Landscape effect

Table of contents

1 Estimate a buffer size of landscape effect on plant and viral richness

1.1 The idea :

The main objective of this part is to find a way to estimate a buffer radius around grids in which landscape could have the most impact on plant richness, viral richness, or any other variables that we want to explain. For that, we will use the package *Siland*. It uses a form of a general linear model for spatial variables.

1.2 Hypothesis

1- Environmental variables that mainly structure the environment are salinity concentration and flooding potential. They are divided by distance from the sea. 2- Because grids are the majority old, plant communities should be resistant to external invasion in such a way that Landscape has no effect on species richness (when environmental variance is considered).

1.3 Implementation

The main complexity is that the landscape cover types and environmental variables are strongly collinear (eg. it is more likely to find agricultural lands in fertile soil with low salt concentration). We must account for variance explained by environmental factors before trying to explain it with the landscape. It can be done easily with the *Siland* function because it accounts for local effects (aka environmental variables) with a classic linear part and a spatial effect part (aka landscapes). The main issue with this is that there are too many environmental and landscape parameters to account for (more than 20 env variables + one cover type

= one parameter ⇒ more than 25 variables for 42 observations). If we give the model all environmental parameters adding over landscape parameters, there is a strong risk of over-fitting our richness or having problems with the convergence of all parameters. To avoid that we must select in advance environmental parameters that account for the most variation of the explained variable. The first option is to compute a PCA of environmental factors and then use the axis that explains the most variation of environmental conditions as co-variable of the spacial model. The *Siland* model will thus attribute a part of the variance to a linear combination of multiple environmental variables and then search for the remaining variance explained by landscape variables (I'm not fully sure of that because of the loop for parameter convergence ...). This is a good way to reduce the variables in the landscape model. However, the PCA doesn't select necessarily environmental variables that explain the most variation in the response variable. Thus, using this method, we assume that the environmental variables that structure the most the study area have the highest contributions to the variance of the plant richness . (This is a strong hypothesis that we must assume , but it is not completely bad because it is honest to assume a priori that richness will vary along those main gradients). The second idea is to use a classic glm with all explanatory environmental variables and then use the residuals of this model as variables to explain. This method doesn't make the strong hypothesis of the first one, but it looks like what we wanted to avoid in the first place: over-fitting and convergence issues. So for now we will stick to the PCA

1.4 Method

1.4.1 Data

Environmental variables

The data has been collected in South France at the Tour du Valat's natural reserve (43° 30 30 N, 4° 40 01 E). There are in total 42 grids of 10m² each one composed of 9 quadra of 20 cm², these latter are distributed in an X shape form in grids. In each quadra an exhaustive identification of plant richness has been done and an estimation of bacterial, fungal, and viral communities of the phyllosphere. Species cover, soil cover (Plant, Litter, Bare Soil) and biomass have also been collected for plants at the quadra scale. Environmental variables have been sampled 3 times (in random locations?) per grid and averaged to limit measure errors.

Environmental variables not integrated for now

We also have information on grazing and mowing in the plots [database ?] and on grazing intensity [database ?]. For meteorological data, we have hydrometry between 2018 and 2022 from [database?]

Landscape variables

The soil occupation shapefile comes from the regional database of soil occupation of PACA and was collected in 2016 [OCSOL 2016].

We have grouped the land-use data into 5 classes: Wetlands are made up of lands that have periods of submergence and can be slightly anthropized (pasture); Non-emitting propagules contain landscapes that are very unlikely to emit propagules that could colonize the salt meadow (e.g. forests, beaches or areas of permanent water); natural landscapes contain lightly anthropized (pasture) land that does not feature in the previous two classes; artificial landscapes contain all types of heavily anthropized cover, such as urban areas or roads, these covers are those that could be of great importance for the import of exotic organisms; cultivated lands are the one used for plantations, such as rice or forage crops.

("Bâti diffus", "Bâti individuel dense", "Tissu urbain continu", "Bâti individuel lâche",
"Bâti isolé", "Décharge", "Bâti collectif", "Terrain vague en zone urbaine",
"Équipement sportif et de loisirs", "Bâti léger ou informel", "Zone d'activité et équipement","Extraction de matériaux",
"Chantier", "Place", "Jardins familiaux", "Espace vert urbain",
"Zone portuaire", "Bâti individuel dans parc paysager", "Cimetière") -> **artificial**

("Marais ouvert", "Feuillu", "Formation arbustive et arborée semi-ouverte", "Ripisylve",
"Conifère", "Formation arbustive et arborée fermée","Plage", "Etang et/ou lagune", "Canal",
"Forêt mélangée") -> **non emitting**

("Riz", "Luzerne", "Prairie temporaire", "Blé", "Tournesol", "Verger, oliveraie", "Friche récente",
"Culture maraichère", "Sorgho, soja", "Vignoble", "Colza", "Maïs") -> **cultivated**

("Prairie naturelle","Coussoul", "Dune embryonnaire",
"Dune végétalisée", "Dune à végétation arbustive") -> **natural landscape**

("Roselière", "Sansouire basse", "Marais ouvert", "Sansouire haute",
"Jonchaie", "Autre marais à végétation émergée", "Marais à marisque", "Sol nu",
"Lagune de pré-concentration", "Table saunante", "Friche salicole récente") -> **wetlands** '

1.4.2 Analysis

To compute optimal buffer radii of soil occupation that affects plant (or viral richness later) we used the package *Siland* (Carpentier and Martin 2021). The model has 2 distinct components: a classic regression part with an intercept and estimators for each variable and a regression part on the spatial component. To summarize, the model tries to minimize the log-likelihood $[y|\theta]$ of 3 parameter types; parameters for size effects of local variables (α_l), parameters for sizes effects of each cover type (β_k), parameters for the shape of the curve of each cover type (δ_k).

$$Y = \mu + \sum_L \alpha_l X + \sum_K \beta_k \sum_R f_{\delta_k}(d_{i,r}) z_r^k$$

L: index of observation point (grids); K: index of cover type; R: landscape mesh index. If the landscape variable k is a presence/absence variable, z_r^k is equal to one or zero. f_{δ_k} Spacial Influence Function (gauss, expo or uniform)

Maximization algorithm of *Siland* :

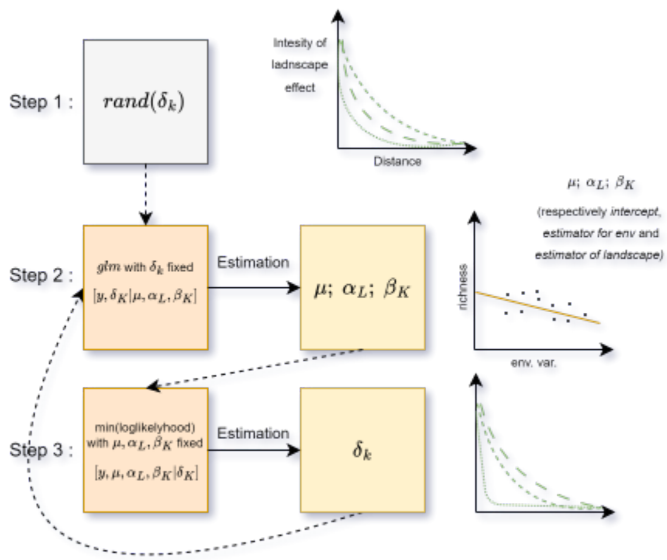

Figure 1: Diagram of *Siland* optimization algorithm. Steps 2 and 3 are repeated until convergence of all parameters (or maximum repetitions is reached)

To avoid catching part of the variance induced by environmental variables we first compute a PCA to extract the axis that represents most of the variance of environmental condition. We then use those linear combinations of environmental variables as descriptors of the local grid's conditions. The model contains in total 8 variables, 3 local variables extracted from the PCA and 5 covers types variables. The model uses a Gaussian function for the landscape effect and uses a Poisson error family. To validate the good convergence of buffer radii, we checked visually that the log-likelihood of all Spatial Influence Functions (SIF) have a minimum that intersects the overall minimal log-likelihoods of the model (as prescribed by the author of the package). Finally, a bootstrap of the local explain variable (eg. richness) coupled with their local explanatory variables (eg. Dim of PCA) is done to assess that the landscape has a statistical significance.

1.4.3 Results

1.4.3.1 PCA

The first axis which accounts for 43.9% of the variation is explained mostly by soil structure and soil fertility. On the left of the PCA the prominent environmental conditions are the

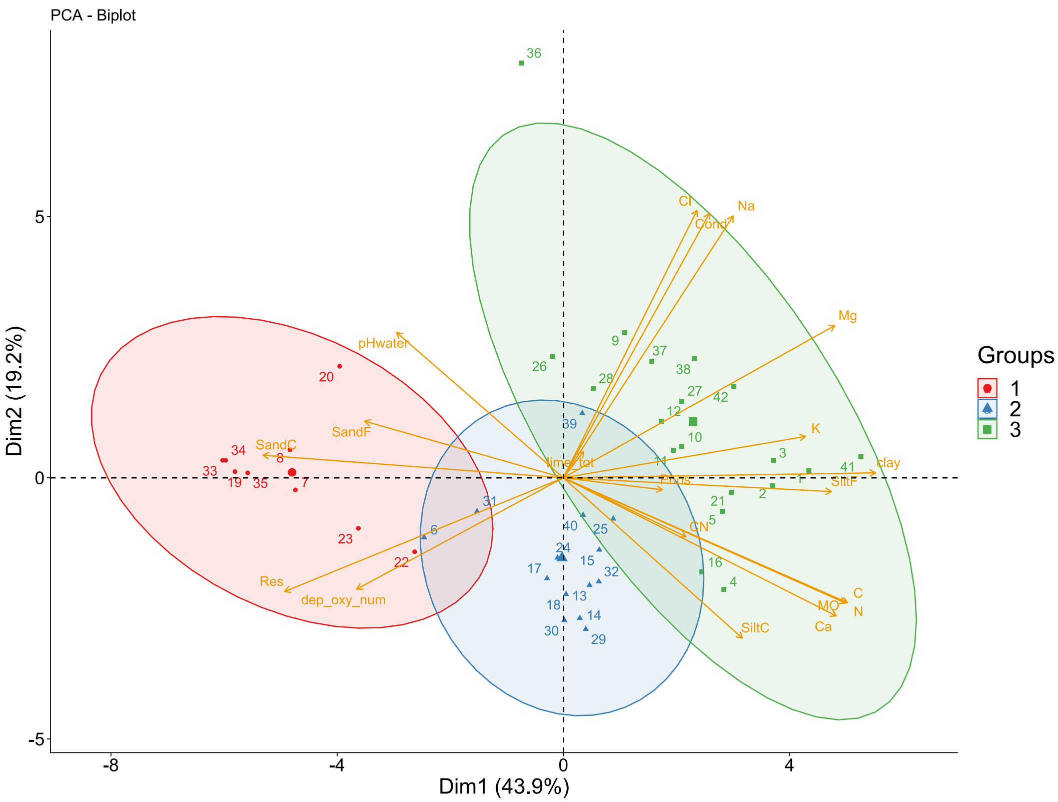

Figure 2: **PCA on environmental variables**. The clustering grids are compared by their Euclidian distance and segregation is done with the ward method