forests

MDPI

Article

REDD+: Quick Assessment of Deforestation Risk Based on Available Data

Giulio Di Lallo [1,2,3,*], Philip Mundhenk [2], Sheila Edith Zamora López [2], Marco Marchetti [1] and Michael Köhl [2,3]

1 DiBT Department of Bioscience and Territory, University of Molise, Contrada Fonte Lappone snc, I-86090 Pesche, Italy; marchettimarco@unimol.it
2 Institute for World Forestry, University of Hamburg, Leuschnerstr. 91, 21031 Hamburg, Germany; Philip.mundhenk@uni-hamburg.de (P.M.); szamora.lopez@gmai.com (S.E.Z.L.); Michael.koehl@uni-hamburg.de (M.K.)
3 The EFI Project Centre SURF—Supporting the Global Implementation of REDD+ and FLEGT, Institute for World Forestry, University of Hamburg, Leuschnerstr. 91, 21031 Hamburg, Germany
* Correspondence: giulio.di.lallo@uni-hamburg.de; Tel.: +49-40-7396-2152

Academic Editor: Timothy A. Martin
Received: 18 October 2016; Accepted: 18 January 2017; Published: 22 January 2017

Abstract: The evaluation of the future dynamics of deforestation is essential to creating the basis for the effective implementation of REDD+ (Reducing Emissions from Deforestation and forest Degradation) initiatives. Such evaluation is often a challenging task, especially for countries that have to cope with a critical lack of data and capacities, higher uncertainties, and competing interests. We present a new modeling approach that makes use of available and easily accessible data sources to predict the spatial location of future deforestation. This approach is based on the Random Forest algorithm, which is a machine learning technique that enables evidence-based, data-driven decisions and is therefore often used in decision-making processes. Our objective is to provide a straightforward modeling approach that, without requiring cost-intensive assessments, can be applied in the early stages of REDD+, for a stepwise implementation approach of REDD+ projects in regions with limited availability of data, capital, technical infrastructure, or human capacities. The presented model focuses on building business-as-usual scenarios to identify and rank potentially suitable areas for REDD+ interventions. For validation purposes we applied the model to data from Nicaragua.

Keywords: REDD+; tropical forests; spatial targeting; random forests; carbon; land-use change modelling

1. Introduction

Deforestation and forest degradation are the largest anthropogenic sources of CO_2 emissions into the atmosphere [1] other than fossil fuel combustion. Tropical forests are the cornerstones of climate change mitigation—they sequester more carbon at faster rates than temperate and boreal forests [2]. Carbon released from loss of forests accounts for at least 12%–20% of the global anthropogenic emissions of greenhouse gasses (GHGs) [3,4]. REDD+ (Reducing Emissions from Deforestation and forest Degradation) aims to mitigate climate change by abating carbon emissions from forests in developing countries (named non-Annex I Parties) through a wide set of activities [5]. A system of economic incentives prompts non-Annex I Parties to participate in the program [6]. Countries willing to participate have to adhere to a REDD+ national strategic plan providing a comprehensive understanding of the political, social, and economic dynamics affecting land-use change. The plan also provides guidance on the efficient allocation of the limited amount of available funding. To efficiently allocate funds, it is necessary to identify areas that are prone to near-future deforestation.

Three key criteria must be considered for effective and operational implementation of REDD+: the level of threat to service provision, the benefits, and the costs [7,8]. The decisive environmental service to consider when designing a REDD+ project on the ground is the reduction of GHG emissions [9]. Such a reduction has to be accretive, i.e., the progress must be achieved as part of the project and not be achievable without it, and it cannot result in leakage. To provide an advantageous cost-to-benefit ratio while ensuring accretion, it is important to locate projects in areas where forest carbon loss has already occurred or where significant deforestation and forest degradation are expected, i.e., areas where the level of threat to service provision is higher. Otherwise, no benefits will be generated, especially in countries in which business-as-usual projections show a low deforestation risk [10–12].

An accurate identification of deforestation risk requires, inter alia, data availability. Although the capacities of tropical non-Annex I countries to monitor forests and forest cover change are likely to improve over the next years, there are still a number of countries unable to implement the basic measures needed in the REDD+ context [13]. Hence, there is a lack of forest-related data for specific assessments of suitable REDD+ activity areas.

Several tools have been developed and used since the 1990s to simulate forest-clearing dynamics and to predict which areas are subject to the risk of losing carbon due to deforestation [14–19]. However, limited data availability can hamper their use in some developing countries [20–22]. In this paper we present a new approach based on the Random Forest algorithm [23]. Random Forest is a decision tree–based method belonging to the family of machine learning. Decision tree–based methods are used in decision-making processes because they enable evidence-based, data-driven decisions. Because a lack of data in developing countries may represent a barrier to the success of REDD+ projects, we adopted an approach that integrates a powerful machine learning technique (such as Random Forest), available geo-spatial layers, and easily accessible data sources. We call our model PREDIT (PREdicting Deforestation In the Tropics). Our approach attempts to overcome some of the current challenges in assessing locations of deforestation risk. Data from Nicaragua were used to evaluate the performance of our approach.

2. Materials and Methods

2.1. Random Forests

We selected a model approach that integrates inputs from different data layers using the Random Forest algorithm. Random Forest is a supervised technique, conceptually simple, and suitable for both regression and classification problems. Decision tree-based models recursively partition the entire dataset (i.e., all the predictors' possible attributes) into fairly homogeneous regions. In the terminology of tree models, such homogeneous regions are referred to as terminal nodes or leaves of the tree (Figure 1). When no further partitioning is required, the process of tree growing is concluded and the tree assigns a class to the dependent variable of interest. In machine learning terminology, it is said the decision tree 'votes' for a class.

While for regression trees the overall objective is to reduce the mean square error (i.e., the difference between the true value and the value predicted by the model), the objective of classification trees is to create nodes having a maximum homogeneity, also called the purity of the node. In fact, having impure nodes increases the probability of misclassification error. Node purity is expressed by the Gini index [24]:

$$G = \sum_{x=1}^{X} p_{mx}(1 - p_{mx}) \tag{1}$$

where p represents the training observations (i.e., those used to calibrate the model) of the X classes in the mth region. The lower the Gini index, the more homogeneous the node is, and consequently the probability to assign an incorrect class to a test observation belonging to that node decreases. Thus, during the process of tree growing, the best split is the alternative providing the lowest Gini index value.

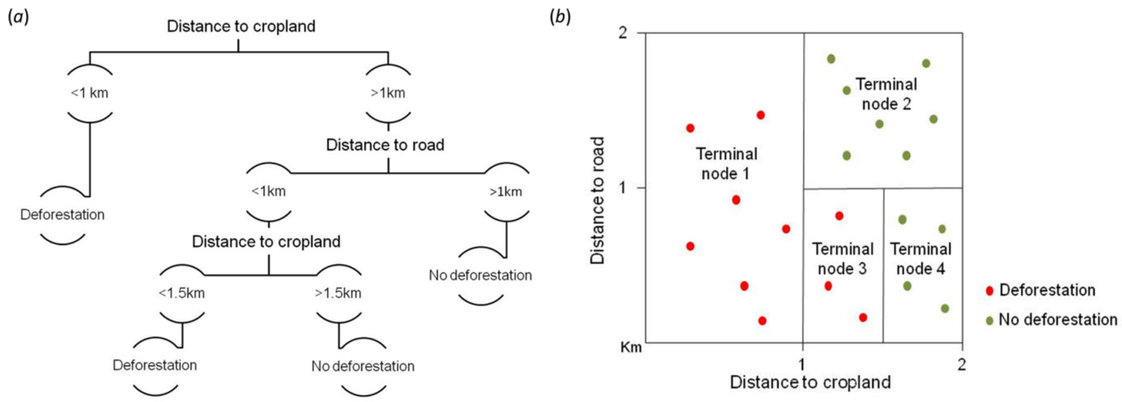

Figure 1. In this illustrative example land cover is predicted based on distance from cropland and road. Both pictures display the same decision problem using different representation systems: (**a**) shows the classification tree with three internal nodes and four terminal nodes; (**b**) shows the partition of the two-dimensional predictor space. In (**b**) the regions are entirely deforested or forested, a situation that is quite unusual in real-world models.

The random forests algorithm basically involves building a large number of classification trees on bootstrapped training samples. It is expressed by:

$$f(x) = \frac{1}{R}\sum_{r=1}^{R} f_r(x) \tag{2}$$

where $f(x)$ is the function of the dependent variable and R is the number of generated bootstrapped training trees, so $f_r(x)$ represents the rth bootstrapped training tree. This can also be expressed in a matrix form:

$$S_n = \begin{bmatrix} Pa_{S1} & Pb_{S1} & \ldots Pk_{S1} \\ \vdots & \vdots & \vdots \\ Pa_{Sn} & Pb_{Sn} & \ldots Pk_{Sn} \end{bmatrix} \tag{3}$$

where S_n is the nth bootstrap dependent variable, P_{aS1} is the predictor a of sample 1, P_{kS1} is the predictor k of sample 1, and $\{P_{aSn} \ldots P_{kSn}\}$ are the respective predictors $\{a \ldots k\}$ of the nth sample.

Finally, the class for which the greatest number of R individual training trees "vote" is used to predict the class for new observations that fall within the same region. Each prediction is expressed as a probability vector.

When using Random Forest, each decision tree is generated based on a random sub-sample (usually two-thirds) of the available observations. The remaining third of the data (not applied to calibrate the model) is called "out-of-bag" and serves as test data for computing an error rate. When trees are built, only a random subset of available k predictors is considered. Typically the number of predictors in the subset equals the square root of the total number of predictors. By defining $m \approx \sqrt{k}$ (where m is the number of predictors considered in a bootstrap sample), randomness is introduced into the tree-growing process, which assigns each predictor the same probability of being selected. This lowers the likelihood that stronger predictors will systematically affect the first split of the trees, a condition that would result in a series of highly similar and correlated trees [25]. The calibrated model that results is then used to predict the out-of-bag observations. The likelihood of classification error is therefore obtained from the out-of-bag estimation, which is an objective cross-validation-based accuracy estimate.

Some parameters must be defined by the user when calibrating Random Forests, including the total number of trees and the number of predictors sampled as candidates at each internal node. Another parameter is the "cut-off", which is the threshold value above which the probability of class

A (e.g., deforestation) occurring is predicted and below which the occurrence of class B (e.g., forest) is predicted.

The Random Forest algorithm is widely used for data mining in many fields, but has been used only relatively recently in ecology and environmental studies [26]. However, to the best of our knowledge, there has so far been no study of its use in predicting the risks of future deforestation. For our study a decision tree-based model was adopted because the underlying relationships between the variables are not linear, and the categorical scale of attributes used is suitable for tree-based evaluation methods [24].

2.2. Data Used and Variable Selection

We used available data sources from the time interval 1983–2011. We included data from t1–t2 (i.e., 1983–2000) to calibrate the model and data from t2–t3 (i.e., 2000–2011) for validation. The whole dataset used for this study included 11 maps of Nicaragua: (1–3) three land cover maps for 1983, 2000, and 2011 respectively; (4–5) two population density maps; (6) a map of protected areas; (7–8) two road network maps; (9) one map of the urban settlements; (10) an elevation map; and (11) a slope map (Table 1). The 1983, 2000, and 2011 land cover maps are referred to as the 't1 reference', 't2 reference', and 't3 reference' maps, respectively. The predicted map for the year 2011 is called 't3 simulated map'.

The t1 reference map has a scale of 1:250,000. It was derived from 1977 and 1978 Landsat images classified by a supervised classification technique and complemented with ground observations [27]. Details of the methods used for adding the ground-based observations to the t1 reference map are not at our disposal. Although the quality of the t1 reference map has some drawbacks in comparison with more recent maps, the t1 reference map is the main map used for estimating land-use changes in Nicaragua of the last 30 years. This map is used by Nicaraguan institutions for official reports and statistics (e.g., the readiness preparation proposal). The t2 reference map was derived from 20 satellite images: 17 Landsat TM5 from many different years, and 3 Landsat TM7 from 2000 [28]. The classification was carried out by the ISODATA (Iterative Self-Organizing Data Analysis Technique Algorithm) unsupervised classification algorithm. Remotely-sensed data were supplemented by 120 field-plot observations [29]. The t3 reference map was derived from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) satellite imagery. The satellite images, from 2011, were not ground-truthed; however, trained staff experienced in interpreting Nicaraguan vegetation from satellite images were employed in creating them [30]. The t1, t2 and t3 reference maps were officially released by the Ministry of Agriculture and Forestry (MAGFOR) and the Ministry of Environment and Natural Resources (MARENA) in Nicaragua.

Altitude and slope maps were extracted from the Digital Elevation Model (DEM) provided by HydroSHEDS [31]. The DEM has a resolution of three arc-seconds (approximately 90 m at the equator). Population density data were obtained from the Gridded Population of the World, Version 3, where density is expressed in terms of persons per square kilometre [32]. The World Database on Protected Areas [33], which is a global database of protected marine and terrestrial areas comprising both spatial vector data and attribute data (i.e., descriptive information), was used for data on protected areas.

Table 1. List of variables used to calibrate and validate the model.

Source Map	Data Format	Years Covered	Variable Extracted	Reference Unit	Sources
Land cover	Vector	1983	- Distance to pasture areas - Distance to cropland areas - Forest type - Forest density	Meters Meters Broadleaved/coniferous Closed forest/open forest	[27]
Land cover	Vector	2000	- Distance to pasture areas - Distance to cropland areas - Forest type - Forest density - Forest cover change	Meters Meters Broadleaved/coniferous Closed forest/open forest Forest/deforestation	[28]
Land cover Digital Elevation Model	Vector Raster	2011 -	- Forest cover change - Altitude	Forest/deforestation Meters above sea level	[30] [31]
Gridded Population of the World	Raster	1990, 2000	- Population density - Slope	Persons/km^2 Degrees	[32]
Protected areas	Raster	From 1980 to 2000	- Presence/absence of protected areas	Protected/No protected	[33]
Road network Urban settlement	Vector Vector	1983, 2000 -	- Distance to road - Distance to urban areas	Meters Meters	[34,35] [36]

Ten independent variables (also called predictors in this article) and one dependent variable were extracted from the maps listed in Table 1. In Random Forest terminology, independent variables are called predictors and dependent variables are called response variables. The dependent variable, derived from the three land cover maps, is categorical and consists of two classes: (i) forest area and (ii) deforested area. Data from the three land cover maps of Nicaragua (1983, 2000, and 2011) were integrated through a GIS polygon-overlay analysis; the maps were overlaid to obtain the 1983–2000 and 2000–2011 forest cover change maps. The nominally-scaled dependent variable, i.e., the category of the dependent variable of each pixel, was binary: it was defined either as "forest" or "deforestation," according to the changes observed, with forests is defined as land with an area of more than 0.5 ha, trees higher than 5 m, and canopy cover of more than 10% [37]. The widely accepted definition of deforestation as "a long-term or permanent conversion of land from forest use to other non-forest uses" was adopted [38].

The independent variables consist of 10 spatially explicit predictors of deforestation. The availability of data was the major limitation in selecting the predictors. The predictors were chosen by reviewing the available literature and according to the author's knowledge of the country. While information on deforestation drivers at the continental level was taken from Hosonuma et al. [39], national information on Nicaragua was derived from the National Forest Inventory [40] and other sources [41–44]. The 10 predictors (Table 1) were chosen on the basis of the supposed relevant drivers of deforestation. Although some of the selected predictors may not be relevant to Nicaragua, we decided to include all of them to test the responsiveness of the model.

The proximity variables were computed using the Euclidean distance from each feature to the closest pixel. Three out of the 10 predictor variables were extracted from the national land cover maps available for 1983, 2000, and 2011: forest type, distance to cropland, and distance to pasture. The other predictors, such as road network and urban settlements, were extracted from remotely-sensed imagery and from publicly available sources (see Table 1 for references).

2.3. Modeling Using 10 and Four Predictors

We performed PREDIT using two sets of predictor variables. The processes of calibration and validation—which are explained in the next two sections—were carried out twice, each time using a different number of predictors (Table 2). Based on our knowledge, and on the of Readiness Preparation Proposal (RPP) developed by the Nicaraguan Ministry of Environment and Natural Resources [43], four predictors were selected as the most useful for inclusion in the model for predicting deforestation dynamics. In the first run, referred to as FourPA (Four Predictors Alternative), we included only these predictors. In the second run, referred to as TenPA (Ten Predictors Alternative), the set of all 10 available predictors was used.

Table 2. Ten predictors used in PREDIT model. The model was run on two alternatives: (i) TenPA, which uses 10 predictors; (ii) FourPA, which uses four predictors substantially associated with the dependent variable.

Screened Predictor Variables	
Used in TenPA	Used in FourPA
Forest density	Altitude
Population density	Distance to cropland areas
Distance to cropland areas	Slope
Protected areas	Distance to pasture areas
Forest type	
Altitude	
Distance to roads	
Distance to urban areas	
Slope	
Distance to pasture areas	

2.4. Model Calibration

The analysis was run using the Random Forest package of R, version 3.2.1 (R Foundation for Statistical Computing, Vienna, Austria) [45] and ArcMap 10.2.2 (Esri, Redlands, CA, USA). Three hundred data trees were grown using the Random Forest. Two and three predictors were sampled as candidates at each internal node to calibrate the model for FourPA and TenPA, respectively. Since our aim was to evaluate the model's ability to predict the risk of deforestation in the time interval t2–t3 (i.e., 2000–2011), we used training data from t1–t2 (i.e., 1983–2000) to calibrate the model (Figure 2). In total, approximately 105,000 pixels were sampled in forest areas from the reference t1 map; the shortest distance allowed between any of them was 90 m. For each sample pixel, we extracted data from the maps of predictor variables at time point t1 (Figure 3) and from the land-use class at time point t2, but no data about forest types at t1 were collected. The model was calibrated using observations for the predictor variables in 1983 and with the assumption that those conditions affected the dependent variable in 2000. An important aspect considered during the calibration phase was the imbalance in the relative frequencies of the classes, i.e., if one class of the dependent variable has fewer observations compared to the other classes. This issue can significantly influence model results. Considering that the area covered by the dependent variable class 'deforestation' is much smaller than the class 'forest', we applied post-stratification based on the dependent variable at t2 to reduce this class imbalance.

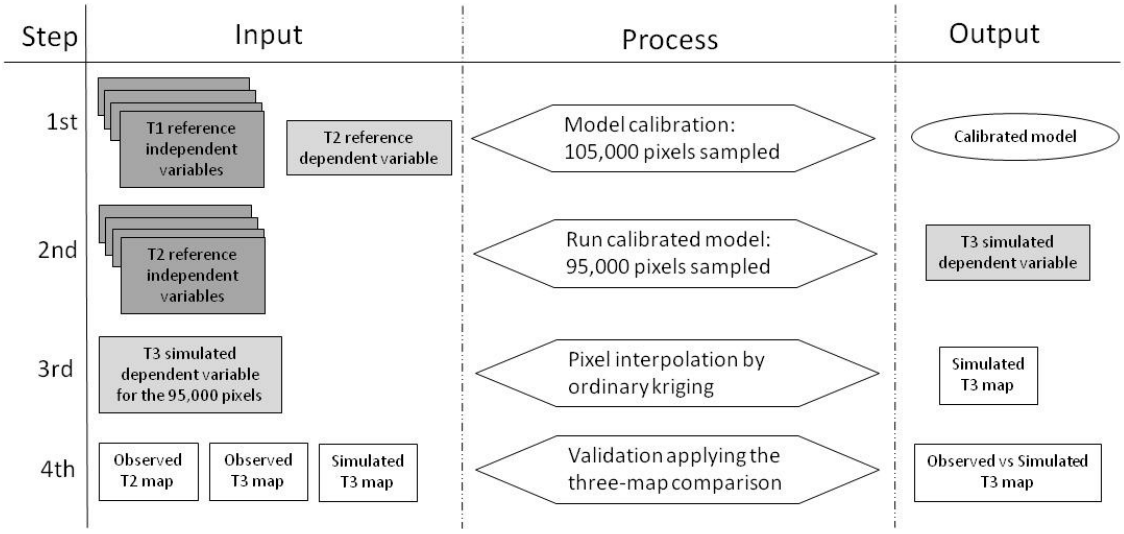

Figure 2. Summary of methods and data used in model building. The four steps were carried out twice: the first time including four and the second time 10 reference independent variables. Dark grey squares represent independent variable maps, light grey squares represent dependent variables, and white squares are maps of forest cover and forest cover change.

2.5. Model Validation

The validation process, which was carried out using data from the time interval t2–t3, included the following steps:

1. Approximately 95,000 sample pixels were generated by adopting a random sampling from the class "forest" at time t2.
2. For each pixel randomly selected, the corresponding value from every map of the independent variables at t2 was extracted.
3. The calibrated model and the fitted parameters used in the calibration procedure were used to predict the dependent variable at time t3 for the 95,000 pixels.
4. The t3-simulated map, which displays the predicted risk of deforestation, was created by interpolating the entire set of pixels using kriging.
5. The performance of the model was assessed by applying the three-map comparison technique and other statistical indicators [46].

Given that the aim of the procedure is to assess the risk of a pixel changing from "forest" to another land cover class, the validity of our model was assessed by random sampling of forest area at t2. Changes between t2 and t3 were estimated using the Random Forest algorithm and data fitted in the calibration phase. To predict the t3 map we used data preceding time t3. The predicted response of the dependent variable was expressed as the risk probability—ranging from 0 to 1—of each pixel in terms of undergoing deforestation. Risk is expressed in four probability classes—very low, low, moderate, and high. The probability threshold values that determine the risk class were applied as follows: very low ($p < 0.2$), low ($0.2 \leq p < 0.4$), moderate ($0.4 \leq p < 0.8$), and high ($0.8 \leq p < 1$). The thresholds were empirically derived considering the maximization of the overall accuracy of prediction for each risk class. In the analysis of the accuracy of the model, pixels belonging to the third and fourth classes of risk (i.e., moderate and high risk, respectively) are considered to be predicted as 'deforested'.

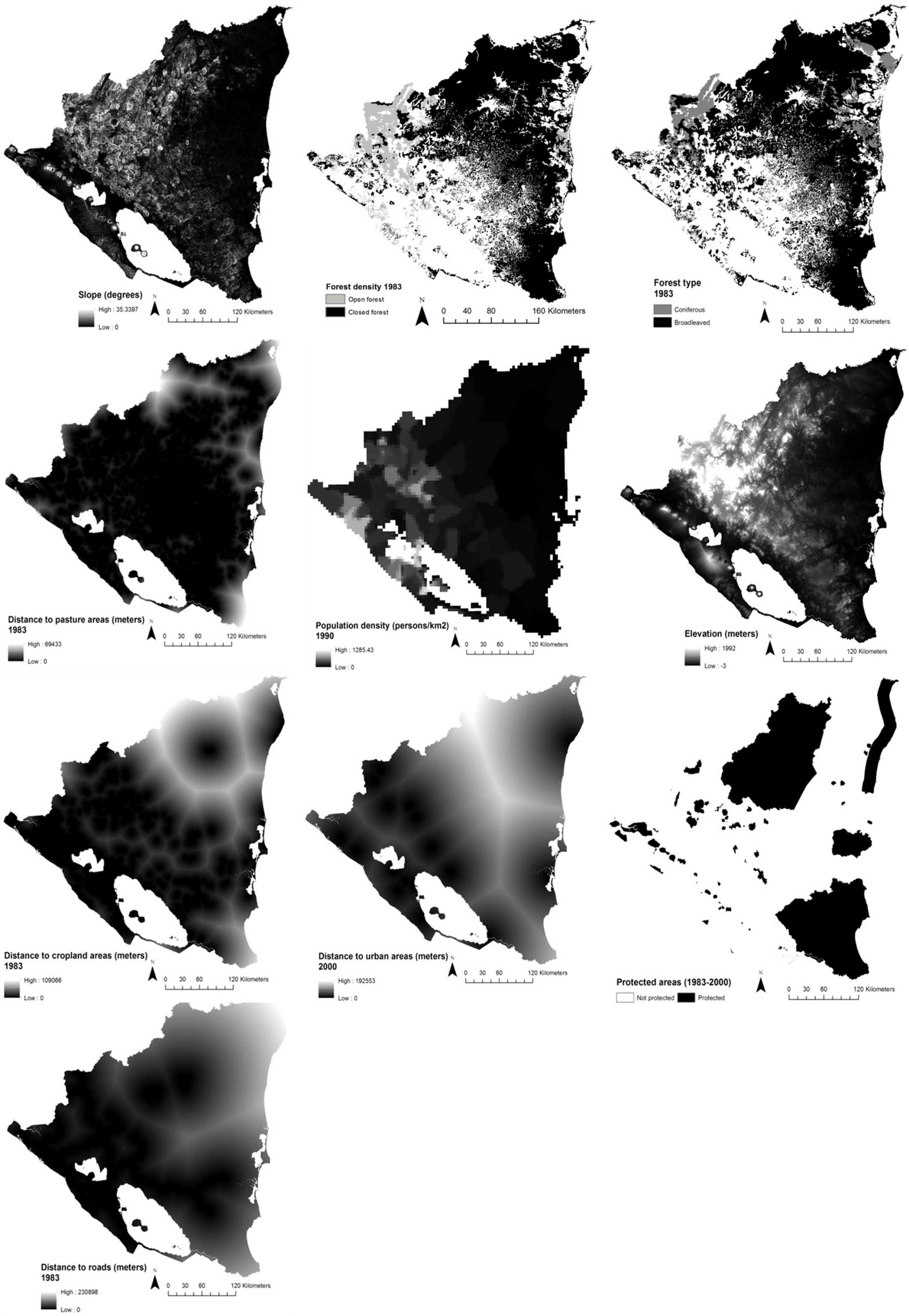

Figure 3. Spatial independent variables of Nicaragua used in the calibration of the model.

Ordinary linear kriging was applied to predict the risk of the dependent variable at a non-sampled location, i.e., for the whole forested areas in 2000 [47]. The simulated t3 map created using kriging has a pixel resolution of 150 m—higher resolution was not possible due to computational limitations. This map was validated using a technique that involves the overlay of three maps, which in this study were the t2 reference map, the t3 reference map, and the t3 simulated map. This three-map comparison [48] provides four components, two of which express correctness while the other two express prediction error: (i) reference deforestation correctly predicted (i.e., hits); (ii) reference forest permanence correctly predicted (i.e., correct rejections); (iii) reference change simulated as forest permanence (i.e., misses); (iv) reference forest permanence simulated as deforestation (i.e., false alarm). This validation technique compares the performance of the developed model with that of a null model that predicts pure persistence (i.e., no deforestation) [49]. Several methods based on these components describe and measure the performance of classification models. In this study we consider the following measures: figure of merit, allocation disagreement, quantity disagreement, producer's accuracy, and user's accuracy [47].

2.6. Study Area

Nicaragua is the largest Central American country, both in terms of land and rainforest area [41]. According to its national forest inventory, forests cover 25% of the total land area [40]. Despite some attempts to preserve its natural heritage, e.g., by establishing a number of protected areas over the past four decades [50], Nicaragua has lost almost half its forest cover since the 1950s and is still affected by deforestation, which has implications for local climate trends and agricultural productions [42,51].

The prime deforestation drivers in Nicaragua are animal husbandry and agriculture expansion, while agroforestry plays a minor role [43,52]. As in other Latin American countries such as Ecuador or Honduras, deforestation mainly follows an agricultural frontier, affecting considerable areas along the Caribbean and in the central north, which are still the regions with the largest stretches of natural forest. In September 2007, Hurricane Felix struck the northern region, or "Región Autónoma del Atlántico Norte"; over one million hectares of forests were affected and 512,165 ha were identified as strongly damaged, i.e., about 15% of the total forest cover of the country. PREDIT does not take into account climatic disturbances, and considering the extraordinarily large impact of Hurricane Felix, the three mostly affected municipalities were not included in the study area, i.e., Prinzapolka, Puerto Cabezas and Rosita.

3. Results

In all, 300 data trees were grown using the Random Forest algorithm. The maps in Figure 4 are the result of the three-map comparison technique (explained in Section 2.6); these maps visually represent model performances by showing the accuracy distribution of the land-change model. Areas characterized by a high risk of deforestation coincide with lowland forests, relatively gentle slopes, dense road networks and proximity to pasture and cropland areas.

Accuracy is also reported in numerical terms in Table 3, which represents the confusion matrix between the simulated land-use changes and the reference changes. The classification error for the two modeling alternatives was similar. The additional six predictors included in TenPA did not improve the overall accuracy, which in both alternatives was 76%. However, the predictions of the two alternatives were different, as described through the measures of accuracy considered in this study. The deforestation correctly predicted by FourPA was lower than that predicted by TenPA (26% versus 30.6%), but the number of "false alarms" (i.e., persisting forest predicted as deforestation) was also higher in TenPA (17.3% versus 10.8%) (Figure 5). This means that TenPA classified more pixels as deforested and, accordingly, detected more deforestation than FourPA. The larger number of pixels classified as "deforestation" by TenPA was also evidenced by calculating the producer's and user's accuracy; in fact, the proportion of pixels incorrectly classified as "deforestation" (as defined by the user's accuracy) was higher in TenPA (Table 4). Figure 5 shows the four components of agreement and disagreement resulting from the validation process; values are expressed as the percentage of forest at t2 and are divided by classes.

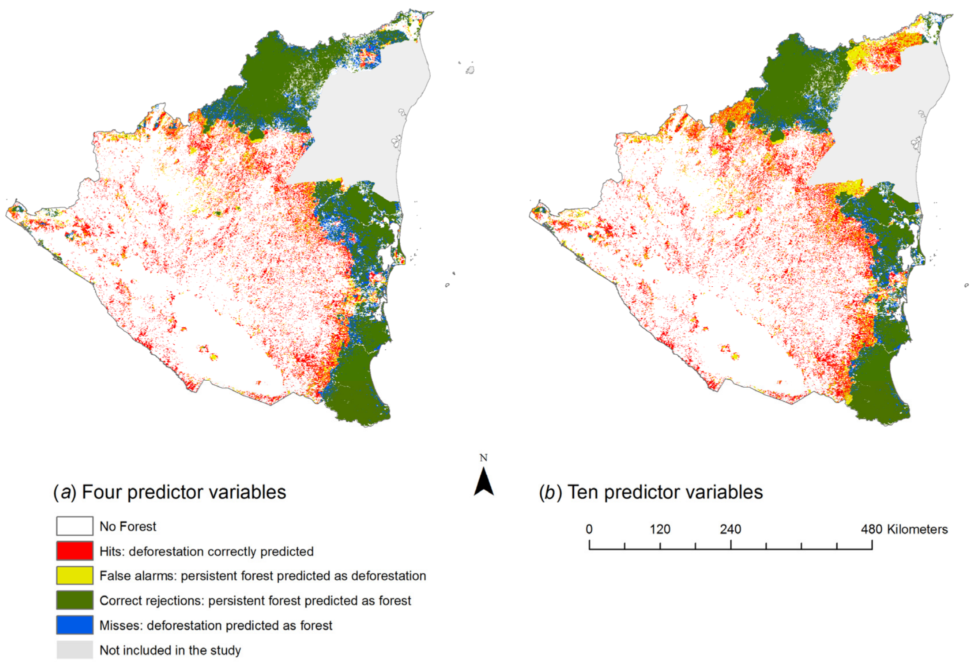

Figure 4. Distribution of agreement and disagreement for Four Predictors Alternative (**a**) and Ten Predictors Alternative (**b**), resulting from the comparison of three maps: the t2 reference map, the t3 reference map, and the t3 simulated map.

Table 3. Error matrix obtained by the three-map comparison for FourPA and TenPA. Values are expressed in percentages.

	Reference					
	TenPA			FourPA		
	Forest	Deforestation	Simulated Total	Forest	Deforestation	Simulated Total
Forest	44.7	7.5	52.2	51.2	12	63.2
Deforestation	17.3	30.5	47.8	10.8	26	36.8
Reference Total	62	38	100	62	38	100

Table 4. Performance of the model in assessing the risk of future deforestation using a set of 10 predictors and using four out of 10 predictors.

	Ten Predictors (TenPA)	Four Predictors (FourPA)
Overall accuracy	76%	76%
Producer's accuracy	0.80	0.69
User's accuracy	0.64	0.71
Figure of merit	55%	53%

The performances obtained from the two variations (FourPA and TenPA) are summarized in Table 4, which shows the measures of the model accuracy assessment derived from the error matrix. The quantity of disagreement was obtained by counting the total simulated pixels that did not match their actual category in the reference maps. The quantity of disagreement accounted for 3% (FourPA) and 25.8% (TenPA) of the total number of pixels classified as observed deforestation in 2011. The higher value of the quantity of disagreement for TenPA is attributable to the overestimation of deforestation by this alternative (depicted in Figure 4 by the yellow-colored "false alarms"). For assessing the allocation disagreement, in addition to the absolute number of pixels, their spatial allocations in their respective categories was also considered. In other words, the allocation disagreement is the discordance between a pixel allocated into the simulated maps and a corresponding pixel in the reference maps. The 2011 map simulated with TenPA had a lower allocation disagreement than the map simulated with FourPA; the difference is attributable to the larger portion of deforestation correctly predicted by TenPA. In absolute terms, there were 674,001 total pixels in the reference map belonging to the category "deforestation"; the allocation disagreement was 383,606 and 264,220 for FourPA and TenPA, respectively.

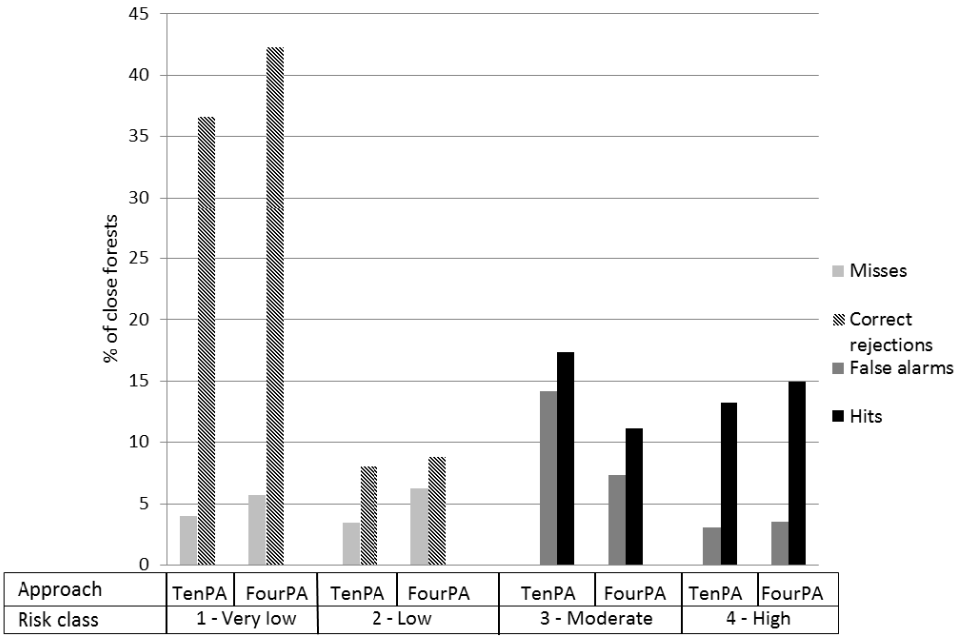

Figure 5. Elements of agreement and disagreement divided by the four classes of risk (i.e., very low, low, moderate and high). The accuracy components for both the modeling alternative with 10 (TenPA) and four (FourPA) predictors are reported for each risk class.

The "figure of merit" is a statistical measurement used to assess the accuracy of land change models; it can range from 0% to 1%, where 100% indicates perfect prediction. In this study it is defined as follows:

$$\text{Figure of Merit} = \text{Hits}/\text{False alarms} + \text{Misses} + \text{Hits} \quad (4)$$

The figure of merit was 55% for TenPA and 53% for FourPA. The 2% difference can be considered negligible in this case; it would be incorrect to judge the accuracy of prediction of one alternative with respect to the other on the basis of this percentage. To comprehensively assess the performance of the model, all the statistical measurements reported in this paper must be considered and evaluated with reference to the final scope of the modeling application. Including 10 predictors does not increase or decrease the accuracy significantly, though one important difference is that in TenPA, the number of "false alarms" is higher. In REDD+ projects, adopting a conservative approach that does not overestimate the risk of deforestation is recommended; therefore, FourPA would be the preferred alternative.

4. Discussion

The risk of deforestation is generally assessed using data about the respective drivers of deforestation. Addressing them involves understanding the complex processes affecting interrelationships among political, institutional, economic, and cultural factors [53]. We present here a novel approach, called PREDIT, which is based on available data and which focuses on building business-as-usual scenarios to predict potentially suitable areas for REDD+ interventions. A strong point of this approach lies in its versatility and potential reproducibility in countries with limited available data or human, technical, or monetary resources. The approach does not focus on capturing the interrelationships of multiple drivers underlying the deforestation processes, which would be both time- and cost-intensive. However, it is an objective and rapid way of checking on potential future forest cover by using available geo-spatial information as a proxy for stakeholder activities.

We applied Random Forests which outperform classical methods (e.g., discriminant analysis or logistic regression) when there are strong interactions among variables, especially if they are non-linear [54]. Spatial autocorrelation, a problem common to parametric linear models, is reduced by the Random Forest algorithm (which is non-parametric). Based on a machine learning technique, PREDIT is highly flexible as it can handle categorical and continuous variables. Flexibility is an appreciable characteristic in spatial prediction models [25,55], though flexibility comes at the expense of interpretability. We are, however, not interested in making inferences or in creating a model that investigates and displays relationships among the dependent variable and the set of predictors for which interpretability is of significance; our goal is a pragmatic prediction of future events.

Performing the process twice, using two sets of predictors, also allowed us to screen the most relevant biophysical and demographic predictors affecting deforestation in Nicaragua. Considering our purposes, using four predictors (FourPA) provided better results than using 10 (TenPA), even though quantity disagreement was higher for FourPA. We seek to accurately identify areas where both deforestation and forest persistence might occur in the future. However, if the study had been aimed at predicting total carbon emissions, without referring to the location where they might occur, then knowing the allocation disagreement (i.e., the area where emissions would take place) would not have been fundamental. The reason including 10 predictors did not improve the prediction accuracy is likely related to the fact that the six predictors added in TenPA were not relevant drivers of deforestation in Nicaragua during the considered time interval. However, some drivers of deforestation can change over time, so a good predictor in 1983 might be a bad predictor in 2000. To analyze and explain all the potential reasons related to the different performances of the two alternatives (FourPA and TenPA), a comprehensive investigation of the land-use dynamics from 1983 to 2011 should be carried out, though such analysis goes beyond the scope of this paper.

To streamline the validation phase, we decided to categorize the probability of deforestation into four classes, though this also meant losing some information concerning risk. The threshold probability values assigned to determine the risk classes are subjective and should be evaluated by considering the final application of the model. For example, if the model is used to project a business-as-usual scenario to build reference levels (RLs), a conservative estimate should be used to minimize the possibility of overestimating deforestation. However, the creation of baseline scenarios useful to preparing RLs involves a series of complex and—often—"stochastic" circumstances. These circumstances increase uncertainties and can undermine the credibility and effectiveness of the mechanism [56]. In order to minimize uncertainties, improve accuracy, and provide completeness, field assessments are strongly recommended when preparing RLs.

We calibrated the model with data covering a time interval of 17 years (1983–2000), during which crucial political, social, and natural events took place in Nicaragua that led to extraordinary land-use changes. Nonetheless, the calibrated model exhibited satisfactory predictive accuracy within its domain of applicability; we expect that applying it in a less unstable context could yield better results. It must be underlined that predictive accuracy is only one of the various criteria used to assess the performance of a land change model and its potential applicability. To evaluate the actual applicability and the predictive performance of PREDIT to another area, several other environmental, political, economic and technical factors that might influence the performance of PREDIT [22] must be considered.

PREDIT has some limitations that could compromise its accuracy. Deforestation drivers may vary over a long time period, e.g., new driving forces not relevant in the calibration phase could, at a later time, become more important, and as a result the predictor variables used might lead to an error in predicting the location of future forest loss. PREDIT also does not predict the risk of deforestation for reforested and afforested areas.

The agreement between the simulated map and the reference map can be considered satisfactory for some payment for ecosystem services purposes, e.g., when identifying potential target areas for REDD+ projects. However, the model may perform differently depending on the location, time, and format of the data [46]. Nicaragua experienced a high deforestation rate for the period considered in this study; this situation might have facilitated a predictive model. In fact, recording small deforestation patches is far more challenging than detecting substantial changes of forest cover [57]. Thus it will be important to also test the model in areas where forest cover changes affect smaller areas.

5. Conclusions

Modeling deforestation is a key first step towards creating the basis for successful REDD+ initiatives, although it is, of course, only one of the numerous circumstances that determine whether a REDD+ project will be effective or not [58,59]. Our research was prompted by the necessity of forecasting the likelihood of deforestation, without reference to further complex assessments such as field measurements, social surveys, and stakeholder involvement. Predictive models focus on the general network of interaction among variables rather than investigating the roles and relationships of each one. This is different from how inference models work; with those models, the main objective is to understand how a dependent variable changes as a function of the independent variables [24,60].

PREDIT can be applied by decision-makers, researchers, and other stakeholders involved in REDD+. Besides its direct use in determining areas that risk losing their ecological importance, it has further potential functions. Using the model approach jointly with tools for estimating carbon stock and emissions will allow the identification of high-value areas where activities advocating forest monitoring should be strengthened. Significantly, the model can support countries involved in the early phases of REDD+. As developing a REDD+ program requires, inter alia, high-quality data, it is hoped that countries lacking data and technical capacity can adopt a gradual approach to engaging in REDD+ [22,61]. The approach presented in this article is suited to a stepwise framework for developing REDD+ [62]. It can aid in the operational implementation of REDD+ projects and in the design of action responses. Its adoption may be effective in the first phase of projects, when a country is still developing the capacity to build its own sound and accurate dataset. Adopting PREDIT or other similar tools would enable wider REDD+ participation; it represents a starting point for countries that grapple with a critical lack of data, higher uncertainties, and competing interests.

Acknowledgments: This paper utilized forest cover data prepared by the National Forest Inventory of INAFOR, Nicaragua. We thank the Department of National Forest Inventory of the National Forest Institute in Nicaragua. We also thank the Government of Nicaragua for making the data available for this study. Thanks to the University of Hamburg which provided funds for covering the costs to publish in open access.

Author Contributions: G.D.L. designed the study, developed the model, analyzed and interpreted the data, drafted the manuscript, and revised it critically; P.M., who conceived the application of Random Forest, also provided important technical content and revised the manuscript; S.E.Z.L. helped to identify the case study, contributed to the interpretation of data, and revised the manuscript; M.M. contributed to the early conception of the idea, helped to coordinate the study, and revised the manuscript; M.K. conceived the idea and helped to structure the study, helped to develop the method, coordinated the study, and substantially contributed to the quality control and critical revision of the manuscript. All the authors have read and approved the final manuscript.

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

References

1. Intergovernmental Panel on Climate Change. Summary for Policymakers. In *Climate Change 2013: The Physical Science Basis*; Contribution of Working Group I to the Fifth Assessment Report of the IPCC; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013.

2. Bonan, G.B. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science* **2008**, *320*, 1444–1449. [CrossRef] [PubMed]
3. Van der Werf, G.R.; Morton, D.C.; DeFries, R.S.; Olivier, J.G.; Kasibhatla, P.S.; Jackson, R.B.; Collatz, G.J.; Randerson, J.T. CO_2 emissions from forest loss. *Nat. Geosci.* **2009**, *2*, 737–738. [CrossRef]
4. Intergovernmental Panel on Climate Change. Carbon and Other Biogeochemical Cycles. In *Climate Change 2013: The Physical Science Basis*; Contribution of Working Group I to the Fifth Assessment Report of the IPCC; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013.
5. United Nations Framework Convention on Climate Change. *Report of the Conference of the Parties on Its Sixteenth Session, Held in Cancun from 29 November to 10 December 2010 Addendum Part Two: Action Taken by the Conference of the Parties 2007 at Its Thirteenth Session*; Distributed 15 March 2011; United Nations Office at Geneva: Geneva, Switzerland, 2011.
6. Streck, C. Financing REDD+: Matching needs and ends. *Curr. Opin. Environ. Sustain.* **2012**, *4*, 628–637. [CrossRef]
7. Lin, L.; Sills, E.; Cheshire, H. Targeting areas for Reducing Emissions from Deforestation and Forest Degradation (REDD+) Projects in Tanzania. *Glob. Environ. Chang.* **2014**, *24*, 277–286. [CrossRef]
8. Wendland, K.J.; Honzák, M.; Portela, R.; Vitale, B.; Rubinoff, S.; Randrianarisoa, J. Targeting and implementing payments for ecosystem services: Opportunities for bundling biodiversity conservation with carbon and water services in Madagascar. *Ecol. Econ.* **2010**, *69*, 2093–2107. [CrossRef]
9. Kim, O.S. An Assessment of Deforestation Models for Reducing Emissions from Deforestation and Forest Degradation (REDD): An Assessment of Deforestation Models for REDD. *Trans. GIS* **2010**, *14*, 631–654. [CrossRef]
10. Köhl, M.; Baldauf, T.; Plugge, D.; Krug, J. Reduced emissions from deforestation and forest degradation (REDD): A climate change mitigation strategy on a critical track. *Carbon Balance Manag.* **2009**, *4*, 10. [CrossRef] [PubMed]
11. Harris, N.L.; Petrova, S.; Stolle, F.; Brown, S. Identifying optimal areas for REDD intervention: East Kalimantan, Indonesia as a case study. *Environ. Res. Lett.* **2008**, *3*, 035006. [CrossRef]
12. Plugge, D.; Baldauf, T.; Köhl, M. The global climate change mitigation strategy REDD: Monitoring costs and uncertainties jeopardize economic benefits. *Clim. Chang.* **2013**, *119*, 247–259. [CrossRef]
13. Romijn, E.; Lantican, C.B.; Herold, M.; Lindquist, E.; Ochieng, R.; Wijaya, A.; Murdiyarso, D.; Verchot, L. Assessing change in national forest monitoring capacities of 99 tropical countries. *For. Ecol. Manag.* **2015**, *352*, 109–123. [CrossRef]
14. Boyd, D.S. A review of *"Modelling Deforestation Processes:"* A review. Trees Series B Report n° l. By E.F. Lambin. *Int. J. Remote Sens.* **1996**, *17*, 1061–1062. [CrossRef]
15. Dudley, R.G. A little REDD model to quickly compare possible baseline and policy scenarios for reducing emissions from deforestation and forest degradation. *Mitig. Adapt. Strateg. Glob. Chang.* **2010**, *15*, 53–69. [CrossRef]
16. Gutierrez-Velez, V.H.; Pontius, R.G. Influence of carbon mapping and land change modelling on the prediction of carbon emissions from deforestation. *Environ. Conserv.* **2012**, *39*, 325–336. [CrossRef]
17. Jong, B.H.J.; Hellier, A.; Castillo-Santiago, M.A.; Tipper, R. Application of the "Climafor" Approach to Estimate Baseline Carbon Emissions of a Forest Conservation Project in the Selva Lacandona, Chiapas, Mexico. *Mitig. Adapt. Strateg. Glob. Chang.* **2005**, *10*, 265–278. [CrossRef]
18. Veldkamp, A.; Fresco, L.O. CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecol. Model.* **1996**, *91*, 231–248. [CrossRef]
19. Kamusoko, C.; Oono, K.; Nakazawa, A.; Wada, Y.; Nakada, R.; Hosokawa, T.; Tomimura, S.; Furuya, T.; Iwata, A.; Moriike, H.; et al. Spatial Simulation Modelling of Future Forest Cover Change Scenarios in Luangprabang Province, Lao PDR. *Forests* **2011**, *2*, 707–729. [CrossRef]
20. Aguilar-Amuchastegui, N.; Forrest, J.L. A Review of Tools and Approaches to Compare Business-as-Usual to REDD+ Scenarios. Available online: http://wwf.panda.org/?209271/REDD-MRV-tools (accessed on 20 January 2017).
21. Aguilar-Amuchastegui, N.; Riveros, J.C.; Forrest, J.L. Identifying areas of deforestation risk for REDD+ using a species modeling tool. *Carbon Balance Manag.* **2014**, *9*, 10. [CrossRef] [PubMed]

22. Huettner, M.; Leemans, R.; Kok, K.; Ebeling, J. A comparison of baseline methodologies for "Reducing Emissions from Deforestation and Degradation". *Carbon Balance Manag.* **2009**, *4*, 4. [CrossRef] [PubMed]
23. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
24. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer Texts in Statistics; Springer: New York, NY, USA, 2013; Volume 103.
25. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
26. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]
27. Ministerio Agropecuario y Forestal. *Cobertura Forestal, Shapefile, Scale 1:250000*; Ministerio Agropecuario y Forestal: Managua, Nicaragua, 1983.
28. Ministerio Agropecuario y Forestal. *Cobertura Forestal, Shapefile, Scale 1:250000*; Ministerio Agropecuario y Forestal: Managua, Nicaragua, 2000.
29. Ministerio Agropecuario y Forestal. *Valoración Forestal Nicaragua, 2000*, 1st ed.; Ministerio Agropecuario y Forestal: Managua, Nicaragua, 2005.
30. Ministerio del Ambiente y Recursos Naturales-Ministerio Agropecuario y Forestal. *Mapa de Uso Actual, Shapefile, Scale 1:250000*; Ministerio del Ambiente y Recursos Naturales-Ministerio Agropecuario y Forestal: Managua, Nicaragua, 2011.
31. Lehner, B.; Verdin, K.; Jarvi, A. *HydroSHEDS Technical Documentation*; World Wildlife Fund US: Washington, DC, USA, 2006.
32. Center for International Earth Science Information Network; Columbia University; Centro Internacional de Agricultura Tropical. *Gridded Population of the World, Version 3 (GPWv3): Population Density Grid*; NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA, 2005.
33. International Union for Conservation of Nature and Natural Resources; United Nations Environment Programme World Conservation Monitoring Centre; United Nations Environment Programme. *The World Database on Protected Areas*; World Conservation Monitoring Centre: Cambridge, UK, 2014.
34. Instituto Nacional de Estudios Territoriales. *Red Nacional de Caminos (Shapefile)*; Instituto Nacional de Estudios Territoriales: Managua, Nicaragua, 1983.
35. Instituto Nacional de Estudios Territoriales. *Red Nacional de Caminos (Shapefile)*; Instituto Nacional de Estudios Territoriales: Managua, Nicaragua, 2000.
36. Instituto Nacional de Estudios Territoriales. *Ciudades y Comarcas (Shapefile)*; Instituto Nacional de Estudios Territoriales: Managua, Nicaragua, 2000.
37. Food and Agricultural Organization. *Global Forest Resources Assessment 2000*; Food and Agricultural Organization: Rome, Italy, 2000.
38. United Nations Framework Convention on Climate Change. Report of the Conference of the Parties serving as the meeting of the Parties to the Kyoto Protocol on its first session, held at Montreal from 28 November to 10 December 2005 Addendum. Part Two: Action taken by the Conference of the Parties serving as the meeting of the Parties to the Kyoto Protocol at its first session. Available online: http://unfccc.int/resource/docs/2005/cmp1/eng/08a04.pdf (accessed on 20 January 2017).
39. Hosonuma, N.; Herold, M.; De Sy, V.; De Fries, R.S.; Brockhaus, M.; Verchot, L.; Angelsen, A.; Romijn, E. An assessment of deforestation and forest degradation drivers in developing countries. *Environ. Res. Lett.* **2012**, *7*, 44009. [CrossRef]
40. Instituto Nacional Forestal. *Resultados del Inventario Nacional Forestal: Nicaragua 2007–2008*; Instituto Nacional Forestal: Managua, Nicaragua, 2009.
41. Zeledon, E.B.; Kelly, N.M. Understanding large-scale deforestation in southern Jinotega, Nicaragua from 1978 to 1999 through the examination of changes in land use and land cover. *J. Environ. Manag.* **2009**, *90*, 2866–2872. [CrossRef] [PubMed]
42. Gourdji, S.; Läderach, P.; Valle, A.M.; Martinez, C.Z.; Lobell, D.B. Historical climate trends, deforestation, and maize and bean yields in Nicaragua. *Agric. For. Meteorol.* **2015**, *200*, 270–281. [CrossRef]
43. Ministerio del Ambiente y Recursos Naturales. *Readiness Preparation Proposal, Formal Version 7*; Ministerio del Ambiente y Recursos Naturales: Managua, Nicaragua, 2013.
44. Redo, D.J.; Grau, H.R.; Aide, T.M.; Clark, M.L. Asymmetric forest transition driven by the interaction of socioeconomic development and environmental heterogeneity in Central America. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8839–8844. [CrossRef] [PubMed]

45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
46. Pontius, R.G.; Boersma, W.; Castella, J.-C.; Clarke, K.; de Nijs, T.; Dietzel, C.; Duan, Z.; Fotsing, E.; Goldstein, N.; Kok, K.; et al. Comparing the input, output, and validation maps for several models of land change. *Ann. Reg. Sci.* **2008**, *42*, 11–37. [CrossRef]
47. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists (Statistics in Practice)*, 2nd ed.; Wiley: Chichester, UK, 2007.
48. Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [CrossRef]
49. Committee on Needs and Research Requirements for Land Change Modeling; Geographical Sciences Committee; Board on Earth Sciences and Resources; Division on Earth and Life Studies; National Research Council. *Advancing Land Change Modeling: Opportunities and Research Requirements*; National Academies Press: Washington, DC, USA, 2014.
50. Ministerio del Ambiente y Recursos Naturales; International Union for Conservation of Nature and Natural Resources. *Estado de la Gestion Compartida de Areas Protegidas en Nicaragua*, 1st ed.; IUCN Mesoamérica: San José, Costa Rica, 2005.
51. Food and Agricultural Organization. *Global Forest Resources Assessment*; Food and Agricultural Organization: Rome, Italy, 2015.
52. Maldidier, C. *Tendencias Actuales de la Frontera Agrícola en Nicaragua*; Nitlapan-UCA: Managua, Nicaragua, 1993.
53. Pacheco, P.; Aguilar-Støen, M.; Börner, J.; Etter, A.; Putzel, L.; Diaz, M.D.C.V. Landscape Transformation in Tropical Latin America: Assessing Trends and Policy Implications for REDD+. *Forests* **2010**, *2*, 1–29. [CrossRef]
54. Cutler, D.R.; Edwards, T.C., Jr.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*, 2783–2792. [CrossRef] [PubMed]
55. Griffith, D.A.; Peres-Neto, P.R. Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. *Ecology* **2006**, *87*, 2603–2613. [CrossRef]
56. Pirard, R.; Karsenty, A. Climate Change Mitigation: Should "Avoided Deforestation" Be Rewarded? *J. Sustain. For.* **2009**, *28*, 434–455. [CrossRef]
57. Sloan, S.; Pelletier, J. How accurately may we project tropical forest-cover change? A validation of a forward-looking baseline for REDD. *Glob. Environ. Chang.* **2012**, *22*, 440–453. [CrossRef]
58. Angelsen, A.; Rudel, T.K. Designing and Implementing Effective REDD + Policies: A Forest Transition Approach. *Rev. Environ. Econ. Policy* **2013**, *7*, 91–113. [CrossRef]
59. Patel, T.; Dhiaulhaq, A.; Gritten, D.; Yasmi, Y.; De Bruyn, T.; Paudel, N.; Luintel, H.; Khatri, D.; Silori, C.; Suzuki, R. Predicting Future Conflict under REDD+ Implementation. *Forests* **2013**, *4*, 343–363. [CrossRef]
60. Angelsen, A.; Ainembabazi, J.H.; Bauch, S.C.; Herold, M.; Verchot, L.; Hänsel, G.; Schueler, V.; Toop, G.; Gilbert, A.; Eisbrenner, K. *Testing Methodologies for REDD+: Deforestation Drivers, Costs and Reference Levels*; ECOFYS: London, UK, 2013.
61. United Nations Framework Convention on Climate Change. Report of the Conference of the Parties on its seventeenth session, held in Durban from 28 November to 11 December 2011 Addendum Part Two: Action taken by the Conference of the Parties at its seventeenth session. Available online: http://unfccc.int/resource/docs/2011/cop17/eng/09a01.pdf (accessed on 20 January 2017).
62. Herold, M.; Angelsen, A.; Verchot, L.V.; Wijaya, A.; Ainembabazi, J.H. A stepwise framework for developing REDD+ reference levels. In *Analysing REDD+: Challenges and Choices*; Angelsen, A., Brockhaus, M., Sunderlin, W.D., Verchot, L.V., Eds.; Center for International Forestry Research: Bogor, Indonesia, 2012; pp. 279–299.