

Moving forward socio-economically focused models of deforestation

CAMILLE DEZÉCACHE¹ , JEAN-MICHEL SALLES², GHISLAIN VIEILLEDENT^{3,4}  and BRUNO HÉRAULT⁵ 

¹Université de la Guyane, UMR EcoFoG (AgroParistech, CNRS, Cirad, Inra, Université des Antilles, Université de la Guyane), Campus agronomique de Kourou, 97310 Kourou, French Guiana, France, ²CNRS, UMR LAMETA (CNRS, Inra, SupAgro, Université de Montpellier), Campus Inra-SupAgro, Bat.26, 2 Place Viala, 34060 Montpellier Cedex 2, France, ³Cirad, UPR Forêts et Sociétés, 34398 Montpellier, France, ⁴JRC, Bio-Economy Unit (JRC.D.1), Joint Research Center of the European Commission, 21027 Ispra, Italy, ⁵Cirad, UMR EcoFoG (AgroParistech, CNRS, Cirad, Inra, Université des Antilles, Université de la Guyane), Campus agronomique de Kourou, 97310 Kourou, French Guiana, France

Abstract

Whilst high-resolution spatial variables contribute to a good fit of spatially explicit deforestation models, socio-economic processes are often beyond the scope of these models. Such a low level of interest in the socio-economic dimension of deforestation limits the relevancy of these models for decision-making and may be the cause of their failure to accurately predict observed deforestation trends in the medium term. This study aims to propose a flexible methodology for taking into account multiple drivers of deforestation in tropical forested areas, where the intensity of deforestation is explicitly predicted based on socio-economic variables. By coupling a model of deforestation location based on spatial environmental variables with several sub-models of deforestation intensity based on socio-economic variables, we were able to create a map of predicted deforestation over the period 2001–2014 in French Guiana. This map was compared to a reference map for accuracy assessment, not only at the pixel scale but also over cells ranging from 1 to approximately 600 sq. km. Highly significant relationships were explicitly established between deforestation intensity and several socio-economic variables: population growth, the amount of agricultural subsidies, gold and wood production. Such a precise characterization of socio-economic processes allows to avoid overestimation biases in high deforestation areas, suggesting a better integration of socio-economic processes in the models. Whilst considering deforestation as a purely geographical process contributes to the creation of conservative models unable to effectively assess changes in the socio-economic and political contexts influencing deforestation trends, this explicit characterization of the socio-economic dimension of deforestation is critical for the creation of deforestation scenarios in REDD+ projects.

Keywords: demography, Guiana shield, REDD+, scenarios, spatially explicit modelling, subsidies

Received 1 August 2016 and accepted 14 December 2016

Introduction

Modelling deforestation to identify the main socio-economic and biophysical drivers of land-use change is a complex and multidisciplinary field, involving economic, environmental and geographical issues (De Pinto & Nelson, 2007). Until the late 1990s, non-spatially explicit models of deforestation were more common than models integrating spatial variables into their conceptual framework (Angelsen & Kaimowitz, 1999), even though the deforestation process is intrinsically a spatial process. The development of more powerful computers, associated with a widespread use of geographical information systems (GIS) and associated land cover data from remote sensing products, has democratized deforestation models taking into account spatial factors.

Creating a spatial model of deforestation implies calculating probabilities of deforestation by spatial units, often satellite-derived pixels. A major question deals with the relevancy of predicting human-induced processes at the pixel scale (Irwin & Geoghegan, 2001). Whilst the pixel unit is convenient for creating deforestation maps or quantifying geographical variables, such as elevation or distance to a given infrastructure, many socio-economic or political variables must be measured at a coarser scale (Agarwal *et al.*, 2005), not only due to technical constraints (cost of a census, for example) but also to achieve a better consistency intrinsic to the variable measured. Indeed, demography, gross domestic product (GDP) or percentage of employment in agriculture are aggregate variables and do not, by nature, make sense over a fine continuous spatial scale. Combining explanatory variables defined at very different resolutions into a single model may lead to an underestimation of the significance of aggregate socio-

Correspondence: Camille Dezécache, tel. +594 5 94 32 92 17, fax 594 5 94 28 43 02, e-mail: Camille.Dezecache@ecofog.gf

economic variables, always available at much coarser scales, in comparison with the pixel-based deforestation process (Vieilledent *et al.*, 2013). To cope with these methodological problems, a clear distinction should be made between the prediction of the location of deforestation (location component) and the prediction of its intensity (intensity component) (Pontius & Millones, 2011; Vieilledent *et al.*, 2013), as proposed by most software used to build spatially explicit models of deforestation, such as Dinamica Ego (Soares-Filho *et al.*, 2002, 2006), CLUE-S (Verbrug *et al.*, 2002), GEOMOD (Pontius *et al.*, 2001) or Land Change Modeller (Kim, 2010).

In most deforestation software and associated studies, however, deforestation intensity is generally estimated based on mean deforestation rates between two dates (Mas *et al.*, 2007) or is predicted based on rough equations not explicitly related to socio-economic variables. As stated by Bax *et al.* (2016), 'non-spatial variables as sociocultural and political drivers are often beyond the scope of the models'. In two recent studies, future deforestation scenarios were built following assumptions made based on a good knowledge of economic pressures or social and political tensions in Bolivia and Brazil (Aguilar *et al.*, 2016; Tejada *et al.*, 2016). When going through the modelling process, however, predicted deforestation trends in the different scenarios are simply expressed as a fraction of past observed deforestation trends or established without an explicit estimation based on socio-economic variables. This lack of attention given to the explicit integration of the underlying forces of land cover change may have contributed to the failure of spatially explicit models to capture deforestation trends, as stated by Dalla-Nora *et al.* (2014). In their review, they identified two main types of spatially explicit land-use change models, each having major drawbacks regarding the formulation of their intensity component. Global models mainly focus on major economic indicators such as economic or population growth, but fail to include intra-regional drivers of deforestation such as political or normative constraints, or variables describing local human activities such as logging or mining. On the contrary, intra-regional models mainly focus on local drivers of deforestation, often implicitly integrated within past deforestation trends used as baselines, with no consideration for major underlying forces (Dalla-Nora *et al.*, 2014).

The main objective of the present study is thus to move forward socio-economically focused models of deforestation, wherein much effort is made to fully integrate the major social, economic and political underlying forces of deforestation intensity, whilst allowing these underlying forces to interact with local direct deforestation drivers. Focusing on the processes leading to deforestation rather than on deforestation

patterns themselves is necessary to provide useful inputs for decision-makers (Brown *et al.*, 2014).

Our location model aims principally at allowing for a realistic representation of deforestation patterns in predicted maps through the use of a machine learning algorithm (Brown *et al.*, 2013). The model of deforestation intensity was divided between several sub-models, each focusing on specific drivers of deforestation in French Guiana over the period 2001–2014. We considered that to efficiently account for both local and global forces of deforestation, such a zoning of the territory was necessary. Indeed, we assumed that the difficulty of integrating underlying global forces of deforestation within intra-regional models was due to the fact that, even if they are global, diffuse or aggregated over large areas, these forces are not acting everywhere. Local political, geographical and economic constraints thus modify how and where global constraints can affect deforestation trends, as can be illustrated with the example of gold mining. Whilst increased gold prices may be a major underlying cause of illegal gold mining expansion in French Guiana, this variable is a driving force of deforestation only in areas where gold deposits are present, whereas gold prices have little to do with the expansion of agriculture or urbanization. Here, local constraints are such that gold mining only occurs where gold deposits are located (geographical constraint) and, eventually, where gold mining is authorized (normative constraint, i.e. imposed by a superior authority).

In our modelling framework, global factors are therefore operating through a local filter, whether it is geographical or normative. If global factors are relevant in a given area, they are taken into account in the intensity model and can eventually interact with local factors of deforestation. Such methodology adds many flexibility to take into account drivers operating at very different temporal scales. Gold mining activity, for example, is expected to be very dynamic in time, following the volatility of gold prices, whereas urban expansion occurs over longer periods of time. Thus, models should be able to take into account those differences and also analyse short-term dynamics. This has recently been made possible by the publication of yearly deforestation data (Hansen *et al.*, 2013), which is critical because deforestation is highly variable in time (Irwin & Geoghegan, 2001) and often driven by commodities booms (Angelsen & Kaimowitz, 2001).

This methodology allowed for the efficient assessment of contrasting deforestation trends amongst the different areas of the territory under pressure of different drivers of deforestation. Strong relationships were evidenced between socio-economic drivers, such as population growth in parallel with the level of agricultural subsidies, and deforestation, which we find

indispensable in the formulation of comprehensive future deforestation scenarios. As a comparison, we built a null deforestation model following more classical methods, where socio-economic variables were spatialized and included in the location component of the model, whilst the intensity component of deforestation was estimated based only on historical deforestation rates. Such a comparison shows that our method is less biased by the underestimation of low deforestation and overestimation of high deforestation areas, suggesting a better integration of socio-economic variables.

Materials and methods

Study area

French Guiana is a French overseas territory located in South America and covering approximately 85 000 sq. km (Fig. 1a).

Along with its neighbouring countries belonging to the Guiana Shield, French Guiana is characterized by a very high forest cover, more than 90% of its territory being covered by a dense tropical rainforest (Rossi *et al.*, 2015), and amongst the lowest rates of deforestation in the world, with an annual deforestation rate of around 0.03% (Hammond, 2005). Non-forested areas are basically restricted to the coast (Fig. 1b), where more than 90% of total inhabitants are concentrated. The remaining villages are distributed along navigable rivers on the borders with Suriname and Brazil. Deforestation in French Guiana in the beginning of the 2000s was mainly due to urbanization, infrastructure building, agricultural expansion and gold mining. Low accessibility to the interior of the forest (main roads are restricted to the coastal areas), and geographical and strong normative constraints *de facto* provoke a separation of the different human activities inducing deforestation (Fig. 1c). Gold mining is restricted to a Greenstone formation (Hammond *et al.*, 2007), i.e. gold-bearing rocks, located in remote areas of the forest and mostly accessed by rivers or helicopters. Logging activities occur within the

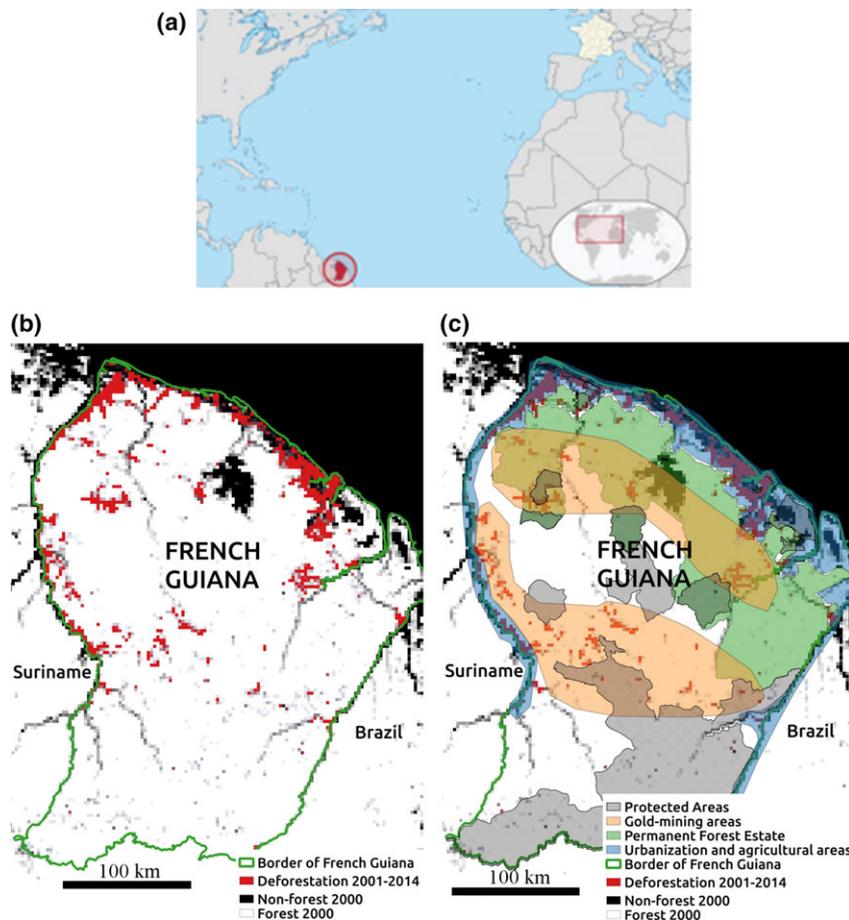


Fig. 1 (a) Overview of the global location of French Guiana. (b) Forest cover and deforestation map. Forest and non-forest areas by 2000 are shown in white and black, respectively. Red areas correspond to main deforestation hot spots during 2001–2014 at 5-km resolution: these are not fully deforested areas, but areas where deforested pixels at 30 m resolution were observed during this period. (c) Overview of the location of the main activities leading to deforestation. Orange polygons roughly identify gold mining areas. The Permanent Forest Estate is represented in green. Purple areas correspond to the location of most cities and villages where deforestation due to urbanization and agriculture is occurring. Grey polygons are integrally protected areas. Full-size maps are presented as (Figures S1 and S2).

Permanent Forest Estate (PFE), where all other types of activities, including agriculture, are forbidden, although legal gold-mining concessions may overlap with the PFE, as well as small-scale illegal gold mining to a lesser extent. Agriculture and urbanization could theoretically occur in the remaining parts of French Guiana, but *de facto* are concentrated in the few accessible coastal areas and along main rivers.

Urbanization and infrastructure expansion in French Guiana are driven by a strong demographic increase: since 2009, the mean annual growth has been 3.9%, six times higher than that of Metropolitan France. Agriculture is also expanding in response to this increasing population, but the territory remains largely dependent upon imports (Hecquet & Moriname, 2008). Integrally protected areas cover a great part of the territory, mostly remote forested areas.

Conceptual approach

The methodology followed in the current study is based on a distinction between the definition of a spatial deforestation potential (location component of the model) and the estimation of the quantity of deforestation (intensity component of the model) (Dalla-Nora *et al.*, 2014), both components being built independently. Once the amount of deforestation has been defined by the intensity component, the corresponding number of pixels is sampled from the pool of pixels having the highest deforestation potential from the deforestation suitability map (hereafter called the deforestation risk map).

We first created a deforestation risk model based on potential spatial predictors of deforestation location over the whole territory of French Guiana. The pool of spatial variables initially included within the location component was chosen based on a thorough knowledge of the local causes of deforestation. At this stage, however, the variable selection procedure was simply based on changes in the model's error rate when adding/removing an explanatory variable, rather than on a theoretical model of the spatial drivers of deforestation. To produce an accurate deforestation risk map, we used Random Forest algorithm (Breiman, 2001) as the classifier, as it is known for its strong predictive accuracy, robustness to noise

(Dietterich, 2000) and ability to take into account complex sets of interactions and nonlinear relationships between deforestation and its predictors (Evans *et al.*, 2011).

Concerning the intensity component of the model, which is the main focus of the present study, we first listed the main previously mentioned drivers of deforestation in French Guiana: agriculture and urban expansion, gold mining and, to a lesser extent, logging activities, mostly through the opening of forest tracks, as only reduced impact logging is practiced in French Guiana (Verger, 2014a). We then created a different sub-model of deforestation intensity for each of these drivers, considering them to be influenced by a different combination of global and local factors, and spatially distributed differently in relation to specific geographical and normative constraints (Table 1) which allowed a zoning of the territory, where each zone is associated with a specific human activity leading to deforestation. Agriculture and urbanization were merged into a unique sub-model, assuming that both drivers are mainly influenced by population growth. Indeed, agriculture in French Guiana is mainly meant for local markets consumption or subsistence, with no expected dependence upon commodities prices. Similar to agriculture, logging activity is mainly dependent on the small local demand for wood, and deforestation due to logging is not likely to be influenced by wood prices. On the contrary, given the boom observed during the beginning of the 2000s, from approximately 300 USD/ounce in 2000 to close to 1900 USD/ounce in 2011 (Hammond *et al.*, 2007; Alvarez-Berrios & Mitchell Aide, 2015), we considered gold prices to be the main underlying driver of deforestation due to gold mining. Ideally, we would have included only global exogenous underlying forces of deforestation in our different sub-models, in interaction with local factors. Due to data constraints or for better models consistency, however, we used some endogenous variables within the logging and gold mining sub-models. More details about this choice are given in the corresponding parts of the Materials and methods section.

A flowchart of the full model is shown in Fig. 2, showing the two independent location and intensity components. Sub-model parameters were estimated using simple and mixed

Table 1 Global and local deforestation drivers acting in each of the zones considered for the intensity sub-models, each zone being defined by a specific set of drivers and some normative and geographical constraints

Sub-model	Agriculture/urbanization	PFE (logging)	Gold mining
Potential global factors	Population growth	Demand for wood	Gold prices
Normative filter	Forbidden in integrally protected areas and in the PFE	Authorized within the boundaries of the PFE	Authorized only within gold-mining concessions but illegal gold mining is a major concern
Geographical filter	Accessibility	Accessibility Presence of remaining forested areas	Restricted to Greenstone areas
Local factors	Rural or urban ways of life Economic incentives for agriculture	Incentives to use wood for construction	Intensity of the fight against illegal mining
Explanatory variables included in the sub-model	Population growth Agricultural subsidies	Wood production	Gold production

Explanatory variables effectively integrated within the different sub-models are also listed.

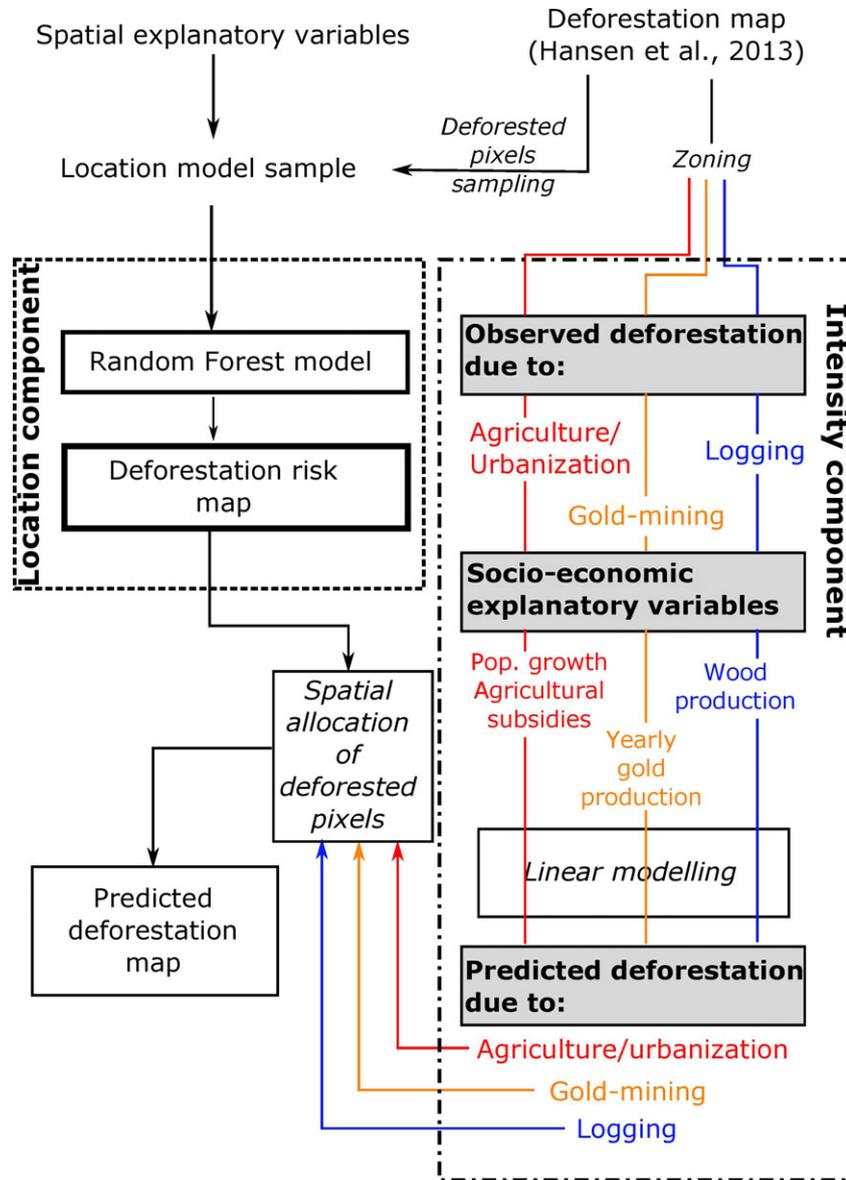


Fig. 2 Flowchart of the full model computed, built based on two independent location and intensity components. The location component is computed for the whole territory at once based on Random Forest algorithm, whereas the intensity component is composed of different sub-models, each focusing on a different type of human activity leading to deforestation.

linear modelling tools, for each of the three areas considered: agriculture and urbanization areas, the PFE and gold mining areas. Finally, for each relevant area of the territory defined, the estimated number of pixels to deforest is sampled amongst the pool of pixels of highest deforestation potential in the deforestation risk map.

Location component

Input variables. Input forest cover and deforestation data—Deforestation data collected between 2001 and 2014 were based on yearly maps from Landsat images (Hansen *et al.*, 2013). After re-projection to EPSG:3857 and re-sampling at

30 m, a 75% crown cover threshold was applied to define forested areas. A neighbouring filter was used to remove noise in the form of isolated deforested pixels, most likely caused by misclassification of satellite images (Mather, 2004) or natural gap-phase forest dynamics. Moreover, large patches of deforestation appearing within swampy areas and mangroves were removed because natural processes, such as coastal dynamics, were assumed to be the cause of these observations.

Spatial explanatory variables—Input maps for spatial explanatory variables (Table 2) were imported into the GIS database as shapefiles and then rasterized to the extent and

Table 2 List of spatial variables included in the location component of the coupled model

Variable name	Resolution (m)	Range	Sources
Dist. to nearest road (Droad)	150	0–170 km	ONF (2014)
Dist. to nearest stream following Strahler classification			
Order 1–3: small (Dstrahler13)	150	0–2 km	Horton (1945), Strahler (1952), USGS (2000)
Order 4–6: intermediate (Dstrahler46)	150	0–15 km	
Order 7 or +: large (Dstrahler7+)	150	0–120 km	
Dist. to closest city (Dcity)	150	0–170 km	Verger (2011)
Dist. to Greenstone (Dgreen)	150	0–65 km	BRGM (2014)
Protected areas (PA)	30	Binary	DEAL Guyane (2015), Joubert (2015)
Permanent Forest Estate (PFE)	30	Binary	Verger (2014b)
Small-scale gold mining authorizations (SS-GM)	30	Binary	DEAL Guyane (2013b)
Permanently flooded areas (flood)	30	Binary	DEAL Guyane (2013a)

Acronyms used in the text are indicated between parentheses next to variable names. The resolution corresponds to the precision adopted for the rasterization of each spatial explanatory variable.

resolution of forest cover and deforestation data. When appropriate, distance rasters were calculated at a coarser scale to avoid excessive calculation times.

Elevation and slope were not included in the model because high values of both variables are found in very remote areas where no-deforestation is likely to occur. However, elevation data were used to compute the stream network with a minimal watershed area of 0.5 sq. km, as applied in Cassard *et al.* (2008). Streams were divided into three classes: small, intermediate and large rivers, following Strahler classification (Strahler, 1952). Indeed, different sizes of streams allow for different types of human activities characterized by different spatial deforestation patterns, from alluvial gold mining along smaller ones, to human settlements close to larger rivers. Three measures of distance to streams were then computed, one for each stream size class. Distance to previously deforested areas as a measure of spatial auto-correlation was excluded to avoid collinearity with other variables, especially distance to nearest road and city.

Random Forest modelling. Instead of classical statistical tools like binary or multinomial logistic regression used in numerous deforestation studies (Overmars *et al.*, 2003; Mahapatra & Kant, 2005), we preferred the use of the Random Forest (RF) algorithm (Breiman, 2001) for computing the location of deforestation. RF, similar to many other ensemble learning algorithms, consists of a collection of decision trees. Used for classification purposes, each tree associates each sample unit (pixel) to a given class, and the majority vote amongst all decision trees in the forest gives the most probable class for each given pixel (Breiman, 2001). Ensemble learning algorithms have received growing interest because of their strong prediction accuracy and robustness to noise (Dietterich, 2000), as well as their ability to take into account complex sets of interactions and nonlinearity between explanatory variables (Evans *et al.*, 2011). RF in particular is also interesting for its unbiased error rates, known as out-of-bag (OOB) error rate, and its variable importance estimation procedure (Rodriguez-Galiano *et al.*, 2012).

Sampling choices. Due to the extreme scarcity of deforestation in the area of interest, defining the appropriate sampling method for building the calibration data set of the model was a major issue. Randomly sampling n pixels may have led to a great underrepresentation of the number of deforested pixels, due to the high probability of sampling only not-deforested pixels with a small n . Another option, as used by Vieilledent *et al.* (2013), would have been to define a minimum sample size n_{\min} necessary to estimate observed deforestation rates within a predefined confidence interval. In our case, however, n_{\min} was too large because of the extremely low deforestation rates. Moreover, using an unbalanced sample of deforested vs. not-deforested pixels would have strongly influenced the model's likelihood towards predicting the no-deforestation dominant class. On the contrary, a deforestation model should be able to accurately predict both classes without advantaging one or the other. Drawing a parallel with species distribution modelling, where presence data may be scarce in a given landscape, using a balanced sample between presence and absence has to be preferred in the case of a RF classification (Barbet-Massin *et al.*, 2012).

In this study, we therefore chose to sample the same amount of deforested and not-deforested pixels to obtain a balanced sample where model calibration would penalize misclassification of deforestation as well as no-deforestation. To avoid excessive data processing, we arbitrarily sampled n pixels corresponding to one-third of the amount of deforested pixels during 2001–2014. This represents a final sample size of approximately 130 000 pixels, half of which are deforested and half not-deforested. Such a sampling method does not affect the results of our models in terms of deforestation intensity, as it is explicitly determined by the intensity model. Here, we are not interested in the prevalence of deforestation, but rather in the relative effect of spatial variables on the location of deforestation. This explains why the extent of the PFE is used as a binary variable in the location component of the model and is also the focus of one of the intensity sub-models, which could give the impression that our location and intensity components are not fully independent. The PFE, however, was not included as a spatial variable in the location model to

simulate a decrease or increase in deforestation risk within the PFE, but rather to enable us to consider different sets of interactions between pixels inside this area compared to pixels outside.

Variable selection. Two measures of variables importance are computed internally within the RF algorithm, as implemented in the R-package *randomForest* (Liaw & Wiener, 2002): the mean decrease in accuracy (MDA) and the mean decrease in Gini index (MDG). The MDA indicates changes in prediction accuracy when values of an explanatory variable of interest are randomly permuted compared to observations. On the other hand, the MDG quantifies the decrease in Gini impurity, or the increase in the homogeneity within each child sub-sample. When a variable is chosen to split a parent sample into two children sub-samples, and if such a variable is important, it should contribute to a higher homogeneity in the children sub-samples, by efficiently discriminating two sets of observed values. MDG was found to be more robust to small changes in the data set (Calle & Urrea, 2011) and was therefore used for the latter variable selection step. Several simulations were led to check for the stability of variables ranking.

In a first step, all explanatory variables were included in the model. The least important variable following MDG was then excluded and OOB error rates were compared to the former model's score. This step was repeated until the removal of the least important variable no longer contributed to a decrease in both error rates associated with deforestation and no-deforestation classes.

Deforestation risk map. Using a stack of rasters corresponding to each explanatory variable and the RF model, a deforestation risk map was computed for all the area of interest. We insist on the fact that the term 'probability map' is not appropriate in our case for characterizing the potential of each pixel to be deforested, as absolute values of predictions rely on the initial balanced distribution of deforested and not-deforested pixels within the calibration sample. As previously mentioned, in the location model we are not interested in the prevalence of deforestation, and as such, we prefer the use of the expressions 'deforestation risk' or 'deforestation potential'. Rather than considering the absolute value of this index, one should instead use it as a way to rank pixels by levels of deforestation risk.

Intensity component

Apart from the location component, an intensity component was built to explicitly estimate, based on socio-economic variables characterizing relevant human activities in diverse areas of the territory, the number of pixels necessary to sample from the deforestation risk map to produce the final predicted deforestation map. Contrary to the methodology set-up for the location component, where all the territory was considered as a whole, several sub-models of deforestation intensity were built independently in three different areas considering that, within each of these areas, very different human activities

were observed to be associated with a completely different set of socio-economic explanatory variables (Fig. 3):

- gold mining areas;
- the PFE, where selective logging activities are taking place;
- agriculture and urbanization areas.

The temporal scales of such activities were also assumed to be different: we focused on annual deforestation due to gold-

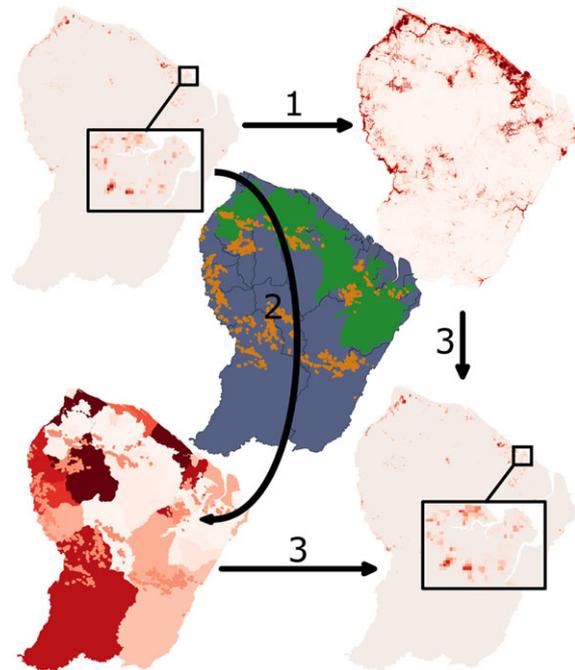


Fig. 3 Spatial flowchart of the coupled model building. (1) Map of observed deforestation during 2001–2014 (top-left) is used to create a sample of deforested and not-deforested pixels. Every pixel of this sample is associated with corresponding value of each spatial explanatory variable included in Table 2. This data set is then used to calibrate a Random Forest model of the location of deforestation, itself enabling the calculation of a deforestation risk map over the entire study area (top-right). (2) A zoning of the country is determined, where each area corresponds to the spatial extent of a specific group of human activities (middle, agriculture and urbanization areas are displayed in blue, the Permanent Forest Estate in green and gold mining areas in orange with a buffer applied for better visibility). Within each zone, a specific deforestation intensity sub-model is built using appropriate socio-economic explanatory variables. The results of these sub-models are displayed in shades of red (bottom-left), from low (light red) to high (dark red) intensities. (3) A deforestation prediction map (bottom-right) is created by coupling the outputs of the intensity model and the deforestation risk map. The corresponding number of deforested pixels predicted by each sub-model is sampled from the pool of pixels of highest deforestation risk value. In this figure, both observed and predicted deforestation maps are a focus in the north-west part of French Guiana, and deforestation was summed over a grid of 1-km cell size for display purposes. Thus, grey areas indicate no-deforestation, light red = low deforestation and dark red = high deforestation.

mining, but on total deforestation over 2001–2014 due to logging, urbanization and agriculture, because gold mining is assumed to be a very dynamic process on a short time scale, whereas the others are associated with longer-term socio-economic processes.

Deforestation intensity is often expressed in terms of rates; however, following Brown & Pearce (1994), only absolute forest cover loss per year should be used in deforestation studies, because deforestation rates are more biased by unmeasured geographical and historical factors, creating artificial differences between administrative subdivisions of very different areas. For example, the district of Cayenne in French Guiana covers only approximately 20 sq. km, whereas the district of Maripasoula is nearly 1000 times larger. A similar-sized area deforested in each district, for a similar reason, would be interpreted very differently in terms of deforestation rates. To avoid such biases, absolute deforestation (number of hectares or pixels deforested) was modelled using mixed linear models in the present study instead of deforestation rates, which are often described by a binomial distribution. As previously mentioned, a neighbouring filter was applied to remove isolated pixels from the deforestation maps. This is particularly important when dealing with huge administrative units such as districts, where errors can accumulate over large areas and sometimes exceed effective deforestation. When dealing with gold mining and the effect of logging, however, such small-scale activities can easily be excluded by the filter used. We therefore chose not to filter the deforestation data for both gold mining and logging sub-models. In the case of gold-mining, the mask used to identify areas impacted by gold-mining was precise enough to avoid the accumulation of noise over large areas. In the case of logging, where forest sectors correspond to areas of thousands of hectares, we controlled for the noise in the data by including sector area as a linear covariate in the model.

Gold mining. The gold mining area was defined using the results of the remote sensing monitoring of gold mining for the year 2014 in the Guiana Shield (Rahm *et al.*, 2015) and previous assessments in 2001 and 2008 (Debarros & Joubert, 2010), each study providing GIS layers of areas impacted by gold mining. These layers were merged and used as a mask for deforestation maps by Hansen *et al.* (2013), after creating a 90-m buffer around each polygon to take into account potential imprecision of the manual digitizing of gold-mining impacts. Yearly deforestation due to gold mining within the area of interest was extracted and a moving average over 3 years was calculated to avoid excessive variability induced by a potential lag between the deforestation event and its detection by satellites in case of cloud persistence. These values were then used to calibrate a linear model of yearly deforestation due to gold mining. According to Hammond *et al.* (2007), gold production over the Guiana Shield is highly correlated with gold prices. We assumed that the level of gold mining activity is positively correlated with deforestation and thus built a model using gold production as an explanatory variable. We also tested the direct effect of gold price on deforestation due to gold mining. A counter-

intuitive significant negative correlation between deforestation and gold prices was found. Even though gold production is endogenous, whereas international gold prices are exogenous and more interesting for modelling of deforestation processes, the sub-model including gold production as the only covariate was selected. Indeed, a repressive policy against illegal gold mining was launched in French Guiana in 2002 (De Rohan *et al.*, 2011) simultaneously with the explosion of gold prices observed in the beginning of the 2000s. This policy may have interfered with the free development of illegal gold mining activities in response to increasing gold prices. This policy effect was impossible to evidence here, however, due to collinearity between the intensity of the repressive policy against illegal gold mining and gold price trends over the period considered. In the absence of official data at the scale of French Guiana, annual gold production was downloaded from global data sets, even though estimates are likely inaccurate due to the importance of informal and illegal activities. Two sources were found to give data on yearly gold production in French Guiana, the first for the period 2002–2012 (Barrientos & Soria, n.d.) and a second for 1992–2012 (24hGold.com n.d.). When the data sets overlapped, we averaged yearly production from both sources. The intensity model in the gold-mining areas was formulated as follows:

$$GM_i \sim N(\beta_{GM_0} + \beta_{GM_P} \times Prod_i; \sigma_{GM}^2)$$

where GM_i is the scaled absolute deforestation (ha) due to gold mining, $Prod_i$ is the estimated gold production (T) at year i and β_{GM_0} and β_{GM_P} are the model parameters.

Permanent Forest Estate. Deforestation intensity was modelled at the forest sector scale, corresponding to areas of very different sizes (min = 0.3 sq. km, max = 1585 sq. km, median = 117 sq. km). Because of the low impact of selective logging in terms of effective deforestation, we modelled total deforestation per sector over the entire period 2001–2014 instead of annual deforestation. Two explanatory variables associated with logging activities in the PFE were initially retained: the amount of wood extracted per forest sector during the period 2001–2012 (ONF Guyane, 2012) and the amount of new forest tracks and roads built within each sector during 2000–2013. This latter variable was estimated based on a detailed map of the roads and tracks network of French Guiana (ONF, 2014). Because of a strong collinearity between those two variables ($R^2 = 0.88$), we chose to use only the first variable tested in the model, assuming that legal logging development determined the extension of the track network and not the opposite. This variable is endogenous, as it is likely to be the consequence of population growth (increased demand for wood) or increased incentives to develop wood production and/or consumption. As these incentives are difficult to estimate, wood production itself is a useful explanatory variable in the perspective of scenarios formulation. Such scenarios would then attempt to answer the question of how much wood would be necessary to meet the demand for wood, given an expected increase in population. As previously mentioned, the surface of each forest sector was used to

take into account noise in the data, in addition to the effect of logging. The model of intensity in the PFE was:

$$PFE_i \sim N(\beta_{PFEV} \times V_{wood_i} + \beta_{PFEz} \times Area_i; \sigma_{PFE}^2)$$

where PFE_i is the total deforestation (ha) between 2001 and 2014, V_{wood_i} is the volume (m^3) of extracted wood, $Area_i$ is the surface (ha) of each forest sector i and β_{PFEV} and β_{PFEz} are the model parameters. We assumed no model intercept, because for a sector of area equal to 0, deforestation is null, as well as wood production.

Agriculture and urbanization areas. This zone corresponds to areas not included in the gold mining and PFE zones where other types of human activities can take place, such as urbanization and agriculture, mainly. Socio-economic data were available at the scale of each district, which corresponds to the smallest administrative subdivision of the territory. Deforestation per district was available for every year, but associated socio-economic variables were not and were thus assumed to be fixed during the whole period. As for the gold mining area, moving averages over 3 years were calculated in order not to overestimate the temporal variability of deforestation.

We included average annual population growth between 1999 and 2009 (INSEE, 2016) and yearly agricultural subsidies for land clearing provided during the period 2000–2014 (DAAF, 2015) as predictors. Population growth was used instead of total population at year i because we assumed that, in a context of high population growth, 'new' inhabitants will need enough space to settle and will cause loss of natural forested habitats. Discussions with a former official of the local Department of Food, Agriculture and Forest indicated that farmers were actually expecting public subsidies to develop their activity, suggesting that the amount of agricultural subsidies is exogenous and that agricultural development would not have occurred the same way in the absence of these subsidies. Agricultural subsidies are provided proportionally to the forested area each farmer declares to convert to agriculture, so we assumed no inter-district variability. A random effect was associated with population growth, however, because we expected large variability in induced deforestation amongst districts, following the type of population considered (urban or rural). Other socio-economic variables (e.g. net income, % of jobs in trade and services, % of jobs in agriculture) were tested in interaction with population growth to modulate the effect of demography following local contexts, but were discarded after automatic model selection based on the AICc criterion [dredge function, R-package *MuMIn* (Barton, 2015)]. To sum up the hypothesis of the agriculture and urbanization sub-model, we assumed that urbanization and subsistence agriculture were mostly driven by population growth (the heterogeneity between those two situations being implicitly taken into account by the random effect), whereas commercial agriculture was boosted mostly by the amount of agricultural subsidies provided. Connectivity to local markets was not explicitly considered here due to the small size of French Guiana and to the fact that Euclidean distance was a

poor estimate of connectivity. Although some districts may be geographically close to main cities, the fact that they are accessible only by boat or airplane makes them far from any market. The final model was the following:

$$AD_{ij} \sim N(\beta_{ADpc,j} \times PopCh_{i,j} + \beta_{ADas} \times AgrSubs_{i,j}; \sigma_{AD}^2)$$

with

$$\beta_{ADpc,j} \sim N(\mu_{ADpc}; \sigma_{ADpc}^2)$$

where AD_{ij} is the deforestation (ha) in district j at year i , $PopCh_{i,j}$ is the increase in population in district j at year i , $AgrSubs_{i,j}$ is the amount of agricultural subsidies in district j at year i and $\beta_{ADpc,j}$ and β_{ADas} are the model parameters. The intercept was assumed to be null, as deforestation is expected to be 0 when the population is stable and no agricultural subsidies are provided.

Coupling and validation

Creation of the predicted deforestation map. A spatial flow-chart of the modelling process is presented in Fig. 3. Observed deforestation maps were used to independently compute a deforestation risk map (step 1) and a deforestation intensity map through the determination of a zoning of the territory (step 2). In a final step, the n pixels with highest deforestation risk on the deforestation risk map were sampled to create the final predicted deforestation map, with n being explicitly estimated by the intensity sub-models for each subdivision of the different zones determined.

Validation of the predicted deforestation map. A confusion matrix was built to compare the final deforestation prediction map with observed deforestation, and several performance indices were computed following definitions from the literature (Pontius *et al.*, 2008; Liu *et al.*, 2011; Vieilledent *et al.*, 2013): overall accuracy (OA), specificity (Spe), sensitivity [i.e. true positive rates (Sen)], figure of merit (FOM) and Cohen's kappa (K). The definition of each index is provided as (Tables S1 and S2). We may, however, want to assess the accuracy of a predicted map at a broader scale than the pixel scale, because we are not interested in predicting the exact location of deforested pixels, but rather areas under pressure. To do so, we tested at which scale the model is accurate enough, based on a series of grids of different cell sizes with widths ranging from 1 to 25 km, which are used to summarize predicted and observed deforestation: plotting the amount of observed against predicted deforested pixels within each cell indicates how well predictions fit the observed data.

Comparing the coupled model with a null model

As a comparison, a null model was created with no explicit determination of the intensity of deforestation using socio-economic variables. Thus, by using a more classical method used in numerous spatially explicit deforestation models, the intensity component of the model simply corresponded to total observed deforestation between 2001 and 2014. The variables

included in the spatial component of the null model were the same as the spatial variables involved in the location component of the first coupled model, but additional socio-economic variables (i.e. population growth and the amount of agricultural subsidies) were included after rasterization at 30 m. Gold production was discarded because yearly production is not spatially dependent and could not be converted into a spatially explicit factor, and wood production was removed during the variables selection procedure.

The same balanced sampling method was applied, and a single RF model was calibrated and used to create a deforestation risk map. n pixels with the highest value of deforestation risk were sampled from the deforestation risk map, n being calculated based on historical deforestation between 2001 and 2014.

The above-mentioned performance indices were estimated, and a grid-based accuracy assessment was performed to compare the performance of the coupled and null models.

All data processing and modelling was carried out using the open-source software R (R Core Team, 2015) and GRASS GIS 7.0 (GRASS Development Team, 2015).

Results

Location component

Variable selection and importance. During the variable selection procedure, permanently flooded area was the only geographical variable excluded from the location model. The most important variables were all distance variables, with the exception of distance to streams of smaller Strahler orders between 1 and 3. Binary variables were all of lower importance (Fig. 4).

Accuracy assessment of the Random Forest model. Out-of-bag error estimates stabilized for a RF size of 100 trees. Mean OOB error rate was 3.2%, with errors associated with deforested and not-deforested pixels of 2.0% and

4.4%, respectively, over the approximately 130 000 pixels of the calibration data set.

Partial plots. Deforestation risk was found to be exponentially and negatively correlated with distance to road (Fig. 5). Overall, distance to city, distance to nearest stream of Strahler order 7 or more and distance to Greenstone areas were also negatively correlated with deforestation risk, but with noisier relationships. In the case of distance to city, for example, increasing distance at a small scale increases the risk of deforestation, but at a larger scale, getting much further from a city decreases the deforestation risk. The pattern is less clear regarding distance to streams of Strahler order between 4 and 6 and distance to nearest small stream of Strahler order between 1 and 3, with a decreasing risk of deforestation for increasing distance at small scale, and a contrary relationship at higher distance, without an overall trend (Fig. 6). Concerning binary variables, protected areas and the Permanent Forest Estate are negatively associated with deforestation, whereas small-scale gold mining authorizations slightly increase the deforestation risk (Fig. 6).

Intensity component

Estimates of the parameters included in the three sub-models forming the intensity component are reported in Table 3. Yearly deforestation due to gold mining was found to be significantly and positively correlated with estimated annual gold production. The volume of logs harvested during the period and the area of each forest sector were both highly correlated with deforestation in the PFE. Mean annual population change between 1999 and 2012 and agricultural subsidies for land clearing between 2000 and 2014 were highly significant positive predictors of deforestation. In one district, predicted

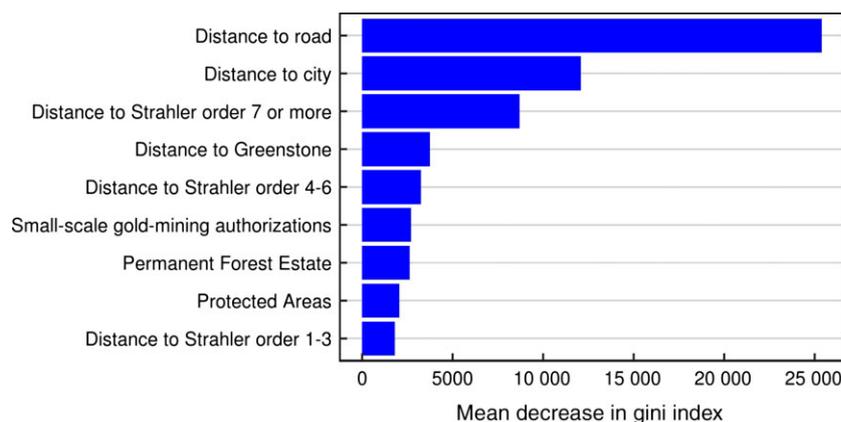


Fig. 4 Ranking of variables included in the location component of the coupled model, based on mean decrease in Gini index. [Colour figure can be viewed at wileyonlinelibrary.com]

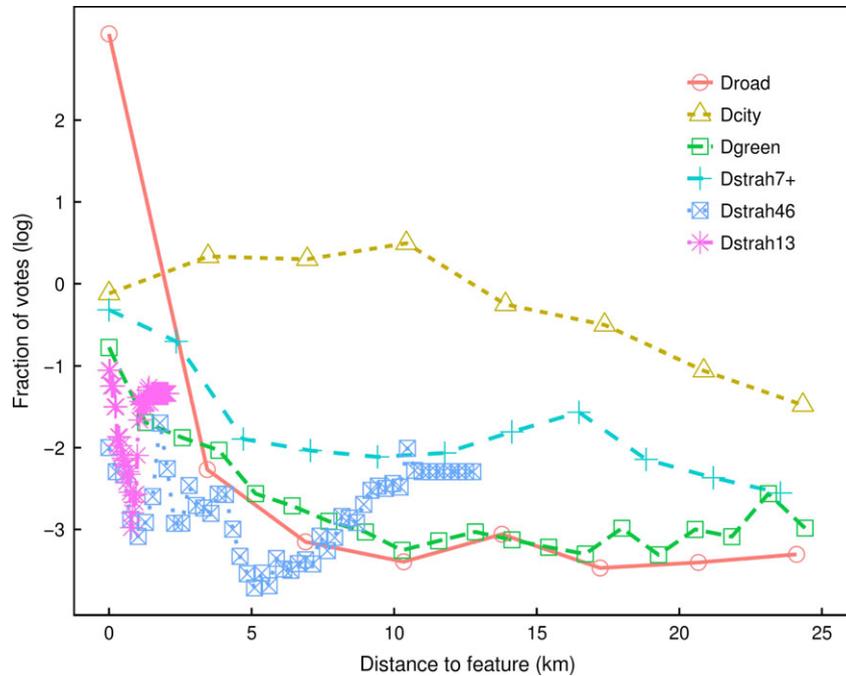


Fig. 5 Partial plot showing the relationship between continuous geographical explanatory variables (distance variables) included in the location component of the coupled model and the log of the fraction of votes by the Random Forest model in favour of classifying a pixel as deforested vs. not-deforested. Droad, Dcity and Dgreen stand for distance to nearest road, city and Greenstone area, respectively. Dstrah7+, Dstrah46 and Dstrah13 stand for distance to nearest stream of Strahler order 7 or more, between 4 and 6 and small streams between 1 and 3, respectively. The variables involved do not all have the same range, which is especially true for Dstrah4 and Dstrah13.

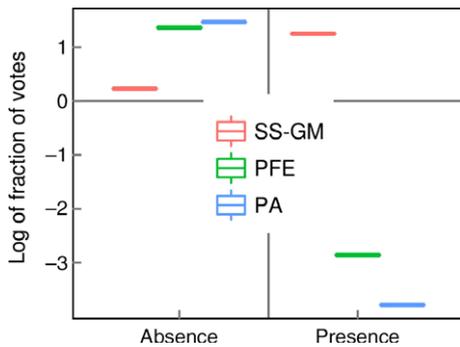


Fig. 6 Partial plots showing the relationship between binary geographical explanatory variables included in the location component of the coupled model and the log of the fraction of votes by the Random Forest model in favour of classifying a pixel as deforested vs. not-deforested. SS-GM, PFE and PA stand for small-scale gold mining authorizations, Permanent Forest Estate and protected areas, respectively.

deforestation was very small but negative, due to a decrease in population over the period, and was replaced by value zero for the model coupling step. Standard error of the random slope associated with annual population change was equivalent to the estimate of the fixed parameter $\beta_{ADPC,j}$ suggesting a strong inter-district variability.

Predictive performance

Visual assessment. A visual comparison between observed and predicted deforestation maps between 2001 and 2014 provided initial insight into our model’s performance. A focus on an agricultural and urbanization hot spot showed that the predicted deforestation map is quite accurate, with most pixels being correctly predicted and mismatched pixels being closed to effectively deforested pixels (Fig. 7a). In gold mining areas, however, deforestation appeared to be more difficult to accurately predict spatially (Fig. 7b).

Pixel-based performance indices. Overall accuracy and Spe were biased by the extremely low rates of deforestation and thus uninformative. In addition, Sen, FOM, and K, which take into account the ability of the model to predict not only the majority class (i.e. no-deforestation), were calculated. On the whole, the null model performed slightly better than the coupled model, wherein all indices were confounded (Table 4).

Grid-based performance assessment. Plots of observed vs. predicted deforestation summed over different grid cell sizes, from 25 to 1 km for each model (Fig. 8),

Table 3 Parameter estimates of the three sub-models included in the intensity component of the coupled model

Intensity sub-model	Intercept	Slope	SD of random slope	Adjusted R-squared
Gold mining areas	$\beta_{GMo} = 604.7^{***}$ (28.6)	$\beta_{GMP} = 0.11^{**}$ (0.03)		0.52
PFE		$\beta_{PFEV} = 1.77^{***}$ (0.13)		0.76
		$\beta_{PFEz} = 0.09^{***}$ (0.01)		
Agriculture and urbanization areas		$\beta_{ADpc,j} = 0.29^{**}$ (0.08)	0.30	Marginal = 0.37
		$\beta_{ADas} = 0.80^{***}$ (0.13)		Conditional = 0.96

SD, standard deviation; PFE, permanent forest estate.

P-values are shown as asterisks ($^{***}P < 0.001$, $^{**}P < 0.01$). Standard errors are presented in parentheses.

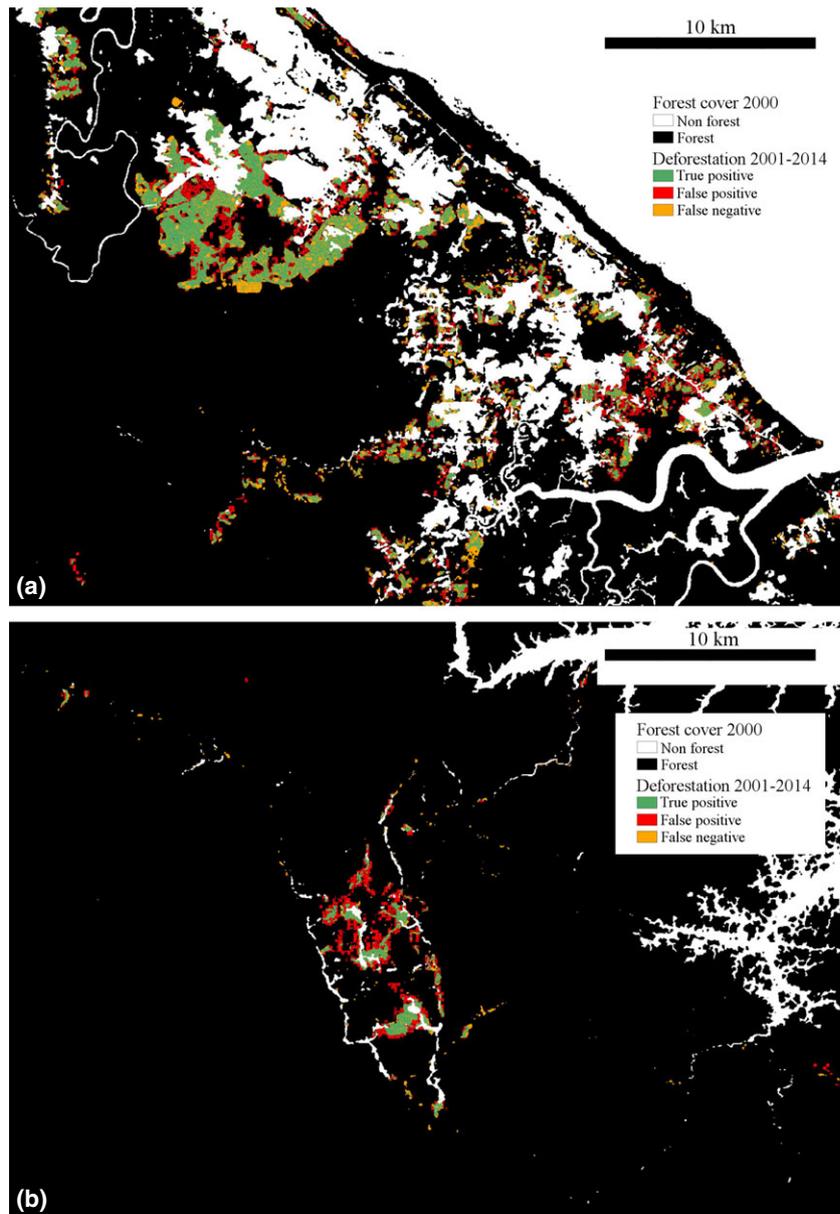


Fig. 7 Visual comparison of predicted vs. observed deforestation map over the period 2001–2014 for an agriculture and urbanization hot spot (a) and a gold mining hotspot (b). Correct predicted pixels are displayed in green, red areas are false positive, and orange pixels are false negative.

Table 4 Values of some common pixel-based performance indices of coupled and null models: overall accuracy (OA), figure of merit (FOM), sensitivity (Sen), specificity (Spe) and Cohen's kappa (K)

Model	OA	FOM	Sen	Spe	K
Coupled	99.5	31.6	47.8	99.8	47.8
Null	99.6	35.6	53.6	99.8	52.8

The null model outperforms the coupled model based on FOM, Sen and K criteria, whereas OA and Spe are biased by the low level of deforestation.

suggested that the null model overestimates high deforestation areas compared to the coupled model, except for in very small grid cell sizes, under 2-km resolution.

A focus on lower deforestation cells indicated an opposite effect, with underestimation of deforestation in the null model compared to the coupled model (Fig. 9).

Discussion

Enriching deforestation models by including expert knowledge: a case study in French Guiana

When focusing on socio-economic drivers of deforestation intensity, it appears necessary to create a typology of zones within the area of interest, corresponding to major differences in socio-economic functioning. To do

so, expert knowledge is indispensable and such zoning must be meaningful. This work can only be carried out in countries where a good level of governance is assumed. On the contrary, the spatial zoning of the different drivers of the intensity model would remain uncertain. A clarification of land rights and increased law enforcement in forested countries, however, are often considered to be a prerequisite to an efficient implementation of the REDD+ mechanism (Karsenty & Ongolo, 2012). As a consequence, we may expect that methodology such as that developed here could be applied to an increasing number of countries.

This necessity to create a zoning *a priori* is likely the main limit of our methodology, especially around the borders between two zones, because any leakage from agricultural areas to the PFE, for example, would be erroneously interpreted as increased deforestation due to logging. Another example of the limits imposed by the zoning is that we cannot predict deforestation due to gold mining outside the rough areas that we know have been previously impacted based on the gold-mining mask used. Emergence of new gold mining areas is thus impossible to model in our case, but it is the consequence of numerous factors, such as discovery of new deposits or depletion of old ones, location and intensity of military interventions against illegal mining, and new mining authorizations. These factors are *per se* difficult to estimate, so it is reasonable not to try to provide maps of emerging deforestation due to gold-mining without a strong theoretical basis.

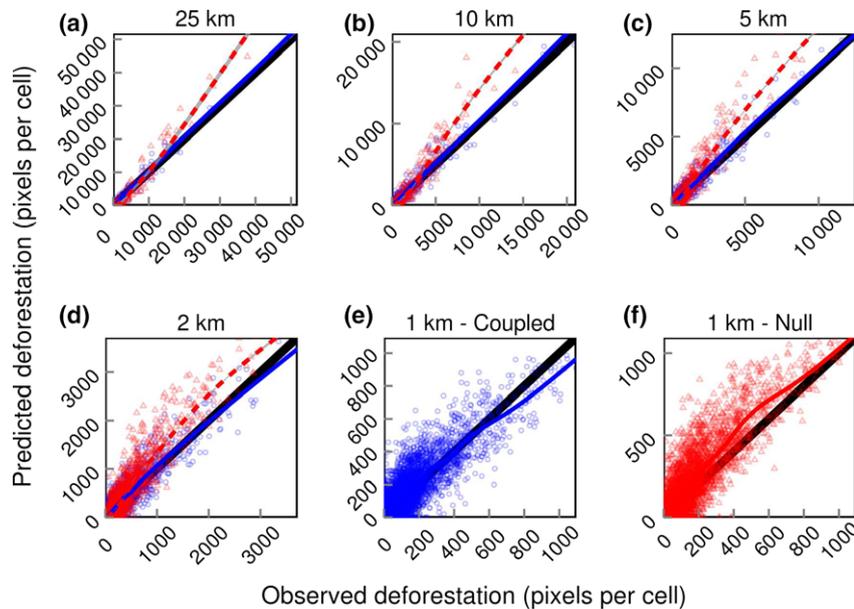


Fig. 8 Plots of predicted vs. observed deforestation (number of predicted/observed pixels per grid cell) summarized at different spatial resolutions, at 25 km (a), 10 km (b), 5 km (c), 2 km (d), and 1 km (e and f). Blue circles and red triangles correspond to the coupled and null models, respectively, and are associated with local regression curves of the same colour. The diagonal corresponding to a perfect prediction is represented by a black line in each plot.

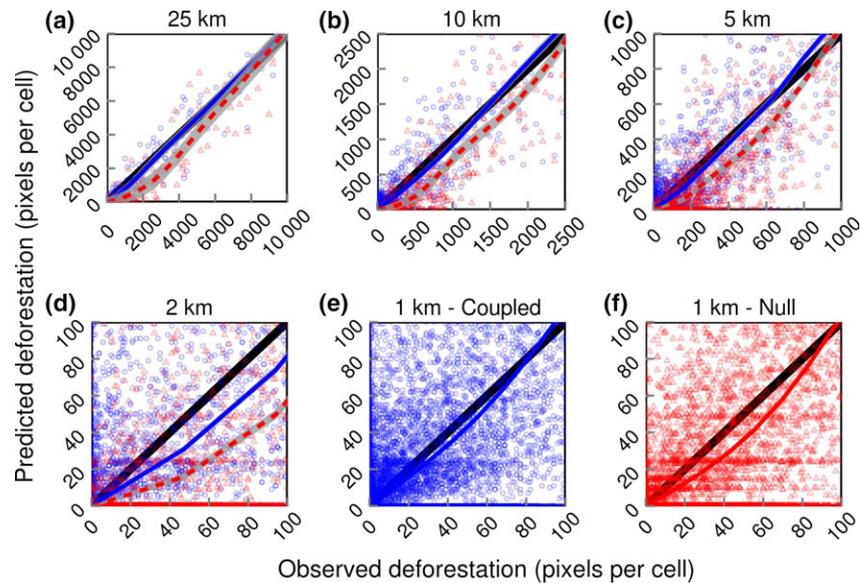


Fig. 9 Plots of predicted vs. observed deforestation (number of predicted/observed pixels per grid cell) summarized at different spatial resolutions, at 25 km (a), 10 km (b), 5 km (c), 2 km (d), and 1 km (e and f), with a focus on low deforestation areas. Blue circles and red triangles correspond to the coupled and null models, respectively, and are associated with local regression curves of the same colour. The diagonal corresponding to a perfect prediction is represented by a black line in each plot.

Gold mining areas. The negative correlation we observed between gold production and gold prices is counter-intuitive compared to the results provided by Hammond *et al.* (2007). Although we may wonder if estimated gold production is accurate due to the importance of illegal gold mining activities, this trend may be influenced by the French military interventions against illegal gold-miners, which were initiated during the period of interest (De Rohan *et al.*, 2011), explaining why annual deforestation due to gold mining decreased during the period whilst gold prices were sharply increasing. This trend emphasizes the need to assess deforestation at a larger scale to test the hypothesis of deforestation leakages. Gold mining activities could have been exported to neighbouring countries (e.g. Suriname) because of the repressive context in French Guiana, as occurred previously between Brazil and French Guiana when the Brazilian government began to regulate gold mining activity (WWF Guianas *et al.*, 2012).

The advantage of defining different intensity sub-models is clear regarding the impact of gold mining, where price volatility is high and susceptible to provoking changes in gold production. Such a flexible method allows the consideration of different human activities responding differently in terms of scale and time, as in the case of gold mining, which is a major improvement, given the contribution of commodities to deforestation trends (Angelsen & Kaimowitz, 2001). On the contrary, using a purely spatial model, important drivers of

deforestation such as gold mining would be inconsistently taken into account, because gold price is global and cannot be spatialized at small scales.

Deforestation due to logging within the Permanent Forest Estate. We estimated that approximately 1.8 ha were deforested for each thousand cubic metres of wood harvested. This represents 1302 ± 196 ha, or $40.3 \pm 6.0\%$, of total predicted deforestation in the PFE, with the remaining being explained by the size effect (i.e. noise in the input deforestation maps). After rasterization of the road layer, we estimated that approximately 500 km of new forest tracks were built within the PFE during the period of interest. Assuming an average track width of 30 m (i.e. one detected deforested Landsat pixel), this would correspond to approximately 1500 ha of deforestation directly due to main logging tracks. This is close to our model's estimates, although in our model forest gaps were not taken into account as well as log yards. Kleinschroth *et al.* (2016) reported a road cover of 0.76% inside selective logging concessions in Central Africa. Given the 52 000 ha of forest logged over the period of interest, this would represent an area of only 395 ha, much less than observed in the present study. Miller *et al.* (2011), on two study sites in Tapajós (Amazonia), found that reduced impact logging techniques, which are also applied in French Guiana (ONF, 2016), provoked a decrease in canopy cover from 96% to 88%. In our case, given the area logged during the period of interest, this would represent around 4000 ha

of lost canopy cover, in addition to areas directly impacted by forest tracks opening. It is unlikely that such logging gaps can be reliably detected in the data set provided by Hansen *et al.* (2013), however, as selective logging is considered to be only marginally detectable using specialized detection algorithms, such as those used by Peres *et al.* (2006), whereas the data set used here has not been calibrated for this purpose. The marginal impact of logging in terms of deforestation is therefore low but statistically significant, although likely underestimated due to data resolution constraints. Given that forest tracks are likely to be the only measurable impact of selective logging and that forest track planning is uncertain, it is likely more robust to predict deforestation due to logging at the scale of the forest sectors relying on explicit socio-economic variables, rather than to refer to high-resolution predicted deforestation maps.

Deforestation in agriculture and urbanization areas. Total predicted deforestation in agriculture and urbanization areas districts was approximately 26 700 ha between 2001 and 2014. Of this amount, one-third corresponded to the contribution of agricultural subsidies, whereas 17 800 ha, or around two-thirds of predicted deforestation, was explained by population growth. Agricultural subsidies provided for land clearing reached an amount of 3200€ per hectare planned to be converted to agriculture. Given the 11M€ provided during the period 2001–2014, this would explain the deforestation of approximately 3700 ha. Our estimation suggested an effect 2.5 times stronger. The effect of the type of subsidies used here (subsidies for land clearing) could have been overestimated due to the absence of other possible types of subsidies (e.g. financial support to agriculture mechanization) in our model. The large standard deviation of the random slope (Table 2) associated with population growth indicates a great heterogeneity of the effect of demography on deforestation amongst agriculture and urbanization areas. The predicted contribution of each additional inhabitant was very different between the densely populated capital, Cayenne, with a value of only 0.009 ha deforested per year per additional inhabitant, and some rural districts where predicted deforestation reaches 1 ha per year per additional inhabitant. This situation likely derives from the coexistence of an officially declared agriculture, the contribution of which to deforestation is captured by the agricultural subsidies variable, with an expanding informal agriculture captured by population growth.

The link between deforestation and population growth has long been debated, often focusing on the impact of shifting cultivation on deforestation in developing countries (Jarosz, 1993). Mixing geographical and

socio-economic variables at very different resolutions in deforestation models has contributed to hiding the effect of demography, in addition to the inherent complexity of taking into account the diversity of consequences of population growth following local socio-economic and political contexts (Geist & Lambin, 2002). Rapid population expansion in developing forested countries, however, is now known to be a clear driver of deforestation (Vieilledent *et al.*, 2013). A model comparison with other rapidly-expanding countries in terms of demography is not easy due to our uncommon use of absolute measures of deforestation and population growth, but the results of the present study confirm this conclusion of a positive correlation between population growth and deforestation such as observed in previous national or global studies (Pahari & Murai, 1999; Agarwal *et al.*, 2005; Vieilledent *et al.*, 2013). Even so, much effort should be directed towards the explication of the random effect associated with population growth, to understand which local forces might influence the way of life of future populations in the perspective of deforestation scenario formulation. Although the amount of agricultural subsidies is a significant exogenous contributor to deforestation, French agricultural policy in French Guiana is *de facto* a response to the high current population increase. Limiting the level of economic incentives to agricultural expansion to decrease deforestation may then create a lack of job opportunities for local populations, which could have unclear environmental consequences, although youngsters in remote districts are likely to be more interested in an urban way of life than in practicing shifting cultivation or commercial agriculture.

Towards socio-economically focused deforestation scenarios

Our coupled model succeeds in predicting the spatial patterns of deforestation, whilst allowing the expression of significant socio-economic variables, which determine its intensity. Compared to the null model, the coupled model is less biased by overestimating high deforestation and underestimating low deforestation intensities (Figs 8 and 9), suggesting a more consistent integration of socio-economic processes. Overestimation of predicted deforestation around correctly predicted deforested pixels was previously mentioned by Chowdhury (2006) and is explained by the strong contribution of several highly important geographical variables, mainly distance to roads, which focuses much deforestation around areas of high deforestation risk, with socio-economic activities occurring in remote areas remaining largely underestimated. This bias explains why accuracy indices are higher for the null

model compared to the coupled model, which evidences the need to assess the accuracy of models not only at the pixel scale, but also at a wider scale, and reinforces the necessity to distinguish the spatial from the intensity processes, as mentioned by Vieilledent *et al.* (2013).

In addition to their knowledge-intensive character, we argue that models based on an explicit socio-economically focused intensity component provide more interesting outputs regarding socio-economic drivers of deforestation. Moving forward spatially explicit socio-economic models of deforestation is difficult because it implies a strong understanding of the different drivers of deforestation within a country or region, and requires a large amount of data. It is critical, however, to formulate useful future deforestation scenarios in a REDD+ perspective. On the contrary, building a model using principally geographical drivers ensures a good fit, even if human activities are not properly understood or taken into account, because of the strength of road accessibility in shaping the development of a territory. Much effort must thus be directed towards a better integration of global and local forces of deforestation which, in our opinion, must be based on a strong knowledge of the drivers of deforestation in the areas considered. By operating a zoning of the territory based on known geographical and normative constraints, we think that such a shift in deforestation modelling methodologies is possible and is consistent with a higher level of law enforcement, environmental governance and land tenure security, factors which are prerequisites of the REDD+ mechanism.

Acknowledgements

We thank the handling editor and the anonymous reviewers for their comments on a previous version of this manuscript. This study is part of a PhD project funded by the French Ministry of Research and associated with the REDD+ for the Guiana Shield project, itself granted by the FFEM, the Région Guyane, the European Union and the Interreg Caraïbes Program and managed by ONF International. It was also part of the GFclim project funded by the PO-Feder Région Guyane. Finally, this work also benefited from an 'Investissement d'Avenir' Grant managed by the Agence Nationale de la Recherche (CEBA: ANR-10-LABEX-0025).

Conflict of interest

Authors confirm having no conflict of interest.

References

- Agarwal DK, Silander JA, Gelfand AE, Dewar RE, Mickelson JG (2005) Tropical deforestation in Madagascar: analysis using hierarchical, spatially explicit, Bayesian regression models. *Ecological Modelling*, **185**, 105–131.
- Aguiar APD, Vieira ICG, Assis TO *et al.* (2016) Land use change emission scenarios: anticipating a forest transition process in the Brazilian Amazon. *Global Change Biology*, **22**, 1821–1840.
- Alvarez-Berrios NL, Mitchell Aide (2015) Global demand for gold is another threat for tropical forests. *Environmental Research Letters*, **10**, 14006.
- Angelsen A, Kaimowitz D (1999) Rethinking the causes of deforestation: lessons from economic models. *The World Bank Research Observer*, **14**, 73–98.
- Angelsen A, Kaimowitz D (2001) *Agricultural Technologies and Deforestation*. CABI Publishing, Oxon, UK.
- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many?. *Methods in Ecology and Evolution*, **3**, 327–338.
- Barrientos M, Soria C (2012) French Guiana Gold Production by Year. Available at: <http://www.indexmundi.com/minerals/?country=gf&product=gold&graph=production> (accessed August 2016).
- Barton K (2015) MuMIn: Multi-Model Inference. R package version 1.13.4. Available at: <http://CRAN.R-project.org/package=MuMIn> (accessed 01 August 2016).
- Bax V, Francesconi W, Quintero M (2016) Spatial modeling of deforestation processes in the Central Peruvian Amazon. *Journal for Nature Conservation*, **29**, 79–88.
- Breiman L (2001) Random forests. *Machine Learning*, **45**, 5–32.
- BRGM (2014) Geological map of French Guiana. Available at: <http://gisguyane.brgm.fr/> (accessed 01 August 2016).
- Brown K, Pearce DW (1994) *The Causes of Tropical Deforestation: The Economic and Statistical Analysis of Factors giving Rise to the Loss of the Tropical Forests*. UBC Press, Vancouver.
- Brown DG, Verburg PH, Pontius RG, Lange MD (2013) Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Environmental Sustainability*, **5**, 452–457.
- Brown DG, Band L, Green K *et al.* (2014) *Advancing Land Change Modeling*. National Academies Press, Washington, DC.
- Calle ML, Urrea V (2011) Letter to the editor: stability of Random Forest importance measures. *Briefings in Bioinformatics*, **12**, 86–89.
- Cassard D, Billa M, Lambert A, Picot JC, Husson Y, Lasserre JL, Delor C (2008) Gold predictivity mapping in French Guiana using an expert-guided data-driven approach based on a regional-scale GIS. *Ore Geology Reviews*, **34**, 471–500.
- Chowdhury RR (2006) Driving forces of tropical deforestation: the role of remote sensing and spatial models. *Singapore Journal of Tropical Geography*, **27**, 82–101.
- DAAF (2015) Subventions à la défriche agricole en Guyane française entre 2000 et 2014. Available at: <http://daf.guyane.agriculture.gouv.fr/> (accessed 01 August 2016).
- Dalla-Nora EL, de Aguiar APD, Lapola DM, Woltjer G (2014) Why have land use change models for the Amazon failed to capture the amount of deforestation over the last decade? *Land Use Policy*, **39**, 403–411.
- De Pinto A, Nelson GC (2007) Modelling deforestation and land-use change: sparse data environments. *Journal of Agricultural Economics*, **58**, 502–516.
- De Rohan J, Dupont B, Berthou J, Antoinette J-E (2011) La Guyane: une approche globale de la sécurité. Available at: <http://www.senat.fr/rap/r10-271/r10-2710.html> (accessed 01 August 2016).
- DEAL Guyane (2013a) Autorisation d'exploitation minière (AEX) de Guyane. DEAL Guyane, Cayenne, French Guiana. Available at: <http://www.geoguyane.fr/geonetwork/srv/fr/finder?uuiid=b62879ad-b32e-4919-ac9d-6a2d79f62718> (accessed 01 August 2016).
- DEAL Guyane (2013b) Atlas des Zones Inondables (2005). DEAL Guyane, Cayenne, French Guiana. Available at: <http://www.geoguyane.fr/geonetwork/srv/fr/finder?uuiid=ce91146b-cdcb-4331-8abb-567b109a9145> (accessed 01 August 2016).
- DEAL Guyane (2015) Réserve Naturelle Nationale de Guyane. DEAL Guyane, Cayenne, French Guiana. Available at: <http://www.geoguyane.fr/geonetwork/srv/fr/finder?uuiid=e6499eb3-13d4-4534-be84-a4fda1902e54> (accessed 01 August 2016).
- Debarros G, Joubert P (2010) *Impact de l'activité aurifère sur le plateau des Guyanes. Rapport final*. ONF – Direction Régionale de Guyane, Cayenne, French Guiana. Available at: http://d2ouvy59p0dg6k.cloudfront.net/downloads/2010_etude_emprise_orpailage_3_guyanes_finale.pdf (accessed 01 August 2016).
- Dietterich TG (2000) An experimental comparison of three methods for constructing ensembles of decision trees. *Machine Learning*, **40**, 139–157.
- Evans JS, Murphy MA, Holden ZA, Cushman SA (2011) Modeling species distribution and change using random forest. In: *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications* (ed. Drew CA, *et al.*), pp. 139–159. Springer Science+Business Media, LLC, New York.
- Geist HJ, Lambin EF (2002) Proximate causes and underlying driving forces of tropical deforestation. *BioScience*, **52**, 143.
- 24hGold.com (2004) French Guyana: history of gold production. Available at: http://www.24hgold.com/english/stat_country_detail.aspx?pays=French+Guyana&deid=24470B1670 (accessed August 2016).

- GRASS Development Team (2015) Geographic Resources Analysis Support System (GRASS) Software, Version 7.0. Open Source Geospatial Foundation. Available at: <http://grass.osgeo.org> (accessed 01 August 2016).
- Hammond DS (2005) *Tropical Forests of the Guiana Shield: Ancient Forests in a Modern World*. CABI Publishing, Cambridge.
- Hammond DS, Gond V, de Thoisy B, Forget P-M, DeDijn BPE (2007) Causes and consequences of a tropical forest gold rush in the Guiana Shield, South America. *Ambio*, **36**, 661–670.
- Hansen MC, Potapov PV, Moore R *et al.* (2013) High-resolution global maps of 21st-century forest cover change. *Science*, **342**, 850–853.
- Hecquet V, Moriame E (2008) Guyane : un développement sous contraintes. Available at: http://www.insee.fr/fr/themes/document.asp?reg_id=25&ref_id=14143 (accessed August 2016).
- Horton RE (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, **56**, 275–370.
- INSEE (2016) Populations légales communales depuis 1968. Available at: <https://www.insee.fr/fr/statistiques/2522602> (accessed 01 August 2016).
- Irwin E, Geoghegan J (2001) Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment*, **85**, 7–24.
- Jarosz L (1993) Defining and explaining tropical deforestation: shifting cultivation and population growth in colonial Madagascar (1896–1940). *Economic Geography*, **69**, 366–379.
- Joubert P (2015) *Carte des vocations 2013. Parc Amazonien de Guyane, Cayenne, French Guiana*. Available at: <http://www.geoguyane.fr/geonetwork/srv/fre/find?uuid=e94ec3cc-8f9d-4efc-b75d-22aa22a8eeb5> (accessed 01 August 2016).
- Karsenty A, Ongolo S (2012) Can “fragile states” decide to reduce their deforestation? The inappropriate use of the theory of incentives with respect to the REDD mechanism. *Forest Policy and Economics*, **18**, 38–45.
- Kim OS (2010) An assessment of deforestation models for reducing emissions from deforestation and forest degradation (REDD). *Transactions in GIS*, **14**, 631–654.
- Kleinschroth F, Healey JR, Sist P, Mortier F, Gourlet-Fleury S (2016) How persistent are the impacts of logging roads on Central African forest vegetation? *Journal of Applied Ecology*, **53**, 1127–1137.
- Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News*, **2**, 18–22.
- Liu C, White M, Newell G (2011) Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography (Cop.)*, **34**, 232–243.
- Mahapatra K, Kant S (2005) Tropical deforestation: a multinomial logistic model and some country-specific policy prescriptions. *Forest Policy and Economics*, **7**, 1–24.
- Mas JF, Paegelow M, De Jong B *et al.* (2007) *Modelling tropical deforestation: a comparison of approaches*. 32rd Symposium on Remote Sensing of Environment, June 2007, San Jose, Costa Rica.
- Mather PM (2004) *Computer Processing of Remotely-Sensed Images: An Introduction*. John Wiley & Sons, Hoboken, NJ, USA.
- Miller SD, Goulden ML, Hutrya LR *et al.* (2011) Reduced impact logging minimally alters tropical rainforest carbon and energy exchange. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 19431–19435.
- ONF (2014) Réseau routier de la Guyane. ONF Guyane, Cayenne, French Guiana. Available at: <http://www.onf.fr/guyane/@index.html> (accessed 01 August 2016).
- ONF (2016) *Charte de l'exploitation forestière à faible impact en Guyane*. ONF Guyane, Cayenne, French Guiana, pp.1–97. Available at: http://www.pefcfrance.org/media/3_standards_guyane_charte_efi_version_finale_fevrier_2016_valid_par_age_pefc_france_21-06-2016.pdf (accessed 01 August 2016).
- ONF Guyane (2012) Coupes au sein du DFP – 1974–2012 ONF Guyane, Cayenne, French Guiana. Available at: <http://www.onf.fr/guyane/@index.html> (accessed 01 August 2016).
- Overmars KP, De KG, Veldkamp A (2003) Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, **164**, 257–270.
- Pahari K, Murai S (1999) Modelling for prediction of global deforestation based on the growth of human population. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**, 317–324.
- Peres CA, Barlow J, Laurance WF (2006) Detecting anthropogenic disturbance in tropical forests. *Trends in Ecology & Evolution*, **21**, 227–229.
- Pontius RG, Millones M (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, **32**, 4407–4429.
- Pontius RG Jr, Cornell JD, Hall CA (2001) Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. *Agriculture, Ecosystems & Environment*, **85**, 191–203.
- Pontius RG, Boersma W, Castella JC *et al.* (2008) Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, **42**, 11–37.
- R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org/> (accessed 01 August 2016).
- Rahm M, Jullian B, Lauger A *et al.* (2015) Monitoring the impact of gold mining on the forest cover and freshwater in the Guiana Shield, 1–60.
- Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **67**, 93–104.
- Rossi V, Dolley T, Cornu G, Guitet S, Hérault B (2015) GuyaSim : un outil d'aide à la décision pour l'aménagement d'un territoire forestier, la Guyane. *Bois Et Forêts Des Tropiques*, **326**, 67–78.
- Soares-Filho BS, Coutinho Cerqueira G, Lopes Pennachin C (2002) DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, **154**, 217–235.
- Soares-Filho BS, Nepstad DC, Curran LM *et al.* (2006) Modelling conservation in the Amazon basin. *Nature*, **440**, 520–523.
- Strahler AN (1952) Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, **63**, 1117–1142.
- Tejada G, Dalla-nora E, Cordoba D, Lafortezza R, Ovando A, Assis T, Paula A (2016) Deforestation scenarios for the Bolivian lowlands. *Environmental Research*, **144**, 49–63.
- USGS (2000) Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global. Distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, SD. Available at: <http://lpdaac.usgs.gov> (accessed 01 August 2016).
- Verbrug PH, Veldkamp WSA, Victoria Espaldon RL, Mastura SSA (2002) Modelling the spatial dynamics of regional land use: the CLUE-S model. *Environmental Management*, **30**, 391–405.
- Verger G (2011) *Villes, villages et lieux dits 500e*. ONF Guyane, Cayenne, French Guiana.
- Verger G (2014a) Drivers of deforestation in the Guiana Shield – French Guiana. REDD+ Workshop. Macapa – 08/20/2014, 1–19 Available at: https://reddguyana.shield.files.wordpress.com/2014/08/wg4_s1_french_guiana.pdf (accessed 01 August 2016).
- Verger G (2014b) *Forêts aménagées de l'ONF*. ONF Guyane, Cayenne, French Guiana.
- Vieilledent G, Grinand C, Vaudry R (2013) Forecasting deforestation and carbon emissions in tropical developing countries facing demographic expansion: a case study in Madagascar. *Ecology and Evolution*, **3**, 1702–1716.
- WWF Guianas (2012) *Living Guianas Report 2012*. WWF Guianas, Paramaribo, Suriname.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Forest cover 2000 and deforestation 2001–2014 map of French Guiana.

Figure S2. Overview of the location of the main activities leading to deforestation.

Table S1. Confusion matrix as input for assessing models performance.

Table S2. Definition of the pixel-per-pixel based performance calculated in this study following (Pontius *et al.*, 2008; Liu *et al.*, 2011; Vieilledent *et al.*, 2013).