forestatrisk: a Python package for modelling and forecasting deforestation in the tropics

Ghislain Vieilledent

The Journal of Open Source Software

ForestAtRisk: A Python package for modelling and forecasting deforestation in the tropics

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 (JRC), I–21027 Ispra (VA), Italy.

DOI: 10.21105/joss.02975

Software

- Review
- Repository
- Archive

Editor: Katy Barnhart

Reviewers:

- @molgor
- @ethanwhite

Submitted: 10 December 2020
Published: 20 January 2021

Summary

The `ForestAtRisk` Python package can be used to model the spatial probability of deforestation and predict future forest cover in the tropics. The spatial data used to model deforestation comes from georeferenced raster files which can be very large (several gigabytes). The functions available in the `ForestAtRisk` package process large rasters by blocks of data, making calculations fast and efficient. This allows deforestation to be modeled over large geographic areas (eg. at the country scale) and at high resolution (eg. $\leq$ 30 m). The `ForestAtRisk` package offers the possibility of using a logistic regression with auto-correlated spatial random effects to model the deforestation process. The spatial random effects make it possible to structure the residual spatial variability in the deforestation process which is not explained by the variables of the model and is often very large. In addition to providing new features, the `ForestAtRisk` Python package is open source (GPLv3 license), cross-platform and scriptable (via Python), user-friendly (functions provided with full documentation and examples), and easily extendable (with statistical models for example). The Python package `ForestAtRisk` was recently used to model deforestation and predict future forest cover by 2100 across the humid tropics.

Statement of Need

Commonly called the "Jewels of the Earth," tropical forests shelter 30 million species of plants and animals representing half of the Earth's wildlife and at least two-thirds of its plant species (Gibson et al., 2011). Through photosynthesis and carbon sequestration, they play an important role in the global carbon cycle, and in regulating the global climate (Baccini et al., 2017). Despite the many ecosystem services they provide, tropical forests are disappearing at an alarming rate (Keenan et al., 2015; Vancutsem et al., 2020), mostly because of human activities. Currently, around 8 Mha (twice the size of Switzerland) of tropical forest are disappearing each year (Keenan et al., 2015). Spatial modelling of the deforestation allows identifying the main factors determining the spatial risk of deforestation and quantifying their relative effects. Forecasting forest cover change is paramount as it allows anticipating the consequences of deforestation (in terms of carbon emissions or biodiversity loss) under various technological, political and socio-economic scenarios, and informs decision makers accordingly (Clark et al., 2001). Because both biodiversity and carbon vary greatly in space (Allnutt et al., 2008; Baccini et al., 2017), it is necessary to provide spatial forecasts of forest cover change to properly quantify biodiversity loss and carbon emissions associated with future deforestation.

The `ForestAtRisk` Python package we present here can be used to model the tropical deforestation spatially, predict the spatial risk of deforestation, and forecast the future forest cover in the tropics (Figure 1). Several other software allow modelling and forecasting forest cover change (Mas et al., 2014). Most famous land cover change software include Dinamica-EGO (Soares-Filho et al., 2002), Land Change Modeller (Eastman & Toledano, 2017), and CLUE (Verburg & Overmars, 2009). Despite the many functionalities they provide, these software are not open source and might not all be cross-platform, scriptable, and completely user-friendly. Moreover, the statistical approaches they propose to model the land cover change do not allow to take into account the residual spatial variability in the deforestation process which is not explained by the model's variables, and which is often very large. The more recent sophisticated algorithms they use (genetic algorithms, artificial neural networks, or machine learning algorithms) might also have the tendency to overfit the data (Mas et al., 2014). Finally, the possibility to use these software on large spatial scales (eg. at the country or continental scale) with high resolution data (eg. $\leq$ 30 m) have not yet been demonstrated (but see Soares-Filho et al. (2006) for a study in the Amazon at 1 km resolution). The `ForestAtRisk` Python package aims at filling some of these gaps and enlarging the range of software available to model and forecast tropical deforestation.

Main functionalities

A set of functions for modelling and forecasting deforestation

The `ForestAtRisk` Python package includes functions to (i) compute the forest cover change raster and the rasters of explanatory variables for a given country from several global datasets (such as OpenStreetMap or the SRTM Digital Elevation Database v4.1 for example) (ii) sample efficiently the forest cover change observations and retrieve the information on spatial explanatory variables for each observation, (iii) estimate the parameters of various statistical deforestation models, (iv) predict the spatial probability of deforestation, (v) forecast the likely forest cover in the future, (vi) validate the models and the projected maps of forest cover change, (vii) estimate the carbon emissions associated to future deforestation, and (viii) plot the results. The `ForestAtRisk` package includes a hierarchical Bayesian logistic regression model with autocorrelated spatial random effects which is well suited for modelling deforestation (see below). But any statistical model class with a `.predict()` method can potentially be used together with the function `forestatrisk.predict_raster()` to predict the spatial risk of deforestation. This allows a wide variety of additional statistical models from other Python packages to be used, such as `scikit-learn` (Pedregosa et al., 2011) for example, for deforestation modeling and forecasting.

Ability to process large raster data

Data for forest cover change and spatial explanatory variables are commonly available as georeferenced raster data. Raster data consists of rows and columns of cells (or pixels), with each cell storing a single value. The resolution of the raster dataset is its pixel width in ground units. Depending on the number of pixels (which is a function of the raster's geographical extent and resolution), raster files might occupy a space of several gigabytes on the hard drive. Processing such large rasters in memory can be prohibitively intensive. Functions in the `ForestAtRisk` package process large rasters by blocks of pixels representing subsets of the raster data. This makes computation efficient, with low memory usage. Reading and writing subsets of raster data is done by using two methods from the GDAL Python bindings (GDAL/OGR contributors, 2020): `gdal.Dataset.ReadAsArray()` and `gdal.Band.WriteArray()`. Numerical computations on arrays are performed with the `NumPy` Python module whose core is mostly made of optimized and compiled C code which runs fast (Harris et al., 2020). This allows the `ForestAtRisk` Python package to model and forecast forest cover change on large spatial scales (eg. at the country or continental scale) using high resolution data (eg. $\leq$ 30 m), even on personal computers with average performance hardware. For example, the `ForestAtRisk` Python package has been used on a personal computer to model and forecast the forest cover change at 30 m resolution for the Democratic Republic of the Congo (total area of 2,345 million km^2), processing large raster files of 71,205 $\times$ 70,280 cells without issues.

Statistical model with autocorrelated spatial random effects

The `ForestAtRisk` Python package includes a method called `forestatrisk.model_binomial_iCAR()` to estimate the parameters of a logistic regression model including auto-correlated spatial random effects. The model considers the random variable y_i which takes value 1 if a forest pixel i is deforested in a given period of time, and 0 if it is not. The model assumes that y_i follows a Bernoulli distribution of parameter θ_i (Equation 1). θ_i represents the spatial relative probability of deforestation for pixel i and 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. The model can include or not an intercept α. To account for the residual spatial variation in the deforestation process, the model includes additional random effects $\rho_{j(i)}$ for the cells of a spatial grid covering the study-area. The spatial grid resolution has to be chosen in order to have a reasonable balance between a good representation of the spatial variability and a limited number of parameters to estimate. Each observation i is associated to one spatial cell $j(i)$. Random effects ρ_j are assumed to be spatially autocorrelated through an intrinsic conditional autoregressive (iCAR) model (Besag et al., 1991). 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'. The variance of the spatial random effects ρ_j is denoted V_ρ. The number of neighbouring cells for cell j, which might vary, is denoted n_j. Spatial random effects ρ_j account for unmeasured or unmeasurable variables (Clark, 2005) which explain a part of the residual spatial variation in the deforestation process that is not explained by the fixed spatial explanatory variables (X_i). Parameter inference is done in a hierarchical Bayesian framework. The `far.model_binomial_iCAR()` method calls an adaptive Metropolis-within-Gibbs algorithm (Rosenthal, 2011) written in C for maximum computation speed.

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

Applications and perspectives

The Python package `ForestAtRisk` was recently used to model the spatial probability of deforestation and predict the forest cover change by 2100 across the humid tropics (https://forestatrisk.cirad.fr/tropics). Future package developments will focus on expanding documentation, case studies, statistical models and validation tools. We are convinced that the "ForestAtRisk" package could be of great help in obtaining estimates of carbon emissions and biodiversity loss under various scenarios of deforestation in the tropics. Such scenarios should help decision-makers take initiatives to tackle climate change and the biodiversity crisis. The results from the `ForestAtRisk` package could contribute to future IPCC and IPBES reports (IPBES, 2019; IPCC, 2014) or help implement the REDD+ mechanism of the Paris Agreement.

Figures

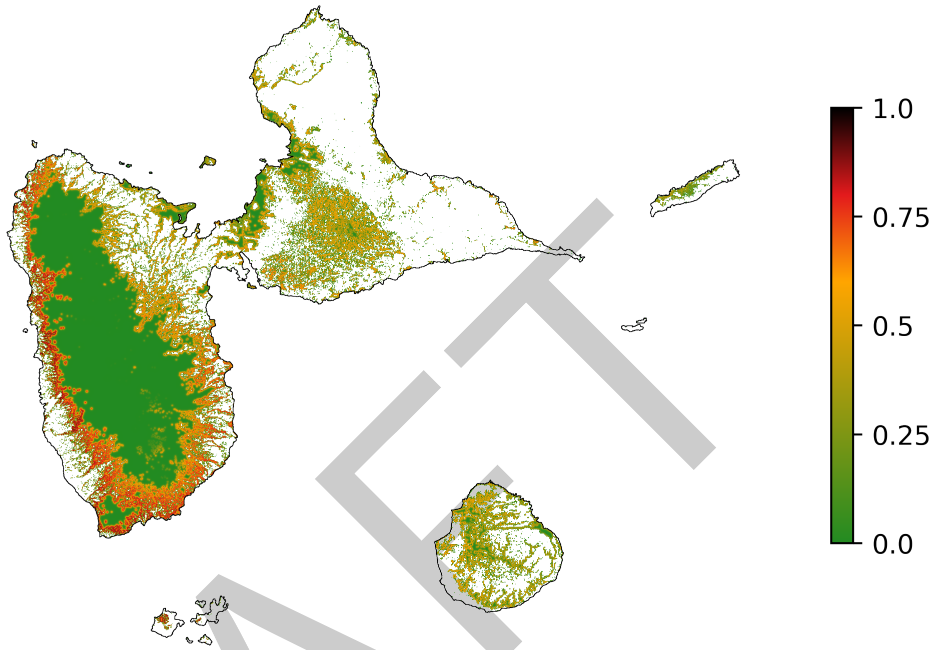

Figure 1: Map of the spatial probability of deforestation in the Guadeloupe archipelago. This map has been produced with the `ForestAtRisk` Python package. Colored pixels represent the extent of the natural old-growth tropical moist forest in 2020. The original map has a 30 m resolution. A relative spatial probability of deforestation was computed for each forest pixel. 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 modelling 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. I would also like to thank Chris Garrard for writing the book *"Geoprocessing with Python"* (Garrard, 2016) which has been of great help for the development of the `ForestAtRisk` package. 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

Allnutt, T. F., Ferrier, S., Manion, G., Powell, G. V. N., Ricketts, T. H., Fisher, B. L., Harper, G. J., Irwin, M. E., Kremen, C., Labat, J.-N., Lees, D. C., Pearce, T. A., & Rakotondrainibe, F. (2008). A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conservation Letters*, *1*(4), 173–181. https://doi.org/10.1111/j.1755-263X.2008.00027.x

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., & Houghton, R. A. (2017). Tropical forests are a net carbon source based on aboveground measurements of gain and loss. *Science*, *358*(6360), 230–234. https://doi.org/10.1126/science.aam5962

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. https://doi.org/10.1007/bf00116466

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

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological forecasts: An emerging imperative. *Science*, *293*(5530), 657–660. https://doi.org/10.1126/science.293.5530.657

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

Garrard, C. (2016). *Geoprocessing with Python*. Manning Publications. ISBN: 978-1617292149

GDAL/OGR contributors. (2020). *GDAL/OGR geospatial data abstraction software library*. Open Source Geospatial Foundation. https://gdal.org

Gibson, L., Lee, T. M., Koh, L. P., Brook, B. W., Gardner, T. A., Barlow, J., Peres, C. A., Bradshaw, C. J. A., Laurance, W. F., Lovejoy, T. E., & Sodhi, N. S. (2011). Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, *478*(7369), 378–381. https://doi.org/10.1038/nature10425

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

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services. The Intergovernmental Platform on Biodiversity and Ecosystem Services, IPBES, Bonn, Germany.* (summary for policy makers). IPBES Secretariat. https://doi.org/10.5281/zenodo.3553579

IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (R. K. Pachauri & L. A. Meyer, Eds.; pp. 3–87). The Intergovernmental Panel on Climate Change, IPCC. https://www.ipcc.ch/report/ar5/syr/

Keenan, R. J., Reams, G. A., Achard, F., Freitas, J. V. de, Grainger, A., & Lindquist, E. (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, *352*, 9–20. https://doi.org/10.1016/j.foreco.2015.06.014

Mas, J.-F., Kolb, M., Paegelow, M., Olmedo, M. T. C., & Houet, T. (2014). Inductive pattern-based land use/cover change models: A comparison of four software packages. *Environmental Modelling & Software*, *51*, 94–111. https://doi.org/10.1016/j.envsoft.2013.09.010

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. https://jmlr.org/papers/v12/pedregosa11a.html

Rosenthal, J. S. (2011). Optimal proposal distributions and adaptive MCMC. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. CRC Press. https://doi.org/10.1201/b10905-5

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

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., & Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature*, *440*(7083), 520–523. https://doi.org/10.1038/nature04389

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragao, L., & Nasi, R. s. (2020). Long-term (1990-2019) monitoring of tropical moist forests dynamics. *bioRxiv*. https://doi.org/10.1101/2020.09.17.295774

Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics in land use modeling: Exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model. *Landscape Ecology*, *24*(9), 1167–1181. https://doi.org/10.1007/s10980-009-9355-7