The Journal of Open Source Software

ForestAtRisk: A Python package for spatial modelling and forecasting of tropical deforestation

Ghislain Vieilledent[1, 2, 3, 4]

1 CIRAD, UMR AMAP, F–34398 Montpellier, France **2** CIRAD, Forêts et Sociétés, F–34398 Montpellier, France. **3** AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France. **4** European Commission, Joint Research Centre, I–21027 Ispra (VA), Italy.

DOI: DOIunavailable

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Reviewers:

- @Pending Reviewers

Submitted: N/A
Published: N/A

Summary

The `ForestAtRisk` Python package can be used to model the tropical deforestation spatially and forecast the future forest cover in the tropics. It includes functions to (i) sample the forest cover change observations and retrieve the information on potential spatial explanatory variables for each observation, (ii) estimate the parameters of various spatial deforestation models, (iii) predict the spatial probability of deforestation, (iv) forecast the likely forest cover in the future, (v) validate the models and the projected maps of forest cover change, and (vi) plot the results. Spatial information is provided through georeferenced raster files which can be very large (several gigabytes). Functions in the `ForestAtRisk` package treat large rasters by blocks of data, making computations fast and efficient. This allows analysis on large geographical extent (e.g. country scale), and at high resolution (e.g. 30 m). The `ForestAtRisk` package offers the possibility to use a logistic regression model with auto-correlated spatial random effects to model the deforestation process. Spatial random effects allow to structure the residual spatial variability (which is often large) within the model. They account for unmeasured or unmeasurable variables that explain a part of the residual spatial variation in the deforestation process that is not explained by the model's explanatory variables. In addition to bringing new features, the `ForestAtRisk` Python package is open source (license GPLv3), cross-platform and scriptable (through Python), user-friendly (functions provided with full documentation and examples), and easily extendable (with additional statistical models for example). The `ForestAtRisk` Python package has been recently used to model the deforestation and forecast the future forest cover by 2100 in the whole humid tropics.

Statement of Need

The `ForestAtRisk` Python package can be used to model the tropical deforestation spatially and forecast the future forest cover in the tropics (Figure 1).

Functions in the `ForestAtRisk` package treat large rasters by blocks of data, making computation fast and efficient (with low memory usage).

Using functions from the `ForestAtRisk` Python package makes computation fast and efficient (with low memory usage) by treating large raster data by blocks. Numerical computations on blocks of data are performed with the NumPy (https://numpy.org) Python module whose core is mostly made of optimized and compiled C code which runs fast (Harris et al., 2020).

Using simple "glm" model is a commonly proposed approach for spatial modelling of deforestation (Eastman & Toledano, 2017; Ludeke et al., 1990; Mas et al., 2007; Mertens & Lambin, 1997; Rosa et al., 2014; B. S. Soares-Filho et al., 2002; Britaldo S. Soares-Filho et al., 2001; Verburg et al., 2002-09-21). The random forest model (Breiman, 2001) is a machine learning approach using an ensemble of random classification trees (where both observations and features are chosen at random to build the classification trees) to predict the deforestation probability for a forest pixel. Random forest has been intensively used for species distribution modelling (Thuiller et al., 2009) and is now also commonly used for spatial modelling of deforestation (Grinand et al., 2020; Santos, 2019; Zanella et al., 2017). The "glm" and "rf" models were fitted using functions `LinearRegression` and `RandomForestClassifier` respectively, both available in the `scikit-learn` Python package (Pedregosa et al., 2011).

Using observations of forest cover change in the period 2010–2020, we modelled the spatial probability of deforestation as a function of the n explanatory variables using a logistic regression. We considered the random variable y_i which takes value 1 if the forest pixel i was deforested in the period 2010–2020 and 0 if it was not. We assumed that y_i follows a Bernoulli distribution of parameter θ_i (Eq. @ref(eq:icar)). In our model, θ_i represents the spatial relative probability of deforestation for pixel i. We assumed that θ_i is linked, through a logit function, to a linear combination of the explanatory variables $X_i\beta$, where X_i is the vector of explanatory variables for pixel i, and β is the vector of effects $[\beta_1, \dots, \beta_n]$ associated to the n variables. All the continuous explanatory variables were normalized before fitting the model. The model includes an intercept α. To account for the residual spatial variation in the deforestation process, we included an additional random effect $\rho_{j(i)}$ for each spatial cell j of a 10 × 10 km grid covering each study-area. This grid resolution was chosen in order to have a reasonable balance between a good representation of the spatial variability of the deforestation process and a limited number of parameters to estimate. A sampled forest pixel i was associated to one cell j and one random effect $\rho_{j(i)}$. We assumed that random effects were spatially autocorrelated through an intrinsic conditional autoregressive (iCAR) model (Banerjee et al., 2014; Besag et al., 1991). This model is denoted "icar" in subsequent sections and results. In an iCAR model, the random effect ρ_j associated to cell j depends on the values of the random effects $\rho_{j'}$ associated to neighbouring cells j'. In our case, the neighbouring cells are connected to the target cell j through a common border or corner (cells defined by the "king move" in chess). The variance of the spatial random effects ρ_j was denoted V_ρ. The number of neighbouring cells for cell j, which might vary, was denoted n_j. Spatial random effects ρ_j account for unmeasured or unmeasurable variables (Clark, 2005) that explain a part of the residual spatial variation in the deforestation process that is not explained by the fixed spatial explanatory variables (X_i).

$$
\begin{aligned}
y_i &\sim \mathcal{B}ernoulli(\theta_i) \\
\text{logit}(\theta_i) &= \alpha + X_i\beta + \rho_{j(i)} \\
\rho_{j(i)} &\sim \mathcal{N}ormal(\sum_{j'} \rho_{j'}/n_j, V_\rho/n_j)
\end{aligned}
\tag{1}
$$

Parameter inference is done in a hierarchical Bayesian framework. Function `model_binomial_iCAR()` from the `forestatrisk` Python package was used for parameter inference. This function calls an adaptive Metropolis-within-Gibbs algorithm (Rosenthal & others, 2011) written in C for maximum computation speed.

Figures

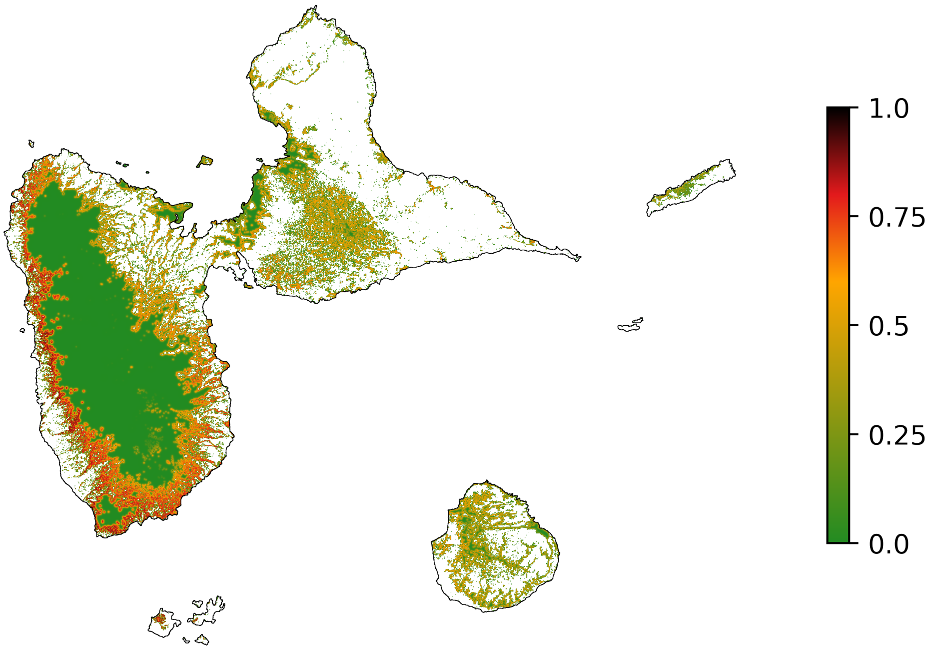

Figure 1: Map of the spatial probability of deforestation in the Guadeloupe archipelago. Colored pixels represent the extent of the natural old-growth tropical moist forest in 2020. A relative spatial probability of deforestation is computed for each forest pixel. The original map has a 30 m resolution. Probability of deforestation is a function of several explanatory variables describing: topography (altitude and slope), accessibility (distances to nearest road, town, and river), forest landscape (distance to forest edge), deforestation history (distance to past deforestation), and land conservation status (presence of a protected area). This map can be reproduced following the Get Started tutorial at https://ecology.ghislainv.fr/forestatrisk.

Acknowledgements

I am grateful to Clovis Grinand, Romuald Vaudry and Matthieu Tiberghien who gave me the opportunity to work on deforestation modeling when we were leading forest conservation projects in Madagascar. I also warmly thank Frédéric Achard and all the members of the IFORCE group for their invaluable support during the first phase of development of the package, while I was seconded to the JRC in Ispra. This work benefited from funding by FRB-FFEM (BioSceneMada project, AAP-SCEN-2013 I), the European Commission (Roadless Forest project), and CNRT (RELIQUES project).

References

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data, second edition*. Chapman; Hall/CRC. https://doi.org/10.1201/b17115

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

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324

Clark, J. S. (2005). Why environmental scientists are becoming bayesians. *Ecology Letters*, *8*(1), 2–14. https://doi.org/10.1111/j.1461-0248.2004.00702.x

Eastman, J. R., & Toledano, J. (2017). A short presentation of the land change modeler (LCM). In *Geomatic approaches for modeling land change scenarios* (pp. 499–505). Springer International Publishing. https://doi.org/10.1007/978-3-319-60801-3_36

Grinand, C., Vieilledent, G., Razafimbelo, T., Rakotoarijaona, J.-R., Nourtier, M., & Bernoux, M. (2020). Landscape-scale spatial modelling of deforestation, land degradation, and regeneration using machine learning tools. *Land Degradation & Development*, *31*(13), 1699–1712. https://doi.org/10.1002/ldr.3526

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Ludeke, A. K., Maggio, R. C., & Reid, L. M. (1990). An analysis of anthropogenic deforestation using logistic regression and GIS. *Journal of Environmental Management*, *31*(3), 247–259. http://www.sciencedirect.com/science/article/pii/S0301479705800386

Mas, J.-F., Paegelow, M., Jong, B. de, Guerrero, G., Follador, M., Olgien, M., Díaz, J. R., Castillo, M. A., & García, T. (2007). Modelling tropical deforestation: A comparison of approaches. *Symposium Remote Sensing of Environment*, 3.

Mertens, B., & Lambin, E. F. (1997). Spatial modelling of deforestation in southern cameroon: Spatial disaggregation of diverse deforestation processes. *Applied Geography*, *17*(2), 143–162. http://www.sciencedirect.com/science/article/pii/S0143622897000325

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rosa, I. M., Purves, D., Carreiras, J. M., & Ewers, R. M. (2014). Modelling land cover change in the brazilian amazon: Temporal changes in drivers and calibration issues. *Regional Environmental Change*, 1–15.

Rosenthal, J. S., & others. (2011). Optimal proposal distributions and adaptive MCMC. *Handbook of Markov Chain Monte Carlo*, *4*(10.1201).

Santos, V. A. B., Fabián AND Graw. (2019). A geographically weighted random forest approach for evaluate forest change drivers in the northern ecuadorian amazon. *PLOS ONE*, *14*(12), 1–37. https://doi.org/10.1371/journal.pone.0226224

Soares-Filho, Britaldo S., Assunção, R. M., & Pantuzzo, A. E. (2001). Modeling the Spatial Transition Probabilities of Landscape Dynamics in an Amazonian Colonization Frontier: Transition probability maps indicate where changes may occur in the landscape, thus enabling better evaluation of the ecological consequences of landscape evolution. *BioScience*, *51*(12), 1059–1067. https://doi.org/10.1641/0006-3568(2001)051%5B1059:MTSTPO%5D2.0.CO;2

Soares-Filho, B. S., Cerqueira, G. C., & Pennachin, C. L. (2002). DINAMICA - a stochastic cellular automata model designed to simulate the landscape dynamics in an amazonian colonization frontier. *Ecological Modelling*, *154*(3), 217–235. https://doi.org/10.1016/S0304-3800(02)00059-5

Thuiller, W., Lafourcade, B., Engler, R., & Araujo, M. B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, *32*(3), 369–373.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., & Mastura, S. S. A. (2002-09-21). Modeling the spatial dynamics of regional land use: The CLUE-s model. *Environmental Management*, *30*(3), 391–405. http://dx.doi.org/10.1007/s00267-002-2630-x

Zanella, L., Folkard, A. M., Blackburn, G. A., & Carvalho, L. M. T. (2017). How well does random forest analysis model deforestation and forest fragmentation in the brazilian atlantic forest? *Environmental and Ecological Statistics*, *24*(4), 529–549. https://doi.org/10.1007/s10651-017-0389-8