Environmental Conservation

cambridge.org/enc

Deforestation risk in the Peruvian Amazon basin

Eduardo Rojas[1], Brian R Zutta[2,3], Yessenia K Velazco[3], Javier G Montoya-Zumaeta[4] and Montserrat Salvà-Catarineu[1]

[1]Department of Geography, University of Barcelona, C. Montalegre, 6, 08001 Barcelona, Spain; [2]Department of Ecology and Evolutionary Biology, University of California (UCLA), Los Angeles, CA 90095, USA; [3]Green Blue Solutions, Surco, Lima, Peru and [4]Crawford School of Public Policy, Australian National University, Canberra ACT 2600, Australia

Report

Cite this article: Rojas E et al. (2021). Deforestation risk in the Peruvian Amazon basin. *Environmental Conservation* **48**: 310–319. doi: 10.1017/S0376892921000291

Received: 20 April 2021
Revised: 10 September 2021
Accepted: 11 September 2021
First published online: 18 October 2021

Keywords:
conservation; Maxent; open-source; species distribution modelling; tropical forest

Author for correspondence:
Eduardo Rojas,
Email: eduardo2188@gmail.com

Summary

The prevention of tropical forest deforestation is essential for mitigating climate change. We tested the machine learning algorithm Maxent to predict deforestation across the Peruvian Amazon. We used official annual 2001–2019 deforestation data to develop a predictive model and to test the model's accuracy using near-real-time forest loss data for 2020. Distance from agricultural land and distance from roads were the predictor variables that contributed most to the final model, indicating that a narrower set of variables contribute nearly 80% of the information necessary for prediction at scale. The permutation importance indicating variable information not present in the other variables was also highest for distance from agricultural land and distance from roads, at 40.5% and 14.3%, respectively. The predictive model registered 73.2% of the 2020 early alerts in a high or very high risk category; less than 1% of forest cover in national protected areas were registered as very high risk, but buffer zones were far more vulnerable, with 15% of forest cover being in this category. To our knowledge, this is the first study to use 19 years of annual data for deforestation risk. The open-source machine learning method could be applied to other forest regions, at scale, to improve strategies for reducing future deforestation.

Introduction

The conversion of tropical forest land to other land uses contributes to nearly 20% of the world's greenhouse gas emissions (Achard et al. 2007), which includes activities such as gross deforestation, forest replacement by agriculture and pastureland and the establishment of monoculture forest plantations, among other activities (Gibbs et al. 2010, Hosonuma et al. 2012, Austin 2017). Several international initiatives were developed to assist developing countries with financial incentives to avoid deforestation and improve forest management, including the REDD+ (Reducing Emissions from Deforestation and Forest Degradation) mechanism, established by the Conference of Parties to the United Nations Framework Convention on Climate Change (UNFCCC) (Corbera & Schroeder 2011), as well as international funding initiatives, including the Carbon Fund of the Forest Carbon Partnership Facility (FCPF) and the Green Climate Fund developed by the UNFCCC. More local initiatives include the National Forest Conservation Program for the Mitigation of Climate Change (Programa Bosques) of the Ministry of the Environment of Peru (MINAM), developed in 2010 within the framework of the free trade agreement between the USA and Peru. As countries increase their capacity to monitor and manage their forests, many systems focus on estimating periodic deforestation from satellite imagery (Potapov et al. 2012, Romijn 2012, Petersen 2018), and a few current systems contain satellite-based near-real-time early alerts of forest loss (Hansen et al. 2016, Musinsky et al. 2018). Techniques such as kernel density mapping (West et al. 2019) and hotspot analysis (Sanchez-Cuervo & Aide 2013, Kalamandeen et al. 2018) are used to track the density of deforestation over time. However, these systems monitor past deforestation events, even if they are considered near-real-time, as satellite images may be available for days to weeks after image capture (Hansen et al. 2016).

The Peruvian Amazon contains the second largest portion of Amazon forest following Brazil, with a complex heterogeneous landscape and a variety of anthropogenic pressures that have increased deforestation in the last decade. These pressures include cacao plantations, which are a historic deforestation driver, particularly in forests at higher altitudes (Bax & Francesconi 2018), where cacao and coffee plantations are favoured at altitudes from 1000 to 2500 m, but they can also be found sporadically in the lowlands. Oil palm plantations represent a more recent driver of deforestation, with significant large-scale impact beginning *c.* 2007 in key lowland rainforest regions of San Martin and Ucayali (Vijay et al. 2018). Illicit artisanal-scale gold mining has had a devastating impact in the last decade, albeit often restricted to the rainforests of Madre de Dios (Asner & Tupayachi 2017, Espejo et al. 2018),

© The Author(s), 2021. Published by Cambridge University Press on behalf of Foundation for Environmental Conservation. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

where cattle ranching is also prevalent in the eastern area, which has similar landscape characteristics to the nearby Brazilian state of Acre (Chávez Michaelsen et al. 2013, Recanati et al. 2015).

Recent approaches in spatial modelling utilize the historic occurrence of local deforestation, along with sets of environmental and socioeconomic conditions, to identify areas at higher risk of deforestation (Aguilar-Amuchastegui et al. 2014, de Souza & De Marco 2014, 2018, Bonilla-Bedoya et al, 2018). These approaches utilize the species distribution modelling (SDM) framework, where the correlation of occurrence or presence points and environmental, bioclimatic or socioeconomic layers are processed through a machine learning-based algorithm that ultimately correlates location and multivariate space to produce a suitability map (de Souza & De Marco 2018). Machine learning algorithms have been applied in order to predict the risk of landslides (Chen 2017) and wildfires (Fonseca et al. 2016), as well as habitat alterations from fire disturbance (Franklin et al. 2014). Among the most popular machine learning approaches in SDM is the use of maximum entropy through the recently open-sourced program Maxent (Phillips 2017).

The current study aims to significantly increase the spatial resolution and modelling scale and leverage 19 years of deforestation data to present a robust deforestation risk model of the entire Peruvian Amazon at a 100m spatial resolution. To our knowledge, no other study has used this range of annual data for deforestation risk modelling and at this geographical scale combined with high spatial resolution. Previous studies on deforestation prediction and risk have either been limited to a 1km spatial resolution of predictor variables (Bax et al. 2016, Bonilla-Bedoya et al. 2018, de Souza & De Marco 2018) or 10 or fewer years of deforestation data (Aguilar-Amuchastegui et al. 2014, Bax & Francesconi 2018), and none combine higher-spatial-resolution data and variables (i.e., ≤100 m) with high-temporal-cadence deforestation data (i.e., >10 years of annual data).

We had three specific objectives. The first was to develop the risk deforestation model for the entire Peruvian Amazon at 100m spatial resolution using Maxent as the SDM framework and using 19 years of official annual deforestation data. Second, we tested the temporal transferability of the model by partitioning data in time in order to test whether a model trained with early data is able to predict later deforestation events. Lastly, we tested the spatial transferability of the model by evaluating whether a model trained with data in one area could predict deforestation in another. We hypothesized that the combined use of high-temporal-cadence deforestation data, along with high-spatial-resolution predictor variables, would produce a robust deforestation risk model that would anticipate deforestation events signalled by an independent early alert (near-real-time) system.

Material and methods

The Peruvian Amazon

The deforestation risk modelling and analysis encompassed the entire Peruvian Amazon biome delineated by the MINAM (2015a) and used in its national and international reporting of deforestation and greenhouse gas emissions. The biome is distributed within 15 departments, with most of the remaining intact forests being humid tropical forests in Loreto, Ucayali and Madre de Dios. The biome extends from the eastern slopes of the Andes from *c.* 3200 m altitude to the lowlands of the Amazon Basin and contains *c.* 690 000 km^2 of forest cover (MINAM 2015b). More than 2.3 million people live in the region, with clear trends in population concentrations and net migration rates, both positive and negative, dependent on the region (Menton & Cronkleton 2019). Deforestation in the Peruvian Amazon has steadily risen over the past 19 years, with spikes occurring in 2005 and 2010 being El Niño-related (Potapov et al. 2014). There have been several national and international initiatives to reduce deforestation (e.g., the Joint Declaration between Norway, Germany and Peru), but there has been no notable reduction in the rate of deforestation for the entire biome thus far.

Deforestation data

Forest cover loss was used as a proxy for deforestation data, with a minimum mapping unit of 30 m and an overall accuracy of 99.4% (Potapov et al. 2014). These date were obtained from the GeoBosques platform (http://geobosques.minam.gob.pe) of the National Forest Conservation Program for Climate Change Mitigation (PNCBMCC) of MINAM, which provides annual forest cover loss data from 2001 to 2019, from a year 2000 forest cover baseline, which is the basis for estimating remaining forest cover. The underlying algorithm and measurement system are based on the tropical forest cover loss system developed by the University of Maryland's Global Land Analysis and Discovery (Potapov et al. 2014). The extent of the Peruvian Amazon biome was delimited by MINAM for emissions reduction reporting, and it excludes forest cover loss due to river meandering (MINAM 2015b). Due to the volume of pixels and computational resource limits, deforestation data were resampled to the 100m pixel spatial resolution using ArcGIS 10.5 (Esri).

Predictor variables

The predictor variables selected for this study were based on potentially direct and indirect drivers of deforestation and units of conservation (Supplementary Annex A1, available online). Continuous land cover variables for the year 2000 include distance from agricultural land, distance from pastureland, distance from mining and distance from non-forest land. These land cover variables are available from the land cover maps of GeoBosques, which are derived from 30m spatial resolution Landsat image mosaics for the year 2000 and are designed to be a complementary dataset to the annual deforestation data. Distance from roads was calculated using the road database from the Ministry of Transportation and Communication (MTC). Roads included major paved, minor paved and unpaved road categories, which were combined into a single road category for this study. Distance from rivers was derived from the land cover map for the year 2000 available from GeoBosques. All distances are Euclidean distances calculated using ArcGIS 10.5 (Esri).

Categorical predictor variables included natural protected areas obtained from the National Service of Natural Areas Protected by the State (https://geo.sernanp.gob.pe) and native community boundaries obtained from the Ministry of Culture (https://geoportal.cultura.gob.pe). Both of these datasets were from the year 2019. Elevation above sea level and slope – both continuous predictor variables – were derived from the void-filled seamless Shuttle Radar Topography Mission (SRTM) digital elevation model (http://srtm.csi.cgiar.org) available at 30 m spatial resolution. All predictor variables were projected to Universal Transverse Mercator (UTM) Zone 18 South and resampled to 100-m pixels.

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

Maxent

We used Maxent (version 3.4.1) running on the R Studio for PC platform (1.2.5033) for the estimation of deforestation risk. Maxent uses presence-only point occurrences together with predictor variables to model distribution over the study area (Phillips et al. 2006). This approach identifies the correlation between multiple presence points and the combination of continuous or categorical layer values in multivariate space (Phillips et al. 2006, Phillips & Dudík 2008). Here, the 'species' is the location of the forest cover loss as a proxy for a deforestation event, as used by the Peruvian government. The location of a deforestation event in any given year is a result of the various interactions between the predictor variables at the given location. Following Aguilar-Amuchastegui et al. (2014), we assume that future deforestation will occur under the same set of conditions, although understanding that several factors can influence the future deforestation rate, including changes in government priorities and climate change. Although true absence data are not included in the Maxent algorithm, we also follow Aguilar-Amuchastegui et al. (2014) in that the true absence of deforestation may be misidentified due to factors that include near-persistent cloud cover in some areas and low-quality pixels in a given year.

Deforestation risk

We used Maxent default settings for regularization and selecting feature classes for all model runs, which include linear, quadratic, threshold and hinge features (Phillips & Dudík 2008), as well as the jack-knife option, which estimates the importance of individual predictor variables. The default settings allow Maxent to select an amount of regularization that is appropriate for the number of point localities and types of features used (Phillips & Dudík 2008). The total amount of deforestation from 2001 to 2019 was more than 21 728 km^2 or more than 23 million point localities in a 30-m resolution grid of the Peruvian Amazon. Therefore, deforestation data were resampled to the 100m pixel spatial resolution, and 10 000 point localities of deforestation were chosen at random from the full dataset. Environmental factors and interactions between predictor variables define the set of features that constrain the geographical distribution of the 'species' being modelled (Phillips et al. 2006). Among several options, Maxent models produce a logistic output format that can be used as a raster file in GIS, with pixel values from 0 to 1 that represent an estimate of relative probability of presence (Phillips & Dudík 2008). Pixels with a logistic value close to 1 are the sites most suitable for a 'species' (Phillips & Dudík 2008), or, in this study, those that most closely resemble the conditions for deforestation.

Test localities were generated by making 10 random partitions of the deforestation point localities for the execution of 10 model runs. Each partition was a random selection of 10 000 georeferenced points of the total deforestation from 2001 to 2019. For each of these model runs, 70% of the points were randomly selected for model training and 30% were selected for model testing for analysis of the spatial accuracy of each model prediction (Phillips et al. 2006, Zutta & Rundel 2017). The relative contribution (%) and permutation importance (%) of each predictive variable to the final predictive model were recorded (Table 1). These two measures indicate the importance of the predictive variable by indicating the amount of information each variable contributes to the final model (contribution %) and which variables contain information not present in the other variables for the model development (permutation importance %) (Bonilla-Bedoya et al. 2018).

Table 1. Contribution (%) and permutation importance (%) of the variables to the model.

Variable	Contribution (%)	Permutation importance (%)
Distance from agricultural land	59.7	40.5
Distance from roads	17.9	14.3
Elevation	6.0	13.0
Natural protected areas	4.9	7.2
Distance from pastureland	4.1	4.2
Distance from non-forest land	4.0	11.8
Distance from mining	1.7	5.2
Slope	1.4	3.0
Native communities	0.2	0.6
Distance from rivers	0.2	0.2

In order to facilitate the interpretation of the final logistic output format, a qualitative classification of deforestation risk is presented. The natural breaks classification method (Wang & Yang 2000, Smith 2015, Bonilla-Bedoya et al. 2018) of the final output values was used to divide the output into five categories: very low (<9.9%), low (10.0–20.9%), medium (21.0–33.8%), high (33.9–48.2%) and very high (48.3–100%).

Model performance

Our study assessed model performance using the threshold-independent test of area under the receiver operating characteristic (ROC) curve (AUC) of the test localities. The ROC curve measures the model's ability to correctly predict the presence and absence of the deforestation event by plotting the model sensitivity (omission rate) against 1 – specificity (commission rate). Pseudo-absences are generated by randomly selecting a set number of background pixels as a substitute for 'true' absence, which the algorithm does not use as an input variable (Phillips et al. 2006). For our study, the default of 10 000 background pixels was randomly chosen as the pseudo-absence for each model run. Consequently, the AUC statistic can be understood as the probability that a presence site of deforestation (i.e., pixel) is ranked above a random background pixel (Phillips et al. 2006). An AUC value for a model will range from 0 to 1.0, with values below 0.5 considered random, up to 1.0 as perfect discrimination. AUC data are presented as means ± SE (n = 10 models) following Phillips et al. (2006) and Zutta and Rundel (2017).

Prior to the general model of deforestation risk with 19 years of information, the spatial and temporal transferability of the models were tested. Following the recommendations of Aguilar-Amuchastegui et al. (2014), deforestation data from 2001 to 2010 were partitioned as training information and then validated with deforestation from 2011 to 2019. Another pertinent aspect was to test spatial transferability; this tested whether a risk model of deforestation from one specific area could predict deforestation in another distant area. For this purpose, the training samples were divided into north and south regions of the Peruvian Amazon.

Model performance was also evaluated by overlaying the January to December 2020 early alerts for forest loss from GeoBosques (Vargas et al. 2019) on the final risk model in order to ascertain the percentage of early alert pixels that were predicted in each deforestation risk category. The 2020 early alert data were not used in the development of the overall predictive model since it is developed through a different methodology.

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

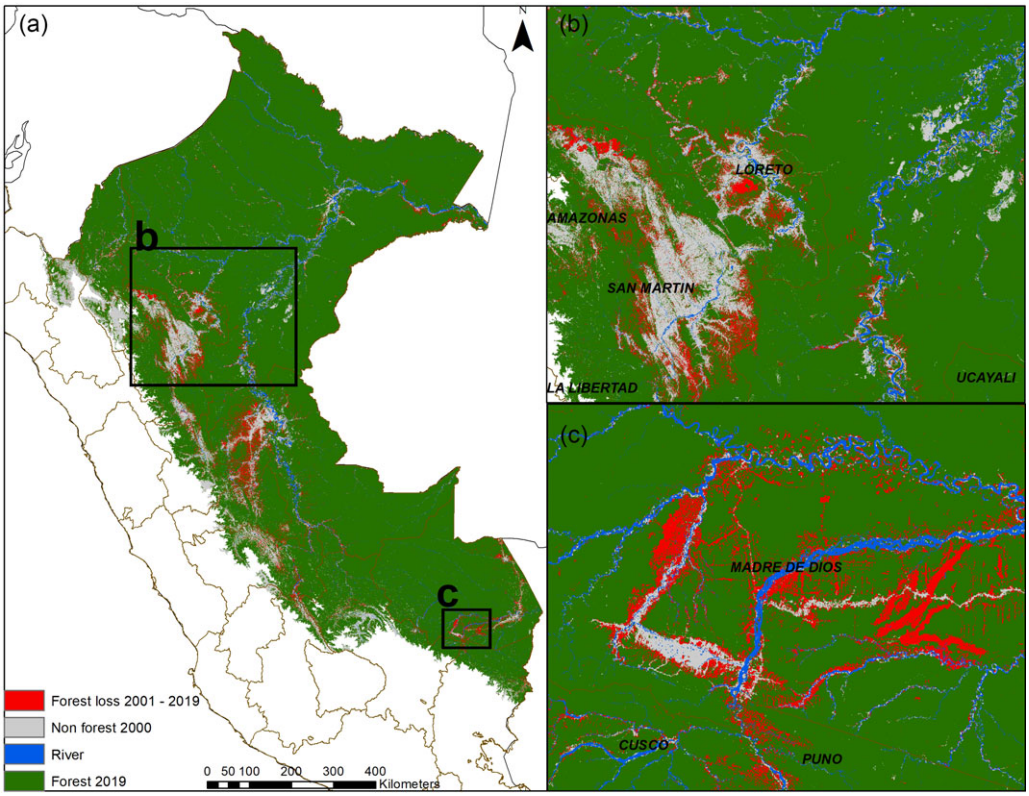


Fig. 1. Peruvian Amazon deforestation. (a) Accumulated deforestation from 2001 to 2019 and remaining forest cover. (b) Deforestation in the department of Loreto San Martin where most large-scale forest loss is due to recent agricultural expansion of oil palm. (c) Deforestation in southern Madre de Dios from a mix of artisanal gold mining and small-scale agriculture.

Results

Historical deforestation

The Peruvian Amazon lost *c.* 24 334 km^2 of forest from 2001 to 2019; this was *c.* 3.4% of the forest cover available in the year 2000, when there was more than 707 000 km^2 of forest. The annual historical average during 2015–2019 was 1560 km^2, compared to the years 2001–2014, which was 1181 km^2. Recent forest loss has been centred on the San Martin and Loreto border (Fig. 1b) and the Madre de Dios regions (Fig. 1c). This can be compared with another hotspot of deforestation between the limits of Huanuco, Pasco and Ucayali, which collectively lost 8721 km^2 of forest from 2001 to 2019. A total of 682 742 km^2 of forest remained in the Peruvian Amazon at the end of 2019.

Deforestation risk

The average test AUC value for predictive models was 0.749 ± 0.003 and the average training AUC value was 0.747 ± 0.001, indicating reasonable success in discriminating forest loss from non-forest loss areas. Prediction values ranged from 0 to 0.722 and were divided into the five categories established by the natural breaks (Fig. 2). Buffer zones were more vulnerable to deforestation than protected areas (Tables 2 & 3).

Generally, deforestation risk was highest near previously deforested areas (Fig. 2b–d). Forested areas, which are often fragmented forests surrounded by other land uses, within the higher-elevational zones of the eastern Andes contain the most area with high and very high deforestation risks (Fig. 2). The area next to the Interoceanic Highway and the area within the artisanal-scale gold mining region of Madre de Dios demonstrated a particularly consolidated area of elevated risk (Fig. 2c). Most very-low-risk areas were in Loreto, which has the largest expanse of forest cover (Fig. 2), the Southern Ucayali and western Madre de Dios region, which contains the Manu and Alto Purus national parks, and smaller pockets in the Loreto, San Martin and Ucayali intersection, which contains the Cordillera Azul National Park (Fig. 2b). In addition, an extensive very-low-risk area was in the southern Madre de Dios and northern Puno region, which contains the Tambopata National Reserve and Bahuaja–Sonene National Park (Fig. 2c).

Protected areas and buffer zones

The total forest cover in natural protected areas at the end of 2019 was 156 234 km^2, with a total loss from 2001 to 2019 of 705 km^2 (Table 2). Protected areas designated as national parks contained 61.2% of the total Amazonian forest cover within protected areas and registered 42.97% of the forest loss during this time. Protected areas within each category with the highest total forest loss from 2001 to 2019 (Table 2) included Alto Mayo (Protected Forest), Cordillera Azul (National Park), El Sira (Communal Reserve), Pacaya–Samiria (National Reserve) and Megantoni (National Sanctuary). The percentage of protected forests remaining in 2019 within the category of very low risk was 66%, low risk was 25%, medium risk was 7%, high risk was 1% and very high was 1%. A few protected areas had much larger areas with high risk of deforestation, notably San Matias–San Carlos (13%),

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

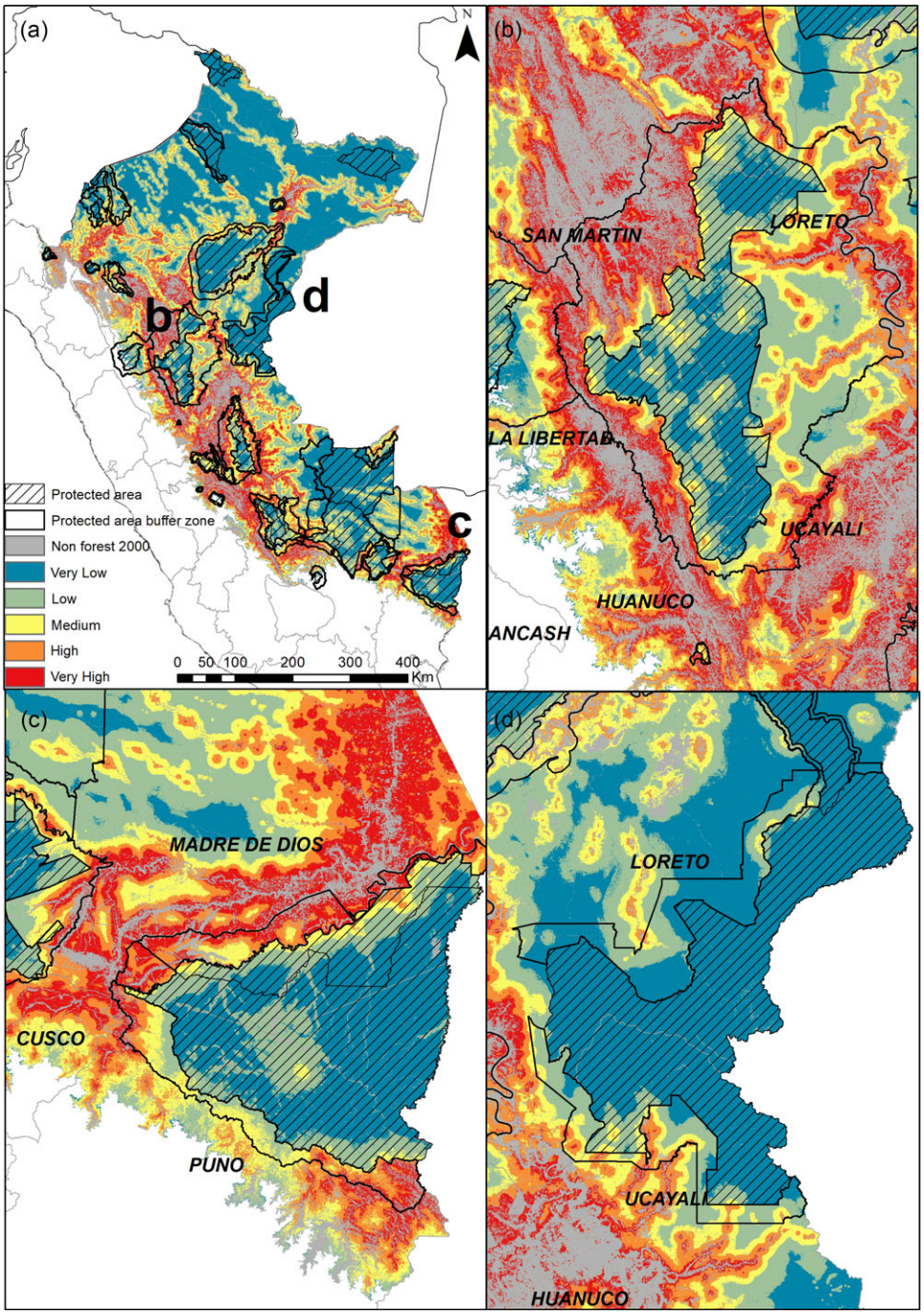


Fig. 2. Peruvian Amazon and five categories of deforestation risk. (a) Categories are based on the natural breaks classification of the final Maxent model for the entire Peruvian Amazon. Natural protected areas containing any amount of Amazon forest are filled with hatch marks and surrounded by their respective buffer zones. (b) Deforestation risk for the Cordillera Azul National Park. (c) The main artisanal gold mining region and interoceanic highway of southern Madre de Dios. (d) The Sierra del Divisor National Park.

Yanachaga–Chemillén (7%), Tingo Maria (50%), Yanesha (41%) and Allpahuayo Mishana (24%), which all have less than 1600 km^2 of forest remaining.

This pattern is in contrast to the 31 buffer zones that contain Amazonian forests. The total forest cover at the end of 2019 was 101 137 km^2, with a forest loss of 5318 km^2, which is 7.6 times greater than within the protected areas (Table 3). The greatest total forest losses in 2001–2019 were in the buffer zones of the Cordillera Azul (2277 km^2), El Sira (923 km^2) and Tambopata (307 km^2). Buffer zone forest remaining in 2019 had a higher predicted risk of deforestation within the categories of high risk (20 397 km^2 or 20%) and very high risk (15 297 km^2 or 15%).

Contribution of predictive variables

Distance from agricultural land and distance from roads contributed nearly 77.6% of the information necessary to predict deforestation (Table 1). Distance from pastureland and elevation contributed 10.1%, and the categorical variables of natural protected areas and native communities contributed 4.9% and 0.2%,

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

Table 2. Forest cover in 2019, total forest loss from 2001 to 2019 and deforestation risk in natural protected areas in the Peruvian Amazon.

Category	Protected area	Forest cover (km^2)	Forest loss (km^2)	Deforestation risk (% of forest cover)				
				vl	lo	md	hi	vh
Protected Forest	Alto Mayo	1592	75	8	35	44	13	0
	San Matias–San Carlos	1361	51	2	14	57	27	0
	Pui Pui	182	1	27	64	9	0	0
National Park	Alto Purús	24 813	13	83	14	3	0	0
	Manu	16 148	65	46	45	9	0	0
	Sierra del Divisor	13 451	51	93	6	1	0	0
	Cordillera Azul	13 269	75	47	45	8	0	0
	Bahuaja–Sonene	10 603	47	72	26	2	0	0
	Yaguas	8657	3	100	0	0	0	0
	Otishi	2874	12	51	46	3	0	0
	Güeppi–Sekime	2023	0	98	2	0	0	0
	Rio Abiseo	1838	25	39	56	5	0	0
	Yanachaga–Chemillén	1056	11	1	32	61	6	0
	Ichigkat Muja–Cordillera del Cóndor	871	1	96	4	0	0	0
	Tingo Maria	41	0	0	0	50	50	0
Communal Reserve	El Sira	5931	90	30	46	17	6	1
	Amarakaeri	3904	10	35	52	13	0	0
	Airo Pai	2452	1	100	0	0	0	0
	Machiguenga	2093	13	10	65	24	1	0
	Purús	1994	1	79	18	3	0	0
	Ashaninka	1734	16	14	59	22	5	0
	Huimeki	1403	2	88	11	1	0	0
	Tuntanain	938	1	84	15	1	0	0
	Yanesha	320	12	0	7	52	41	0
	Chayu Nain	221	0	43	49	8	0	0
National Reserve	Pacaya–Samiria	19 973	94	70	21	7	2	0
	Pucacuro	6307	4	100	0	0	0	0
	Matses	4189	3	98	2	0	0	0
	Tambopata	2683	21	36	40	19	5	0
	Allpahuayo Mishana	527	5	11	25	40	24	0
National Sanctuary	Megantoni	1996	11	12	62	24	2	0
	Cordillera de Colán	359	4	52	47	1	0	0
	Tabaconas–Namballe	233	3	18	70	12	0	0
	Pampa Hermosa	97	1	2	61	34	3	0
	Machu Picchu	103	2	43	55	2	0	0
Total		156 234	705	66	25	7	1	1

vl = very low; lo = low; md = medium; hi = high; vh = very high.

respectively, to the final model. The remaining land covers, often directly associated with forest loss, distance from non-forest and distance from mining, contributed 4.0% and 1.7%, respectively, while slope and distance from rivers together contributed 1.6% to the final model.

The ranking of permutation importance followed a similar pattern to the contribution of each predictor variable to the final model (Table 1). Distance from agricultural land had the highest percentage contribution (40.5%), followed by distance from non-forest land, distance from roads and elevation, which combined had a permutation importance of 39.1%. The remaining continuous predictive variables – distance from mining, distance from pastureland, distance from rivers and slope – accounted for 12.4%. Natural protected areas and native communities had permutation importance values of 7.2% and 0.2%, respectively.

Temporal and spatial transferability

Regarding temporal transferability, the deforestation risk model using data from the years 2001–2010 registered 86.4% of the remaining forest in the high-risk and very-high-risk categories for 2011, followed by 82.7% and 80.3% in the same categories for 2012 and 2013, respectively. For 2018 and 2019, the model registered 72.5% and 63.1% of the remaining forest for the same categories (Fig. 3b). Logically, the models predicted a greater percentage of deforestation for the years closest to the last observation.

The mean percentage (± SD) of monthly early alerts from January to December 2020 that were registered under the very high deforestation risk category was 44.7% (± 7.5%). More than 35% of monthly early alerts, except for in February, occurred in very-high-risk areas (Fig. 3a). Indeed, the mean percentage of monthly alerts that were registered as either high risk or very high risk was nearly 80%. The other risk categories were also fairly consistent throughout the year, with 15.4% ± 2.8% for medium risk, 8.7% ± 2.9% for low risk and 2.7% ± 1.1% for very low risk. The mean fractions of early alerts that registered as very high risk were also similar throughout the lows and peak of the deforestation season, with 41.1% ± 7.3% from January to June, when the number of alert counts were at their lowest, 44.3% ± 8.2% from July to September, the peak forest loss season, and 47.5% ± 8.0% from October to December, when the number of alerts returned to pre-peak levels.

There was far less predictive ability when deforestation localities were divided into north and south regions. Northern deforestation could only predict less than 10% for the mean, high-risk and very-high-risk categories for the southern half of the Peruvian Amazon. Similarly, southern deforestation models could only predict less than 5% of these categories in the northern half of the Peruvian Amazon.

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

Table 3. Buffer zone forest cover in 2019, total forest loss from 2001 to 2019 and deforestation risk in natural protected areas in the Peruvian Amazon.

Category	Protected area buffer zone	Forest area (km^2)	Forest loss (km^2)	Deforestation risk (% of forest area)				
				vl	lo	md	hi	vh
Protected Forest	Alto Mayo	1601	152	5	23	21	26	25
	San Matias–San Carlos	932	150	0	2	7	24	67
	Pui Pui	298	3	5	20	42	25	8
National Park	Alto Purús	16 788	9	48	37	12	3	0
	Manu	5929	53	24	28	25	17	6
	Sierra del Divisor	6109	59	43	39	12	5	1
	Cordillera Azul	17 265	2277	0	27	25	23	25
	Bahuaja–Sonene	2301	204	0	20	30	26	24
	Otishi	317	3	7	43	43	7	0
	Rio Abiseo	4081	212	10	30	29	16	15
	Yanachaga–Chemillén	417	27	0	1	8	60	31
	Ichigkat Muja–Cordillera del Cóndor	1329	4	58	37	5	0	0
	Tingo Maria	27	5	0	0	4	29	67
Communal Reserve	El Sira	8035	923	1	9	23	31	36
	Amarakaeri	2217	156	0	18	30	30	22
	Machiguenga	4329	80	2	14	48	32	4
	Purús	2159	8	6	48	29	16	1
	Ashaninka	2837	184	0	6	24	38	32
	Tuntanain	2784	17	38	37	18	6	1
	Yanesha	276	75	0	0	0	6	94
	Chayu Nain	116	0	1	35	43	20	1
National Reserve	Pacaya–Samiria	9362	219	8	28	33	23	8
	Pucacuro	3356	4	84	13	2	1	0
	Matses	2205	22	45	32	15	7	1
	Tambopata	1381	307	0	0	17	33	50
	Allpahuayo Mishana	461	31	15	14	23	21	27
National Sanctuary	Megantoni	2351	101	0	4	26	37	33
	Cordillera de Colán	163	3	22	51	24	3	0
	Machu Picchu	336	12	15	16	38	25	6
	Pampa Hermosa	43	1	2	14	36	36	12
	Tabaconas–Namballe	332	17	5	29	27	34	5
Total		101 137	5318	18	25	22	20	15

vl = very low; lo = low; md = medium; hi = high; vh = very high.

Discussion

Here, we leveraged 19 years of annual, high-spatial-resolution deforestation data, with high-spatial-resolution predictor variables, representing drivers of deforestation, within an open-source, machine learning SDM framework to produce a robust deforestation risk model for the Peruvian Amazon. In addition, we were able to test the temporal transferability of the methodology, similar to testing and training forest loss data from different periods in Aguilar-Amuchastegui et al. (2014), and we tested spatial transferability by using test and training datasets from different Peruvian Amazon regions, which has not been explored in other studies. We were also able to test the risk model accuracy using data from an independent, high-temporal-cadence monitoring system (i.e., early alerts).

The approach provided a clear, geospatially explicit correlation between deforestation and distance from agricultural land and roads. Other studies that have applied Maxent to the Ecuadorian (Bonilla-Bedoya et al. 2018) and the Brazilian Amazon (de Souza & De Marco 2018), at different spatial and temporal resolutions from our study, could identify similar sets of variables that describe most of the deforestation modelled in their region. Maxent, as a predictive modelling tool option, provides several advantages over other predictive modelling tools, including the availability of graphical user interface (GUI) and open-source versions (Phillips 2017), its relative ease of implementation (Aguilar-Amuchastegui et al. 2014), its lack of requirement for presence and absence data where true absence can be difficult to confirm (Philips 2006), its provision of accurate predictions with relatively few presence data and high computing efficiency and its ability to enable the use of large-scale high-resolution data layers (Buermann et al. 2008).

Distance from agricultural land and distance from roads were the key predictors of deforestation in the Peruvian Amazon. Agricultural expansion in the past 19 years has been driven by a complex interaction of periodic immigration from Andean and inter-Amazonian regions, changes in commodity and gold prices, the expansion of industrial agriculture (most often from oil palm production) and a variety of other factors (Menton & Cronkleton 2019). Artisanal-scale gold mining in Madre de Dios has a particularly devastating impact at the local scale and has increased dramatically since 2011, but it has not had the overall impact of small- to medium-scale agriculture. Conversion of forest to pastureland is a major factor in deforestation in the eastern portion of Madre de Dios in the area that is closest to the Brazilian state of Acre, and it shares similar conditions for cattle ranching; this driver is particular to this region and is not as prevalent in other Amazonian regions of Peru. Rivers are an important form of transportation and are thought to be as important a driver of deforestation as roads, as they represent important transportation hubs in many of these regions where roads and connectivity to towns are limited (Bax et al. 2016). However, the model's outputs reflect that previous deforestation contributed most to predicting future deforestation as these deforested areas will be nearer to the next expansion of forest removal.

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

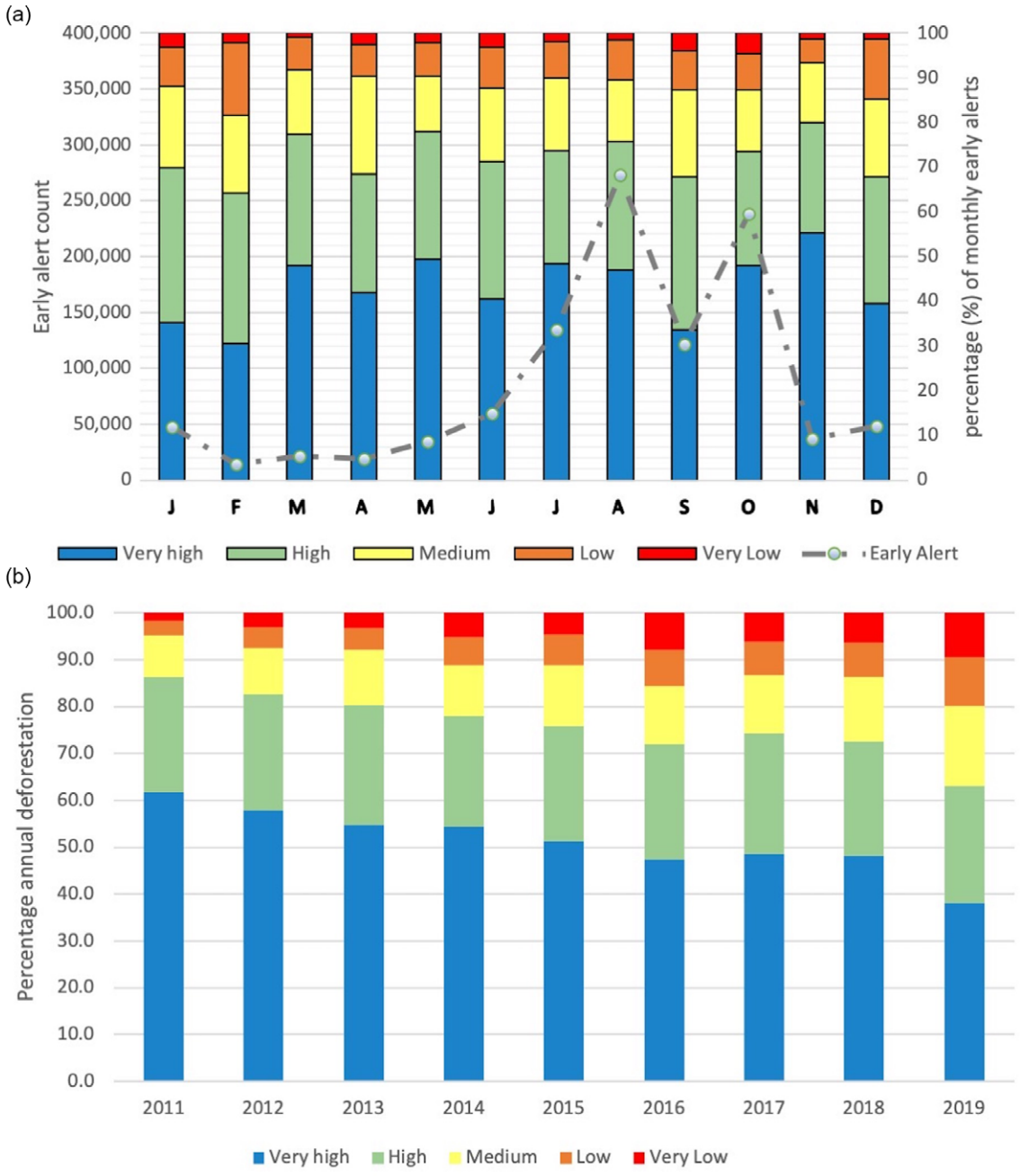


Fig. 3. Monthly early alert count in 2020. (a) The percentage of monthly early alerts that was registered under each deforestation risk category (very low, low, medium, high and very high). (b) Annual deforestation from 2011 to 2019 (%) recorded under each category of deforestation risk (very low, low, medium, high and very high).

The distance from native communities, although important for conservation priorities, was not a significant contributor to the deforestation risk model. This is probably due to the encroachment on forest that is currently occurring within native communities, most often due to illegal selective logging and other deforestation activities. According to national statistics, native communities contained the second highest amount of forest loss in 2019 below untitled lands, accounting for almost 300 km^2 or 19% of the total deforestation in that year (MINAM 2020). Therefore, the native community boundaries do not represent as concrete a deterrent to deforestation as national park boundaries. Peruvian naturally protected areas were among the least deforested of all the land tenure categories monitored by the Peruvian government from 2001 to 2019, with over 30 km^2 or 2.1% of the total Amazonian deforestation in 2019 (MINAM 2020). The deforestation patterns currently seen are in response to the location and ability of national and local authorities to limit deforestation in many of these protected areas.

In order to fully apply Maxent as an approach for predicting risk, we need to address key factors that may limit the overall accuracy of the models. First, the accuracy of the predictive model will be dependent on the quality of the point locality data, or in this case the forest loss locations, which are derived from satellite data. Frequent cloud cover, rough topography and availability of Landsat images may affect the number of pixels available for the final annual deforestation estimation (Potapov et al. 2014).

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

It is therefore possible that deforestation that is not identified in a given year is counted in a subsequent year, when quality pixels become available. In addition, the quality of the predictor variables, including the accuracy of road maps and the correct identification of land cover, may play a role in increasing the accuracy of the predictive model.

Conclusion

Using open-source machine learning algorithms to identify areas of moderate to high risk for deforestation can ultimately help to reduce deforestation in tropical forests. The approach is simple and easy to use and interpret for a wide range of end users, from technical experts who wish for a relatively quick risk assessment of large areas to less technical users wanting to explore deforestation risk using a reliable and robust approach. To the best of our knowledge, this is the first study of deforestation risk using 19 years of historical annual data for the entire Peruvian Amazon. Indeed, the large dataset and high spatial resolution require higher-end computational power, which points to developing a more interactive platform to reach an even wider audience interested in exploring deforestation risk due to the alarming recent increase in forest loss in rainforests across the tropics. The potential of this approach is highlighted by a static version of this risk model hosted by the National Forest Conservation Program for Climate Change Mitigation (PNCBMCC-MINAM) of the Peruvian Ministry of the Environment (http://geobosques.minam.gob.pe/geobosque/visor).

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0376892921000291.

Acknowledgements. We would like to thank Craig Wayson from SilvaCarbon, Monica Romo from USAID and Daniel Castillo from PNCBMCC-MINAM.

Financial support. The project was supported by USAID and the US Government inter-agency program SilvaCarbon.

Conflict of interest. None.

Ethical standards. None.

References

Achard F, Defries R, Eva H, Hansen M, Mayaux P, Stibig HJ (2007) Pan-tropical monitoring of deforestation. *Environmental Research Letters* 2: 045022.

Aguilar-Amuchastegui N, Riveros JC, Forrest JL (2014) Identifying areas of deforestation risk for REDD+ using a species modeling tool. *Carbon Balance and Management* 9: 1–10.

Asner GP, Tupayachi R (2017) Accelerated losses of protected forests from gold mining in the Peruvian Amazon. *Environmental Research Letters* 12: 094004.

Austin KG (2017) Trends in size of tropical deforestation events signal increasing dominance of industrial-scale drivers. *Environmental Research Letters* 12: 054009.

Bax V, Francesconi W (2018) Environmental predictors of forest change: an analysis of natural predisposition to deforestation in the tropical Andes region, Peru. *Applied Geography* 91: 99–110.

Bax V, Francesconi W, Quintero M (2016) Spatial modeling of deforestation processes in the Central Peruvian Amazon. *Journal for Nature Conservation* 29: 79–88.

Bonilla-Bedoya S, Estrella-Bastidas A, Molina JR, Herrera MÁ (2018) Socioecological system and potential deforestation in Western Amazon forest landscapes. *Science of the Total Environment*, 644: 1044–1055.

Buermann W, Saatchi S, Smith TB, Zutta BR, Chaves JA, Milá B, Graham CH (2008) Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography* 35: 1160–1176.

Chávez Michaelsen A, Huamani Briceño L, Fernandez Menis R, Bejar Chura N, Valera Tito F, Perz S et al. (2013) Regional deforestation trends within local realities: land-cover change in southeastern Peru 1996–2011. *Land* 2: 131–157.

Chen W (2017) Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques. *Geoderma* 305: 314–327.

Corbera E, Schroeder H (2011) Governing and implementing REDD+. *Environmental Science and Policy* 14: 89–99.

de Souza RA, De Marco P (2014) The use of species distribution models to predict the spatial distribution of deforestation in the western Brazilian Amazon. *Ecological Modelling* 291: 250–259.

de Souza RA, De Marco P (2018) Improved spatial model for Amazonian deforestation: an empirical assessment and spatial bias analysis. *Ecological Modelling* 387: 1–9.

Espejo JC, Messinger M, Román-Dañobeytia F, Ascorra C, Fernandez LE, Silman M (2018) Deforestation and forest degradation due to gold mining in the Peruvian Amazon: a 34-year perspective. *Remote Sensing* 10: 1–17.

Fonseca MG, Aragão LEOC, Lima A, Shimabukuro YE, Arai E, Anderson LO (2016) Modelling fire probability in the Brazilian Amazon using the maximum entropy method. *International Journal of Wildland Fire* 25: 955–969.

Franklin J, Regan HM, Syphard AD (2014) Linking spatially explicit species distribution and population models to plan for the persistence of plant species under global change. *Environmental Conservation* 41: 97–109.

Gibbs HK, Ruesch AS, Achard F, Clayton MK, Holmgren P, Ramankutty N, Foley JA (2010) Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America* 107: 16732–16737.

Hansen MC, Krylov A, Tyukavina A, Potapov PV, Turubanova S, Zutta B et al. (2016) Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters* 11: 34008.

Hosonuma N, Herold M, De Sy V, De Fries RS, Brockhaus M, Verchot L et al. (2012) An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters* 7: 044009

Kalamandeen M, Gloor E, Mitchard E, Quincey D, Ziv G, Spracklen D et al. (2018) Pervasive rise of small-scale deforestation in Amazonia. *Scientific Reports* 8: 1–10.

Menton M, Cronkleton P (2019) *Migration and Forests in the Peruvian Amazon: A Review*. Bogor, Indonesia: Center for International Forestry Research.

MINAM (2015a) *Memoria Descriptiva del Mapa de Bosque/No Bosque año 2000 y Mapa de pérdida de los Bosques Húmedos Amazonicos del Perú 2000-2011*. Lima, Peru: MINAM.

MINAM (2015b) *Protocolo de Clasificación de pérdida en los bosques Húmedos Amazonicos*. Lima, Peru: MINAM.

MINAM (2020) *Monitoreo de la pérdida de bosques húmedos amazónicos en el año 2019*. Lima, Peru: MINAM.

Musinsky J, Tabor K, Cano CA, Ledezma JC, Mendoza E, Rasolohery A, Sajudin ER (2018) Conservation impacts of a near real-time forest monitoring and alert system for the tropics. *Remote Sensing in Ecology and Conservation* 4: 189–196.

Petersen R (2018) *Tropical Forest Monitoring: Exploring the Gaps between What Is Required and What Is Possible for REDD+ and Other Initiatives*. Bogor, Indonesia: Center for International Forestry Research.

Phillips SJ (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Phillips SJ (2017) Opening the black box: an open-source release of Maxent. *Ecography* 40: 887–893.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.

Potapov PV, Dempewolf J, Talero Y, Hansen MC, Stehman SV, Vargas C et al. (2014) National satellite-based humid tropical forest change

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press

assessment in Peru in support of REDD+ implementation. *Environmental Research Letters* 9: 124012.

Potapov PV, Turubanova SA, Hansen MC, Adusei B, Broich M, Altstatt A et al. (2012) Remote sensing of environment quantifying forest cover loss in Democratic Republic of the Congo, 2000–2010, with Landsat ETM + data. *Remote Sensing of Environment* 122: 106–116.

Recanati F, Allievi F, Scaccabarozzi G, Espinosa T, Dotelli G, Saini M (2015) Global meat consumption trends and local deforestation in Madre de Dios: assessing land use changes and other environmental impacts. *Procedia Engineering* 118: 630–638.

Romijn E (2012) Assessing capacities of non-Annex I countries for national forest monitoring in the context of REDD+. *Environmental Science and Policy* 19–20: 33–48.

Sanchez-Cuervo AM, Aide TM (2013) Identifying hotspots of deforestation and reforestation in Colombia (2001–2010): implications for protected areas. *Ecosphere* 4: 1–21.

Smith M (2015) Geospatial analysis. *Voprosy Kurortologii, Fizioterapii, i Lechebno? Fizichesko? Kultury* 1: 4–10.

Vargas C, Montalban J, Leon AA (2019) Early warning tropical forest loss alerts in Peru using Landsat. *Environmental Research Communications* 1: 121002.

Vijay V, Reid CD, Finer M, Jenkins CN, Pimm SL (2018) Deforestation risks posed by oil palm expansion in the Peruvian Amazon. *Environmental Research Letters* 13: 114010.

Wang D, Yang L (2000) Visual hiding scheme using one secret image. *Jisuanji Xuebao/Chinese Journal of Computers* 23: 943–948.

West TAP, Börner J, Fearnside PM (2019) Climatic benefits from the 2006–2017 avoided deforestation in Amazonian Brazil. *Frontiers in Forests and Global Change* 2: 52.

Zutta BR, Rundel PW (2017) Modeled shifts in *Polylepis* species ranges in the Andes from the last glacial maximum to the present. *Forests* 8: 1–16.

https://doi.org/10.1017/S0376892921000291 Published online by Cambridge University Press