

RESEARCH ARTICLE

A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon

Fabián Santos^{1*}, Valerie Graw², Santiago Bonilla^{1,3}

1 Research Center for the Territory and Sustainable Habitat, Universidad Tecnológica Indoamérica, Quito, Ecuador, **2** Center of Remote Sensing of Land Surfaces (ZFL), University of Bonn, Bonn, Germany, **3** Departament of Forest Engineering, E.T.S.I.A.M., Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, Córdoba, Spain

* fabian_santos@hotmail.com



OPEN ACCESS

Citation: Santos F, Graw V, Bonilla S (2019) A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon. PLoS ONE 14(12): e0226224. <https://doi.org/10.1371/journal.pone.0226224>

Editor: Zhenlong Li, University of South Carolina, UNITED STATES

Received: July 10, 2019

Accepted: November 21, 2019

Published: December 23, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0226224>

Copyright: © 2019 Santos et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data from MAE are available from: <http://suia.ambiente.gob.ec/web/suia/anexos-nivel-referencia> All data from IGM are available from: <http://www.geoportaligm.gob.ec/portal/index.php/cartografia-de-libre-acceso->

Abstract

The Tropical Andes region includes biodiversity hotspots of high conservation priority whose management strategies depend on the analysis of forest dynamics drivers (FDDs). These depend on complex social and ecological interactions that manifest on different space–time scales and are commonly evaluated through regression analysis of multivariate datasets. However, processing such datasets is challenging, especially when time series are used and inconsistencies in data collection complicate their integration. Moreover, regression analysis in FDD characterization has been criticized for failing to capture spatial variability; therefore, alternatives such as geographically weighted regression (GWR) have been proposed, but their sensitivity to multicollinearity has not yet been solved. In this scenario, we present an innovative methodology that combines techniques to: **1)** derive remote sensing time series products; **2)** improve census processing with dasymetric mapping; **3)** combine GWR and random forest (RF) to derive local variables importance; and **4)** report results based in a clustering and hypothesis testing. We applied this methodology in the northwestern Ecuadorian Amazon, a highly heterogeneous region characterized by different active fronts of deforestation and reforestation, within the time period 2000–2010. Our objective was to identify linkages between these processes and validate the potential of the proposed methodology. Our findings indicate that land-use intensity proxies can be extracted from remote sensing time series, while intercensal analysis can be facilitated by calculating population density maps. Moreover, our implementation of GWR with RF achieved accurate predictions above the 74% using the out-of-bag samples, demonstrating that derived RF features can be used to construct hypothesis and discuss forest change drivers with more detailed information. In the other hand, our analysis revealed contrasting effects between deforestation and reforestation for variables related to suitability to agriculture and accessibility to its facilities, which is also reflected according patch size, land cover and population dynamics patterns. This approach demonstrates the benefits of integrating remote sensing–derived products and socioeconomic data to understand coupled socioecological systems more from a local than a global scale.

escala-50k/ All data from Natural Earth are available from: <https://www.naturalearthdata.com/> All census data from INEC are available from: <https://www.ecuadrencifras.gob.ec/institucional/home/>

Funding: SENESCYT: data collection and analysis. Universidad Indoamérica: decision to publish and preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

1.1 Introduction

The Tropical Andes is a mountainous region at the base of the Andes ridge. Due to its altitudinal gradient, it is characterized by 23 ecoregions and 8 bioregions [1], and it provides important economic and ecological services to almost 40 million inhabitants [2]. The region is recognized as an endangered biodiversity hotspot of high conservation priority [3,4], where population growth and agriculture expansion [5,6] are the major driving forces of deforestation, contributing to potential impacts of climate change [7]. On the other hand, large-scale reforestation has been detected in some areas of Latin America [8], especially along old colonization fronts [9]. However, these areas are less studied or understood, and their role in forest recovery and restoration of important environmental services is ignored [10,11]. Therefore, analyzing forest dynamics drivers (FDDs), i.e., deforestation and reforestation, in the Tropical Andes is very important for conservation, climate change adaptation, and sustainability. This knowledge is decisive for countries like Ecuador, where most of the remaining native forests are located and deforestation rates have been the highest in South America for some years [12,13].

Forest dynamics are shaped by complex societal and ecological interactions, or drivers. Geist and Lambin [14] proposed a conceptual framework to facilitate the understanding of these drivers of land dynamics, classifying them as follows:

1. proximate causes (local level, direct agents);
2. underlying causes (different levels, socioeconomic processes); and
3. other causes (determined by environmental factors and social triggering events).

These drivers have been accepted by countries participating in Reducing Emissions from Deforestation and Forest Degradation (REDD+), but recent research has recognized that underlying causes are less frequently analyzed in Latin America [6,15]. Proximate causes are mostly identified through remote sensing-based techniques [16], while underlying causes can be more complex, as they rely on socioeconomic data. These data are frequently not available or reliable at the scale needed [17]. Moreover, impacts of globalization [18] and economic development [11,19] generate more complex scenarios.

In Ecuador, previous studies combined remote sensing products and socioeconomic data to identify FDDs. For instance, Southgate et al. [20] analyzed thematic cartography and census data in a regression analysis to highlight agricultural rents, spontaneous settlements, and land tenure insecurity as deforestation drivers in eastern Ecuador. Following a similar approach but adding survey data, Rudel et al. [9] discussed reforestation drivers observed among ethnic groups and their relationships between land-use practices, cultural backgrounds, and distance to roads in southern Ecuador. Later, Mena et al. [21] combined thematic cartography, census, and survey data in a spatial regression model to conclude that road accessibility and population density were the most important deforestation drivers in northern Ecuador. Similarly, Walsh et al. [22] identified that reforestation drivers were motivated by land security and distance to roads. More recently, Bonilla-Bedoya et al. [23] related deforestation processes to legal timber harvesting, road expansion, and poverty indices. From these studies, it can be observed that deforestation is not commonly associated with reforestation. In this paper, the evaluation of contrasting driving forces (e.g., population growth/decay, agricultural expansion/contraction) with regard to possible linkages defines the first research interest.

Processing of multivariate data for FDD analysis has made significant progress in recent years. For instance, advances in remote sensing and open access to satellite archives [24] have contributed to a better understanding of global land-cover and land-use changes [25]. As a

result, products derived from time series (e.g., spectral trends, class-level metrics) has been increasingly applied to explain driving forces, making it possible to identify direct drivers [26] or better understand ecosystem fragmentation [27]. Furthermore, censuses are common sources of socioeconomic data, while processing of these data is not common in FDD analysis of underlying causes. Obstacles such as boundary changes [28] and scale effects [29] are probably the most challenging, and different approaches have been proposed to solve them, including areal interpolation [30,31] and statistical modeling [32–34]. Among them, areal interpolation with dasymetric mapping is perhaps the most popular [35], as it can combine land-cover maps and census data to model population distribution more precisely than other methods [36]. Other advances are related to capturing the spatial variability of FDDs. In this regard, geographically weighted regression (GWR) [37] has been demonstrated to satisfy this objective [38,39], but is sensitive to local collinearity and can produce unreliable results [40]. Nonparametric algorithms such as random forest (RF) [41] have interesting applications for high-dimensional problems with correlated variables [42]. Moreover, implementation of RF with GWR has recently proposed [43] but further applications explaining variables relationships are yet to be evaluated. The design of an innovative methodology for analyzing FDDs using these techniques constitutes the second research interest of this paper.

As the Tropical Andes constitutes a complex mosaic of landscapes, a workflow to analyze its FDDs is presented in this paper. We conducted our research in the Northwestern Ecuadorian Amazon (NEA), a study area located in an altitudinal gradient that includes different bioregions and colonization fronts with heterogeneous socioeconomic settings. Our main objective was to explore a set of variable groups to observe how they influenced deforestation and reforestation rates in the NEA in 2000–2010. This period is known as the beginning of the dollarization and economic stabilization in Ecuador [44]. For this purpose, we designed and implemented an experimental methodology for FDD analysis that benefits from the novel techniques mentioned above. Two research questions guided our work:

- What are the theoretical and empirical implications of our experimental methodology in FDD analysis?
- What could be the linkages between the driving forces of deforestation and reforestation in the NEA during 2000–2010?

To answer these questions, we (i) explain how we calculated the forest change rates and time series-derived products, (ii) conduct dasymetric mapping for intercensal analysis, and (iii) briefly describe the variable groups before (iv) explaining our implementation of GWR and RF, together with a clustering and hypothesis test to summarize our results. The discussion considers the benefits and limitations of the proposed methodology and its contribution to the current knowledge of FDDs in the NEA.

1.2 Study area

The NEA covers 21,857 km² over an altitudinal gradient from 200 to 2,800 m.a.s.l. on the western slopes of the Andean Range (Fig 1). It includes 16 cantons (second-level administrative units in Ecuador), which are used in this research to identify specific zones in the NEA. According to Olson et al. [1], two ecoregions exist in the NEA: the Napo moist forests and the Eastern Cordillera real montane forest. The latter is of the highest conservation priority in Ecuador as it covers less than 33% of its original area [45]. Moreover, the NEA is characterized by extraordinary biodiversity, intense annual precipitation (1,500–4,500 mm), and a multitude of ecosystems [46]. Most of the soils are ferric, with low fertility and high aluminum toxicity, although volcanic and alluvial soils can be an exception [47,48]. Under these conditions, the

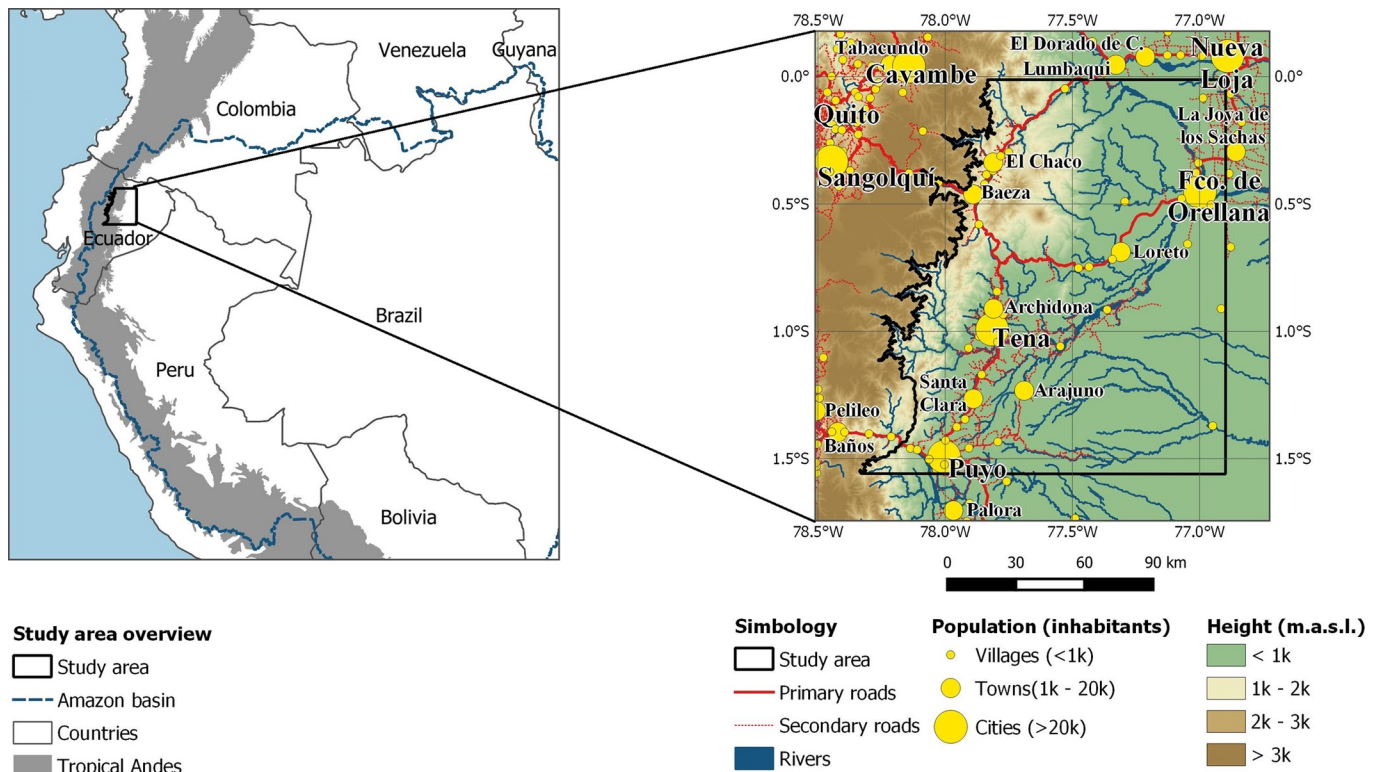


Fig 1. Study area and its location in the Amazon basin. Data from Natural Earth [52] and Instituto Geográfico Militar (IGM) [53].

<https://doi.org/10.1371/journal.pone.0226224.g001>

agricultural limitations are well known; however, this does not prevent the native people from co-evolving with their natural environment [49]. Dramatic changes began in the 1970s with the exploration and extraction of oil, generating accelerated economic growth and industrialization [50]. Extensive road construction and the Agrarian and Colonization Reform of 1964 stimulated in-migration and rapid settlement over the whole Ecuadorian Amazon. According to Brown et al. [51], its population grew by 432% from 1950 to 1990, resulting in an urban system that followed the discovery of petroleum and the related economic opportunities. This led to disorganized and arbitrary colonization where land conflicts between the *colonos* (mestizo colonists) and native people were common and traditional land-use practices were replaced by extensive agriculture and cattle ranching [55].

Forest clearing in the Ecuadorian Amazon peaked during 1970–1990, when the deforestation rate was one of the highest in South America [12]. In the NEA, the forested areas experienced an 19.6% reduction (4130 km²) by the end of 2014, principally due to pasture expansion for cattle ranching [56] (Fig 2). However, this was less intense than in the northeast NEA, where oil fields were located [57]. The declaration of protected areas, which accounted for 29% of the area and few oil discoveries [58], contributed to a reduced interest in colonization and to deforestation. Improved road connections between Quito and Nueva Loja and recent oil discoveries motivated further colonization of remote areas [59]. Despite this, reports indicate a drop of deforestation from 92,800 to 74,000 ha/year⁻¹ in Ecuador since 1990 [60].

Later, financial instability led to a crisis that ended with the dollarization of the Ecuadorian economy in September of 2000. A reduction in the inflation rate from 96 to 7% was seen as an important sign of economic stabilization for the period 2000–2014 [44]. As consequence, Ecuador experienced an unprecedented wave of emigration, especially between 2000 and 2007

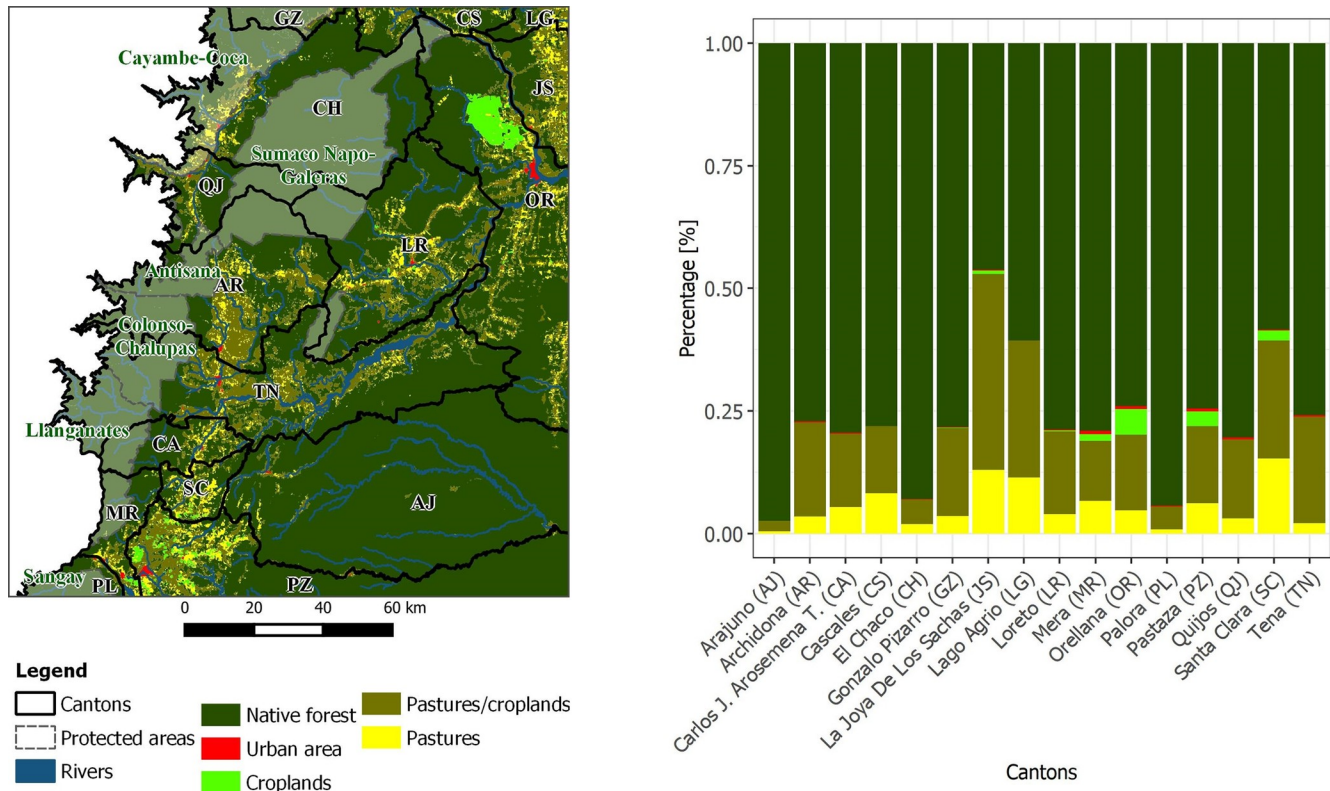


Fig 2. Land cover for 2008 in the study area. Data from Ministerio del Ambiente (MAE) [54].

<https://doi.org/10.1371/journal.pone.0226224.g002>

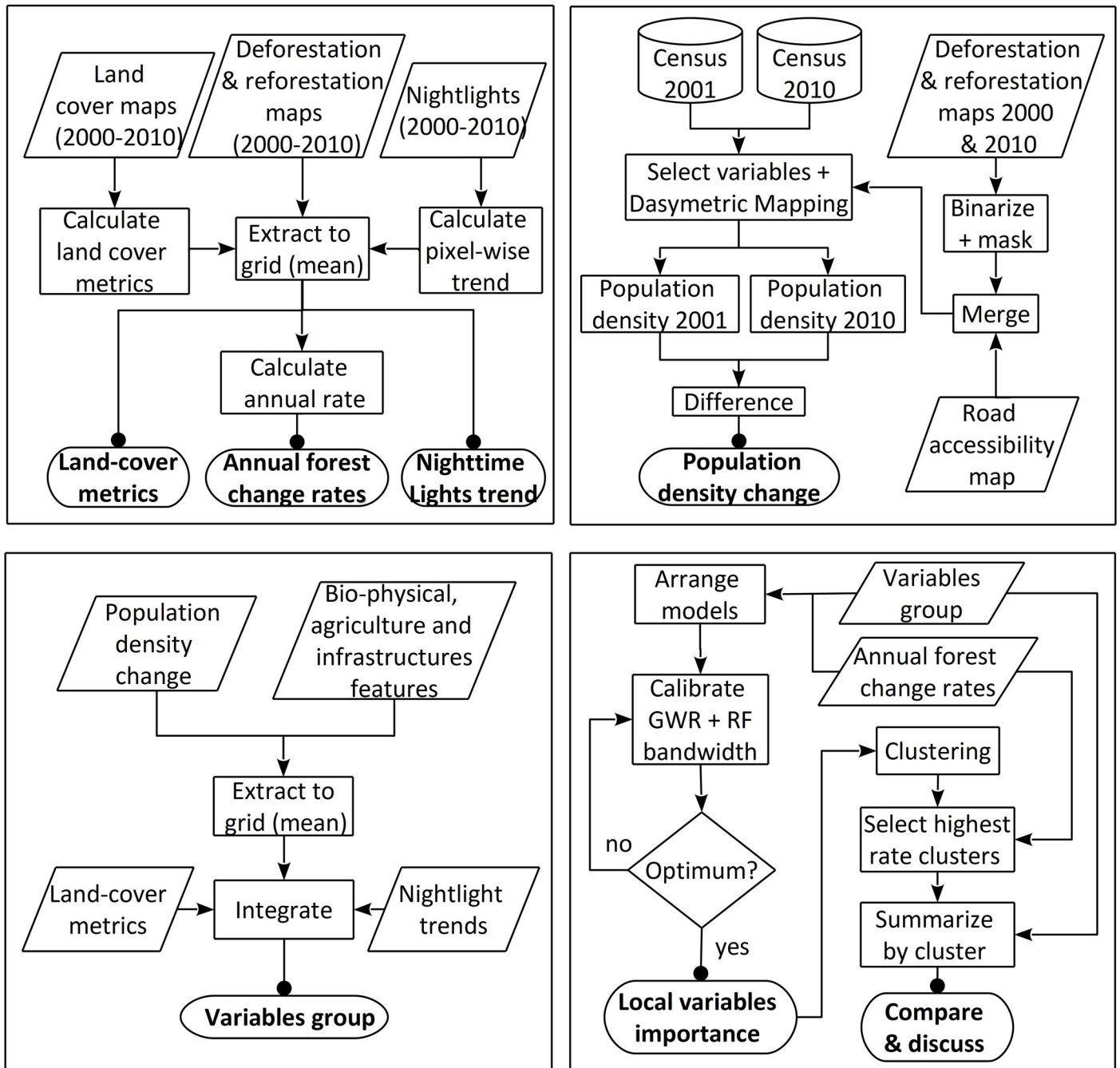
(around 483.000 migrants) [61]. Nevertheless, the effects of migration and remittances through land-use change have been associated with an increase of agriculture activities rather than land abandonment and forest transition in Ecuador [62].

1.3 Methods

This research was fully implemented using R programming language [63] and integrating specific libraries for spatial data [64,65], database management [66], parallel processing [67], and data visualization [68]. For mapmaking, we used QGIS 3.4.3 Madeira [69]. Fig 3 shows the workflow of the proposed methodology.

1.3.1 Annual forest change rates and remote sensing time series-derived products

We collected a set of land-cover and land-cover change maps generated biannually for the period 1990–2014 in the NEA. They were generated for previous research to monitor long-term forest dynamics with scarce data [70], reporting an overall accuracy above 70%. Specifically, this approach uses the Landsat surface reflectance time-series product [71] to reduce it into cloud-free biennial composites. Then, it trains and executes a supervised-classification algorithm to derive land-cover maps, classified into 4 classes: evergreen forest, bamboo forest (*guadua* spp.), bare soil/infrastructure, and pasture/cropland. This collection of maps is post-processed and the classes are aggregated into forest and non-forest binaries to derive deforestation and reforestation areas. In the case of deforestation, the algorithm simply flags the date of conversion from forest to non-forest, while for reforestation it first considers a minimum time



Legend

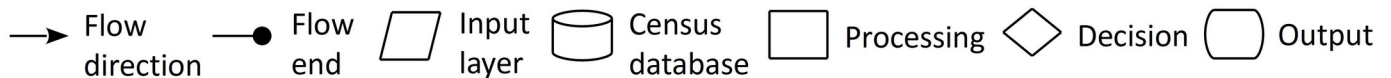


Fig 3. Flowchart of the methodology: (a) derivation of forest change rates and time-series products; (b) census processing with dasymetric mapping; (c) data integration; and (d) implementation of GWR and RF.

<https://doi.org/10.1371/journal.pone.0226224.g003>

classified as forest after a disturbance (i.e., 4 years) to flag it as reforestation if the area remains as forest until the end of the time series. To derive the annual forest change rate, we first

removed areas less than 1 ha from deforestation and reforestation maps, as they frequently represent misclassified pixels [26]. Then, the annual forest change rate q according to the UN Food and Agriculture Association (FAO) [72] was calculated, considering A_1 and A_2 as forest cover for time periods t_1 and t_2 :

$$q = \left(\frac{A_2}{A_1}\right)^{1/(t_2-t_1)} - 1$$

To operate this, we created an analysis grid with a cell size of 400 ha and extracted for each cell the deforested and reforested areas between 2000 and 2010. These years were selected to match the census years used in this research (2001 and 2010) as well as the cell size to reduce processing time during calculation. The resulting deforestation and reforestation rates constituted the set of dependent variables analyzed (Fig 4), a summary of which is shown in Table 1.

Furthermore, historical land use influence ecological landscape functions and link cause-connection patterns [73]. For this, we derived land-cover frequencies of deforested and reforested areas to pasture/cropland and bare soil/infrastructure classes. This provided further information about LCLUC dynamics and helped us to determine the permanence or semipermanence of a specific land-cover class Z . For this, we stacked the land-cover maps used in deforestation and reforestation mapping to obtain time series $M_{1:n}$, which was split to match deforestation or reforestation date i . This gave us 2 segments, defining the conditions (1) after a change $M_a = M_{i:n}$ and (2) before a change $M_b = M_{1:(i-1)}$. For deforestation, the segment M_a was used to determine the land-cover frequency f of class Z after a deforestation event D_{fz} . For

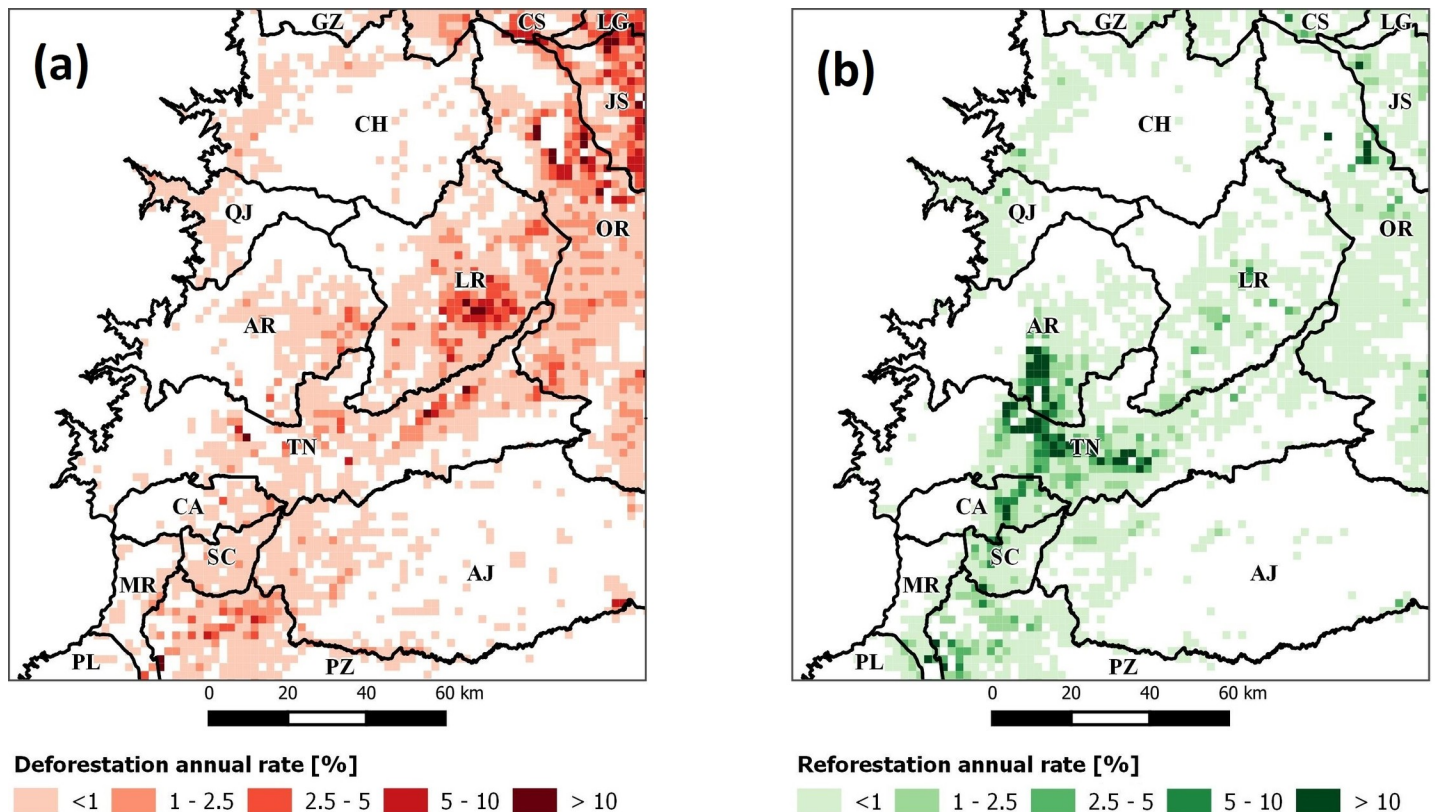


Fig 4. Annual forest change rates for: (a) deforestation and (b) reforestation. Data from Santos et al. (2018).

<https://doi.org/10.1371/journal.pone.0226224.g004>

Table 1. Descriptive statistics of dependent variables.

Variable	Prefix	q (%)		Analysis grid (no. cells)	Total area (ha)	Data source
		Mean	SD ²			
Annual deforestation rate ¹	DEF	-1.29	6.02	2418	967,200	[70]
Annual reforestation rate	REF	2.08	11.23	1998	799,200	

¹ To facilitate map reading, absolute values from deforestation rates were used.

² SD, standard deviation.

<https://doi.org/10.1371/journal.pone.0226224.t001>

this, we summed up the class occurrences in M_a and divided the sum by the extent of the segment:

$$D_{jz} = \frac{|\{Z \in M_a\}| * 100}{|M_a|}$$

For reforestation, segment M_b was used to determine the land-cover frequencies of class Z before a reforestation event R_{jz} happened. Similarly, it was calculated by adding up their occurrences in M_b and dividing the sum by the extent of the segment:

$$R_{jz} = \frac{|\{Z \in M_b\}| * 100}{|M_b|}$$

These calculations gave us 4 layers in total, describing the land-cover frequencies for pasture/cropland and bare soil/infrastructure as percentages, for both deforestation (Fig 5A) and reforestation.

A final time series-derived product was obtained from the Visible Infrared Imaging Radiometer Suite (VIIRS) and its Nighttime Lights Time Series cloud-free composites (nightlights 2000–2010) [74]. This dataset constitutes a measure of visible and near-infrared emission sources at night (e.g., cities, towns, gas flares, and other sources of persistent lighting), which can refer to access to electricity and human development factors such as: access to education [75], emissions of CO₂ [76] or socioeconomic trends [77]. The latter motivated its use in this research to enrich socioeconomic parameters. For this, we used Google Earth Engine [78] to calculate a pixel-wise linear trend map for 2000–2010 using the stable light band from this dataset. From the resulting pixel-wise linear model, we extracted its slope to determine its trend (Fig 5B).

1.3.2 Dasymetric mapping and population change

We processed the 2001 and 2010 population censuses published by the National Institute of Statistics and Census of Ecuador (INEC) [79,80] to generate population density surfaces with dasymetric mapping. This technique allows redistribution of population counts from a set of areal units into a grid using land-cover maps. To implement it, we extracted the most detailed level of census information (census blocks) to avoid aggregation and bias effects [29]. Moreover, since rural population better explains conversion from forest to agricultural land [81], we used census blocks from rural areas with an average size of $4,995 \pm 10,250$ ha and filtered the population by working age (15–72 years). Following Mennis [82], we adapted his dasymetric mapping with areal weighting approach to operate with rural populations. We first binarized deforestation maps from 2000 and 2010 to produce 2 non-forest masks to represent areas where rural populations were mostly settled in a specific year. Since other features not related to rural population were also present in these masks (e.g., water bodies, cliffs, urban areas), we

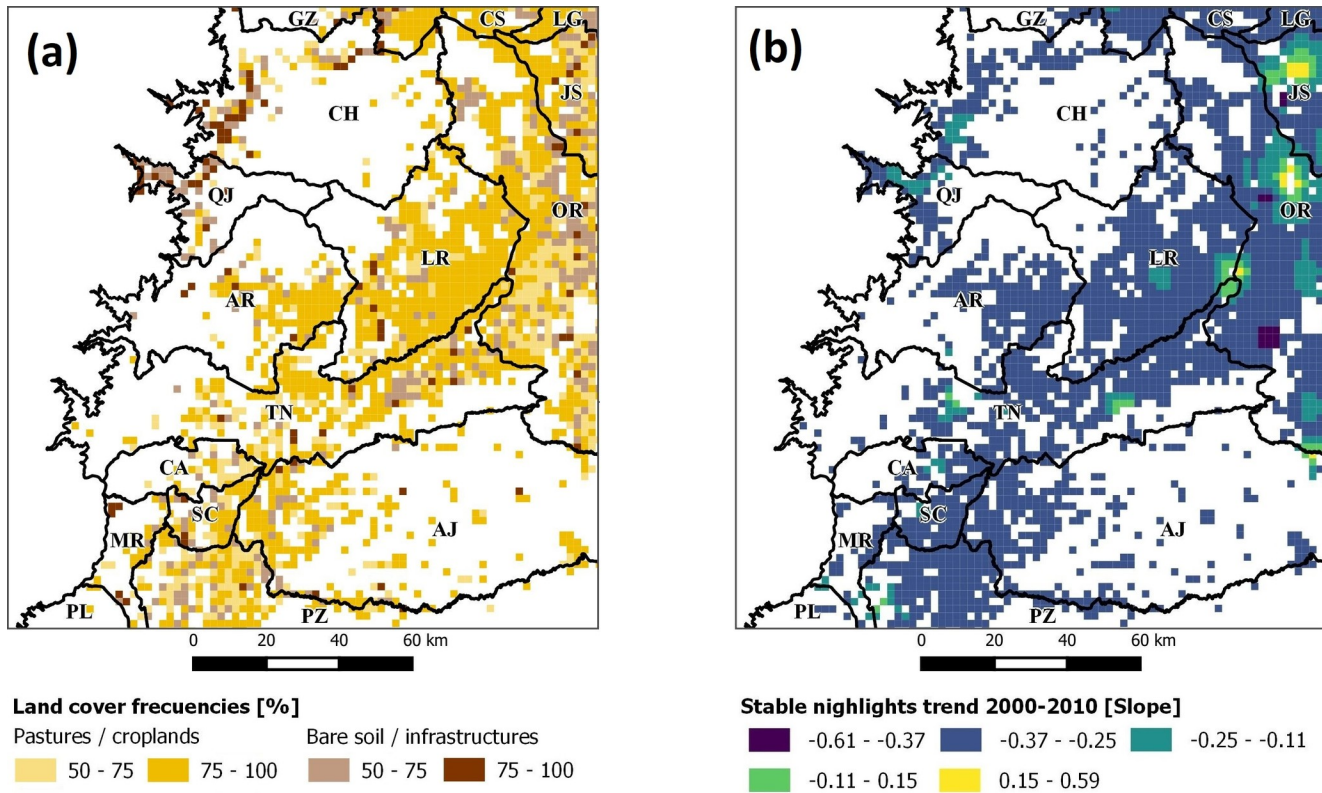


Fig 5. Time series-derived products for deforested areas: (a) land-cover frequencies and (b) stable nightlight changes.

<https://doi.org/10.1371/journal.pone.0226224.g005>

identified and removed them in order to be consistent with our focus on rural populations. As further details were required to locate rural populations, a road accessibility model [83] was added to these masks to facilitate their identification. We assumed that higher rural population density would occur in areas with better road accessibility [21,84]; therefore, we reclassified the road accessibility map into 3 travel time ranges (high: 0–1 h; medium: 1–3 h; low: >3 h) to obtain what is called rural density classes. We considered only 3 classes, as the algorithm proposed by Mennis [82] was not tested with more than 3. In the next step, we identified census blocks that were almost completely within each rural density class to calculate their densities (Table 2).

Table 2. Sampled census blocks and their population values.

Year	Locations of census blocks (cantons)	Rural density classes	Population (no. persons)	Area (ha)	Population density (persons/ha)	Sum density	Population density fraction
2001	LR ¹	High	36	4293	0.008	0.021	0.394
	AJ ²	Medium	34	3195	0.010	0.021	0.501
	AR ³	Low	23	10420	0.002	0.021	0.103
2010	LR	High	68	4297	0.015	0.030	0.745
	AJ	Medium	42	3269	0.012	0.030	0.604
	AR	Low	18	10389	0.001	0.030	0.081

¹ Refer to Loreto canton.

² Refer to Arajuno canton.

³ Refer to Archidona canton.

<https://doi.org/10.1371/journal.pone.0226224.t002>

We then averaged their population density fraction to obtain the next values: 0.570, 0.553, and 0.092, which correspond to high, medium, and low rural density classes, respectively. These values are dimensionless and are obtained by:

$$d_{uc} = \frac{p_{uc}}{p_{hc} + p_{mc} + p_{lc}}$$

where d_{uc} is the population density fraction of rural class u in census block c , and p_{uc} is the population density (persons/ha) of rural class u in census block c and is divided by the sum of all rural density classes (high h , medium m , and low l) and their census blocks. After calculating with all census blocks, the next step in the algorithm is to evaluate the area ratio. This operation divides the area of class u by 33.3% to adjust densities equally according to the difference in area of each rural density class within that census block. This can be expressed as:

$$a_{ub} = \frac{n_{ub}/n_b}{0.33}$$

where a_{ub} is the area ratio of rural class u and n_{ub} is the number of grid cells of rural class u in census block b , and n_b is the number of grid cells in census block b . The next step is to calculate the total fraction by multiplying d_{uc} and a_{ub} and dividing that result by the result of that same expression for all 3 rural classes (h, m, l) in that census block:

$$f_{ubc} = \frac{(d_{uc} * a_{ub})}{[(d_{hc} * a_{hb}) + (d_{mc} * a_{mb}) + (d_{lc} * a_{lb})]}$$

where f_{ubc} is the total fraction of rural class u in census block b of spatial unit c . The final step in the algorithm is to assign each rural class to grid cells within that census block. This was done by dividing the population assigned to the rural class evenly among the grid cells in the census block that has that rural class. This can be expressed as:

$$pop_{ubc} = \frac{(f_{ubc} * pop_b)}{n_{ub}}$$

where pop_{ubc} is the population assigned to one grid cell of rural class u in census block b of spatial unit c . The result is a population density surface (Fig 6), which represents the number of persons by pixel area (which was set as 1 ha to avoid census block elimination during vector-to-raster conversion). We iterated the algorithm with 20 census variables (see Section 1.3.3) to obtain population density surfaces. In all cases, the pycnophylactic property [85] was verified by adding population density surface pixels and comparing them with their original values. Only incomplete census blocks were observed as suspicious, because their geometry was modified during the study area extraction. Consequently, their counts were adjusted proportional to their original areas before calculation. In the final step, population density surfaces for 2001 and 2010 were subtracted for each variable to obtain their change.

1.3.3 Variable groups

With the processed censuses, we defined a set of variable groups related to demographic features and their change between 2001–2010. These were selected based on similar research [86,87] and included the categories: age composition (Age), literacy level (Education), gender distribution (Gender), household structure (Household), spoken languages (Language), and work sectors (Work). These constituted the Socioeconomic and Sociocultural macro levels (Table 3). Another set of variable groups was defined following similar research [6] to include biophysical (Biophysical) and land cover (Land Cover) features, constituting the Landscape

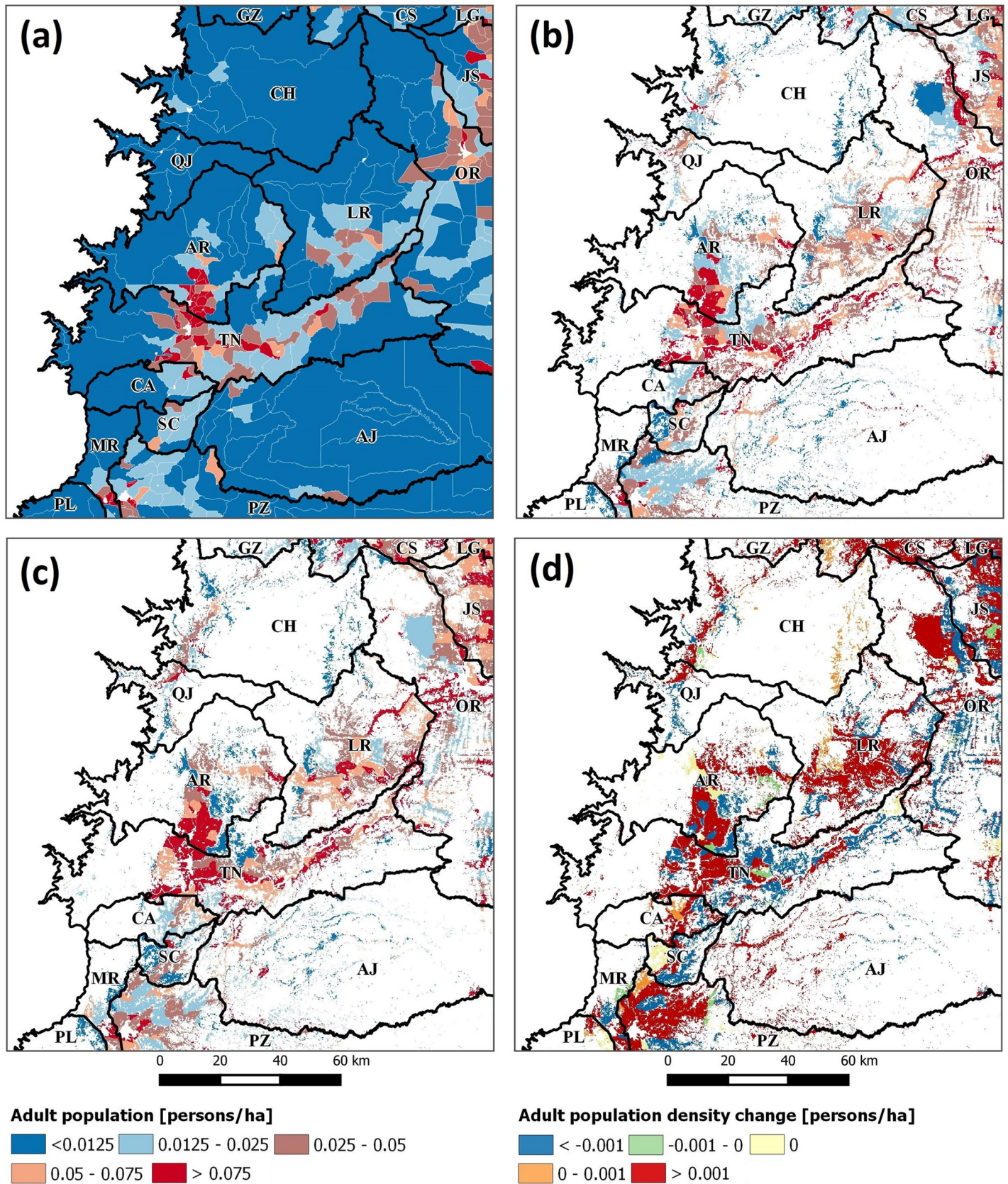


Fig 6. Adult population density surfaces using: (a) census 2001, (b) dasymetric mapping 2001, and (c) dasymetric mapping 2010. The subtraction of the last two to derive adult population density change 2001–2010 is shown in (d).

<https://doi.org/10.1371/journal.pone.0226224.g006>

macro-level (Table 4). Moreover, cost-distance models of agricultural collection facilities (palm oil, coffee, cacao, fruit, and milk products) were used as proxies of commercial agricultural activities (Agriculture). Another variable group (Infrastructures) considered the Euclidean distance to human-built infrastructures such as oil wells and mining blocks established between 2000 and 2010. Additionally, we include here the linear trend map from the night-lights 2001–2010 dataset. These variable groups formed the Commodities macro level (Table 4). A total of 34 variables were organized in 10 groups and 4 macro levels, with their values extracted into the analysis grid cells as averages.

1.3.4 Geographically weighted regression (GWR) and random forest (RF)

GWR is a statistical method to model spatial relationships under the assumption of spatial non-stationarity and location interdependency. It was conceived as an extension of linear regression analysis incorporating local estimates and surface representations of relationships among dependent and independent variables [37]. A GWR model can be specified as:

$$\gamma_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} \chi_{ki} + \varepsilon_i; \quad i = \{1, \dots, n\}$$

where γ_i is the dependent variable at i location, β_{i0} is the estimated intercept at i location, χ_{ki} is a vector of $k = \{1, \dots, m\}$ independent variables at i location and ε_i is the error term of the estimation at i location. Since i is considered as an $n \times n$ (with n number of observations) diagonal matrix in GWR, its formulation for local parameter estimates at i location is more conveniently expressed as:

$$\hat{\beta}(i) = [X^T W(i) X]^{-1} X^T W(i) \gamma$$

where $\hat{\beta}(i)$ is a vector of spatially weighted estimates for the k -th independent variables at i location, X is a $n \times k$ matrix of independent variables, $W(i)$ is the $n \times n$ diagonal weights matrix which ensure that observations near i have the largest weight values rather than those further away [90], and γ is a vector of k observations of the dependent variable. To define $W(i)$ a weighting function is declared considering: (1) the type of distance between i and its neighbors, (2) a kernel function specifying the weighting scheme, and (3) the bandwidth distance to control the number of observations within the kernel. Commonly, the Euclidean distance and the exponential kernel function are used as the weighting scheme [91]. The latter is defined by:

$$W_{ij} = \exp\left(-\frac{|d_{ij}|}{bw}\right)$$

where W_{ij} is the weight assigned to observation j for the estimation of i , d_{ij} is the distance between j and i , and bw is the bandwidth. The latter defines GWR mapping sensibility, as large values result in global regression estimates, while small ones introduce randomness [92]. Moreover, bw can be set as a fixed (constant distance) or an adaptive kernel (constant number of local observations). The latter is recommended, as it ensures a sufficient flow of information for each local calibration, while its size can be determined through cross-validation [93]. We implemented an adaptive exponential kernel function using Euclidean distance but tested different values for $bw = \{100, 200, \dots, 800\}$ to determine it. Then, for each i and its neighbors, we constructed an RF model instead of a linear regression.

Table 3. Mean population densities change (2001–2010) for variables groups in socioeconomic and sociocultural macro levels using deforestation (DEF) and reforestation (REF) datasets.

Macro level	Variable group	Prefix	Variables ¹	DEF	REF	Data source
Socioeconomic	Age	D_ygr	Young population (15–25 y)	0.006	0.005	[79,80]
		D_adt	Adult population (26–45 y)	0.008	0.008	
		D_old	Older adult population (45–72 y)	0.009	0.007	
	Education	E_ilt	Illiterate	-0.001	-0.002	
		E_pri	Primary education (1–6 y)	-0.0009	-0.004	
		E_sec	Secondary education (7–12 y)	0.018	0.020	
		E_hgr	Higher education (>13 y)	0.003	0.004	
	Work	W_agr	Agricultural workers	0.010	0.008	
		W_ind	Industrial workers	-0.002	-0.002	
W_ser		Service workers	0.0006	0.001		
Sociocultural	Gender	G_chm	Chief male household	0.008	0.007	
		G_pom	Male population	0.012	0.010	
		G_chf	Chief female household	0.002	0.002	
		G_pof	Female population	0.012	0.011	
	Household	H_sma	Small families (1–2 children)	0.002	0.002	
		H_med	Medium families (3–5 children)	0.004	0.004	
		H_lar	Large families (>5 children)	-0.014	-0.016	
	Language	L_spa	Speak Spanish ²	0.046	0.046	
		L_kcw	Speak <i>Kichwa</i> ³	0.029	0.032	
L_wao		Speak <i>Huaio Tededo</i> ⁴	0.0007	-0.0002		

¹ All variables are reported as population densities (persons/ha).

² Most commonly spoken language by *colonos* in the NEA.

³ Second most commonly spoken language and ethnicity in the NEA.

⁴ The language of *Huaorani* people.

<https://doi.org/10.1371/journal.pone.0226224.t003>

Table 4. Mean values for variables groups in landscape and commodities macro levels using deforestation (DEF) and reforestation (REF) datasets.

Macro level	Variable group	Prefix	Variables and units	DEF	REF	Data source
Landscape	Biophysical	B_alt	Altitude (m.a.s.l.)	656	716	[88]
		B_fer	Soil fertility (% organic matter)	1–2	1–2	
		B_rfl	Annual rainfall (mm)	3588	3658	
	Land cover	C_bsl	Bare soil (% frequency)	31	25	[70]
		C_pas	Pasture (% frequency)	73	53	
		C_sze	Mean patch size (ha)	6.5	4.9	
Commodities	Agriculture	A_plm	Accessibility to oil palm extraction facilities (h) ¹	1–3	1–3	[83]
		A_cao	Accessibility to coffee and cacao collection centers (h) ¹	0.5–1	0.5–1	
		A_fru	Accessibility to fruit collection centers (h) ¹	1–3	1–3	
		A_mlk	Accessibility to milk product collection centers (h) ¹	1–3	1–3	
	Infrastructure	I_oil	Distance to oil wells (perforated between 2000 and 2010) (m) ¹	250	299	[88,89]
		I_min	Distance to mining blocks (assigned between 2000 and 2010) (m) ¹	109	95	
		I_ngt	Stable nightlights trend 2000–2010 (slope)	-0.27	-0.27	

¹ These variables were normalized (0–1) and inverted (i.e., maximum distance and travel time were assigned values near zero, contrary values near one) during GWR modelling to facilitate interpretation but were transformed to their original units for reporting.

<https://doi.org/10.1371/journal.pone.0226224.t004>

RF is an ensemble learning method for classification and regression that produces multiple decision trees using bagging to select subsets of training samples and random feature selection to split them [41]. It is easy to compute and is tolerant to missing and multicollinear data [94], moreover it provides error estimates without requiring a validation dataset. During the training phase, it randomly sample with replacement, about two-third of the training samples (referred to as *in-bag* samples) for a given training set $T = \{1, \dots, t\}$ to grow a specified number of trees to the largest extent possible, selecting randomly a number of variables $V = \{1, \dots, v\}$ at each node to determine their split. This gives an ensemble of classification or regression trees, if T and V are bagged repeatedly B times to grow trees with these samples. After training, it averages predictions from all individual regression trees or by taking the majority vote in classification. This can be summarized as:

For $b = \{1, \dots, B\}$:

1. Sample randomly, with replacement, n training samples and variables from T, V . Set them as T_b, V_b .
2. Train a classification or regression tree on T_b, V_b . Set it as RF_b .

End for

3. Average individual RF_b results in regression or by taking the majority vote in classification and calculate model performance

To monitor error, the remaining one-third samples (referred to as *out-of-bag* samples or OBB) are used in an internal cross-validation technique [95], which computes the number of correct predictions. Other accuracy metrics are also possible to derive from OBB (e.g. Kappa, R-square, etc.). In addition, variables predictive power (or importance) can be calculated through different approaches (e.g. Gini index, accuracy decrease, permutation) but permutation is mostly recommended [96].

While the vast majority of RF problems can be solved with a unique (or global) model; here, we followed the approach of Georganos et al. [43] to combine GWR and RF to derive multiple spatially weighted (or local) RF models. This is possible if during the bagging step of RF, we assign to the neighbor observations of i a sampling probability based on the distance weights or $W(i)$. For this, we can reshape step one from the previous RF workflow as follows:

For $b = \{1, \dots, B\}$:

1. Apply $W(i)$ probabilities in sampling, with replacement, for n training samples from T . Sample n random variables from V . Set them as T_b, V_b .

⋮

Contrary to a global model, in this approach, an ensemble of spatially weighted (or local) RF models are obtained. Their features can be mapped and among them we can mention: 1) local variables importance (*LVI*), which shows variables predictive power locally (or spatially), for each variable in V ; 2) prediction results, which can be also reported as probabilities; and 3) model performance. The latter includes specific metrics for classification (e.g. kappa index, confusion matrix) or regression (e.g. r-squared, mean absolute deviation). To compute these features, in the next section we explain how we implemented GWR and RF.

1.3.5 Implementation of GWR and RF

To operate this GWR and RF as an algorithm, we used the ranger [42] and GWmodel [97] packages. The first case, is a fast C++ and R implementation of RF that allows weights for

sampling training observations. This parameter, called case weights in the ranger function, is used to define the spatial weights $W(i)$ for observations near i . The second case, is a complete toolbox for the geographically weighted approach, including functions for: regression analysis, spatial metrics, weight-decay functions, among others. Since the proposed algorithm (named from now as GWRF) involves multiple steps, we summarized them as a pseudocode:

```

Algorithm 1: Geographically weighted random forest (GWRF)
INPUTS
Sp: spatial dataset (with dependent and independent variables); Dep:
dependent variable name; Kfun: kernel function; Ktyp: kernel type; Kbw:
kernel bandwidth;
OUTPUTS
LVI: variables importance; YHAT: prediction probabilities; ACC: accu-
racy metrics;
PROCEDURE
1: READ Sp; SET Dep as dependent variable; SET Outputs as an empty
list
FOR each i element IN Sp DO
    2: CALCULATE distances from all elements in Sp to i; SET them as Di
    3: SORT Di AND select those within Kbw; SET selected observations as
iobs
    4: REMOVE variables with zero variance in iobs
    IF Dep is categorical (classification problem)
        5: UPSAMPLE unbalanced classes in iobs
    END IF
    6: CALCULATE spatial weights for iobs applying the Kfun; SET it as W
(i)
    7: TRAIN Random Forest with iobs applying W(i) as sampling probabili-
ties; SET it as RFi
    8: EXTRACT LVIi from RFi; REMOVE variables with negative scores in
LVIi
    IF number of variables in LVIi are not equal to input Sp variables
number DO
        9: TRAIN Random Forest with iobs applying W(i) as sampling prob-
abilities; UPDATE object RFi
        10: EXTRACT LVIi from RFi; REMOVE variables with negative scores
in LVIi
    END IF
    11: SET removed variables in LVIi as zero
    12: EXTRACT predictions (probabilities, predicted value); SET it
as YHATi
    13: CALCULATE accuracy metrics (Kappa, R-squared, prediction fail-
ure, residual standard error); SET it as ACCi
    14: SAVE LVIi, YHATi, ACCi into Outputsi
    END FOR
15: MERGE Outputs{i1,...,in}; SAVE outputs as LVI, YHAT, ACC spatial datasets
END PROCEDURE

```

Note that the algorithm inputs require to define the dependent variable and we set it as *DEF* and *REF* to refer to the annual deforestation and reforestation rates (See Table 1). As the algorithm assume that the rest of variables are independent, we can express them according to their variables groups names:

$$Vars = \{Landscape, Commodities, Socioeconomic, Sociocultural\}$$

Where *Vars* refers to all variables described in Table 3 and Table 4. Now, we can represent the calibration of GWR for *DEF* as:

$$\begin{aligned} GWRF(Sp = DEF | Vars, Dep = DEF, K_{fun} = exponential, K_{typ} = adaptative, K_{bw} \\ = \{100 \dots, 800\}) \end{aligned}$$

And similarly, for *REF* as:

$$\begin{aligned} GWRF(Sp = REF | Vars, Dep = REF, K_{fun} = exponential, K_{typ} = adaptative, K_{bw} \\ = \{100 \dots, 800\}) \end{aligned}$$

After reading and preparing models (step 1), a loop is defined for process each element in the spatial dataset *Sp*. This processing included operations such as:

- Data cleaning (steps 4, 5, 8, 10 and 11), following Genuer et al. [98] for recommended practices in RF variables selection analysis;
- GWR calculations (steps 2, 3 and 6), following Gao et al. [91], Farber et al. [93] and Gollini et al. [90] to define kernel type K_{typ} as adaptive, and its function K_{fun} as exponential;
- RF training (steps 7 and 9), following Breiman [41] and Wright et al. [42] for decide default RF calibration (i.e. 500 decision trees, square root number of variables to split at in each node, and permutation method for *LVI* calculation);
- Accuracy assessment (step 12 and 13), calculating Kappa and prediction failure in classification; and R-squared and residual standard error in regression. This assessment is conducted with the OBB samples;
- and Storing outputs (step 14 and 15). These included: *LVI* score for each variable in *Vars*, predictions and probabilities (*YHAT*), and models accuracies (*ACC*).

Since all above calculations were computing demanding, we implemented GWR for parallel computing but processing time depended of kernel bandwidth K_{bw} (See S1 Appendix). Furthermore, we tested GWR for classification, reclassifying *DEF* and *REF* rates into 5 classes (see Fig 4), while for regression we maintained them as continuous values. To decide the best approach, we compared results of *ACC* for different values of K_{bw} (see section 1.3.4).

1.3.6 DEF and REF linkages assesment

Since *LVI* results were extensive, we first plotted $LVI_{DEF|VARS}$ and $LVI_{REF|VARS}$ in a radar plot [99] to observe how forest change rates were influenced by *Vars*. This facilitated identification of variables with opposite predictive power in *DEF* and *REF*, which were selected to map and observe with more detail. Moreover, we calculated variables correlation with forest change rates to further explore their similitude with *LVI*. For the next step, we followed Freitas et al. [39] and clustered $LVI_{DEF|VARS}$ and $LVI_{REF|VARS}$. For this, we used the expectation-maximization algorithm [100], as it allows continuous and categorical data. We determined 2 clusters based on the gap statistics [101] to later select the one with the highest rate. We assume that these areas represent active forest change fronts with similar variables importance, which distil their driving forces. We named them as $CLUS_{DEF}$ and $CLIS_{REF}$ groups and extracted *Vars* to

test the next hypothesis:

$$H0 : CLUS_{DEF|Vars} = CLUS_{REF|Vars}$$

$$H1 : CLUS_{DEF|Vars} \neq CLUS_{REF|Vars}$$

We applied the Wilcoxon rank sum test [102], which computes P-values that test the $H0$ hypothesis that the two groups have the same distribution. If $H0$ was rejected ($P\text{-value} > 0.05$), we assumed a difference and subtracted the median values to observe if $CLUS_{DEF|Vars}$ was higher or lower than $CLUS_{REF|Vars}$. This operation allows us to classify results for each variable in $Vars$ according three categories:

- DEF and REF were equal ($H0$ is accepted; similar DEF and REF medians);
- DEF was greater ($H0$ is rejected; DEF is higher than REF median);
- DEF was lower ($H0$ is rejected; DEF is lower than REF median).

Additionally, we calculated the Cliff's Delta [103] to observe the effect size of $Vars$ in DEF and REF . We applied the thresholds provided in Romano [104] to classify this metric into 4 classes (i.e. negligible, small, medium and large) and complete our analysis.

1.4 Results

1.4.1 Comparison between dasymetric mapping and censuses

We processed 21 census variables from 2001 to 2010 with dasymetric mapping (DAS) to compare their population densities with those derived from unprocessed censuses (CEN). For this, we calculated the population density for each variable and census year, using DAS and CEN data sources to plot them and highlight their differences (Fig 7A). DAS exceeded CEN for variables with larger values (e.g., L_spa, G_pom, G_pof) but fell behind for those with smaller values (e.g., L_wao, H_sma, E_lit). On average, DAS obtained 0.06 ± 0.04 and 0.15 ± 0.12 persons/ha for the population density in 2001 and 2010, while CEN obtained 0.02 ± 0.01 and 0.09 ± 0.07 persons/ha, respectively. This means that DAS exceeded CEN by 145% in 2001 and 58% in 2010. Furthermore, we subtracted population densities in 2001 and 2010 in both sources to obtain their change (Fig 7B). Similarly, it was observed that DAS exceeded CEN for variables with large values (e.g., L_spa, G_pom, E_sec) but was inferior for those with small values (e.g., H_lar, L_wao, E_lit). Averaging all variables, DAS obtained 0.08 ± 0.08 persons/ha for the population density change between 2001 and 2010, while it was 0.07 ± 0.08 persons/ha for CEN; i.e., an increase of 27% by DAS. Interestingly, by DAS, variables H_lar and L_wao showed decreases of -23.2% and -31.2%, respectively, for the population density change between 2001 and 2010.

1.4.2 Accuracy assessment of GWR

We analyzed GWR according to eight bw values and two approaches (classification and regression) to decide which one achieved the highest accuracy. It can be seen that the classification improved the Kappa as the size of bw increased, around 400 observations stabilized and variability was reduced. Moreover, results for REF were better than for DEF , showing in both cases a logarithmic curve with increasing bw (Fig 8A). This was shown by different regression results, as the R-squared diminished with $bw > 100$ for DEF but was not seen as relevant for REF (Fig 8B). This was interpreted as unexpected, as it is known that accuracy increases with larger values of bw [93] until its value is large enough to cover all the study area and become a

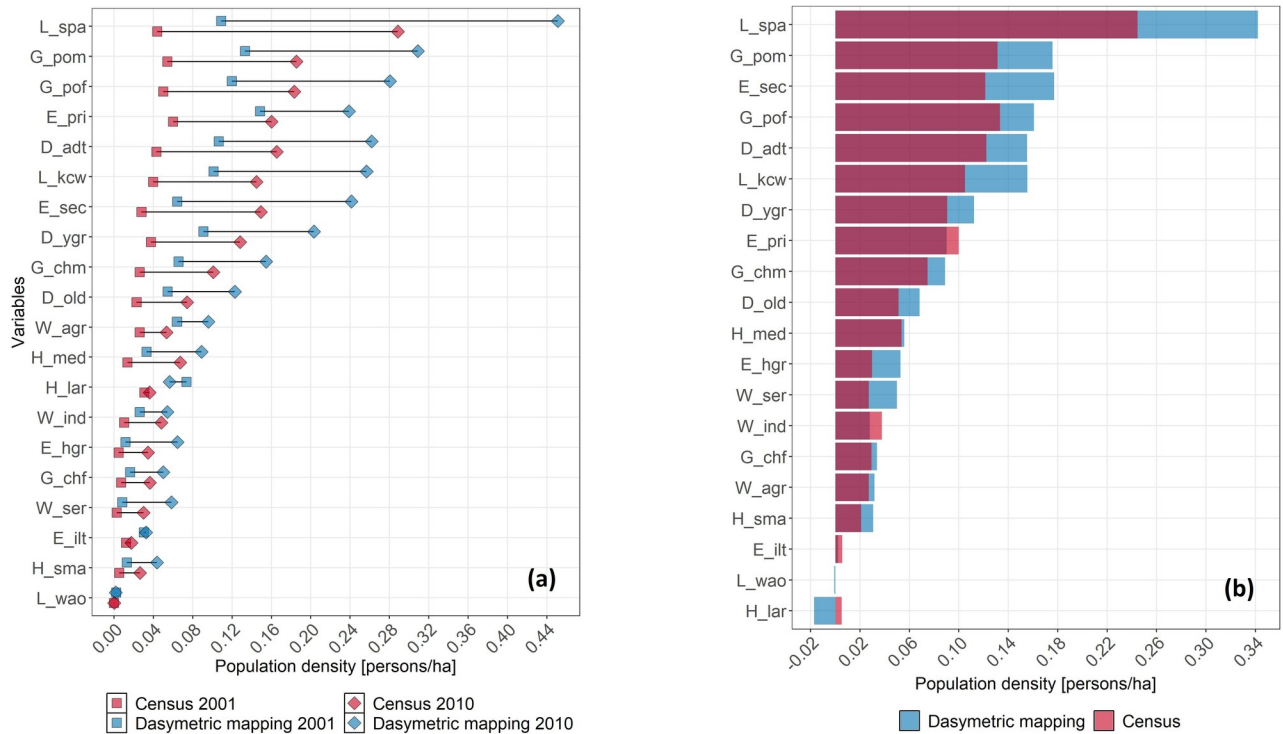


Fig 7. (a) Population density differences between dasymetric mapping and census data for 2001 and 2010; and (b) subtraction to derive population density change between 2001 and 2010. In both cases, variables were ordered according to magnitude and transparent color was used to see overlapping areas (red is over blue).

<https://doi.org/10.1371/journal.pone.0226224.g007>

global average. For these reasons, we decided to use a classification approach for $bw = 400$, as larger values did not significantly improve results, achieving a Kappa of $96 \pm 2\%$ for *DEF* and $97 \pm 1\%$ for *REF*. This indicated that the classification approach resulted in adequate predictions for all classes considered but not in regression. This was probably due to the imbalanced sampling introduced by the kernel during calculations. Finally, we mapped the Kappa for GWRf with $bw = 400$ to observe its spatial distribution (Fig 9). Here we could see that the lowest relative Kappa values (74–95%) covered areas with the largest rates (>2.5%) in *DEF* and *REF*. This implies that GWRf resulted in poor predictions in areas where rates varied (e.g., JS or La Joya de los Sachas in *DEF*, and TN or Tena in *REF*) than in areas with homogeneous rate intervals or where few high rate peaks were observed.

1.4.3 LVI comparison and visualization

After the RF classification achieved acceptable results, we created radar plots using the *LVI* results for the four macro levels considered. In the case of Landscape (Fig 10A), it can be seen that variables related to land cover were important for both *DEF* and *REF* but those related to Biophysical seem to be more in *DEF*. Among them, *C_pas*, *C_bls*, and *B_alt* were more important in *REF*, while *B_fer* was only in *DEF*. Commodities (Fig 10B) indicates that variables related to Infrastructures were more important in *REF* but those related to Agriculture were more important in *DEF*. Among them, variables *I_min* and *I_ngt* were more important in *REF*, while *A_cao*, *A_fru* and *A_plm* were more important in *DEF*. The Socioeconomic (Fig 10C) indicate that Work was more important in *REF* but Education was more important in *DEF* with the exception of *E_ilt*. Furthermore, the variable group Age showed a similar result

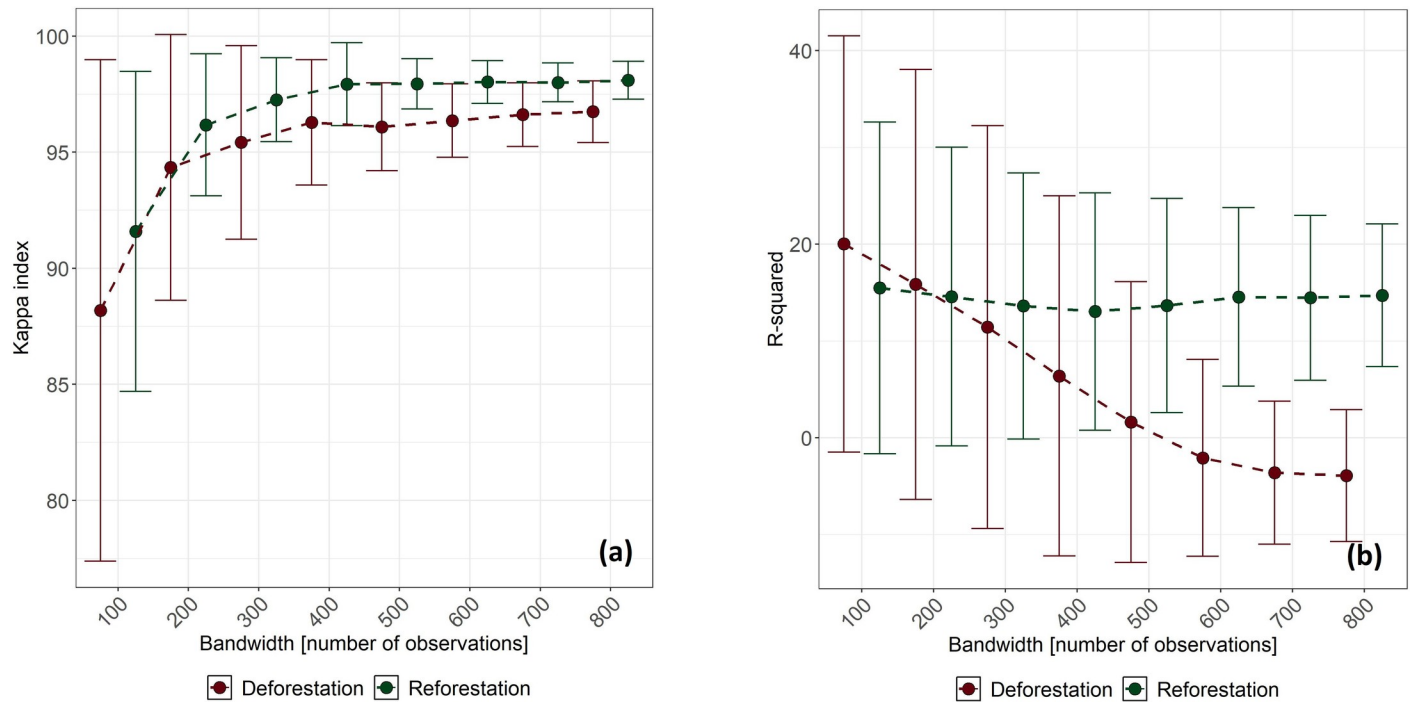


Fig 8. Accuracy metrics error bars for different bw sizes in GWRF: (a) classification (kappa) and (b) regression (R-squared). Points connect mean values in Kappa and R-squared, while bars indicate standard deviations.

<https://doi.org/10.1371/journal.pone.0226224.g008>

between *DEF* and *REF*, but variable *D_old* appears to be more important in *REF*. Finally, the Sociocultural (Fig 10D) indicate that all of them (Gender, Household, and Language) were more important in *DEF*, with a few exceptions. It was seen that *G_chf* and *L_wao* variables were only important in *REF*, while the rest of the variables were important in *DEF*.

Following, we mapped variables: *B_fer*, *I_ngt*, *E_sec* and *H_med*; as they showed opposite predictive power. Fig 11 shows results for *DEF* and here it can be seen that high *LVI* values are spatially related to high *DEF* rates as well. In *REF*, this was different as its observed low to medium *LVI* values (Fig 12) except for variables *E_sec* and *I_ngt* whose values are higher in areas with also high *REF* rates. This is particularly interesting, as these variables seem to be good predictors in both *DEF* and *REF* when correlation exists. Following, results from correlations between *LVI* and forest change rates (see S2 Appendix) indicated that *E_pri* and *W_agr* were also good predictors (correlation > 0.131), while worse predictors were *A_fru* and *L_wao* (correlation < 0.061). It was observed, that the latter were associated to zero *LVI* values (see Fig 10) as this was the result of the cleaning routine of GWRF (see section 1.3.5). Furthermore, it was observed that variables such as *G_chm*, *E_hgr*, *H_med*, *L_kcw*, *B_alt* and *G_chf* generated the opposed effect in *DEF* and *REF*. These variables have inverse relationships and highlights the complex structure of these land cover change dynamics. To facilitate the analysis of these dynamics, in the next section, we report results from the clustering and the hypothesis testing.

1.4.4 Clustering and hypothesis testing

After clustering *LVI* into two groups, we tested our hypothesis. In the case of $CLUS_{DEF|Vars}$ we observed that its rate q achieved $2.0 \pm 7.5\%$ and included 1160 grid cells (464,000 ha), while in $CLUS_{REF|Vars}$ the rate achieved $4.1 \pm 16.4\%$ and included 722 grid cells (288,800 ha). The

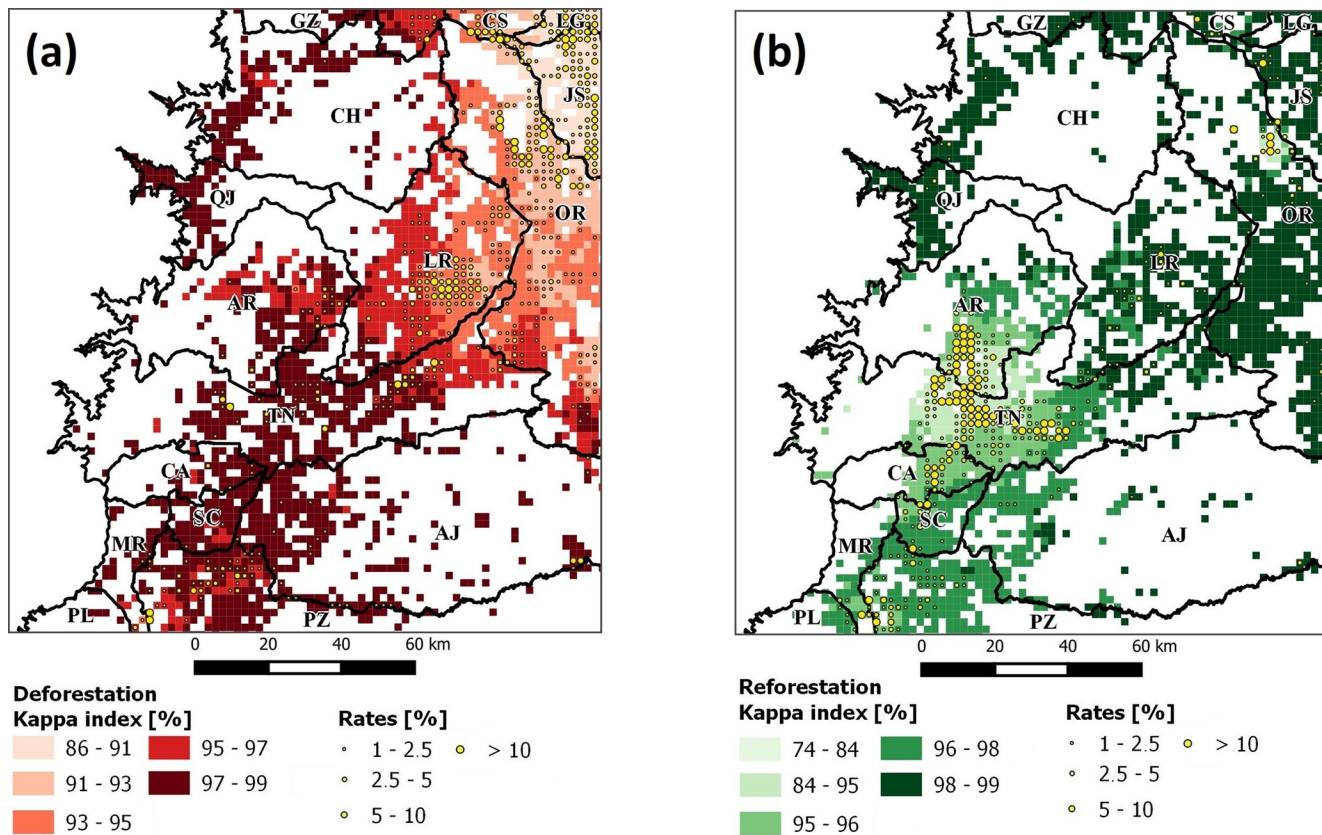


Fig 9. Kappa for GWRf classification with $b = 400$ in (a) DEF and (b) REF datasets. Rates were filtered to values greater than 1% to enhance visualization.

<https://doi.org/10.1371/journal.pone.0226224.g009>

$CLUS_{DEF|Vars}$ was larger than $CLUS_{REF|Vars}$ with 438 grid cells (175200 ha) but also its rate was lesser by 2.1%. Regarding their locations, both groups matched cantons where higher rates were observed (Fig 13). Interestingly, a boundary between $CLUS_{DEF|Vars}$ and $CLUS_{REF|Vars}$ appears at ~482 m.a.s.l. at cantons CH, LR, TN, and AJ (El Chaco, Loreto, Tena and Arajuno), indicating the limit between these regions and their governing forest change phenomena. Moreover, an overlap was seen between $CLUS_{DEF|Vars}$ and $CLUS_{REF|Vars}$ but it was marginal as it included only three grid cells (or 1200 ha).

Following, we describe the results of the hypothesis. We first show results for variables where $CLUS_{DEF|Vars}$ and $CLUS_{REF|Vars}$ medians were equal. This is summarized in the Table 5 and it can be noted that only variables from the Socioeconomic and Sociocultural macro levels were present. These similitudes were also evidenced by the Cliff's delta, as resulting effect sizes were negligible. We identified the next variable groups: Education (E_hgr), Gender (G_chm), Household (H_med, H_sma), Language (L_kcw, L_spa and L_wao) and Work (W_agr) and could observe that variables H_sma (small families) and W_agr (agriculture workers) were close to be significant but their Cliff's delta indicates that their magnitude effects were still negligible.

The next section belongs to variables whose H_0 was rejected. We first report variables whose median value was lower in $CLUS_{DEF|Vars}$. This is shown in Table 6 and here its seen variables from Commodities, Landscape and Socioeconomic macro levels exclusively. They were represented by variables groups: Work (W_ind, W_ser), Agriculture (A_mlk, A_cao, A_fru), Infrastructure (I_oil) and Biophysical (B_alt, B_rfl). From them, variables with the largest

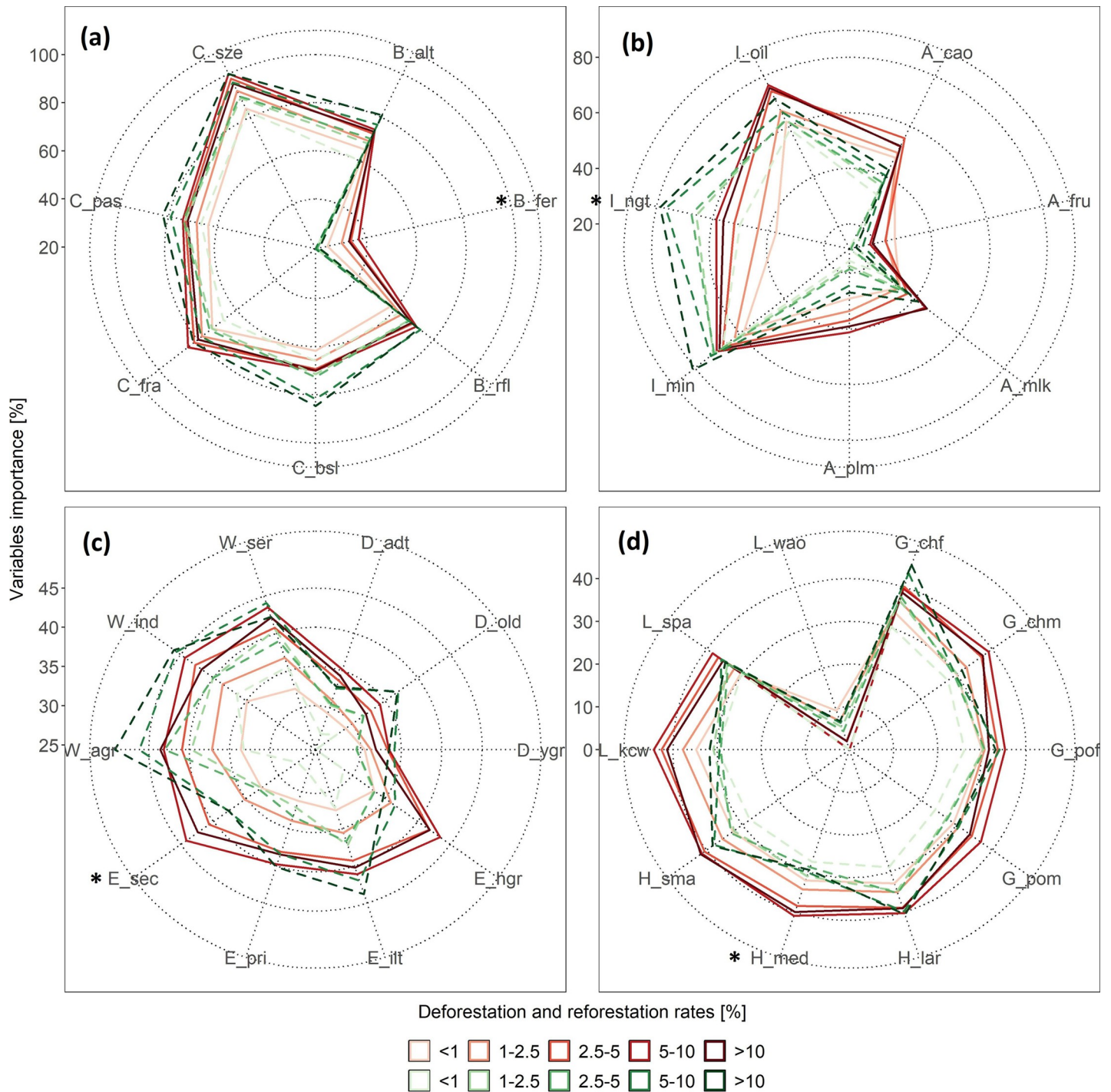


Fig 10. LVI radar plot for: (a) landscape, (b) commodities, (c) socioeconomic, and (d) sociocultural macro levels. The asterisk (*) highlight mapped variables.

<https://doi.org/10.1371/journal.pone.0226224.g010>

difference meant that $CLUS_{DEF|Vars}$ was characterized by: higher accessibility to oil palm extraction facilities (A_plm), closer distance to oil wells (I_oil), lower altitudes (B_alt) and lower annual rainfall (B_rfl). Following, higher accessibility to fruit, coffee and cacao collection centers (A_fru and A_cao) was also seen in $CLUS_{DEF|Vars}$ but their differences with $CLUS_{REF}$

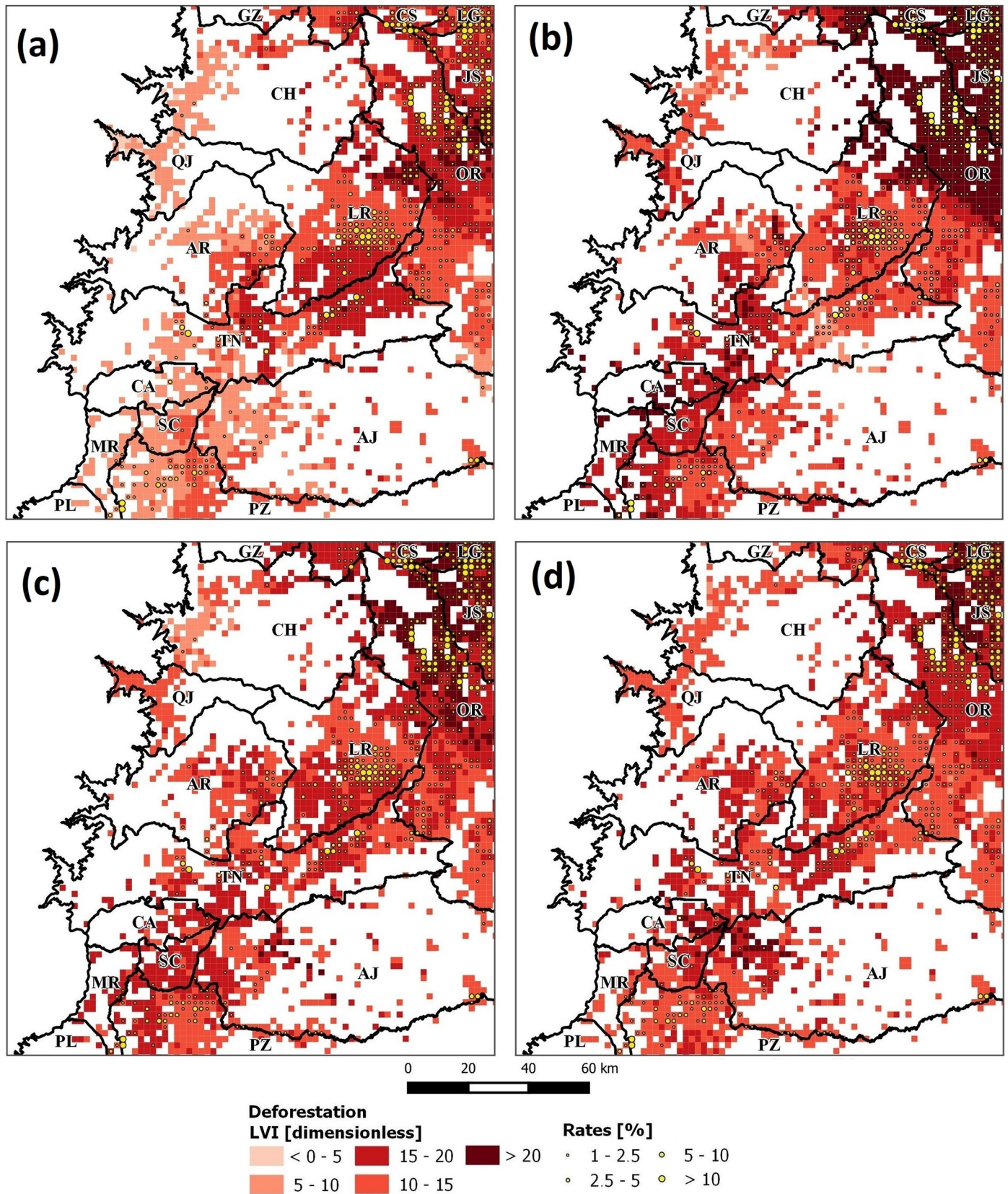


Fig 11. $LVI_{DEF|Vars}$ maps for selected variables: a) B_fer (soil fertility), b) I_ngt (stable nightlights), c) E_sec (secondary education); and d) H_med (medium families). High values refer to increased predictive power. Rates were filtered to values greater than 1% to enhance visualization.

<https://doi.org/10.1371/journal.pone.0226224.g011>

$Vars$ were smaller. The opposite picture can be inferred from the abovementioned to describe $CLUS_{REF|Vars}$ characteristics. On the other hand, the zero value of L_wao variable can be attributed to its absence in these regions. Finally, variables A_mlk, W_ind and W_ser seem to achieved negligible differences; therefore, we qualified them as not relevant for our analysis. Contrasting these results, we now report variables whose median value was higher in $CLUS_{DEF|Vars}$. This is shown in Table 7 and in this case, all macro levels were observed. Among identified variable groups, we can mention: Infrastructure (I_ngt, I_min), Biophysical (B_fer), Land cover (C_frac, C_pas, C_size, C_bsl), Age (D_adt, D_ygr, D_old), Education (E_pri, E_ilt, E_sec), Gender (G_pof, G_pom, G_chf) and Household (H_lar). From them, the only variable that achieved a large difference was bare soil frequency (C_bsl). This indicated that $CLUS_{DEF|Vars}$ was more prone to experience land clearing after tree removal. This is reasonable if we consider that forest succession implies vegetation regrowth after land clearing. Following, variables with medium effect size meant that $CLUS_{DEF|Vars}$ was characterized by: larger patches sizes (C_size) and higher pasture frequency (C_pas). In addition, other variables with small effect size indicated higher population density for: secondary education (E_sec), older adults (D_old), Illiterate (I_ilt), and female chief household (G_chf). Remaining also in small magnitude, larger distance to mining infrastructures (I_min) seems to also characterize $CLUS_{DEF|Vars}$. For the remaining variables, i.e. I_ngt, B_fer, C_fra, D_adt, D_ygr, E_pri, G_pof, G_pom and H_lar, a negligible effect size is observed and were less informative to identify differences between $CLUS_{DEF|Vars}$ and $CLUS_{REF|Vars}$.

1.5 Discussion

1.5.1 Utility of remote sensing time series–related products for FDD analysis

Some studies have successfully identified proximate causes in FDD analysis by using remote sensing time series–related products [105,106]. In this research, we extended these applications through the use of (i) grid-based rate calculations, (ii) derivation of land-cover metrics, and (iii) trend analysis of Nighttime Lights Time Series to explore correlations with DEF and REF rates. Except for the (i), due its simplicity, the other two deserve further discussion as they represent innovative approaches which are not well documented to our current knowledge. Derivation of land-cover metrics indicates that it is possible to extract additional information from remote sensing time series that can be useful for determining the degree of land-use intensity from previous or posterior land-cover change events. This has been done using spectral trajectories [107,108], but here we show how they can be derived from land-cover maps, with a less sophisticated approach and comparatively limited results. With more dense optical and radar time series availability, it might be possible to more precisely detect land-cover classes that are usually not identifiable by their spectral features (e.g., coffee, cacao, forest plantations) but rather by their spectrotemporal signatures, as other studies have demonstrated [109–111]. This could help to improve forest monitoring, but also improve agriculture-related accessibility models, which are more difficult to derive and validate. Furthermore, trend analysis of Nighttime lights Time Series has shown that despite the low spatial resolution, they are still useful for investigating unknown patterns that strengthen model predictions (see Section 1.4.3). This was made possible thanks to free cloud-based platforms that allow processing of vast amounts of data from remote sensing time series and derivation of unprecedented

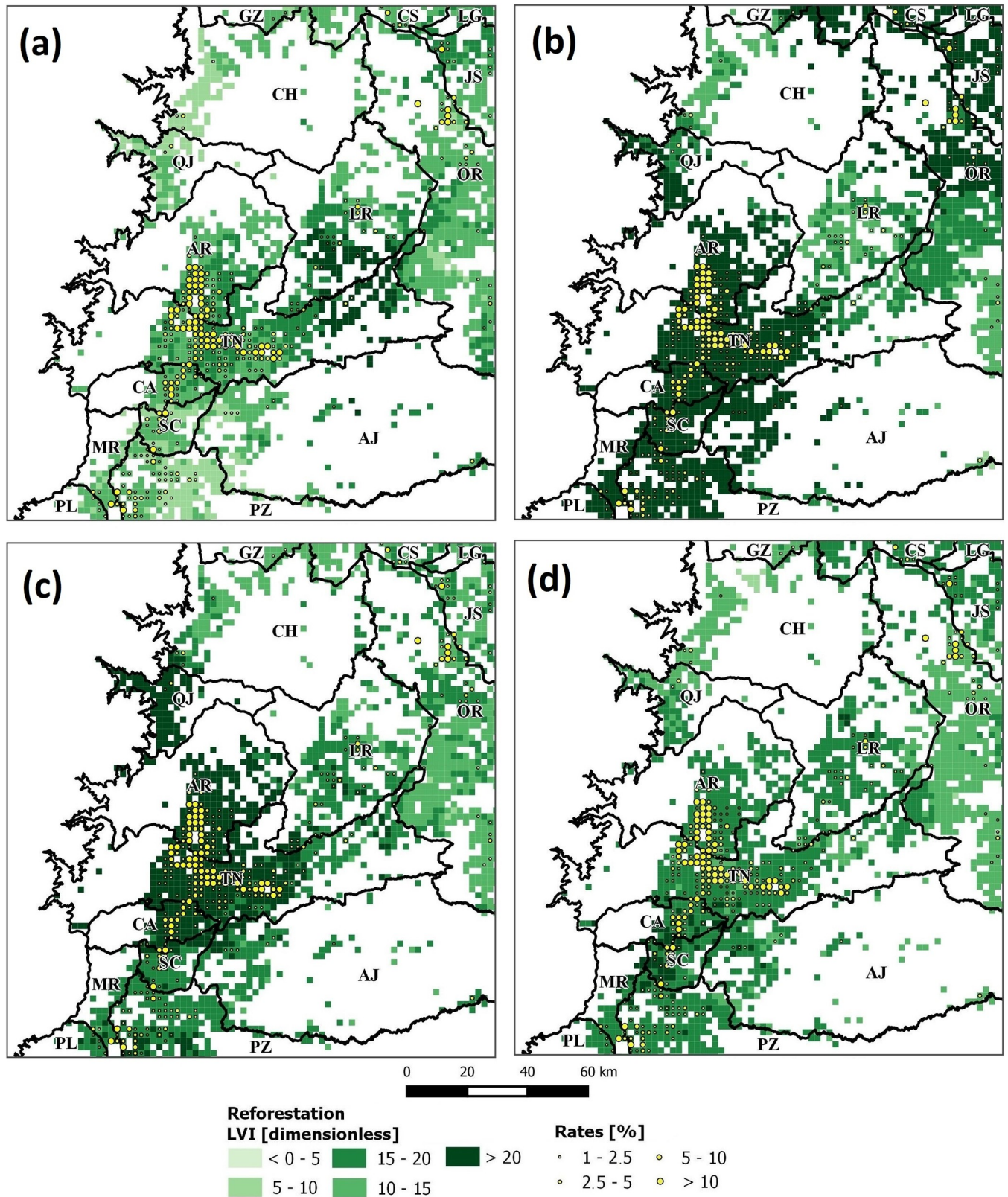


Fig 12. $LVI_{REFVars}$ maps for selected variables: a) B_{fer} (soil fertility), b) I_{ngt} (stable nightlights), c) E_{sec} (secondary education); and d) H_{med} (medium families). High values refer to increased predictive power. Rates were filtered to values greater than 1% to enhance visualization.

<https://doi.org/10.1371/journal.pone.0226224.g012>

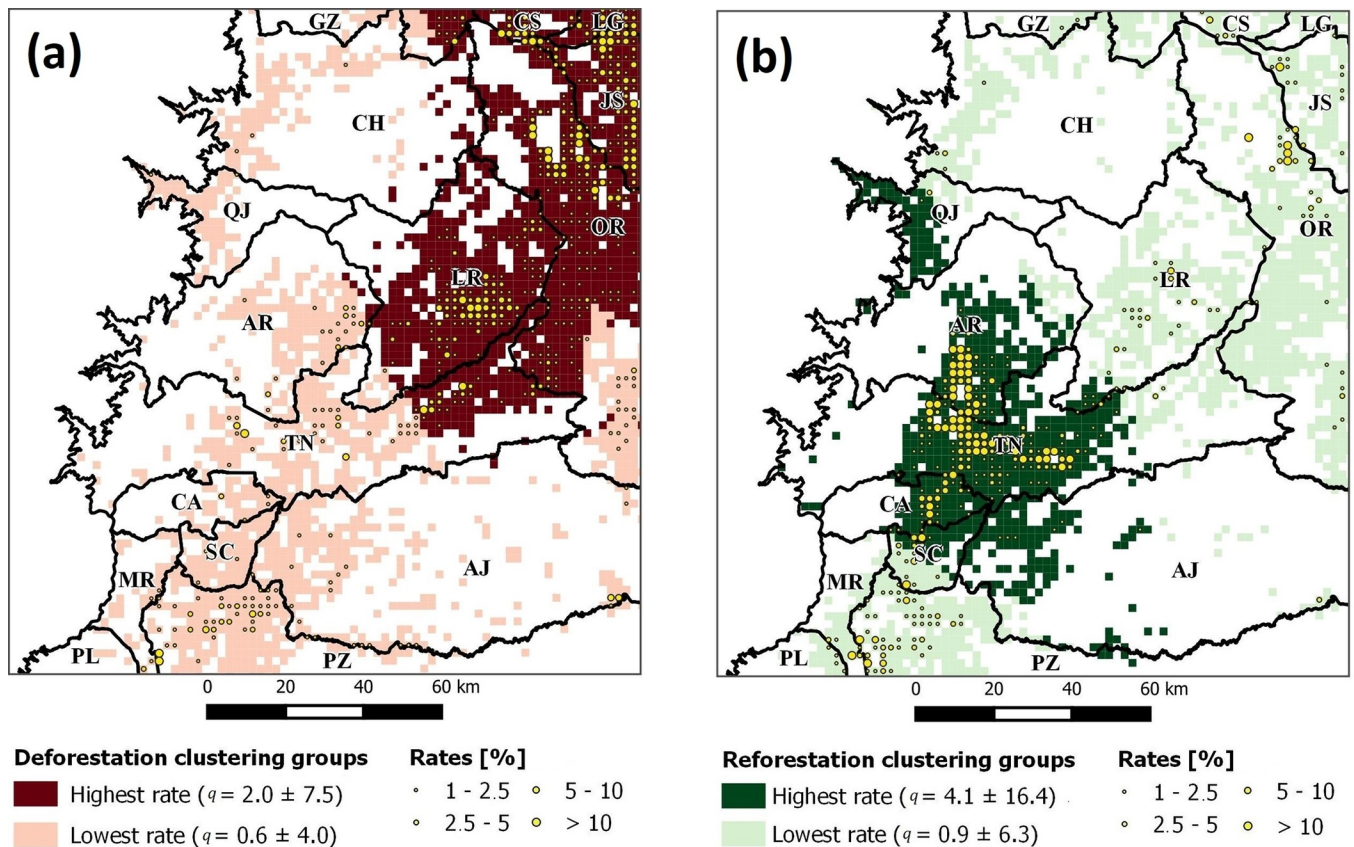


Fig 13. Local variable importance (LVI) clustering groups for (a) deforestation (DEF) and (b) reforestation (REF). Rates were filtered to values greater than 1% to enhance visualization.

<https://doi.org/10.1371/journal.pone.0226224.g013>

products [25,112]. This opens new possibilities for future research and reformulating known limitations due to processing capabilities and data availability. This does not mean that field data, novel algorithms, and local knowledge can be shared to such platforms naively, as sensitive information could be exposed and distributed without any ethical concern [113].

Table 5. Wilcox test and Cliff’s delta magnitude calculation for variables where DEF and REF were equal.

Variable	Median values			Wilcox test		Cliff’s Delta	
	DEF	REF	DEF-REF	P-value (p)	Significance ¹	Estimate (d)	Magnitude ²
E_hgr	0.0012	0.0011	0.0001	0.2753	*	0.0299	negligible
G_chm	0.0072	0.0029	0.0042	0.3264	*	0.0269	negligible
H_med	0.0021	0.0018	0.0004	0.8309	*	0.0058	negligible
H_sma	0.0018	0.0007	0.0011	0.0507	**	0.0535	negligible
L_kcw	0.0099	0.008	0.002	0.408	*	0.0227	negligible
L_spa	0.0294	0.0264	0.003	0.5073	*	-0.0182	negligible
L_wao	0	0	0	0.1773	*	-0.0218	negligible
W_agr	0.0071	0.0007	0.0064	0.0528	**	0.053	negligible

¹ Significance thresholds: * ($p > 0.1$), ** ($p < 0.1$), and *** ($p < 0.05$).

² Based in Romano 2006: negligible ($d < 0.33$), small ($d < 0.474$), medium ($d < 0.474$), and large ($d > 0.474$).

<https://doi.org/10.1371/journal.pone.0226224.t005>

Table 6. Wilcox test and Cliff’s delta magnitude calculation for variables where DEF is lower.

Variable	Median values			Wilcox test		Cliff’s Delta	
	DEF	REF	DEF-REF	P-value	Significance ¹	Estimate	Magnitude ²
B_rfl	3460,5	3927,8	-467,3	6.93e-96	***	-0,5688	large
B_alt	332,9	639,5	-306,6	1,89e-210	***	-0,8475	large
I_oil	89,1	391,6	-302,5	2,61e-195	***	-0,8162	large
A_plm	1-3	>3	-	1,13e-147	***	-0,6697	large
A_fru	1-3	1-3	-	4,96e-37	***	-0,2846	small
A_cao	0,5-1	1-3	-	1,86e-33	***	-0,3214	small
A_mlk	1-3	1-3	-	0,002	***	-0,0771	negligible
W_ind	0,0001	-0,0005	-0,0004	3,89e-7	***	0,1389	negligible
W_ser	0,0006	0,0004	0,0002	0,0346	***	0,0578	negligible

¹ Significance thresholds: * ($p > 0.1$), ** ($p < 0.1$), and *** ($p < 0.05$).

² Based in Romano 2006: negligible ($d < 0.33$), small ($d < 0.474$), medium ($d < 0.474$), and large ($d > 0.474$).

<https://doi.org/10.1371/journal.pone.0226224.t006>

1.5.2 DAS achievements and failures in intercensal analysis

Previous studies applying DAS have shown its effectiveness and improved performance for census processing in urban areas [114,115], specifically its capability to harmonize data and allow intercensal analysis. However, few studies have explored DAS in rural areas, as it was done in this study to enhance our comprehension of underlying causes in FDDs. This is because DAS reduces uncertainty in rural population mapping, as census blocks in rural areas are generally large in their extension, depending mostly on larger administrative units and have small population counts. Therefore, their population density calculations result in low

Table 7. Wilcox test and Cliff’s delta magnitude calculation for variables where DEF is greater.

Variable	Median values			Wilcox test		Cliff’s Delta	
	DEF	REF	DEF-REF	P-value	Significance ¹	Estimate	Magnitude ²
C_bsl	30,9	13,5	17,3823	2,97e-108	***	0,6035	large
C_sze	4,7	2,4	2,3	2,096e-43	***	0,3782	medium
C_pas	81,8	71	10,8	3,13e-56	***	0,4325	medium
G_chf	0,0017	0,0002	0,0014	1,49e-19	***	0,2476	small
E_sec	0,0137	0,0072	0,0065	5,25e-9	***	0,1598	small
E_ilt	0,0002	-0,0016	0,0018	2,78e-14	***	0,2083	small
D_old	0,0068	0,0031	0,0037	2,01e-9	***	0,1642	small
I_min	88,6	60,2	28,4	2,49e-16	***	0,2244	small
I_ngt	-0,2907	-0,2906	-0,0001	0,0023	***	0,0709	negligible
B_fer	1-2	1-2	-	1,04e-7	***	0,1254	negligible
C_fra	1,0754	1,0707	0,0047	1,911e-6	***	0,1304	negligible
D_adt	0,0077	0,0024	0,0053	0,0002	***	0,1011	negligible
D_ygr	0,006	-0,0008	0,0068	7,97e-8	***	0,1469	negligible
E_pri	-0,0008	-0,003	0,0022	0,0035	***	0,08	negligible
G_pof	0,0077	0,0016	0,0061	1,46e-5	***	0,1187	negligible
G_pom	0,0112	0,003	0,0082	3,73e-7	***	0,1391	negligible
H_lar	-0,0116	-0,0125	0,0009	1,63e-5	***	0,118	negligible

¹ Significance thresholds: * ($p > 0.1$), ** ($p < 0.1$), and *** ($p < 0.05$).

² Based in Romano 2006: negligible ($d < 0.33$), small ($d < 0.474$), medium ($d < 0.474$), and large ($d > 0.474$).

<https://doi.org/10.1371/journal.pone.0226224.t007>

figures that tend to obscure negative trends. Assuming that the data were collected properly, we saw this effect with CEN results, as it hid negative trends for the variables *L_wao* and *H_lar* (i.e., variables with small counts), contradicting DAS as well as other research observations in this region [116]. This is particularly important, as future research may consider more precise mapping approaches than choropleths to perform more reliable population density calculations.

Furthermore, as we used a road accessibility model to enhance its location, some observations are worth mentioning. First, this input data incorporated restrictions on non-forest masks with regard to areas less likely to be inhabited. Therefore, their use is valid under the assumption that road accessibility and rural populations are related. However, other transportation sources (e.g., rivers, airfields) may attract rural populations, especially among indigenous groups in the Amazon [117]. This can generate a bias effect that forces allocation of populations to exclusively road-related intervention areas. Moreover, errors in non-forest masks (e.g., confusion with nonanthropogenic deforestation events, misclassified pixels) could add additional noise that may explain why populations were allocated to areas not known to be occupied (see eastern side of canton CH in Fig 6B, which is a ridge). While identifying and eliminating these artifacts are important tasks in this approach, our recommendation is that future research must improve the methods of rural population mapping before applying DAS, such as using products from Nightlights products with higher spatial resolution than the used in this research.

1.5.3 Advantages and limitations of GWR and RF

GWR is a proven methodology for capturing spatial nonstationary relationships, not as a global overview but as a local estimate [118]. However, the use of this approach depends on its calibration (especially for the *bw* parameter) and variable selection to reduce its sensitivity to multicollinearity. While some studies have proposed different ways to do this [119,120,121], in this study we present a novel approach using the RF algorithm. Even though it was not possible to determine the impact of variables directly from *LVI*, we could analyze all proposed variables, no matter their multicollinearity, noise, or even type. This represents an advantage in overcoming multicollinearity in GWR, but also selection bias effects [122], which are more complex to control in multivariate problems. Moreover, *LVI* and its mapping showed the predictive power of selected variables that helped not only to identify those relevant for modeling but also their spatial extent. This subsequently facilitated to extract regions with similar physical and human impact characteristics, which allowed us to discuss with more detail where spatial determinants were relevant to *DEF* and *REF*. This strengthens the need for better strategies in land planning. Also future work is needed to explore more applications of GWRFC, as we do not discuss other additional results (e.g. prediction probabilities) or RF model interpretation approaches (e.g. partial dependence plots), which are also possible to derive using this methodology (See S3 Appendix). Furthermore, clustering of *LVI* spatial representations to later extract the impact of variables without any transformation of original values should be considered as another advantage. While Wilcoxon rank sum test allowed to identify a similitude or difference between rates and variables; it was the Cliff's Delta test, which gave additional detail to quantify these findings. Critics of null hypothesis significance testing [123, 124] and GWR [125], may have found this procedure more convenient than application of parametric approaches in GWR, as assumptions failures and bias effects are less relevant to non-parametric algorithms such as RF. Its further *LVI* clustering and identification of focus areas allowed to conduct exploratory analysis of original data and statistical tests to better discriminate specific variables interactions.

Nevertheless, some limitations of the proposed methodology are important to mention. Our approach does not constitute a GWR but rather a geographically weighted RF classification (GWRFC; see [S3 Appendix](#)). As seen in Section 1.4.2, RF regression achieved a relatively poor performance with respect to GWRFC, forcing our FDD analysis from a regression to a classification problem. That is why we discretized our independent variables into classes and up-sampled unbalanced cases during RF training. The latter operation allowed the GWRFC to obtain acceptable results, as other studies also found [[126](#)]; however, is unknown whether unbalanced sampling affected RF regression as well. Breiman [[41](#)] also warned the limited performance of RF regression that may be also applied to our results. This highlights the need for further research, and our recommendation is that experimentation with other nonparametric algorithms, especially for regression analysis (e.g., support vector machines, neural networks) should be considered, as novel studies has shown [[127](#)]. However, we must caution that GWR is more accepted as an exploratory or interpolation technique rather than a predictive tool [[128,129](#)], something already known in the literature but with only marginal discussion [[130](#)].

1.5.4 Linkages between DEF and REF in the NEA

Prior works on FDD analysis in Latin America have documented contrasting dynamics where population growth, socioeconomic development, and agricultural expansion affect *DEF* and *REF* differently [[131](#)]. In our research, we extend these findings to identify more specific and localized FDDs, which enrich the explanations from those already known in the NEA (see Section 1.1). With respect to their location, our results indicated two hotspots that highlighted the tendency for *DEF* and *REF* to be spatially clustered, which supports Fagua et al. [[132](#)], showing that forest change is not an accidental process, but rather is determined by geographical location and intensity. In this regard, *DEF* showed a relationship with intense land-use changes associated with oil extraction, increasing nightlight intensity, suitability for commercial agriculture [[133](#)], and accesibility to facilities (especially for palm oil, coffee and cacao). This landscape verifies the expansion of the oil industry, economic oportunities, and colonization of the northern and western Ecuadorian Amazon [[21,57,134](#)]. This is in contrast to *REF*, as its biophysical setting (>482 m.a.n.m.) indicated less suitability for commercial agriculture (except for coffee and cacao) and diminished accesibility to their facilities. Moreover, an increasing distance from oil wells but less distance from mining blocks indicates other natural resource extraction interests [[135](#)]. Here, agroforestry systems with patches less than 4 ha combining secondary forests, cacao, and coffee plantations seem to dominate the landscape. This is similar to the traditional “chakra” land-use system described by Torres et al. [[136](#)] and to naturally regenerated forests as a consequence of the abandonment of degraded pastures due to nutritional limitations of soils in the region [[137,138](#)]. The latter may explain why accessibility to milk production facilities was better than other related agriculture products in *REF*, but also compares well with Rudel et al. [[9](#)] for the highest probability of *REF* at the shortest distances to roads. This was manifested especially in abandoned pastures where *colonos* experienced an important out-migration from the late 1980s, as is also described by Carr [[81](#)] as a regional trend in Latin America. Nevertheless, future research may consider incorporate migration censuses to corroborate these findings and identify where they manifest locally.

These differences between *DEF* and *REF* landscapes were also reflected in their demographic structure. Despite people of all ages (especially older adults, i.e. 45–72 y) from *colonos* and *Kichwas* groups related to agricultural activities and secondary education showing more of a link to *DEF* than *REF*, we found some variables that highlighted their particularities. In this respect, the diminishing trend of high fertility and large families, which favors more *REF* than *DEF*, is remarkable. This resembles the demographic and forest transition theories

[11,139] that fit well considering the economic development after the discovery of oil in the Ecuadorian Amazon. Another finding in this direction is increasing education years that seems to have a positive effect on *DEF* and *REF* but the latter only when is related to higher education. This suggests environmental externalities produced by education, as has also been reported in other regions [140,141]. Furthermore, a slight *DEF* association was seen where the male population exceeded the female population, a phenomenon observed in other studies [142,143], but it was not true at all in our case, as chief female households were more strongly associated with *DEF*, similar to the results of Sellers [144]. However, we found that the number of chief male households exceeded female households for both *DEF* and *REF* indicated different proportions. This suggests that land-tenure and land-use decision-making is mostly dominated by males in the NEA, but this could be different among ethnicities, since a diminished effect in both *DEF* and *REF* was observed for the *Kichwa* group respect to *colonos* but little or no effect compared with the *Huaorani* group. This agrees with the results of Sierra et al. [145], who reported levels of *DEF* and *REF* of 42.7% and 35.7% in the *Kichwa* territory and 0.3% and 0.4% in the *Huaorani* territory for almost the same period of time (2000–2008). This indicates that language (as a proxy of ethnicity) together with gender should be considered in future research to better characterize and discuss FDDs, as other studies have also suggested [146,147].

1.6 Conclusions

This research underlines the importance of downscaling global problems to the local scale and assessing individual drivers of land-use change in coupled socioecological systems. Applying an experimental methodology fusing remote sensing time series products, dasymetric mapping, and GWRFC, we were able to support the analysis of the spatial distribution of the population and forest dynamics in the Ecuadorian Amazon in more detail. Our findings reveal that at the local scale, key FDDs identified at the global scale can be better described. This was demonstrated in our study, as different groups played different roles in forest change, with varying impact in different regions in the NEA. Accessibility to agricultural collection centers and distance to infrastructure had an influence on both *DEF* and *REF*. However, biophysical and land-cover variable groups demonstrated that they could not be minimized, since they are ancillary sources that support and corroborate findings focused on them, i.e., describing suitable conditions for agriculture or natural resource extraction. Furthermore, socioeconomic and sociocultural variable groups had a strong influence on untangling population dynamics and their relationship with forest change, which made interpreting the results challenging and final statements fuzzier. Nevertheless, combining forest dynamics and population information in a geospatial environment underlines their variable complexity and extent. Combining aspects of livelihood patterns can be more meaningful than using proxies to represent individual aspects. The results of this study also highlight the roles of education, gender, and language in forest dynamics, which are more studied in social sciences but therefore show a strong relevance also for environmental studies. Interdisciplinary expertise and transdisciplinary exchange are needed to foster a better understanding of coupled socioecological systems from local to global scales. This can only be facilitated by inter- and transdisciplinary research.

Supporting information

S1 Appendix. Processing time for multiple bandwidth sizes using a sample dataset of 1000 obs. with 34 variables.

(JPG)

S2 Appendix. Correlations between LVI, DEF and REF rates.

(JPG)

S3 Appendix. Source code of GWRFC algorithm. <https://github.com/FSantosCodes/GWRFC>.

(DOCX)

Acknowledgments

We want to thanks to our colleagues from the Center for Development Research (ZFL), Center for Remote Sensing of Land Surfaces (ZEF) and Remote Sensing Research Group (RSRG) in Bonn-Germany. We also thanks to colleagues from the AI Saturdays-Quito, Universidad Tecnológica Indoamérica, and Ministerio del Ambiente in Quito-Ecuador for their comments during this research. Special thanks to the Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) and Universidad Tecnológica Indoamérica for funding this research. The authors also thanks to R, OSGeo, USGS and Google Earth Engine communities for sharing their knowledge and provide no-cost tools and data. Thanks also to two anonymous reviewers and José Jara for their comments on an earlier version of this paper.

Author Contributions**Conceptualization:** Fabián Santos, Valerie Graw.**Data curation:** Fabián Santos.**Formal analysis:** Fabián Santos.**Funding acquisition:** Fabián Santos, Santiago Bonilla.**Investigation:** Fabián Santos.**Methodology:** Fabián Santos, Valerie Graw, Santiago Bonilla.**Project administration:** Fabián Santos.**Resources:** Fabián Santos, Valerie Graw, Santiago Bonilla.**Software:** Fabián Santos.**Supervision:** Fabián Santos, Valerie Graw.**Validation:** Fabián Santos, Valerie Graw, Santiago Bonilla.**Visualization:** Fabián Santos.**Writing – original draft:** Fabián Santos.**Writing – review & editing:** Valerie Graw, Santiago Bonilla.**References**

1. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. 2001; 51: 933–938.
2. Armenteras D, Rodríguez N, Retana J, Morales M. Understanding deforestation in montane and lowland forests of the Colombian Andes. *Reg Environ Chang*. 2011; 11: 693–705. <https://doi.org/10.1007/s10113-010-0200-y>
3. Brooks TM, Mittermeier AR, da Fonseca GBA, Gerlach J, Hoffmann M, Lamoreux JF, et al. Global Biodiversity Conservation Priorities. 2006; 313: 58–62.
4. Myers N, Mittermeier R, Mittermeier C, Fonseca G da, Kent J. Biodiversity hotspots for conservation priorities. *Nature*. 2000; 403: 853–858. <https://doi.org/10.1038/35002501> PMID: 10706275

5. Cincotta RP, Wisniewski J, Engelman R. Human population in the biodiversity hotspots. *Nat.* 2000; 404: 990–992. <https://doi.org/10.1038/35010105> PMID: 10801126
6. Armenteras D, María J, Rodríguez N, Retana J. Deforestation dynamics and drivers in different forest types in Latin America: Three decades of studies (1980–2010). *Glob Environ Chang.* 2017; 46: 139–147. <https://doi.org/10.1016/j.gloenvcha.2017.09.002>
7. Buytaert W, Cuesta-Camacho F, Tobón C. Potential impacts of climate change on the environmental services of humid tropical alpine regions. *Glob Ecol Biogeogr.* 2011; 20: 19–33. <https://doi.org/10.1111/j.1466-8238.2010.00585.x>
8. Grau HR, Aide M. Globalization and land-use transitions in Latin America. *Ecol Soc.* 2008; 13. <https://doi.org/10.1057/9780230603554>
9. Rudel TK, Bates D, Machinguishi R. A tropical forest transition? Agricultural change, out-migration, and secondary forests in the Ecuadorian Amazon. *Ann Assoc Am Geogr.* 2002; 92: 87–102. <https://doi.org/10.1111/1467-8306.00281>
10. Nagendra H. Drivers of reforestation in human-dominated forests. *Proc Natl Acad Sci.* 2007; 104: 15218–15223. <https://doi.org/10.1073/pnas.0702319104> PMID: 17881576
11. Rudel TK, Coomes OT, Moran E, Achard F, Angelsen A, Xu J, et al. Forest transitions: Towards a global understanding of land use change. *Glob Environ Chang.* 2005; 15: 23–31. <https://doi.org/10.1016/j.gloenvcha.2004.11.001>
12. Mosandl R, Günter S, Stimm B, Weber M. Ecuador Suffers the Highest Deforestation Rate in South America. In: Beck E, Bendix J, Kottke I, Makeschin F, Mosandl R, editors. *Gradients in a Tropical Mountain Ecosystem of Ecuador.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. pp. 37–40. https://doi.org/10.1007/978-3-540-73526-7_4
13. FAO. *State of the World's Forests.* 2007.
14. Geist HJ, Lambin EF. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *Bio-science.* 2002; 52: 143. [https://doi.org/10.1641/0006-3568\(2002\)052\[0143:PCAUDF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2)
15. Salvini G, Herold M, De Sy V, Kissinger G, Brockhaus M, Skutsch M. How countries link REDD+ interventions to drivers in their readiness plans: implications for monitoring systems. *Environ Res Lett.* 2014; 9: 074004. <https://doi.org/10.1088/1748-9326/9/7/074004>
16. Da Ponte E, Fleckenstein M, Leinenkugel P, Parker A, Oppelt N, Kuenzer C. Tropical forest cover dynamics for Latin America using Earth observation data: a review covering the continental, regional, and local scale. *Int J Remote Sens.* 2015; 36: 3196–3242. <https://doi.org/10.1080/01431161.2015.1058539>
17. Grainger A. Difficulties in tracking the long-term global trend in tropical forest area. *Proc Natl Acad Sci.* 2008; 105: 818–823. <https://doi.org/10.1073/pnas.0703015105> PMID: 18184819
18. Meyfroidt P, Lambin EF, Erb KH, Hertel TW. Globalization of land use: Distant drivers of land change and geographic displacement of land use. *Curr Opin Environ Sustain.* 2013; 5: 438–444. <https://doi.org/10.1016/j.cosust.2013.04.003>
19. Mertens B, Sunderlin WD, Ndoye O, Lambin EF. Impact of macroeconomic change on deforestation in South Cameroon: Integration of household survey and remotely-sensed data. *World Dev.* 2000; 28: 983–999. [https://doi.org/10.1016/S0305-750X\(00\)00007-3](https://doi.org/10.1016/S0305-750X(00)00007-3)
20. Southgate D, Sierra R, Brown L. The causes of tropical deforestation in Ecuador: A statistical analysis. *World Dev.* 1991; 19: 1145–1151. [https://doi.org/10.1016/0305-750X\(91\)90063-N](https://doi.org/10.1016/0305-750X(91)90063-N)
21. Mena CF, Bilborrow RE, McClain ME. Socioeconomic drivers of deforestation in the Northern Ecuadorian Amazon. *Environ Manage.* 2006; 37: 802–815. <https://doi.org/10.1007/s00267-003-0230-z> PMID: 16555027
22. Stephen J, Walsh, Yang Shao, Carlos F. Mena, McCleary and AL. Integration of Hyperion Satellite Data and Household Social Survey to Characterize the Causes and Consequences of Reforestation Patterns in the Northern Ecuadorian Amazon. *Photogramm Eng Remote Sens.* 2008; 74: 725–735.
23. Bonilla-Bedoya S, Estrella-Bastidas A, Molina JR, Herrera MÁ. Socioecological system and potential deforestation in Western Amazon forest landscapes. *Sci Total Environ.* 2018; 644: 1044–1055. <https://doi.org/10.1016/j.scitotenv.2018.07.028> PMID: 30743818
24. Wulder MA, Hilker T, White JC, Coops NC, Masek JG, Pflugmacher D, et al. Virtual constellations for global terrestrial monitoring. *Remote Sens Environ.* 2015; 170: 62–76. <https://doi.org/10.1016/j.rse.2015.09.001>
25. Hansen MC, Potapov P, Moore R, Hancher M, Turubanova S, Tyukavina A, et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science (80-).* 2013; 342: 850–853. <https://doi.org/10.1126/science.1244693> PMID: 24233722

26. Kennedy RE, Yang Z, Braaten J, Copass C, Antonova N, Jordan C, et al. Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. *Remote Sens Environ.* 2015; 166: 271–285. <https://doi.org/10.1016/j.rse.2015.05.005>
27. Hernández A, Miranda MD, Arellano EC, Dobbs C. Landscape trajectories and their effect on fragmentation for a Mediterranean semi-arid ecosystem in Central Chile. 2016; 127: 74–81. <https://doi.org/10.1016/j.jaridenv.2015.10.004>
28. Logan JR, Xu Z, Stults B. 1970 to 2010: A Longitudinal Tract Database. *Prof Geogr.* 2014; 66: 412–420. <https://doi.org/10.1080/00330124.2014.905156> PMID: 25140068
29. Holt DT, Steel DG, Tranmer M, Wrigley N. Aggregation and Ecological Effects in Geographically Based Data. *Geogr Anal.* 1996; 28: 244–261. <https://doi.org/10.1111/j.1538-4632.1996.tb00933.x>
30. Krivoruchko K, Gribov A, Krause E. Multivariate areal interpolation for continuous and count data. *Procedia Environ Sci.* 2011; 3: 14–19. <https://doi.org/10.1016/j.proenv.2011.02.004>
31. Semenov Tian-Shansky. *Metody dazimetrii (Methods of Dasymetric Mapping)*. Dazimetrichskaya Karta Evropeiskoi. Petrograd: Scientific Chemistry and Technology Publishing; 1923. pp. 18–26.
32. Tobler WR. Smooth pycnopylactic interpolation for geographical regions. *J Am Stat Assoc.* 1979; 74: 519–530. <https://doi.org/10.1080/01621459.1979.10481647> PMID: 12310706
33. Kraus SP, Senger LW, Ryerson JM. Estimating population from photographically determined residential land use types. *Remote Sens Environ.* 1974; 3: 35–42. [https://doi.org/10.1016/0034-4257\(74\)90036-4](https://doi.org/10.1016/0034-4257(74)90036-4)
34. Stevens FR, Gaughan AE, Linard C, Tatem AJ. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS One.* 2015; 10: 1–22. <https://doi.org/10.1371/journal.pone.0107042> PMID: 25689585
35. Petrov A. One Hundred Years of Dasymetric Mapping: Back to the Origin. *Cartogr J.* 2012; 49: 256–264. <https://doi.org/10.1179/1743277412Y.0000000001>
36. Zandbergen PA, Ignizio DA. Comparison of Dasymetric Mapping Techniques for Small-Area Population Estimates. *Cartogr Geogr Inf Sci.* 2010; 37: 199–214. <https://doi.org/10.1559/152304010792194985>
37. Brunsdon C, Fotheringham A, Charlton ME. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geogr Anal.* 1996; 28: 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
38. Pineda Jaimes NB, Bosque Sendra J, Gómez Delgado M, Franco Plata R. Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. *Appl Geogr.* 2010; 30: 576–591. <https://doi.org/10.1016/j.apgeog.2010.05.004>
39. de Freitas MWD, Santos JR dos, Alves DS. Land-use and land-cover change processes in the Upper Uruguay Basin: Linking environmental and socioeconomic variables. *Landsc Ecol.* 2013; 28: 311–327. <https://doi.org/10.1007/s10980-012-9838-9>
40. Wheeler D, Tiefelsdorf M. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *J Geogr Syst.* 2005; 7: 161–187. <https://doi.org/10.1007/s10109-005-0155-6>
41. Breiman L. Random forests. *Mach Learn.* 2001; 45: 5–32. <https://doi.org/10.1023/A:1010933404324>
42. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. 2015; 77: 1–17. <https://doi.org/10.18637/jss.v077.i01>
43. Georganos S, Grippa T, Niang Gadiaga A, Linard C, Lennert M, Vanhuysse S, et al. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* 2019; 0: 1–16. <https://doi.org/10.1080/10106049.2019.1595177>
44. Anderson A. Dollarization: A Case Study of Ecuador. *Imp J Interdiscip Res.* 2016; 2: 2454–1362. Available: <http://www.onlinejournal.in>
45. Sierra R, Campos F, Chamberlin J. Assessing biodiversity conservation priorities: Ecosystem risk and representativeness in continental Ecuador. *Landsc Urban Plan.* 2002; 59: 95–110. [https://doi.org/10.1016/S0169-2046\(02\)00006-3](https://doi.org/10.1016/S0169-2046(02)00006-3)
46. MAE. Metodología para la representación Cartográfica de los Ecosistemas del Ecuador Continental. Subsecretaría de Patrimonio Natural, editor. Quito—Ecuador: Ministerio del Ambiente del Ecuador (MAE); 2013.
47. Huttel C, Zebrowski C, Gondard P. Paisajes Agrarios del Ecuador. IGM; 1999.
48. Eberhart N. Transformaciones agrarias en el frente de colonización de la Amazonia ecuatoriana. 1998; 186 pages.

49. Coq-Huelva D, Torres-Navarrete B, Bueno-Suárez C. Indigenous worldviews and Western conventions: Sumak Kawsay and cocoa production in Ecuadorian Amazonia. *Agric Human Values*. 2017; 0: 1–17. <https://doi.org/10.1007/s10460-017-9812-x>
50. Pierre G, Juan L, Paola S, Chiriboa M, Cuví M, Emmanuel F, et al. *Transformaciones Agrarias En El Ecuador*. Juan L, Pierre P, editors. Centro Ecuatoriano de Investigación Geográfica. Quito—Ecuador: IGM; 1988.
51. Brown L, Smith R. Frameworks of Urban System Evolution in Frontier Settings and the Ecuador Amazon. 1994; 72–96.
52. Natural Earth. Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales. 2019 [cited 24 Jan 2019]. Available: <https://www.naturalearthdata.com/>
53. IGM. Base escala 1:50.000 y 250.000. Quito—Ecuador: IGM; 2011.
54. MAE-MAGAP. Protocolo metodológico para la elaboración del Mapa de cobertura y uso de la tierra del Ecuador continental 2013–2014, escala 1:100.000. Ministerio del Ambiente (MAE), Ministerio de Agricultura, Ganadería, Acuacultura y Pesca (MAGAP); 2015. <https://doi.org/10.1017/CBO9781107415324.004>
55. Perreault T. Developing Identities: Indigenous Mobilization, Rural Livelihoods, and Resource Access in Ecuadorian Amazonia. *Ecumene*. 2001; 8.
56. MAE. Documentation of the information used for the establishment of Ecuador's Forest Reference Emission Level. 2017 [cited 16 Aug 2017]. Available: <http://suia.ambiente.gob.ec/web/suia/anexos-nivel-referencia>
57. Sierra R. Dynamics and patterns of deforestation in the western Amazon: The Napo deforestation front, 1986–1996. *Appl Geogr*. 2000; 20: 1–16. [https://doi.org/10.1016/S0143-6228\(99\)00014-4](https://doi.org/10.1016/S0143-6228(99)00014-4)
58. Wasserstrom R, Southgate D. Deforestation, Agrarian Reform and Oil Development in Ecuador, 1964–1994. *Nat Resour*. 2013; 04: 31–44. <https://doi.org/10.4236/nr.2013.41004>
59. Barrera D. Gestión del territorio y manejo de bienes comunes en contextos extractivos: una aproximación al caso de las comunidades Kichwas del Cantón Arajuno en la Provincia de Pastaza, Ecuador. Facultad Latinoamericana de Ciencias Sociales (FLACSO). 2014.
60. FAO. Global Forest Resources Assessment 2015: Desk Reference. 2015. <https://doi.org/10.1002/2014GB005021>
61. Bertoli S, Moraga JFH, Ortega F. Immigration policies and the Ecuadorian exodus. *World Bank Econ Rev*. 2011; 25: 57–76. <https://doi.org/10.1093/wber/lhr004>
62. Gray CL, Bilsborrow RE. Consequences of out-migration for land use in rural Ecuador. *Land use policy*. 2014; 36: 182–191. <https://doi.org/10.1016/j.landusepol.2013.07.006> PMID: 24187416
63. R Development Core Team. The R Project for Statistical Computing, Version 3.4.3. GNU project; 2017. Available: <http://www.r-project.org/>
64. Pebesma E, Bivand R, Rowlingson B, Gomez-Rubio V, Hijmans R, Sumner M, et al. *sp: Classes and Methods for Spatial Data*, Version 1.2–5. 2017. Available: <https://github.com/edzer/sp/> <https://edzer.github.io/sp/>
65. Hijmans R, Etten J van, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, et al. *raster: Geographic Data Analysis and Modeling*, Version 2.6–7. 2017. Available: <http://www.rspatial.org/>
66. Dowle M, Srinivasen A, Gorecki J, Short T, Lianoglou S, Antonyan E. *data.table: Extension of "data frame"*, Version 1.10.4–3. 2017. Available: <http://r-datatable.com>
67. Revolution Analytics, Weston S. *foreach: Provides Foreach Looping Construct for R*, Version 1.4.3. 2015. Available: <https://cran.r-project.org/package=foreach>
68. Wickham H, Chang W. Package 'ggplot2', version 2.2.1. 2016. <https://doi.org/10.1093/bioinformatics/btr406>
69. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project; 2019. Available: <https://www.qgis.org/en/site/index.html>
70. Santos F, Meneses P, Hostert P. Monitoring long-term forest dynamics with scarce data: a multi-date classification implementation in the Ecuadorian Amazon. *Manuscr Submitt Publ*. 2018.
71. USGS. Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) On Demand Interface. 2014 [cited 14 Jan 2017]. Available: <https://espa.cr.usgs.gov>
72. Puyravaud JP. Standardizing the calculation of the annual rate of deforestation. *For Ecol Manage*. 2003; 177: 593–596. [https://doi.org/10.1016/S0378-1127\(02\)00335-3](https://doi.org/10.1016/S0378-1127(02)00335-3)
73. Lambin EF, Turner BL, Geist HJ, Agbola SB, Angelsen A, Folke C, et al. The causes of land-use and land-cover change: moving beyond the myths. 2001; 11: 261–269.

74. NOAA. Version 4 DMSP-OLS Nighttime Lights Time Series. 2019 [cited 5 Mar 2019]. Available: <https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html>
75. Ghosh T, Anderson SJ, Elvidge CD, Sutton PC. Using nighttime satellite imagery as a proxy measure of human well-being. *Sustain*. 2013; 5: 4988–5019. <https://doi.org/10.3390/su5124988>
76. Oda T, Maksyutov S. A very high-resolution (1km×1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmos Chem Phys*. 2011; 11: 543–556. <https://doi.org/10.5194/acp-11-543-2011>
77. Proville J, Zavala-Araiza D, Wagner G. Night-time lights: A global, long term look at links to socio-economic trends. *PLoS One*. 2017; 12: 1–12. <https://doi.org/10.1371/journal.pone.0174610> PMID: 28346500
78. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens Environ*. 2017; 202: 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
79. INEC. Censo de Población y Vivienda 2010. Instituto Nacional de Estadísticas y Censos; 2010. Available: <http://www.inec.gob.ec/estadisticas/>
80. INEC. Censo de Población y Vivienda 2001. Instituto Nacional de Estadísticas y Censos; 2001. Available: <http://www.inec.gob.ec/estadisticas/>
81. Carr D. Population and deforestation: Why rural migration matters. *Prog Hum Geogr*. 2009; 33: 355–378. <https://doi.org/10.1177/0309132508096031> PMID: 20485541
82. Mennis J. Generating Surface Models of Population Using Dasymetric Mapping. *Prof Geogr*. 2003; 55: 31–42.
83. SIGTIERRAS U-M-P. Metodología Accesibilidad. Proyecto: “Levantamiento de cartografía temática escala 1:25000, lote 2”. Temáticas Nacionales. 2015. Available: <http://metadatos.sigtierras.gob.ec:8080/geonetwork/srv/spa/catalog.search#/search?resultType=details&any=accesibilidad&from=1&to=20&sortBy=relevance>
84. Pan W, Walsh SJ, Bilsborrow RE, Frizzelle BG, Erlien CM, Baquero F. Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. *Agric Ecosyst Environ*. 2004; 101: 117–134. <https://doi.org/10.1016/j.agee.2003.09.022>
85. Langford M, Unwin DJ. Generating and mapping population density surfaces within a geographical information system. *Cartogr J*. 1994; 31: 21–26. <https://doi.org/10.1179/000870494787073718> PMID: 12288211
86. Gray CL, Bilsborrow RE, Bremner JL, Lu F. Indigenous Land Use in the Ecuadorian Amazon: A Cross-cultural and Multilevel Analysis. *Hum Ecol*. 2008; 36: 97–109. <https://doi.org/10.1007/s10745-007-9141-6>
87. Jian W. The relationship between culture and language. *ELT J*. 2000; 54/4: 328–334.
88. SNI. Archivos de Información Geográfica—Sistema Nacional de Información (SNI). 2017 [cited 20 Dec 2017]. Available: <http://sni.gob.ec/coberturas>
89. NOAA. Version 4 DMSP-OLS Nighttime Lights Time Series. 2019.
90. Gollini I, Lu B, Charlton M, Brunsdon C, Harris P. GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. 2013. <https://doi.org/10.1080/10095020.2014.917453>
91. Gao J, Li S. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Appl Geogr*. 2011; 31: 292–302. <https://doi.org/10.1016/j.apgeog.2010.06.003>
92. Guo L, Ma Z, Zhang L. Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Can J For Res*. 2008; 38: 2526–2534. <https://doi.org/10.1139/X08-091>
93. Farber S, Páez A. A systematic investigation of cross-validation in GWR model estimation: Empirical analysis and Monte Carlo simulations. *J Geogr Syst*. 2007; 9: 371–396. <https://doi.org/10.1007/s10109-007-0051-3>
94. Chiu D. *Machine Learning with R Cookbook*. Birmingham, UK: Packt Publishing Ltd.; 2015.
95. Belgiu M, Drăgu L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J Photogramm Remote Sens*. 2016; 114: 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
96. Strobl C, Malley J, Gerhard T. An Introduction to Recursive Partitioning: Rationale, Application Psychol Methods. *Psychol Methods*. 2009; 14: 323–348. <https://doi.org/10.1037/a0016973> PMID: 19968396
97. Binbin L, Harris P, Charlton M, Brunsdon C, Nakaya T, Gollini I. GWmodel: Geographically-Weighted Models, Version 2.0–5. 2017. Available: <http://gwr.nuim.ie/>

98. Genuer R, Poggi J, Tuleau-malot C. Variable selection using Random Forests. *Pattern Recognit Lett*. 2012; 31: 2225–2236.
99. Saary MJ. Radar plots: a useful way for presenting multivariate health care data. *J Clin Epidemiol*. 2008; 61: 311–317. <https://doi.org/10.1016/j.jclinepi.2007.04.021> PMID: 18313553
100. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J*. 2016; 8: 289–317. Available: <https://www.ncbi.nlm.nih.gov/pubmed/27818791> PMID: 27818791
101. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J Stat Softw*. 2014; 61. <https://doi.org/10.18637/jss.v061.i06>
102. Wilcoxon F. Individual Comparisons by Ranking Methods. *Biometrics Bull*. 1945; 1: 80–83. <https://doi.org/10.2307/3001968>
103. Cliff N. *Ordinal Methods for Behavioral Data Analysis*. Taylor & Francis Group; 2016. Available: <https://books.google.com.ec/books?id=yPvbjwEACAAJ>
104. Romano J, Kromrey JD, Coraggio J, Skowronek J, Devine L. Exploring methods for evaluating group differences on the NSSE and other surveys: Are the t-test and Cohen's d indices the most appropriate choices? *Annu Meet South Assoc Institutional Res*. 2006; 14–17. <https://doi.org/10.1017/CBO9781107415324.004>
105. Hermosilla T, Wulder MA, White JC, Coops NC, Hobart GW. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sens Environ*. 2015; 158: 220–234. <https://doi.org/10.1016/j.rse.2014.11.005>
106. Oeser J, Pflugmacher D, Senf C, Heurich M, Hostert P. Using intra-annual Landsat time series for attributing forest disturbance agents in Central Europe. *Forests*. 2017; 8. <https://doi.org/10.3390/f8070251>
107. Rufin P, Müller H, Pflugmacher D, Hostert P. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. *Int J Appl Earth Obs Geoinf*. 2015; 41: 1–10. <https://doi.org/10.1016/j.jag.2015.04.010>
108. Griffiths P, Nendel C, Pickert J, Hostert P. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sens Environ*. 2019; 1–12. <https://doi.org/10.1016/j.rse.2019.03.017>
109. Sirro L, Häme T, Rauste Y, Kilpi J, Hämäläinen J, Gunia K, et al. Potential of different optical and SAR data in forest and land cover classification to support REDD+ MRV. *Remote Sens*. 2018; 10. <https://doi.org/10.3390/rs10060942>
110. Dostálová A, Wagner W, Milenković M, Hollaus M. Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. *Int J Remote Sens*. 2018; 39: 7738–7760. <https://doi.org/10.1080/01431161.2018.1479788>
111. Hu Q, Sulla-Menashe D, Xu B, Yin H, Tang H, Yang P, et al. A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *Int J Appl Earth Obs Geoinf*. 2019; 80: 218–229. <https://doi.org/10.1016/j.jag.2019.04.014>
112. Ravanelli R, Nascetti A, Cirigliano RV, Di Rico C, Leuzzi G, Monti P, et al. Monitoring the impact of land cover change on surface urban heat island through Google Earth Engine: Proposal of a global methodology, first applications and problems. *Remote Sens*. 2018; 10: 1–21. <https://doi.org/10.3390/rs10091488>
113. Myers A. Camp Delta, Google Earth and the ethics of remote sensing in archaeology. *World Archaeol*. 2010; 42: 455–467. <https://doi.org/10.1080/00438243.2010.498640>
114. Lo CP, Yang X. Drivers of Land-Use / Land-Cover Changes and Dynamic Modeling for the Atlanta, Georgia Metropolitan Area. October. 2002; 68: 1073–1082. Cited By (since 1996) 5\rExport Date 23 November 2011
115. Reibel M, Agrawal A. Areal interpolation of population counts using pre-classified land cover data. *Popul Res Policy Rev*. 2007; 26: 619–633. <https://doi.org/10.1007/s11113-007-9050-9>
116. Pappalardo SE, De Marchi M, Ferrarese F. Uncontacted Waorani in the Yasuní Biosphere Reserve: Geographical Validation of the Zona Intangible Tagaeri Taromenane (ZITT). *PLoS One*. 2013; 8: 21–25. <https://doi.org/10.1371/journal.pone.0066293> PMID: 23840436
117. Schillinger K, Lycett SJ. The Flow of Culture: Assessing the Role of Rivers in the Inter-community Transmission of Material Traditions in the Upper Amazon. *J Archaeol Method Theory*. 2019; 26: 135–154. <https://doi.org/10.1007/s10816-018-9369-z>
118. Fotheringham A, Charlton ME, Brunson C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ Plan A*. 1998; 30: 1905–1927. <https://doi.org/10.1068/a301905>

119. Tu J. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Appl Geogr.* 2011; 31: 376–392. <https://doi.org/10.1016/j.apgeog.2010.08.001>
120. Wheeler D. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environ Plan A.* 2007; 39: 2464–2481. <https://doi.org/10.1068/a38325>
121. Harris P, Fotheringham AS, Juggins S. Robust Geographically Weighted Regression: A Technique for Quantifying Spatial Relationships Between Freshwater Acidification Critical Loads and Catchment Attributes. *Ann Assoc Am Geogr.* 2010; 100: 286–306. <https://doi.org/10.1080/00045600903550378>
122. Ellenberg JH. Selection bias in observational and experimental studies. *Stat Med.* 1994; 13: 557–567. <https://doi.org/10.1002/sim.4780130518> PMID: 8023035
123. J C. The Earth is Round ($p < .05$). *Am Psychol.* 2004; 49: 997–1003. Available: http://d.wanfangdata.com.cn/NSTLQK_10.1037-0003-066X.49.12.997.aspx%0Apapers2://publication/livfe/id/199
124. Kim J, Bang H. Three common misuses of P values. *Dent Hypotheses.* 2016; 7: 73. <https://doi.org/10.4103/2155-8213.190481> PMID: 27695640
125. Cho S, Lambert DM, Kim SG, Jung S. Extreme Coefficients in Geographically Weighted Regression and Their Effects on Mapping. *GIScience Remote Sens.* 2009; 46: 273–288. <https://doi.org/10.2747/1548-1603.46.3.273>
126. Tantithamthavorn C, Hassan AE, Matsumoto K. The Impact of Class Rebalancing Techniques on the Performance and Interpretation of Defect Prediction Models. *IEEE Trans Softw Eng.* 2018; 1–20. <https://doi.org/10.1109/TSE.2018.2876537>
127. Li L. Geographically weighted machine learning and downscaling for high-resolution spatiotemporal estimations of wind speed. *Remote Sens.* 2019; 11. <https://doi.org/10.3390/rs11111378>
128. Páez A, Farber S, Wheeler D. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environ Plan A.* 2011; 43: 2992–3010. <https://doi.org/10.1068/a44111>
129. Wheeler DC. Geographically weighted regression. *Handb Reg Sci.* 2019; 1–27.
130. Harris P, Brunsdon C, Fotheringham AS. Links, comparisons and extensions of the geographically weighted regression model when used as a spatial predictor. *Stoch Environ Res Risk Assess.* 2011; 25: 123–138. <https://doi.org/10.1007/s00477-010-0444-6>
131. Aide TM, Clark LM, Grau R, López-Carr D, Levy AM, Redo D, et al. Deforestation and Reforestation of Latin America and the Caribbean (2001–2010). *Biotropica.* 2012; 1: 1–10.
132. Fagua JC, Baggio JA, Ramsey RD. Drivers of forest cover changes in the Chocó-Darien Global Ecoregion of South America. *Ecosphere.* 2019; 10: e02648. <https://doi.org/10.1002/ecs2.2648>
133. Pirker J, Mosnier A, Kraxner F, Havlík P, Obersteiner M. What are the limits to oil palm expansion? *Glob Environ Chang.* 2016; 40: 73–81. <https://doi.org/10.1016/j.gloenvcha.2016.06.007>
134. Castro M, Sierra R, Calva O, Camacho J, Lopez F. Zonas de Procesos Homogéneos de Deforestación del Ecuador. Factores promotores y tendencias al 2020. Quito—Ecuador: Programa GESOREN-GIZ y Ministerio de Ambiente del Ecuador.; 2013. <https://doi.org/10.13140/2.1.3210.2081>
135. Finer M, Jenkins CN. Proliferation of hydroelectric dams in the andean amazon and implications for andes-amazon connectivity. *PLoS One.* 2012; 7: 1–9. <https://doi.org/10.1371/journal.pone.0035126> PMID: 22529979
136. Torres B, Maza OJ, Aguirre P, Hinojosa L, Günter S. The Contribution of Traditional Agroforestry to Climate Change Adaptation in the Ecuadorian Amazon: The Chakra System. In: Leal Filho W, editor. *Handbook of Climate Change Adaptation.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2015. pp. 1973–1994. https://doi.org/10.1007/978-3-642-38670-1_102
137. Custode E, Sourdat M. Paisajes y suelos de la Amazonía ecuatoriana: entre la conservación y la explotación. *Cult del Banco Cent Ecuador.* 1986; 8: 325–338.
138. Bonilla-Bedoya S, López-Ulloa M, Vanwalleghem T, Herrera-Machuca MÁ. Effects of Land Use Change on Soil Quality Indicators in Forest Landscapes of the Western Amazon. *Soil Sci.* 2017; 182. Available: https://journals.lww.com/soilsci/Fulltext/2017/04000/Effects_of_Land_Use_Change_on_Soil_Quality.2.aspx
139. Kirk D. Demographic Transition Theory. *Popul Stud (NY).* 1996; 50: 361–387. <https://doi.org/10.1080/0032472031000149536> PMID: 11618374
140. Godoy R, Groff S, O'Neill K. The Role of Education in Neotropical Deforestation: Household Evidence from Amerindians in Honduras. *Hum Ecol.* 1998; 26. <https://doi.org/10.1007/s10988-000-0649-9>
141. Dolisca F, McDaniel JM, Teeter LD, Jolly CM. Land tenure, population pressure, and deforestation in Haiti: The case of Forêt des Pins Reserve. *J For Econ.* 2007; 13: 277–289. <https://doi.org/10.1016/j.jfe.2007.02.006>

142. Moran EF, Siqueira A, Brondizio E. Household Demographic Structure and Its Relationship to Deforestation in the Amazon Basin. In: Fox J, Rindfuss RR, Walsh SJ, Mishra V, editors. *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Boston, MA: Springer US; 2003. pp. 61–89. https://doi.org/10.1007/0-306-48130-8_3
143. Barbieri AF, Carr DL. Gender-specific out-migration, deforestation and urbanization in the Ecuadorian Amazon. *Glob Planet Change*. 2005; 47: 99–110. <https://doi.org/10.1016/j.gloplacha.2004.10.005> PMID: 19657469
144. Sellers S. HHS Public Access. 2018; 38: 424–447. <https://doi.org/10.1007/s11111-017-0275-1.Family>
145. Sierra R. Patrones y factores de deforestación en el Ecuador continental, 1990–2010. Y un acercamiento a los próximos 10 años. Conservación Internacional Ecuador, Forest Trends and Ministry of Environment of Ecuador; 2013. p. 51. Available: http://www.forest-trends.org/documents/files/doc_3396.pdf
146. Villamor GB, Desrianti F, Akiefnawati R, Amaruzaman S, van Noordwijk M. Gender influences decisions to change land use practices in the tropical forest margins of Jambi, Indonesia. *Mitig Adapt Strateg Glob Chang*. 2014; 19: 733–755. <https://doi.org/10.1007/s11027-013-9478-7>
147. Hutchison HC, Vallejo I. La deforestación y la participación de mujeres en el manejo de recursos naturales: una comparación de casos de estudio de comunidades indígenas y colonas en la provincia de Napo, Ecuador. 2016.