land

MDPI

Article

Towards a Reproducible LULC Hierarchical Class Legend for Use in the Southwest of Pará State, Brazil: A Comparison with Remote Sensing Data-Driven Hierarchies

Mariane S. Reis [1,*], Maria Isabel S. Escada [2], Luciano V. Dutra [2], Sidnei J. S. Sant'Anna [2] and Nathan D. Vogt [3]

1 Earth System Science Center, National Institute for Space Research–INPE, São José dos Campos 515–12245–970, SP, Brazil

2 Image Processing Department, National Institute for Space Research–INPE, São José dos Campos 515–12245–970, SP, Brazil; isabel@dpi.inpe.br (M.I.S.E.); dutra@dpi.inpe.br (L.V.D.); sidnei@dpi.inpe.br (S.J.S.S.)

3 University of Paraíba Valley, São José dos Campos 12244-000, SP, Brazil; ndvogt@gmail.com

* Correspondence: reis@dpi.inpe.br; Tel.: +55-12-3208-6781

Received: 12 April 2018; Accepted: 11 May 2018; Published: 18 May 2018

check for updates

Abstract: Land Use and Land Cover (LULC) classes defined by subjective criteria can diminish the significance of a study, hindering the reproducibility and the comparison of results with other studies. Having a standard legend for a given study area and objective could benefit a group of researchers focused on long-term or multidisciplinary studies in a given area, in the sense that they would be able to maintain class definition among different works, done by different teams. To allow for reproducibility, it is important that the classes in this legend are described using quantifiable elements of land cover, which can be measured on the ground, as is recommended by Land Cover Meta Language (LCML). The present study aims to propose LCML formalized hierarchical legends for LULC classes, focusing on the southwest of Pará state, within the Brazilian Amazon. In order to illustrate the potential of these legends, a secondary objective of the current study is to analyze classification results using legends derived from a particular Remote Sensing dataset and compare these results with the classification obtained using the LCML hierarchical legend proposed. To perform this analysis, firstly, we proposed a conceptual class model based on existing classification systems for the upland Brazilian Amazon Biome. From this model, 16 LULC classes were described in LCML, using quantifiable and easily recognizable physiognomic characteristics of land cover classes measured on the lower Tapajós river, in Pará state. These classes were grouped into legends with different levels of detail (number of classes), based on our model or on the image and clustering algorithms. All legends were used in supervised classification of a Landsat5/TM image. Results indicate that it is necessary to incorporate multi-temporal knowledge for class definition as well as the proposed thresholds (height and cover proportion of soil, litter, herbaceous vegetation, shrubs, and trees) in order to properly describe classes. However, the thresholds are useful to delimit classes that happen in a successive way. Classification results revealed that classes formed by the same elements of land cover with similar thresholds present high confusion. Additionally, classifications obtained using legends based on the class separability in a given Remote Sensing image tend to be more accurate but not always useful because they can hide or mix important classes. It was observed that the more generalized the legend (those with few details and number of classes), the more accurate the classifications results are for all types of legends.

Keywords: land use and land cover(LULC); Brazilian Amazon; remote sensing; class definition; field data collection; land cover meta language(LCML); classification system

1. Introduction

Human-induced changes in Land Use and Land Cover (LULC) affect the biophysics, biogeochemistry, and biogeography of the terrestrial surface [1]. Changes in land systems and resulting land cover changes go beyond local alterations, affecting many environmental processes, such as primary production, the water cycle, biogeochemical cycles, climate system and biodiversity [2]. Therefore, the understanding of the distribution and dynamics of LULC has been recognized as an important scientific field. Remote Sensing imagery has been an important source of information about LULC over time, in varying scales [3–14]. As a result, many regional and global LULC maps have been produced and are available [13–22]. However, these maps have been created with different objectives, class definition and resolution, producing different maps useful for specific studies, rendering the harmonization and integration of the selected ones very challenging [23–25]. In this study, we will focus on problems associated with class definition, mainly within the Brazilian Amazon.

There are three commonly used ways to define the classes to be considered in a study. The first one is simply establishing the classes that would be sufficient for the study objective. The second one is to consider class separability based on the available Remote Sensing imagery potential and classification techniques. The third way is to derive classes from a classification system or to adopt a standardized legend. A classification system is a logical framework that holds the names of the classes, the criteria used to distinguish them and the relationships between classes [26]. A legend, however, is the set of classes used for a specific purpose, which may or may not be derived from a classification system [26,27].

When establishing classes considering only the study objective, it is important to highlight that studies conducted with different objectives probably will present a distinct definition of classes, data collection protocol, and analytical techniques. For instance, there is a fundamental difference between data used for local and global studies. According to McConnell and Moran [26], global researchers are often interested in the representation of biophysical aspects and temporal patterns linked to climate change (e.g., carbon dynamic and sequestration rates), usually focusing on major land cover changes across large spatial extents. Local researchers, instead, focus on the characterization of human-induced modifications to land cover at finer scales. Even considering similar scales and objectives, the straightforward comparison between results is sometimes impossible because of slight to moderate differences in class definition [28]. Besides precluding the straightforward comparison between data, different definitions of classes can also influence the problem being analyzed. According to Hansen and Loveland [29], most large area, medium spatial resolution, land cover monitoring products are focused on forest cover change mapping, with special attention to changes occurring in tropical forests. 'Forest', however, can be a challenging concept to define and the lack of consensus on this concept may have serious consequences in conservation, development, climate, livelihood, biodiversity and ecosystem services [30,31]. Ref. Chazdon et al. [32] point out that the simple change in the definition of the term forest can alter management policies or inventoried forested areas. For instance, Refs. Putz and Redford [30], Chazdon et al. [32] and Morales-Barquero et al. [31] cite examples in their works in which an area nearly devoid of trees would still be considered forested.

The definition of classes considering the data used can solve some accuracy problems since the image classification is tailored to present accurate maps, but many types of inconveniences may arise. The most obvious one is that studies with similar study area and objectives can be carried out with different legends because they are based on different sets of data, or they use the same type of data in distinct times or because of differences in field data collection, processing methodologies, or the expected quality of the final mapping. It is the case of the works carried out by [6,11,12,33–40] within an area in the Lower Tapajós region, Pará state, within the Brazilian Amazon. The resultant maps are rarely directly comparable and their usefulness for studies with other types of data is diminished.

The use of a unique classification system or standardized legend has the advantage of allowing direct comparison among class sets [41]. On the other hand, such classification system would not be applicable to every study objective or area of interest [42]. While none of the classification systems previously proposed has been fully and internationally accepted [42], having a standard legend could benefit a group of researchers focused on long-term or multidisciplinary studies in a given area, in the sense that they would be able to maintain long-term class definition in different works, done by different teams.

To our experience and based on the above, a useful standard legend should be: (1) easily understandable; (2) reproducible; (3) data-independent; (4) constructed around a common objective for a specific geographical area; (5) open to the inclusion of necessary not predicted classes and (6) hierarchical, to allow for studies in varied scales and, to a certain degree, to accommodate different types of data. In order to satisfy the three first criteria, one solution is describing classes using quantifiable elements of land cover that can be measured on the ground. This is the main purpose of the Land Cover Classification System (LCCS) [42], developed by the Food and Agriculture Organization (FAO) and the United Nations Environment Programme (UNEP). LCCS was proposed as a flexible system, in which the use of common land cover classifiers can be used to standardize how these classes are described and not the categories (classes). Since the first concepts endorsed in 1996 [42], this classification system has evolved to the Land Cover Meta-Language (LCML). The LCML is a Unified Modeling Language (UML) meta-model that allows for describing land cover classes using their quantifiable physiognomic aspects. This language was used to describe the classes from some well-known land cover classification systems [24], as well as to define the legends employed in some studies across the globe [43–46] and to define classes used in FAO's projects. In addition, LCML has been recognized as an instrument for harmonization, with respect to meso-level integration of global and local data [26] and to harmonize land cover datasets [47].

Among the previously cited legends and classification systems described in LCML [24,43–46], only the one used in Terraclass [13,46]—Land Use and Land Cover mapping of Deforested Areas in the Legal Amazon Project—is specific for studies in the Brazilian Amazon. However, the class definition is mainly based on used data, as evidenced by the existence of the class 'Mosaic of Uses', defined as "*land cover units that, due to the spatial resolution of the satellite images, cannot be broken down further into specific components*". Moreover, this legend is not hierarchical and LCML description was done after the legend definition in a non-systematic way.

The objective of the present work is to propose LCML formalized hierarchical legends for LULC classes focusing on the southwest of Pará state, within the Brazilian Amazon, which is an intensively studied area in relation to land use and land cover. Vegetated classes were defined considering data collected in the field and a proposed conceptual model based on commonly used legends and classification systems [41,42,48]. These classes were also described in LCML, in order to allow the easy and objective understanding of their meaning. This step includes the description of classes using quantifiable characteristics of the vegetation of these classes, allowing for the objective identification of the defined classes in the field. The defined classes were further organized in hierarchical legends to encompass studies in different scales and types of Remote Sensing data while being independent of data and classification methodologies.

A secondary aim of this work is to analyze classification results of a particular Remote Sensing dataset using legends constructed considering the separability of classes on this data, and compare these results with classification results obtained with the LCML hierarchical legend proposed in the current work. Therefore, this work contributes to the scientific field by offering an Amazon-specific framework from which researchers could derive their legends and by presenting fully described LULC classes that may help researchers to recognize these classes in the field. This framework can be expanded and new classes can be created to attend other studies. Additionally, we also present tools to construct hierarchical legends based on data and a concise analysis of their usability. This work is divided as follows. Firstly, the studied area is illustrated in Section 2, followed by Section 3, in which the data collected in the field, Remote Sensing image and collected land use and land cover samples are presented. The methods used to define the proposed class system and LCML LULC classes, as well as the image classification steps, are presented in Section 4. The resulting products are presented in

Section 5. The applicability of these products is discussed in Section 6 while the final considerations, recommendations, and future perspectives are presented in Section 7.

2. Study Area

The studied area in this work corresponds to the lower Tapajós basin region, Pará State, Brazilian Amazon. Figure 1 shows this area, in relation to its political and natural limits. The region presents a humid tropical climate, with average annual rainfall around 1820 mm and higher precipitations occurring from January to May. The annual average temperature is around 25 °C, varying less than 5 °C throughout the year. The region is inserted into two morpho-structural units, the Planalto Rebaixado da Amazônia , with an average altitude of around 100 m above sea level, and the Planalto Tapajós–Xingu , where the altitude ranges between 120 and 170 m. The original vegetation is of humid tropical rainforest, with the presence of woody lianas, palm trees and epiphytes [49].

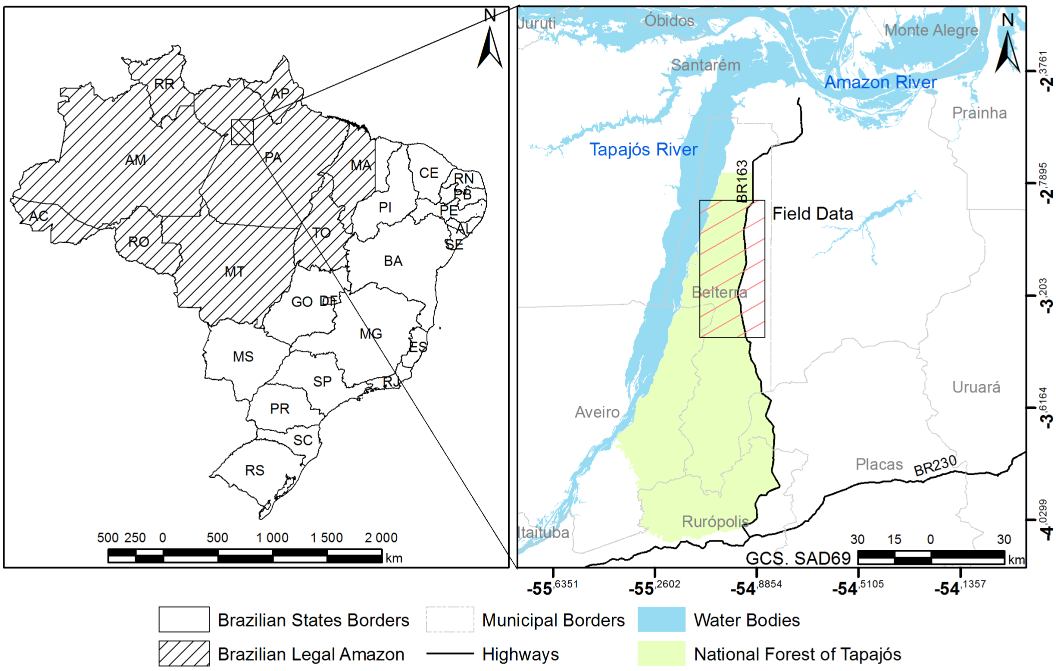


Figure 1. Study area.

Although the region of Tapajós river was colonized by Europeans in the 17th century, the more salient modification of land cover began in 1970s, with the construction of the Cuiabá–Santarém highway (BR 163) and subsequent governmental occupation projects. As a result of the occupation, there are areas of secondary vegetation in diverse stages of development, pasture and agriculture (mechanized and for subsistence) within a forest matrix. The National Forest of Tapajós, a major forest reserve created by the Act number 73,684 of 19 February 1974, with approximately 600,000 ha, located on the east side of the Tapajós river and west side of Cuiabá–Santarém highway is also an important element in the study area. Besides agricultural activities cited, there are reports of logging, mining, hunting, fishing and inadequate use of fire. This is also an attractive area for researchers, due to the available infrastructure, like military buildings and the research camp built through the Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA). Many types of studies were conducted in the area through the time and a large amount of data has been produced from the area over the past 20 plus years [50–55].

3. Materials

3.1. Field Data

Field data (ground truth) were collected along km 60 to 120 of Cuiabá–Santarém highway (BR-163) and adjacent areas, as illustrated by the hatched rectangle in the Figure 1. The data, consisting of geo-located photographs and land cover descriptions, was collected in September of 2009 and 2010, August of 2013 and March of 2015. Subjective estimations of proportion and height (when applicable) of land cover elements were obtained for features in 2013 and 2015. These estimations were made in the field and mainly in a visual way, based on the observation of at least two analysts, considering the majority of the features identified in the field and on high-resolution Remote Sensing images (from SPOT 5, SPOT 6 and Rapideye sensors). For 2013 field data, trees were randomly measured using a TruPulse 200B Rangefinder (Laser Technology Inc., Centennial, CO, USA). Major land cover classes in the area were previously identified using TerraClass data for the years 2008 and 2010, PRODES data for the years 2012 and 2014, and also high-resolution images covering the study area (images from SPOT 5, SPOT 6 and Rapideye sensors). Considering the major land cover classes encountered, the amount of samples of each class with information about cover proportion and height of land cover elements, for each date, is presented in Table 1. For agricultural land cover classes, there is also the information about the type of culture and state of development. When possible, information about the history of the area was acquired from local inhabitants.

Table 1. Amount of samples with information about cover proportion and height of land cover elements.

Land Cover Class	2013 Field Data	2015 Field Data
Pasture	114	116
Annual Agriculture	41	69
Secondary Vegetation	30	23
Modified Forest	6	4
Mature Forest	5	0
Shifting Cultivation	3	0
Woody Plantation	1	0

3.2. Remote Sensing Image and Class Samples

A Landsat5/TM image was used. This image was acquired over the studied area on 29 June 2010, path/row 227/62, with 8 bits, 7 spectral channels (0.45–0.52 µm; 0.52–0.60 µm; 0.63–0.69 µm; 0.76–0.90 µm; 1.55–1.75 µm; 10.4–12.5 µm; and 2.08–2.35 µm) and 30 m of spatial resolution (120 m on the thermal band). This image was registered to an orthorectified Landsat5/TM image from GeoCover project and radiometrically corrected, using the method described in Green et al. [56]. This is a medium spatial resolution multispectral image, commonly used for land cover classification.

As explained in Section 4, a set of classes for the studied area was defined in this work and organized in a hierarchical way. Labeled samples of the classes in the lower and more detailed level of this set of classes were collected over the Landsat5/TM image, using information derived from the field work of 2010 (class description and photographs). Two sets of labeled samples were collected: the training set and the test set. The training set was used to perform feature selection of the bands of the image to be used, to define data-driven legends (Section 4.3) and in a later step to train the supervised classifier (Section 4.4). The test set was used to test the accuracy of the classifications. Note that, although information of the structure of the features in 2010 was not collected, the classes themselves were defined in a similar way as those in 2013 and 2015, so the class definition is maintained. In later steps, these samples were also grouped accordingly to each legend for classification and accuracy assessment.

A feature selection process was executed on bands 1–5 and 7 of Landsat5/TM data, using the minimum Jeffries–Matusita (JM) distance between all possible pairing of the initial classes set. JM distance is given by [57]:

$$JM_{ij} = \sqrt{2(1 - e^{-B_{ij}})}, \tag{1}$$

in which B_{ij} is the Bhattacharyya distance between the classes i and j. For Gaussian distributions, it is given by:

$$B_{ij} = \frac{1}{8}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \frac{1}{2} ln \frac{\left|\frac{\Sigma_i + \Sigma_j}{2}\right|}{\sqrt{|\Sigma_i|\,|\Sigma_j|}}. \tag{2}$$

μ_k is the mean of class k and Σ_k is the covariance matrix of class k. We selected the set with three bands, due to the intrinsic dimensionality of Landsat5/TM data.

Bands 3–5 were selected because those presented the highest minimum JM distance between all possible pairing of the initial classes set considering combinations of three bands (0.44), while the difference to the highest values for combinations with four bands (0.47) was small.

4. Methods

The proposed methodology can be settled out in the main steps illustrated in Figure 2. The methods used for each step are described in the following sections.

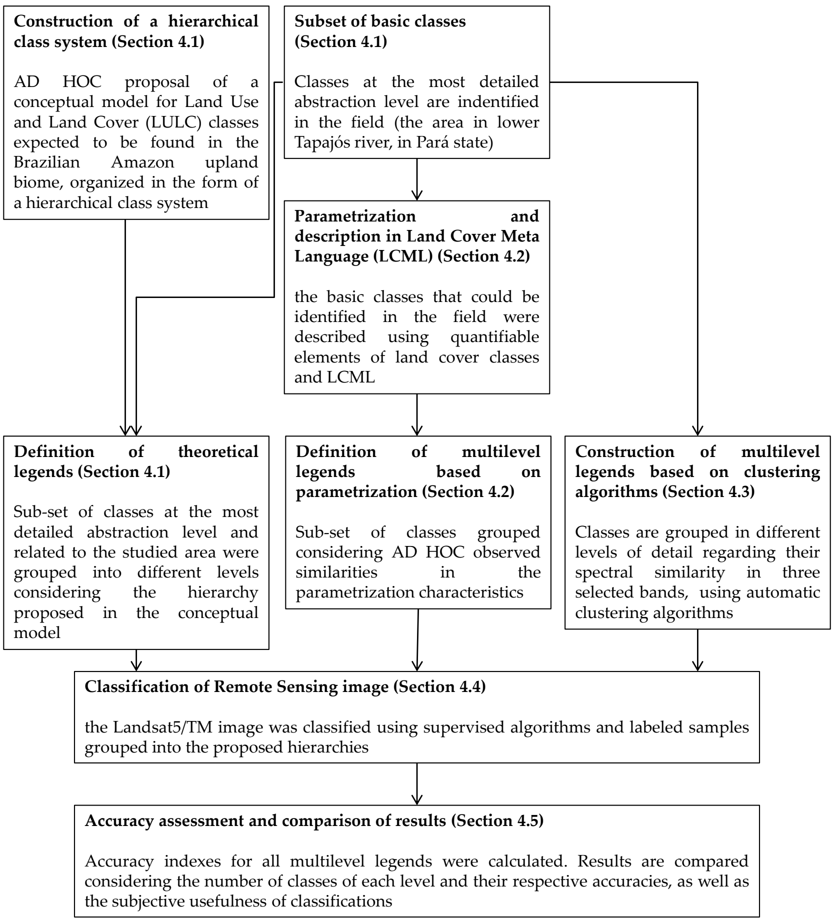


Figure 2. Main steps of the proposed methodology.

The methodology is composed of two main streams. The first one is the construction of a hierarchical conceptual model, in a top-down fashion, based on the experience of the authors and published literature about LULC classification (Section 4.1). The second stream is focused on finding hierarchical relationships of classes as defined by similarity on vegetation characteristics or spectral response on the Remote Sensing data (Sections 4.2 and 4.3), in a bottom-up approach.

4.1. Construction of a Hierarchical Class System for the Brazilian Amazon Upland Biome

The hierarchical class system was proposed based on an ad hoc analysis of well-known classification systems and legends used in Brazilian Amazon studies, the author's previous field and LULC classification experience as well as classes adopted by local inhabitants. Nomenclature and definitions are similar to those used by the LCCS version 2 dichotomous phase [42] and the legend from TerraClass project [13]. This system represents a conceptual model for LULC classes expected to be found in the Brazilian Amazon upland biome, from which researchers are able to derive LULC classes for related studies. This was the first step to identify classes in the studied area.

This system was organized in a hierarchical way. Therefore, there are super-classes that can be detailed in subclasses until what we refer to basic classes, which are those in the most detailed abstraction level. The basic classes identified in the studied area in the lower Tapajós region were singled out and used as the basis for the next steps in this work. Moreover, those that were identified in the subset of the Landsat5/TM image from 2010 (Section 3.2), ten LULC classes, were used as the basis of our proposed multilevel legends and are referred as L1(10) in the following sections. These classes were grouped to form the other two legends, considering the hierarchy proposed in the system. The resulting three-tier legend is referred in this work as theoretical legends.

4.2. Parametrization of LULC Classes and Translation to LCML

In order to allow the reproducibility of class definition during data collection, it is useful to define each class based on quantitative criteria. Following the proposal of LCML, LULC classes may be characterized by the physiognomy and structure of elements of the cover for each class. We used the mean height and the cover proportion of 5 elements (soil, litter, herbaceous vegetation, shrubs, and trees) to describe the basic classes proposed in the hierarchical class system. These elements are described as:

- **Soil:** exposed soil. This feature height equals to zero. Therefore, only the proportion in relation to other features in the same stratum is considered;
- **Litter:** organic debris. It occurs in the same stratum as 'Soil' and a height equal to zero was considered;
- **Herbaceous vegetation:** plants that have no persistent woody stem above ground. Height may vary and it can be in the same stratum than 'Soil' and 'Litter' or not, depending on height and structure. The height usually varies up to 2 m;
- **Shrubs:** plants that have persistent woody stem above ground and height smaller than 5 m;
- **Trees:** plants with elongated woody stem and higher than 5 m. In the presence of emergent trees, two strata may be composed of trees.

These elements may be distributed in different strata, so it is possible, and even probable, that the sum of the cover proportion of all elements surpasses 100%, as established in [42]. Mean height and cover proportion were selected among LCML classifiers because they are easily understandable parameters for non-specialist analysts.

The basic classes from the hierarchical class system, identified on the studied area were described by the definition of thresholds for the presented quantifiable elements of land cover, based on field data (Section 3.1), literature [58–60] and previous knowledge of the classes. This characterization was also done in LCML, for standardization purposes, using the FAO Land Cover Classification System 3 software, version 1.8.0 [61].

Moreover, the selected initial set of basic classes, L1(10), was grouped in order to form other two legends with different number of classes. These legends were derived by the grouping of classes with similar values for cover proportion and height of the elements of land cover.

4.3. Construction of Multilevel Legends Based on Automatic Clustering Algorithms

The initial set of classes was also organized into two hierarchical legends based on the Landsat5/TM Remote Sensing image, labeled samples of these classes and two automatic hierarchical clustering algorithms, the 'Single Minimum Link Dendrogram' (SMLD) and the 'Double Minimum Link Dendrogram' (SMLD). These algorithms were selected because they are very simple. More complex algorithms may be used in future approaches for comparison purposes.

The automatic hierarchical clustering algorithms, denominated 'Single Minimum Link Dendrogram' (SMLD), is based on the Single Link Hierarchical Clustering (SLHC) algorithm. SLHC was also used in Negri [8], in a similar way. Consider an image to be classified into N classes and a set of labeled pixels for each class. Consider that from these samples it is possible to determine a representative probability density function for each class and to calculate a dissimilarity measure between pairs of these functions. In the first iteration of the algorithm, the pair of classes that presents the lowest dissimilarity value is aggregated, generating a set of $N - 1$ classes. For the second iteration, dissimilarity values are recalculated considering this new set of classes and, again, the pair with the minimum value is aggregated. This process is iteratively made until some established stop criterion is achieved or only two classes are left. We opted for the latter criteria in this work. Results are then organized as a dendrogram, in which different set of classes are determined by a given threshold, herein the minimum dissimilarity value between the pairs of classes in that set. The second algorithm is called 'Double Minimum Link Dendrogram' (DMLD) and is very similar to SMLD. The main difference is that instead of aggregating one pair of classes at each iteration, in DMLD, the two pair of classes that presents the lowest dissimilarity values are merged. Note that if the lowest dissimilarity value was obtained by classes i and j, for instance, and the second lowest by j and l, although i and l may have a high dissimilarity value, the resulting aggregated class would be the union of i, j and l in DMLD. Since a Gaussian distribution can be assumed for optical data, the dissimilarity measure used in this work is the Jeffries–Matusita (JM) distance.

4.4. Classification of Remote Sensing Image

In order to illustrate the potential of the defined classes for land cover classification, we classified the Landsat 5/TM image described in Section 3.2 using supervised algorithms and a Monte Carlo approach, based on the one proposed on Reis et al. [62]. Bands 3, 4 and 5 of the Landsat 5/TM image were classified using all the proposed legends. Firstly, training samples were merged according to the legend and then used to train three supervised algorithms implemented in the R software: the Tree Decision-based algorithm J48 (RWeka package [63–65]), k-Nearest Neighbor (k-NN) (based on the spectral k nearest neighbors, as implemented in RWeka package [63–65]) and Maximum Likelihood (ML) based on a Gaussian distribution (Rasclass package [66]). For classification using the J48 algorithm, we adopted the default pruning confidence threshold (0.25), seed (1) and number of folds (3) and varied the minimum number of instances from 2 to 10, in steps of 1. For k-NN, we varied the number of neighbors from 1 to 20, in steps of 1.

Considering each legend and algorithm separately, 100 pixels of each class were randomly selected from the training set and used to train the classifier and obtain one classification. This process was repeated 100 times, creating 100 classifications for each legend and algorithm. Each pixel in the final classification was assigned the most frequent (mode) label in the 100 classifications. This process was done to supposedly obtain image classifications more independent from the sample collection and to avoid problems of spatial correlation [62].

We observed that similar results were obtained using all classifiers, although the best configuration for J48 or k-NN varied depending on the legend level. Since the ML classifier is robust to be applied in uni-modal data such as ours and obtained similar results as other classifiers without the need of parameter tuning, this classifier was selected for further analysis.

4.5. Accuracy Assessment and Comparison of Results

For accuracy assessment of the final classifications, 50 pixels for each class from the test set were randomly selected. With these samples, a confusion matrix was constructed, and the Overall Accuracy and Kappa Index [67,68] were calculated, as well as the Producer's and User's Accuracy for each class. This process was repeated 1000 times, resulting in 1000 confusion matrices for each classification and 1000 values for each index. The average value of the resulting 1000 indexes values was selected as the expected value for that index, and selected as the representative of that classification. Because of the varied test samples used, it is also possible to calculate the standard deviation of index values and indexes' average values were compared using unpaired *t*-test with 1% of significance level. Moreover, the average confusion matrix (each cell in this matrix is the mean value of the 1000 values of this cell in the matrices replication) was also analyzed.

5. Results

5.1. Proposed Hierarchical Class System for the Brazilian Amazon Biome

In this section, we present the proposed hierarchical class system for the Brazilian Amazon biome (see Section 4.1). Each defined class can be further detailed when needed, such as natural land cover classes proper from biomes other than Amazon. The system is illustrated in Figure 3, with some of the major LULC classes described. More detailed classes and the criteria for subdivision are explained as follows.

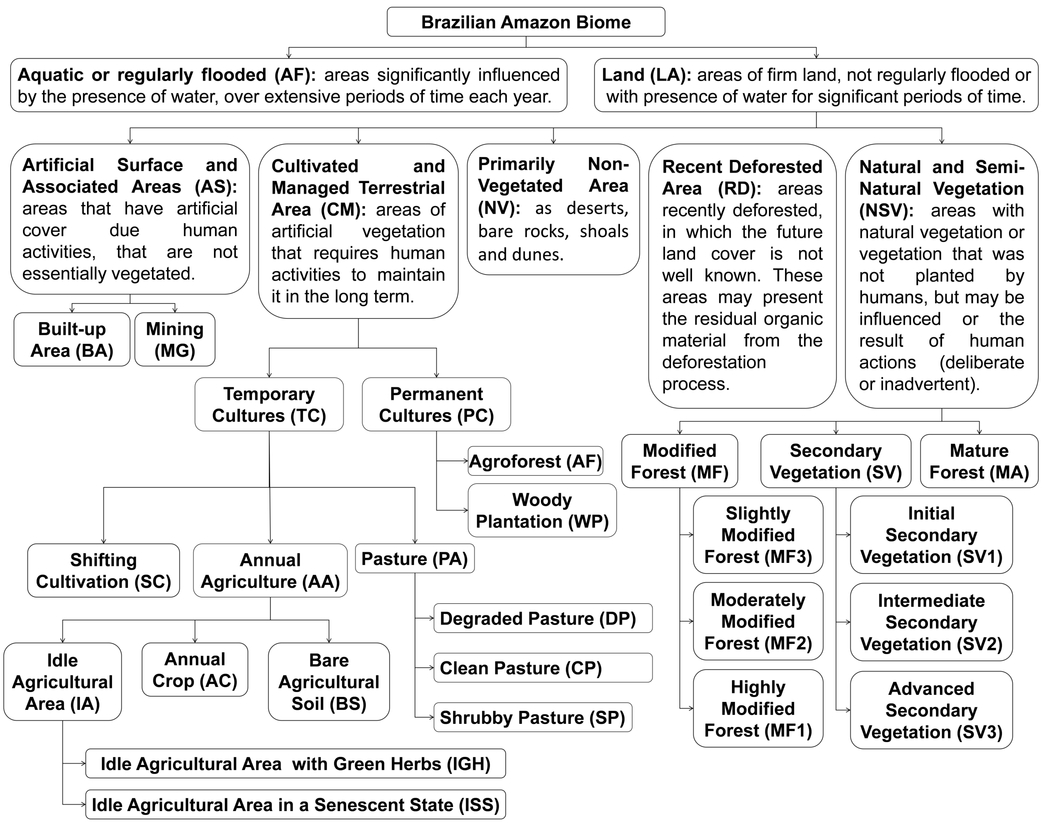


Figure 3. Proposed hierarchical class system for the Brazilian Amazon Biome.

Firstly, this work is focused only on the 'Land' classes. 'Land' category is further divided into five classes.

Some considerations must be made about the 'Natural and Semi-Natural Vegetation' class. Generally, it is very difficult to determine a single and well-accepted definition of 'forest', mainly in regard to what a 'pristine' forest is and, consequently, of alterations it has suffered. It is well accepted that a closed forest may be a class in which more than 70% of the area is covered by three crowns [30,42,46]. Given that it may be very difficult to define exactly what a pristine forest is in the Amazon biome, the term 'Mature Forest' was adopted to describe well structured, climax forests, with small to no evidence of alteration. If a forest is modified to the point of losing its original characteristics, it may be classified as 'Modified Forest'. This class covers both forests being degraded by factors as logging, fire, invasion by exotic species or other natural or anthropic activities, as well as forests being managed by silviculture practices employed in order to favor economically valuable species.

In relation to originally covered areas of forest, there is also any kind of semi-natural vegetation (not planted by humans, but resultant of human activities) that has grown in areas in which the original vegetation was completely removed. These areas were classified as 'Secondary Vegetation'. The classes 'Modified Forest' and 'Secondary Vegetation' may present similar characteristics in Remote Sensing data, so it is common to find works in which the distinction between these classes is not made. However, as stated in Putz and Redford [30], there are enough differences in structure, composition, dynamics and management suitable practices to evoke a differentiation between these classes. These classes can be further divided in regard to the intensity of modification or stage of development. Although these subclasses occur in continuum gradients, it is feasible to state one class for the beginning of the processes, one for the final stage and an intermediary one. For Secondary Vegetation, these stages may be summarized in 'Initial Secondary Vegetation', 'Intermediate Secondary Vegetation' and 'Advanced Secondary Vegetation'. Similar categorization was used in works like Lu et al. [3], Vieira et al. [69] and Salomão et al. [70]. In an analog way, Modified Forests also may be divided into 'Slightly Modified Forest', 'Moderately Modified Forest' and 'Highly Modified Forest'.

The 'Cultivated and Managed Terrestrial Area' class was divided into different subclasses considering the differences in management and structure. Firstly, they were divided into 'Permanent Culture' and 'Temporary Culture'. The first one encompasses classes that are dominated by shrubs and trees and does not need regular replanting. The second class refers to areas composed mainly of herbaceous cultures, which present some temporal sequencing depending behavior within one year. 'Permanent Culture' can be further divided into 'Woody Plantation' and 'Agroforest'. Agricultural areas of permanent, arboreal, cultures were included in 'Woody Plantation', as well as those not used for agricultural purposes but that are equally cultivated. 'Agroforest' represents areas in which husbandry and cropping practiced among natural, semi-natural or domesticated trees. The 'Woody Plantation' class can be further divided into two other classes, depending on use and species, which are 'Orchard' and 'Forest Plantation'. 'Orchard' represents areas of permanent culture of fruit trees, considering one or multiple species. 'Forest Plantation' encompasses areas with human-planted trees, be it for ecosystems' management purposes or future cutting, in any stage of development. These classes may be further divided as the need arises.

The 'Temporary Culture' class was subdivided into 'Annual Agriculture', 'Pasture' and 'Shifting Cultivation'. 'Annual Agriculture' are areas used for cultivating annual crops. This class may be divided into 'Annual Crop' (culture currently cultivated in the area, in any stage of development), 'Bare Agricultural Soil' and Idle Agricultural Area' (fallow annual agriculture areas, not being used for cultivation at the time of analysis). 'Idle Agricultural Area' may be split into other two classes, with respect to the state of the vegetation covering the area at the time being analyzed: 'Idle Agricultural Area in a Senescent State', in which haulm or invasive plants are expected to be present and, in their majority, senescent, and 'Idle Agricultural Areas with Green Herbs' in which invasive plants must be, in their majority, still green. Though 'Annual Agriculture' classes represent mainly mechanized, extensive areas, some small low productivity plantations are also expected. 'Shifting Cultivation' is practiced in small areas and has a mosaic and dynamic footprint consisting of cultivated fields and fallow lands. Cultivated fields are typically managed for two or three years before being left to fallow, and fallow lands are lands unused for three to fifteen years.

The 'Pasture' class represents areas mainly covered by Gramineae. Mainly because of management practices, the structure of these classes may vary over time. This class was divided into: 'Clean Pasture', 'Shrubby Pasture' and 'Degraded Pasture'. 'Clean Pasture' are pasture areas, mainly covered by Gramineae, in which shrubs invasive plants can be found in small quantities. 'Shrubby Pasture' are covered by Gramineae but with a significant presence of shrubs invasive plants. 'Degraded Pasture' are pasture areas with reduced productivity and/or with pasture grass with decreased vigor. Trees or palms may also be present in these classes. For further characterization, these classes can be also classified by the presence or not of original remaining trees or cultivated palms, as illustrated in Figure 4.

(**a**) CP with Palms (**b**) SP without Palms (**c**) OP with Palms

In which CP = Clean Pasture and SP = Shrubby Pasture.

Figure 4. Examples of pasture classes.

Regarding 'Artificial Surfaces and Associated Area', the main two classes found in Amazon were 'Built-up Area' and 'Mining'. 'Built-up Area' is defined as an area that was constructed by humans. For field characterization or further detailed studies, it may be of interest to classify this area in 'Impervious Surface' (roofs, paved streets or roads) and 'Non-Impervious Surface' (gardens, parks, non-paved roads, among others). It may also be interesting to typify these elements in the type of material, size, among other characteristics. 'Mining' are areas used for any type of mineral extraction.

Besides the classes defined, it is also possible to find areas in which the LULC class is being changed at the time, like a secondary vegetation area being cleaned for pasture purposes, for instance. This type of class may belong to the 'Transition area' class. Note that this class may be present as some type of 'Bare Soil', and that the defined 'Bare Agricultural Soil' fits into this description. Furthermore, given the frequency in which fires occur, any class, including those pertaining to 'Transition area', may be further classified as 'Burned'.

5.2. Parametrization of LULC Classes and LCML Translation

From the classes defined in the previous subsection, it was possible to identify 13 land cover classes in the lower Tapajós region: Bare Agricultural Soil (BS), Idle Agricultural Area in a Senescent State (ISS), Idle Agricultural Area with Green Herbs (IGH), Annual Crop (AC), Shifting Cultivation (SC), Clean Pasture (CP), Shrubby Pasture (SP), Woody Plantation (WP), Initial Secondary Vegetation (SV1), Intermediate Secondary Vegetation (SV2), Advanced Secondary Vegetation (SV3), Slightly Modified Forest (MF3) and Mature Forest (MA). From these and the previous knowledge of the region, it was also possible to extrapolate the characteristics of the other three classes: Highly Modified Forest (MF1), Moderately Modified Forest (MF2) and Recent Deforested Area (RD). The proportion and the mean height thresholds of land cover elements are defined in Table 2. In this table, the mean height of 'Soil' and 'Litter' are not specified because it was previously defined as zero. The mean height of 'Herbaceous' was not decisive to define any of the classes, so it was also suppressed from Table 2.

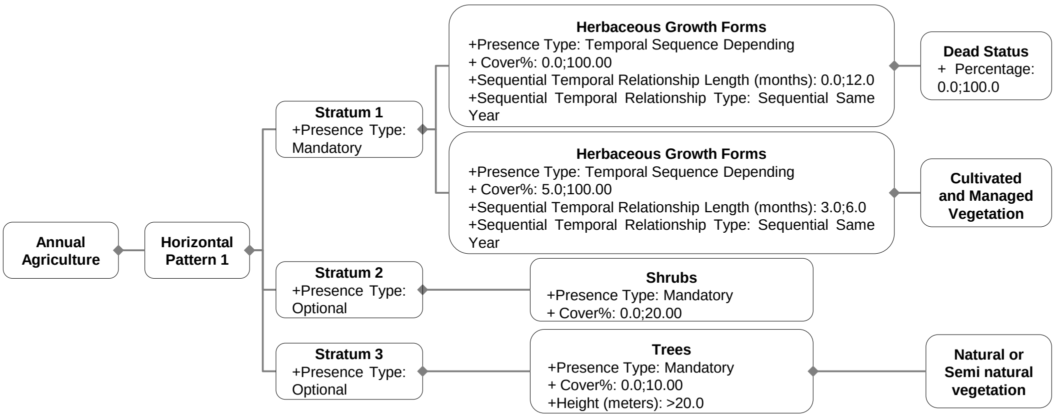
Table 2. Thresholds of land cover elements: cover proportion (in %) and mean height (in m).

Land Cover Classes	Herbaceous Min	Herbaceous Max	Shrubs Min	Shrubs Max	Trees Min	Trees Max	Litter Min	Litter Max	Soil Min	Soil Max	Shrubs Min (Mean Height)	Shrubs Max (Mean Height)	Trees Min (Mean Height)	Trees Max (Mean Height)
BS	0	10	0	5	0	10 [a]	0	20	80	100	*	*	*	*
ISS	0	10	0	20	0	10 [a]	20	100	0	80	*	*	*	*
IGH	20	100	0	20	0	10 [a]	0	20	0	80	*	*	*	*
AC	5	100	0	5	0	10 [a]	0	20	0	95	*	*	*	*
RD	0	15	0	15	30 [a,b]	100 [a,b]	0	70	0	70	*	*	*	*
SC	30	80	0	40	0	0	0	40	0	40	*	*	*	*
CP	60	100	0	15	0	10 [a]	*	*	0	40	*	*	*	*
SP	20	85	15	40	0	10 [a]	*	*	*	*	*	*	*	*
WP	*	*	*	*	60	100	*	*	*	*	*	*	*	*
SV1	*	*	40	100	0	10 [a]	*	*	0	40	0.5	5	*	* [a]
SV2	*	*	0	40	60	100 [a]	*	*	*	*	2	5	5	15 [a]
SV3	*	*	0	40	60	100 [a]	*	*	*	*	2	5	15	20 [a]
MF1	*	*	*	*	10	40	*	*	*	*	*	*	20	**
MF2	*	*	*	*	40	80	*	*	*	*	*	*	20	**
MF3	*	*	*	*	80	90	*	*	*	*	*	*	20	**
MA	*	*	*	*	90	100	*	*	*	*	*	*	20	**

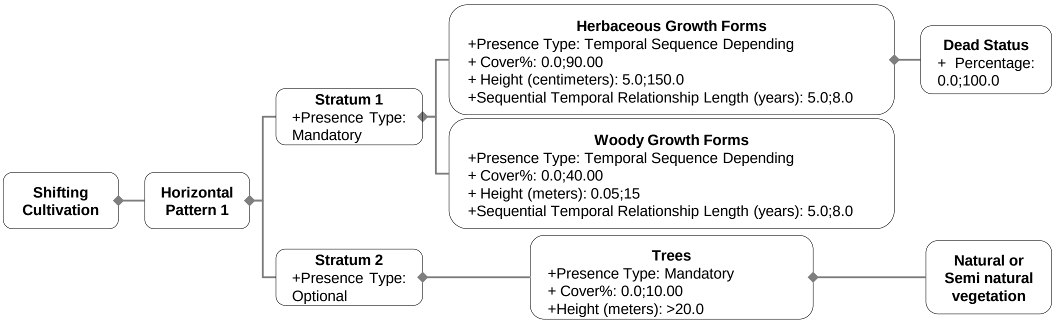
In which * denotes that the proportion of an element or height does not interfere with the definition of the respective class and ** that the element may reach its maximum height, but this height may vary. [a] 10% of remaining natural trees, higher than 20 m, allowed. [b] Cut trees or burnt standing ones. BS = Bare Agricultural Soil, ISS = Idle Agricultural Area in a Senescent State, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, RD = Recent Deforested Area, SC = Shifting Cultivation, CP = Clean Pasture, SP = Shrubby Pasture, WP = Woody Plantation, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF1 = Highly Modified Forest, MF2 = Moderately Modified Forest, MF3 = Slightly Modified Forest and MA = Mature Forest.

The thresholds presented in Table 2 are intended to help to identify classes in the field in a reproducible way, and not as an absolute descriptor. Furthermore, there are some classes in which these thresholds are superimposed, as happens with Shifting Cultivation and Shrubby Pasture, Woody plantation and some secondary vegetation/modified forest classes and also between Idle Agricultural Area with Green Herbs and Clean Pasture. Although these classes may not be fully distinguishable considering the defined thresholds, they present different spatial pattern, composition and/or temporal behaviors, which can be fully described in LCML. Some examples of LCML formalized classes are shown in Figures 5 and 6.

For instance, consider the classes Idle Agricultural Area with Green Herbs and Clean Pasture, exemplified in Figure 7. As can be seen, differentiating these classes may be a confusing task, even in the field. However, besides the difference in use, these classes also present different temporal behaviors. While one pasture area may present Clean Pasture and/or Shrubby Pasture along the year, one annual agricultural area would probably present the classes Annual Crop, Bare Soil and Idle Agricultural Area in succession. Therefore, this time sequencing component of Annual Agriculture may be useful to distinguish this class. Spectral variation in images from temporal series is already used in the TerraClass project to classify the corresponding 'Annual Agriculture' class. This difference in temporal behavior for agricultural classes was formalized in LCML, as illustrated in Figure 5a. Note that all agricultural classes (BS, IGH, ISS, and AC) were modeled as a single class, Annual Agriculture, in LCML.

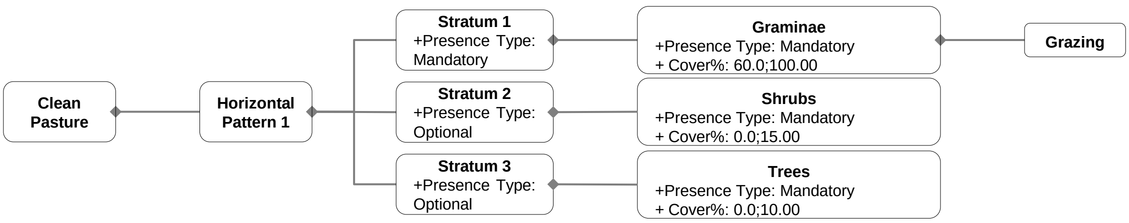


(**a**) Annual Agriculture

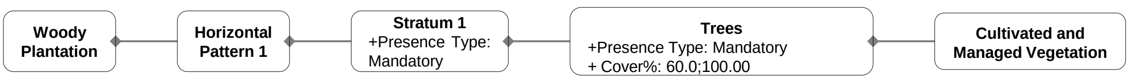


(**b**) Shifting Cultivation

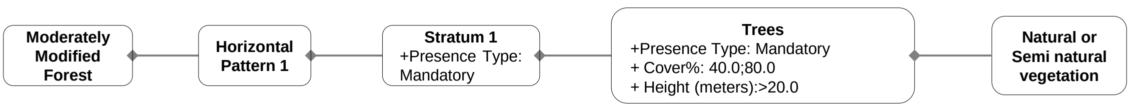
Figure 5. Examples of LCML(land cover meta language) modeling of classes with temporal sequencing behavior.



(**a**) Clean Pasture



(**b**) Woody Plantation



(**c**) Moderately Modified Forest

Figure 6. Other examples of LCML modeling.

(**a**) IGH

(**b**) CP

Figure 7. Examples of areas of Idle Agricultural Area with Green Herbs (IGH) and Clean Pasture (CP).

The classes Shifting Cultivation and Shrubby Pasture, although also presenting differences in temporal behavior, can be distinguished given their very different patterns. Shifting Cultivation areas present no infrastructure for cattle and the species cultivated are generally for consumption and somewhat diversified, while in Shrubby Pasture the shrubs are invasive plants. In LCML modeling, this can be differentiable by the inclusion of a 'grazing' classifier for Graminae and 'natural or semi-natural vegetation' for Shrubs, both in Shrubby Pasture. In Remote Sensing images, they may be distinguished by the format and size of features, since pastures are usually larger than areas of Shifting Cultivation.

The Woody Plantation (WP) class also presents some overlapping to the classes Intermediate Secondary Vegetation, Advanced Secondary Vegetation, Moderately Modified Forest, Slightly Modified Forest and Mature Forest. However, distribution pattern and, in some cases, number of species, may be easy to distinguish, even to an inexperienced observer. Since the trees were actually human-planted in Woody Plantation, they may be regularly distributed and present a homogeneous texture or be of specific, easily recognizable, commercial species. In secondary vegetation, modified or mature forests a regular pattern in distribution or single species are not expected. To illustrate this point,

Figure 8 presents a subset of high resolution optical Remote Sensing image from an area in Brasil Novo, also in Pará state. This is a SPOT-6 image, acquired on 19 August 2015, in which the spectral bands (6 m of spatial resolution) were fused with the panchromatic band (1.5 m of spatial resolution). In this figure, it is possible to verify the aforementioned difference in pattern from human-planted trees (indexes b and c in the figure) to trees not planted by humans in a uniform pattern (index a). This difference may also be noticeable in medium resolution images because of differences in texture. Note that it is possible to discriminate between 'natural or semi-natural vegetation' and 'cultivated vegetation' in LCML, as illustrated in Figures 5 and 6.

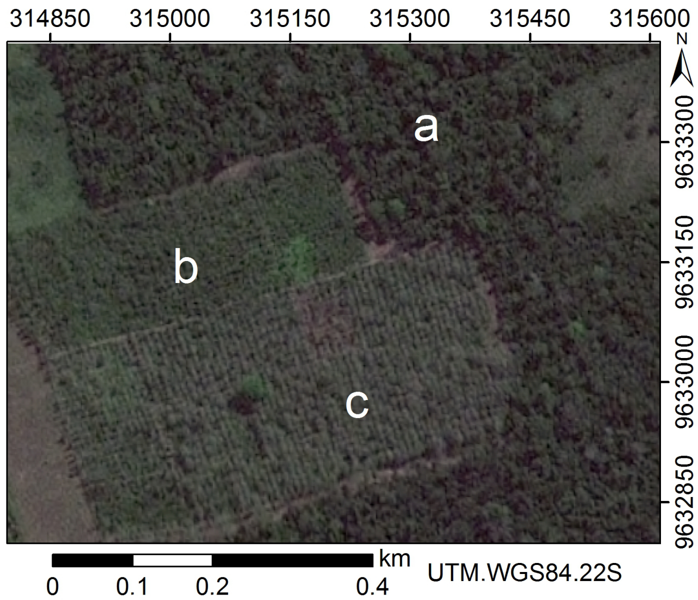


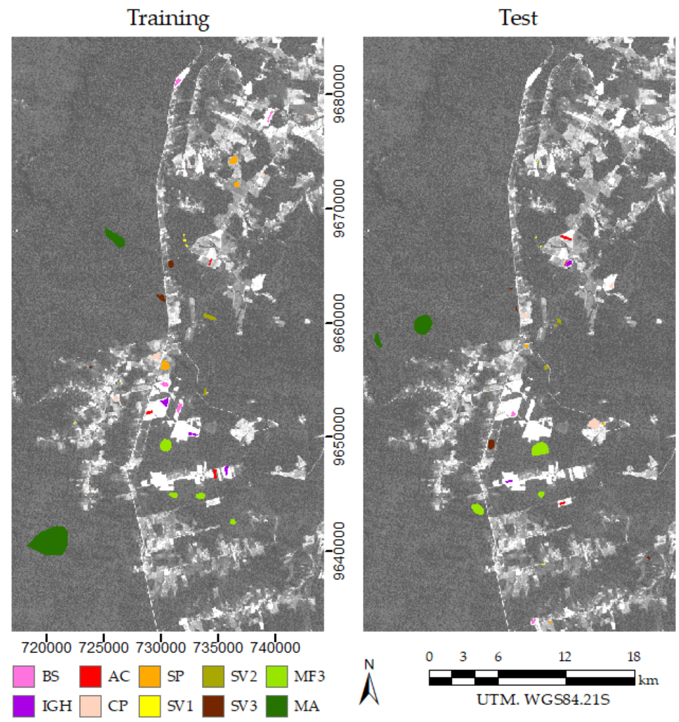
Figure 8. A subset of a SPOT-6 image from an area in Brasil Novo, Pará state, Brazil, R(3)G(2)B(1) color composition: (**a**) area of mature forest; (**b**) area of planted cocoa; (**c**) area of mixed woody plantation.

Distribution pattern also seems to be useful to distinguish different types of alteration that lead to modified forests. Some examples of different patterns of modified forests and how it alters the appearance of the classes in Remote Sensing images are illustrated in works like Asner et al. [71] and Pinheiro et al. [72].

Overall, the defined limits are interesting to be used for data characterization in the field of some classes. For once, the thresholds separating classes that occur in a gradient are well defined. This means that, although the land cover percentage was defined mainly in a subjective way, these thresholds can be useful to help to separate Clean Pasture from Shrubby Pasture, or the point in which a Shrubby Pasture turns into an Initial Secondary Vegetation, as well as distinguishing stages of secondary vegetation or modified forests.

5.3. Defined Legends

In the 2010 Landsat5/TM image, ten LULC classes were identified: Bare Soil, Idle Agricultural Area with Green Herbs, Annual Crop, Clean Pasture, Shrubby Pasture, Initial Secondary Vegetation, Intermediate Secondary Vegetation, Advanced Secondary Vegetation, MF3 and Mature Forest. Samples of these classes can be seen in Figure 9, with band 5 of the Landsat5/TM image for reference, as well as the number of samples collected for each class (number of pixels considering the image spatial resolution).



# of samples in each set		
Class	Training	Test
BS	940	256
IGH	694	344
AC	459	407
CP	831	991
SP	1175	224
SV1	158	167
SV2	516	332
SV3	599	609
MF3	2032	2727
MA	8880	2898

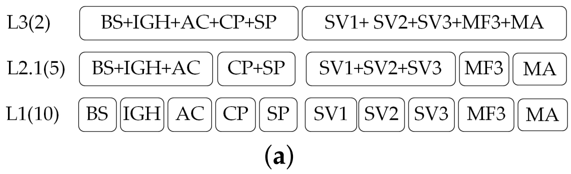
In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest.

Figure 9. Training and test samples sets, superposed on band 5 of the Landsat5/TM image for reference.

The ten LULC classes were then organized into three types of legends: theoretical, based on the parametrization and based on automatic clustering algorithms. The first two legends are illustrated in Figure 10. In this figure, each legend is named as $L_i(n)$, with i being an identifier and n representing the number of classes in the legend. Note that two of the three layers of these legends are equal, so the same nomenclature was adopted. Each legend is explained as follows:

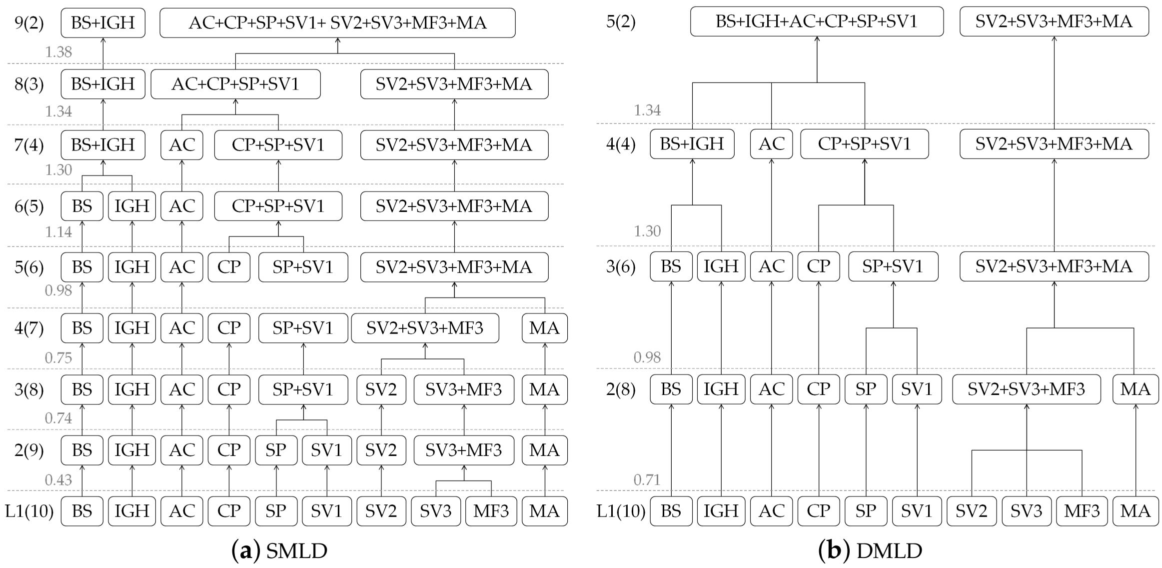
- **L1(10):** the ten identified LULC classes. This is the most detailed legend and may be useful for data field collection and mappings with diverse objectives;
- **L2.1(5):** first grouping of the ten LULC classes, following the relationships established by the hierarchical class system. This legend has the potential to yield better classification results than L1(10) while maintaining significant classes for decision-makers;
- **L2.2(4):** grouping of the ten LULC classes based on similarity in the parametrization. Good classification results Remote Sensing data classification are expected with this legend [62];
- **L3(2):** this legend possesses only two classes, which can be achieved either by following the hierarchical class system or the classifiers in LCML and for which the most accurate classification results are expected. This legend may be particularly interesting for deforestation related studies.

The legends based on the automatic clustering algorithms are depicted in Figure 11. In these legends, tiers are presented as 'Legend_n', in which 'Legend' can be either 'SMLD' or 'DMLD' and 'n' is the associated level. The number of classes is reported in parentheses after the legend's name. In Figure 11, the minimum dissimilarity value between the pairs of classes in each tier is registered in SMLD and the second minimum in DMLD. Notice that, in the first tier, in all legends are the same set of ten classes (L1(10)), and that SMLD_5(6) and DMLD_3(6) are the same, and so are SMLD_7(4) and DMLD_4(4).



In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest.

Figure 10. Legends proposed based on (**a**) the hierarchical class system and (**b**) parametrization.



In which SMLD = Single Minimum Link Dendrogram, DMLD = Double Minimum Link Dendrogram, BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest.

Figure 11. Legends based on spectral separability of data.

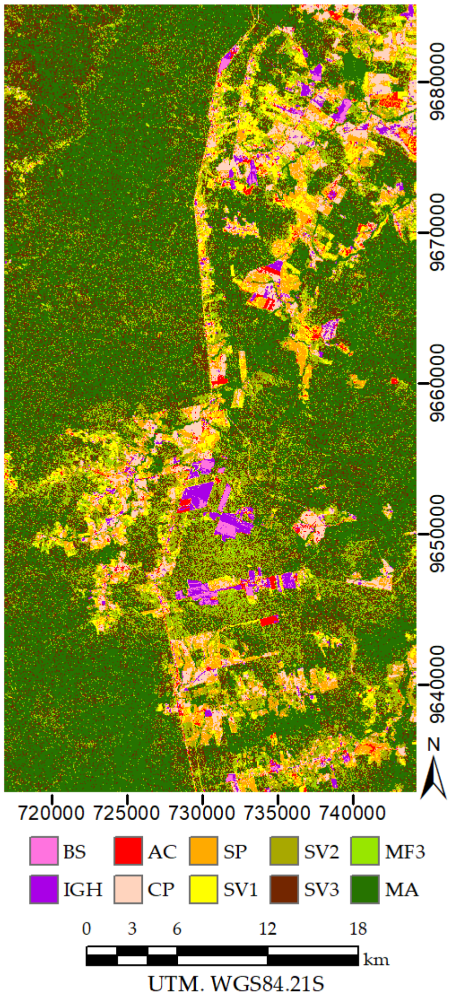
It is possible to notice that, with the obvious exception of L1(10), no tier of the legends based on automatic clustering algorithms is equal to a tier of the other defined legends. Actually, only from the legends SMLD_2(9) and DMLD_2(8), it is possible to obtain L2.2(4) or L3(2) because of the union of Shrubby Pasture (SP) and Initial Secondary Vegetation (SV1) in less detailed legends based on data. Since these classes pertain to different major LULC classes according to the model presented in Section 5.1 (Clean Pasture and Shrubby Pasture pertain to 'Cultivated and Managed Terrestrial Area' and Initial Secondary Vegetation to 'Natural and Semi-Natural Vegetation'), these are never unified in the proposed theoretical legend. However, besides the proximity in the DN values of Shrubby Pasture and Initial Secondary Vegetation samples in the feature space, it is common for Initial Secondary Vegetation to be characterized as a transition class from Shrubby Pasture to more developed secondary vegetation classes in LULC studies. This may be exemplified by the existence of a class 'Secondary Vegetation with Pasture', in the TerraClass legend. Therefore, although these classes may present very different meanings depending on the study being carried, the legends SMLD_7(4) and DMLD_4(4), or the lower ones, may still be useful for LULC mapping.

Another difference between legends with the same number of classes is that, in the legends based on the hierarchical class system or the parametrization, the classes Bare Agricultural Soil (BS), Idle Agricultural Area with Green Herbs (IGH) and Annual Crop (AC) were aggregated in the second level. Since these classes present different spectral characteristics in the selected features of Landsat5/TM data, in the legends based on data, BS+IGH were only merged with AC in

DMLD_5(2) legend. In SMLD_8(3), AC was merged with CP+SP+SV1, which leads to the grouping of AC+CP+SP+SV1 and SV2+SV3+MF3+Ma in SMLD_9(2). In DMLD both AC and BS+IGH were merged to CP+SP+SV1, which leads to a two class legend somewhat useful for some types of studies, given the previously discussed relation between SP and SV1.

5.4. Image Classification

As previously explained, samples illustrated in Figure 9 were grouped in each legend and used to classify bands 3, 4 and 5 of the LANDAT5/TM image using a Monte Carlo approach and ML classifier. We will first focus on the classification results obtained by the legends that are independent of the Remote Sensing image, i.e., those based on the hierarchical class system and on parametrization. The classified images, as well as the confusion matrix (in %) and accuracy indexes, are presented in Figures 12–15.



Mean Confusion Matrix

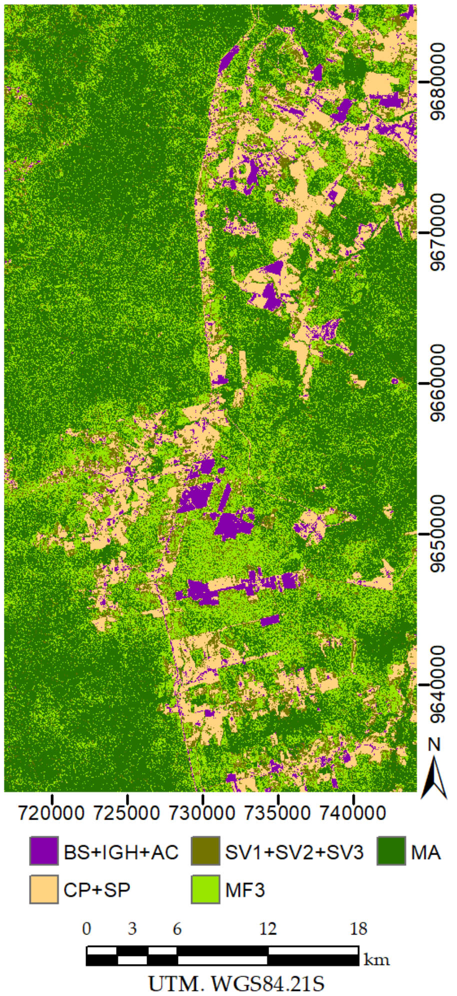
Classification \ Reference	BS	IGH	AC	CP	SP	SV1	SV2	SV3	MF3	MA
BS	**98**	0	0	0	0	0	0	0	0	0
IGH	2	**42**	0	6	9	0	0	0	0	0
AC	0	5	**100**	7	1	3	0	0	0	0
CP	0	53	0	**77**	18	0	0	0	0	0
SP	0	0	0	10	**43**	8	2	0	0	0
SV1	0	0	0	0	29	**87**	2	1	3	0
SV2	0	0	0	0	0	1	**44**	52	29	1
SV3	0	0	0	0	0	0	21	**22**	23	19
MF3	0	0	0	0	0	1	30	21	**34**	4
MA	0	0	0	0	0	0	1	4	11	**76**

Mean Accuracy Indexes

Class	PA	UA
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.42 ± 0.07	0.71 ± 0.07
AC	1.00 ± 0.00	0.87 ± 0.03
CP	0.78 ± 0.06	0.52 ± 0.07
SP	0.43 ± 0.06	0.68 ± 0.04
SV1	0.87 ± 0.04	0.72 ± 0.04
SV2	0.44 ± 0.06	0.35 ± 0.04
SV3	0.22 ± 0.06	0.26 ± 0.06
MF3	0.34 ± 0.07	0.38 ± 0.06
MA	0.76 ± 0.06	0.83 ± 0.05
Overall Accuracy	0.62 ± 0.02	
Kappa	0.58 ± 0.02	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, MA = Mature Forest, PA = Producer's Accuracy and UA = User's Accuracy.

Figure 12. Classification results for the ML(Maximun Likelihood) algorithm using the L1(10) legend, related mean confusion matrix (in %) (with the main diagonal in bold font) and accuracy indexes (presented as $m \pm sd$ in which m is the mean value of the indexes and sd is the standard deviation).



Mean Confusion Matrix

	Reference				
Classification	BS+IGH+AC	CP+SP	SV1+SV2+SV3	MF3	MA
BS+IGH+AC	**98**	16	1	0	0
CP+SP	2	**84**	11	0	0
SV1+SV2+SV3	0	0	**27**	19	2
MF3	0	0	56	**66**	13
MA	0	0	5	15	**85**

Mean Accuracy Indexes

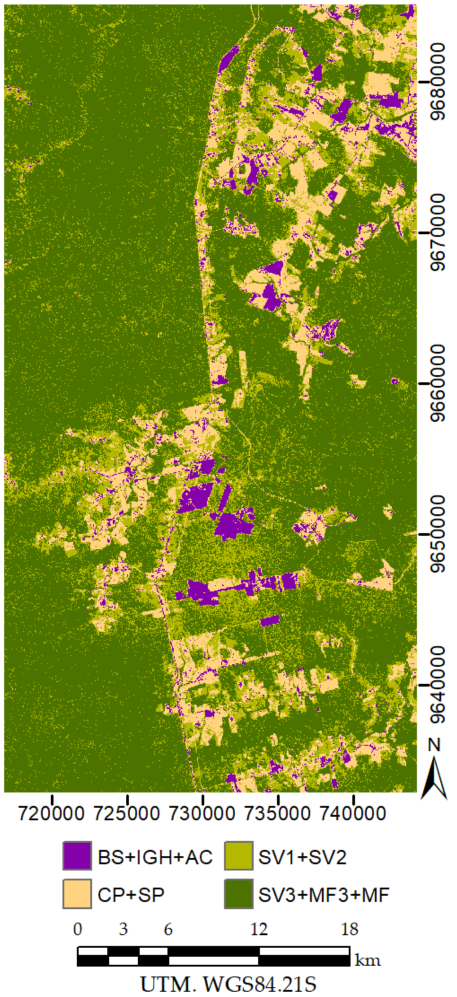
Class	PA	UA
BS+IGH+AC	0.99 ± 0.02	0.86 ± 0.04
CP+SP	0.84 ± 0.05	0.87 ± 0.04
SV1+SV2+SV3	0.27 ± 0.06	0.56 ± 0.09
MF3	0.67 ± 0.07	0.49 ± 0.04
MA	0.85 ± 0.05	0.81 ± 0.05
Overall Accuracy	0.72 ± 0.02	
Kappa	0.65 ± 0.03	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest, PA = Producer's Accuracy and UA = User's Accuracy.

Figure 13. Classification results for the ML algorithm using the L2.1(5) legend, related mean confusion matrix (in %) (with the main diagonal in bold font) and accuracy indexes (presented as $m \pm sd$ in which m is the mean value of the indexes and sd is the standard deviation).

These classifications were obtained with a simple classification algorithm, only one image for only one date, and a pixel-wise approach. Therefore, it was expected that classes formed by the same elements of land cover in similar proportions would present some confusion in the classification results. This is clear in the analysis of results from L1(10) legend, given the high amount of samples of Idle Agricultural Area of Green Herbs classified as Clean Pasture, or the confusion between Slightly Modified Forest and Intermediate Secondary Vegetation/ Advanced Secondary Vegetation. Classes without superposing thresholds but with successional behavior, as Shrubby Pasture/ Initial Secondary Vegetation or Intermediate Secondary Vegetation/ Advanced Secondary Vegetation, also presented high confusion. Despite the high confusion between some classes, the use of L1 legend resulted in the good classification of classes as Bare Agricultural Soil, Annual Crop, and, to a lesser degree, Initial Secondary Vegetation and Mature Forest.

The mean accuracy indexes (Overall Accuracy, Kappa, Producer's Accuracy and User's Accuracy) of classifications using the legends obtained by the clustering algorithms are depicted in Table 3, along with the standard deviation of values. In this table, indexes related to classifications with a proximate number of classes to the theoretical legends are presented first, followed by the remaining ones.



Mean Confusion Matrix

	Reference			
Classification	BS+IGH+AC	CP+SP	SV1+SV2	SV3+MF3+MF
BS+IGH+AC	**98**	16	1	0
CP+SP	2	**83**	11	0
SV1+SV2	0	1	**69**	33
SV3+MF3+MF	0	0	19	**67**

Mean Accuracy Indexes

Class	PA	UA
BS+IGH+AC	0.98 ± 0.02	0.86 ± 0.04
CP+SP	0.83 ± 0.05	0.87 ± 0.04
SV1+SV2	0.69 ± 0.06	0.67 ± 0.05
SV3+MF3+MF	0.67 ± 0.06	0.78 ± 0.05
Overall Accuracy	0.79 ± 0.03	
Kappa	0.73 ± 0.03	

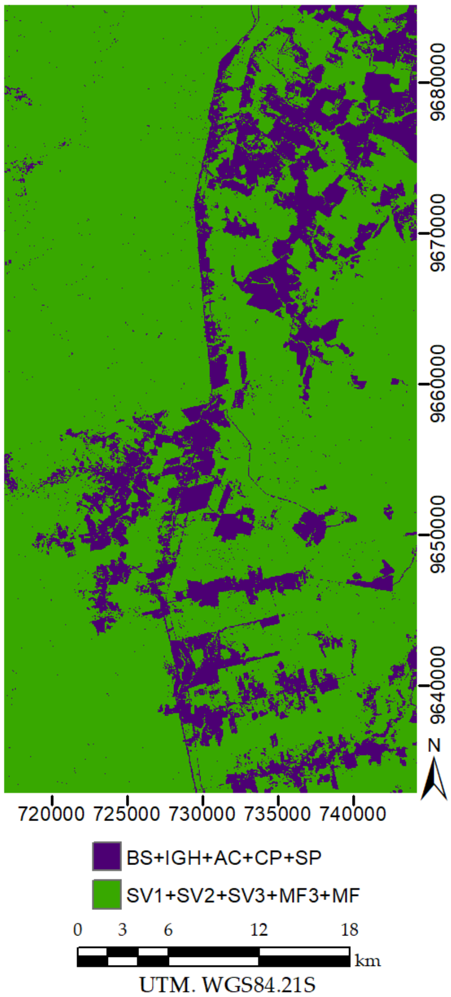
In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest, PA = Producer's Accuracy and UA = User's Accuracy.

Figure 14. Classification results for the ML algorithm using the L2.2(4) legend, related mean confusion matrix (in %) (with the main diagonal in bold font) and accuracy indexes (presented as $m \pm sd$ in which m is the mean value of the indexes and sd is the standard deviation).

Higher accuracy values were obtained using legends based on data separability. However, it is important to highlight that the legends based on the hierarchical class system or on parametrization are data independent, so they do not present some of the problems indicated in Section 1 (e.g., obtaining different legends for each dataset).

Although the usefulness of legends based on Remote Sensing data depends on the study being carried on, it is interesting to note that the legend SMLD_2(9) has led to higher accuracy values than L2.1(5), with higher detail. Given that some LULC studies do not distinguish between modified and secondary forests, the merging of the classes Slightly Modified Forest (MF3) and Advanced Secondary Vegetation (SV3) can be inconsequential to the objective, but necessary to achieve better accuracy values. Considering regional studies based on structure, SMLD_7(4)/ DMLD_4(4) seems to present a good compromise between accuracy and detail, with agricultural, pasture and developed forest classes separated. The main problem of these legends is the union of Initial Secondary Vegetation (SV1) with Clean Pasture (CP) and Shrubby Pasture (SP), as previously discussed. Considering the legends with the lower level of detail (2 classes), we have very good accuracy values for both L4(2) and DMLD_5(2), but lower values for SMLD_9(2). Although DMLD_5(2) has shown higher accuracy than L3(2), this difference is small. There is the same possible problem of grouping SV1 to the agricultural classes in DMLD_5(2).

In general, classes that presented the lowest values of mean Producer's and User's Accuracy using one legend have been merged into another class to form the next tier of the legends SMLD and DMLD. Notice that, although the high Producer Accuracy and User Accuracy values for the classes BS+IGH+AC of both L2.1(5) and L2.2(4), the class AC has been well classified even using 10 classes (L1(10)), and this good classification was maintained until this class was merged to another in either SMLD or DMLD. The same happened to the BS class. These results suggest that the use of automatic clustering methods may be useful not only to form legends, but also to determine which subclasses could remain separated from the class of interest in a study, in order to preserve a higher level of detail for other possible uses of the generated classification.



Mean Confusion Matrix

		Reference	
		BS+IGH+ AC+CP+SP	SV1+SV2+ SV3+MF3+MF
Classification	BS+IGH+AC+CP+SP	**100**	4
	SV1+SV2+SV3+MF3+MF	0	**96**

Mean Accuracy Indexes

Class	PA	UA
BS+IGH+AC+CP+OP	1.00 ± 0.00	0.97 ± 0.02
SV1+SV2+SV3+MF3+MF	0.96 ± 0.03	1.00 ± 0.00
Overall Accuracy	0.98 ± 0.01	
Kappa	0.96 ± 0.03	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest, PA = Producer's Accuracy and UA = User's Accuracy.

Figure 15. Classification results for the ML algorithm using the L3(2) legend, related mean confusion matrix (in %) (with the main diagonal in bold font) and accuracy indexes (presented as $m \pm sd$ in which m is the mean value of the indexes and sd is the standard deviation).

Table 3. Mean Producer's Accuracy (PA), mean User's Accuracy (PU) and mean Overall Accuracy and Kappa Coefficient values for classifications of the Landsat5/TM image, followed by ± standard deviation of values, using the proposed legends based on automatic clustering algorithms.

	PA	**UA**
SMLD_5(6)/DMLD_3(6)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.76 ± 0.07
AC	1.00 ± 0.00	0.88 ± 0.044
CP	0.79 ± 0.05	0.56 ± 0.03
SP+SV1	0.82 ± 0.05	0.88 ± 0.04
SV2+SV3+MF3+MA	0.98 ± 0.02	1.00 ± 0.00
Overall Accuracy	0.83 ± 0.02	
Kappa	0.80 ± 0.02	
SMLD_6(5)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.46 ± 0.07	0.87 ± 0.06
AC	1.00 ± 0.00	0.90 ± 0.03
CP+SP+SV1	0.89 ± 0.05	0.63 ± 0.03
SV2+SV3+MF3+MA	0.97 ± 0.02	1.00 ± 0.00
Overall Accuracy	0.86 ± 0.02	
Kappa	0.83 ± 0.02	
SMLD_7(4)/DMLD_4(4)		
BS+IGH	0.71 ± 0.06	0.90 ± 0.04
AC	1.00 ± 0.00	0.93 ± 0.03
CP+SP+SV1	0.87 ± 0.05	0.75 ± 0.04
SV2+SV3+MF3+MA	0.97 ± 0.02	1.00 ± 0.00
Overall Accuracy	0.89 ± 0.02	
Kappa	0.85 ± 0.03	
SMLD_9(2)		
BS+IGH	0.88 ± 0.04	0.96 ± 0.03
AC+CP+SP+SV1+SV2+SV3+MF3+MA	0.97 ± 0.03	0.89 ± 0.03
Overall Accuracy	0.92 ± 0.03	
Kappa	0.85 ± 0.05	
DMLD_5 (2)		
BS+IGH+AC+CP+SP+SV1	0.99 ± 0.01	0.98 ± 0.02
SV2+SV3+MF3+MA	0.98 ± 0.02	0.99 ± 0.01
Overall Accuracy	0.99 ± 0.01	
Kappa	0.97 ± 0.02	
SMLD_2(9)		
BS	0.97 ± 0.02	1.00 ± 0.00
IGH	0.42 ± 0.07	0.71 ± 0.07
AC	1.00 ± 0.00	0.86 ± 0.04
CP	0.78 ± 0.05	0.52 ± 0.03
SP	0.43 ± 0.06	0.70 ± 0.07
SV1	0.88 ± 0.04	0.73 ± 0.04
SV2	0.48 ± 0.07	0.55 ± 0.06
SV3+MF3	0.48 ± 0.07	0.43 ± 0.05
MA	0.84 ± 0.05	0.86 ± 0.04
Overall Accuracy	0.70 ± 0.02	
Kappa	0.66 ± 0.02	

Table 3. *Cont.*

	PA	UA
SMLD_3(8)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.78 ± 0.07
AC	1.00 ± 0.00	0.88 ± 0.04
CP	0.79 ± 0.06	0.56 ± 0.04
SP+SV1	0.81 ± 0.06	0.87 ± 0.04
SV2	0.47 ± 0.06	0.56 ± 0.06
SV3+MF3	0.51 ± 0.07	0.43 ± 0.05
MA	0.84± 0.05	0.86 ± 0.04
Overall Accuracy	0.73 ± 0.02	
Kappa	0.69 ± 0.02	
DMLD_2(8)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.71 ± 0.07
AC	1.00 ± 0.00	0.87 ± 0.04
CP	0.79 ± 0.06	0.53 ± 0.04
SP	0.43 ± 0.06	0.70 ± 0.07
SV1	0.87 ± 0.04	0.73 ± 0.04
SV2+SV3+MF3	0.84 ± 0.05	0.85 ± 0.04
MA	0.87 ± 0.05	0.88 ± 0.04
Overall Accuracy	0.77 ± 0.02	
Kappa	0.74 ± 0.02	
SMLD_4(7)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.74 ± 0.07
AC	1.00 ± 0.00	0.87 ± 0.04
CP	0.78 ± 0.06	0.56 ± 0.04
SP+SV1	0.81 ± 0.05	0.89 ± 0.04
SV2+SV3+MF3	0.86 ± 0.05	0.86 ± 0.04
MA	0.86 ± 0.05	0.88 ± 0.04
Overall Accuracy	0.82 ± 0.02	
Kappa	0.79 ± 0.02	
SMLD_8(3)		
BS+IGH	0.68 ± 0.06	0.95 ± 0.04
AC+CP+SP+SV1	0.95 ± 0.03	0.74 ± 0.04
SV2+SV3+MF3+MA	0.98 ± 0.02	0.99 ± 0.01
Overall Accuracy	0.87 ± 0.02	
Kappa	0.81 ± 0.04	

In which PA = Producer's Accuracy, UA = User's Accuracy, BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest and MA = Mature Forest, PA = Producer's Accuracy and UA = User's Accuracy.

6. Discussion

In this work, a set of hierarchically organized LULC classes for the Brazilian Amazon biome was proposed, as a hierarchical class system. This system, however, is not finished, in the sense that new classes or subdivisions may be included if needed. For instance, it can be important for certain studies to differentiate the agent of forest disturbance, so the analyst can include the distinction between burned and logged forests in the 'Modified Forest' category, similarly to the studies carried out by Souza Jr. et al. [73] and Diniz et al. [15]. As suggested by McConnell and Moran [26], researchers may find how to compare maps obtained at global and local scales at the regional level. In this sense, the use of the hierarchical class system could be interesting to help to determine which classes must be grouped or further detailed to generate data at the regional level.

With field data and previous knowledge, we also defined 16 detailed LULC classes (derived from the lowest level of the hierarchical class system) in LCML. As suggested by Di Gregorio and Jansen [42], these classes may be easily understandable and recognized. Furthermore, very simple classifiers (mean height and proportion of cover of five elements of land cover classes) were used. However, since many of these classifiers presented overlapping thresholds between very different classes, it is clear that they can not be configured as absolute descriptors of each class. For instance, analysts may need different class definition in their studies, or their definition may be based on criteria like above-ground biomass, Net Primary Productivity, forest stand volume, basal area, average stand height, diameter at breast height, age and/or many others, as used by studies such Lu et al. [3], Vieira et al. [69], Salomão et al. [70] and Romero-Sanchez and Ponce-Hernandez [74] to justify the subdivision of secondary vegetation classes. Nonetheless, many of these classifiers are implemented and the users are free to further describe the classes or even to include new subclasses, without compromising the use of the proposed hierarchical class system. Moreover, the simple use of multi-temporal data can easily differentiate between classes with overlapping thresholds, as is the case of Idle Agricultural Area with Green Herbs and Clean Pasture. There are also methods based on multi-temporal Remote Sensing images focused on detecting and measuring forest disturbances [75,76].

Despite this lacking characteristic of the proposed thresholds, it must be highlighted that they are very easily understood by non-specialist analysts, and may be particularly useful for studies that lay on Community-Based Observations [77]. Additionally, there is potential to measuring them directly from very-high-resolution images (cover proportion) and LIDAR data (mean height), which would greatly diminish the subjectivity of studies. Most importantly, the use of the proposed thresholds is useful to delimit classes that occur in a successive way in a given area, which tend to be the most difficult ones to describe. Note that the difference among Clean Pasture, Shrubby Pasture, Initial Secondary Vegetation, Intermediate Secondary Vegetation, Advanced Secondary Vegetation is clearly defined, as well as the difference among Mature Forest and the Modified Forest classes. We understand that one can argue that the difference between 15% and 16% of shrubs to qualify an area as Shrubby Pasture is arbitrary and it is practically impossible to be determined precisely in the field or even in very-high-resolution images. This problem is the core of the Fuzzy set theory [78]. Although the necessity of separating the classes in Boolean sets for field data collection or Remote Sensing analysis can be argued [79], it is still the most common approach because it is operational and simpler than fuzzy methodologies. Furthermore, to use the Boolean class definition, clear thresholds must be defined. In this sense, some misplacement between very similar classes is expected, since measures contain errors. Another perceived advantage of the use of LCML formalized legends is the possibility to compare results obtained with different class definition, as presented by Reis et al. [80]. However, although LCML formalized legends are easily understandable, they may not be directly comparable and a harmonization process would still be necessary [25].

Regarding Landsat5/TM classification results, the comparison among the mean Kappa values demonstrates that more generalized legends (those with few details and number of classes) result in more accurate classifications. This behavior was expected and had been previously documented [8,35,62]. The legends based on data normally lead to the most accurate maps, even though these are not always the most useful ones. However, it is important to highlight once more that the use of different data, be it the image (type, date or pre-processing) or the initial set of classes or the samples collection, would lead to different legends when these are derived using automatic clustering algorithms. Therefore, it is fundamental to verify if the chosen classes are really significant for a determined study objective, or if those are just the ones that yield better accuracy figures. Nonetheless, automatic clustering algorithms based on data may be useful to analyze the data being used. They are also able to point subclasses that could be preserved in order to generate classification with higher detail level that could still be generalized to a theoretical legend, not greatly compromising the accuracy. The legends L2.2(4) and L3(2), that are based on the structural characteristics of classes and data independent, returned high mean Kappa values using the ML algorithm (respectively 0.73 and 0.98) and may be interesting for future studies.

7. Conclusions

This work showed a reproducible hierarchical legend for LULC classes tailored for usage in the southwest of Pará state and compared it to two possible hierarchical legends derived from commonly used Remote Sensing data. These results are useful to guide researchers in the definition of vegetated classes from the upland Brazilian Amazon biome focusing on either classification results or reproducibility, as well as explaining LULC classes to non-specialists with objectivity, mainly in classes happening in a successive manner in a given area. Moreover, the use of the defined classes enables reproducibility of data acquisition for multi-temporal works and also can be an important tool for harmonization of different products and for the translation of field data into information applicable to local, regional or global scale studies. We could also perceive that the collection of field data with estimates of mean height and cover proportion of soil, litter, herbaceous vegetation, shrubs, and trees is important, both for the use of the proposed class definition or to allow the researchers to define their own thresholds.

There are many aspects of this work that may be further explored. For instance, only the upland classes were developed, which leaves mainly in the development of legends for the aquatic or regularly flooded areas. New classes could also be added if necessary. Regarding the classification results and as previously discussed, the use of multi-temporal data may yield better results. It is also feasible to consider the further processing of the image (feature extraction, fusion to other data) and more refined classification methods (region based and object-oriented) to improve classification accuracy. Future works also include testing the applicability of the legends for classification of other types of images, always maintaining a reference legend scheme for comparison, and expansion to other study areas. It is also interesting to harmonize the proposed legends to the ones used in the available Remote Sensing products, in order to allow the comparison among this products. It may also be interesting to estimate the cover proportion of elements from very-high resolution Remote Sensing images and improve the threshold estimates. Additionally, similar studies for other biomes (such as the Brazilian savannah) are important to fully understand the dynamics of forested areas.

Author Contributions: M.S.R. and L.V.D. designed the general framework of the experiment. M.S.R., M.I.S.E. and N.D.V. determined the thresholds applied to class definition. M.S.R and M.I.S.E. proposed the conceptual model and the theoretical legends. M.S.R., L.V.D. and S.J.S.S. designed the automatic clustering and classification methodology. N.D.V. pre-processed the Remote Sensing images. M.S.R. and S.J.S.S. provided the field data. M.S.R. executed the experiments and wrote the paper. All authors revised the paper.

Funding: This research was funded CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), grants #312753/2015-2, #401528/2012-0 and #309135/2015-0.

Acknowledgments: The authors are thankful to Rogério Galante Negri for providing the clustering algorithms and to the Earth System Science Center of the Brazilian National Institute for Space Research. Authors are also thankful to ICMBio (Instituto Chico Mendes de Conservação da Biodiversidade) for authorizations #48186-4, #48186-2 e #38157-2, the LBA program and the Monitoramento Ambiental por Satélite no Bioma Amazônia project, process #1022114003005-MSA-BNDES.

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

References

1. Pielke, R.A.; Pitman, A.; Niyogi, D.; Mahmood, R.; McAlpine, C.; Hossain, F.; Goldewijk, K.K.; Nair, U.; Betts, R.; Fall, S.; et al. Land use/land cover changes and climate: modeling analysis and observational evidence. *Wiley Interdiscip. Rev. Clim. Change* **2011**, *2*, 828–850. [CrossRef]
2. Verburg, P.H.; Erb, K.H.; Mertz, O.; Espíndola, G. Land System Science: between global challenges and local realities. *Curr. Opin. Environ. Sustain.* **2013**, *5*, 433–437. [CrossRef] [PubMed]

3. Lu, D.; Mausel, P.; Brondízio, E.; Moran, E. Classification of successional forest stages in the Brazilian Amazon basin. *For. Ecol. Manag.* **2003**, *181*, 301–312. [CrossRef]
4. Fujiki, S.; Aoyagi, R.; Tanaka, A.; Imai, N.; Kusma, A.D.; Kurniawan, Y.; Lee, Y.F.; Sugau, J.B.; Pereira, J.T.; Samejima, H.; et al. Large-Scale Mapping of Tree-Community Composition as a Surrogate of Forest Degradation in Bornean Tropical Rain Forests. *Land* **2016**, *5*, 45. [CrossRef]
5. Deus, D. Integration of ALOS PALSAR and Landsat Data for Land Cover and Forest Mapping in Northern Tanzania. *Land* **2016**, *5*, 43. [CrossRef]
6. Angelis, C.F.; Freitas, C.C.; Valeriano, D.M.; Dutra, L.V. Multitemporal analysis of land use/land cover JERS-1 backscatter in the Brazilian tropical rainforest. *Int. J. Remote Sens.* **2002**, *23*, 1231–1240. [CrossRef]
7. Viet Nguyen, L.; Tateishi, R.; Kondoh, A.; Sharma, R.C.; Thanh Nguyen, H.; Trong To, T.; Ho Tong Minh, D. Mapping Tropical Forest Biomass by Combining ALOS-2, Landsat 8, and Field Plots Data. *Land* **2016**, *5*, 31. [CrossRef]
8. Negri, R.G. Avaliação de Dados Polarimétricos Do Sensor ALOS PALSAR Para Classificação da Cobertura da Terra da Amazônia. Master's Thesis, Brazilian National Institute for Space Research, São José dos Campos, Brazil, 2009.
9. Elz, I.; Tansey, K.; Page, S.E.; Trivedi, M. Modelling Deforestation and Land Cover Transitions of Tropical Peatlands in Sumatra, Indonesia Using Remote Sensed Land Cover Data Sets. *Land* **2015**, *4*, 670–687. [CrossRef]
10. Newby, J.; Cramb, R.; Sakanphet, S. Forest Transitions and Rural Livelihoods: Multiple Pathways of Smallholder Teak Expansion in Northern Laos. *Land* **2014**, *3*, 482–503. [CrossRef]
11. Pereira, L.O.; Freitas, C.C.; Sant'Anna, S.J.S.; Lu, D.; Moran, E.F. Optical and radar data integration for land use and land cover mapping in the Brazilian Amazon. *Gisci. Remote Sens.* **2013**, *50*, 301–321.
12. Silva, W.B.; Pereira, L.O.; Sant'Anna, S.J.S.; Freitas, C.C.; Guimarães, R.J.P.S.; Frery, A.C. Land cover discrimination at Brazilian Amazon using region based classifier and stochastic distance. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, BC, Canada, 24–29 July 2011; pp. 2900–2903.
13. Almeida, C.A.d.; Coutinho, A.C.; Esquerdo, J.C.D.M.; Adami, M.; Venturieri, A.; Diniz, C.G.; Dessay, N.; Durieux, L.; Gomes, A.R. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazon.* **2016**, *46*, 291–302. [CrossRef]
14. Arino, O.; Gross, D.; Ranera, F.; Leroy, M.; Bicheron, P.; Brockman, C.; Defourny, P.; Vancutsem, C.; Achard, F.; Durieux, L.; et al. GlobCover: ESA service for global land cover from MERIS. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Barcelona, Spain, 23–28 July 2007; pp. 2412–2415.
15. Diniz, C.G.; Souza, A.A.A.; Santos, D.C.; Dias, M.C.; Luz, N.C.; Moraes, D.R.V.; Maia, J.S.; Gomes, A.R.; Narvaes, I.S.; Valeriano, D.M.; et al. DETER-B: the new Amazon near real-time deforestation detection system. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3619–3628. [CrossRef]
16. Matthews, E. Global vegetation and land-use: New-high resolution data bases for climate studies. *J. Clim. Appl. Meteorol.* **1983**, *22*, 474–487. [CrossRef]
17. Loveland, T.; Belward, A. the International Geosphere Biosphere Programme Data and Information System global land cover data set (DISCover). *Acta Astronaut.* **1997**, *41*, 681–689. [CrossRef]
18. Loveland, T.R.; Reed, B.C.; Brown, J.F.; Ohlen, D.O.; Zhu, Z.; Yang, L.; Merchant, J.W. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [CrossRef]
19. Friedl, M.; McIver, D.; Hodges, J.; Zhang, X.; Muchoney, D.; Strahler, A.; Woodcock, C.; Gopal, S.; Schneider, A.; Cooper, A.; et al. Global land cover mapping from MODIS: Algorithms and early results. *Remote Sens. Environ.* **2002**, *83*, 287–302. [CrossRef]
20. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [CrossRef]
21. Townshend, J.R.; Masek, J.G.; Huang, C.; Vermote, E.F.; Gao, F.; Channan, S.; Sexton, J.O.; Feng, M.; Narasimhan, R.; Kim, D.; et al. Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *Int. J. Digit. Earth* **2012**, *5*, 373–397. [CrossRef]

22. 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]
23. Herold, M.; Latham, J.S.; Gregorio, A.D.; Schmullius, C.C. Evolving standards in land cover characterization. *J. Land Use Sci.* **2006**, *1*, 157–168. [CrossRef]
24. Herold, M.; Hubald, R.; Gregorio, A.D. *Translating and Evaluating Land Cover Legends Using the UN Land Cover Classfication System (LCCS)*; GOFC-GOLD Report No. 43. Technical Report 189p; Global Observation of Forest and Land Cover Dynamics (GOFC-GOLD): Jena, Germany, 2009.
25. Jansen, L.J.; Groom, G.; Carrai, G. Land-cover harmonisation and semantic similarity: some methodological issues. *J. Land Use Sci.* **2008**, *3*, 131–160. [CrossRef]
26. McConnell, W.; Moran, E. *Meeting in the Middle: The Challenge of Meso-Level Integration*; LUCC Report Series No. 5. Technical Report 62p; Land Utilization Coordinating Council: Rome, Italy, 2001.
27. Herold, M.; Woodcock, C.; di Gregorio, A.; Mayaux, P.; Belward, A.; Latham, J.; Schmullius, C. A joint initiative for harmonization and validation of land cover datasets. *IEEE Trans. Geosci. Remote Sensi.* **2006**, *44*, 1719–1727. [CrossRef]
28. Ahlqvist, O. Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: A study of 1992 and 2001 U.S. National Land Cover Database changes. *Remote Sens. Environ.* **2008**, *112*, 1226–1241. [CrossRef]
29. 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]
30. Putz, F.E.; Redford, K.H. The Importance of Defining 'Forest': Tropical Forest Degradation, Deforestation, Long-term Phase Shifts, and Further Transitions. *Biotropica* **2010**, *42*, 10–20. [CrossRef]
31. Morales-Barquero, L.; Skutsch, M.; Jardel-Peláez, E.J.; Ghilardi, A.; Kleinn, C.; Healey, J.R. Operationalizing the Definition of Forest Degradation for REDD+, with Application to Mexico. *Forests* **2014**, *5*, 1653–1681. [CrossRef]
32. Chazdon, R.L.; Brancalion, P.H.S.; Laestadius, L.; Bennett-Curry, A.; Buckingham, K.; Kumar, C.; Moll-Rocek, J.; Vieira, I.C.G.; Wilson, S.J. When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. *Ambio* **2016**, *456*, 1–13. [CrossRef] [PubMed]
33. Freitas, C.C.; Soler, L.S.; Sant'Anna, S.J.S.; Dutra, L.V.; Santos, J.R.; Mura, J.C.; Correia, A.H. Land use and land cover mapping in Brazilian Amazon using polarimetric airborne P-band SAR data. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2956–2970. [CrossRef]
34. Dutra, L.V.; Scofield, G.B.; Aboud Neta, S.R.; Negri, R.G.; Freitas, C.C.; Andrade, D. *Land Cover Classification in Amazon Using Alos Palsar Full Polarimetric Data*; Simpósio Brasileiro de Sensoriamento Remoto, 14. (SBSR); Instituto Nacional de Pesquisas Espaciais (INPE): São José dos Campos, Brazil, 2009; pp. 7259–7264.
35. Pantaleão, E.; Dutra, L.V.; Sandri, S.A. Scenario analysis for image classification using multi-objective optimization. *InfoComp* **2012**, *11*, 15–22.
36. Negri, R.G.; Sant'Anna, S.J.S.; Dutra, L.V. A new contextual version of Support Vector Machine based on hyperplane translation. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, VIC, Australia, 21–26 July 2013; pp. 3116–3119.
37. Reis, M.S.; Torres, L.; Sant'Anna, S.J.S.; Freitas, C.C.; Dutra, L.V. Evaluation of SAR-SDNLM filter for change detection classification. In Proceedings of the 2014 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 2042–2045.
38. Dutra, L.V.; Negri, R.G.; Sant'Anna, S.J.S.; Lu, D. *Development of Dissimilarity Functions Using Stochastic Distances for Region-Based Land Cover Classification: A Case Study Near Tapajós Flona, Pará State, Brazil*; Simpósio Brasileiro de Sensoriamento Remoto, 17. (SBSR); Instituto Nacional de Pesquisas Espaciais (INPE): São José dos Campos, Brazil, 2015; pp. 1655–1662.
39. Anjos, D.S.; Lu, D.; Dutra, L.V.; Sant'Anna, S. Change Detection Techniques Using Multisensor Data. In *Remotely Sensed Data Characterization, Classification, and Accuracies*; Crc Press: London, UK, 2015; Volume 1, pp. 375–395.
40. Pereira, L.O.; Freitas, C.C.; Sant'Anna, S.J.S.; Reis, M.S. ALOS/PALSAR Data Evaluation for Land Use and Land Cover Mapping in the Amazon Region. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 5413–5423. [CrossRef]

41. Anderson, J.R. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; US Government Printing Office: Washington, DC, USA, 1976; Volume 964.
42. Di Gregorio, A.; Jansen, L. *Land Cover Classificiation System*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2005.
43. Uddin, K.; Shrestha, H.L.; Murthy, M.; Bajracharya, B.; Shrestha, B.; Gilani, H.; Pradhan, S.; Dangol, B. Development of 2010 national land cover database for the Nepal. *J. Environ. Manag.* **2015**, *148*, 82–90. [CrossRef] [PubMed]
44. Arafat, S.M.; Saleh, N.S.; Aboelghar, M.; Elshrkawy, M. Mapping of North Sinai land cover according to FAO-LCCS. *J. Remote Sens. Space Sci.* **2014**, *17*, 29–39. [CrossRef]
45. Cord, A.; Conrad, C.; Schmidt, M.; Dech, S. Standardized FAO-LCCS land cover mapping in heterogeneous tree savannas of West Africa. *J. Arid Environ.* **2010**, *74*, 1083–1091. [CrossRef]
46. Coutinho, A.; Almeida, C.; Venturieri, A.; Esquerdo, J.; Silva, M. *Uso e Cobertura da Terra Nas Áreas Desflorestadas da AMazônia Legal: TerraClass, 2008*; Technical Report; Embrapa, INPE: Brasília, Belém, 2013; 116p.
47. Kosmidou, V.; Petrou, Z.; Bunce, R.G.; Mücher, C.A.; Jongman, R.H.; Bogers, M.M.; Lucas, R.M.; Tomaselli, V.; Blonda, P.; Padoa-Schioppa, E.; et al. Harmonization of the Land Cover Classification System (LCCS) with the General Habitat Categories (GHC) classification system. *Ecol. Indic.* **2014**, *36*, 290–300. [CrossRef]
48. IBGE. *Manual Técnico de Uso da Terra—3 Edição*; Technical Report; Instituto Brasileiro de Geografia e Estatística (IBGE): Rio de Janeiro, Brazil, 2015; 171p.
49. IBAMA. *Floresta NAcional do TApajós Plano de Manejo: Volume I—Informações Gerais*; Technical Report; Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis: Federal District, Brazil, 2004; 76p.
50. Shimabukuro, Y.; Amaral, S.; Ahern, F.; Pietsch, R. Land Cover Classification from RADARSAT Data of the Tapajós National Forest, Brazil. *Can. J. Remote Sens.* **1998**, *24*, 393–401. [CrossRef]
51. Asner, G.P.; Bustamante, M.M.; Townsend, A.R. Scale dependence of biophysical structure in deforested areas bordering the Tapajós National Forest, Central Amazon. *Remote Sens. Environ.* **2003**, *87*, 507–520. [CrossRef]
52. Keller, M.; Varner, R.; Dias, J.D.; Silva, H.; Crill, P.; de Oliveira, R.C., Jr.; Asner, G.P. Soil–Atmosphere Exchange of Nitrous Oxide, Nitric Oxide, Methane, and Carbon Dioxide in Logged and Undisturbed Forest in the Tapajos National Forest, Brazil. *Earth Interact.* **2005**, *9*, 1–28. [CrossRef]
53. Alves, D.A. Science and technology and sustainable development in Brazilian Amazon. In *Stability of Tropical Rainforest Margins*; Environmental Science and Engineering; Tscharntke, T., Leuschner, C., Zeller, M., Guhardja, E., Bidin, A., Eds.; Springer: Heidelberg/Berlin, Germany, 2007; pp. 491–510.
54. Keller, M.; Palace, M.; Hurtt, G. Biomass estimation in the Tapajos National Forest, Brazil: Examination of sampling and allometric uncertainties. *For. Ecol. Manag.* **2001**, *154*, 371–382. [CrossRef]
55. Lei, Y.; Treuhaft, R.; Keller, M.; dos Santos, M.; Gonçalves, F.; Neumann, M. Quantification of selective logging in tropical forest with spaceborne SAR interferometry. *Remote Sens. Environ.* **2018**, *211*, 167–183. [CrossRef]
56. Green, G.; Schweik, C.; Hanson, M. *Radiometric Calibration of LANDSAT Multi-Spectral Scanner and Thematic Mapper Images: Guidelines for the Global Change Community*; Technical Report; CIPEC Working Paper CWP-02-03; Bloomington Center for the Study of Institutions, Population, and Environmental Change (CIPEC), Indiana University: Bloomington, IN, USA, 2002.
57. Schowengerdt, R. *Remote Sensing*, 3rd ed.; Academic Press: Washington, DC, USA, 2006; p. 560.
58. Moran, E.; Brondízio, E.; Tucker, J.; da Silva-Forsberg, M.; Falesi, I.; McCracken, S. Land-use change after deforestation in Amazônia. In *People and Pixels: Linking Remote Sensing and Social Science*; Liverman, D., Moran, E., Rindfuss, R., Stern, P., Eds.; National Academy Press: Washington, DC, USA, 1998; pp. 129–149.
59. Moran, E.; Brondízio, E. Strategies for Amazonian forest restoration: Evidence for afforestation in five regions of the Brazilian Amazon. In *Amazônia at the Crossroads: The Challenge of Sustainable Development*; Hall, A., Ed.; Institute for Latin American Studies, University of London: London, UK, 2000; pp. 94–120.
60. Moran, E.F.; Brondizio, E.S.; Tucker, J.M.; da Silva-Forsberg, M.C.; McCracken, S.; Falesi, I. Effects of soil fertility and land-use on forest succession in Amazônia. *For. Ecol. Manag.* **2000**, *139*, 93–108. [CrossRef]
61. Food and Agriculture Organization. *FAO Land Cover Classification System 3*, (Version 1.8.0); Food and Agriculture Organization: Rome, Italy, 2015.

62. Reis, M.S.; Dutra, L.V.; Sant'Anna, S.J.S.; Escada, M.I.S. Examining Multi-Legend Change Detection in Amazon with Pixel and Region Based Methods. *Remote Sens.* **2017**, *9*, 77. [CrossRef]
63. Hornik, K.; Buchta, C.; Zeileis, A. Open-Source Machine Learning: R Meets Weka. *Comput. Stat.* **2009**, *24*, 225–232. [CrossRef]
64. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann: San Francisco, CA, USA, 2005.
65. Hornik, K.; Buchta, C.; Hothorn, T.; Karatzoglou, A.; Meyer, D.; Zeileis, A. RWeka: R/Weka Interface, Version 3.9.2; 2018. Available online: https://www.google.com.hk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwikt_Krg4fbAhUMkpQKHdgkBQkQFgg0MAE&url=ftp%3A%2F%2Fcran.r-project.org%2Fpub%2FR%2Fweb%2Fpackages%2FRWeka%2FRWeka.pdf&usg=AOvVaw1wK2Z8V4vD-W3_PUV4kKRO (accessed on 14 May 2018).
66. Wies, D. Rasclass: Supervised Raster Image Classification, Version 0.2.2; 2016. Available online: https://cran.r-project.org/web/packages/rasclass/index.html (accessed on 14 May 2018).
67. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2008.
68. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]
69. Vieira, I.C.G.; de Almeida, A.S.; Davidson, E.A.; Stone, T.A.; de Carvalho, C.J.R.; Guerrero, J.B. Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazônia. *Remote Sens. Environ.* **2003**, *87*, 470–481. [CrossRef]
70. Salomão, R.P.; Vieira, I.C.G.; Brienza, S., Jr.; Amaral, D.D.; Santana, A.C. Sistema Capoeira Classe: Uma proposta de sistema de classificação de estágios sucessionais de florestas secundárias para o estado do Pará. *Mus. Para. Emilio Goeldi* **2012**, *7*, 297–317.
71. Asner, G.P.; Keller, M.; Silva, J.N. Spatial and temporal dynamics of forest canopy gaps following selective logging in the eastern Amazon. *Glob. Chang. Biol.* **2004**, *10*, 765–783. [CrossRef]
72. Pinheiro, T.F.; Escada, M.I.S.; Valeriano, D.d.M.; Hostert, P.; Gollnow, F.; Muller, H. Forest degradation associated with logging frontier expansion in the Amazon: the BR-163 region in southwestern Pará, Brazil. *Earth Interact.* **2016**, *20*, 17. [CrossRef]
73. Souza, C., Jr.; Firestone, L.; Silva, L.M.; Roberts, D. Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models. *Remote Sens. Environ.* **2003**, *87*, 494–506. [CrossRef]
74. Romero-Sanchez, M.E.; Ponce-Hernandez, R. Assessing and Monitoring Forest Degradation in a Deciduous Tropical Forest in Mexico via Remote Sensing Indicators. *Forests* **2017**, *8*, 302. [CrossRef]
75. Tritsch, I.; Sist, P.; Narvaes, I.S.; Mazzei, L.; Blanc, L.; Bourgoin, C.; Cornu, G.; Gond, V. Multiple Patterns of Forest Disturbance and Logging Shape Forest Landscapes in Paragominas, Brazil. *Forests* **2016**, *7*, 315. [CrossRef]
76. Jarron, L.R.; Hermosilla, T.; Coops, N.C.; Wulder, M.A.; White, J.C.; Hobart, G.W.; Leckie, D.G. Differentiation of Alternate Harvesting Practices Using Annual Time Series of Landsat Data. *Forests* **2017**, *8*, 15. [CrossRef]
77. Pratihast, A.K.; DeVries, B.; Avitabile, V.; de Bruin, S.; Kooistra, L.; Tekle, M.; Herold, M. Combining Satellite Data and Community-Based Observations for Forest Monitoring. *Forests* **2014**, *5*, 2464–2489. [CrossRef]
78. Zadeh, L. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
79. Fisher, P.F. Remote Sensing of land cover classes as type 2 fuzzy sets. *Remote Sens. Environ.* **2010**, *114*, 309–321. [CrossRef]
80. Reis, M.S.; Escada, M.I.S.; Sant'Anna, S.J.S.; Dutra, L.V. *Harmonização de Legendas Formalizadas em Land Cover Meta Language-LCML*; Gherardi, D.F.M., Aragão, L.E.O.e.C.d., Eds.; Simpósio Brasileiro de Sensoriamento Remoto, 18. (SBSR); Instituto Nacional de Pesquisas Espaciais (INPE): São José dos Campos, Brazil, 2017; pp. 863–870.