SUPPLEMENTARY MATERIALS

Forecasting tropical moist forest cover change in the 21st century under a "business-as-usual" scenario

Ghislain VIEILLEDENT[1,2,3,4,⋆] Christelle VANCUTSEM[1]

and Frédéric ACHARD[1]

[1] **European Commission**, JRC, Bio-economy Unit, I-21027 Ispra (VA), ITALY
[2] **CIRAD**, UPR Forêts et Sociétés, F-34398 Montpellier, FRANCE
[3] **CIRAD**, UMR AMAP, F-34398 Montpellier, FRANCE
[4] AMAP, **Univ Montpellier**, CIRAD, CNRS, INRAE, IRD, Montpellier, FRANCE

[⋆] **Corresponding author:** \E-mail: ghislain.vieilledent@cirad.fr \Phone: +33 4 67 61 49 09

1 Materials and Methods

1.1 Historical forest cover change maps

We derived historical forest cover change maps on two periods of time: 1st January 2000 – 1st January 2010, and 1st January 2010 – 1st January 2020 from the forest cover change annual product by Vancutsem *et al.* (2020). Vancutsem *et al.* (2020) annual product classifies Landsat image pixels at 30 m resolution in 16 categories for each year (on the 31st of December) between 1982 and 2019 and allows identifying moist tropical forest pixels at each date (Table S1). This classification is based on an expert model analyzing time-series data at the pixel level extracted from the full Landsat satellite image archive on the period 1982–2019. The expert model has been built on Google Earth Engine (Gorelick *et al.*, 2017). For our forest definition, we only considered natural old-growth moist tropical forest, disregarding plantations and regrowths. We included degraded forests (not yet deforested) in our forest definition. As a consequence, we considered pixels in the following categories: 1, 2, 3, 4, 5, 13, or 14 in the annual product, to be natural old-growth moist tropical forest pixels (simply abbreviated "forest" in this manuscript). Because several decades are usually necessary to reach the state of old-growth forest, we assumed every pixels classified as "forest" at a given date between 1999 and 2019 to be also classified as "forest" in the previous years of that period of time. We thus obtained three forest cover maps for the dates 1st January 2000, 1st January 2010, and 1st January 2020. We combined these three maps to obtained high-resolution moist tropical forest cover change maps on the periods 2000–2010–2020 at 30 m resolution at the global scale.

We did not consider potential forest regrowth in our forest definition for three main reasons. First, throughout the humid tropics, forest regeneration involves much smaller areas than deforestation (Vancutsem *et al.*, 2020, see also 2000-2012 tree cover gain in Hansen *et al.* (2013)). Second, there is little evidence of natural forest regeneration in the long term in the tropics (Grouzis *et al.*, 2001). This can be explained by several ecological processes following deforestation such as soil erosion (Grinand *et al.*, 2017) and reduced seed bank due to fire-induced deforestation and soil loss (Grouzis *et al.*, 2001). Moreover, in areas where forest regeneration is ecologically possible, young forest regrowth are more easily re-burnt for agriculture and pasture (Vieilledent *et al.*, 2020). Third, young secondary forests generally provide more limited ecosystem services compared to old-growth natural forests in terms of biodiversity (Gibson *et al.*, 2011) and carbon storage (Blanc *et al.*, 2009).

1.2 Environmental explanatory variables

To explain the observed spatial deforestation on the period 2010–2020, we considered a set of spatial explanatory variables describing: topography (altitude and slope), accessibility (distances to nearest road, town, and river), forest landscape (distance to forest edge), deforestation history (distance to past deforestation), and land conservation status (belonging to a protected area). Characteristics of each explanatory variable are summarized in Table S2.

Elevation (in m) and slope (in degree) at 90 m resolution were obtained from the SRTM Digital Elevation Database v4.1 (http://srtm.csi.cgiar.org/). Distances (in m) to nearest road, town and river at 150 m resolution were obtained from the OpenStreetMap (OSM) project (https://www.openstreetmap.org/). OSM country data were downloaded from two websites: Geofabric (http://download.geofabrik.de/) and OpenStreetMap.fr (https://download.openstreetmap.fr/extracts/). To obtain the road network in each country, we considered the "motorway", "trunk", "primary", "secondary" and "tertiary" categories for the "highway" key in OSM. To obtain the network of populated places in each country (that we simply call "towns" in the present study), we considered the "city", "town" and "village" categories for the "place" key in OSM. To obtain the river network in Madagascar, we considered the "river" and "canal" categories for the "waterway" key in OSM. For a more detailed description of each category, see the OSM wiki page (https://wiki.openstreetmap.org/wiki/Tags). OSM data have been downloaded during the period January–March 2020 for all countries. Distance to forest edge was computed at 30 m resolution from the forest cover map in 2010. Distance to past deforestation in 2010 was computed at 30 m resolution from the 2000–2010 forest cover change map. To minimize border effect for the computation of distance to forest edge and distance to past deforestation, a buffer of 10 km around each country extent was considered. Country borders were obtained from version 3.6 of the GADM data (https://gadm.org). Data on protected areas were obtained from the World Database on Protected Areas (https://www.protectedplanet.net, UNEP-WCMC & IUCN (2020)) using the `pywdpa` Python package (https://pypi.org/project/pywdpa/). WDPA data have been downloaded during the period January–March 2020 for all countries. All protected areas defined by at least one polygon were considered in the analysis, but protected areas defined by a point were not taken into account. Data included protected areas of all IUCN categories (from Ia to VI) and of all types defined at the national level (e.g. National Parks, Reserves), even if the type and IUCN category were not reported. Polygons representing protected areas were rasterized at 30 m resolution.

In total, we obtained 8 spatial explanatory variables to model the spatial probability of deforestation Table S2.

1.3 Point sampling

We performed a stratified sampling between (i) forest pixels in 2010 which have been deforested on the period 2010–2020 and (ii) forest pixels in 2010 which have not been deforested on that period and which represent the remaining forest in 2020. To maximize the representativity of the data, the total number of forest pixels sampled in each country for the year 2010 has been chosen proportional to the area of forest in 2010 in that country (2000 points for 1 Mha of forest), with the condition that this number had to be between 20,000 (minimal number of observations to model the deforestation process) and 100,000 (to limit the computation time). For countries with less than 20,000 forest pixels in 2010, all forest pixels were included in the data-set. For each sampled pixel, we retrieved information regarding the 8 computed explanatory variables at their original spatial resolution. When the information was not complete for a given pixel (eg. elevation and slope data missing for a forest pixel located close to the sea border), the observation was removed from the data-set. Missing information affects a minority of pixels for each country. The global data-set included a total of 3,186,698 observations (1,601,810 of non-deforested pixels and 1,584,888 of deforested pixels, corresponding to an area of 144,163 ha and 142,647 ha, respectively) spread into 119 territories and representing 92 countries (Table S3). Each country is defined by one unique three-letter country code following the ISO 3166-1 standard.

- Brazil divided in X states
- India in three independent regions

1.4 Models

- deforestation process supposed independent, and different, for each study area.

Using observations of the deforestation on the period 2010–2020, we modelled the spatial probability of deforestation as a function of the height explanatory variables. compared two deforestation models. The first model described in Eq. (1) is a simple logistic regression model, a special case of generalized linear model (GLM) for binary data. This model is denoted "glm" in subsequent sections and results. We considered the random variable y_i which takes value 1 if the forest pixel i is deforested on the period 2000–2010 and 0 if it is not. We assumed that y_i follows a Bernoulli distribution of parameter θ_i. In our model, θ_i represents the spatial probability of deforestation for pixel i. We assumed that θ_i is linked, through a logit function, to a linear combination of the explanatory variables $X_i\beta$, where X_i is the vector of explanatory variables for pixel i and β is the vector of model parameters to be estimated.

$$\begin{aligned} y_i &\sim \mathcal{B}ernoulli(\theta_i) \\ \text{logit}(\theta_i) &= X_i\beta \end{aligned} \tag{1}$$

The second model described in Eq. (2) includes additional random effects ρ_j for each spatial cell j of a 10 × 10 km grid covering Madagascar. This grid resolution was choosen in order to have a reasonable balance between a good representation of the spatial variability of the deforestation process and the number of parameters. We assumed that random effects were spatially autocorrelated through an intrinsic conditional autoregressive (iCAR) model (Besag *et al.*, 1991; Banerjee *et al.*, 2014). This model is denoted "icar" in subsequent sections and results. In a iCAR model, the random effect ρ_j associated to cell j depends on the values of the random effects $\rho_{j'}$ associated to neighbouring cells j'. In our case, the neighbouring cells are cells connected to the target cell j through a common border or corner (cells defined by the "king move" in chess). The number of neighbouring cells for cell j, which might vary, is denoted n_j.

$$\begin{aligned} y_i &\sim \mathcal{B}ernoulli(\theta_i) \\ \text{logit}(\theta_i) &= X_i\beta + \rho_j \\ \rho_j &\sim \mathcal{N}ormal(\sum_{j'} \rho_{j'}/n_j, V_\rho/n_j) \end{aligned} \tag{2}$$

The first model can be viewed as a "process-based" model for which variables are selected on an *a priori* knowledge of the deforestation process. For example, we assumed that the risk of deforestation decreases with the distance to road and forest edge, and is lower in protected areas. The second model can be viewed as a model combining a "process-based" part and a "pattern-oriented" part. Additional spatial random effects ρ_j account for unmeasured or unmeasurable variables (Clark, 2005) that explain a part of the residual spatial variation in the deforestation process (the residual spatial "pattern") that is not explained by the fixed environmental variables (X_i). While the first model has only 9 parameters to be estimated (one intercept parameter plus 8 slope parameters for the explanatory variables), the second model has 6,266 parameters to be estimated, including the 6,257 spatial random effects for the 10 $\times$ 10 km cells covering whole Madagascar (for which lands cover 587 000 km^2).

We used a random sample of 20,000 forest pixels in 2000 to fit the two models. The sample was stratified between 10,000 deforested pixels in 2000–2010 and 10,000 non-deforested pixels. A balanced sample between deforested and non-deforested pixels is preferable in our case (Dezécache *et al.*, 2017). First, deforestation events are rare (~1 %/yr) and a non-stratified sample would lead to very few observations of deforestation events, rendering difficult a good estimation of the slope parameters for the explanatory variables. Second, only the value of the linear model intercept is affected by this balanced sampling, which is not the case for the slope or random parameters. In our case, a biased estimate of the intercept is not an issue, as we are not interested in estimating the intensity of deforestation but the relative probability of deforestation between pixels. Function `sample()` from the `forestatrisk` Python package was used for fast stratified sampling and for extracting variable values at each point.

Parameter inference was done in a hierarchical Bayesian framework. Non-informative priors were used for all parameters: $\beta \sim \mathcal{N}ormal(\text{mean} = 0, \text{var} = 10^6)$ and $V_\rho \sim 1/\mathcal{G}amma(\text{shape} = 0.05, \text{rate} = 0.0005)$. We run a Markov Chain Monte Carlo (MCMC) of 7000 iterations. We discarded the first 2000 iterations (burn-in phase) and we thinned the chain each 5 iterations (to reduce autocorrelation between samples). We obtained 1000 estimates for each parameter. MCMC convergence was visually checked looking at MCMC traces and parameter posterior distributions. Function `model_binomial_iCAR()` from the `forestatrisk` Python package was used for parameter inference. This function calls an adaptive Metropolis-within-Gibbs algorithm (Rosenthal *et al.*, 2011) written in C for maximum computation speed.

1.5 Model comparison

We computed the deviance $\mathcal{D}$ of the two models with the formula $\mathcal{D} = -2\log\mathcal{L}$, $\mathcal{L}$ being the likelihood of the model, i.e. the probability of observing the data given the model and estimated parameters. We compared the deviance of the two models with the deviances of both the "null" model and the "full" model. The "null" model assumes a constant probability of deforestation for all theobservations and has only one parameter, the intercept of the linear relationship. At the other extreme, the "full" model has as many parameters as there are observations. We then computed the percentage of deviance explained by each model, considering that the "null" model explains 0% of the deviance and the "full" model explains 100% of the deviance.

We also performed a cross-validation to compare models using an independent validation data-set of 20,000 forest pixels in 2000. Again, the sample was stratified between 10,000 deforested pixels in 2000–2010 and 10,000 non-deforested pixels. We used the fitted models to predict the deforestation probability of all the pixels of the validation data-set. To transform the deforestation probabilities into binary values, we identified the probability threshold respecting the percentage of deforested pixels (eg. the mode of the probabilities for a deforestation rate of 50%). Using model predictions and observations, we computed several accuracy indices: the Area Under the ROC Curve (AUC), the Figure of Merit (FOM), the Overall Accuracy (OA), the Expected Accuracy (EA), the Kappa of Cohen (K), the Specificity (Spe), the Sensitivity(Sen), and the True Skill Statistics (TSS). A detailed description of these indices can be found in Pontius *et al.* (2008) (for the FOM) and Liu *et al.* (2011) (for all the other indices). Formulas used to compute these indices are presented in Appendix 1.

Because the value of these indices depends on the deforestation rate (Pontius *et al.*, 2008), we computed the accuracy indices for various percentage of deforested pixels: 1, 5, 10, 25 and 50%. To do so, we selected subsamples of the deforested pixels in our validation data-set at random.

1.6 Computing the spatial probability of deforestation in 2010

For the "icar" model, before computing the predictions of the deforestation probability, the spatial random effects at 10 km were interpolated at 1 km using a bicubic interpolation method. This was done in order to obtain spatial random effects at a resolution closer to the original forest raster resolution of 30 m, and to smooth the deforestation probability spatially.

Deforestation probabilities (float values in the interval $[0, 1]$) were transformed as integer values on the interval $[\![1, 65535]\!]$. This allowed us to record the large raster of probabilities as UInt16 type and save space on disk. We then obtained a map of the relative probability of deforestation for the year 2010 at 30 m resolution.

In 2010, Madagascar was covered by 9.3 Mha of natural forest corresponding to more than 104 M pixels at 30 m resolution. Predictions were computed using functions `predict_raster*()` from the `forestatrisk` Python package which make computation fast and efficient (with low memory usage) by treating raster data by blocks.

1.7 Forecasting forest cover change on the period 2010–2017

We computed the observed deforestation D (in ha) on the period 2010–2017 from the forest cover maps at these two dates. To forecast the forest cover change in 2010–2017 with our models, we used the previously derived maps of relative probability of deforestation in 2010. The resolution of these maps is $r = 30$ m, equivalent to $r_{\text{ha}} = 0.09$ ha. We computed a probability threshold θ_T in the interval $[\![1, 65535]\!]$ identifying the n forest pixels in 2010 with the highest probability of deforestation so that $nr_{\text{ha}} = D + \epsilon$. Because deforestation probabilities have finite values in $[\![1, 65535]\!]$, some forest pixels might have the same deforestation probability and it might not be possible to identify θ_T such that $\epsilon = 0$. We thus selected the threshold θ_T minimizing ϵ. We obtained negligible ϵ ($<$ 32,000 ha) compared to D (874,211 ha) for both models. We considered those n forest pixels in 2010 as deforested on the period 2010–2017 and derived the forest cover change map on that period.

2 Tables

2.1 Categories of the forest cover annual product

Table S1: **Categories of the forest cover annual product by Vancutsem *et al.* (2020)**. The forest cover annual product classifies Landsat image pixels in 16 categories for each year (on the 31st of December) between 1982 and 2019 and allows identifying moist tropical forest pixels at each date.

Class	Definition
1	Tropical moist forest (TMF including bamboo-dominated forest and mangroves)
2	TMF converted later in a tree plantation
3	NEW degradation
4	Ongoing degradation (disturbances still detected)
5	Degraded forest (former degradation, no disturbances detected anymore)
6	NEW deforestation (may follow degradation)
7	Ongoing deforestation (disturbances still detected)
8	NEW Regrowth
9	Regrowthing
10	Other land cover (not water)
11	Permanent Water (Pekel et al, 2015)
12	Seasonal Water (Pekel et al, 2015)
13	Init period without valid data - Init class = TMF
14	Init period with min 1 valid obs - Init class = TMF
15	Nodata - Init class = other LC
16	Init period without valid data - Init class = Plantation

2.2 Variables

Table S2: **Set of explicative variables used to model the spatial probability of deforestation**. A total of height variables were tested. They indicate topography, forest accessibility, forest landscape, deforestation history, and conservation status.

Product	Source	Variable derived	Unit	Resolution (m)	Date
Forest maps (2000-2010-2020)	Vancutsem et al. 2020	distance to forest edge	m	30	–
		distance to past deforestation	m	30	–
Digital Elevation Model	SRTM v4.1 CSI-CGIAR	altitude	m	90	–
		slope	degree	90	–
Highways	OSM-Geofabrik	distance to roads	m	150	Jan-Mar 2020
Places		distance to towns	m	150	Jan-Mar 2020
Waterways		distance to river	m	150	Jan-Mar 2020
Protected areas	WDPA	presence of protected area	–	30	Jan-Mar 2020

2.3 Sample size

Table S3: **Number of observations used for the spatial model of deforestation for each study area**. The table includes the number of non-deforested (nfor) and deforested (ndef) pixels per study area. These numbers include the forest pixels with full information regarding the explanatory variables. The corresponding number of hectares is also provided (nfHa and ndHa, respectively).

Continent	iso3	nfor	ndef	nfHa	ndHa
Africa	AGO	10,000	10,000	900	900
Africa	BDI	10,000	9,999	900	900
Africa	BEN	9,981	9,995	898	900
Africa	CAF	10,000	10,000	900	900
Africa	CIV	10,000	10,000	900	900
Africa	CMR	22,983	22,987	2,068	2,069
Africa	COD	50,000	50,000	4,500	4,500
Africa	COG	23,412	23,411	2,107	2,107
Africa	COM	9,990	9,984	899	899
Africa	ETH	10,000	10,000	900	900

Africa	GAB	23,986	23,966	2,159	2,157
Africa	GHA	9,999	10,000	900	900
Africa	GIN	9,978	9,998	898	900
Africa	GMB	9,988	9,999	899	900
Africa	GNB	9,882	9,977	889	898
Africa	GNQ	9,997	9,988	900	899
Africa	KEN	9,986	9,998	899	900
Africa	LBR	9,999	9,997	900	900
Africa	MDG	9,994	9,998	899	900
Africa	MUS	9,968	9,969	897	897
Africa	MWI	10,000	10,000	900	900
Africa	MYT	9,963	9,983	897	898
Africa	NGA	9,984	9,995	899	900
Africa	REU	9,996	9,995	900	900
Africa	RWA	10,000	10,000	900	900
Africa	SEN	9,892	9,976	890	898
Africa	SLE	9,997	9,999	900	900
Africa	SSD	5,737	7,613	516	685
Africa	TGO	10,000	9,999	900	900
Africa	TZA	9,983	9,970	898	897
Africa	UGA	10,000	10,000	900	900
Africa	ZMB	10,000	10,000	900	900
America	ATG	9,840	6,336	886	570
America	BHS	9,923	9,946	893	895
America	BLZ	9,995	9,999	900	900
America	BOL	30,476	30,476	2,743	2,743
America	BRB	9,958	8,251	896	743
America	COL	49,995	49,999	4,500	4,500
America	CRI	9,997	9,991	900	899
America	CUB	9,955	9,958	896	896
America	DMA	9,991	9,878	899	889
America	DOM	9,992	9,997	899	900
America	ECU	14,395	14,395	1,296	1,296
America	GLP	9,979	9,923	898	893
America	GRD	9,974	9,928	898	894
America	GTM	9,999	9,999	900	900
America	GUF	10,000	10,000	900	900
America	GUY	18,511	18,511	1,666	1,666
America	HND	9,997	9,999	900	900
America	HTI	9,952	9,970	896	897
America	JAM	9,988	9,996	899	900
America	KNA	9,988	4,614	899	415
America	LCA	9,994	9,967	899	897
America	MAF	3,156	3,808	284	343
America	MEX	9,994	9,997	899	900
America	MSR	9,982	1,258	898	113
America	MTQ	9,972	9,973	897	898
America	NIC	9,999	9,999	900	900
America	PAN	9,996	9,994	900	899
America	PER	50,000	50,000	4,500	4,500
America	PRI	9,993	9,984	899	899
America	PRY	10,000	10,000	900	900
America	SLV	9,965	9,982	897	898
America	SUR	13,699	13,698	1,233	1,233

America	SXM	1,305	1,396	117	126
America	TTO	9,993	9,984	899	899
America	VCT	9,981	9,792	898	881
America	VEN	41,916	41,911	3,772	3,772
America	VGB	9,869	9,851	888	887
America	VIR	9,945	9,874	895	889
Asia	AUS-QLD	9,979	9,984	898	899
Asia	BGD	9,962	9,993	897	899
Asia	BRN	9,989	9,997	899	900
Asia	BTN	10,000	10,000	900	900
Asia	FJI	9,983	9,957	898	896
Asia	IDN	49,967	49,984	4,497	4,499
Asia	IND-AND	9,950	9,908	896	892
Asia	IND-WEST	9,989	9,990	899	899
Asia	IND-EAST	9,997	10,000	900	900
Asia	KHM	9,993	9,999	899	900
Asia	LAO	10,000	10,000	900	900
Asia	LKA	10,000	9,993	900	899
Asia	MMR	16,076	16,078	1,447	1,447
Asia	MYS	20,199	20,209	1,818	1,819
Asia	NCL	9,961	9,932	896	894
Asia	PHL	13,251	13,251	1,193	1,193
Asia	PNG	39,738	39,691	3,576	3,572
Asia	SGP	9,904	9,961	891	896
Asia	SLB	9,939	9,853	895	887
Asia	THA	9,990	9,997	899	900
Asia	TLS	9,994	9,966	899	897
Asia	VNM	9,998	10,000	900	900
Asia	VUT	9,977	9,925	898	893
Brazil	BRA-AC	13,164	13,164	1,185	1,185
Brazil	BRA-AL	9,996	9,998	900	900
Brazil	BRA-AM	50,000	50,000	4,500	4,500
Brazil	BRA-AP	11,466	11,469	1,032	1,032
Brazil	BRA-BA	9,986	9,998	899	900
Brazil	BRA-CE	9,996	9,999	900	900
Brazil	BRA-ES	9,989	9,999	899	900
Brazil	BRA-GO	10,000	10,000	900	900
Brazil	BRA-MA	9,987	9,997	899	900
Brazil	BRA-MG	10,000	10,000	900	900
Brazil	BRA-MS	10,000	10,000	900	900
Brazil	BRA-MT	31,678	31,678	2,851	2,851
Brazil	BRA-PA	49,999	49,999	4,500	4,500
Brazil	BRA-PB	9,973	10,000	898	900
Brazil	BRA-PE	9,964	9,999	897	900
Brazil	BRA-PI	10,000	10,000	900	900
Brazil	BRA-PR	9,996	10,000	900	900
Brazil	BRA-RJ	9,994	9,991	899	899
Brazil	BRA-RN	9,949	9,992	895	899
Brazil	BRA-RO	12,964	12,964	1,167	1,167
Brazil	BRA-RR	15,548	15,548	1,399	1,399
Brazil	BRA-RS	10,000	9,999	900	900
Brazil	BRA-SC	9,999	10,000	900	900
Brazil	BRA-SE	9,970	9,999	897	900
Brazil	BRA-SP	9,997	9,997	900	900

Brazil	BRA-TO	10,000	10,000	900	900
TOTAL		1,601,810	1,584,888	144,163	142,647

2.4 Mathematical formulas for accuracy indices

Table S4: **Confusion matrix used to compute accuracy indices**. A confusion matrix can be computed to compare model predictions with observations.

		Observations		Total
		0 (non-deforested)	1 (deforested)	
Predictions	0	n_{00}	n_{01}	n_{0+}
	1	n_{10}	n_{11}	n_{1+}
Total		n_{+0}	n_{+1}	n

Table S5: **Formulas used to compute accuracy indices**. Several accuracy indices can be computed from the confusion matrix to estimate and compare models' predictive skill. We followed the definitions of Pontius *et al.* (2008) for the FOM and Liu *et al.* (2011) for the other indices. Note that the AUC relies on the predicted probabilities for observations 0 (non-deforested) and 1 (deforested), not on the confusion matrix.

Index	Formula
Overall Accuracy	$\text{OA} = (n_{11} + n_{00})/n$
Expected Accuracy	$\text{EA} = (n_{1+}n_{+1} + n_{0+}n_{+0})/n^2$
Figure Of Merit	$\text{FOM} = n_{11}/(n_{11} + n_{10} + n_{01})$
Sensitivity	$\text{Sen} = n_{11}/(n_{11} + n_{01})$
Specificity	$\text{Spe} = n_{00}/(n_{00} + n_{10})$
True Skill Statistics	$\text{TSS} = \text{Sen} + \text{Spe} - 1$
Cohen's Kappa	$\text{K} = (\text{OA} - \text{EA})/(1 - \text{EA})$
Area Under ROC Curve	$\text{AUC} = 1/(n_{+1}n_{+0}) \sum_{i=1}^{n_{+0}} \sum_{j=1}^{n_{+1}} \phi(\delta_i, \theta_j)$
	where $\phi(\delta_i, \theta_j)$ equals 1 if $\theta_j > \delta_i$, 1/2 if $\theta_j = \delta_i$, and 0 otherwise
	δ_i and θ_j are the predicted probabilities for $Y_i = 0$ and $Y_j = 1$

3 Figures

References

Banerjee, S., Carlin, B.P. & Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. Chapman and Hall/CRC.
URL https://doi.org/10.1201%2Fb17115

Besag, J., York, J. & Mollié, A. (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–20.

Blanc, L., Echard, M., Herault, B., Bonal, D., Marcon, E., Chave, J. & Baraloto, C. (2009) Dynamics of aboveground carbon stocks in a selectively logged tropical forest. *Ecological Applications*, **19**, 1397–1404.

Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters*, **8**, 2–14.
URL https://doi.org/10.1111/j.1461-0248.2004.00702.x

Dezécache, C., Salles, J.M., Vieilledent, G. & Hérault, B. (2017) Moving forward socio-economically focused models of deforestation. *Global Change Biology*, **23**, 3484–3500. ISSN 1365-2486.
URL http://dx.doi.org/10.1111/gcb.13611

Gibson, L., Lee, T.M., Koh, L.P., Brook, B.W., Gardner, T.A., Barlow, J., Peres, C.A., Bradshaw, C.J.A., Laurance, W.F., Lovejoy, T.E. & Sodhi, N.S. (2011) Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, **478**, 378–381. ISSN 0028-0836.
URL http://dx.doi.org/10.1038/nature10425

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, **202**, 18–27.
URL https://doi.org/10.1016/j.rse.2017.06.031

Grinand, C., Le Maire, G., Vieilledent, G., Razakamanarivo, H., Razafimbelo, T. & Bernoux, M. (2017) Estimating temporal changes in soil carbon stocks at ecoregional scale in Madagascar using remote-sensing. *International Journal of Applied Earth Observation and Geoinformation*, **54**, 1–14.

Grouzis, M., Razanaka, S., Le Floc'h, E. & Leprun, J.C. (2001) Évolution de la végétation et de quelques paramètres édaphiques au cours de la phase post-culturale dans la région d'Analabo. *Sociétés paysannes, transitions agraires et dynamiques écologiques dans le Sud-Ouest de Madagascar, Antananarivo, IRD/CNRE*, pp. 327–337.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O. & Townshend, J.R.G. (2013) High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, **342**, 850–853.
URL http://www.sciencemag.org/content/342/6160/850.abstract

Liu, C., White, M. & Newell, G. (2011) Measuring and comparing the accuracy of species

distribution models with presence-absence data. *Ecography*, **34**, 232–243. ISSN 1600-0587.
URL http://dx.doi.org/10.1111/j.1600-0587.2010.06354.x

Pontius, R., Boersma, W., Castella, J.C., Clarke, K., de Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T., Veldkamp, A. & Verburg, P. (2008) Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, **42**, 11–37. ISSN 0570-1864.
URL http://dx.doi.org/10.1007/s00168-007-0138-2

Rosenthal, J.S. *et al.* (2011) Optimal proposal distributions and adaptive MCMC. *Handbook of Markov Chain Monte Carlo*, **4**.

UNEP-WCMC & IUCN (2020) Protected Planet: The World Database on Protected Areas (WDPA). On-line; 01-03/2020, Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net.
URL https://www.protectedplanet.net

Vancutsem, C., Achard, F., Pekel, J.F., Vieilledent, G., Carboni, S., Simonetti, D. & Gallego, J. (2020) High-resolution mapping of tropical moist forest cover dynamics over the last 30 years.
URL https://doi.org/10.5194/egusphere-egu2020-13834

Vieilledent, G., Nourtier, M., Grinand, C., Pedrono, M., Clausen, A., Rabetrano, T., Rakotoarijaona, J.R., Rakotoarivelo, B., Rakotomalala, F.A., Rakotomalala, L., Razafimpahanana, A., Ralison, J.M. & Achard, F. (2020) It's not just poverty: unregulated global market and bad governance explain unceasing deforestation in Western Madagascar.
URL https://doi.org/10.1101/2020.07.30.229104