REDD+ FOR THE GUIANA SHIELD PROJECT

STUDY REPORT

METHODOLOGICAL APPROACH FOR MODELLING DEFORESTATION IN THE GUIANA SHIELD

SEPTEMBER 2015

Camille Dezécache

Abstract

Modelling deforestation involves a combination of multiple scientific fields (Economy, Ecology, Geography, Political Sciences, etc.), spatial scales (micro, regional, national or global) and economic agents interacting with each other in very complex ways. Since decades, hundreds of models have been created to assess this issue. The aim of this paper is not to review exhaustively all the work done, but to propose a concise synthesis of the objectives, strengths or weaknesses of the main deforestation models, based on published reviews and original studies. No national or regional study publicly available specifically focused on the Guiana Shield, but a review of models produced elsewhere may lead to relevant conclusions for that region. The main conclusions of this work are the following:

- no universal drivers of deforestation can be brought out. Instead, a focus on local context (present and historical) is a necessity;
- absolute deforestation (area deforested per year) is the dependent variable that would preferentially be used to build models of deforestation;
- in order to take into account the differences between the local geographic processes influencing the spatial potential of deforestation at small scales and the socio-economic processes leading demand for land at a broader scale, a distinction between factors influencing the location of deforestation and factors influencing its intensity can be made;
- this distinction implies the use of different types of modelling frameworks for considering apart the location of deforestation and its intensity: Random Forest classifier offers good perspectives for the first part, whereas Poisson regression and associated quasi-poisson or negative binomial are of great interest for modelling count data like absolute deforestation.

This study was led with the financial and technical support of the REDD+ for the Guiana Shield project. The REDD+ for the Guiana Shield is a cooperation project between Suriname, Guyana, French Guiana and Amapa State of Brasil. It is funded by the European ERDF/FEDER Interreg Caraibes IV funds, FFEM, ONF and Région Guyane.

Introduction

Deforestation is a complex and multidisciplinary process. Modelling deforestation is an economic issue, as it deals with human activities producing land use changes, as well as an object of interest for geographers, dealing with spatial organization of the territory, or environmental scientists, concerned about the human impact on ecosystems (De Pinto and Nelson, 2007). It also involves political studies, because public policies strongly influence the development of a given region which in turn affects its deforestation rates. Finally, the historical dimension must not be forgotten (Geist and Lambin, 2002), as past political decisions and all kind of events can become long term drivers of development trends and deforestation.

A review by Angelsen and Kaimowitz (1999) introduced different categories of economic models of deforestation based on the scale (household, sub-national or national) and on the methodology used (analytical, simulation or regression models). This classification is general and is applicable for other models that are not explicitly economic. It helps to focus on the types of models which are relevant in the frame of a given study. However, these categories are not independent. The firsts mainly focus on the processes leading to deforestation, and as such concern more frequently household studies where these processes can be analyzed at small scales, whereas the others try in the first place to explicit where and how much deforestation did (regression) / will (simulation) occur, generally at a broader scale (Angelsen and Kaimowitz, 1999), although their conclusions may help to formulate some hypothesis concerning the ultimate drivers of deforestation.

To be consistent with the scale and objectives of the REDD+ Guiana Shield Project, this report will mainly focus on national studies using regression (econometric models) or simulation models. However, sub-national and analytical studies are also taken into account when useful and necessary, as larger scale regression analysis cannot capture the impact of economic agents behaviour and the local heterogeneity in the drivers of deforestation. In a first section, general insights on the drivers of deforestation will be presented and debated. Next, model structuration will be discussed, with a focus on methodological issues which are a consequence of the specificities of the Guiana Shield. Finally, main modelling tools allowing to create a model for the location of deforestation and another for its intensity will be presented.

1 Identifying the drivers of deforestation

1.1 Proximate causes and underlying drivers of deforestation

Identifying the drivers of deforestation is a prerequisite to modelling this process. Two types of potential drivers are generally distinguished: proximate (direct) and underlying causes (indirect) of deforestation (Angelsen and Kaimowitz, 1999; Geist and Lambin, 2002).

Agricultural expansion is an example of proximate cause of deforestation. Agriculture is a land consuming activity: in forested areas, its expansion directly provokes deforestation. Population can be considered as an underlying force of deforestation: its growth may favour an increase in the demand for agricultural products which in turn increases deforestation rates. Forest logging, depending on the context, can be considered as either a direct or indirect deforestation driver (Angelsen and Kaimowitz, 1999): unsustainable logging leads to forest degradation and deforestation; but even sustainable logging, by opening tracks inside the forest, may increase its accessibility and have the same consequences.

Proximate causes are preferred in micro-scale studies, whereas underlying forces are more common in macro-studies where they provide more conclusive results (Kaimowitz and Angelsen, 1998). However, a strict distinction between those two types is still subject to discussions, as some factors can be at a same time proximate or underlying factors depending on the point of view: agricultural subsidies, for example, are an underlying factor of deforestation when they respond to local demand for land. But they could be seen as a direct cause of deforestation when they create a demand for land.

Figure 1 page 4 shows a schematical representation of the links between some drivers and deforestation. Underlying processes lead to deforestation indirectly, through the expression of direct factors, and local potential of deforestation expresses through the effect of socio-economic drivers.

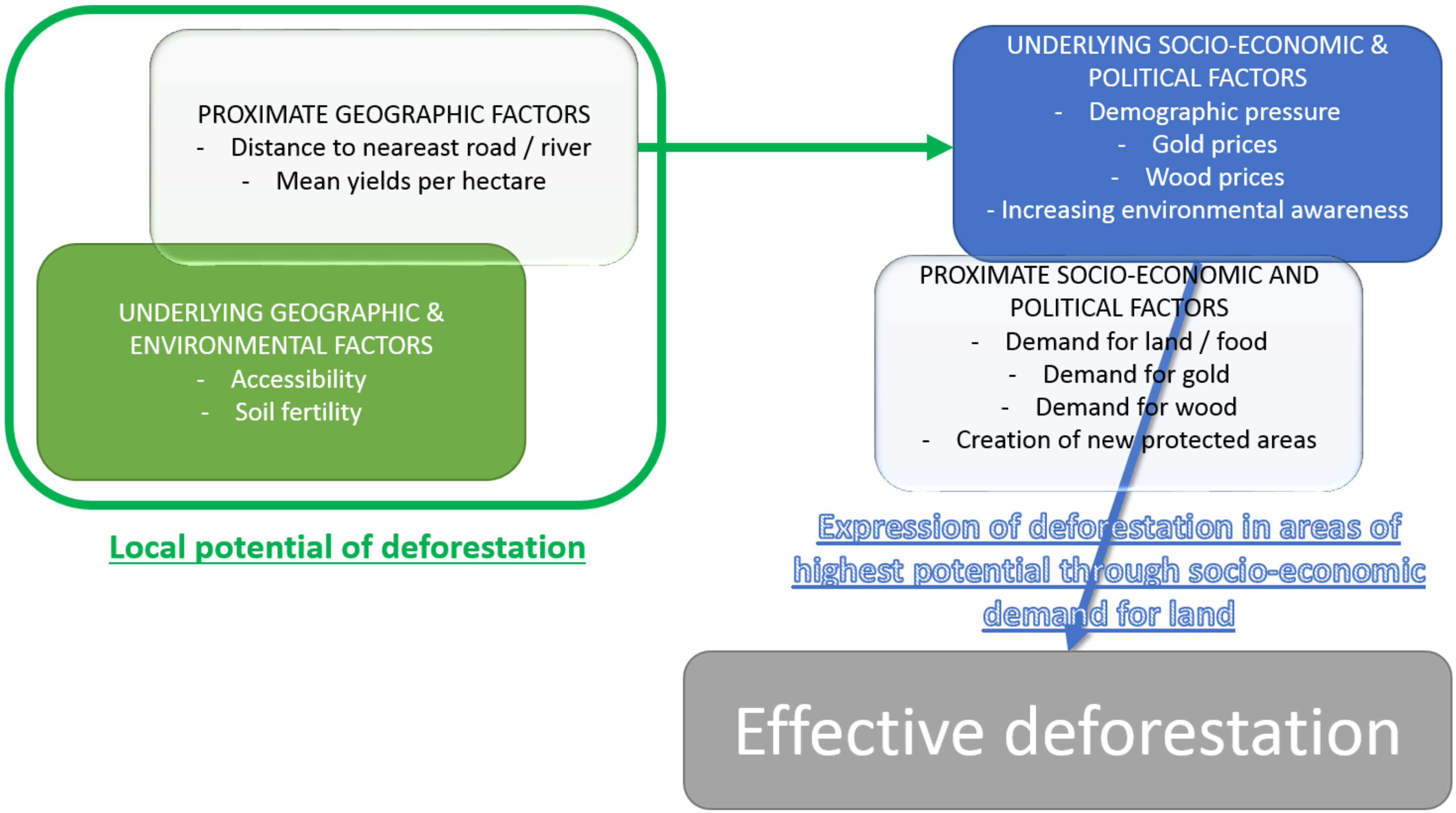


Figure 1: Representation of the links between some drivers and the process of deforestation. Underlying processes lead to deforestation indirectly, through the expression of direct factors, and local potential of deforestation expresses through the effect of socio-economic drivers.

1.2 Defining universal factors of deforestation?

The review of around 150 economic models of deforestation by Kaimowitz and Angelsen (1998) concluded that:

> *Deforestation tends to be greater when: forested lands are more accessible; agricultural and timber prices are higher; rural wages are lower; and there are more opportunities for long distance trade. Population and migration both affect deforestation rates, but in a complex fashion that cannot simply be reduced to saying population growth promotes deforestation. Major doubts remain regarding the relationships between deforestation and productivity growth, input prices, land markets, land and forest tenure security, and household income (poverty) [...].*

Despite much concern on economic variable, other geographic, political or geophysical variables are also used as explanatory variables. Soil suitability for agriculture may also be a relevant deforestation driver, as mentioned by Chomitz and Gray (1996) and Müller et al. (2012): slope, rainfall (excessive or deficient), fertility may favour or prevent agricultural development and deforestation. Besides, political aspects, like the creation of national parks or the recognition of indigenous territories may also impede deforestation (Müller et al., 2012; Soares-Filho et al., 2006; Rosa et al., 2013), although their efficiency is not absolute (Soares-Filho et al., 2006).

However, following Geist and Lambin (2002), no universal causal effect can be identified, most of the drivers of deforestation being region specific.

Mahapatr and Kant (2005) identified multiple drivers of deforestation, each having potential opposite effects on the dependant variable deforestation. Each factor belongs to one of six sectors: forest, demography, macro-economy, agriculture, infrastructure and politics:

- the percentage of forest cover is an indicator of the accessibility of a territory. For a same infrastructure network, the forest ecosystems of a country with low forest cover will face more pressure than in a country with a high remaining forest cover. However, with a high forest cover, a 'free common good attitude' can emerge and lead to less protection and more deforestation;

- population growth could create a Malthusian (increasing pressure on agricultural products create incentives to the expansion of agriculture) or Boserup (more people could enhance creativity and the development of new technologies to face the problem of deforestation) effect in different contexts;

- economic growth could have negative effects on deforestation rates if the hypothesis of the environmental Kuznet's curve is true: poor people would destroy their environment in order to survive, while a growing economy would create off-farm employment and allow availability of capitals for forest protection. However, more investments in remote region could also favour deforestation, especially if the demand for agricultural and forest products is increasing because of the favourable economic situation;

- credits can be used for importing products or developing new technologies in the alternate energy sector thus leading to less deforestation. However, an increasing debt service could also lead to short term decisions forgetting the long term scale which is necessary for protecting the environment;

- agricultural growth could lead to an expansion of agriculture or an intensification of the production, which would have opposite impacts on deforestation;

- road development promotes a better accessibility enhancing deforestation, but may also allow better control, protection and forest management;

- finally, in high level of democracy countries, checks and balance would avoid illegal deforestation, and public pressure may promote environmental protection measures. But in this situation, less fear of punishment could also lead to the opposite observation.

This analysis may explain why defining common explaining variables in a heterogeneous context is so difficult, and why it is necessary to take into account each regional or local context.

1.3 The case of the Guiana Shield

Concerning the Guiana Shield, a review of local models of deforestation is impossible, as no published national or regional study were found. A regional study by Miranda et al. was led but no published work was available. Moreover, this work was done using expensive softwares confirming the need to provide an open source and publicly available model of deforestation for the region. Instead, the case of Amazon deforestation is well documented but the scale of the ongoing processes is then very different than for the Guianas.

The deforestation model presented in the study by Soares-Filho et al. (2006) includes French Guiana, Guyana and Suriname. However, because

> *systematic deforestation map series [were] not available for [...] subregions [outside Brazil] [...], deforestation rates and their annual variation were assigned by applying figures from subregions of Brazil that were considered similar in frontier type and age (see article's supplementary information).*

As we see, given the importance of considering local contexts, it is unlikely that this model would provide consistent predictions for the Guiana Shield. This is particularly true if we consider the specific context of the Guianas, where deforestation rates are among the lowest in the world, whereas annual deforestation in Brazilian Amazon was much higher.

Moreover, in the case of the Guianas, beside more classic forces of deforestation included in many studies as logging, infrastructure building or agricultural expansion, mining (legal or illegal) in forest is a huge and somewhat original deforestation driver (Plouvier et al., 2012). This particular driver may be susceptible to world gold prices (Hammond et al., 2007) and might be influenced by political decisions that could favour or not new exploitation authorizations. A particular attention and specific explanatory variables should probably be used in that case, as no known deforestation model specifically focused on that driver of deforestation. For example, the criteria of accessibility, which is always considered as an important driver of deforestation, might impede instead of encouraging this activity, especially in the case of illegal gold-mining.

2 Model's structuration

2.1 Choosing the appropriate dependent variable

Once potential drivers of deforestation are identified, an appropriate dependent variable still has to be defined before starting the modelling process. In order to measure deforestation, forest cover maps are produced and compared at different periods, relying on an existing forest definition. A 30% threshold is usually considered as an international standard (IGBP, 1992), however we can wonder if this threshold is always appropriate given the huge heterogeneity of forest ecosystems around the world. Calculating forest extent for different forest thresholds between 5 and 99% in French Guiana, we showed that this choice had little impact on total forest cover for values lower than 90%, confirming that for dense tropical rainforests this choice didn't significantly affect results on forested area (see Figure 2 page 9).

In a review of different studies focusing on deforestation by Brown and Pearce (1994), the possible dependent variables are:

- remaining forest cover (percentage or absolute);
- deforestation as a percentage of a region's area;
- absolute deforestation (area of forest cover lost per year).

Choosing the appropriate unit for measuring deforestation is not a trivial issue, although remaining forest cover is necessarily correlated to deforestation.

Current forest cover in a given country (absolute or percentage) depends on the original forest surface (forest cover before human settlement began) and on the history of land use change. Modelling deforestation would allow a comparison between different countries/regions only if the original forest area is the same, and if the history of deforestation began at the same period and followed the same trends. An alternative would be to be able to estimate the original forest cover for each country/region (which could be possible for the Guianas) and to know their respective deforestation process history with low uncertainties. The latter is a more unrealistic assumption. Moreover, the results of these studies are entirely expected, as mentioned by Brown and Pearce (1994): 'very few human settlements are found in forests

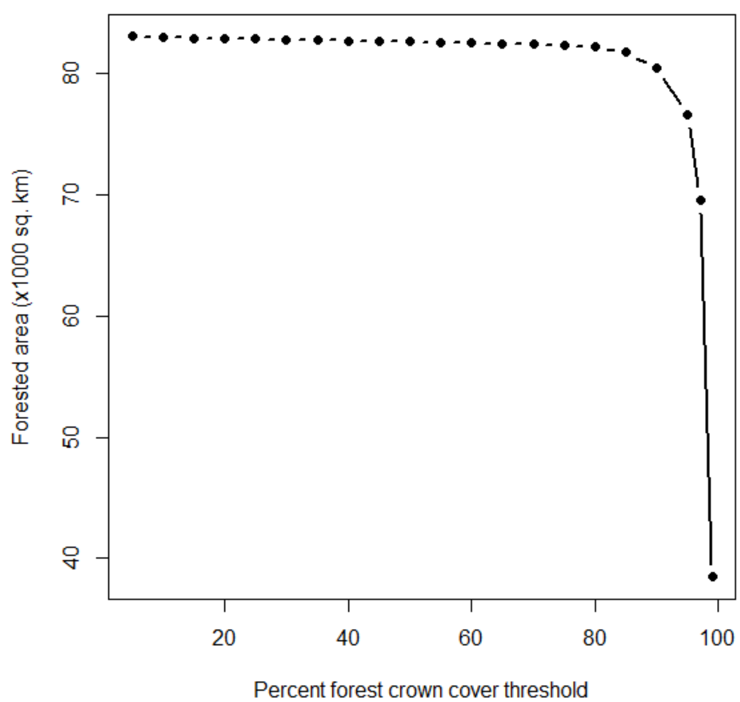


Figure 2: Estimated forested area in thousand square kilometers in function of the forest crown cover threshold defined.

and very few forests are found in human settlements; as such, we would expect that there will always be a negative relationship between forest cover and, for example, roads and population. This is simply logical but it does not say anything about the causes of ongoing deforestation.'

Following Brown and Pearce (1994), only the last one (absolute forest cover loss per year) should be used in deforestation studies, because remaining forest cover and deforestation rates are more biaised by unmeasured geographic and historical factors, compared to a direct measure of absolute deforestation. Today's context explains directly how forest cover loss (deforestation is a function of economic, demographic, political contexts) is changing, but indirectly explains how much forest cover remains (remaining forest cover is a function of original forest cover, current deforestation, and past deforestation which is mainly unknown).

More concretely, using percentages to characterize deforestation in the specific case of the Guiana Shield might be inaccurate just because of the very unequal administrative subdivisions of the different countries. Indeed, taking the example of French Guiana, comparing a percentage of deforestation in Cayenne district(administrative area of around 20 sq.km) and in Maripasoula district (around 18'000 sq.km) would not make any sense. One could object that the same problem might occur with absolute deforestation, because of the impossibility of observing huge deforestation in a small district: absolute deforestation in Cayenne cannot exceed a few square kilometers by definition of the administrative boundaries, whereas in Maripasoula one could imagine that thousands of square kilometers might be deforested. However, it is not unlikely to observe a same urban structure (same types of activities, same population) with very different admnistrative delimitations, as large districts are not large because they have a big potential for deforestation, but because no immediate historical pressure imposed their subdivision. Let's say that a city of 20'000 inhabitants may exist whatever its territory's size, with similar pressures and thus similar absolute deforestation observed, but that by definition their relative deforestation rates will necessarily be different.

Forest scarcity might influence deforestation by increasing land prices or inducing a 'free common good attitude '(Mahapatr and Kant, 2005), but this variable could be included implicitly as an explanatory variable in the model in order to take into account land scarcity while still using absolute deforestation as the dependent variable.

2.2 Defining an appropriate resolution: study scale and technical issues related to sampling

Following Geist and Lambin (2002), no universal driver exist, and different deforestation processes may imply a different scale for deforestation. A recent study on Amazonia (Rosa et al., 2013) used a resolution of 5km x 5km for the variables used in the model. However, using MODIS data at 500m resolution (which would correspond to 100 more pixels for a same area covered) we were unable to detect any deforestation in French Guiana because deforestation occurs at much smaller scales in this region, in particular because of the small demographic pressure and the inexistence of large scale agriculture. LANDSAT images allow to work at 30m resolution and to detect medium to small scale deforestation, but in that case the total number of pixels would be much higher for a same extent, inducing technical problems (calculation time, storage capacity, etc.). Deforestation maps resolution must then be consistent with the processes ocurring in a given area, with a finer resolution needed when deforestation is lower and/or at smaller scales.

When a high resolution is preferred, then technical issues raise: the huge number of pixels makes it impossible to consider all of them in the latter statistical process. To solve this problem, Vieilledent et al. (2013) defined a minimum theoretical sample size necessary to estimate the deforestation rates with less than $\pm 0.1\%$ error per year, given an observed mean deforestation rate in the area of interest (in Madagascar) of 1% per year. However, when deforestation rates are very low, as in the Guiana Shield (rates lower than 0.1% per year), these approximations cannot be used because the sample size would need to be almost as big as the total number of pixels to approach observed deforestation rates with low uncertainties. Moreover, with very low deforestation rates, random sampling pixels would create a huge imbalance between classes: not deforested pixels will be overrepresented compared to deforested ones. In that case, the likelihood of a model would be too much influenced by not deforested pixels, whereas a deforestation model should precisely predict the opposite, deforested pixels. In more concrete terms, a model predicting an absence of deforestation would be evaluated as a good model, because it would predict well the greatly dominant class: no deforestation. A balanced sampling might be preferred in that case in order to penalize errors in deforestation as well as no deforestation classification.

2.3 The need for a distinction between spatial prediction of deforestation and its intensity

Creating a spatial model of deforestation implies calculating probabilities of deforestation for a sample of spatial units (pixels). However, we must question the relevancy of this scale for predicting human induced processes. If human pressure increases in a given district, deforestation will probably follow the same trend, but this deforestation will not occur the same way anywhere. People will settle and create activities where more appropriate, easier or less costly.

As such, the probability of deforestation computed for each pixel could be more precisely interpreted as a potential, a risk of deforestation, or as well as the opportunity cost of settling in a given place compared to another. This being said, we can operate a distinction between factors influencing the location of deforestation and the factors influencing its intensity. This distinction is somewhat artificial because, at local scales, areas not suitable for deforestation (remote areas for example) will suffer low deforestation. But at larger scales (national or regional), deforestation will occur first on more accessible areas, but less favourable zones could be deforested later if the deforestation pressure continues to increase.

Based on this principle, we can make the assumption that spatial distribution of deforestation is driven only by geographic and environmental factors (as well as historical factors that shaped a given region in a particular context, creating existing road network or villages for example), whereas its intensity (demand for land) is driven only by political and socio-economic factors. If local demand for land exists, the more suitable areas will be deforested first. If local pressures continue to increase, then less suitable areas will be deforested. When only poor quality land remains, demand might shift to another area (for example another administrative district, or any administrative boundary) where easily accessible areas still exist. In that case we could imagine that population stops to increase in a given district due to a relocation in neighbouring areas, as a results to its saturation.

3 Modelling methods

3.1 Logistic regression for computing a probability of deforestation

Logistic (or logit) regressions are common in deforestation models (Overmars et al., 2003) and can be very useful for calculating a deforestation risk (i.e. a potential for deforestation) for each pixel of a given area of interest.

Binary logit regression is useful when it deals with predicting the state of a given pixel which can be either deforested or not. In this situation, the response variable (i.e. the dependent variable, here deforestation) can only take two values, 0 and 1, corresponding to the states deforested or not deforested, allowing to calculate the probability π_i of deforestation of a given pixel i. However, the independent variables can take any positive or negative value between $-\infty$ and $+\infty$, whereas the response variable is binary.

To be coherent with the range of values of each dependent and independent variable, the logit transformation was introduced:

$$\eta_i = logit(\pi_i) = \log \frac{\pi_i}{1 - \pi_i}$$

or

$$\pi_i = logit^{-1}(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

where η_i is a linear combination of independent variables (see Figure 3 page 19 to graphically observe the correspondence between the probability and the logit values).

3.2 Random forest algorithm as a pixel classifier

Apart from usual logistic regression, other methods can be used to classify pixels based on their probability of deforestation. Random Forest (Breiman, 2001) is an example of powerful classifiers based on decision trees. It computes a chosen number of trees forming a "forest" of trees. During the construction of each tree, the sample (in this case a certain number of deforested and not deforested pixels) is divided into smaller and smaller sub-samples. The explanatory variables are used to split these sub-samples, forming at each branch, more and more homogeneous groups of pixels. When the tree is completely grown, each pixel is then classified as deforested or not deforested, forming the leaves of the tree. The "random" part of the algorithm is due to the fact that:

- a sub-sample (around 2/3) of the complete sample of pixels is used to build each tree, the remaining part allowing to automatically calculate error rates (called Out-of-bag error, or OOB error) and a confusion matrix;
- at each node, a sub-sample of the complete set of explanatory variables is randomly chosen, to which the best variable for splitting the node is kept.

The main disadvantage of Random Forest compared to usual logit regression is the fact that it is not a variable-explicit method, so there is no mathematical formalism able to explain how the method is classifying the data.

However, this technique has a lot of advantages compared to logistic models:

- the algorithm is very flexible and able to fit very sharp increase / decrease in the probability of deforestation, even in the presence of noise;
- it is able to implicitly take into account very complex sets of interactions between explanatory variables, which is not the case of generalized linear modelling frameworks;
- it is able to display graphically the relationship between each explanatory variable of interest and the dependent variable, in a graph called "partial plot";
- it automatically calculates robust error rates estimates as mentioned above;
- it also calculates indices of variables importance, allowing to have a better understanding of which are the main explanatory variables influencing the model.

In any case, in the absence of unanimously admitted statistical tests for validating the model produced by random forest algorithm, it is even more important to discuss each partial plot and make sure that the relationship between each explanatory variable and the dependent variable (here deforestation) is interpretable.

As an example, figure 4a page 20 shows the relationship between distance to closest area previously deforested and a measure of the probability of deforestation (the log of the ratio between the number of time a pixel is classified as 1 and 0 by the forest of decision trees) for a sample of 100'000 deforested and 100'000 not deforested pixels in French Guiana for the period 2004-2008. When the value of the index on the y-scale is equal to 0, a pixel has a same probability of being classified as deforested or not deforested. When this value is positive, the pixel has a higher probability of being deforested than not deforested and vice-versa. Here we observe a sharp decrease of the probability of deforestation when going further from previously deforested area. This spatial autocorrelation is a major driver of the location of deforestation, this process being contagious.

Figure 4b page 20 shows the relationship between distance to nearest road and the same measure of the probability of deforestation. Here, the pattern is similar than before but less clear. The probability of deforestation decreases less constantly when going further from a road, meaning that some deforestation might still occur far from roads. This is particularly true in very remote areas when settlements focus near rivers, or for illegal gold-mining activities for which inaccessibility is important in order to avoid police intervention.

Finally, figure 4c page 20 shows the dependance plot of variable slope. In that case, the pattern is much more complex and difficult to interpret. Increasing slope for low values between 0 and around 10% would increase the probability of classifying a pixel as deforested. For higher slopes, the probability of deforestation decrease until around 40% and starts to raise again after that threshold. Here the use of this variable to calibrate a model is questionable, because this result could be driven by a complex natural pattern but also by an attempt of the algorithm to fit the data better using a new variable of adjustment, and in a sense over-fitting the data.

3.3 Statistical framework for the intensity model

Once a model predicting the deforestation potential for each pixel has been built, the next step is to compute the intensity of deforestation on a territory or each area of a given territory, based on political or socio-economic variables. This intensity of deforestation will then be applied spatially, selecting a predicted number of pixels to deforest, by decreasing order of deforestation potential (i.e. probability of deforestation derived from logit model or random forst for example).

The intensity model focuses on divisions of a territory considered as relevant for studying a process of interest leading to deforestation, but also taking into account data availability as such scales. A lot of socio-economic data exist concerning administrative divisions of a territory. For agricultural or urban expansion for example, countries districts could be the most relevant scale, as it is generally at this scale that demographic data or information on the economic structure are gathered. For foresty or mining, on the contrary, national policies might prevail on local decisions, obliging to consider these areas as a whole.

Focusing more on the statistical methodology now, the intensity model has to take into account the specificity of the variable of interest itself. As mentionned in 2.1 page 8, considering that absolute deforestation only can be used as a metric of deforestation, its value (unless considering forest regrowth) cannot be negative. Moreover, when absolute deforestation is measured as a sum of deforested pixels, not only its value is always positive but it should also be an integer. This type of count data can be modelled using methods like Poisson regression for example.

However, poisson regression is defined by a parameter λ equal to the mean of the value of the dependent variable (here deforestation) and which is assumed to be equal to its variance. This is a strong assumption, because areas experiencing low deforestation are expected to show low variance, but increasing deforestation might also increase its variance. This can be due, for example, to the fact that if dry season is wetter than expected, deforestation might not be detected or might be delayed, producing an artificially low value of deforestation for a given year, and an artificially high value for the next period. In that case, overdispersion (i.e. increasing variance when predicted value of the dependent variable increases) can be integrated into the model using quasi-poisson or negative binomial regression (Hoef and Boveng, 2007).

Let Y be a poisson variable of parameter λ. By definition its mean $E(Y)$ and variance $var(Y)$ are equal to λ. In the case of quasi-poisson now, $E(Y)$ is still equal to λ, but the variance is proportional to λ: $var(Y) = \theta\lambda$. In negative-binomial models on the contrary, $var(Y) = \lambda + k\lambda^2$ where k is a multiplicative coefficient. When in the first case the variance is directly proportional to the mean, in the case of negative binomial regression, variance is a polynomial function of the mean. Following the existing relationship between mean and variance in the case of heteroskedasticity, one must carefully choose between those two types of modelling frameworks. Using a linear model to fit the relationship between mean and variance and comparing AIC score might contribute to that choice.

An example in French Guiana gives the relationship between observed mean and variance, and the same relationship modelled by a quasi-poisson regression and a negative-binomial regression (Figure 5 page 21). Heteroskedasticy is obvious for the observations: variance increase for increasing mean of deforestation per district. Both models take into account this heteroskedasticity, but in the case of quasi-poisson model, variance is linear to the mean, contrary to negative-binomial regression where variance is quadratic to the mean. Graphical observation cannot easily discriminate the best model in that case, but a comparison of AIC scores gave preference for the negative-binomial case in that example.

Conclusion

After reviewing relevant scientific literature concerning spatially explicit deforestation models, a certain number of choices can be done in order to create an appropriate model of deforestation for the Guiana Shield.

The specificities of this region, in terms of forest cover (dominant forest cover), of drivers (importance of gold-mining), of deforestation scale (mainly small scale deforestation) and intensity (low deforestation) make it essential to adapt a methodology to this particular context.

Using equilibrate samples to cope with low deforestation rates, a powerful classifier like Random Forest (Breiman, 2001) to deal with the complex set of interactions influencing the location of deforestation, and more usual tools like poisson or negative-binomial regressions would provide a simple and intuitive framework to obtain a prediction of the most probable location of future deforestation, and some insights on how political and socio-economic drivers might affect the intensity of deforestation in the years to come.

The main steps of the modelling process are summarized in the flowchart below (Figure 6 page 22). The location and the intensity parts of the model are computed independently, but merge in the final step for combining the predicted intensity of deforestation (corresponding to the demand for land) to its spatial expression.

Making the assumption that spatial processes are static (i.e. spatial drivers of deforestation will not change in the future), it is then possible to focus more on the intensity part of the model to provide scenarios concerning future deforestation trends, taking into account changes in demographic pressure, or socio-economic incentives for deforestation for example.

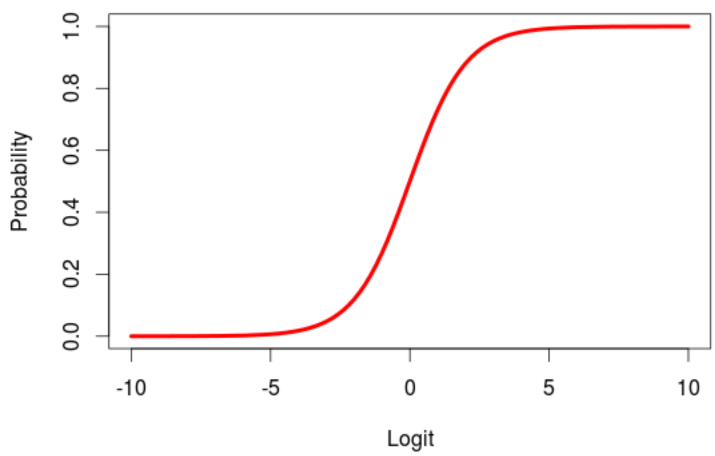
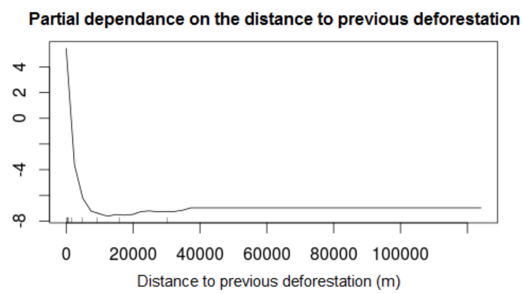
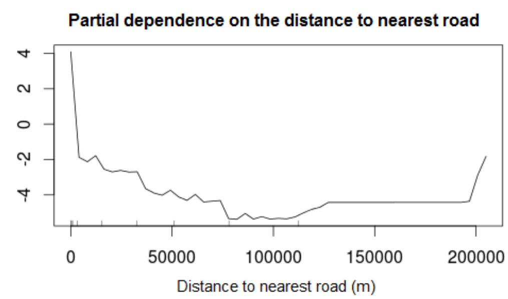


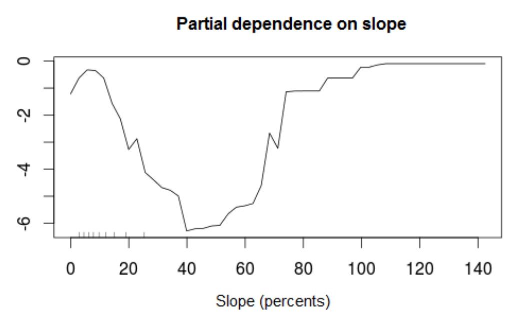
Figure 3: Probability as a function of the logit value



(a)



(b)



(c)

Figure 4: Partial dependence plots between the probability of deforestation (measured as the log of the ratio between the number of times a pixel is classified as deforested versus not-deforested by the forest of decision trees) and distance to previous deforestation in meters (a), distance to nearest road in meters (b) and slope in percent (c)

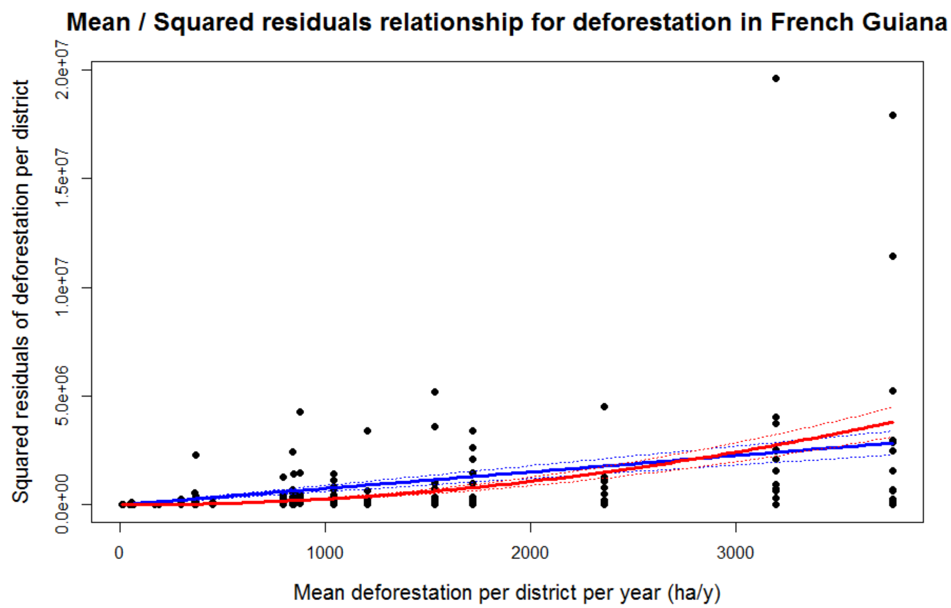


Figure 5: Relationship between observed mean and variance of yearly deforestion in all districts of French Guiana (black points), and linear model predictions for quasi-poisson (blue line) and negative-binomial regressions (red line). 95% confidence intervals are shown with dashed lines for each model. Heteroskedasticy is obvious for the observations. Both models take into account this increase in variance with increasing mean of deforestation. In the case of quasi-poisson model, variance is linear to the mean. For negative-binomial regression, variance is quadratic to the mean.

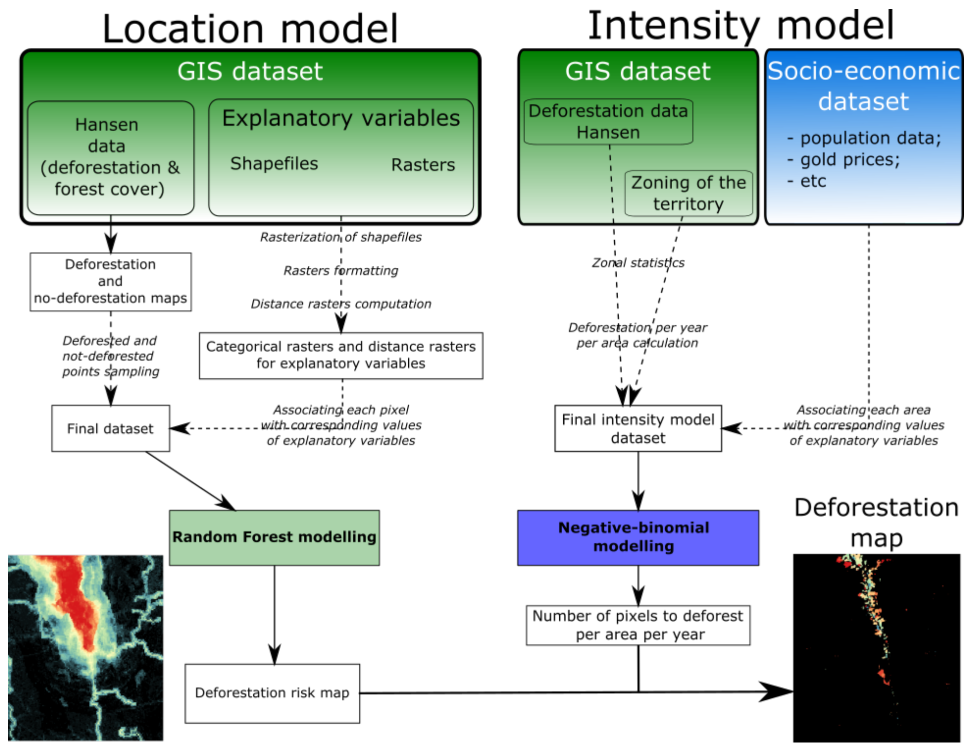


Figure 6: Flowchart of the modelling process. The location and the intensity parts of the model are computed independently, but merge in the final step for combining the predicted intensity of deforestation (corresponding to the demand for land) to its spatial expression.

References

Angelsen, A. and Kaimowitz, D. (1999). Rethinking the causes of deforestation: lessons from economic models. *The World Bank research observer*, 14(1):73–98.

Breiman, L. (2001). Random forests. *Machine learning*, pages 5–32.

Brown, K. and Pearce, D. W. (1994). *The causes of tropical deforestation: the economic and statistical analysis of factors giving rise to the loss of the tropical forests*. UBC Press.

Chomitz, K. M. and Gray, D. a. (1996). Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. *The World Bank Economic Review*, 10(3):487–512.

De Pinto, A. and Nelson, G. C. (2007). Modelling deforestation and land-use change: Sparse data environments. *Journal of Agricultural Economics*, 58(3):502–516.

Geist, H. J. and Lambin, E. F. (2002). Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *BioScience*, 52(2):143.

Hammond, D. S., Gond, V., de Thoisy, B., Forget, P.-M., and DeDijn, B. P. E. (2007). Causes and consequences of a tropical forest gold rush in the Guiana Shield, South America. *Ambio*, 36(8):661–670.

Hoef, J. M. V. and Boveng, P. L. (2007). Quasi-Poisson VS. Negative Binomial Regression: How Should We Model Overdispersed Count Data ?

IGBP (1992). IGBP Improved global data for land applications. *J.R.G. Townshend (Ed.)*.

Kaimowitz, D. and Angelsen, A. (1998). *Economic Models of Tropical Deforestation A Review*, volume 6.

Mahapatr, K. and Kant, S. (2005). Tropical deforestation: A multinomial logistic model and some country-specific policy prescriptions. *Forest Policy and Economics*, 7:1–24.

Müller, R., Müller, D., Schierhorn, F., Gerold, G., and Pacheco, P. (2012). Proximate causes of deforestation in the Bolivian lowlands: An analysis of spatial dynamics. *Regional Environmental Change*, 12:445–459.

Overmars, K. P., De Koning, G. H. J., and Veldkamp, a. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164:257–270.

Plouvier, D., Gomes, L., Verweij, P., and Verlinden, N. (2012). *Living Guianas Report 2012*.

Rosa, I. M. D., Purves, D., Souza, C., and Ewers, R. M. (2013). Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. *PLoS ONE*, 8(10).

Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P., and Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature*, 440(March):520–523.

Vieilledent, G., Grinand, C., and 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.