



**HAL**  
open science

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

Ghislain Vieilledent

► **To cite this version:**

Ghislain Vieilledent. forestatrisk: a Python package for modelling and forecasting deforestation in the tropics. *Journal of Open Source Software*, 2021, 6 (59), 10.21105/joss.02975 . hal-03162430v1

**HAL Id: hal-03162430**

**<https://hal.inrae.fr/hal-03162430v1>**

Submitted on 8 Mar 2021 (v1), last revised 8 Apr 2021 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License




# 1 ForestAtRisk: A Python package for modelling and 2 forecasting deforestation in the tropics

3 **Ghislain Vieilledent**<sup>1, 2, 3, 4</sup>

4 **1** CIRAD, UMR AMAP, F-34398 Montpellier, France **2** CIRAD, Forêts et Sociétés, F-34398  
5 Montpellier, France. **3** AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France.  
6 **4** European Commission, Joint Research Centre (JRC), I-21027 Ispra (VA), Italy.

DOI: [10.21105/joss.02975](https://doi.org/10.21105/joss.02975)

## Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Katy Barnhart](#) 

## Reviewers:

- [@molgor](#)
- [@ethanwhite](#)

Submitted: 10 December 2020

Published: 20 January 2021

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## 7 Summary

8 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.

## 23 Statement of Need

24 Commonly called the “Jewels of the Earth,” tropical forests shelter 30 million species of  
25 plants and animals representing half of the Earth’s wildlife and at least two-thirds of its plant  
26 species ([Gibson et al., 2011](#)). Through photosynthesis and carbon sequestration, they play an  
27 important role in the global carbon cycle, and in regulating the global climate ([Baccini et al.,  
28 2017](#)). Despite the many ecosystem services they provide, tropical forests are disappearing at  
29 an alarming rate ([Keenan et al., 2015](#); [Vancutsem et al., 2020](#)), mostly because of human  
30 activities. Currently, around 8 Mha (twice the size of Switzerland) of tropical forest are  
31 disappearing each year ([Keenan et al., 2015](#)). Spatial modelling of the deforestation allows  
32 identifying the main factors determining the spatial risk of deforestation and quantifying their  
33 relative effects. Forecasting forest cover change is paramount as it allows anticipating the  
34 consequences of deforestation (in terms of carbon emissions or biodiversity loss) under various  
35 technological, political and socio-economic scenarios, and informs decision makers accordingly  
36 ([Clark et al., 2001](#)). Because both biodiversity and carbon vary greatly in space ([Allnutt et al.,  
37 2008](#); [Baccini et al., 2017](#)), it is necessary to provide spatial forecasts of forest cover change to  
38 properly quantify biodiversity loss and carbon emissions associated with future deforestation.

39 The ForestAtRisk Python package we present here can be used to model the tropical  
40 deforestation spatially, predict the spatial risk of deforestation, and forecast the future forest

41 cover in the tropics (Figure 1). Several other software allow modelling and forecasting forest  
42 cover change (Mas et al., 2014). Most famous land cover change software include *Dinamica-*  
43 *EGO* (Soares-Filho et al., 2002), *Land Change Modeller* (Eastman & Toledano, 2017), and  
44 *CLUE* (Verburg & Overmars, 2009). Despite the many functionalities they provide, these  
45 software are not open source and might not all be cross-platform, scriptable, and completely  
46 user-friendly. Moreover, the statistical approaches they propose to model the land cover  
47 change do not allow to take into account the residual spatial variability in the deforestation  
48 process which is not explained by the model's variables, and which is often very large. The  
49 more recent sophisticated algorithms they use (genetic algorithms, artificial neural networks,  
50 or machine learning algorithms) might also have the tendency to overfit the data (Mas et al.,  
51 2014). Finally, the possibility to use these software on large spatial scales (eg. at the country,  
52 or continental scale) with high resolution data (eg.  $\leq 30$  m) have not yet been demonstrated  
53 (but see Soares-Filho et al. (2006) for a study in the Amazon at 1 km resolution). The  
54 *ForestAtRisk* Python package aims at filling some of these gaps and enlarging the range of  
55 software available to model and forecast tropical deforestation.

## 56 Main functionalities

### 57 A set of functions for modelling and forecasting deforestation

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

### 73 Ability to process large raster data

74 Data for forest cover change and spatial explanatory variables are commonly available as  
75 georeferenced raster data. Raster data consists of rows and columns of cells (or pixels), with  
76 each cell storing a single value. The resolution of the raster dataset is its pixel width in ground  
77 units. Depending on the number of pixels (which is a function of the raster's geographical  
78 extent and resolution), raster files might occupy a space of several gigabytes on the hard  
79 drive. Processing such large rasters in memory can be prohibitively intensive. Functions in  
80 the *ForestAtRisk* package process large rasters by blocks of pixels representing subsets of  
81 the raster data. This makes computation efficient, with low memory usage. Reading and  
82 writing subsets of raster data is done by using two methods from the GDAL Python bindings  
83 (GDAL/OGR contributors, 2020): `gdal.Dataset.ReadAsArray()` and `gdal.Band.Write`  
84 `Array()`. Numerical computations on arrays are performed with the NumPy Python module  
85 whose core is mostly made of optimized and compiled C code which runs fast (Harris et al.,  
86 2020). This allows the *ForestAtRisk* Python package to model and forecast forest cover  
87 change on large spatial scales (eg. at the country or continental scale) using high resolution

88 data (eg.  $\leq 30$  m), even on personal computers with average performance hardware. For  
89 example, the ForestAtRisk Python package has been used on a personal computer to model  
90 and forecast the forest cover change at 30 m resolution for the Democratic Republic of the  
91 Congo (total area of 2,345 million km<sup>2</sup>), processing large raster files of 71,205  $\times$  70,280 cells  
92 without issues.

### 93 Statistical model with autocorrelated spatial random effects

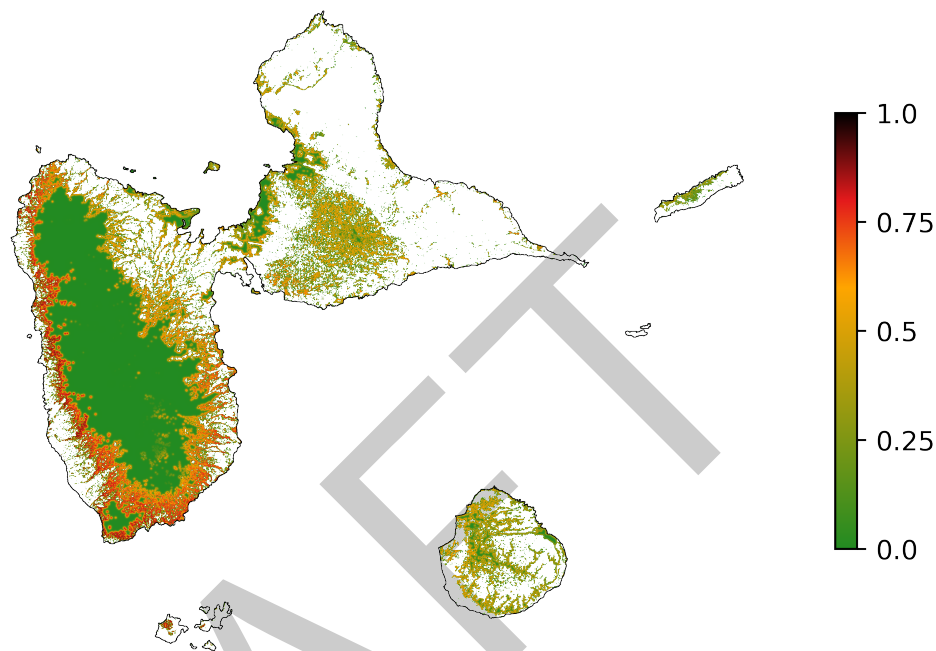
94 The ForestAtRisk Python package includes a method called `forestatrisk.model_binomi`  
95 `al_iCAR()` to estimate the parameters of a logistic regression model including auto-correlated  
96 spatial random effects. The model considers the random variable  $y_i$  which takes value 1 if a  
97 forest pixel  $i$  is deforested in a given period of time, and 0 if it is not. The model assumes  
98 that  $y_i$  follows a Bernoulli distribution of parameter  $\theta_i$  (Equation 1).  $\theta_i$  represents the spatial  
99 relative probability of deforestation for pixel  $i$  and is linked, through a logit function, to a  
100 linear combination of the explanatory variables  $X_i\beta$ , where  $X_i$  is the vector of explanatory  
101 variables for pixel  $i$ , and  $\beta$  is the vector of effects  $[\beta_1, \dots, \beta_n]$  associated to the  $n$  variables.  
102 The model can include or not an intercept  $\alpha$ . To account for the residual spatial variation in  
103 the deforestation process, the model includes additional random effects  $\rho_{j(i)}$  for the cells of  
104 a spatial grid covering the study-area. The spatial grid resolution has to be chosen in order  
105 to have a reasonable balance between a good representation of the spatial variability and a  
106 limited number of parameters to estimate. Each observation  $i$  is associated to one spatial  
107 cell  $j(i)$ . Random effects  $\rho_j$  are assumed to be spatially autocorrelated through an intrinsic  
108 conditional autoregressive (iCAR) model (Besag et al., 1991). In an iCAR model, the random  
109 effect  $\rho_j$  associated to cell  $j$  depends on the values of the random effects  $\rho_{j'}$  associated to  
110 neighbouring cells  $j'$ . The variance of the spatial random effects  $\rho_j$  is denoted  $V_\rho$ . The  
111 number of neighbouring cells for cell  $j$ , which might vary, is denoted  $n_j$ . Spatial random  
112 effects  $\rho_j$  account for unmeasured or unmeasurable variables (Clark, 2005) which explain a  
113 part of the residual spatial variation in the deforestation process that is not explained by the  
114 fixed spatial explanatory variables ( $X_i$ ). Parameter inference is done in a hierarchical Bayesian  
115 framework. The `far.model_binomial_iCAR()` method calls an adaptive Metropolis-within-  
116 Gibbs algorithm (Rosenthal, 2011) written in C for maximum computation speed.

$$\begin{aligned} y_i &\sim \text{Bernoulli}(\theta_i) \\ \text{logit}(\theta_i) &= \alpha + X_i\beta + \rho_{j(i)} \\ \rho_{j(i)} &\sim \text{Normal}\left(\sum_{j'} \rho_{j'} / n_j, V_\rho / n_j\right) \end{aligned} \quad (1)$$

### 117 Applications and perspectives

118 The Python package ForestAtRisk was recently used to model the spatial probability of  
119 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 docu-  
120 mentation, case studies, statistical models and validation tools. We are convinced that the  
121 “ForestAtRisk” package could be of great help in obtaining estimates of carbon emissions  
122 and biodiversity loss under various scenarios of deforestation in the tropics. Such scenarios  
123 should help decision-makers take initiatives to tackle climate change and the biodiversity cri-  
124 sis. The results from the ForestAtRisk package could contribute to future IPCC and IPBES  
125 reports (IPBES, 2019; IPCC, 2014) or help implement the REDD+ mechanism of the Paris  
126 Agreement.  
127

128 **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>.

129 **Acknowledgements**

130 I am grateful to Clovis Grinand, Romuald Vaudry and Matthieu Tiberghien who gave me  
131 the opportunity to work on deforestation modelling when we were leading forest conserva-  
132 tion projects in Madagascar. I also warmly thank Frédéric Achard and all the members of  
133 the IFORCE group for their invaluable support during the first phase of development of the  
134 package, while I was seconded to the JRC in Ispra. I would also like to thank Chris Garrard  
135 for writing the book “*Geoprocessing with Python*” (Garrard, 2016) which has been of great  
136 help for the development of the `ForestAtRisk` package. This work benefited from fund-  
137 ing by FRB-FFEM (BioSceneMada project, AAP-SCEN-2013 I), the European Commission  
138 (Roadless Forest project), and CNRT (RELIQUES project).

139 **References**

- 140 Allnutt, T. F., Ferrier, S., Manion, G., Powell, G. V. N., Ricketts, T. H., Fisher, B. L.,  
141 Harper, G. J., Irwin, M. E., Kremen, C., Labat, J.-N., Lees, D. C., Pearce, T. A., &  
142 Rakotondrainibe, F. (2008). A method for quantifying biodiversity loss and its application  
143 to a 50-year record of deforestation across Madagascar. *Conservation Letters*, 1(4), 173–  
144 181. <https://doi.org/10.1111/j.1755-263X.2008.00027.x>
- 145 Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., & Houghton, R. A.  
146 (2017). Tropical forests are a net carbon source based on aboveground measurements of  
147 gain and loss. *Science*, 358(6360), 230–234. <https://doi.org/10.1126/science.aam5962>
- 148 Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications  
149 in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.  
150 <https://doi.org/10.1007/bf00116466>
- 151 Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecology Letters*,  
152 8(1), 2–14. <https://doi.org/10.1111/j.1461-0248.2004.00702.x>
- 153 Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D.  
154 M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O.,  
155 Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological forecasts: An emerging  
156 imperative. *Science*, 293(5530), 657–660. <https://doi.org/10.1126/science.293.5530.657>
- 157 Eastman, J. R., & Toledano, J. (2017). A short presentation of the Land Change Mod-  
158 eler (LCM). In *Geomatic approaches for modeling land change scenarios* (pp. 499–505).  
159 Springer International Publishing. [https://doi.org/10.1007/978-3-319-60801-3\\_36](https://doi.org/10.1007/978-3-319-60801-3_36)
- 160 Garrard, C. (2016). *Geoprocessing with Python*. Manning Publications. ISBN: 978-  
161 1617292149
- 162 GDAL/OGR contributors. (2020). *GDAL/OGR geospatial data abstraction software library*.  
163 Open Source Geospatial Foundation. <https://gdal.org>
- 164 Gibson, L., Lee, T. M., Koh, L. P., Brook, B. W., Gardner, T. A., Barlow, J., Peres, C. A.,  
165 Bradshaw, C. J. A., Laurance, W. F., Lovejoy, T. E., & Sodhi, N. S. (2011). Primary  
166 forests are irreplaceable for sustaining tropical biodiversity. *Nature*, 478(7369), 378–381.  
167 <https://doi.org/10.1038/nature10425>
- 168 Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau,  
169 D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk,  
170 M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., ... Oliphant,  
171 T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- 173 IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity  
174 and ecosystem services. The Intergovernmental Platform on Biodiversity and Ecosystem  
175 Services, IPBES, Bonn, Germany.* (summary for policy makers). IPBES Secretariat. <https://doi.org/10.5281/zenodo.3553579>
- 177 IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II  
178 and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*  
179 (R. K. Pachauri & L. A. Meyer, Eds.; pp. 3–87). The Intergovernmental Panel on Climate  
180 Change, IPCC. <https://www.ipcc.ch/report/ar5/syr/>
- 181 Keenan, R. J., Reams, G. A., Achard, F., Freitas, J. V. de, Grainger, A., & Lindquist, E.  
182 (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources  
183 Assessment 2015. *Forest Ecology and Management*, 352, 9–20. <https://doi.org/10.1016/j.foreco.2015.06.014>
- 184



- 185 Mas, J.-F., Kolb, M., Paegelow, M., Olmedo, M. T. C., & Houet, T. (2014). Inductive pattern-  
186 based land use/cover change models: A comparison of four software packages. *Environ-*  
187 *mental Modelling & Software*, 51, 94–111. <https://doi.org/10.1016/j.envsoft.2013.09.010>
- 188 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel,  
189 M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau,  
190 D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in  
191 Python. *Journal of Machine Learning Research*, 12, 2825–2830. [https://jmlr.org/papers/  
192 v12/pedregosa11a.html](https://jmlr.org/papers/v12/pedregosa11a.html)
- 193 Rosenthal, J. S. (2011). Optimal proposal distributions and adaptive MCMC. In S. Brooks,  
194 A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*.  
195 CRC Press. <https://doi.org/10.1201/b10905-5>
- 196 Soares-Filho, B. S., Cerqueira, G. C., & Pennachin, C. L. (2002). DINAMICA - a stochastic  
197 cellular automata model designed to simulate the landscape dynamics in an Amazonian  
198 colonization frontier. *Ecological Modelling*, 154(3), 217–235. [https://doi.org/10.1016/  
199 S0304-3800\(02\)00059-5](https://doi.org/10.1016/S0304-3800(02)00059-5)
- 200 Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C.  
201 A., Voll, E., McDonald, A., Lefebvre, P., & Schlesinger, P. (2006). Modelling conservation  
202 in the Amazon basin. *Nature*, 440(7083), 520–523. <https://doi.org/10.1038/nature04389>
- 203 Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego,  
204 J., Aragao, L., & Nasi, R. s. (2020). Long-term (1990-2019) monitoring of tropical moist  
205 forests dynamics. *bioRxiv*. <https://doi.org/10.1101/2020.09.17.295774>
- 206 Verburg, P. H., & Overmars, K. P. (2009). Combining top-down and bottom-up dynamics  
207 in land use modeling: Exploring the future of abandoned farmlands in Europe with the  
208 Dyna-CLUE model. *Landscape Ecology*, 24(9), 1167–1181. [https://doi.org/10.1007/  
209 s10980-009-9355-7](https://doi.org/10.1007/s10980-009-9355-7)