UNIVERSITÉ DE GENÈVE
INSTITUT DES SCIENCES DE L'ENVIRONNEMENT

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES DE LA SOCIÉTÉ
Département de géographie et environnement

CONSERVATOIRE ET JARDIN BOTANIQUES

CERTIFICATE OF GEOMATICS

SPECIES DISTRIBUTION MODELLING OF THE *CAPURODENDRON* AUBRÉV. (SAPOTACEAE) GENUS IN MADAGASCAR

MEMOIR

RHÉA GARRATT

JUNE 2015

Director : Dr. Laurent Gautier, Conservatoire et Jardin Botaniques de la Ville de Genève, et Laboratoire de Systématique végétale et Biodiversité

Co-director : Dr. Anthony Lehmann, EnviroSPACE, Institut des Sciences de l'Environmment

TABLE OF CONTENTS

1. INTRODUCTION

1.1. MADAGASCAR AND STUDIES ON ITS FLORA

Madagascar is a tropical island in the Indian Ocean, with an area of approximately 587'000 km^2. It is an exceptional reservoir of animal and vegetal biodiversity due to its particular conditions, for instance its edaphic and climatic diversity (Myers, et al. 2000). Furthermore, the endemic species rate is high on the island, as a direct consequence of its tectonic and climatic history: an early separation and isolation of the island, approximately 100 millions of years ago (De Wit 2003, Gautier et Goodman 2003).

The island of Madagascar has been the subject of botanical explorations since two centuries. H. Perrier de la Bâthie and H. Humbert made the first big contributions by characterizing vegetation types and producing vegetation maps of the island (Perrier de la Bâthie 1921, H. Humbert 1965, Koechlin, Guillaumet et Morat 1974, Gautier et Goodman 2003, H. Humbert 1955, Perrier de la Bâthie 1936) (Cornet et Guillaumet 1976). To this day, it is still widely used, but it has to be noted that the WWF recently proposed a new classification system of the different regions of the island (Gautier et Goodman 2003).

Pressure on species habitat is growing and the need to make information about the extinction risk of species rapidly available to the public, scientists and policy makers is urgent to improve the efficiency of protected area coverage (Willis, Moat et Paton 2003, Callmander, Schatz et Lowry II 2005, Callmander, Schatz, et al. 2007). A possibility would be to concentrate on areas such as global biodiversity hotspots and their endemic species (Schatz 2009, Callmander, Schatz, et al. 2007).

This work has for main objective the modelling of the distribution of the Capurodendron genus species distribution, using occurrence-only data extracted from herbaria specimen. This work will be integrated in a Master thesis on new approaches in conservation: the improvement of the knowledge of plant species threat level, for instance the IUCN indexes for the *Capurodendron* genus, thanks to a fast and objective methodology involving species distribution modelling and the development of new approaches to set conservation priorities thanks to molecular information, such as species age, in countries with low botanical explorations resources like Madagascar.

1.2. THE SAPOTACEAE FAMILY AND *CAPURODENDRON* GENUS

The reference of the *Capurodendron* genus figures in the work of Aubréville (1974) on the Sapotaceae family, part of the series titled "Flore de Madagascar et des Comores" and founded by H. Humbert. Indeed, the genus, whose name stems from the surname Capuron which belonged to the botanist who collected the major part of the herbaria concerning the *Capurodendron* Aubrév. genus, is strictly endemic to Madagascar. It includes 23 tree and shrub species (Figure 1) whose flowers present a homogenous structure, unlike the seeds. This last fact explains the difficulty of the previous botanists to bring the collection of species under the same genus. The

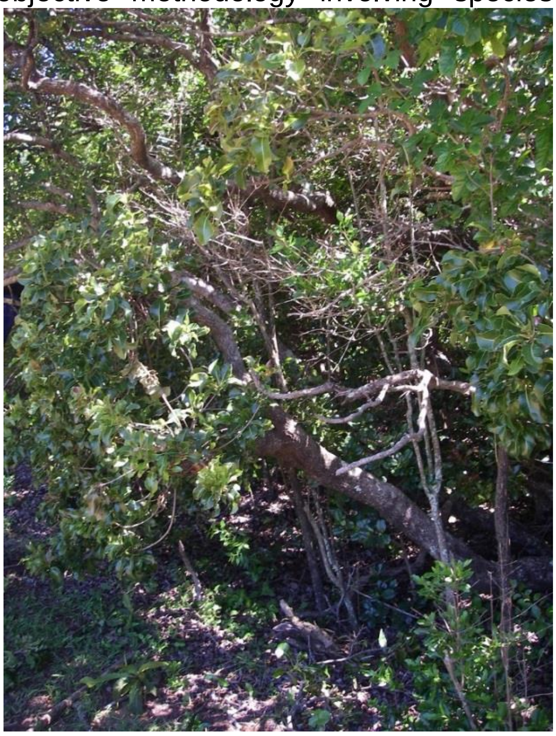


Figure 1. Photograph of the plant species Capurodendron greveanum Aubrév., by Martin W. Callmander (http://www.tropicos.org/Image/76441)

species of *Capurodendron* are spread throughout the whole island (Figure 2), although two tendencies can be identified within the genus: species populating the dry forests (northeast, south and west of the island), or the dense humid forests (east of the island). This genus is one of the largest endemic genus in Madagascar (Gautier, Naciri, et al. 2013).

A Madagascar endemic like the *Capurodendron* genus from an island widely known as harbouring an invaluable part of the world's biodiversity is a way to increase the efficiency of assessing the Malagasy flora (Callmander, Schatz et Lowry II 2005).

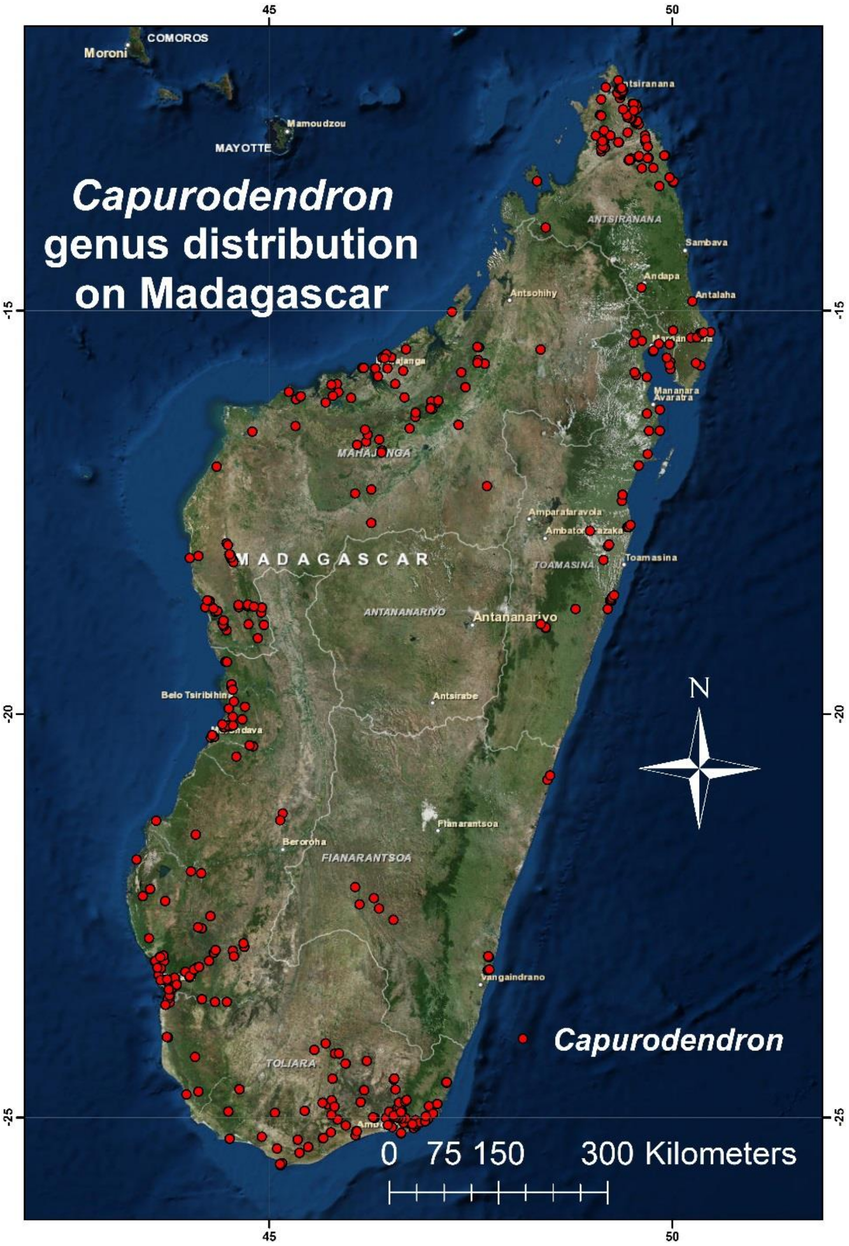


Figure 2. Map of the distribution of the Capurodendron genus on Madagascar

1.3. IUCN RED LIST CATEGORIES

In 2002, governments part of the Sixth Conference of the Parties (COP6) to the Convention on Biological Diversity (CBD) endorsed the Global Strategy for Plant Conservation (GSPC) for 2010, which was accepted again for 2020. The program enounces 16 targets, revised in the

updated version of the GSPC 2011-2020, which mostly address the need to prevent the loss of plant diversity (IUCN Plant Conservation Sub Comittee 2011).
More specifically, the second target stipulates that all known plant species should be assessed according to their conservation status to guide conservation efforts, which calls for the efforts of plant taxonomists. However, studies recognize the huge task that awaits and call for the development of priority-setting techniques helping to speed up the process (Schatz 2009, Rivers, Taylor, et al. 2011, Miller, et al. 2013).

A well-known conservation tool that informs about the threat level of a species is the International Union for Conservation of Nature (IUCN) Red List Categories and Criteria (2000). The aim was to produce a classification system (Figure 3) based on the biological indicators of populations that was easy to understand, explicit, objective and applicable to a variety of organisms.
In 2004, the IUCN Red List contained only 3% of the known plants. The number rose to 5% in 2011, but is in itself only a fraction of the world plant diversity which has yet to be discovered (Callmander, Schatz et Lowry II 2005, Miller, et al. 2013).

As showed in Figure 3, a species has to be classified as vulnerable (VU), endangered (EN) or critically endangered (CR) to be defined as under threat.

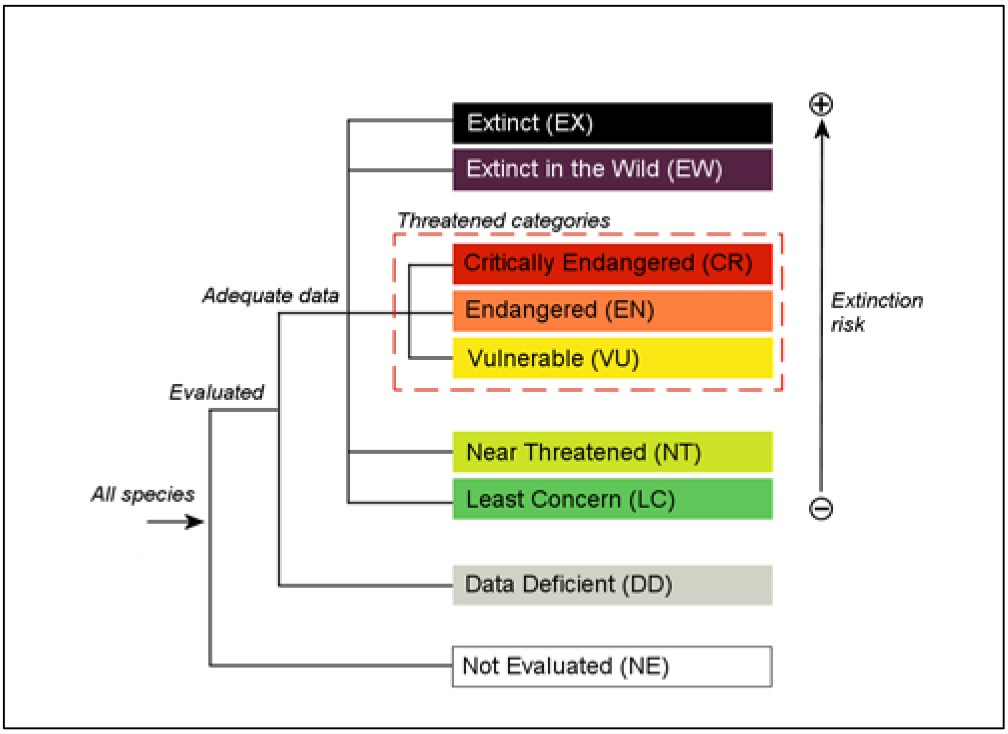


Figure 3. Structure of the categories (IUCN Species Survival Commission 2000).

According to the IUCN Species Survival Commission (2000) and IUCN Standards and Petitions Subcommittee (2014), five criteria are used to evaluate the level of threat of a taxon:

A. Declining population (past, present and/or projected)
B. Geographic range size, and fragmentation, decline, or fluctuations
C. Small population size and fragmentation, decline, or fluctuations
D. Very small population or very restricted distribution
E. Quantitative analysis of extinction risk (e.g., Population Viability Analysis)

See **Erreur ! Source du renvoi introuvable.** page **Erreur ! Signet non défini.** for the definition of the criteria.

For a taxon to be considered under threat, only one of the five criteria has to be fulfilled. For documentation purposes, it is however best to check if the taxon meets any other criterion.

For each criteria, the thresholds are quantitative, but the combination with inference and projection enables the system to be effective on taxa for which there is little information, to prevent for the taxa not to be assessed.
Also, the criteria permit the classification of a taxon even though the exact threat has not yet been identified. Indeed, they are aimed at the symptoms of endangerment, and not the causes.

This review focuses on criterion B. It is determined by two concepts (Figure 4):

- Extent of occurrence (EOO) : given a certain number of known occurrences, the EOO is “the area contained within the shortest continuous imaginary boundary which can be drawn to encompass all the known, inferred or projected sites of present occurrence of a taxon, excluding cases of vagrancy” (IUCN Species Survival Commission 2000). The objective of the measure is to assess the potential spreading of a threat in the geographical distribution of a taxon, and not the distribution of the taxon itself (IUCN Standards and Petitions Subcommittee 2014).
- Area of occupancy (AOO): given a certain number of known occurrences and a certain scale, the area of the AOO is the sum of the pixels containing the occurrences.

It is possible to exclude obviously unsuitable habitats from the EOO measure, but it is discouraged as it is the AOO that represents those factors in a species distribution.

According to the quality of the information, herbarium specimens are usually not the sole source used for a criterion The IUCN gives the possibility of calculating the EOO and the AOO with predicted data from the known occurrences in the case those are very few (IUCN Standards and Petitions Subcommittee 2014). In this work, two preliminary calculations of the EOO will be done based on the known occurrences and potential distribution.

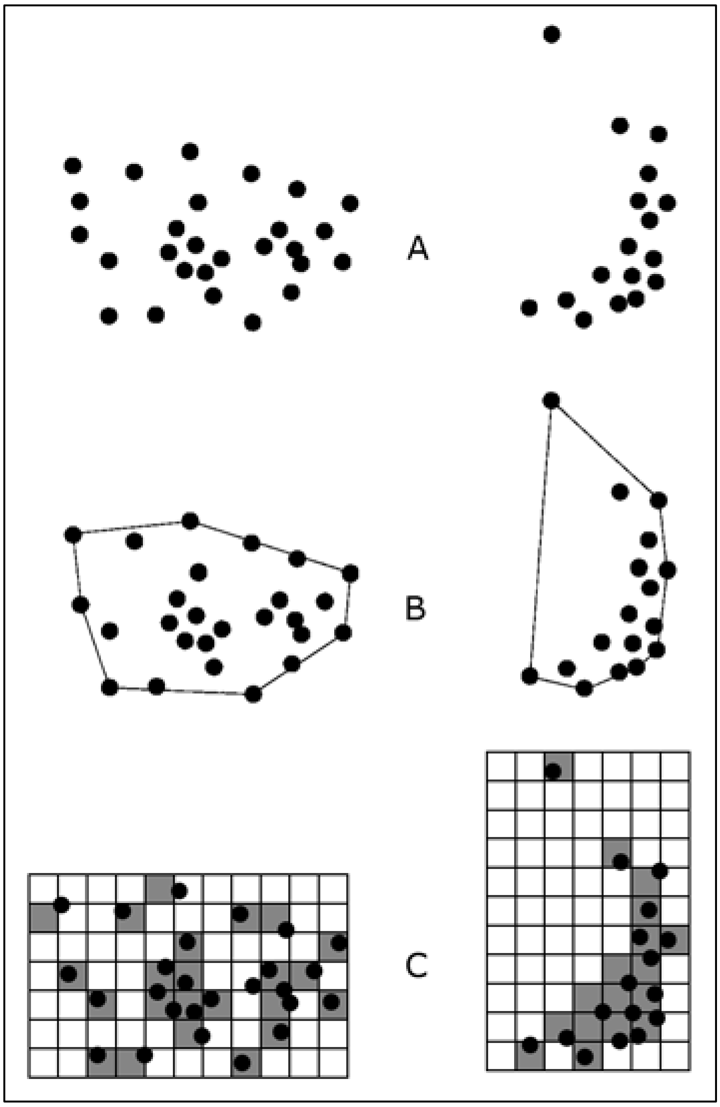


Figure 4. Two examples of the distinction between extent of occurrence and area of occupancy. (A) is the spatial distribution of known, inferred or projected sites of present occurrence. (B) shows one possible boundary to the extent of occurrence, which is the measured area within this boundary. (C) shows one measure of area of occupancy which can be achieved by the sum of the occupied grid squares. (IUCN Species Survival Commission 2000)

2. METHODOLOGY

In ArcGIS (ESRI 2013), the provided tools are essentially Python scripts and each user can create his own set, either by creating new ones or by modifying pre-existing tools to fit his own purposes. In this work, the latter was extensively done and when data was being processed on ArcGIS, most of the steps were documented in a toolbox in ArcGIS which contains Model Builder models, Model Builder being an application available in the ArcMap interface, and Python scripts.
The files were separated into two stages: the preparation of the data used as input for the species distribution model (SDM), and the processing and analysing of the output SDM, the species modelling being conducted by another program in this case.
The toolbox, named “SDM inputs and outputs” for this particular project, is available, has been thoroughly documented and the use of its contents was made compatible with distribution to the public. Indeed, the initial input data that has to be selected when launching the Python script, such as the files containing species occurrences, environmental variables or the output of the SDM, determines the rest of the operations.
The format of the input files for the tools is fixed, but the choice of the SDM, as for the species and environmental data of interest, is up to the user.

Although useful to begin the process, it has to be noted that the Model Builder models only serve the limited purpose of showing the steps taken, as an educational aid, the needed operations being in reality much more complex and converted into a Python Script to address multiple species and multiple layers at the same time by creating loops.
Further details are viewable in the metadata of the tools.

This work was conducted with ArcGIS 10.0 and a full license provided by the University of Geneva.

This chapter retraces the various steps taken before, during and after the species modelling.

2.1. SPECIES DISTRIBUTION MODEL

To model species distribution with occurrence-only data, only a selection of models accept that kind of input, and MaxEnt is one of them (Phillips, Dudík et Schapire 2004, Elith, Graham, et al. 2006). The software is freely available on the Department of Computer Science of Princeton University website (https://www.cs.princeton.edu/~schapire/maxent/).
With pixels with known species presence and environmental variables as inputs, MaxEnt is a maximum entropy species distribution model which finds the probability distribution of maximum entropy, or that is the closest to uniform, to estimate a target probability distribution on the pixels of a given study area.
According to Elith et al. (2011), a model such as MaxEnt that uses presence-only data presents some limitations: the prevalence (proportion of occupied sites) is hard to calculate, a selection bias (sampling effort), spatial and temporal scale bias (presence of a plant species may depend on the time and place) and unknown survey method (sampling unit size). The latter can be corrected by assuming a determined grid cell size.
However, it present various serious advantages like, to name a few of them, requiring presence-only data of course, incorporating continuous or categorical variables, creating a continuous output which gives a lot of freedom concerning the threshold (Phillips, Anderson et Schapire 2006).

As mentioned, such a model needs only two types of inputs: species distribution data, which is the *Capurodendron* genus known occurrences, and environmental data. All of the layers need to be previously processed in a GIS because MaxEnt needs all layers to cover the exact same area for example, and then exported in CSV and ASCII format to be read by MaxEnt.

2.2. DATA PREPARATION

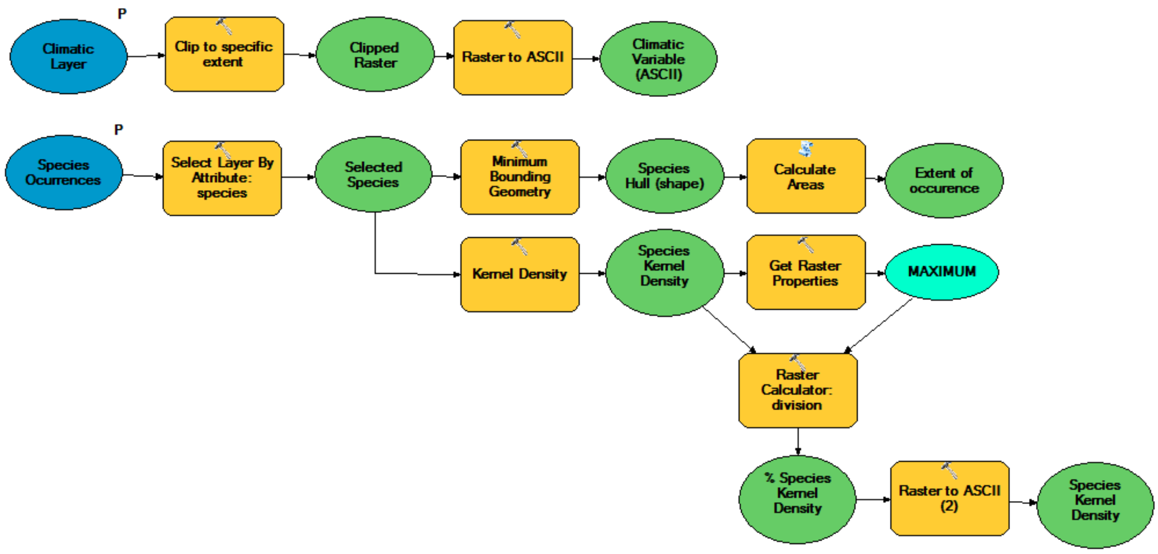


Figure 5. Model of the data preparation steps, made with the Model Builder application available in the ArcMap interface.

As shown in Figure 5, the different layers were processed in ArcMap (ESRI 2013), mainly to restrict them to a specific extent and adapt cell size, which is Madagascar in this case. Extent of occurrence and the kernel densities variables were created from the species occurrences data.

- **SPECIES OCCURRENCE DATA**

Building a database
Explorations on Madagascar have been conducted by various institutions and most of the specimens are recorded on the online Catalogue of Vascular Plants of Madagascar, a project lead by the Missouri Botanical Garden (Madagascar Catalogue 2013). After extraction of the data listed on the website, it was completed with specimens from other databases, such as the catalogue and herbarium of The Muséum National d'Histoire Naturelle in Paris (https://science.mnhn.fr/taxon/genus/capurodendron), and herbarium specimen of the Conservatoire et Jardin Botaniques de la Ville de Genève.

Georeferencing
All herbaria specimen are not accompanied by geographical coordinates or even an accurate description of the collection site. The Gazetteer is an online database constantly being updated, linking known locations on Madagascar to geographical coordinates (Schatz et Lescot 2003), and helped reference geographically collection sites shown on the herbaria specimen sheet.
As of 1980, the GPS was introduced and a good part of the specimens collected as of this date contain more or less accurate coordinates.
Then, the degree-minute-second (DMS) coordinates were converted in decimal degree (DD) coordinates to be used in the SDM, MaxEnt, and the geographical information system (GIS), ArcGIS.

Finally, the georeferencing of the layer was checked according to the extent of the bioclimatic layers, to make sure every occurrence point has a bioclimatic value (e.g. move a point that lands in the sea).

For this work, three species with different and seemingly well-defined known occurrences patterns were selected as an example: *Capurodendron androyense*, *Capurodendron nodosum* and *Capurodendron pervillei* (See Figure 6).

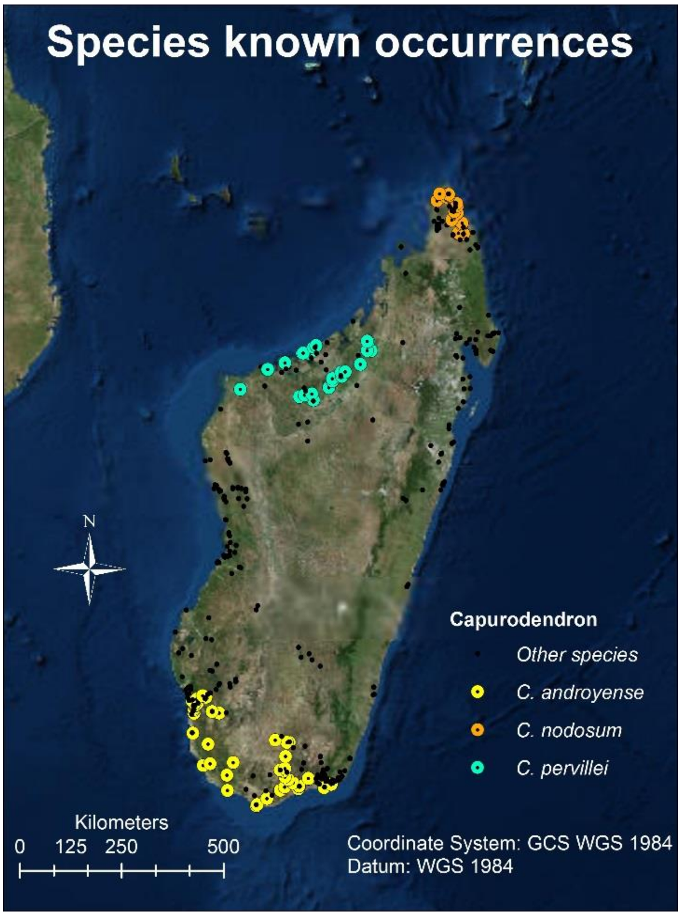


Figure 6. Map of the known occurrences of the Capurodendron genus on Madagascar, with C. androyense (yellow), nodosum (orange) and pervillei (light blue) as the chosen species.

Extent of occurrence (EOO) of known occurrences

The smallest polygon possible englobing all of the occurrences (points) of one species was created, using the ArcGIS tool "Minimum Bounding Geometry". Its area was then calculated, which constitutes the extent of occurrence EOO as defined by the IUCN (see above).

- **ENVIRONMENTAL DATA**

Bioclimatic data

In their WorldClim project, Hijmans & al. (2005) have compiled various bioclimatic variables for with temperature and precipitation measure for the whole world that are freely available on their website (http://www.worldclim.org/). After their download in raster data format, with 1 km-resolution, the layers were clipped on ArcMap to the extent of Madagascar. Each following layers in this work have been adapted to the same extent.

After that, a Principal Component Analysis (PCA) was conducted to put aside bioclimatic variables that show strong correlation to each other (see Annex 1). This way, only eight (out of nineteen) layers were kept (see Figure 7):

- BIO1 = Annual Mean Temperature [°C*10]
- BIO3 = Isothermality (BIO2/BIO7) (* 100) [°C*10]

- BIO4 = Temperature Seasonality (standard deviation *100) [°C*10]
- BIO6 = Min Temperature of Coldest Month [°C*10]
- BIO7 = Temperature Annual Range (BIO5-BIO6) [°C*10]
- BIO12 = Annual Precipitation [mm]
- BIO16 = Precipitation of Wettest Quarter [mm]
- BIO18 = Precipitation of Warmest Quarter [mm]

The layers were then reduced to the area of interest with the "Clip" in ArcGIS, used as a reference to fix the environment of the other layers regarding projection, extent (-9°S to -27°S, 42°E to 52°E) and cell size (1x1km), and exported as a ASCII file.

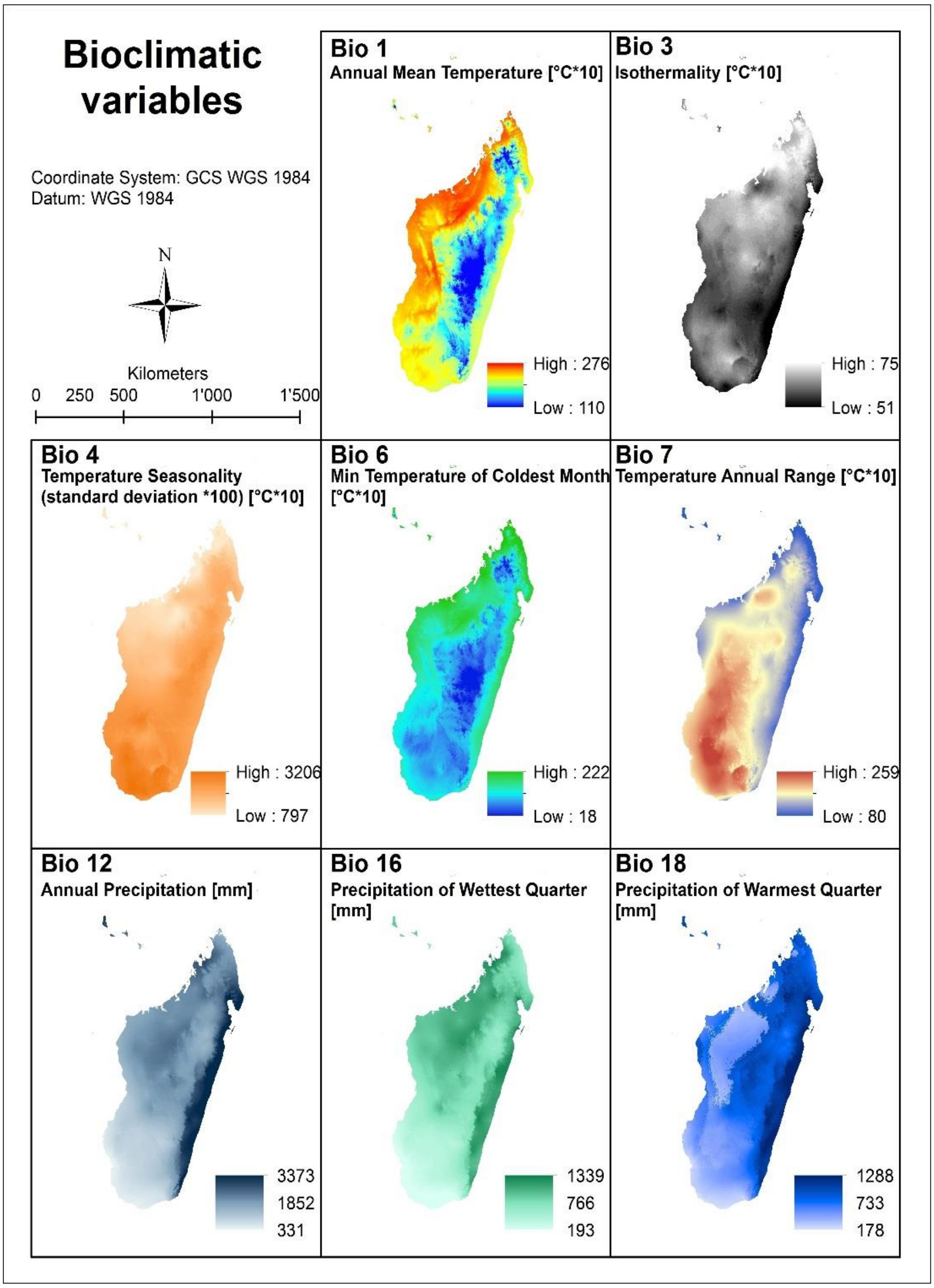


Figure 7. Image of the eight chosen bioclimatic variables: BIO1, 3, 4, 6, 7, 12, 16, 18.

Kernel density

To prevent the modelling from predicting the presence of a species in an area at a great distance from the known occurrences and to thus assign more weight to the known locations, a layer was produced using an ArcGIS tool called “Kernel density”. The layer can actually be compared to a simplified representation of plant dispersal radius, although the radius chose here is fictitious.

For a given radius, the tool calculates a raster in which a surface is fitted to each occurrence point. The smaller the radius value, the more the layer puts an emphasis on the known occurrence areas and, once converted into a ASCIIi format and input to the species distribution model, probably the more the model will constrict its prediction in the area defined by the radius, and vice-versa. However, the second possibility only was tested, which is is to multiply this layer with the potential distribution modelled by MaxEnt for each species. Especially for this purpose, the layer is first converted into one with values ranging from 0 to 1, to assimilate it the percentages, before being converted in the ASCII format.

Geology data

As well as the climatic conditions, vegetation type is also determined by the soil. Du Puy & Moat (1996) have published a simplified geology map (see Annex 2) and it is freely available on the Kew Royal Botanical Gardens website (http://www.kew.org/gis/papers/bio1996.html) in vector format.
From it, a layer distinguishing soils composed of sedimentary rocks of limestone from the other soil types has been produced on ArcMap (see Figure 8). First, the areas containing limestone (“Mesozoic Limestones + Marls (inc. "Tsingy")” and “Tertiary Limestones + Marls & Chalks”), were selected. Then a field was added to the attribute table where two values differentiated the two soil types: “limestone” and “not limestone”. Finally, the layer was converted to a raster layer.

Before being input to the SDM, the extent of the layer has to be exactly the same as the other environmental variables, as well as the projection and the cell size, parameters which were set in the script. Focusing on the extent, a series of steps involving masks and the “Nibble” tool enabled to make sure that every pixel with a non-zero value in the bioclimatic variables is also a non-zero value in the geology layer.

For a summary of the steps, see Annex 3 for the Model Builder model. There is no script for this particular layer, for the operations were done manually.

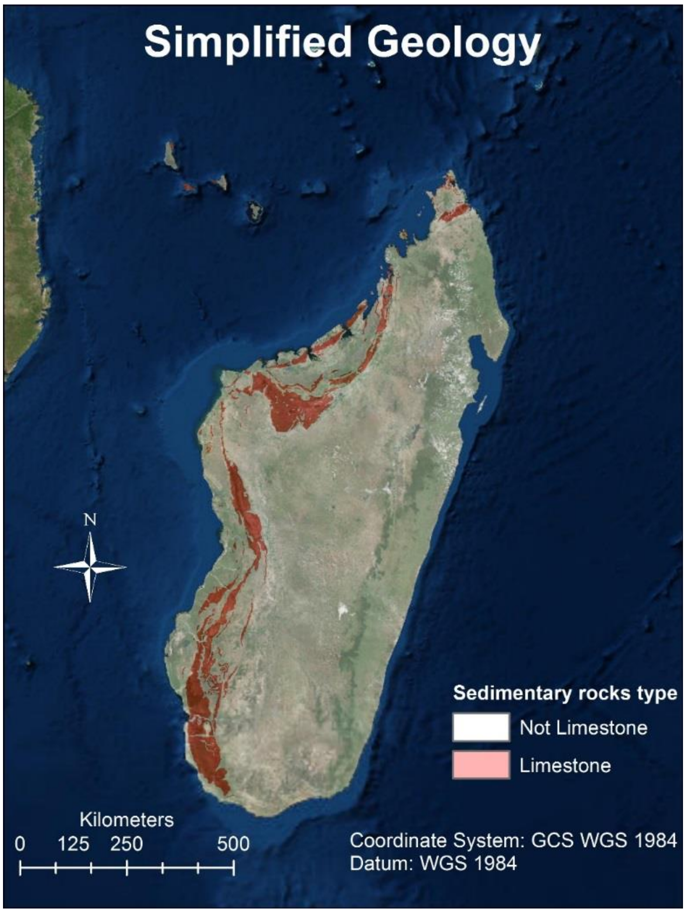


Figure 8. Map of the simplified geology layer, distinguishing areas with a limestone-type sedimentary rock from areas not containing limestone in their sedimentary rock.

2.3. RUNNING THE MODEL

Inputs

MaxEnt offers the possibility of running multiple species with the same set of environmental variables, but only the results of one species, *Capurodendron androyense*, were chose to be exposed here, as an example.

The environmental variables types are continuous for the bioclimatic data and categorical for the geology information.

Output formats

There are three available output formats (logistic, cumulative and raw), but the logistic was chosen for it is easier to interpret as it gives to each pixel a value between 0 and 1.

Statistical analysis

The convergence threshold, maximum iterations and regularization value are set to their default values (respectively 10^{-5}, 1000 and 10^{-4}).

Having a small number of sample records, no random test percentage was set. The model cannot do any testing.

For more information about the parameters, see the tutorial provided with the executable file of MaxEnt.

2.4. OUTPUT PROCESSING

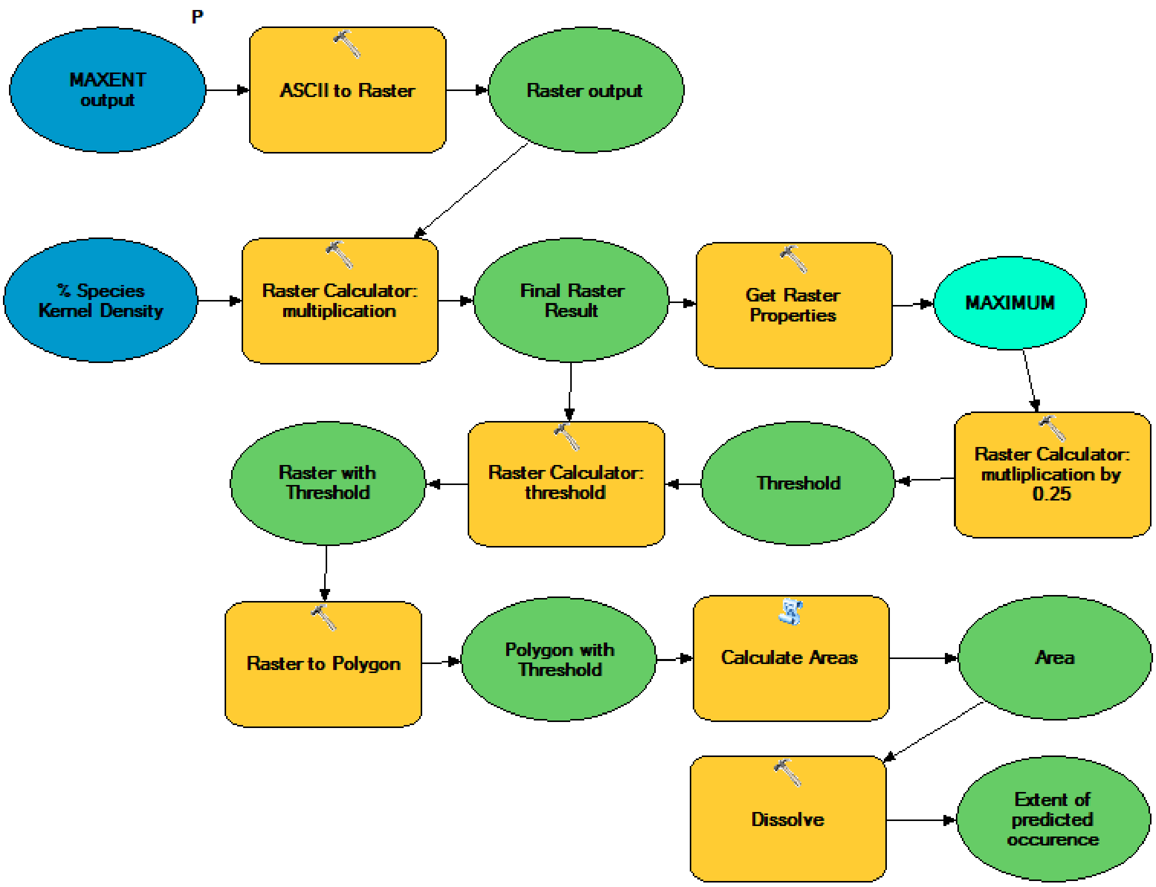


Figure 9. Model of the processing steps taken to analyse output created by MAXENT, made with the Model Builder application available in the ArcMap interface.

After the MaxEnt analysis on the data, the ASCII output can be viewed on ArcMap by converting it to a raster layer. It can constitute a result on its own or be combined to the kernel densities, which is the next step. So the final potential distribution layer is achieved by simply multiplicating the two layers.

For this final layer, a fictitious threshold was chosen, the highest 25%, under which the probability of presence is ignored. The pixels with values above the threshold were then selected, converted into polygons and theirs areas calculates.

Finally, the “Dissolve” tool allowed to create a layer composed of two polygons only containing the former pixel values above and under 25% of the maximum probability. The tool also merges the information contained in the attribute table and adds up the areas. The final area corresponds to the extent of occurrence EOO calculated with predicted distribution (see Figure 9).

3. RESULTS

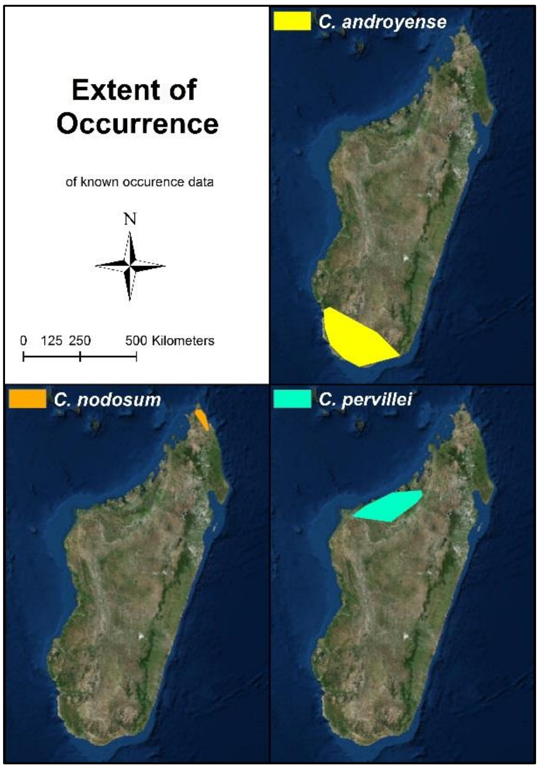


Figure 10. Extent of occurence based on the known occurrences of the Capurodendron species: C. androyense (yellow), C. nodosum (orange) and C. pervillei (blue).

The extent of occurrence with known occurrences

The extent of occurrence polygon *of Capurodendron androyense, nodosum and pervillei* is at the same location than that of the known occurrences (Figure 10). Their respective areas are of 43'740.46 km^2, 2'880.93 km^2 and 23'118.59 km^2.

The kernel densities

As expected, the occurrence points are surrounded by an area with high value decreasing with the increase in distance to those points (see Figure 11).

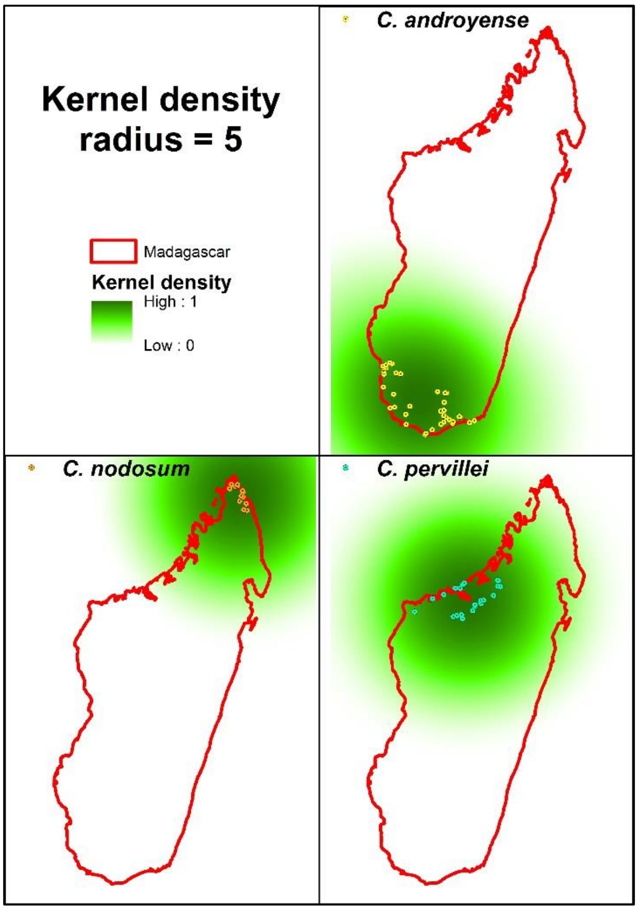


Figure 11. Kernel density (vert) reduced to values between 0 and 1 for each of the four selected species : C. androyense (yellow), C. nodosum (orange) and C. pervillei (blue).

The MaxEnt prediction

The MaxEnt model calculated that the predicted suitable conditions for *Capurodendron androyense* are found along the southern coast of Madagascar with a high probability. There also appears to be a suitable area on the north-eastern coast, although of very low probability.

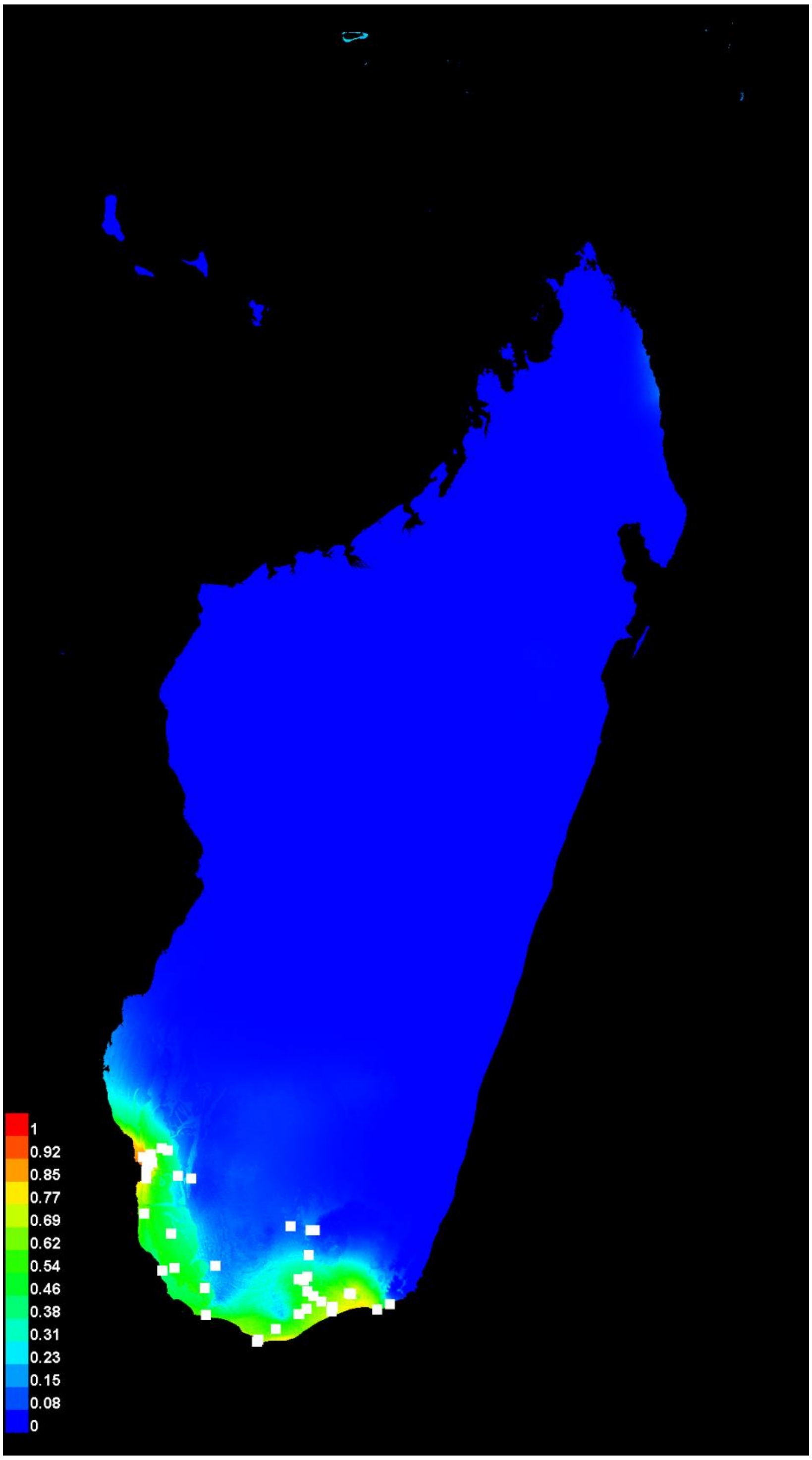


Figure 12.Representation in logistic format of the MaxEnt model for Capurodendron androyense, with areas with high (red) low (light blue) or no (blue) probability of suitability of the species.

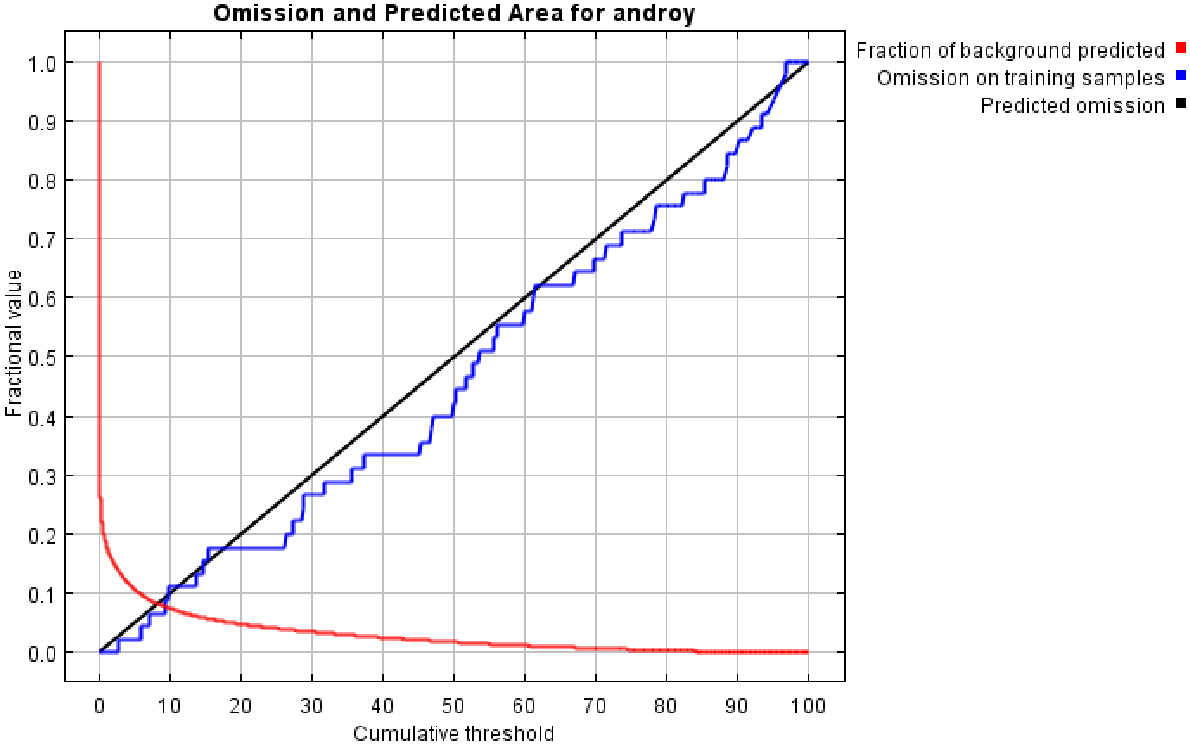


Figure 13. Omission and Predicted Area for Capurodendron androyense.

Statistical analysis

Figure 13 shows how testing and training omission and predicted area vary with the choice of cumulative threshold. The omission on training samples is a good match to the predicted omission rate.

The area under the ROC curve (AUC) for the training data is 97.4% (0.974) (see Annex 4).

Variable importance

Which variables is contributing to fitting the model is a data of which the MAXENT model keeps track while it is being trained. The gain of the model rises every time the coefficient for one feature is changed, and this increase is attributed to an environmental layer and converted to percentages (Table 1). BIO16 appears to contribute the most information by far, the other variables showing a contribution percentage smaller than 10%, and BIO1 and BIO6 bring no contribution at all.

Then, the model is revaluated for each environmental variable, the training AUC drop is measured and the decrease is converted to percentages, which is the permutation importance.

Table 1. Table of the relative contributions of the environmental variables to the MaxEnt model of Capurodendron androyense.

Variable	Percent contribution	Permutation importance
BIO16	81.9	94.8
BIO12	8.3	4.2
BIO18	6.6	0
BIO4	1.5	0.2
BIO3	0.9	0.2
Geolsimp	0.6	0
BIO7	0.3	0.5
BIO1	0	0.1
BIO6	0	0

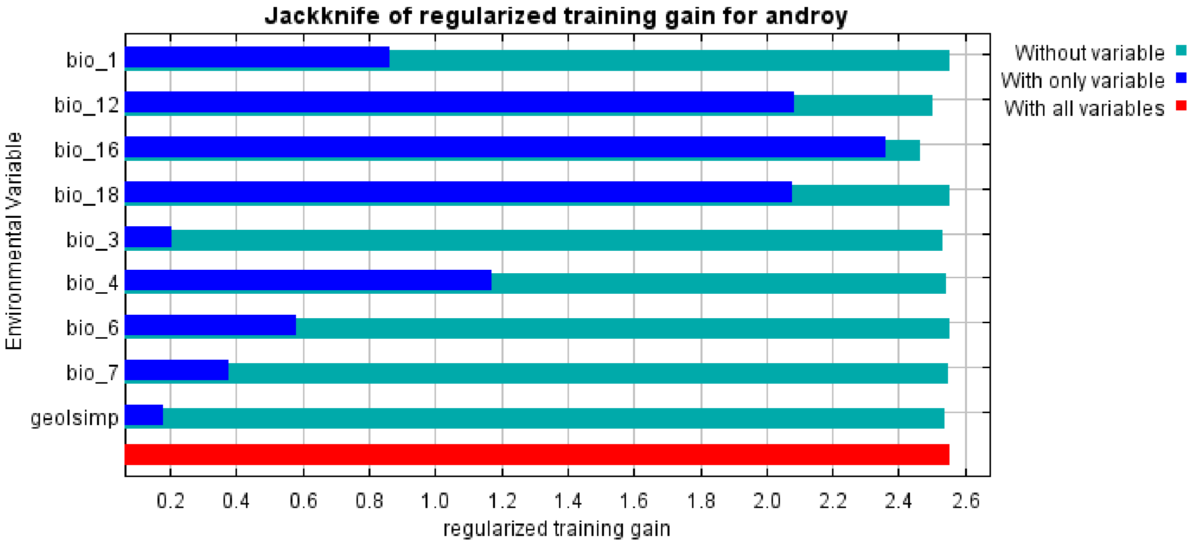


Figure 14. Plot of the jackknife of regularized training gain for Capurodendron androyense without a variable (lignt blue), with only that vairable (dark blue) and with all of the variables (red) for: BIO1, 3, ,4 ,6 ,7, 12, 16, 18, geology (geolsimp).

The jackknife of variable importance (see Figure 14) shows high gains for BIO12, BIO16 and BIO18, and very low gains for BIO3 and the geology (geolsimp), when the variables are taken separately.

The predicted suitability decreases with BIO12, 16 and 18 (Figure 15), and briefly increases with BIO1, 4, 6 and 7. With the geology, it is at its highest in zone 1.

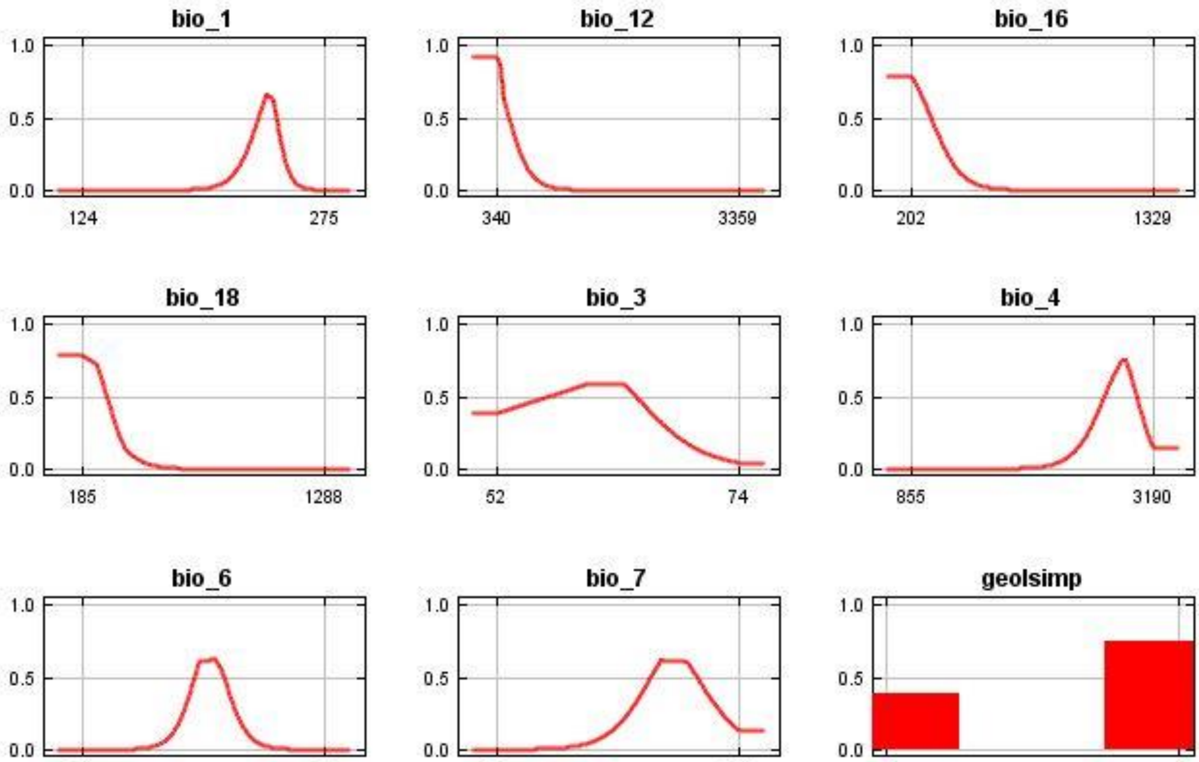


Figure 15. Plots of the response curves representing the predicted suitability (Y axis) in relation to a selected variable (X axis): BIO1, 3, 4, 6, 7, 12, 16, 18.

Prediction and kernel density combination

When the predicted suitable conditions are combined with a kernel density, the predicted distribution area for *Capurodendron androyense* is confined to the southern coast of Madagascar, with two spots that show a very high probability (red) (Figure 16).

Extent of occurrence with predicted distribution

The polygons englobing the former pixels englobing a probability of presence higher the 75% of the total are situated along the southern coast, in the north and in the north-west respectively for *Capurodendron androyense, nodosum* and *pervillei* (Figure 17).

The area of these polygons, or extent of occurrence, is respectively of 33’724.17 km^2, 5’380.67 km^2 and 46204.17 km^2.

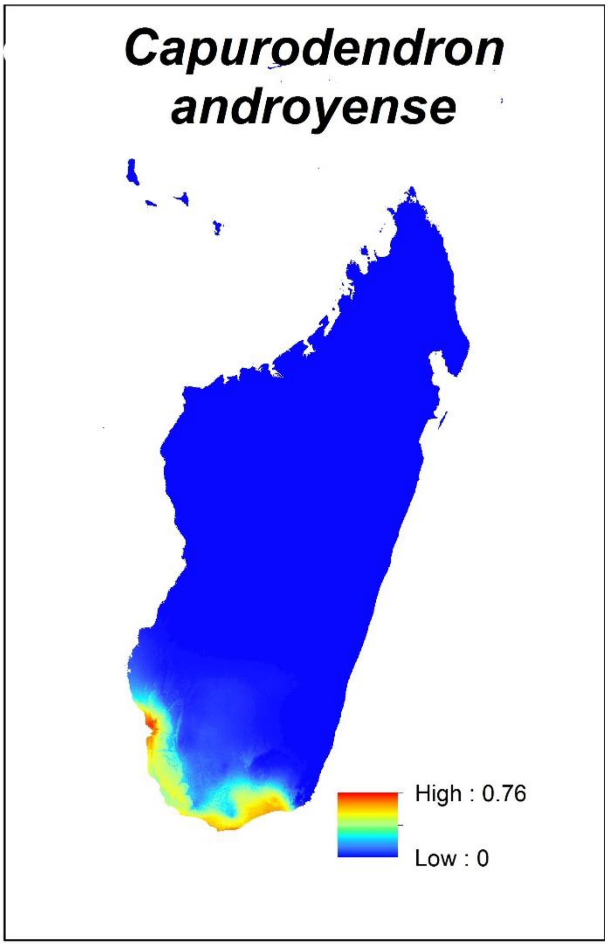


Figure 16. Image of the predicted suitable conditions for Capurodendron androyense on Madagascar, combined with a kernel density of radius =5, with high probability in red and low in blue.

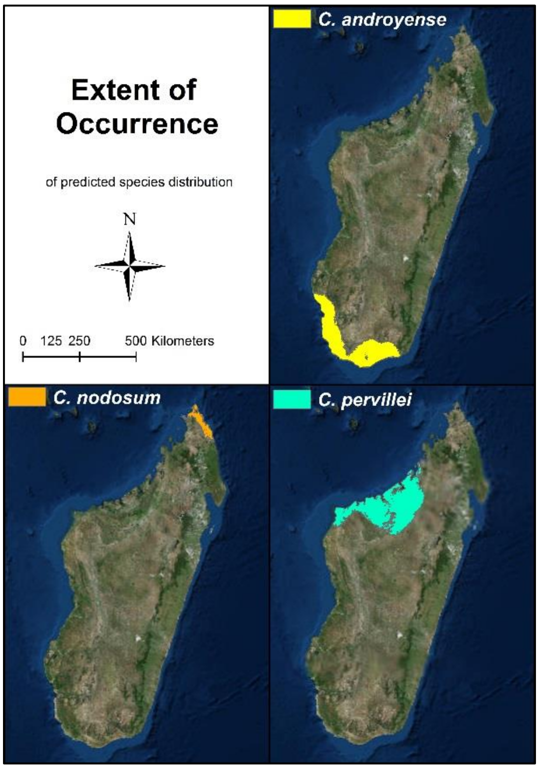


Figure 17. Maps of the extent of occurrence, calculated for each of the three species Capurodendron androyense (yellow), nodosum (orange) and pervillei (light blue) with predicted distribution.

4. DISCUSSION

The extent of known occurrences
The polygons are indeed the smallest polygons possible that contain all of the species occurrences (Figure 10).

The kernel densities
The radius of 5 appears big, maybe too much so, and one of the reasons is that the layer is in geographical coordinates, and not in projected, which make the units angles, and not meters. This is a mistake that needs to be corrected. It does not however prevent it from being useful, just makes it an even more fictitious situation (the radius size already having been chosen arbitrarily).

Statistical analysis
No percentage of the sample records was put aside for testing by MaxEnt, due to their small number. The model thus could not calculate the omission on test samples (Figure 13), the AUC on the test data (**Erreur ! Source du renvoi introuvable.**) and the statistical significance of the prediction using a binomial test of omission.

With an area under the ROC curve (AUC) of 0.974, which means that the model produced predictions that are significantly better than random (AUC = 0.5), so the MaxEnt show a good ability to discriminate suitable from unsuitable areas for the species. In addition, it seems that with presence-only data, the AUC is always smaller than 1 (Phillips, Anderson et Schapire, Maximum entropy modeling of species geographic distributions. 2006).

Variable importance
It is accepted that the geology variable, locating areas with limestone and those without, is uncorrelated to the bioclimatic variables.
However, the bioclimatic variable are extrapolated from a same set of precipitation and temperature measures, and although they were chosen according to the results of a PCA, as explained above, correlation between them cannot be excluded. Also, some already showed a negative correlation in the PCA. For example, BIO 6 and 7, whose contributions are null (Table 1), are respectively negatively correlated to BIO12 and BIO4.

The jackknife of variable importance (see Figure 14) shows that BIO12, BIO18 and more particularly BIO16, are very useful by themselves for the estimation of the distribution of *Capurodendron androyense, because they achieve a lot of gain (dark blue bars)*. On the contrary, BIO3 and the geology, if taken by themselves, do not allow a good fit for the training data.

The predicted suitability decreases with BIO12, 16 and 18 (Figure 16), and briefly increases with BIO1, 4, 6 and 7. With the geology, it is at its highest in zone 1.
As another example, we see that predicted suitability is negatively correlated with annual precipitation (pre6190_ann), if all other variables are held fixed. In other words, once the effect of all the other variables has already been accounted for, the marginal effect of increasing annual precipitation is to decrease predicted suitability.

Prediction and kernel density combination
Combining the MaxEnt output to a kernel density limits the possible suitable areas by a determined radius, which was fixed at 5 in our case (Figure 16). In comparison to the raw output (Figure 12), the predicted zone stops at a radius of 5 from the occurrence points, and, for instance, the area on the northern coast is now excluded.
The kernel density information can actually be compared to a measure of the species dispersion distance, by pollen dissemination for example. However, the choice of its radius is either random, like in this case, or needs previous knowledge about the pollination process.

As mentioned above, the kernel density data is not in the right coordinate system and the chose radius is not in the right units (meters). However, it anyway allows to measure the effect it has when combined with the predictions anyway.

Predicted extent of occurrence

With a fictive threshold at 25%, the thus selected probable suitable areas were kept to create a polygon whose area is the extent of occurrence (EOO) (Figure 17) whose area is necessary amongst other parameters, to calculate a Red List Index for the species. However, the polygon per specie should not contain any holes and this is a task that should be accomplished before using the area value.

However, as expected, the values of the EOOs calculated with predicted distribution are higher than with known occurrence data (Table 2).

Table 2. Comparison of the EOOs: with known occurrences and with predicted distribution.

	EOO with known occurrences	**EOO with predicted distribution**
C. androyense	43’740.46 km^2	33’724.17 km^2
C. nodosum	2’880.93 km^2	5’380.67 km^2
C. pervillei	23’118.59 km^2	46’204.17 km^2

On one hand, the known occurrences reflect only a small part of the species distribution, but on the other hand, the potential distribution is usually not realized for various reasons, such as geographic barriers, human influence or biotic interactions, so the real EOO would actually be situated somewhere in between.

5. CONCLUSION

The species modelling with MaxEnt is a good alternative for presence-only data such as ours. The parameters in this case were for the most part set on default and the model yields fairly easily interpreted results. It would however be interesting to change the random test percentage setting to set aside a number of species occurrences for testing.
As mentioned by Mackey and Lindenmayer, “Bioclimatic and soil-type variables measure availability of the fundamental primary resources of light, heat, water and mineral nutrients.” The inclusion of bioclimatic variables is supported, but the choice of which to select is debatable and should be studied more closely to make a more informed choice, maybe by including expert opinion. The geology layer is very simple, being composed of only two classes, presence or absence of limestone in the sedimentary rocks, but seems to be well included in the analysis which proves its usefulness. In addition, there may be other environmental variable that have not yet taken into account but which could be fundamental for explaining the distribution of a specific species.
Anyway, the realized distribution of the species can only be estimated afterwards, by incorporating knowledge specific or not to the species in question. However, additional data such as landcover classification may pose a problem with herbaria specimen, since the actual landcover does not necessarily correspond to what was in place at the collection time (Phillips, Anderson et Schapire 2006).

The integration of other parameter for the modelling, such as biotic interactions (competitions, symbiosis…) should also be taken into account when predicting a plant species distribution pattern. However, such information is not always easily accessible as it necessitates a lot of field work, but could be added to the result after acquisition (Austin 2002, Guisan et Thuiller 2005). The kernel density data is a simplified way of taking into account other parameters for the modelling than climatic data, since it can be considered as a simplified representation of the dispersion radius of a plant species. However, it would seem that in this work, the chosen radius might be exaggerated. Again, a more informed choice needs to be made.

In the future, the choice of the threshold of the withheld suitable area, set here at 25%, needs to be more complex by, for example, calculating the best Kappa (Fielding et Bell 1997). Also, the final EOO polygon should be whole and not contain any holes before having its area calculated, as mentioned above. It is however a step closer towards the calculation of the Red List Index for the species.

It goes without saying that the analysis conducted in this work was successfully applied to the other species of *Capurodendron*, when the known occurrences proved to be enough, but their results are not shown here.

In conclusion, although some parameters need to be looked at more closely, the overall methodology yields good results and will be used within the framework of the aforementioned Master thesis.

6. REFERENCES

Anderberg, A. A., and U. Swenson. “Evolutionary lineages in Sapotaceae (Ericales): A cladistic analysis based on ndhF sequence data.” *Int. J. Pl. Sci.*, 2003: 763-773.

Aubréville, A. “Sapotacées.” In *Flore de Madagascar et des Comores*, by H. Humbert, 1-128. Paris: Museum National d'Histoire Naturelle, 1974.

Austin, M. P. “Spatial projection of species distribution: an itnerface between ecological theory and statistical modelling.” *Ecological Modelling*, 2002: 101-118.

Aydin, Z., T. Marcussen, A. S. Ertekin, and B. Oxelman. “Marginal Likelihood Estimate Comparisons to Obtain Optimal Species Delimitations in Silene sect. Cryptoneurae (Caryophyllaceae).” *PLoS ONE*, 2014.

Buerki, S., M. W. Callmander, S. Bachman, J. Moat, J.-N. Labat, and F. Forest. “Incorporating evolutionary history into conservation planning in biodiversity hotspots.” *Phli. Trans. R. Soc. B*, 2015.

Callmander, M. W., et al. “Identification of priority areas for plant conservation in Madagascar using Red List criteria: rare and threatened Pandanaceae indicate sites in need of protection.” *Oryx*, April 2007: 168–176.

Callmander, M. W., G. E. Schatz, and P. P. Lowry II. “IUCN Red List assessment and the Global Strategy for Plant Conservation: taxonomists must act now.” *Biodiversity & Conservation*, November 2005: 1047-1050.

Cornet, A. “Essai de cartographie bioclimatique à Madagascar.” *ORSTOM*, 1974: 1-55.

Cornet, A., and J.-L. Guillaumet. “Divisions floristiques et étages de végétation à Madagascar.” *Cahiers de l'ORSTOM*, 1976: 35-42.

De Wit, Maarten J. “MADAGASCAR: Head It's a Continent, Tails It's an Island.” *Annual Review of Earth and Planetary Sciences*, 1 May 2003: 213-248.

Degnan, J. H., and N. A. Rosenberg. “Gene tree discordance, phylogenetic inference and the multispecies coalescent.” *Trends in Ecology and Evolution*, 2009: 332-340.

Degnan, J. H., M. DeGiorgio, D. Bryant, and N. A. Rosenberg. “Properties of Consensus Methods for Inferring Species Trees from Gene Trees.” *Properties of Consensus Methods for Inferring Species Trees from Gene Trees*, 2009: 35–54.

Devey, D. S., F. Forest, F. Rakotonasolo, P. Ma, B. T.M. Dentinger, and S. Buerki. “A snapshot of extinction in action: The decline and imminent demise of the endemic Eligmocarpus Capuron (Caesalpinioideae, Leguminosae) serves as an example of the fragility of Madagascan ecosystems.” *South African Journal of Botany*, 2013.

Drummond, A. J., and A. Rambaut. “BEAST: Bayesian evolutionary analysis by sampling trees.” *BMC Evol. Biol.*, 2007.

Du Puy, D., and J. Moat. “A refined classification of the primary vegetation of Madagascar based on the underlying geology: using GIS to map its distribution and to assess its conservation status.” In *Proceedings of the International Symposium on the Biogeography of Madagascar*, by W.R. Lourenço (editor), 205-218 + 3 maps. Paris: Editions de l’ORSTOM, 1996.

Elith, J., et al. “Novel methods improve prediction of species' distributions from occurrence data.” *Ecography*, 2006: 00-00.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. “A statistical explanation of MaxEnt for ecologists.” *Diversity and Distributions*, 2011: 43-57.

ESRI. 2013. http://www.esri.com/.

Faith, D. P. “Conservation evaluation and phylogenetic diversity.” *Biol. Conserv.*, 1992: 1-10.

Ferrier, S., G. Watson, and J. Pearce. “Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling.” *Biodiversity and COnservation*, 2002: 2275-2307.

Fielding, A. H., and J. F. Bell. “A review of methods for the assessment of prediction errors in conservation presence/absence models.” *Environmental Conservation*, 1997: 38-49.

Gautier, L., and S. M. Goodman. “Introduction to the FLora of Madagascar.” In *The Natural History of Madagascar*, by S. M. Goodman and J. P. Benstead, 229-250. Chicago: The University of Chicago Press, 2003.

Gautier, L., Y. Naciri, A. A. Anderberg, J. E. E. Smedmark, R. Randrianaivo, and U. Swenson. “A new species, genus and tribe of Sapotaceae, endemic to Madagascar.” *Taxon*, 2013: 972–983.

Guisan, A., and N. Zimmermann. “Predictive habitat distribution models in ecology.” *Ecological Modelling*, 2000: 147-186.

Guisan, A., and W. Thuiller. “Predicting species distribution: offering more than simple habitat models.” *Ecology Letters*, 2005: 993-1009.

Hannah et al., L. “Climate change adaptation for conservation in Madagascar.” *Biol. Lett.*, 2008: 590-594.

Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. “Very high resolution interpolated climate surfaces for global land areas.” *International Journal of Climatology* (John Wiley & Sons, Ltd.) 25, no. 15 (September 2005): 1965-1978.

Humbert, H. “Les territoires phytogéographiques de Madagascar.” *Année biologique*, 1955: 439-448.

Humbert, H. “Description des types de végétation.” In *Notice de la carte de Madagascar*, by H. Humbert and G. Cours-Dame, 46-78. Pondichéry: Insitut Français de Pondichéry, 1965.

Humbert, Henri. “La destruction d'une flore insulaire par le feu. Principaux aspects de la végétation à Madagascar. Documents photographiques et notices.” *Mémoires de l'Académie Malgache - Fascicule V*, 1927.

Isaac, N. J. B., S. T. Turvey, B. Collen, C. Waterman, and J. E. M. Baillie. “Mammals on the EDGE: Conservation Priorities Based on Threat and Phylogeny.” *PLoS ONE*, 2007.

IUCN Plant Conservation Sub Comittee. “The Global Strategy for Plant Conservation (GSPC) 2011-2020 …...and what it means for you.” IUCN Plant Conservation Sub Comittee Factsheet, IUCN PLANT CONSERVATION SUB COMMITTEE FACTSHEET, 2011.

IUCN Species Survival Commission. *IUCN Red List Categories and Criteria: Version 3.1. Second edition.* Gland, Switzerland and Cambridge, UK: IUCN, 2000.

IUCN Standards and Petitions Subcommittee. *Guidelines for Using the IUCN Red List Categories and Criteria. Version 11.* Prepared by the Standards and Petitions, 2014.

Kembel, S. W., et al. “Picante: R tools for integrating phylogenies and ecology.” *Bioinformatics*, 2010: 1463–1464.

Koechlin, J., J.-L. Guillaumet, and P. Morat. *Flore et végétation de Madagascar.* J. Cramer, Vaduz, 1974.

Lehmann, A., J. McC. Overton, and J. R. Leathwick. “GRASP: generalized regression analysis and saptial prediction.” *Ecological Modelling*, 2002: 189-207.

Lehmann, A., J. R. Leathwick, and J. McC. Overton. *GRASP v.3.0 User's Manual.* Swiss Centre for Faunal Cartography, Switzerland, 2004.

Madagascar Catalogue. *Catalogue of the Vascular Plants of Madagascar.* Missouri Botanical Garden, St. Louis, U.S.A. & Antananarivo, Madagascar. 2013. http://www.efloras.org/madagascar (accessed April 2015).

Maggini, R., A. Lehmann, and N. E. Zimmermann. “Improving generalized regression analysis for the spatial prediction of forest communities.” *Journal of Biogeography*, 2006: 21.

Miller, J. S., et al. “Toward Target 2 of the Global Strategy for Plant Conservation: An Expert Analysis of the Puerto Rican Flora to Validate New Streamlined Methods for Assessing Conservation Status.” *Annals of the Missouri Botanical Garden*, 2013: 199-205.

Myers, N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, and J. Kent. “Biodiversity hotspots for conservation priorities.” *Nature*, 24 Février 2000: 853-858.

Naciri, Y., and P. Linder. “Species delimitation and relationships: The dance of the seven veils.” *Taxon*, 2015: 14.

Pennington, T. D. *The Genera of Sapotaceae.* Whitstable, Kent: Whitstable Litho Ltd., 1991.

Perrier de la Bâthie, H. "La végétation malgache." *Annales du Musée Colonial de Marseille*, 1921: 1-268.

—. *Biogéographie des plantes de Madagascar.* Paris: Société d'éditions géographiques, maritimes et coloniales, 1936.

Phillips, S. J., M. Dudík, and R. E. Schapire. "A Maximum Entropy Approach to Species Distribution Modeling." *Proceedings of the Twenty-First International Conference on Machine Learning*, 2004: 655-662.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. "Maximum entropy modeling of species geographic distributions." *Ecological Modelling*, 2006: 231-259.

Purvis, A., P.-M. Agapow, J. L. Gittleman, and G. M. Mace. "Nonrandom Extinction and the Loss of Evolutionary History." *Science*, 2000: 328-330.

Rakotoarinivo, M., J. Dransfield, S. P. Bachman, J. Moat, and W. J. Baker. "Comprehensive Red List Assessment Reveals Exceptionally High Extinction Risk to Madagascar Palms." *PLoS ONE*, July 2014.

Rivers, M. C., L. Taylor, N. A. Brummit, T. R. Meagher, D. L. Roberts, and E. N. Lughadha. "How many herbarium specimens are needed to detect threatened species?" *Biological Conservation*, 2011.

Rivers, M. C., S. P. Bachman, T. R. Meagher, E. N. Lughadha, and N. A. Brummit. "Subpopulations, locations and fragmentation: applying IUCN red list criteria to herbarium specimen data." *Biodiversity Conservation*, 2010: 2071–2085.

Sato, H. "Frugivory and seed dispersal by brown lemurs in a Malagasy tropical dry forest." *Biotropica*, 2012: 479-488.

Schatz, G. E. "Plants on the IUCN Red List: setting priorities to inform conservation." *Trends in Plant Science*, 2009: 638-642.

Schatz, G. E., and M. Lescot. *Gazetteer to Malagasy Botanical Collection Localities.* 2003. http://www.mobot.org/mobot/research/madagascar/gazetteer/.

Schrest, W., et al. "Hotspots and the conservation of evolutionary history." *PNAS*, 2002: 2067–2071.

Smedmark, J. E. E., U. Swenson, and A. A. Anderberg. "Accounting for variation of substitution rates through time in Bayesian phylogeny reconstruction of Sapotoideae (Sapotaceae)." *Molec. Phylgen. Evol.*, 2006: 706-721.

Swenson, U., and A. A. Anderberg. "Phylogeny, character evolution and classification of Sapotaceae (Ericales)." *Cladistics*, 2005: 101-130.

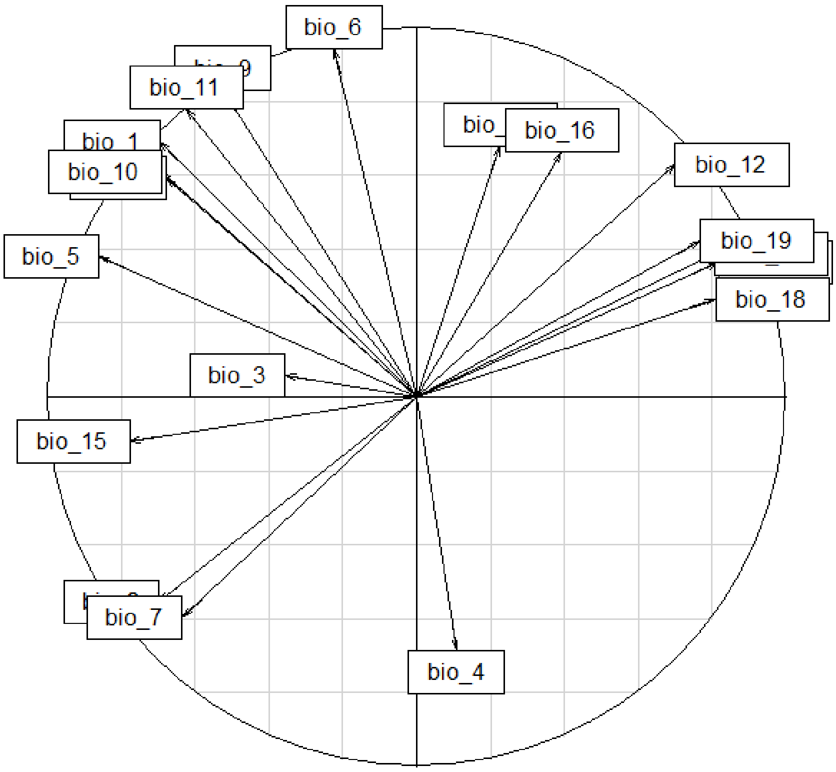
The R Foundation. *The R Project for Statistical Computing.* 2015. http://www.r-project.org/.

Thuiller, W., D. Georges, and R. Engler. *Package.* 2015.

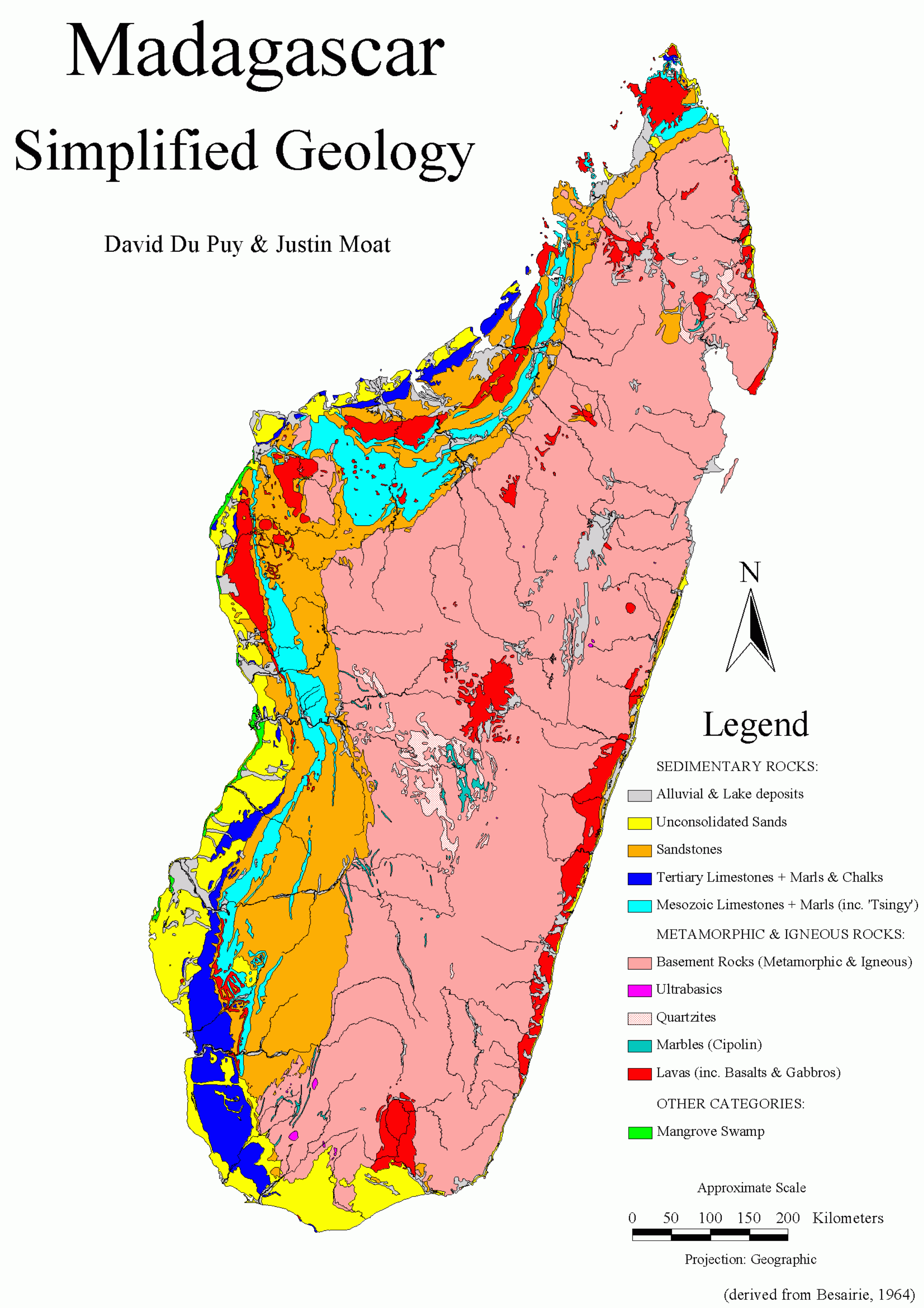
Willis, F., J. Moat, and A. Paton. "Defining a role for herbarium data in Red List assessments: a case study of Plectranthus from eastern and southern tropical Africa." *Biodiversity and Conservation*, 2003: 1537–1552.

Wulff, A. S., et al. "Conservation Priorities in a Biodiversity Hotspot: Analysis of Narrow Endemic Plant Species in New Caledonia." *PLoS ONE*, 2013.

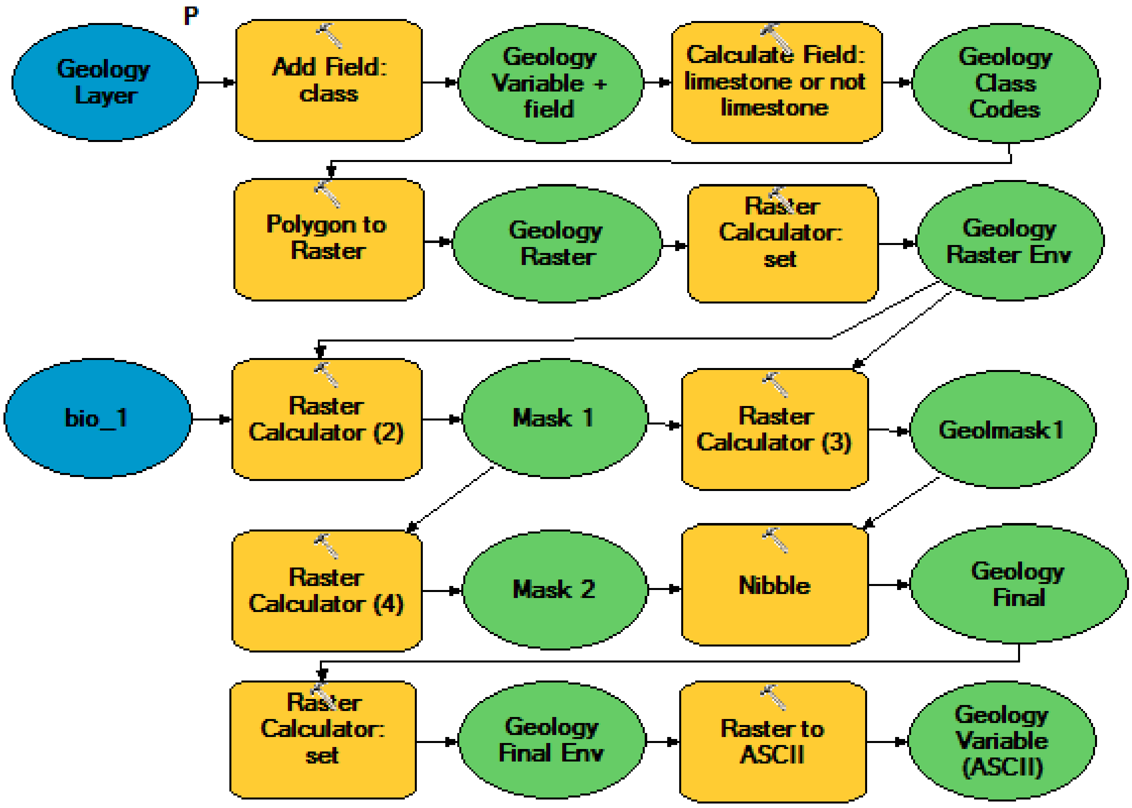
7. APPENDIX



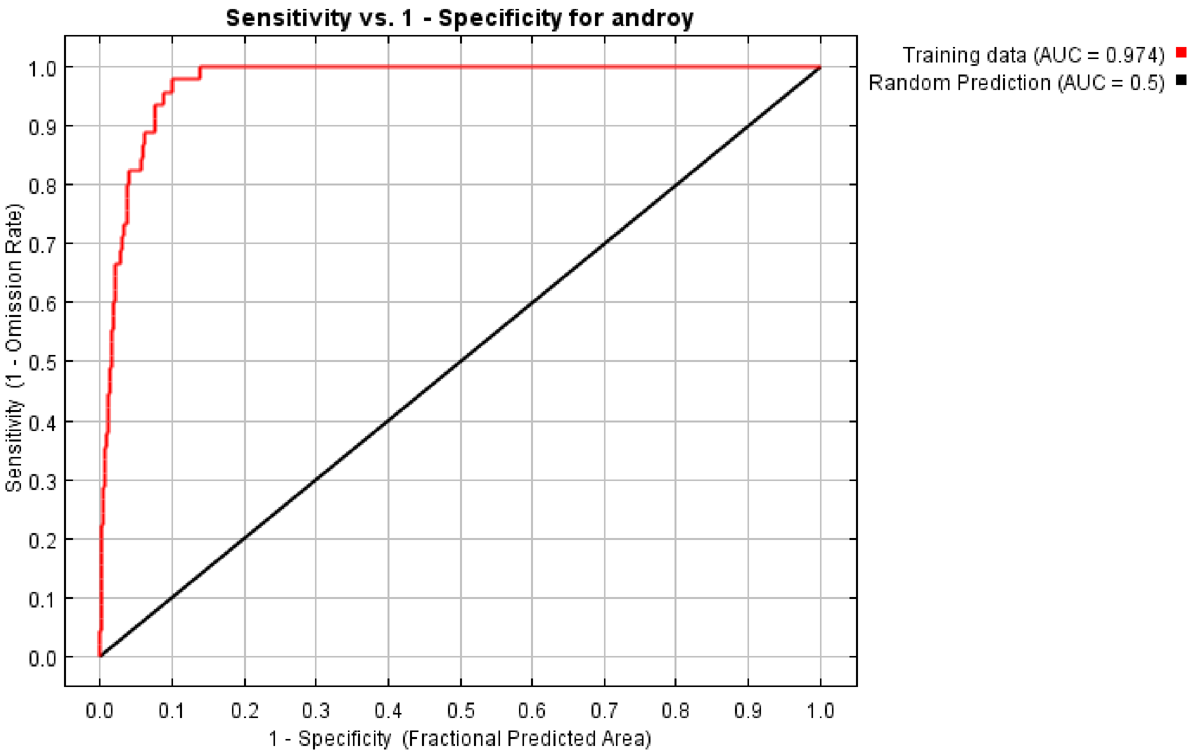
Annex 1. Graph of the results of the Principal Components Analysis (PCA) conducted with R.



Annex 2. Simplified geology map by Du Puy & Moat (1996).



Annex 3. Model of the processing steps taken to process the simplified geology and make it into an input for MaxEnt, made with the Model Builder application available in the ArcMap interface.



Annex 4. Sensitivity vs. 1-Specificity for Capurodendron androyense.

```
# ---------------------------------------------------------------------------
# KernelDensityScript.py
# Created on: 2015-05-27 00:57:54.00000
#   (generated by ArcGIS/ModelBuilder)
# Description:
# ---------------------------------------------------------------------------

# Import arcpy module
import arcpy

from arcpy import env
from arcpy.sa import *

# Check out any necessary licenses
arcpy.CheckOutExtension("spatial")

# Set Geoprocessing environments
arcpy.env.outputCoordinateSystem = "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',6378137.0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.0174532925199433]],VERTCS['Unknown VCS',VDATUM['Unknown'],PARAMETER['Vertical_Shift',0.0],PARAMETER['Direction',1.0],UNIT['Meter',1.0]]"
arcpy.env.extent = arcpy.Extent(41.9916782444343, -27.0083316126838, 52.0000120997429, -8.99999734014273)
arcpy.env.cellSize = "8.33333376795054E-03"

##Local variable: WorldClim
bioclim = ["bio_1", "bio_2", "bio_3", "bio_4", "bio_5", "bio_6", "bio_7", "bio_8", "bio_9", "bio_10", "bio_11", "bio_12", "bio_13", "bio_14", "bio_15", "bio_16", "bio_17", "bio_18", "bio_19"]
for NBBIO in bioclim:
    bios = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\WorldClim\\bio\\"+NBBIO
    bios2 = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\EnvLayers\\RasterFinal\\"+NBBIO
    bios_ASC = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\EnvLayers\\"+NBBIO+".ASC"

    # Process: Clip
    arcpy.Clip_management(bios, "42 -27 52 -9", bios2, "", "", "NONE")

    # Process: Raster to ASCII
    arcpy.RasterToASCII_conversion(bios2, bios_ASC)

## Local variables: Hull
Capurodendron = "Species\\Capurodendron"
Output = "Species\\Capurodendron"

sp = ["androy", "ankara", "anton", "apoll", "bak_var_a", "bak_var_b", "bak_temp", "cost", "delph", "gracili", "grev", "grev_term", "indet", "ludiif", "mada", "mand", "micr_andr", "micro", "nod", "nod_perr", "perr", "perr_cf", "perr_obl", "perv", "perv_cf", "pseudo", "rubro", "rubro_cf", "ruf", "ruf_cf", "saka", "sp", "aff_perv", "sp_1", "sp_10", "sp_11", "sp_12", "sp_13", "sp_2", "sp_3", "sp_4", "sp_6", "sp_7", "sp_8", "sp_9", "suar", "tamp_cf", "tam_var_a", "tam_var_t", "term"]
for SP in sp:
    hull = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\Hull\\h"+SP+".shp"
    hull_shp = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\Hull\\HullArea\\h"+SP+".shp"
```

```
    # Process: Select Layer By Attribute
    arcpy.SelectLayerByAttribute_management(Capurodendron, "NEW_SELECTION",
"\"Species\" = '"+SP+"'")

    # Process: Minimum Bounding Geometry
    arcpy.MinimumBoundingGeometry_management(Output, hull, "CONVEX_HULL", "ALL",
"", "MBG_FIELDS")

    # Process: Calculate Areas
    arcpy.CalculateAreas_stats(hull, hull_shp)


## Local variables: 2KernelScriptTot
Capurodendron = "Species\\Capurodendron"
Output = "Species\\Capurodendron"

sp = ["androy", "nod", "perv", "rubro"]
for SP in sp:
    k1_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k1_"+SP
    k2_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k2_"+SP
    k3_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k3_"+SP
    k4_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k4_"+SP
    k5_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k5_"+SP
    k6_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k6_"+SP
    k7_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k7_"+SP
    k8_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k8_"+SP
    k9_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k9_"+SP
    k10_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k10_"+SP
    k20_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\k20_"+SP

    # Process: Select Layer By Attribute
    arcpy.SelectLayerByAttribute_management(Capurodendron, "NEW_SELECTION",
"\"Species\" = '"+SP+"'")

    # Process: Kernel Density
    arcpy.gp.KernelDensity_sa(Output, "NONE", k1_androy, "8.33333376795054E-03", "1",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k2_androy, "8.33333376795054E-03", "2",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k3_androy, "8.33333376795054E-03", "3",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k4_androy, "8.33333376795054E-03", "4",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k5_androy, "8.33333376795054E-03", "5",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k6_androy, "8.33333376795054E-03", "6",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k7_androy, "8.33333376795054E-03", "7",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k8_androy, "8.33333376795054E-03", "8",
"SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k9_androy, "8.33333376795054E-03", "9",
"SQUARE_MAP_UNITS")
```

```
    arcpy.gp.KernelDensity_sa(Output, "NONE", k10_androy, "8.33333376795054E-03", "10", "SQUARE_MAP_UNITS")
    arcpy.gp.KernelDensity_sa(Output, "NONE", k20_androy, "8.33333376795054E-03", "20", "SQUARE_MAP_UNITS")

## Local variables: 3Kernel0a1
sp = ["k1_androy", "k1_nod", "k1_perv", "k1_rubro", "k2_androy", "k2_nod", "k2_perv", "k2_rubro", "k3_androy", "k3_nod", "k3_perv", "k3_rubro", "k4_androy", "k4_nod", "k4_perv", "k4_rubro", "k5_androy", "k5_nod", "k5_perv", "k5_rubro", "k6_androy", "k6_nod", "k6_perv", "k6_rubro", "k7_androy", "k7_nod", "k7_perv", "k7_rubro", "k8_androy", "k8_nod", "k8_perv", "k8_rubro", "k9_androy", "k9_nod", "k9_perv", "k9_rubro", "k10_androy", "k10_nod", "k10_perv", "k10_rubro", "k20_androy", "k20_nod", "k20_perv", "k20_rubro"]
for SP in sp:

    Kernels = Raster("C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels\\"+SP)
    Kernels_ASC = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Envlayers\\"+SP+".ASC"

    ## Process: Get Raster Properties
    MaxVal = arcpy.GetRasterProperties_management(Kernels, "MAXIMUM")
    MaxValue = MaxVal.getOutput(0)
    print(MaxValue)
    type(MaxValue)

    ## Process: Raster Calculator
    Kernels0a1 = Kernels/float(MaxValue)
    Kernels0a1.save("C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels0a1\\"+SP)

    ## Process: Raster to ASCII
    arcpy.RasterToASCII_conversion(Kernels0a1, Kernels_ASC)
```

Annex 5. Python script of the data preparation in ArcGIS.

```
# ---------------------------------------------------------------------------
# Kernel0a1Script.py
# Created on: 2015-06-08 22:55:28.00000
#   (generated by ArcGIS/ModelBuilder)
# Description:
# ---------------------------------------------------------------------------

# Import arcpy module
import arcpy

from arcpy import env
from arcpy.sa import *

## Check out any necessary licenses
arcpy.CheckOutExtension("spatial")

## Set Geoprocessing environments
arcpy.env.outputCoordinateSystem = "GEOGCS['GCS_WGS_1984',DATUM['D_WGS_1984',SPHEROID['WGS_1984',6378137.
```

```
0,298.257223563]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.0174532925199433]],VERT
CS['Unknown
VCS',VDATUM['Unknown'],PARAMETER['Vertical_Shift',0.0],PARAMETER['Direction',1.0],
UNIT['Meter',1.0]]"
arcpy.env.extent = arcpy.Extent(41.9916782444343, -27.0083316126838,
52.0000120997429, -8.99999734014273)
arcpy.env.cellSize = "8.33333376795054E-03"

# Local variables:
sp = ["androy", "nod", "perv", "rubro"]
for SP in sp:
   ## Pour ASCII To Raster
   androy_asc = "C:\\Users\\Rhéa
Garratt\\Documents\\UNIGE\\4Sp\\Modelling\\Resultat\\"+SP+".asc"
   androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Modelling\\Resultat\\"+SP
   k1_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels0a1\\k1_"+SP
   k5_androy = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Kernels0a1\\k5_"+SP

   # Process: ASCII To Raster
   Rasters = arcpy.ASCIIToRaster_conversion(androy_asc, androy, "FLOAT")
   Rastersp = Rasters.getOutput(0)
   print(Rastersp)

   # Process: Raster Calculator
   MAxK_androy = k1_androy*Raster(Rastersp)
   MxK5_androy = k5_androy*Raster(Rastersp)
   MAxK_androy.save("C:\\Users\\Rhéa
Garratt\\Documents\\UNIGE\\4Sp\\MaxTrend\\k1_"+SP)
   MxK5_androy.save("C:\\Users\\Rhéa
Garratt\\Documents\\UNIGE\\4Sp\\MaxTrend\\k5_"+SP)

## Local variables:

sp = ["k1_androy", "k1_nod", "k1_perv", "k1_rubro", "k5_androy", "k5_nod", "k5_perv",
"k5_rubro"]
for SP in sp:

   MaxTrend = Raster("C:\\Users\\Rhéa
Garratt\\Documents\\UNIGE\\4Sp\\MaxTrend\\"+SP)

   ## Process: Get Raster Properties
   MaxVal = arcpy.GetRasterProperties_management(MaxTrend, "MAXIMUM")
   MaxValue = MaxVal.getOutput(0)
   print(MaxValue)
   type(MaxValue)

   ## Process: Raster Calculator
   Seuil = 0.25*float(MaxValue)
   print(Seuil)
   k1_androy2 = MaxTrend > Seuil
   k1_androy2.save("C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Seuil\\"+SP)

   # Process: Raster to Polygon
   Seuil_shp = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Seuil\\"+SP+".shp"
   arcpy.RasterToPolygon_conversion(k1_androy2, Seuil_shp, "SIMPLIFY", "VALUE")
```

```
# Process: Calculate Areas
calcarea = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Seuil\\Area\\"+SP+".shp"
arcpy.CalculateAreas_stats(Seuil_shp, calcarea)

# Process: Dissolve
dissolve_shp = "C:\\Users\\Rhéa Garratt\\Documents\\UNIGE\\4Sp\\Seuil\\Dissolve\\"+SP+".shp"
arcpy.Dissolve_management(calcarea, dissolve_shp, "GRIDCODE", "F_AREA SUM", "MULTI_PART", "DISSOLVE_LINES")
```

Annex 6. Python script of the data processing of the MaxEnt output.