ISEC – 5 July 2018

Accounting for spatial autocorrelation in deforestation modelling (and forecasting)

Ghislain Vieilledent[1,2] and Frédéric Achard[1]

[1] **EU Joint Research Center**, Ispra, ITALY
[2] **Cirad**, Montpellier, FRANCE

European Commission

cirad

Motivation

Consequences of deforestation

- Biodiversity loss
- Carbon emissions and climate change

Modelling deforestation

- Scenarios of biodiversity (IPBES) and CO_2 emissions (IPCC)
- Anticipate and take action (ex. policies)

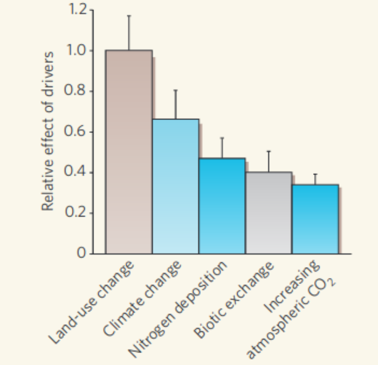


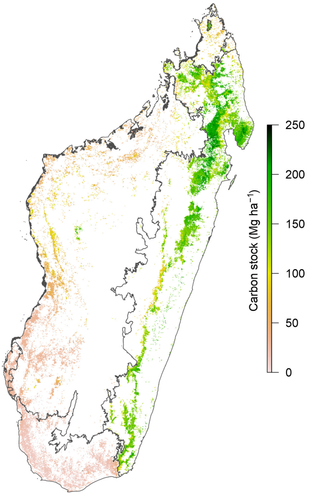
Figure 1 | The main factors, or 'drivers', affecting biodiversity.

Thuiller 2007, *Nature*

Focus

Location of deforestation

- ~~How much deforestation ? (deforestation rates)~~
- Where deforestation occurs preferentially ?
- Biodiversity and carbon stock vary spatially



Vieilledent et al. 2016, *J. of Ecol.*

State of art

Model

- $Y_i \in \{0, 1\}$
- $Y_i \sim \mathcal{B}ernoulli(\theta_i)$
- $\text{logit}(\theta_i) = f(\text{spatial factors}_i)$

Spatial factors

- **Landscape** : distance to forest edge
- **Accessibility** : altitude, slope, distance to road, town
- **Land-policy** : protected area network

Research gap

Unmeasured factors

- Many other explicative spatial factors
- Population density, soil, geographical barriers, controls...
- Some **not measured**, some **unmeasurable**

Proposed model

- $\text{logit}(\theta_{ij}) = f(\text{spatial factors}_i) + \rho_j$
- Spatial random effects ρ_j to account for unmeasured factors

Objective

Model comparison

- Model 1 : $\text{logit}(\theta_{ij}) = f(\text{spatial factors}_i)$
- Model 2 : $\text{logit}(\theta_{ij}) = f(\text{spatial factors}_i) + \rho_j$

Challenge

- Use the model for forecasting (not only for inference)
- On large spatial scale

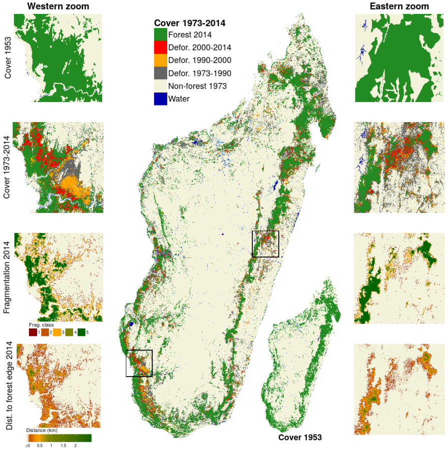
Data

Deforestation process in Madagascar. Current rate : 100,000 ha/yr.

Data

Response variable

- $Y_i \in \{0, 1\}$
- Deforestation in Madagascar on the period 2000-2010
- 30 m resolution
- 20,000 points i (deforested/non-deforested)



Vieilledent et al. 2018 *Biol. Conserv.*

Data

Explicative variables

Product	Source	Variable derived	Unit	Resolution (m)
Deforestation maps (1990-2000-2010)	Vieilledent et al. 2018	distance to forest edge	m	30
		distance to past deforestation	m	30
Digital Elevation Model	SRTM v4.1 CSI-CGIAR	altitude	m	90
		slope	°	90
Highways	OSM - Geofabrik	distance to roads	m	150
Places		distance to towns	m	150
Waterways		distance to river	m	150
Protected areas	Rebioma	presence of protected area	-	30

Model

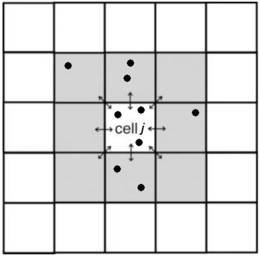
Spatial model

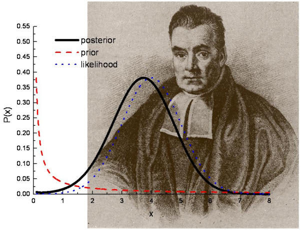
- $\text{logit}(\theta_{ij}) = f(\text{spatial factors}_i) + \rho_j$
- ρ_j : 10 km resolution ($\sim$ 1500 cells j)

Intrinsic CAR

$$p(\rho_j | \rho_{j'}) = \mathcal{N}ormal(\mu_j, V_\rho / n_j)$$

μ_j : mean of $\rho_{j'}$ in the neighborhood of j.
V_ρ : variance of the spatial random effects.
n_j : number of neighbors for cell j.





Software

Python `deforestprob` package

- Efficient geoprocessing on (very) large rasters
- Sampling, inference, and spatial projection
- Gibbs sampler in C using Metropolis algorithm
- GitHub : `https://github.com/ghislainv/deforestprob`

Usable in R

- `reticulate` R package
- Access to Python functions and objects
- Plotting

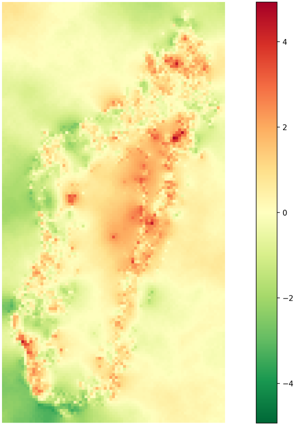
Parameter estimates

	GLM		iCAR	
Parameter	Mean	SE	Mean	SE
Intercept	0.0112	0.020	-0.705	0.108
protected area	-0.3747	0.034	-0.549	0.0694
altitude	-0.1774	0.019	-0.662	0.0713
slope	-0.1166	0.018	-0.147	0.0255
dist. defor	-0.7537	0.031	-0.869	0.0546
dist. defor^2	0.0750	0.005	0.0734	0.00595
dist. edge	-0.3711	0.024	-0.51	0.0366
dist. road	-0.0006	0.016	-0.101	0.0464
dist. town	-0.0938	0.016	-0.0168	0.0424
Vrho	-	-	7.49	0.512

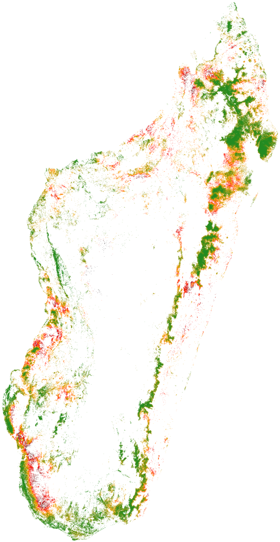
Small changes in parameter estimates, but same sign ($\pm$) and relative magnitude.
Intuitive effects. Both models are rather good for explanatory modeling.

Spatial random effects

- Hotspots of deforestation
- Not explained by the fixed env. factors



Spatial probability of deforestation



- Computed at 30 m resolution
- Greener : lower probability
- Darker red : higher probability

Model performance

Based on the 20,000 observations

- GLM explains only 8% of the deviance
- iCAR model better whatever the index
- GLM closer to the null model

Index	null	GLM	iCAR
Deviance expl. (%)	0	8	30
OA (%)	50	62	79
Kappa (%)	0	24	58

Forecasting power

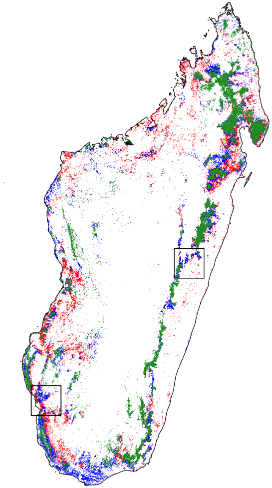
- Map of probability of deforestation in 2010 + known deforested area on 2010-2014
- Observed vs. projected deforestation on 2010-2014
- Area deforested in 10 × 10 km areas

Index	GLM	iCAR
Pearson corr.	12%	31%

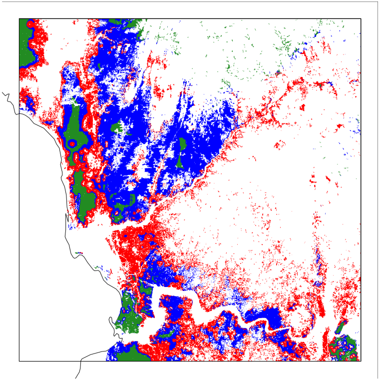
Comparing long term forecast

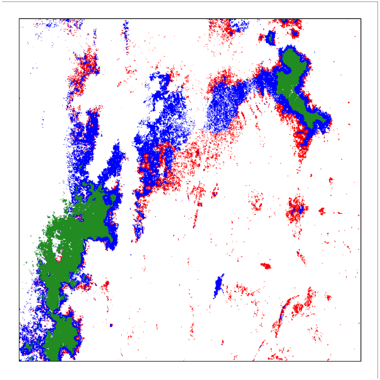
Deforestation 2010-2050

- Assuming deforestation of 100,000 ha/yr (current rate)
- green : residual forest in 2050
- red : deforested area 2010-2050
- blue : differences in model predictions



Comparing long term forecast





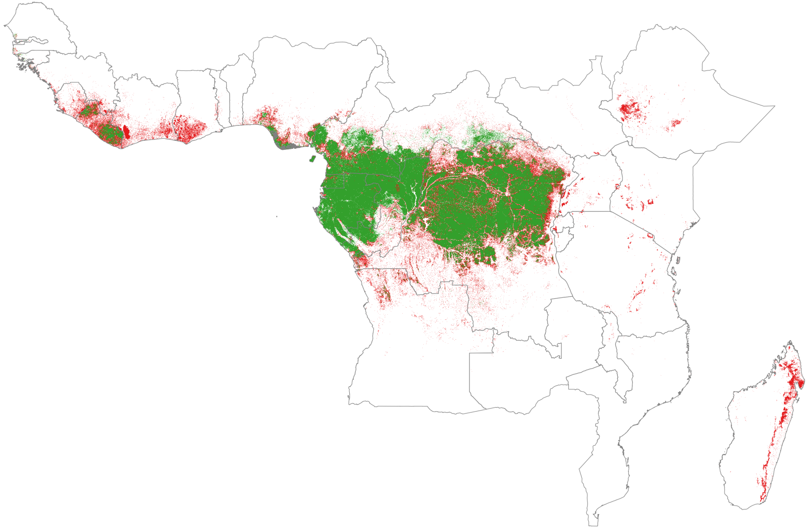
Advantages of the iCAR model

- Account for unmeasured or unmeasurable factors
- Spatial random effect for all cells in the landscape
- Model performance is higher
- More accurate projections

Limitations

- Better, but still with a lot of uncertainty
- Random effects can stand for many different factors
- Do they last in time (ex. people migration) ?
- It is always better to include fixed effects if possible

Extending work to the world humid tropical forest



Tropical forest conservation

- Tropical forests are disappearing at an alarming rate
- Engage for forest and biodiversity conservation
- Tarzan will tank you !

... Thank you for attention ...