
34. Kelley, L.C.; Evans, S.G.; Potts, M.D. Richer histories for more relevant policies: 42 years of tree cover loss and gain in Southeast Sulawesi, Indonesia. *Glob. Chang. Biol.* **2016**. [[CrossRef](#)] [[PubMed](#)]
35. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
36. Gaertner, J.; Genovese, V.B.; Potter, C.; Sewake, K.; Manoukis, N.C. Vegetation classification of Coffea on Hawaii Island using WorldView-2 satellite imagery. *J. Appl. Remote Sens.* **2017**, *11*, 046005. [[CrossRef](#)]
37. Cooley, S.W.; Smith, L.C.; Stepan, L.; Mascaro, J. Tracking Dynamic Northern Surface Water Changes with High-Frequency Planet CubeSat Imagery. *Remote Sens.* **2017**, *9*, 1306. [[CrossRef](#)]
38. Perfecto, I.; Vandermeer, J.; Wright, A. *Nature's Matrix: Linking Agriculture, Conservation and Food Sovereignty*, 1st ed.; Earthscan: London, UK, 2009.
39. Cerda, F.; Flores, N.; Toval Hernández, K.I. Diversidad y Usos de la Fauna Silvestre en el Parque Ecológico Municipal Cerro Santa Gallo, Telpaneca, Condega, Nicaragua. Ph.D. Thesis, Universidad Nacional Agraria, Managua, Nicaragua, 2009.
40. Kauth, R.J.; Thomas, G.S. The Tasseled Cap—A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, IN, USA, 29 June–1 July 1976; pp. 4b41–4b51.
41. Huang, C.; Wylie, B.; Yang, L.; Homer, C.; Zylstra, G. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance. *Int. J. Remote Sens.* **2002**, *23*, 1741–1748. [[CrossRef](#)]
42. Yarbrough, L.D.; Easson, G.; Kuzmaul, J.S. Proposed workflow for improved Kauth–Thomas transform derivations. *Remote Sens. Environ.* **2012**, *124*, 810–818. [[CrossRef](#)]
43. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
44. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [[CrossRef](#)]
45. Taylor, B.W. An outline of the vegetation of Nicaragua. *J. Ecol.* **1963**, *51*, 27–54. [[CrossRef](#)]
46. INIFOM Municipios Instituto Nicaragüense de Fomento Municipal. Available online: <http://www.inifom.gob.ni/municipios/municipios.html> (accessed on 10 July 2015).
47. Gourdj, S.; Laderach, P.; Martinez Valle, A.; Zelaya Martinez, C.; Lobell, D. Historical climate trends, deforestation, and maize and bean yields in Nicaragua. *Agric. For. Meteorol.* **2015**, *200*, 270–281. [[CrossRef](#)]
48. USGS (United States Geological Survey). Landsat Collection 1 Level 1 Product Definition. Version 1.0. United States Geological Survey, Department of Interior 2017, Version 1.0. Available online: [https://landsat.usgs.gov/sites/default/files/documents/LSDS-1656\\_Landsat\\_Level-1\\_Product\\_Collection\\_Definition.pdf](https://landsat.usgs.gov/sites/default/files/documents/LSDS-1656_Landsat_Level-1_Product_Collection_Definition.pdf) (accessed on 1 May 2018).
49. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903. [[CrossRef](#)]
50. Chander, G.; Markham, B.L.; Barsi, J.A. Revised Landsat-5 thematic mapper radiometric calibration. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 490–494. [[CrossRef](#)]

51. Song, C.; Woodcock, C.E.; Seto, K.C.; Lenney, M.P.; Macomber, S.A. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sens. Environ.* **2001**, *75*, 230–244. [[CrossRef](#)]
52. Baig, M.H.A.; Zhang, L.; Shuai, T.; Tong, Q. Derivation of a tasseled cap transformation based on Landsat 8 at-satellite reflectance. *Remote Sens. Lett.* **2014**, *5*, 423–431. [[CrossRef](#)]
53. Liu, Q.; Liu, G.; Huang, C.; Xie, C. Comparison of tasseled cap transformations based on the selective bands of Landsat 8 OLI TOA reflectance images. *Int. J. Remote Sens.* **2015**, *36*, 417–441. [[CrossRef](#)]
54. Tasseled Cap Transformation of Landsat 8 Surface Reflectance. Available online: [https://groups.google.com/forum/?utm\\_source=digest&utm\\_medium=email#!searchin/google-earth-engine-developers/TOA%20cap%20transformation%20%7Csort:date/google-earth-engine-developers/sj\\_IY7ZXipw/MZ8QsKK2BwAJ](https://groups.google.com/forum/?utm_source=digest&utm_medium=email#!searchin/google-earth-engine-developers/TOA%20cap%20transformation%20%7Csort:date/google-earth-engine-developers/sj_IY7ZXipw/MZ8QsKK2BwAJ) (accessed on 28 March 2018).
55. Dong, J.; Xiao, X.; Menarguez, M.A.; Zhang, G.; Qin, Y.; Thau, D.; Biradar, C.; Moore, B., III. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens. Environ.* **2016**, *185*, 142–154. [[CrossRef](#)] [[PubMed](#)]
56. Orengo, H.A.; Petrie, C.A. Large-scale, multi-temporal remote sensing of paleo-river networks: A case study from Northwest India and its implications for the Indus Civilisation. *Remote Sens.* **2017**, *9*, 735. [[CrossRef](#)]
57. Landsat Composite Methods and Earth Engine Algorithms. Available online: <https://docs.google.com/document/d/14fNqbm8-oguRylapdhmif-tLckYI-87RbKiPfc5N0/edit#heading=h.c855pzu7ubf3> (accessed on 30 May 2018).
58. Hansen, M.C.; Stehman, S.V.; Potapov, P.V.; Loveland, T.R.; Townshend, J.R.G.; DeFries, R.S.; Pittman, K.W.; Arunarwati, B.; Stolle, F.; Steininger, M.K.; et al. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9439–9444. [[CrossRef](#)] [[PubMed](#)]
59. Potapov, P.V.; Turubanova, S.A.; Hansen, M.C.; Adusei, B.; Broich, M.; Altstatt, A.; Mane, L.; Justice, C.O. Quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM + data. *Remote Sens. Environ.* **2012**, *122*, 106–116. [[CrossRef](#)]
60. Rodriguez-Galiano, V.R.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier to land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104. [[CrossRef](#)]
61. Crist, E.P.; Cicone, R.C. A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap. *IEEE Trans. Geosci. Remote Sens.* **1984**, *3*, 256–263. [[CrossRef](#)]
62. Lobser, S.E.; Cohen, W.B. MODIS Tasseled Cap: Land Cover Characteristics Expressed through Transformed MODIS Data. *Int. J. Remote Sens.* **2007**, *28*, 5079–5101. [[CrossRef](#)]
63. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*. [[CrossRef](#)]
64. Time Series Analysis. Available online: <http://earthenginesummit2016.earthoutreach.org/training-materials> (accessed on 28 March 2018).
65. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [[CrossRef](#)] [[PubMed](#)]
66. Ding, Y.; Zhao, K.; Zheng, X.; Jiang, T. Temporal dynamics of spatial heterogeneity over cropland quantified by time-series NDVI, near infrared and red reflectance of Landsat 8 OLI imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *30*, 139–145. [[CrossRef](#)]
67. Glenn, E.P.; Huete, A.R.; Nagler, P.L.; Nelson, S.G. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors* **2008**, *8*, 2136–2160. [[CrossRef](#)] [[PubMed](#)]
68. Dalponte, M.; Orka, H.O.; Gobakken, T.; Gianelle, D.; Naesset, E. Tree species classification in boreal forests with hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2632–2645. [[CrossRef](#)]
69. Jin, H.; Stehman, S.V.; Mountrakis, G. Assessing the impact of training sample selection on accuracy of an urban classification: A case study in Denver, Colorado. *Int. J. Remote Sens.* **2014**, *35*, 2067–2081.
70. Hua, J.P.; Xiong, Z.C.; Lowey, J.; Suh, E.; Dougherty, E.R. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* **2005**, *8*, 1509–1515. [[CrossRef](#)] [[PubMed](#)]
71. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.

72. Google Earth Engine API. Available online: [https://developers.google.com/earth-engine/api\\_docs](https://developers.google.com/earth-engine/api_docs) (accessed on 28 May 2018).
73. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 56–66. [[CrossRef](#)]
74. Lee, J.S.H.; Wich, S.; Widayati, A.; Koh, L.P. Detecting industrial oil palm plantations on Landsat images with Google Earth Engine. *Remote Sens. Appl.: Soc. Environ.* **2016**, *4*, 219–224. [[CrossRef](#)]
75. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [[CrossRef](#)]
76. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2009.
77. Sexton, J.O.; Song, X.P.; Feng, M.; Noojipady, P.; Anand, A.; Huang, C.; Kim, D.H.; Collins, K.M.; Channan, S.; DiMiceli, C.; et al. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS Vegetation Continuous Fields with lidar-based estimates of error. *Int. J. Digit. Earth* **2013**. [[CrossRef](#)]
78. Läderach, P.; Ramirez-Villegas, J.; Navarro-Racines, C.; Zelaya, C.; Martinez-Valle, A.; Jarvis, A. Climate change adaptation of coffee production in space and time. *Clim. Chang.* **2017**, *141*, 47–62. [[CrossRef](#)]
79. Maurer, E.P.; Roby, N.; Stewart-Frey, I.T.; Bacon, C.M. Projected twenty-first-century changes in the Central American mid-summer drought using statistically downscaled climate projections. *Reg. Environ. Chang.* **2017**, *17*, 2421–2432. [[CrossRef](#)]
80. Millard, K.; Richardson, M. On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sens.* **2015**, *7*, 8489–8515. [[CrossRef](#)]
81. Bandeira, F.P.; Martorell, C.; Meave, J.A.; Caballero, J. The role of rustic coffee plantations in the conservation of wild tree diversity in the Chinantec region of Mexico. *Biodivers. Conserv.* **2005**, *14*, 1225–1240. [[CrossRef](#)]
82. Soto-Pinto, L.; Romero-Alvarado, Y.; Caballero-Nieto, J.; Segura Warnholtz, G. Woody plant diversity and structure of shade-grown-coffee plantations in Northern Chiapas, Mexico. *Rev. Biol. Trop.* **2001**, *49*, 977–987. [[PubMed](#)]
83. Soto-Pinto, L.; Villalvazo-López, V.; Jiménez-Ferrer, G.; Ramírez-Marcial, N.; Montoya, G.; Sinclair, F.L. The role of local knowledge in determining shade composition of multistrata coffee systems in Chiapas, Mexico. *Biodivers. Conserv.* **2007**, *16*, 419–436. [[CrossRef](#)]



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).