RESEARCH PROGRAM ON Forests, Trees and Agroforestry

Consultancy Report number 6, by Ophélie Ratel – October 2020

Implementation of a statistical model following a Bayesian approach

Objective

Express the quantity of biomass recovered over time by including co-variables (landscape, climate, topography, soil, functional component) and developing our model in a Bayesian framework, which is particularly adapted when there is little data, thanks to the addition of information that could be described as "non-pure data" or "priors". These priors can be based on previous studies or expert knowledge: they are a way to make sure that our predictions are within the range of acceptable values given our prior knowledge on similar processes, and are thus especially important when data is scarce. Moreover, the Bayesian approach allows a rigorous estimation of parameters correlation and uncertainty.

Hypothesis

Landscape structure (configuration and composition) and dynamic -> impact on the regeneration potential of forests (proximity to crops, fragmentation dynamic, etc.).

The more complex the structure of the landscape, the more negatively it influences the rate of forest regeneration.

The increase in the number of interfaces between forests and intensive agriculture reduces the regeneration potential of forests.

The higher the dynamics of landscape fragmentation, the lower the potential for forest regeneration.

Method

Implementation of the model was done using R Studio and RStan package (Stan Development Team, 2020[1]). The model is written in two parts. First, all the theoretical model is built within a RStan file (under R Studio). Then, this model is used in a R script, adding the dataset avec all the covariables data.

The several steps to build the model are as following:

1. Describe every input data as vectors
    - Plots number
    - Inventories number
    - Exploited plots
    - AGB value
    - Year of inventory + year of disturbance
    - Covariables
2. Describe the parameters of the model following the Michaelis & Menten equation
    - AGBmax, which represents the maximum value of biomass
    - Mu and sigma AGBmax, for the mean and standard deviation
    - Km, which is the necessary time to reach half of AGBmax

---

[1] Stan Development Team, 2020. RStan: the R interface to Stan. R package version 2.21.2. http://mc-stan.org/

- T0, for exploited forests which do not start with an AGB = 0
- Lambda, which represents a random effect acting on AGBmax, assessing that there may be have variation between plots.

3. Then, the model has to be estimated.
    - Here, we want to explain the AGB production according to variables, following a distribution law of mu and sigma parameters. We chose a normally distributed law but we may change for a log normally distribution.
    - We then added some priors with a high level of uncertainty that can be subsequently reduced according to our *a priori*

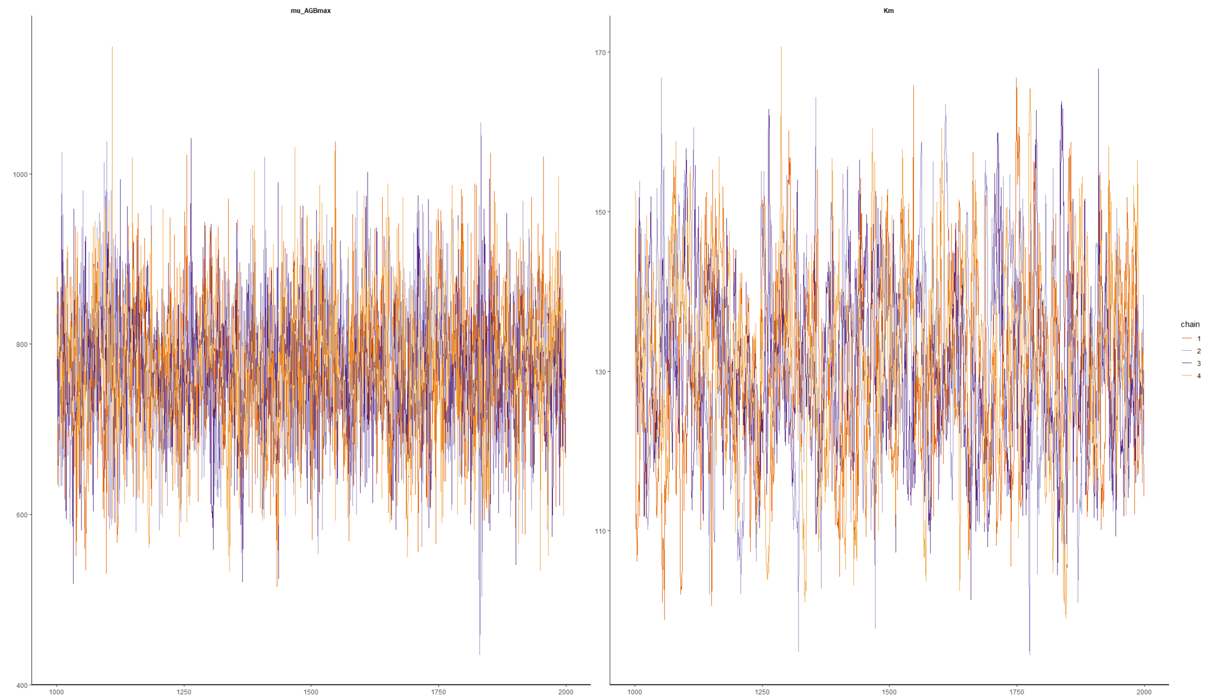
Once these steps completed, the model is run with the data. For now we didn't add any real covariables to see how the model is working, calibrate the data and observe how the biomass evolves according to time. We added some lines within the model to explain how the covariables will be added.

R Script: *data_AGB-2*

RStan file: *AGB_model.stan*

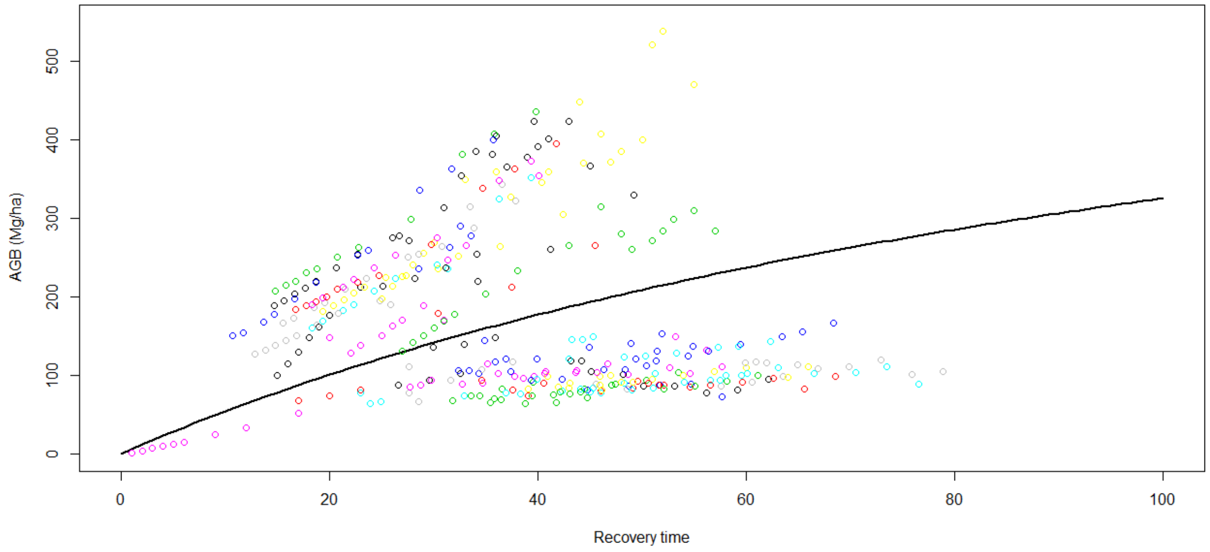
Results

After running the model, the first thing to check is the value Rhat, which is the potential scale reduction factor on split chains. In other terms, the more Rhat near to 1, the more the model converges. Here, for every of our parameters, Rhat =1, which means that the model converged, and so that the results are consistent.



This graphs shows the model iterations and parameters stabilisation for both AGBmax and Km parameters. According to this graph the model estimates a mean AGBmax value near 800 Mg/ha, for a mean Km value of 130 years, which means that it would take on average 130 years for the forests to reach 400 Mg/ha. But for now we didn't add any covariables and our priors are quite permissive.

We then made a prediction curve using the maximum likelihood for every parameter, showing biomass recovery according to time.



This graph shows in black the prediction curve of Biomass recovery according to time from the year of disturbance. Each colour represents one plot.

Perspectives

Further work will aim to integrate the different covariates into the model in order to assess the impact of its different components on this biomass production. Our hypothesis is that landscapes features will have an impact on how forests regenerate and their speed of regeneration. For example, we could ask ourselves whether a patch of forest surrounded by banana or pineapple crops would be more or less productive than a patch of forest in the middle of a heterogeneous landscape (mixing forest, crops and pastures).