

RÉPUBLIQUE DU SÉNÉGAL

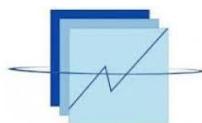


Un Peuple - Un But - Une Foi

Ministère de l'Economie, du Plan et de la Coopération



LA RECHERCHE AGRONOMIQUE
POUR LE DÉVELOPPEMENT



ANSO

Agence Nationale de
la Statistique et de la Démographie



Centre de Suivi Ecologique



ensaé

École Nationale de la Statistique et de l'Analyse Économique

Mémoire de fin d'étude

MODÉLISATION DE LA DISTRIBUTION DES PRINCIPALES ESPÈCES LIGNEUSES DANS
DEUX PARCS AGROFORESTIERS DU BASSIN ARACHIDIER SÉNÉGALAIS

Rédigé par :

Aboubacar HEMA

Élève Ingénieur des Travaux Statistiques en quatrième année

Encadreur Académique

Dr. Souleymane DIAKITE

Ingénieur Statisticien Economiste

Maître de stage

Dr Louise LEROUX

UR AIDA, Cirad

Encadreur

M Babacar NDAO

CSE

Mars - Juillet 2020

Modélisation de la distribution des principales espèces ligneuses dans deux parcs agroforestiers du bassin arachidier sénégalais.

Document rédigé par :

Aboubacar HEMA

Élève Ingénieur des Travaux Statistiques, ITS4.

Sous l'encadrement de :

Encadreur Académique

Dr. Souleymane DIAKITE

Ingénieur Statistique Economiste

Maître de stage

Dr Louise LEROUX

UR AIDA, Cirad

Encadreur

M Babacar NDAO

CSE

AVERTISSEMENT

L'Ecole Nationale de la Statistique et de l'Analyse Économique (ENSAE), le Centre de Suivi Ecologique (CSE) et le Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) n'entendent donner aucune approbation ni improbation aux opinions émises dans ce document. Nous, auteurs de ce document, assumons l'entièvre responsabilité des propos tenus et des opinions émises.

AVANT-PROPOS

L'Ecole Nationale de la Statistique et de l'Analyse Economique (ENSAE) fait partie du Réseau des Écoles Africaines de Statistiques (RESA) qui intervient dans la formation de cadres statisticiens africains. Elle forme des Ingénieurs Statisticiens Economistes (ISE), des Ingénieurs des Travaux Statistiques (ITS) et des Techniciens Supérieurs de la Statistique (TSS). Le stage de fin de formation des élèves ITS est obligatoire et s'effectue sur une durée minimum de quatre mois. Il est sanctionné par la rédaction et la soutenance d'un mémoire de recherche.

C'est dans ce cadre que nous avons effectué notre stage au CSE, sur une période de six mois. Ce mémoire, qui est écrit à cet effet, porte sur la "**Modélisation de la distribution des principales espèces ligneuses dans deux parcs agroforestiers du bassin arachider sénégalais**".

REMERCIEMENTS

Il nous plaît en ce moment d'adresser nos remerciements à tous ceux qui, d'une manière ou d'une autre, ont contribué à l'aboutissement de ce travail.

Qu'il nous soit permis, à travers ces quelques lignes, d'exprimer avant tout notre profonde gratitude à l'État burkinabé, à travers le CIOSPB et l'INSD, pour avoir accepté de financer notre formation de la première à la quatrième année.

La justesse et la rigueur des méthodes utilisées, des conditions de travail favorables et un bon relationnel dans la structure d'accueil constituent des facteurs décisifs à la bonne réalisation du stage de fin d'études sanctionné par le présent mémoire.

Nous tenons donc à remercier tout le corps professoral de l'ENSAE pour les sacrifices réalisées afin de nous offrir une formation de qualité. Nous témoignons ici notre reconnaissance au directeur M. Abdou DIOUF, notre responsable de filière Dr. Souleymane FOFANA, à l'ancien coordonnateur des études de l'ENSAE M. Mady DANSOKHO, au chef d'unité de formation des Ingénieurs Statisticiens Economistes (ISE) M. Idrissa DIAGNE ainsi que M. Mamadou BALDE, M. Oumar SENE et M. Oumar DIOP.

Nous tenons particulièrement à remercier notre encadreur Dr. Souleymane DIAKITE, pour non seulement sa grande disponibilité mais également tout ce qu'il nous a apporté durant notre formation.

Nous remercions également au Directeur du CSE Dr Assize TOURE, pour avoir autorisé notre stage au sein de sa structure. Nous remercions particulièrement nos maîtres de stage Dr Louise LEROUX et M Babacar NDAO qui n'ont ménagé aucun effort pour mettre à notre disposition les outils nécessaires à la réalisation de ce travail.

Un grand merci à nos camarades de la future 9^e promotion des ITS de l'ENSAE. Très chers camarades, **DIEUREUDIEUF**

Sommaire

Remerciements	iv
Sigles et Abbréviations	x
Résumé	xi
Abstract	xii
Introduction Générale	1
1 Revue de la littérature	6
1.1 Modélisation de la distribution des espèces (SDM)	6
1.2 Méthodes de modélisation	7
1.3 Techniques de Modélisation	8
1.4 Contexte biophysique du Sénégal	12
2 Méthodologie	15
2.1 Zone d'étude	15
2.2 Données utilisées	16
2.3 Traitement des données et modélisation de la distribution des espèces	19
2.4 Validation des modèles	28
3 Résultats et Discussion	29
3.1 Résultats	29
3.2 Discussion	54
Conclusion et Perspectives	57
Annexes	i
A Modélisation de la distribution des espèces	ii
B Analyse spatiale descriptive	x

Table des figures

2.1	Zone d'étude dans le parc à <i>Faidherbia albida</i>	15
2.2	Abondance des espèces dans le parc à <i>Faidherbia albida</i>	21
2.3	Abondance des espèces dans le parc à <i>Cordyla pinnata</i>	21
2.4	Analyse factorielle de la niche écologique (ENFA) de la distribution de <i>Faidherbia albida</i>	25
2.5	Analyse factorielle de la niche écologique (ENFA) de la distribution de <i>Cordyla pinnata</i>	26
2.6	Autocorrélation spatiale des covariables dans le parc à <i>Faidherbia albida</i>	27
2.7	Autocorrélation spatiale des covariables dans le parc à <i>Cordyla pinnata</i>	27
2.8	Séparation des données de test et d'entraînement de <i>Faidherbia albida</i>	28
3.1	Abondance des espèces dans le parc à <i>Faidherbia albida</i> suivant le type de sol .	29
3.2	Abondance des espèces dans le parc à <i>Cordyla pinnata</i> suivant le type de sol .	30
3.3	Corrélation entre les prédicteurs dans le parc à <i>Faidherbia albida</i>	32
3.4	Corrélation entre les prédicteurs dans le parc à <i>Cordyla pinnata</i>	33
3.5	Graphique des valeurs propres	36
3.6	Biplot symétrique et colonnes colorées selon leur contribution	36
3.7	Contributions en lignes	37
3.8	Contributions en colonnes	38
3.9	Classification Hiérarchique sur Composantes Principales des espèces suivant le type de sol dans le parc à <i>Faidherbia albida</i>	39
3.10	Graphique des valeurs propres	40
3.11	Biplot symétrique et colonnes colorées selon leur contribution	41
3.12	Contributions en lignes	41
3.13	Contributions en colonnes	42
3.14	Classification Hiérarchique sur Composantes Principales des espèces suivant le type de sol dans le parc à <i>Cordyla pinnata</i>	43
3.15	Variables explicatives de la distribution des espèces étudiées	44
3.16	Cartes de probabilité des espèces étudiées dans le parc à <i>Faidherbia albida</i> issues du modèle MaxEnt	46
3.17	Cartes de probabilité des espèces étudiées dans le parc à <i>Cordyla pinnata</i> issues du modèle MaxEnt	46

3.18	Cartes de présence des espèces étudiées dans le parc à <i>Faidherbia albida</i> issues du modèle MaxEnt	47
3.19	Cartes de présence des espèces étudiées dans le parc à <i>Cordyla pinnata</i> issues du modèle MaxEnt	48
3.20	Réponse écologique de <i>Faidherbia albida</i>	49
3.21	Réponse écologique de <i>Balanites aegyptiaca</i>	49
3.22	Réponse écologique d' <i>Anogeissus leiocarpus</i>	50
3.23	Réponse écologique d' <i>Adansonia digitata</i>	50
3.24	Réponse écologique d' <i>Acacia nilotica</i>	51
3.25	Réponse écologique de <i>Cordyla pinnata</i>	51
3.26	Réponse écologique d' <i>Azadirachta indica</i>	52
3.27	Réponse écologique de <i>Ficus capensis</i>	53
3.28	Réponse écologique de <i>Prosopis africana</i>	53
A.1	Analyse factorielle de la niche écologique (ENFA) de la distribution de <i>Balanites aegyptiaca</i>	ii
A.2	Analyse factorielle de la niche écologique (ENFA) de la distribution d' <i>Anogeissus leiocarpus</i>	iii
A.3	Analyse factorielle de la niche écologique (ENFA) de la distribution d' <i>Adansonia digitata</i>	iii
A.4	Analyse factorielle de la niche écologique (ENFA) de la distribution d' <i>Acacia nilotica</i>	iv
A.5	Analyse factorielle de la niche écologique (ENFA) de la distribution de <i>Azadirachta indica</i>	iv
A.6	Analyse factorielle de la niche écologique (ENFA) de la distribution de <i>Prosopis africana</i>	v
A.7	Analyse factorielle de la niche écologique (ENFA) de la distribution de <i>Ficus capensis</i>	v
A.8	Comparaison des modèles pour <i>Faidherbia albida</i>	vi
A.9	Comparaison des modèles pour <i>Balanites aegyptiaca</i>	vi
A.10	Comparaison des modèles pour <i>Anogeissus leiocarpus</i>	vii
A.11	Comparaison des modèles pour <i>Adansonia digitata</i>	vii
A.12	Comparaison des modèles pour <i>Acacia nilotica</i>	vii
A.13	Comparaison des modèles pour <i>Cordyla pinnata</i>	viii
A.14	Comparaison des modèles pour <i>Azadirachta indica</i>	viii
A.15	Comparaison des modèles pour <i>Ficus capensis</i>	viii
A.16	Comparaison des modèles pour <i>Prosopis africana</i>	ix
B.1	Réprésentation graphique des prédicteurs dans le parc à <i>Faidherbia albida</i> (bio1-bio6)	x

B.2	Réprésentation graphique des prédicteurs dans le parc à <i>Faidherbia albida</i> (bio7-bio12)	xi
B.3	Réprésentation graphique des prédicteurs dans le parc à <i>Faidherbia albida</i> (bio13-bio18)	xi
B.4	Réprésentation graphique des prédicteurs dans le parc à <i>Faidherbia albida</i> (bio19-CLYPPT)	xii
B.5	Réprésentation graphique des prédicteurs dans le parc à <i>Faidherbia albida</i> (ORCDRC-Slo1)	xii
B.6	Réprésentation graphique des prédicteurs dans le parc à <i>Faidherbia albida</i> (Len1-Elev)	xiii
B.7	Réprésentation graphique des prédicteurs dans le parc à <i>Cordyla pinnata</i> (bio1-bio6)	xiii
B.8	Réprésentation graphique des prédicteurs dans le parc à <i>Cordyla pinnata</i> (bio7-bio12)	xiv
B.9	Réprésentation graphique des prédicteurs dans le parc à <i>Cordyla pinnata</i> (bio13-bio18)	xiv
B.10	Réprésentation graphique des prédicteurs dans le parc à <i>Cordyla pinnata</i> (bio19-CLYPPTC)	xv
B.11	Réprésentation graphique des prédicteurs dans le parc à <i>Cordyla pinnata</i> (ORCDRC-Slo1)	xv
B.12	Réprésentation graphique des prédicteurs dans le parc à <i>Cordyla pinnata</i> (Len1-Elev)	xvi

Liste des tableaux

2.1	Variables bioclimatiques	17
2.2	Les variables des propriétés du sol : AfSoilGrids250m	18
3.1	Test du Chi Deux entre les espèces et le type de sol	31
3.2	Test de Joincount de la distribution des espèces	31
3.3	Profile ligne des espèces dans le parc à <i>Faidherbia albida</i>	35
3.4	Profile ligne des espèces dans le parc à <i>Cordyla pinnata</i>	40
3.5	Meilleur modèle selon l’AUC	44
3.6	Meilleur modèle selon l’AUC et avec l’ajout de variables à moyenne résolution spatiale	45
3.7	Le seuil de positivité fixé par le modèle des espèces étudiées	47

Sigles et Abbréviations

- ★ **ANSD** : Agence Nationale de la Statistique et de la Démographie ;
- ★ **AUC** : Area Under the Curve ;
- ★ **BCT** : Boosted Classification Trees ;
- ★ **CIOSPB** :Centre National de l'Information, de l'Orientation Scolaire et Professionnelle et des Bourses ;
- ★ **CNFA** : Climate Niche Factor Analysis ;
- ★ **ENFA** : Ecological Niche Factor Analysis ;
- ★ **ENSAE** : École Nationale de la Statistique et de l'Analyse Économique ;
- ★ **ENSEA** :École Nationale supérieure de Statistique et d'Économie Appliquée ;
- ★ **FAO** : Organisation des Nations Unies pour l'Alimentation et l'Agriculture ;
- ★ **INS** Institut Nationale de la Statistique ;
- ★ **INSD** : Institut Nationale de la Statistique et de la Démographie ;
- ★ **ISSEA** :Institut Sous régional de Statistique et d'Économie Appliquée ;
- ★ **ISE** : les Ingénieurs Statisticiens Économistes ;
- ★ **ITS** :les Ingénieurs de Travaux Statistiques ;
- ★ **MaxEnt** : Maximum d'Entropie ;
- ★ **SDM** : Species Distribution Models
- ★ **TSS** : Techniciens Supérieurs de la Statistique ;

RÉSUMÉ

De nos jours, la prédition de la distribution des espèces est au centre de diverses applications en écologie et joue également un rôle clé dans l'évaluation de l'impact des changements globaux sur les écosystèmes. Les modèles de distributions d'espèces (SDM) sont couramment utilisés pour prédire la présence d'espèces. Dans cette étude, nous avions pour objectif de modéliser la distribution des principales espèces ligneuses dans deux parcs agroforestiers du bassin arachidier sénégalais (le parc à *Faidherbia albida* à Niakhar et le parc à *Cordyla pinnata* à Nioro du Rip). Spécifiquement, il s'agissait de déterminer les facteurs explicatifs de la distribution spatiale des espèces étudiées, évaluer la capacité de prédition de différents algorithmes utilisés dans les modèles de distribution des espèces ligneuses, élaborer des cartes de probabilité de présence des espèces étudiées et mettre en évidence au moyen du meilleur modèle la réponse écologique de chacune des espèces étudiées.

Pour cela, deux bases de données géoréférencées sur les espèces ligneuses, à savoir 9258 arbres à Niakhar et 6611 arbres à Nioro, ont été utilisées. Egalement, différents types de variables environnementales essentiellement dérivées de données et techniques géospatiales ont été utilisés. Les variables explicatives de la distribution des espèces étudiées ont été sélectionnées par une méthode d'analyse factorielle de la niche écologique (ENFA). Pour la modélisation des niches écologiques, nous avons utilisé deux approches : la modélisation avec les blocs spatiaux et la modélisation sans les blocs spatiaux. Pour chaque approche de modélisation, sept modèles ont été testés.

Cette étude révèle que la modélisation avec les blocs spatiaux améliorent de manière générale les modèles de distribution des espèces. Le meilleur modèle parmi les sept testés est MaxEnt. Le modèle MaxEnt a été mis à contribution pour apprécier la réponse écologique des espèces étudiées vis-à-vis des variables climatiques et écologiques. Les résultats de notre étude montrent que les espèces ligneuses répondent différemment aux conditions environnementales et climatiques. Nous avons aussi montré que l'ajout des variables à moyenne résolution spatiale n'améliorent pas toujours les modèles de distribution des espèces. Cela dépend de l'espèce étudiée.

Mots clés : Modèles de distribution des espèces (SDM), espèces ligneuses, MaxEnt, réponse écologique, analyse factorielle des niches écologiques (ENFA).

ABSTRACT

Nowadays, predicting the species distribution is at the focus of various applications in ecology and also has a major significance in assessing the impact of global changes on ecosystems. Species Distribution Models (SDM) are commonly used to predict the occurrence of species. In this study, we aimed to model the distribution of the main woody species in two agroforestry parks in the Senegalese groundnut basin (*Faidherbia albida* in Niakhar and *Cordyla pinnata* in Nioro du Rip). More specifically, the aim was to determine the factors explaining the spatial distribution of the species studied, to evaluate the predictive capacity of different algorithms used in woody species distribution models, to develop probability maps of the presence of the species studied and to highlight the ecological response of each of the species studied using the best model. For this purpose, two geo-referenced databases on woody species, including 9258 trees in Niakhar and 6611 trees in Nioro. In addition, different types of environmental variables derived mainly from geospatial data and techniques were used. Explanatory variables for the species distribution studied were selected by an Ecological Niche Factor Analysis Method (ENFA). Two approaches were used for modelling ecological niches : modelling with spatial blocks and modelling without spatial blocks. Seven models for each modeling approach have been tested. Research shows that modelling with spatial blocks improves species distribution models. MaxEnt is the best of the seven models tested. MaxEnt was used to evaluate the ecological response of the studied species to climatic and ecological variables. Research results show that woody species respond variously to environmental and climatic conditions. Furthermore, we have also shown that the addition of variables with spatial medium resolution does not always improve species distribution models. It depends on the species being investigated.

Keywords : Species Distribution Models (SDM), woody species, MaxEnt, ecological response, Ecological Niche Factor Analysis (ENFA).

INTRODUCTION GÉNÉRALE

L'arbre fait partie intégrante des systèmes de production en petite agriculture familiale africaine et est considérée comme l'un des leviers d'amélioration de la sécurité alimentaire et des moyens de subsistance des populations tout en améliorant leur résilience face aux changements à venir (Garrity et al., 2010; Mbow et al., 2014). Ces parcs agroforestiers sont par conséquent une des voies de l'intensification durable de l'agriculture africaine. En effet, en fournissant des aliments, du bois de chauffe ou des plantes médicinales aux populations, les parcs agroforestiers jouent un rôle stratégique dans la diversification des sources de revenus ou d'alimentation. Par ailleurs, ils contribuent également à l'amélioration de la fertilité du sol, de la productivité des cultures ou la conservation de la biodiversité. En plus selon le rapport IPBES de 2019, plus de 2 milliards de personnes utilisent du combustible ligneux pour répondre à leurs besoins primaires en énergie, environ 4 milliards se soignent principalement avec des remèdes naturels, et 70% des médicaments utilisés pour traiter les cancers sont des produits naturels ou des produits de synthèse inspirés par la nature. Par le biais de ses processus écologiques et évolutifs, la nature maintient la qualité de l'air, des eaux douces et des sols dont l'humanité est tributaire, distribue l'eau douce, régule le climat, assure la pollinisation, lutte contre les ravageurs et atténue l'impact des aléas naturels.

La nature est essentielle pour la réalisation des objectifs de développement durable. Cependant, sachant que les objectifs de développement durable sont intégrés et indissociables, et qu'ils sont mis en œuvre au niveau des pays, les tendances négatives qui prévalent actuellement concernant la biodiversité et les écosystèmes compromettent la progression vers 35 des 44 cibles (80%) définies dans les objectifs relatifs à la pauvreté, à la faim, à la santé, à l'eau, aux villes, au climat, aux océans et aux terres (objectifs de développement durable 1, 2, 3, 6, 11, 13, 14, et 15). Cette biodiversité, produit de plus de trois milliards d'années d'évolution, constitue un patrimoine naturel et une ressource vitale pour toute l'humanité (Barbault, 2005).

En 2009, les signataires de l'accord de Copenhague ont souligné l'impact critique du changement du climat sur les pays du Sud, riche en biodiversité, qui auront donc besoin d'une assistance particulière pour établir un programme adéquat d'adaptation (Neuenschwander et al., 2011). Les espèces végétales sont universellement reconnues comme une composante fondamentale de cette biodiversité. Or selon le rapport de l'IPBES de 2019, en moyenne, 25% des espèces appartenant aux groupes d'animaux et de végétaux évalués sont menacés, ce qui suggère qu'environ 1 million d'espèces sont déjà menacées d'extinction, beaucoup dans les décennies à venir, à moins que des mesures ne soient prises pour réduire l'intensité des facteurs à l'origine de la perte de biodiversité. Faute de mesures, l'augmentation du taux global d'espèces

menacées d'extinction va encore s'accélérer, alors qu'il est déjà au moins des dizaines voire des centaines de fois plus élevé que la moyenne sur les 10 millions d'années écoulés.

En Afrique subsaharienne, plus de 2000 espèces forestières sont considérées comme éteintes ou menacées de disparition (Hilton-Taylor, 2000) sans qu'elles ne bénéficient de stratégies de conservation efficaces.

Il est donc important d'établir des priorités parmi les espèces concernées afin d'identifier les espèces prioritaires (Maxted et al., 2006). Cette priorisation est basée sur une évaluation multicritère qui traditionnellement utilise des critères d'endémicité, de rareté, de protection et particulièrement le statut (de menace) de l'espèce.

Au-delà du fait que la perte d'habitats de la flore serait actuellement due à la progression de l'agriculture, la destruction des milieux naturels liée à l'urbanisation, l'exploitation du bois et la récolte abusive de plantes médicinales, les changements climatiques pourraient affecter la distribution et entraîner l'extinction au cours du siècle prochain de nombreuses espèces végétales (Bradshaw et al., 2009). Plusieurs chercheurs ont démontré que les évolutions prédictives dans les concentrations en CO₂ atmosphérique et les changements climatiques qui en résultent sont susceptibles d'affecter la distribution et l'abondance de la plupart des espèces (Schroter et al., 2005). Selon Boko et al. (2007), l'Afrique a de fortes chances de connaître des changements climatiques marqués au cours du 21ème siècle par la sécheresse et le réchauffement dans la plupart des régions subtropicales. Selon les mêmes auteurs, l'Afrique est également l'un des continents les plus vulnérables en raison du manque de capacités adaptatives et des défis auxquels elle est confrontée tels que la pauvreté, les faiblesses technologiques, les conflits politiques.

Quelques études ont dressé les conséquences potentielles des changements de climat sur la biodiversité en Afrique. Ainsi, McClean et al. (2005) estiment que plus de 5000 espèces de plantes africaines pourraient connaître des pertes d'habitats avant 2080. Parmi les taxa les plus vulnérables à la réduction d'habitat adéquat, se trouvent les espèces qui : (1) ont un âge de maturité élevé, (2) produisent des diaspores en faibles quantités, (3) sont disséminées sur de faibles distances, (4) sont déjà soumises à une pression anthropique importante (Araujo et al., 2006 ; Penman et al., 2010).

Ces espèces semblent en outre être des espèces idéales pour des études de modélisation de distribution de niche (niche-based modelling). Une telle modélisation à l'échelle de leur aire de répartition peut d'une manière plus générale servir à évaluer la sensibilité des espèces d'arbres au climat et les impacts éventuels des changements climatiques (Mbatudde et al., 2012).

La préservation de l'environnement est aussi le 7ème des huit Objectifs du Millénaire pour le Développement, considéré par l'ONU comme crucial pour la réussite des autres objectifs énoncés dans la Déclaration du Sommet du Millénaire en 1995. Avec l'évolution de la perception et

de la conscience des communautés, l'environnement a acquis une valeur de bien commun, de support de vie nécessaire à toutes les espèces, et de patrimoine à exploiter en tenant compte des intérêts des générations futures.

Au cours des dernières décennies au Sénégal, les défis environnementaux générés par l'intensité des interactions de facteurs naturels (changements climatiques) et anthropiques (migrations, urbanisation et activités humaines) ont posé des enjeux énormes de préservation de l'environnement. Aujourd'hui, la prise de conscience de la nécessité de préserver l'environnement est une réalité à tous les niveaux de décision en fonction des problèmes environnementaux spécifiques à résoudre : pouvoir central, décideurs locaux, acteurs privés, organisations non gouvernementales, associations communautaires et populations locales. D'ailleurs, l'Etat du Sénégal qui est signataire de la plupart des accords et conventions internationaux sur l'environnement, la biodiversité et les ressources naturelles, s'est engagé dans une politique d'articulation de tous les programmes, projets et interventions en faveur du développement avec les préoccupations de protection et de sauvegarde de l'environnement. Cette volonté politique est en effet clairement manifestée à travers l'article L4 du code de l'environnement (loi 2001-01 du 15 janvier 2001).

De nos jours, la prédiction de la distribution des espèces est au centre de diverses applications en écologie, agriculture, horticulture, foresterie et conservation (Corsi et al., 1999 ; Welk et al., 2002 ; Elith et al., 2006) et joue également un rôle clé dans l'évaluation de l'impact des changements globaux sur les écosystèmes (Salinger et al., 2005). Selon Gaston (1996), la prédiction et la cartographie des habitats potentiellement adéquats pour les espèces menacées et en danger est nécessaire pour : (1) le contrôle et la restauration de leurs populations d'origine et de leurs habitats naturels, (2) les introductions artificielles, (3) la sélection des sites prioritaires de conservation, (4) la gestion de leur habitat naturel. Les outils de modélisation de la distribution des espèces sont devenus de plus en plus populaires en écologie et les modèles de distribution de niche deviennent largement utilisé dans bon nombre d'applications écologiques. Ces modèles établissent des relations entre l'apparition des espèces et les conditions biophysiques et environnementales de la zone d'étude.

Il existe de nombreuses méthodes de modélisation de la distribution des espèces en vu de prédire leur potentiel habitat adéquat (Guisan et Zimmermann, 2000 ; Guisan et al., 2007 ; Wisz et al., 2008). Cependant, très peu de modèles de prédiction sont adaptés pour les espèces de plantes rares et menacées (Engler et al., 2004). Généralement, deux types de modèles sont utilisés : ceux qui nécessitent des données de présence des espèces uniquement et ceux qui utilisent à la fois des données de présence et d'absence des espèces (Robertson et al., 2001). La plupart de ces méthodes de modélisation de la distribution des espèces est sensible à la dimension de l'échantillon (Wisz et al., 2008). Ils permettent dans les conditions actuelles d'application de prévoir généralement avec une bonne précision la distribution d'habitat pour des espèces.

Ainsi la modélisation de la distribution spatiale des principales espèces arborées peut permettre d'aider dans l'évaluation spatialisée et à la gestion des services écosystémiques associés. Les modèles de distributions d'espèces (SDMs) sont couramment utilisés pour prédire la présence d'espèces en se basant sur des corrélations entre des variables environnementales et des données géoréférencées d'enregistrement d'espèces. Cela permet notamment d'analyser les variables qui expliquent la présence des espèces mais également de délimiter la distribution potentielle des espèces.

Toute bonne recherche vise à répondre à une question précise. La nôtre est la suivante : **Comment modéliser la répartition spatiale des principales espèces ligneuses dans les parcs agroforestiers du bassin arachidier sénégalais à partir des modèles de distribution d'espèces ?**

La réponse à cette question pourrait permettre de mieux connaître les aires de distribution potentielle et des facteurs environnementaux conditionnant ces répartitions ; d'identifier les facteurs écologiques régissant la répartition géographique et la cartographie de son aire potentielle sous les conditions climatique et écologique actuelles et ce, pour une meilleure planification de leurs gestions.

Objectifs

L'objectif de notre étude est de modéliser la distribution des principales espèces ligneuses pour deux parcs agroforestiers sénégalais (le parc à *Faidherbia albida* dans la région de Niakhar et le parc à *Cordyla pinnata* dans la région de Nioro du Rip), en intégrant des variables à moyenne résolution spatiale décrivant le paysage.

Il s'agit spécifiquement de :

- ① déterminer les facteurs explicatifs de la distribution spatiale des espèces étudiées ;
- ② évaluer la capacité de prédiction des différents algorithmes utilisés dans les modèles de distribution des espèces ligneuses ;
- ④ élaborer des cartes de probabilité de présence des espèces ;
- ③ mettre en évidence au moyen des modèles la réponse écologique de chacune des espèces étudiées ;

Hypothèses de recherche

Traditionnellement, les SDM sont basés sur des variables climatiques globales, comme la température ou les précipitations ou d'autres données environnementales comme l'occupation du sol à large échelle et à basse résolution spatiale (supérieur à 1km). Dans notre présente étude, nous proposons d'y ajouter des variables à moyenne résolution spatiale (250 m), comme l'Eva-poTranspiration et Interception (ETI) réelle, la productivité annuelle nette en eau de biomasse (NBWP), le pH du sol. Ainsi, nous postulons deux hypothèses (H1 et H2) de recherche :

Hypothèses

- ❶ **H1 :** *Les variables à moyenne résolution spatiale améliorent les SDM ;*
- ❷ **H2 :** *Les espèces ligneuses répondent différemment aux conditions environnementale et climatique.*

Pour atteindre l'objectif général cité ci-dessus et confirmer ou infirmer les hypothèses, nous avons structuré le document en trois chapitres. Le premier chapitre de ce travail porte sur la revue de la littérature. Nous y retrouvons les concepts, les hypothèses et les algorithmes appliqués pour les modèles de distribution des espèces. Quant au deuxième chapitre, il présente la méthodologie adoptée dans la présente étude. Le dernier, expose les résultats de l'étude et la discussion.

Révue de la littérature

1.1 MODÉLISATION DE LA DISTRIBUTION DES ESPÈCES (SDM)

Les modèles de distribution des espèces (SDM) sont des outils numériques qui combinent les observations de l'occurrence ou de l'abondance des espèces avec les estimations environnementales. Ils sont utilisés pour obtenir des informations écologiques et évolutives et pour prédire les distributions à travers les paysages.

Les SDM sont désormais largement utilisés dans les domaines terrestre, d'eau douce et marin (Elith et Leathwick, 2009).

1.1.1 Concepts

La modélisation de la distribution des espèces (SDM) est également connue sous d'autres noms, notamment la modélisation de l'enveloppe climatique, la modélisation de l'habitat et la modélisation de niche (environnementale ou écologique). Le but des SDMs est d'estimer la similitude des conditions sur n'importe quel site avec les conditions aux endroits d'occurrence connue (et peut-être de non-occurrence) d'un phénomène. Une application courante de cette méthode est de prédire les aires de répartition des espèces avec des données climatiques comme prédicteurs.

Les étapes des SDMs sont :

- ① connaître les lieux d'occurrence des espèces ;
- ② extraire les valeurs des variables prédictives environnementales (telles que le climat) à ces emplacements ou mesurer les variables environnementales à ces endroits ;
- ③ utiliser les valeurs environnementales pour ajuster un modèle afin d'estimer la similitude avec les sites d'occurrence, ou une autre mesure telle que l'abondance de l'espèce ;
- ④ le modèle est utilisé pour prédire la variable d'intérêt dans une région d'intérêt.

L'objectif n'est pas alors de décrire complètement l'hypervolume dans un espace multidimensionnel mais plutôt d'utiliser comme approximation de la niche fondamentale (Tingley et al., 2009). En reliant de manière empirique des données d'occurrences à des variables environnementales prédictives ; on caractérise essentiellement la niche réalisée de l'espèce puisque les données d'occurrences sont contraintes par les interactions biotiques. Enfin, pour approximer au mieux la niche fondamentale, il est recommandé de calibrer le modèle avec des données d'occurrences provenant d'une large étendue géographique, à travers l'ensemble de la réparti-

tion de l'espèce (Araújo and Guisan, 2006 ; Barbet-Massin et al., 2010 ; Phillips et al., 2006).

1.1.2 Hypothèses

Les hypothèses clefs des SDM sont :

① Le postulat d'équilibre

Les modèles de niche font l'hypothèse que les espèces sont en équilibre avec leur environnement, c'est-à-dire qu'elles sont présentes dans tous les habitats favorables et absentes de tous les habitats défavorables (Araújo and Pearson, 2005).

② Le bon échantillonage

Les données de présences représentent un bon échantillonnage de l'espace environnemental. L'échantillonnage des données de présence et des variables environnementales est souvent limité dans le temps et/ou l'espace, les modèles reflètent un instantané de la relation entre l'espèce et son environnement (Guisan and Thuiller, 2005).

1.2 MÉTHODES DE MODÉLISATION

Un grand nombre d'algorithmes ont été utilisés dans la modélisation de la distribution des espèces. Ils peuvent être classés en :

- * méthodes de profil : domain, la distance de Mahalanobis et bioclim ;
- * méthodes de régression : GLM, les modèles additifs généralisés(GAM) ;
- * méthodes d'apprentissage automatiques tels que le Random Forest, les machines à vecteurs de support (SVM) , les arbres de régression boostés(BRT), MaxEnt.

1.2.1 Méthodes de profils

Les méthodes de profil ne prennent en compte que les données de présence, pas les données d'absence.

1.2.2 Méthodes de régression

Les méthodes de régression utilisent à la fois des données de présence et d'absence.

1.2.3 Méthodes d'apprentissage automatique

Comme les méthodes de régression, les méthodes d'apprentissage utilisent à la fois des données de présence et d'absence.

Elles sont des modèles de régression flexibles non paramétriques.

1.3 TECHNIQUES DE MODÉLISATION

La modélisation de la distribution des espèces (SDM) est un outil populaire en écologie, en partie parce qu'elle est capable de produire des prédictions spatiales de la distribution des espèces. Une composante importante du processus de modélisation est l'évaluation des résultats (Guisan, Thuiller et Zimmermann, 2017).

Lors de l'évaluation des performances SDM, il est courant de tester dans quelle mesure les prédictions correspondent aux observations à un ensemble d'emplacements (Franklin, 2010 ; Guisan et al., 2017 ; Peterson et al., 2011). Alors que les premières applications de la SDM avaient tendance à se concentrer sur les mesures statistiques de l'ajustement du modèle aux données utilisées pour former le modèle, l'attention s'est progressivement tournée vers les tests sur des données indépendantes (Elith et Leathwick, 2009). Étant donné que les données totalement indépendantes (telles que les nouvelles enquêtes) sont rarement disponibles, une approche courante consiste à sous-échantillonner les données disponibles pour la modélisation. En écologie, cela implique généralement de diviser les données en sous-ensembles pour la formation et les tests (également appelés respectivement étalonnage et validation) (Franklin, 2010 ; Radosavljevic et Anderson, 2014). Les données de formation sont utilisées pour ajuster le modèle et les données de test pour évaluer les performances. C'est ce qu'on appelle la validation du modèle (Hastie, Tibshirani et Friedman, 2009), avec des variantes basées sur des stratégies de fractionnement et de test. Par exemple, une ou plusieurs divisions peuvent être effectuées et les tests peuvent être effectués une fois ou répétés d'une certaine manière (par exemple, la validation croisée k- fold).

Encadré 1 : Validation croisée, blocs, plis et données sur les espèces

Validation croisée et plis

La validation croisée (cv) est une technique d'évaluation des modèles prédictifs. Il partitionne les données en k parties (plis), en utilisant une partie pour les tests et les autres (k - 1 plis) pour l'ajustement du modèle. Dans k -fold cv, le processus est répété jusqu'à ce que tous les plis aient été utilisés pour les tests. Si k est égal au nombre d'enregistrements, il est appelé n-fold, ou validation croisée avec omission (Hastie et al., 2009 , p. 241). Lorsque les plis ne sont pas choisis au hasard et qu'une sorte de stratégie environnementale, temporelle ou spatiale est utilisée pour construire les plis, cela s'appelle la validation croisée par blocs (Roberts et al., 2017).

Blocs et plis

D'une manière générale, les blocs sont des regroupements ayant des caractéristiques similaires. Ce sont des unités de zone géographique (par exemple des rectangles, des polygones spatiaux et des tampons de distance spécifique) ou des clusters environnementaux. Au sein de ces unités, toutes les données sur les espèces sont traitées ensemble, par exemple, allouées au même pli.

Données sur les espèces

Les données de présence uniquement sont des endroits où une espèce a été observée. Ceux-ci peuvent être couplés à un large échantillon de points à travers le paysage, connus sous le nom de données de fond (Renner et al., 2015). Les données de présence-absence comprennent à la fois les emplacements où l'espèce a été observée et où elle a été recherchée, mais non trouvée (Guillera - Arroita et al., 2015). Suivant la convention commune, notre code s'attend à ce que les données de présence soient représentées par un 1 et l'absence ou l'arrière-plan par un 0.

Conceptuellement, chaque bloc est une unité dans laquelle toutes les données sur les espèces sont traitées de la même manière. Plusieurs blocs peuvent être attribués à un pli de cv, il faut donc prendre des décisions sur la façon de regrouper les blocs en plis pour la validation croisée (encadré 1). La mise en œuvre pratique de cette étape de «blocs en plis» est affectée par les données sur les espèces.

L'ensemble suppose que les données spatiales des espèces et les variables prédictives (rasters) ont la même projection et une étendue similaire, c'est-à-dire que les prédicteurs s'étendent au moins sur toutes les données d'espèces.

1.3.1 Les différentes techniques de blockages

Il existe trois stratégies pour séparer les données de formation et de test : le blocage spatial, la mise en mémoire tampon spatiale et le blocage environnemental, chacun étant expliqué en

détail ci-dessous. Nous expliquons d'abord les stratégies, puis discutons des approches pour choisir la taille du bloc ou du rayon.

1.3.1.1 Le blockage spatial

Une stratégie générale pour tenir compte de l'autocorrélation spatiale lors de l'évaluation des modèles consiste à diviser les données spatialement en blocs. Il existe plusieurs méthodes pour créer des blocs spatiaux. L'une consiste à construire des blocs spatiaux carrés d'une taille spécifiée (c'est-à-dire une largeur). La position spatiale de ces blocs peut être décalée verticalement ou horizontalement, ce qui permet d'évaluer la sensibilité des métriques d'évaluation du modèle à des dispositions de blocs spécifiques et de diviser la zone d'étude en bacs verticaux ou horizontaux d'une hauteur ou d'une largeur donnée (blocs rectangulaires), comme l'ont utilisé Wenger et Olden (2012) et Bahn et McGill (2012) respectivement. Enfin, les blocs peuvent être spécifiés par une couche de polygones spatiaux.

L'attribution de blocs aux plis (encadré 1) peut se faire de manière aléatoire, systématique ou en damier. Les blocs à plis sont l'une des étapes clés de la modélisation des espèces, car les données sur les espèces sont rarement réparties uniformément sur les paysages.

Lorsque la sélection aléatoire des plis est choisie, des contraintes peuvent être définies pour éviter les plis avec peu ou pas de données de présence ou (le cas échéant) d'absence. Des techniques sont également mises en œuvre pour trouver des allocations de blocs à plis qui permettent une répartition plus uniforme des données sur les espèces entre les plis (par exemple, un nombre similaire d'enregistrements de présence et d'absence dans chaque pli).

Dans l'allocation systématique, les blocs sont numérotés et affectés séquentiellement aux plis. Le motif en damier n'a que deux plis, mais est utile pour appliquer aucun bloc adjacent dans un pli. Nous notons que dans tous les scénarios de blocage spatial, toutes les données des plis de test (y compris les points d'arrière-plan, le cas échéant) sont exclues des jeux de données d'apprentissage.

1.3.1.2 Mise en mémoire tampon

La stratégie de mise en mémoire tampon génère des plis d'entraînement et de test séparés dans l'espace en considérant des tampons circulaires de rayon spécifié autour de chaque point d'observation (Le Rest, Pinaud, Monestiez, Chadoeuf et Bretagnolle, 2014). Cette approche est liée à la validation croisée avec omission unique (encadré 1) et peut être utilisé que si l'on souhaite s'assurer qu'aucune donnée de test ne se substitue aux données de formation. Dans ce cas, il n'est pas nécessaire de faire la distinction entre les blocs (points tamponnés) et les plis, car chaque point laissé de côté équivaut à un pli. L'approche varie légèrement selon le type de données sur les espèces disponibles.

Pour les données de présence-absence , des plis sont créés en fonction de tous les enregistrements, à la fois les présences et les absences. Chaque observation cible (présence ou absence)

forme un point de test qui peut être considéré comme le centre du tampon ; tous les autres points de présence et d'absence dans le tampon sont ignorés. L'ensemble d'entraînement comprend tous les points de présence et d'absence à l'extérieur du tampon.

Quant aux données de présence , les plis de test sont déterminés par les données de présence. Un tampon est défini autour d'un point de présence cible central en utilisant le rayon (plage) spécifié.

Tous les points de présence non cibles à l'intérieur du tampon sont exclus à la fois des ensembles d'entraînement et de test. Tous les points (présence et arrière-plan) en dehors du tampon sont utilisés comme ensemble d'apprentissage.

1.3.1.3 Blocage environnemental

Cette stratégie utilise des méthodes de regroupement (k - moyens ; Hartigan et Wong, 1979) pour spécifier des ensembles de conditions environnementales similaires en fonction des covariables d'entrée et du nombre choisi de grappes dans l'espace environnemental (éventuellement multivarié). Les données sur les espèces au sein de chaque cluster sont affectées à un pli, de sorte que le nombre de plis est par défaut égal au nombre de clusters choisi par l'utilisateur. Le regroupement peut être basé sur toutes les cellules raster ou uniquement sur les valeurs aux points d'espèce. Lorsqu'ils sont basés sur les rasters complets, les grappes seront cohérentes dans toute la région et à travers toutes les espèces considérées dans cette région. Cependant, cela ne garantit pas que tous les clusters contiennent des enregistrements d'espèces, en particulier lorsque les données sur les espèces ne sont pas dispersées dans tous les environnements. Ainsi, les plis résultants dans la pratique peuvent être inférieurs au k spécifié . Alternativement, le regroupement peut être effectué uniquement sur la base des valeurs des prédicteurs aux points de présence et d'absence (ou de fond) de l'espèce. Dans ce cas, le nombre de plis sera toujours k. Bien que le regroupement soit répété plusieurs fois pour obtenir les regroupements les plus homogènes, les regroupements résultants peuvent changer dans différentes séries.

1.3.2 Choix de la technique de blockage

La meilleure stratégie à suivre dans une situation particulière dépendra de l'objectif de la modélisation et de la structure environnementale de la région (Roberts et al., 2017). Par exemple, les blocs environnementaux et les blocs spatiaux horizontaux pourraient créer des groupes climatiques distincts lors de la formation et des tests de jeux de données, ce qui peut être utile pour évaluer des modèles visant à prévoir de nouvelles conditions climatiques.

Différents choix auront différentes implications sur les estimations des performances prédictives. Par exemple, la mise en mémoire tampon est considérée comme utile pour imposer une séparation spatiale entre les plis d'entraînement et de test (Le Rest et al., 2014 ; Pohjankukka, Pahikkala, Nevalainen et Heikkonen, 2017), mais, en fonction de la taille relative du tampon et de la région, il peut produire des ensembles d'apprentissage à travers des répétitions plus

similaires que ceux produits par d'autres stratégies de blocage. Un inconvénient des ensembles d'apprentissage similaires est que l'estimateur d'erreur a tendance à avoir une variance élevée, c'est-à-dire qu'il est sensible à l'ensemble de données utilisé pour estimer l'erreur (Hastie et al., 2009 , p. 242). La mise en mémoire tampon applique également autant d'ensembles d'entraînement que de points de présence ou de présence-absence, ce qui peut donc coûter assez cher en termes de calcul.

1.3.3 Choix de la taille du bloc

L'un des défis de l'utilisation de blocs spatiaux ou de la validation croisée tamponnée est de choisir la taille optimale des blocs ou des tampons (Trachsel et Telford, 2016). La plage d'autocorrélation spatiale dans les résidus du modèle a été utilisée pour définir la distance de séparation optimale entre les ensembles d'entraînement et d'essai (Roberts et al., 2017 ; Telford et Birks, 2009 ; Trachsel et Telford, 2016). Il s'agit de la plage sur laquelle les résidus sont approximativement indépendants et peut être caractérisée à l'aide du variogramme empirique, un outil géostatistique fondamental pour mesurer l'autocorrélation spatiale. Le variogramme empirique décrit la structure de l'autocorrélation spatiale en mesurant la variabilité entre toutes les paires de points (O'Sullivan & Unwin, 2010), et la plage peut être utilisée comme taille de bloc. Il est facile de définir les résidus pour les données de présence-absence (par exemple Bio, De Becker, De Bie, Huybrechts, & Wassen, 2002 ; Dunn & Smyth, 1996) mais moins évident pour les données de présence uniquement (Baddeley, Turner, Møller, Et Hazelton, 2005).

Cependant, avoir les résidus implique d'avoir déjà ajusté le modèle, ce qui lie l'évaluation au modèle ajusté. Une option suggérée pour les données de présence-absence consiste à ajuster un variogramme aux données sur les espèces brutes (Roberts et al., 2017 et voir Bio et al., 2002) et à utiliser la distance résultante de l'analyse comme taille de bloc. Alternativement, pour prendre en charge un premier choix de taille de bloc, avant tout ajustement de modèle, l'autocorrélation existante dans les prédicteurs peut être choisie, comme une indication de la structure spatiale du paysage. L'intérêt principal de choix qu'il permet d'ajuster automatiquement les modèles de variogramme à chaque trame continue et en trouvant la plage efficace d'autocorrélation spatiale.

1.4 CONTEXTE BIOPHYSIQUE DU SÉNÉGAL

Pays sahélien situé à l'extrême ouest du continent africain, le Sénégal s'étend sur une superficie de 196 722 km² et partage ses frontières avec la République Islamique de Mauritanie au nord, le Mali à l'est, la République de Guinée au sud-est et la Guinée-Bissau au sud. La Gambie, située entre les régions naturelles du Sine Saloum et de la Casamance, forme une enclave sur le cours inférieur du fleuve du même nom. Le Sénégal est un pays plat avec une altitude inférieure à 50 m sur près de $\frac{3}{4}$ du territoire. Le point culminant (581 m) est situé à

l'extrême sud-est, sur les contreforts du Fouta Djallon. Le Sénégal, compte tenu de sa position géographique (extrême ouest du continent africain) et de sa façade maritime de plus de 700 km, connaît des différences climatiques entre la zone côtière et les régions de l'intérieur. La circulation atmosphérique, facilitée par un relief sans obstacles, expose le territoire en partie ou entièrement à l'influence de l'alizé maritime, de l'harmattan et de la mousson.

Ces masses d'air déterminent un climat de type soudanosahélien avec deux saisons :

- ★ une saison des pluies (de juin à octobre) pendant laquelle souffle la mousson, vent chaud et humide issu de l'anticyclone de Sainte-Hélène ;
- ★ une saison sèche (de novembre à mai) caractérisée par la prédominance des alizés du nord (alizé maritime issu de l'anticyclone des Açores et harmattan issu de l'anticyclone libyen).

La pluviométrie est très variable dans le temps et dans l'espace. Aussi, note-t-on une grande disparité hydro-climatique entre le sud humide (avec une pluviométrie annuelle supérieure à 1000 mm) et le nord sec (recevant moins de 500 mm de pluie par an). Trois principales zones de pluviométrie correspondant aux trois zones climatiques sont ainsi déterminées : une zone forestière au sud, une savane arborée au centre et une zone semidésertique au nord. Sur un plan temporel, on note une forte variabilité interannuelle des précipitations annuels qui met en évidence une succession de périodes plus ou moins marquées d'années sèches et d'années humides.

La situation du Sénégal est donc fortement marquée par un déficit pluviométrique et son impact sur les ressources naturelles et l'environnement.

Le réseau hydrographique de régime tropical est marqué par des différences importantes de débit entre la saison des pluies et la saison sèche qui peuvent entraîner parfois le tarissement des rivières. Les principaux bassins hydrographiques du Sénégal sont ceux des fleuves Sénégal et Gambie dont les eaux proviennent, en majeure partie, des montagnes du Fouta Djallon. La réalisation des grands barrages de l'OMVS contribue au développement socioéconomique du pays, notamment dans les secteurs de l'agriculture, de l'élevage, de la navigation et de l'approvisionnement des populations en eau potable et en énergie. La variété de ces conditions bioclimatiques et la diversité de son substratum géologique font que le Sénégal recèle plusieurs types de sols aux aptitudes inégales (sols secs et sablonneux du Sahel jusqu'aux sols latéritiques dans la région soudanaise au sud). Dans l'ensemble, les sols sont peu fertiles et souvent fragiles, sensibles à l'érosion éolienne et hydrique.

1.4.1 Les parcs agroforestiers du sénégal

Dans les régions semi-arides de l'Afrique de l'Ouest, particulièrement au Sud du Bassin Arachidier sénégalais, les parcs agroforestiers jouent un rôle fondamental pour les populations rurales (Diatta et al., 2016). Ils procurent aussi des produits utilisés dans l'alimentation humaine et du bétail, la pharmacopée, le bois de service et d'énergie (Sène, 2004 ; Yameogo et al., 2013 ;

Laouali et al., 2014) et procurent des sources de revenus (Boffa, 2000) aux populations locales. D'un point de vue écologique, ils contribuent à l'atténuation des changements climatiques par leur capacité de séquestration du carbone (Mbow, 2009 ; Kumar et Nair, 2011) et jouent un rôle important dans la protection (Samba et al., 2012) et l'amélioration de la fertilité des sols (Samba, 2001 ; Saidou et al., 2012).

1.4.1.1 Parcs à *Faidherbia albida*

Dans la région, les mauvaises pratiques culturelles combinées aux effets de la sécheresse et à l'explosion démographique ont conduit à l'épuisement des terres. Les principaux problèmes de l'agriculture sont : baisse de fertilité du sol, érosion éolienne, perte de surface arable, baisse des productions agricoles, difficulté d'accès à l'eau, élagage, plantes envahissantes.

Pour maintenir la fertilité dans leurs champs, les populations de perpétuent une pratique traditionnelle, la culture sous parc arboré. *Faidherbia albida* contribue fortement à reconstituer la fertilité du sol en restituant l'Azote atmosphérique. Par ailleurs, sa présence dans les champs réduit la vitesse du vent et par conséquent, permet de stabiliser le sol. Pour accroître l'efficacité de cette pratique, les populations y associent la rotation culturale et l'apport de fumure organique.

1.4.1.2 Parcs à *Cordyla pinnata*

Cependant, les systèmes de production agricole dans le Bassin Arachidier du Sénégal, en particulier le parc agroforestier à *Cordyla pinnata*, sont marqués par une dégradation progressive au fil des années (Herrmann et Tappan, 2013). Les principales contraintes du parc sont les coupes pour le bois d'énergie et le bois d'œuvre, les divers prélèvements (écorce, racine, feuilles...) de plantes médicinales, les feux de brousse et l'avancée du front pionnier agricole (Lykke, 2000). A cela s'ajoute la faible régénération des espèces et la lenteur de la croissance des principales espèces agroforestières à haute valeur ajoutée (Diop et al., 2011). Aujourd'hui, l'état de dégradation du parc agroforestier à *Cordyla pinnata* ne semble pas le permettre de fournir tous les services associés à ce système d'utilisation et de gestion des arbres.

Méthodologie

2.1 ZONE D'ÉTUDE

L'étude est menée dans le bassin arachidier du Sénégal, précisément dans les parcs à *Faidherbia albida* et *Cordyla pinnata*.

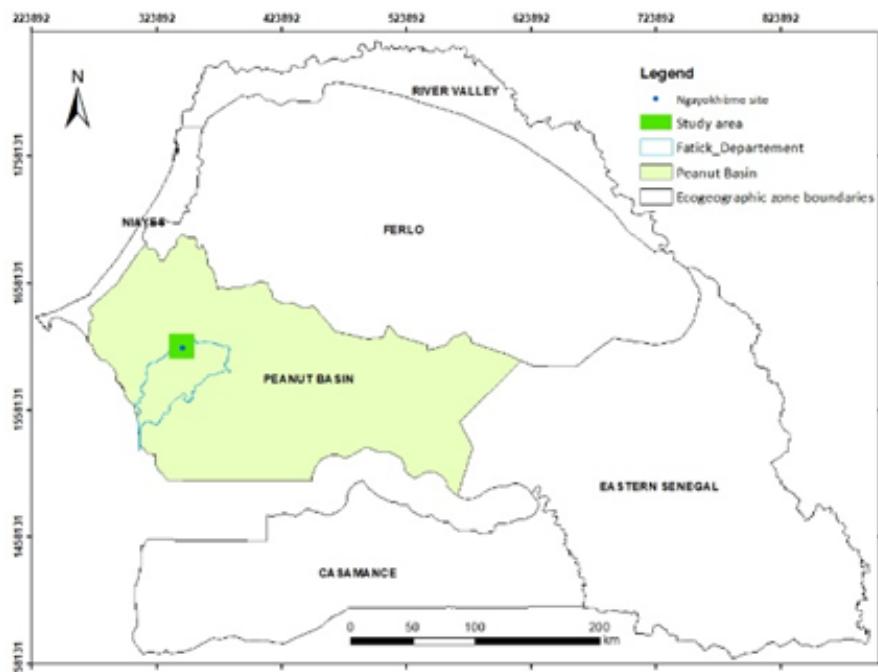


FIGURE 2.1 – Zone d'étude dans le parc à *Faidherbia albida*

Source : Babacar et al., 2019

Les coordonnées limites du parc à *Faidherbia albida* sont ($xmin = -16.53864$, $xmax = -16.35454$, $ymin = 14.45461$, $ymax = 14.63543$) et celles du parc à *Cordyla pinnata* sont ($xmin = -15.89166$, $xmax = -15.7052$, $ymin = 13.67262$, $ymax = 13.84793$).

La zone d'étude est caractérisée par une température moyenne annuelle de $27^{\circ}C$, des précipitations annuelles comprises entre 490 et 540 mm dans le parc à *Faidherbia albida*, entre 660 et 700 mm dans le parc à *Cordyla pinnata*. Elle est caractérisée également par un pH du sol inférieur à 7, une EvapoTranspiration et Interception réelle comprise entre 3000 et 5000 dans parc à *Faidherbia albida*, entre 4000 et 8000. La productivité annuelle nette en eau de biomasse dans le parc à *Faidherbia albida* est comprise entre 1100 et 1500 tandis qu'elle est comprise

entre 600 et 1500 dans le parc à *Cordyla pinnata* .

Le parc à *Faidherbia albida* est caractérisé par trois type de sol : ferrugineux tropicaux,hydromorphes et hydromorphes salés (Annexe B, section 2.0).

Les sols ferrugineux tropicaux sont caractérisés par une profondeur de sol variable (en tout cas moins de 3 m) ; une nette horizonation des profils pédologiques, horizon de surface riche en matière organique ; une présence, parfois, d'un horizon de transition lessivé (30-40 cm) et d'un horizon d'accumulation en argile et en fer avec ou sans concrétions et nodules ; le rapport limon fin sur argile est toujours supérieur à 0,20 et celui SiO₂ / Al₂O₃ à 2 ; le drainage interne généralement bon est parfois mauvais ; et une présence d'horizon induré (surtout dans les bassins-versants). Quant aux sols hydromorphes, ils montrent des marques physiques de saturation régulière en eau. Les sols hydromorphes salés, en plus d'avoir une saturation régulière en eau, ils sont salés.

Quant au parc à *Cordyla pinnata* , il est caractérisé par les sols ferrigineux tropicaux rubefiés, ferrigineux tropicaux lessives, lithosols et hydromorphes.

2.2 DONNÉES UTILISÉES

2.2.1 Données de terrain

Il s'agit de deux bases de données sur les arbres. Elles ont été acquises à l'issu de deux campagnes de collectes de données dans les deux sites étudiés. La première campagne s'est déroulé en juin 2018 à Niakhar et la seconde en février 2019 à Nioro. La collecte est faite dans les deux sites suivant un protocole d'échantillonnage stratifié tenant compte de l'hétérogénéité spatiale du paysage (Ndao et al., 2019). A Niakhar une base de données de 9258 arbres géoréférencés répartis en 63 espèces a été collectée. De même, à Nioro c'est une base de 6611 arbres géoréférencés répartis en 60 espèces. Dans chaque base de données sur les arbres, la variable sur le type de sol a été renseignée.

2.2.2 Données environnementales

Nous avons utilisé différents types de variables pouvant caractériser l'environnement de la zone d'étude et son fonctionnement notamment la répartition spatiale des espèces. Ces variables sont présentées ici selon leurs sources et leurs natures.

2.2.2.1 Variables bioclimatiques

Les variables bioclimatiques utilisées dans le cadre de notre étude sont dix-neuf (19). Elles sont de résolution 1 km, et dérivent des valeurs mensuelles de la température et des précipitations pour générer des variables plus significatives sur le plan biologique. Utilisées dans la

modélisation de la répartition des espèces et les techniques de modélisation écologique, les variables bioclimatiques représentent les tendances annuelles (la température annuelle moyenne, les précipitations annuelles, ...), la saisonnalité (par exemple, la plage annuelle de température et de précipitations) et les facteurs environnementaux extrêmes ou limitants (par exemple, la température du mois le plus froid et le plus chaud, et les précipitations des quartiers secs). Ces variables sont téléchargeables à partir de la base WorldClim.

WorldClim est un ensemble de couches climatiques mondiales (données climatiques maillées). Dans le cadre de notre étude, nous avons utilisé la version 2.0 pour télécharger les données bioclimatiques (<https://www.worldclim.org/>).

Le tableau 2.1 donne la description des variables bioclimatiques.

TABLE 2.1 – Variables bioclimatiques

Code	Variables	Unité
bio1	température moyenne annuelle	°C
bio2	plage diurne moyenne (moyenne mensuelle (temp max - temp min))	°C
bio3	isotherme (bio2 / bio7) (* 100)	
bio4	saisonnalité de la température (écart-type * 100)	°C
bio5	température maximale du mois le plus chaud	°C
bio6	Température minimale du mois le plus froid	°C
bio7	Plage annuelle de température (bio5-bio6)	°C
bio8	Température moyenne du trimestre le plus humide	°C
bio9	Température moyenne du trimestre le plus sec	°C
bio10	Température moyenne du trimestre le plus chaud	°C
bio11	Température moyenne du trimestre le plus froid	°C
bio12	Précipitations annuelles	mm
bio13	Précipitation du mois le plus humide	mm
bio14	Précipitation du mois le plus sec	mm
bio15	Saisonnalité des précipitations (coefficient de variation)	
bio16	Précipitation du trimestre le plus humide	mm
bio17	Précipitation du trimestre le plus sec	mm
bio18	Précipitation du trimestre le plus chaud	mm
bio19	Précipitation du trimestre le plus froid	mm

Source : Données WorlClim

2.2.2.2 Variables sur les propriétés du sol (2010 à 2019)

Sept variables sur les propriétés du sol ont été utilisées dans le cadre de notre étude comme l'illustre le tableau 2.2.

TABLE 2.2 – Les variables des propriétés du sol : AfSoilGrids250m

Description	Code	Unité
Proportion d'Argile	CLYPPT	g /100 g
Carbone Organique	ORCDRC	g/Kg
Proportion de Limon	SLTPPT	g /100 g
pH du sol	PHIOX	
Total d'Azote	NTO	g / Kg
Proportion de Sable	SNDPPT	g/ 100 g
Phosphore	P	mg/ 100 Kg

Source : Données AfSoilGrids250m

Ces variables sont issues de la base de données AfSoilGrids250m téléchargeables sur <https://www.isric.org/projects/soil-property-maps-africa-250-m-resolution>. Elles sont à l'échelle africaine et de résolution 250 mètres.

2.2.2.3 Variables sur la productivité de l'eau (2010 à 2019)

Deux variables sur la productivité de l'eau ont été utilisées. Il s'agit de l'EvapoTranspiration et Interception (ETI) réelle et la productivité annuelle nette en eau de biomasse (NBWP). L'EvapoTranspiration et Interception réelle est la somme de l'évaporation du sol (E), de la transpiration de la canopée (T) et de l'évaporation des précipitations interceptées par les feuilles (I). La valeur de chaque pixel représente l'ETIa au cours d'une année donnée.

La productivité annuelle nette en eau de biomasse exprime la quantité de production (production totale de biomasse) par rapport au volume d'eau effectivement consommé (par transpiration du couvert végétal) dans l'année, et donc net de l'évaporation du sol.

Ces variables de résolution 250 mètres sont issues de WaPOR (https://wapor.apps.fao.org/home/WAPOR_2/1), le portail de la FAO pour surveiller la productivité de l'eau grâce au libre accès aux données dérivées de télédétection, surveille et rend compte de la productivité de l'eau agricole en Afrique et au Proche-Orient.

2.2.2.4 Variables phénologiques

Les métriques phénologiques ont téléchargées avec le logiciel TIMESAT lors du stage de Louise Lemettais (Lemettais 2020) portant sur l'analyse des tendances de la végétation au Sahel. Il s'agit de :

- * SOS ou Start of the Season, qui exprime le début de la période de croissance de la végétation ;
- * SINT ou Small Integrale, qui est le cumul de l'indice de végétation NDVI durant la période de croissance c'est-à-dire entre le début de la croissance et la fin de la croissance. Il exprime la productivité de la végétation durant la saison de production.

2.2.2.5 Variables topographiques

Ce sont des variables décrivant les attributs du bassin et des sous bassins versants de la zone. Elles sont dérivées du modèle agro hydrologique SWAT (Soil and Water Assessment Tool) qui permet de modéliser le fonctionnement hydrologique des bassins versants. Un modèle numérique de terrain (MNT) de résolution 30 cm a été utilisé pour effectuer les calculs sous ArcSWAT (le module de SWAT dans ArcGIS). Six variables de fonctionnement hydrologique ont été testé :

- * wid1 et dep1, qui représentent la largeur et la profondeur du canal secondaire (sous bassin versant);
- * Len1 et Elev, qui traduisent le plus long chemin d'écoulement et l'élevation du bassin versant;
- * Csl et sll, qui sont des dérivées de la pente.

2.3 TRAITEMENT DES DONNÉES ET MODÉLISATION DE LA DISTRIBUTION DES ESPÈCES

2.3.1 Prétraitement

Les étapes suivantes ont été adoptées pour le traitement des données utilisées :

- ❶ l'ensemble des variables bioclimatiques et environnementales ont été projetées dans le même système de projection (WGS84 : système de référence mondiale);
- ❷ les variables ont été découpées sur les deux parcs en utilisant leurs coordonnées limites comme définies dans la présentation de la zone d'étude ;
- ❸ les variables topographiques ont été rastérisées ;
- ❹ les variables ont été rééchantillonnées pour obtenir la même résolution spatiale de 250 mètres.

L'ensemble des travaux de traitement des données ont été réalisés en utilisant les logiciels QGIS (version 3.10.0) et R (version 3.6).

2.3.2 Préparations des données

2.3.2.1 Données sur l'occurrence des espèces

Les discussions sur la modélisation de la répartition des espèces se concentrent souvent sur la comparaison des méthodes de modélisation, mais si vous traitez des espèces avec des enregistrements peu nombreux et incertains, votre objectif devrait probablement être d'améliorer la qualité des données d'occurrence (Lobo, 2008). Toutes les méthodes fonctionnent mieux si vos données d'occurrence sont non biaisées et exemptes d'erreur (Graham et al. , 2007) et que vous

disposez d'un nombre relativement important d'enregistrements (Wisz et al. , 2008). Les données sur les occurences des espèces constituent notre variable d'intérêt. Elle est définie comme suit :

$$Y_i = \begin{cases} 1 & \text{si l'espèce } i \text{ est présente , } \forall i \in \{1, 2, .., 9\} \\ 0 & \text{sinon} \end{cases}$$

La présence ou l'absence d'une espèce est connue par ses coordonnées géographiques (longitude et latitude).

2.3.2.2 Enregistrements en double

Fait intéressant, un autre problème de la qualité des données est l'enregistrement en double ou doublons pour la même espèce et certaines coordonnées dans les données. Après vérification nous avons recalculé à partir de QGIS les coordonnées géographiques des espèces en prenant six chiffres après la virgule.

2.3.2.3 Biais d'échantillonnage

Un biais d'échantillonnage est fréquemment présent dans les enregistrements d'occurrences (Hijmans et al. , 2001). On peut tenter de supprimer une partie du biais en sous-échantillonnant les enregistrements. Cependant, le sous-échantillonnage réduit le nombre d'enregistrements et ne peut pas corriger les données pour les zones qui n'ont pas du tout été échantillonnées. Pour éviter de réduire le nombre d'enregistrements ou d'occurrences des espèces, nous avons étendu la zone de découpage en prenant deux chiffres après la virgule sur les coordonnées limites.

2.3.3 Choix des espèces étudiées

L'analyse des figures 2.1 et 2.2 révèlent que l'espèce la plus abondante dans le parc à *Faidherbia albida* est *Faidherbia albida* avec une proportion de 42%. Elle est suivie respectivement par *Balanites aegyptiaca* (11%), *Anogeissus leiocarpus* (9%) et *Adansonia digitata* (6%). L'espèce ligneuse la moins abondante dans le parc à *Faidherbia albida* parmi les espèces étudiées est *Acacia nilotica* avec une proportion de 4%. Les autres espèces (58) représentant 28% des espèces dans le parc, sont moins abondantes avec une proportion de moins de 3%.

L'espèce la plus abondante dans le parc à *Cordyla pinnata* est *Cordyla pinnata* avec une proportion de 43%, suivi d'*Azadirachta indica*(20%). Les espèces les moins abondantes parmi celles étudiées dans ce parc sont *Prosopis africana* et *Ficus capensis* avec 5% chacune. 56 espèces dans ce parc sont moins abondantes, représentant 27% des espèces.

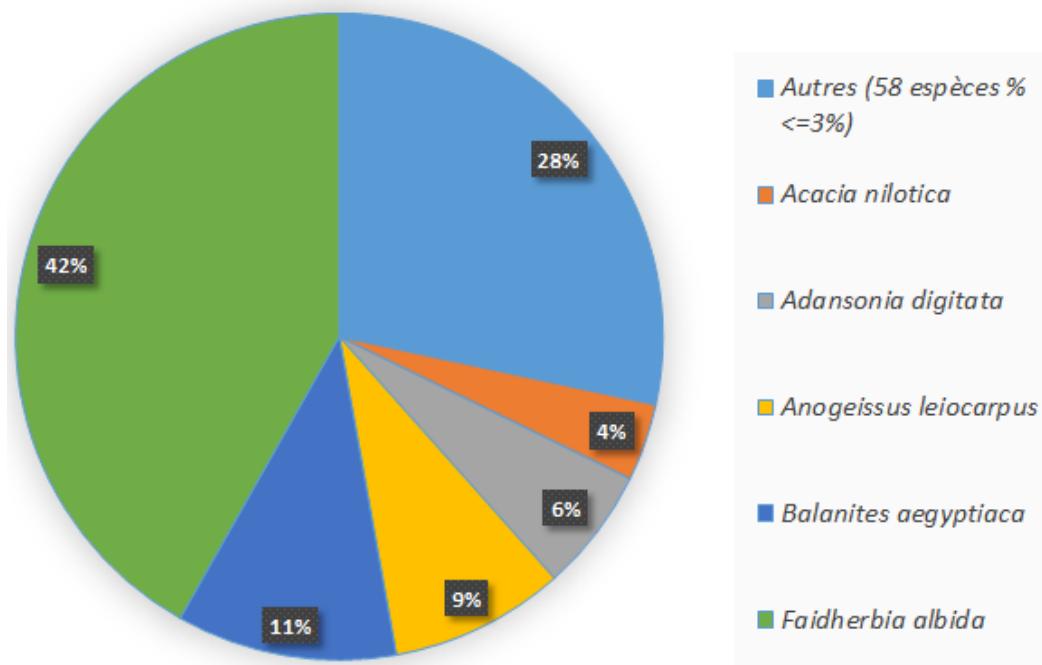


FIGURE 2.2 – Abondance des espèces dans le parc à *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

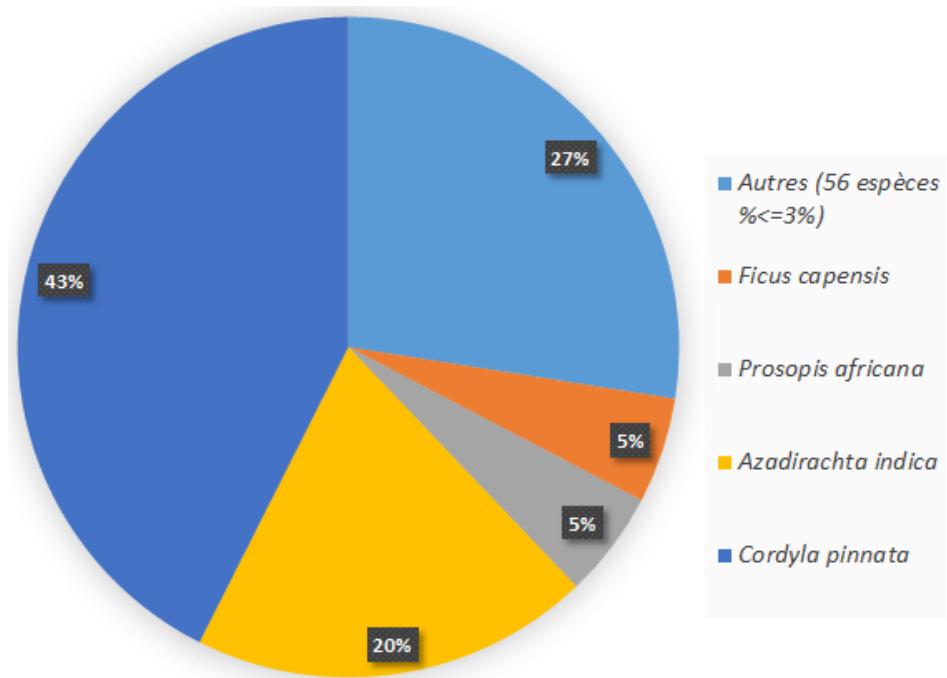


FIGURE 2.3 – Abondance des espèces dans le parc à *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Les espèces majoritaires dans chaque parc ont été choisies dans le cadre de notre étude. En effet, les cinq espèces majoritaires (*Faidherbia albida*, *Balanites aegyptiaca*, *Anogeissus leiocarpus*, *Adansonia digitata* et *Acacia nilotica*) ont été choisies dans le parc à *Faidherbia albida*

et les quatre espèces majoritaires (*Cordyla pinnata*, *Azadirachta indica*, *Prosopis africana* et *Ficus capensis*) dans le parc à *Cordyla pinnata*.

Les espèces étudiées dans le parc à *Faidherbia albida* représentent 72% des espèces dans ce parc et 73% pour celles dans le parc à *Cordyla pinnata*.

Ensuite nous avons effectuer une analyse bi-variée afin d'apprécier l'association entre la présence ou l'absence d'une espèce et la variable, le type de sol. Pour apprécier cette association, nous effectuerons des tests du « **Khi-deux** ».

Encadré 1 :

Le test du Khi-deux (χ^2) est un test statistique qui permet d'accepter ou de rejeter l'hypothèse H_0 « il n'y a aucune relation entre les variables qualitatives, nominales ou ordinaires dans la population dont est issu l'échantillon ».

Elle consiste à examiner l'existence éventuelle d'une relation entre la présence ou l'absence d'une espèce et chacune des variables qualitatives. A l'aide de la statistique du khi deux, il s'agit pour nous de vérifier si la relation entre deux variables est significative ou non.

2.3.4 Autocorrélation spatiale

L'autocorrélation mesure la corrélation d'une variable avec elle-même, lorsque les observations sont considérées avec un décalage dans le temps (autocorrélation temporelle) ou dans l'espace (autocorrélation spatiale). On définit l'autocorrélation spatiale comme la corrélation, positive ou négative, d'une variable avec elle-même du fait de la localisation spatiale des observations.

D'un point de vue statistique, de nombreuses analyses (analyse des corrélations, régressions linéaires, etc.) reposent sur l'hypothèse d'indépendance des variables. Lorsqu'une variable est spatialement autocorrélée, l'hypothèse d'indépendance n'est plus respectée, remettant ainsi en cause la validité des hypothèses sur la base desquelles ces analyses sont menées. D'autre part, l'analyse de l'autocorrélation spatiale permet une analyse quantifiée de la structure spatiale du phénomène étudié(présence/absence d'une espèce dans notre étude). En présence d'autocorrélation spatiale, on observe que la valeur d'une variable pour une observation est liée aux valeurs de cette même variable pour les observations voisines.

- ★ l'autocorrélation spatiale est positive lorsque des valeurs similaires de la variable à étudier se regroupent géographiquement ;
- ★ l'autocorrélation spatiale est négative lorsque des valeurs dissemblables de la variable à étudier se regroupent géographiquement : des lieux proches sont plus différents que des lieux éloignés ;
- ★ en l'absence d'autocorrélation spatiale, on peut considérer que la répartition spatiale des observations est aléatoire.

Lorsque la variable d'intérêt n'est pas continue, mais catégorielle (Présence ou Absence), on mesure le degré d'association locale grâce à une analyse des statistiques des join count (ZHUKOV 2010).

On observe :

- ★ une autocorrélation spatiale positive si le nombre de liaisons Présence-Absence est significativement inférieur à ce que l'on aurait obtenu à partir d'une répartition spatiale aléatoire ;
- ★ une autocorrélation spatiale négative si le nombre de liaisons Présence-Absence est significativement supérieur à ce que l'on aurait obtenu à partir d'une répartition spatiale aléatoire ;
- ★ aucune autocorrélation spatiale si le nombre de liaisons Présence-Absence est approximativement identique à ce que l'on aurait obtenu à partir d'une répartition spatiale aléatoire.

S'il y a n observations, n_p observations de présence et $n_a = n - n_p$ observations d'absence, la probabilité de présence de l'espèce est : $P_p = \frac{n_p}{n}$ et sa probabilité d'absence est : $P_a = 1 - P_p$. En l'absence d'autocorrélation spatiale, les probabilités de présence et d'absence dans deux polygones voisins sont $P_{pp} = P_p P_p = P_p^2$ et $P_{aa} = (1 - P_a)(1 - P_a) = (1 - P_a)^2$.

La probabilité que l'espèce soit présente ou absente dans l'un des deux polygones est $P_{pa} = P_p(1 - P_p) + P_p(1 - P_p) = 2P_p(1 - P_p)$.

Sous l'hypothèse d'une répartition spatiale aléatoire des observations, $\frac{1}{2} \sum_i \sum_j w_{ij}$ mesure le nombre de liaisons existantes, on peut écrire :

$$E[pp] = \frac{1}{2} \sum_i \sum_j w_{ij} P_p^2 \quad (2.1)$$

$$E[aa] = \frac{1}{2} \sum_i \sum_j w_{ij} (1 - P_p)^2 \quad (2.2)$$

$$E[pa] = \frac{1}{2} \sum_i \sum_j w_{ij} 2P_p(1 - P_p) \quad (2.3)$$

Si l'on désigne par $y_i = 1$ lorsque l'espèce est présente et par $y_i = 0$ dans le cas contraire , les contre-parties empiriques (valeurs observées) de ces espérances mathématiques peuvent s'écrire :

$$pp = \frac{1}{2} \sum_i \sum_j w_{ij} y_i y_j \quad (2.4)$$

$$aa = \frac{1}{2} \sum_i \sum_j w_{ij} (1 - y_i)(1 - y_j) \quad (2.5)$$

$$pa = \frac{1}{2} \sum_i \sum_j w_{ij} (y_i - y_j)^2 \quad (2.6)$$

Dans ce cas, la statistique de test permettant d'évaluer la significativité de l'autocorrélation spatiale repose sur l'hypothèse qu'en l'absence d'autocorrélation spatiale, les statistiques de join count (pp , aa et pa) suivent une loi normale. Nous écrivons alors :

$$\frac{pa - E[pa]}{\sqrt{var(pa)}} \text{ suit } N(0, 1) \quad (2.7)$$

$$\frac{pp - E[pp]}{\sqrt{var(pp)}} \text{ suit } N(0, 1) \quad (2.8)$$

$$\frac{aa - E[aa]}{\sqrt{var(aa)}} \text{ suit } N(0, 1) \quad (2.9)$$

2.3.5 Sélection des variables par la méthode ENFA

Le concept de niche, tel que défini par Hutchinson (1957), considère la niche écologique d'une espèce comme un hypervolume dans l'espace multidimensionnel défini par des variables environnementales, au sein desquelles les populations d'une espèce peuvent persister.

L'analyse écologique des facteurs de niche a été développée par Hirzel et al. (2002) pour analyser la position de la niche dans l'espace écologique.

La méthode utilisée (ENFA) est spécifique pour l'identification des niches écologiques et permet de sélectionner les variables contribuant à la niche écologique des espèces.

Le principal avantage des modèles de niches écologiques, par rapport aux autres approches de modélisation de régression traditionnelles, est qu'ils ne nécessitent que des données de présence et qu'ils évaluent efficacement la probabilité de présence d'espèces, ou l'adéquation écologique relative d'une unité spatiale, dans la zone d'étude.

L'ENFA fait ressortir pour une espèce donnée, le sous-espace de l'espace globale qui lui est favorable ainsi que les variables qui expliquent de manière significative la présence de celle-ci dans ce sous-espace. Le premier axe de l'ENFA est la marginalité : c'est à dire la direction dans laquelle l'espèce se niche le plus possible par rapport aux conditions disponibles dans la zone globale. Une marginalité positive signifie que l'espèce préfère une valeur supérieure à la moyenne sur la variable écologique. Le deuxième axe est la spécialisation qui indique dans quelle mesure la niche de l'espèce est restreinte par rapport à la zone d'étude.

La marginalité représente la distance au carré du barycentre de niche par rapport à l'habitat disponible moyen. Une grande spécialisation correspond à une niche étroite par rapport aux conditions d'habitat disponibles pour l'espèce.

Pour obtenir les facteurs de spécialisation, le système de référence est modifié afin de transformer l'ellipsoïde de l'espèce dans une sphère, dont la variance est égale à l'unité dans chaque direction. Dans ces nouvelles mesures, le premier facteur de spécialisation est celui qui maximise la variance de la distribution globale (bien qu'orthogonal au facteur de marginalité).

Les figures 2.3 et 2.4 illustrent l'application d'ENFA pour les deux espèces majoritaires dans les deux parcs. Le polygone gris clair montre les conditions environnementales dispo-

nibles dans la zone d'étude, le polygone gris foncé montre les conditions environnementales où les espèces ont été observées (représentation de la niche observée), et le petit cercle blanc correspond au barycentre de sa distribution. Toutes les variables qui se trouvent dans le polygone gris foncé expliquent de manière significative la présence de l'espèce.

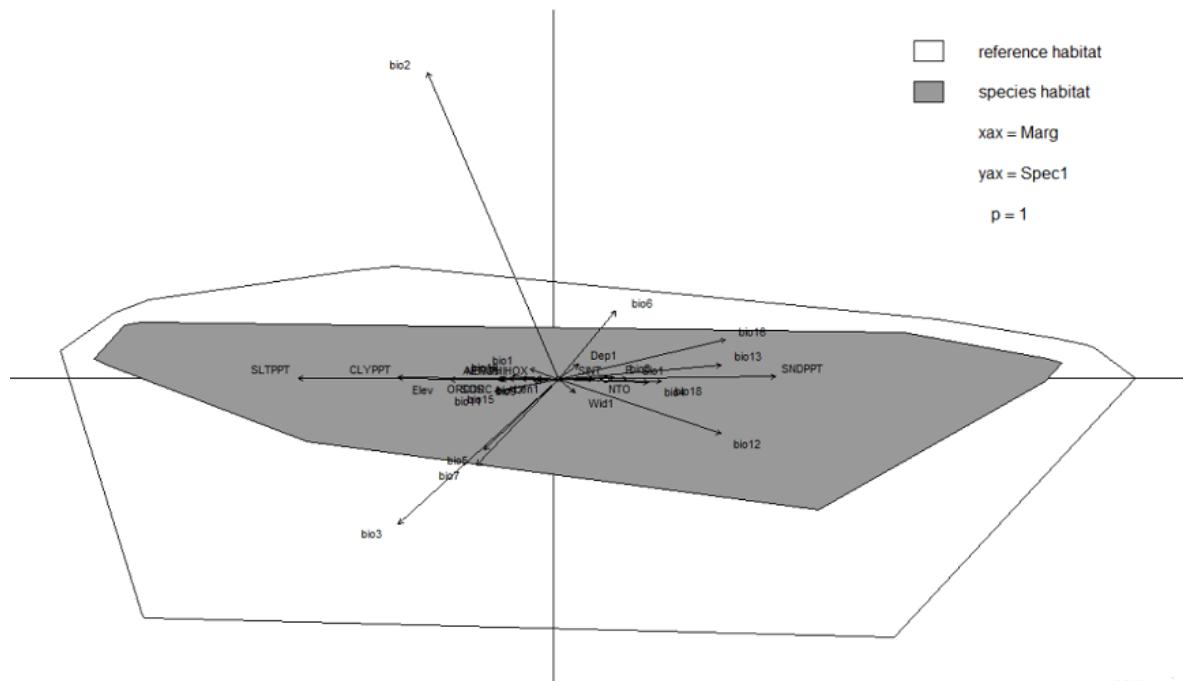


FIGURE 2.4 – Analyse factorielle de la niche écologique (ENFA) de la distribution de *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

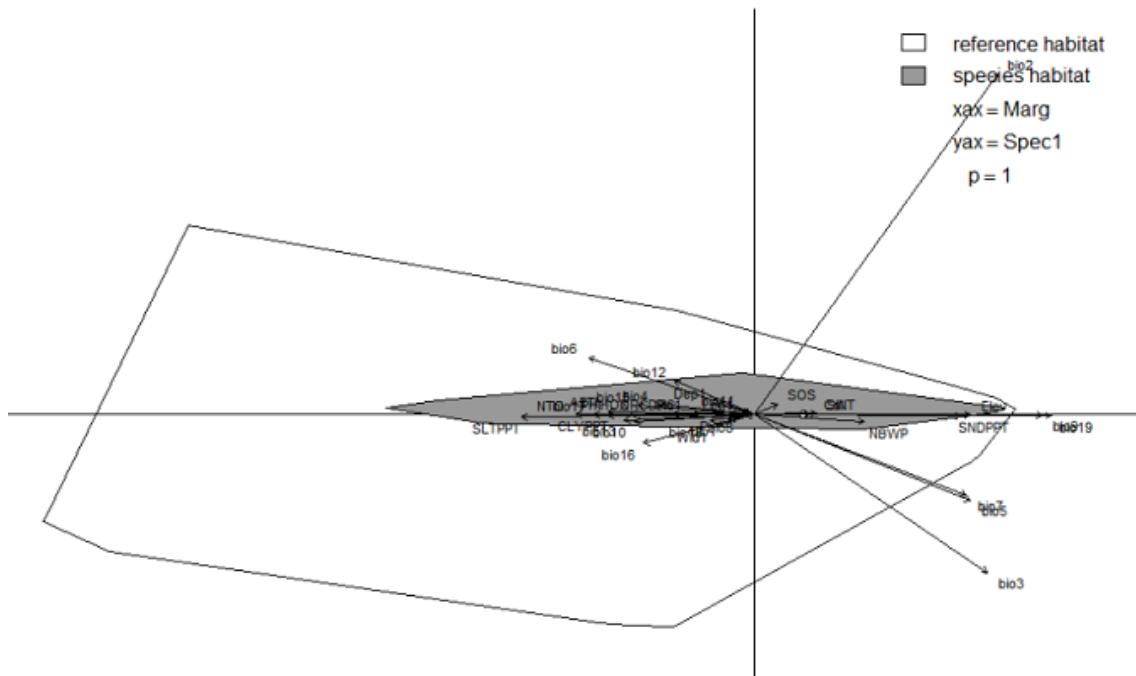


FIGURE 2.5 – Analyse factorielle de la niche écologique (ENFA) de la distribution de *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

2.3.6 Technique de modélisation

Les tracés de sortie (figures 2.5 et 2.6) montrent les plages d'autocorrélation spatiale des covariables en entrée dans chaque parc et le bloc spatial qui a été créé en fonction de la médiane (en pointillé) de ces plages.

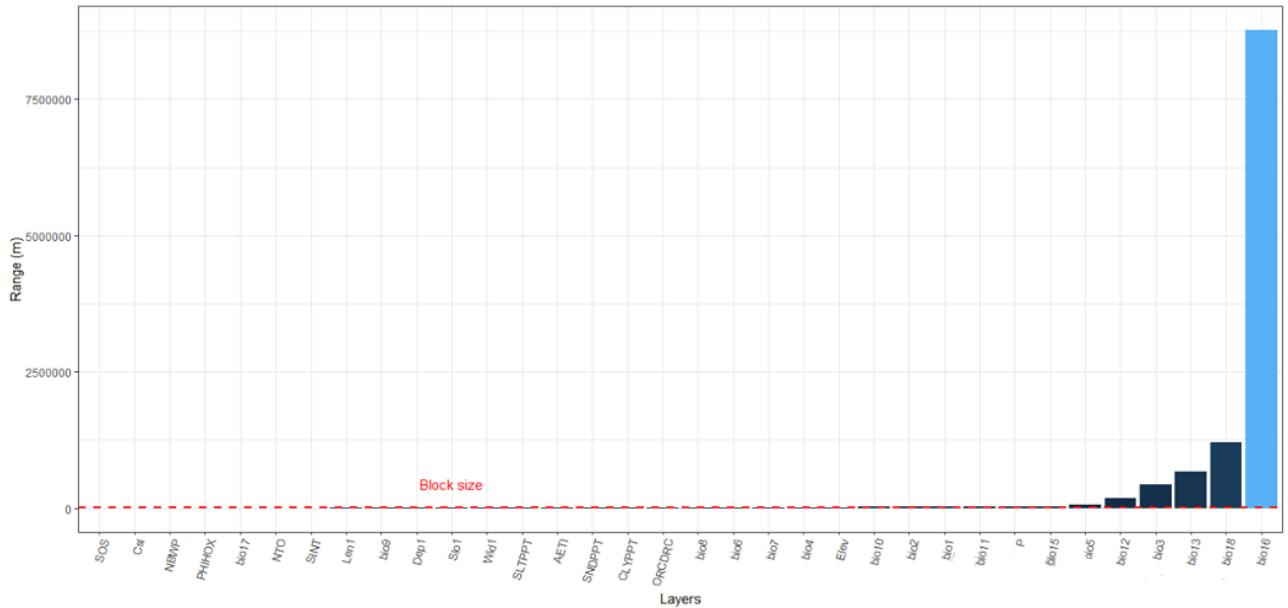


FIGURE 2.6 – Autocorrélation spatiale des covariables dans le parc à *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

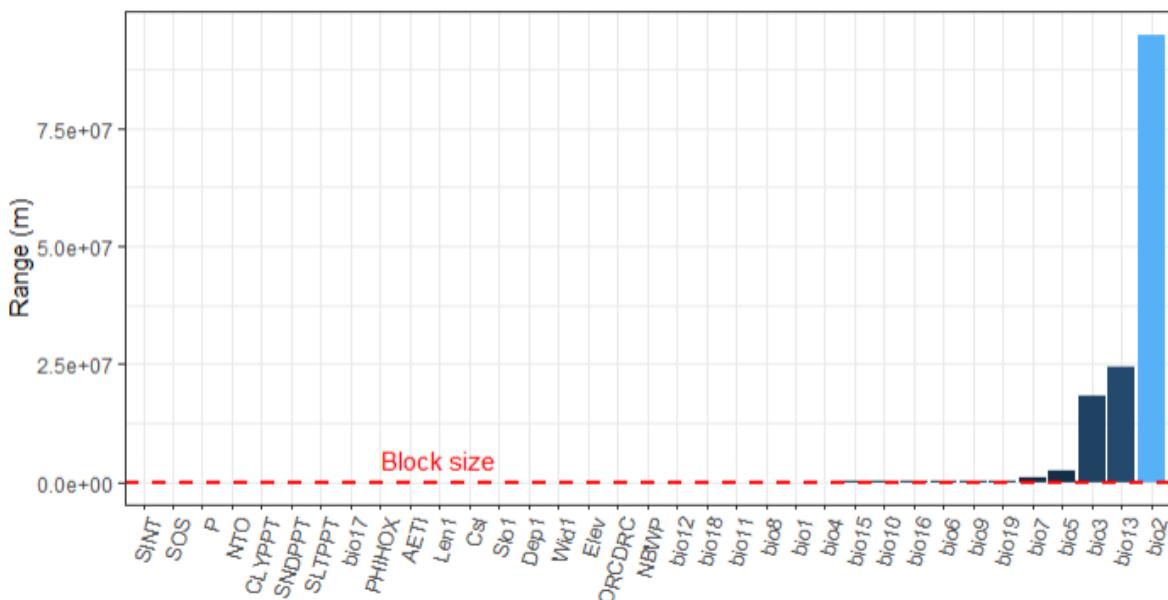


FIGURE 2.7 – Autocorrélation spatiale des covariables dans le parc à *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

A partir de l'autocorrélation spatiale des covariables dans les deux parcs, nous définissons les blocs spatiaux pour la séparation des données de test et d'entraînement. La séparation de ces données pour l'espèce *Faidherbia albida* est illustrée dans la figure ci-dessous.

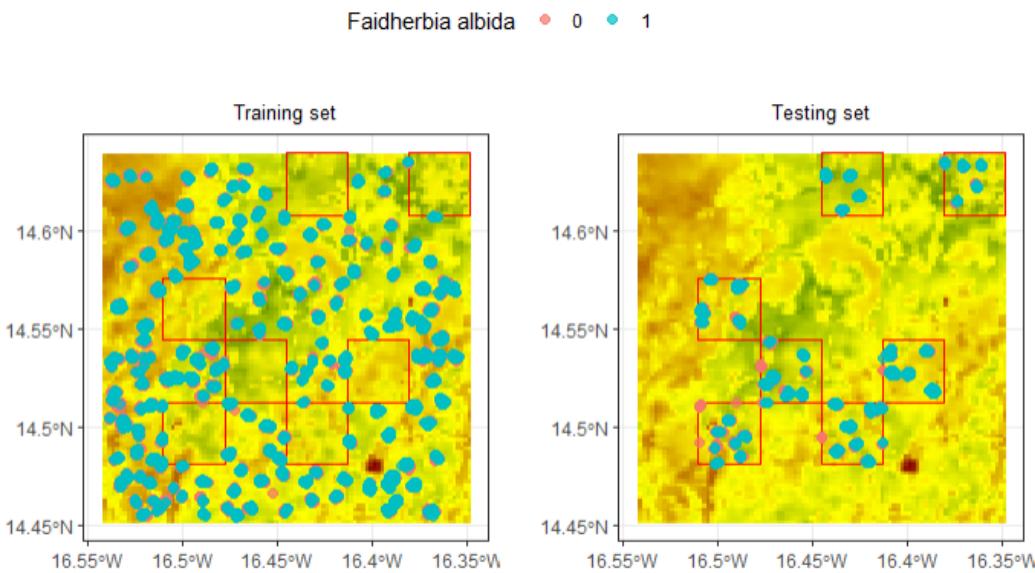


FIGURE 2.8 – Séparation des données de test et d’entraînement de *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l’auteur

2.4 VALIDATION DES MODÈLES

La modélisation a été effectuée avec les blocs spatiaux et sans les blocs c'est à dire en prenant 20% pour données de test et 80% pour données d'entraînement. Pour chacune des étapes de la modélisation, nous avons utilisé les variables explicatives issues de ENFA pour chaque espèce sans ajout des variables à haute résolution spatiale et avec ajout de ces dernières.

Ensuite la validation du modèle a été réalisée sur base de l'AUC (Area Under ROC 20 Curve). L'AUC est reconnu comme étant l'un des évaluateurs indépendants pertinents de seuil de pouvoir discriminant des modèles (Fielding et Bell, 1997); il varie entre 0 et 1. Une valeur de AUC de 0,5 indique que le modèle à une prédiction aléatoire (pas mieux que le hasard). Les modèles avec un AUC de 0,7 sont acceptables, tandis qu'un AUC=0,8 signifie que dans les zones d'occurrence de l'espèce, dans 80 % des cas, les valeurs prédictives sont plus élevées comparativement à celles des absences. Pour une valeur de AUC comprise entre 0,8 et 0,9 nous sommes dans le cas d'une bonne discrimination, et lorsque AUC est comprise entre 0,9 et 1 nous sommes dans le cas d'une excellente discrimination (Kleinbaum et Klein, 2010).

Résultats et Discussion

3.1 RÉSULTATS

3.1.1 Analyse descriptive

3.1.1.1 Analyse de la distribution des espèces suivant le type de sol

Le parc à *Faidherbia albida* est caractérisé par trois types de sol : ferrugineux tropicaux, hydromorphes et hydromorphes salés.

La figure 3.1 montre la proportion des espèces étudiées dans le parc à *Faidherbia albida* suivant les trois types de sol.

Les sols ferrugineux tropicaux sont majoritairement occupés par *Faidherbia albida* avec une proportion d'environ 23%, suivi des sols hydromorphes (environ 18.5%) et enfin par les sols hydromorphes salés (environ 0.5%).

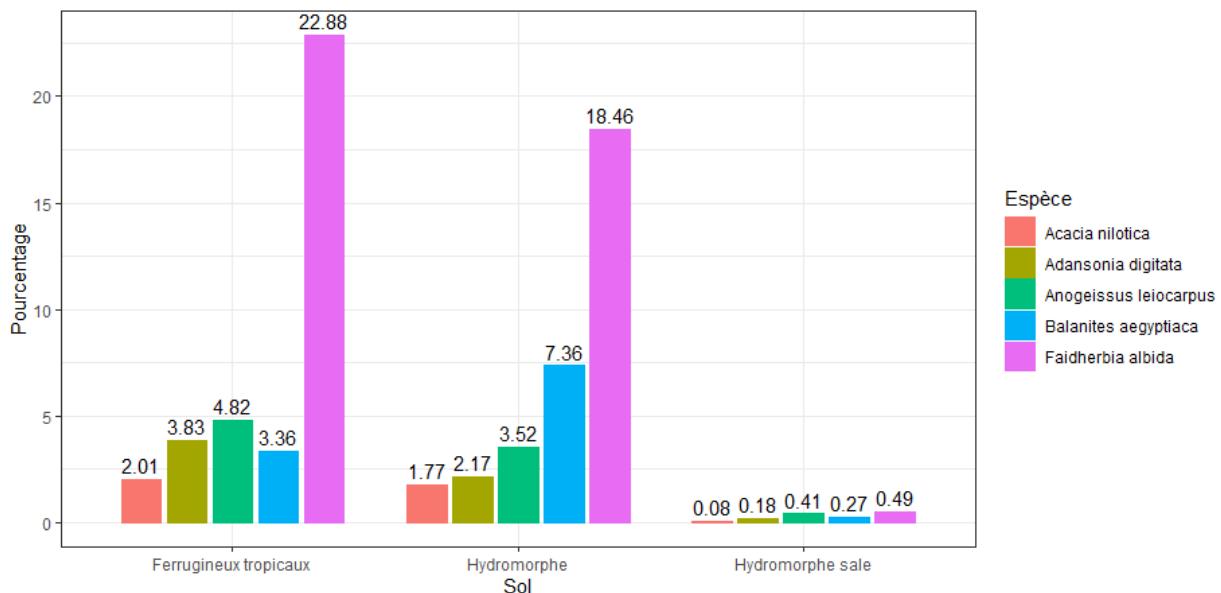


FIGURE 3.1 – Abondance des espèces dans le parc à *Faidherbia albida* suivant le type de sol

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Le parc à *Cordyla pinnata* est caractérisé par quatre types de sols : lithosols, ferrugineux tropicaux lessivés, hydromorphes et ferrugineux tropicaux rubifiés.

Cordyla pinnata occupe majoritairement les sols ferrugineux tropicaux rubifiés avec une proportion d'environ 25.7%, suivi des sols ferrugineux tropicaux lessivés (10.8%) et des sols lithosols(6%).

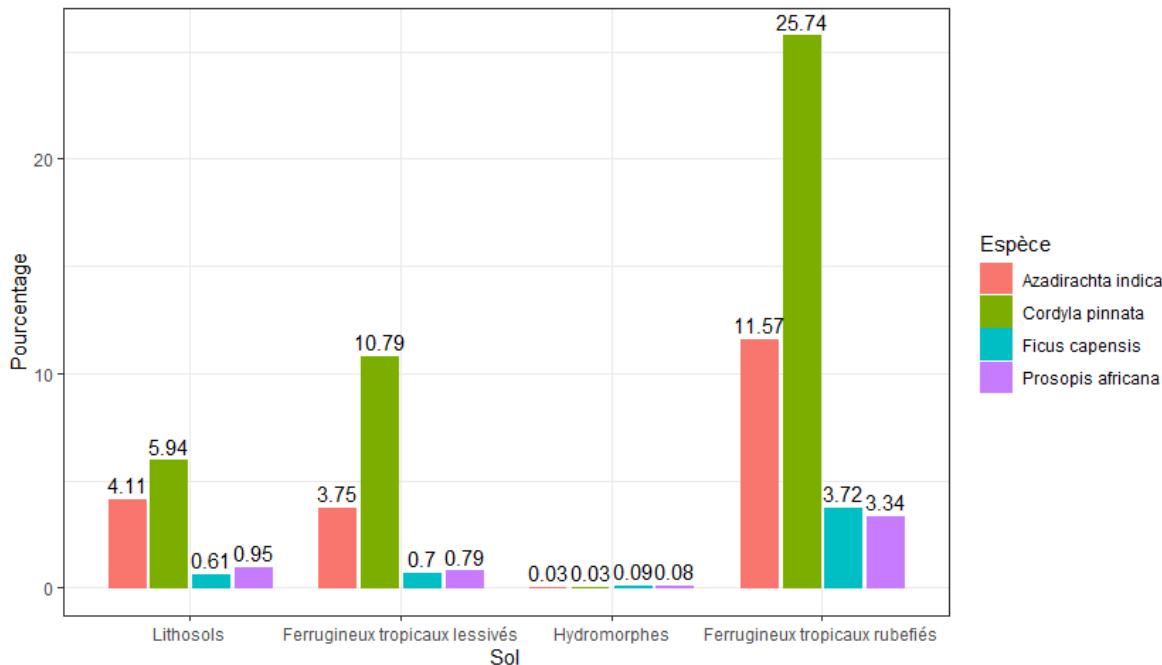


FIGURE 3.2 – Abondance des espèces dans le parc à *Cordyla pinnata* suivant le type de sol

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

En examinant l'existence éventuelle d'une relation entre la présence ou l'absence d'une espèce et le type de sol, le tableau 3.1 nous permet de dire que la présence ou l'absence des espèces ne sont pas toutes liées aux types de sols dans la zone d'étude. En effet, *Acacia nilotica* est la seule espèce où sa présence ou son absence est indépendante du type de sol ($p\text{-value}=0.9436225 > 5\%$, donc on accepte l'hypothèse nulle au seuil 5%). La présence ou l'absence d'*Azadirachta indica* est liée aux types de sols dans le parc à *Cordyla pinnata* au seuil 5% et 10% mais cette liaison n'est pas significative au seuil 1%. Pour les autres espèces étudiées, cette liaison est significative au seuil 10%, donc au seuil 1% et 5%.

TABLE 3.1 – Test du Chi Deux entre les espèces et le type de sol

Espèce	Significativité	p-value
<i>Faidherbia albida</i>	* ; ** ; ***	$7.086493 * 10^{-12}$
<i>Balanites aegyptiaca</i>	* ; ** ; ***	$1.404496 * 10^{-43}$
<i>Anogeissus leiocarpus</i>	* ; ** ; ***	$3.081575 * 10^{-09}$
<i>Adansonia digitata</i>	* ; ** ; ***	5.65839310^{-08}
<i>Acacia nilotica</i>	pas significatif	0.9436225
<i>Cordyla pinnata</i>	* ; ** ; ***	6.48347810^{-09}
<i>Azadirachta indica</i>	** ; *	0.0105
<i>Prosopis africana</i>	* ; ** ; ***	$8.1247 * 10^{-05}$
<i>Ficus capensis</i>	* ; ** ; ***	$5.0251 * 10^{-06}$

*** = < 1% ; ** = < 5% ; * = < 10%

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.1.2 Analyse d'autocorrélation spatiale de la distribution des espèces

De l'analyse du tableau ci-dessous, nous constatons une autocorrélation spatiale négative car le nombre de liaisons Présence-Absence (colonne Liaison) est significativement supérieur à ce que l'on aurait obtenu à partir d'une répartition spatiale aléatoire(colonne z-value). Les valeurs dissemblables (présence et absence) des espèces étudiées dans les deux parcs se regroupent géographiquement : des lieux proches sont plus différents que des lieux éloignés.

TABLE 3.2 – Test de Joincount de la distribution des espèces

Espèce	Liaison	Statistique	Espérance	Variance	z-value
<i>Faidherbia albida</i>	Présence :Absence	1592.30	2252.85	195.97	-47.19
<i>Balanites aegyptiaca</i>	Présence :Absence	671.60	905.38	54.01	-31.81
<i>Anogeissus leiocarpus</i>	Présence :Absence	593.00	739.21	41.60	-22.67
<i>Adansonia digitata</i>	Présence :Absence	387.20	537.59	28.06	-28.39
<i>Acacia nilotica</i>	Présence :Absence	309.60	343.27	16.56	-8.27
<i>Cordyla pinnata</i>	Présence :Absence	856.30	1616.45	140.28	-64.18
<i>Azadirachta indica</i>	Présence :Absence	528.10	1039.66	75.21	-58.99
<i>Prosopis africana</i>	Présence :Absence	239.00	327.04	17.82	-20.85
<i>Ficus capensis</i>	Présence :Absence	258.90	324.36	17.66	-15.58

Source : Projets SERENA et LYSA(2017-2019), Calcul de l'auteur

3.1.1.3 Analyse de la corrélation des prédicteurs

Le coefficient de corrélation linéaire τ donne une mesure de l'intensité et du sens de la relation linéaire entre deux variables (prédicteurs). On s'intéresse ici à son interprétation :

- ★ le coefficient de corrélation est compris entre -1 et $+1$;
- ★ plus le coefficient est proche de 1 , plus la relation linéaire positive entre les variables est forte ;

- * plus le coefficient est proche de -1 , plus la relation linéaire négative entre les variables est forte ;
 - * plus le coefficient est proche de 0 , plus la relation linéaire entre les variables est faible.
- Des figures 3.3 et 3.4, nous notons une forte corrélation entre les variables bioclimatiques. En effet, dans les deux parcs le coefficient de corrélation linéaire τ entre les variables bioclimatiques est proche de -1 ou de 1 .

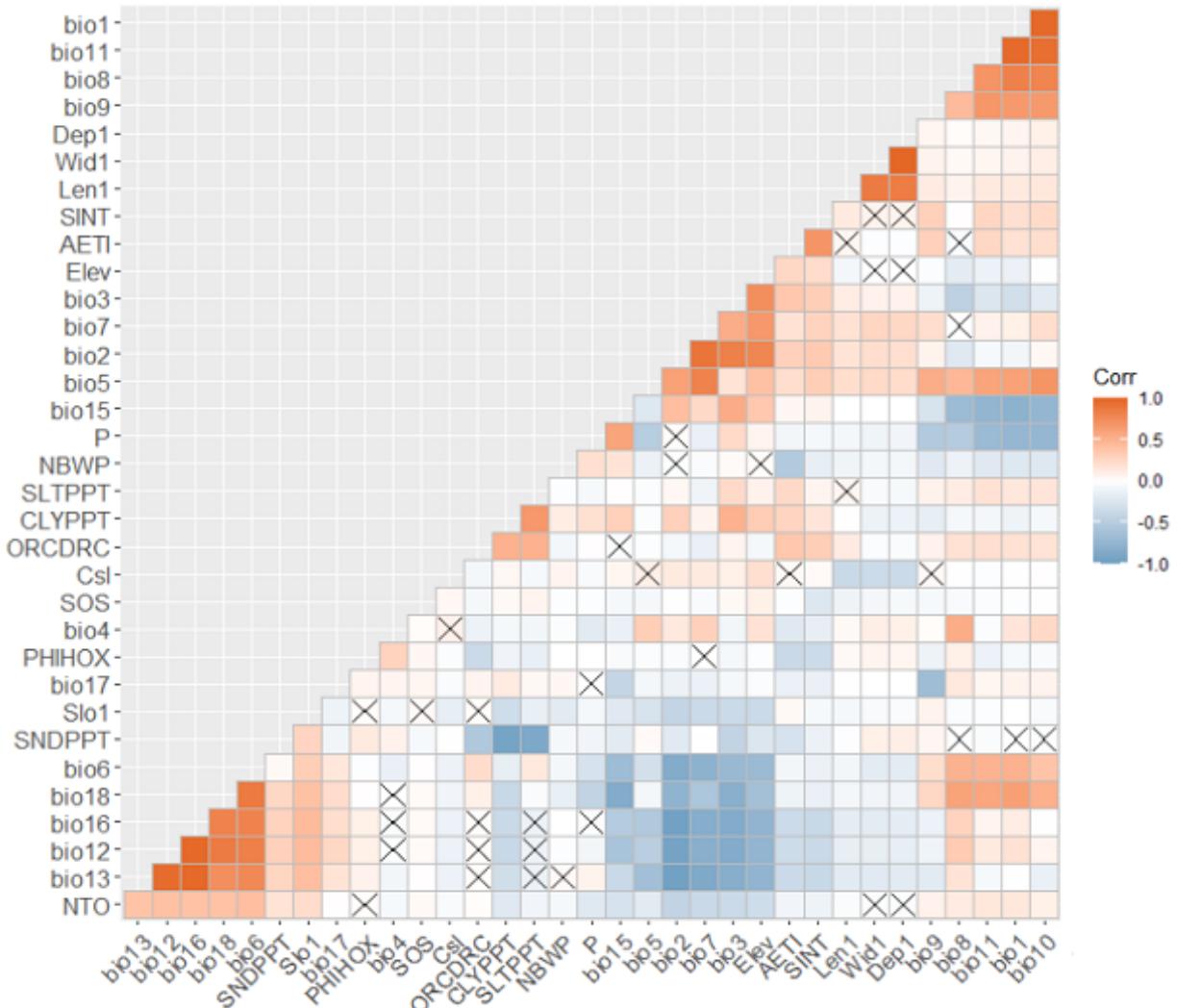


FIGURE 3.3 – Corrélation entre les prédicteurs dans le parc à *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

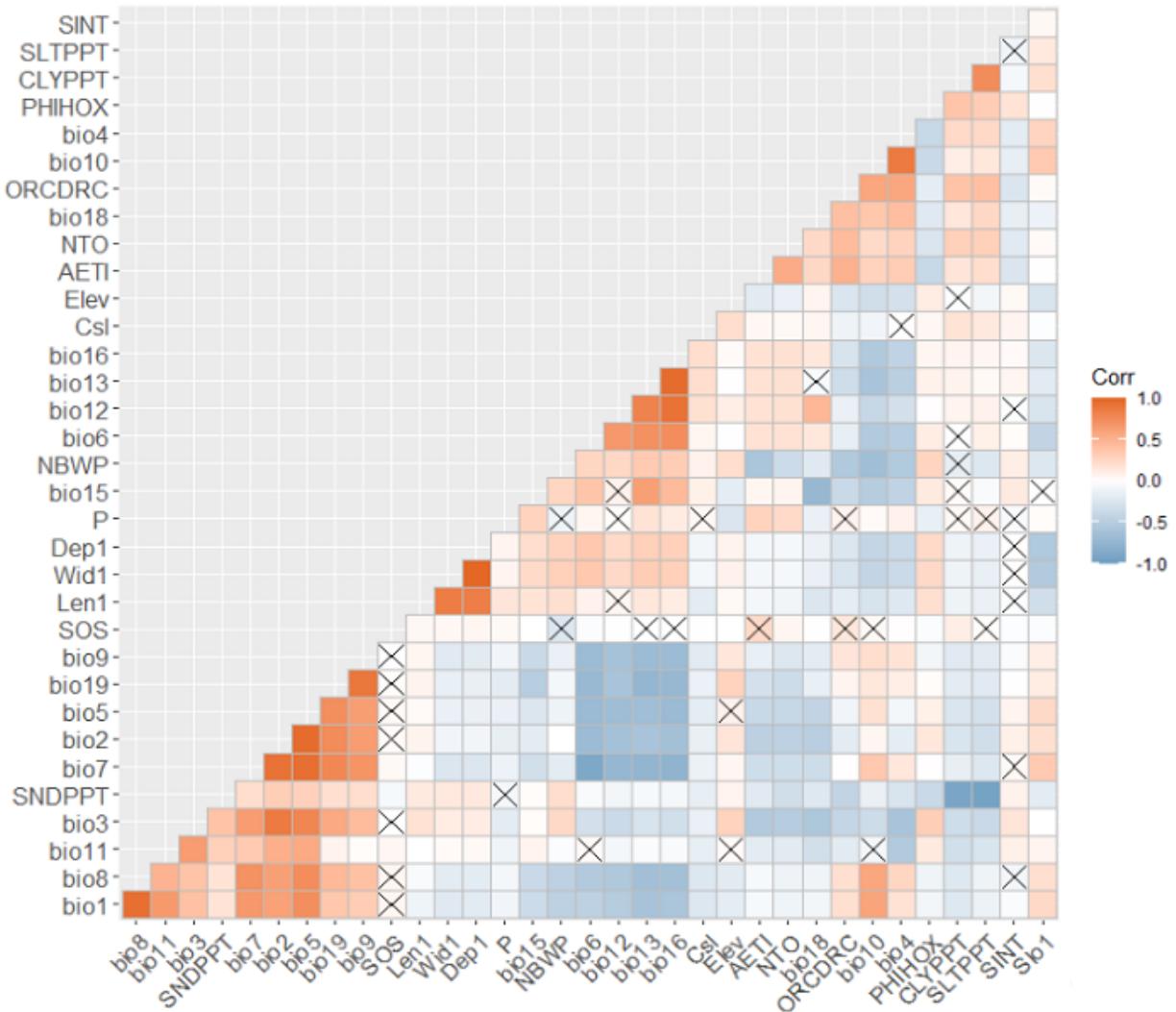


FIGURE 3.4 – Corrélation entre les prédicteurs dans le parc à *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Le test de significativité du coefficient de corrélation linéaire permet de décider si ce lien est significatif, autrement dit si la corrélation observée entre les prédicteurs (dans l'échantillon) existe bel et bien au sein de la population à l'étude.

Dans la logique d'un test d'hypothèses, nous postulons :

$$\begin{cases} H_0 : \tau = 0 \\ H_1 : \tau \neq 0 \end{cases}$$

Si la p-value est supérieure à 5%, nous acceptons l'hypothèse nulle (H_0) et concluons que la corrélation observée entre deux variables est due au hasard. Dans le cas contraire, nous rejetons l'hypothèse nulle et concluons qu'une corrélation entre deux variables existe bel et bien dans la zone d'étude autrement dit la corrélation est significative entre les deux variables.

Dans les figures 3.3 et 3.4, la significativité de la corrélation entre deux variables est illustrée par une croix.

3.1.2 Analyses factorielles

3.1.2.1 Analyse factorielle des correspondances simples dans le parc à *Faidherbia albida*

L' analyse factorielle des correspondances (AFC ou CA pour correspondence analysis en anglais) est une extension de l'analyse en composantes principales (ACP)) pour analyser l'association entre deux variables qualitatives (ou catégorielles).

L'AFC permet de résumer et de visualiser l'information contenue dans le tableau de contingence formé par les deux variables catégorielles. Le tableau de contingence contient les fréquences formées par les deux variables.

L'AFC retourne les coordonnées des éléments des colonnes et des lignes du tableau de contingence. Ces coordonnées permettent de visualiser graphiquement l'association entre les éléments de lignes et de colonnes dans un graphique à deux dimensions.

Lors de l'analyse d'un tableau de contingence, une question typique est de savoir si certains éléments lignes sont associés à certains éléments colonnes. L'analyse factorielle par correspondance est une approche géométrique pour visualiser les lignes et les colonnes d'une table de contingence dans un graphique en nuage de points, de sorte que les positions des points lignes et celles des points colonnes correspondent à leurs associations dans le tableau. Les données correspondent à un tableau de contingence contenant les espèces et leur répartition suivant le type de sol :

- * les lignes sont les différentes espèces ;
- * les colonnes sont le type de sol ;
- * les valeurs sont les fréquences des espèces suivant le type de sol.

Pour interpréter l'AFC, la première étape consiste à évaluer s'il existe une dépendance significative entre les lignes et les colonnes.

Une méthode rigoureuse consiste à utiliser la statistique de χ^2 pour examiner l'association entre les modalités des lignes et celles des colonnes. Dans notre exemple, l'association est très significative (χ^2 : 413.52, $p\text{-value} = 0$).

TABLE 3.3 – Profile ligne des espèces dans le parc à *Faidherbia albida*

Espèce	Ferrugineux tropicaux	Hydromorphe	Hydromorphe sale	Total
<i>Acacia nilotica</i>	52.10	45.94	1.96	100.00
<i>Acacia seyal</i>	28.97	66.36	4.67	100.00
<i>Adansonia digitata</i>	61.95	35.08	2.97	100.00
<i>Anogeissus leiocarpus</i>	55.06	40.25	4.69	100.00
<i>Azadirachta indica</i>	36.40	60.29	3.31	100.00
<i>Balanites aegyptiaca</i>	30.58	66.96	2.46	100.00
<i>Bauhinia rufescens</i>	53.02	43.26	3.72	100.00
<i>Borassus aethiopum</i>	63.71	35.48	0.81	100.00
<i>Celtis integrifolia</i>	60.81	37.84	1.35	100.00
<i>Diospiros mespiliformis</i>	61.76	34.56	3.68	100.00
<i>Faidherbia albida</i>	54.70	44.14	1.16	100.00
<i>Piliostigma reticulatum</i>	34.22	64.00	1.78	100.00
<i>Prosopis juliflora</i>	47.66	52.34	0.00	100.00
<i>Sclerocarya birrea</i>	58.17	39.87	1.96	100.00
<i>Tamarindus indica</i>	52.63	45.39	1.97	100.00
<i>Ziziphus mauritiana</i>	37.21	60.93	1.86	100.00
Total	50.93	46.99	2.08	100.00

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Le tableau de profile ligne nous permet d'avoir :

- * les espèces *Acacia nilotica*, *Adansonia digitata*, *Anogeissus leiocarpus*, *Bauhinia rufescens*, *Borassus aethiopum*, *Celtis integrifolia*, *Diospiros mespiliformis*, *Faidherbia albida*, *Sclerocarya birrea* et *Tamarindus indica* préfèrent vivre dans les sols ferrugineux tropicaux. En effet, ces espèces ont une proportion plus de 50.93%, ce qui représente la part des sols ferrugineux tropicaux dans le parc ;
- * les espèces *Acacia seyal*, *Azadirachta indica*, *Balanites aegyptiaca*, *Piliostigma reticulatum*, *Prosopis juliflora* et *Ziziphus mauritiana* préfèrent les sols hydromorphes ;
- * *Acacia seyal*, *Adansonia digitata*, *Anogeissus leiocarpus*, *Azadirachta indica*, *Balanites aegyptiaca*, *Bauhinia rufescens* et *Diospiros mespiliformis* préfèrent les sols hydro-morphes salés.

L'examen des valeurs propres permet de déterminer le nombre d'axes principaux à considérer. Les valeurs propres correspondent à la quantité d'informations retenue par chaque axe. Les valeurs propres peuvent être utilisées pour déterminer le nombre d'axes à retenir. Il n'y a pas de «règle générale» pour choisir le nombre de dimensions à conserver pour l'interprétation des données.

Une des méthodes pour déterminer le nombre de dimensions est de regarder le graphique des valeurs propres, ordonnées de la plus grande à la plus petite valeur. Le nombre d'axes est déterminé par le point , au-delà duquel les valeurs propres restantes sont toutes relativement petites et de tailles comparables.

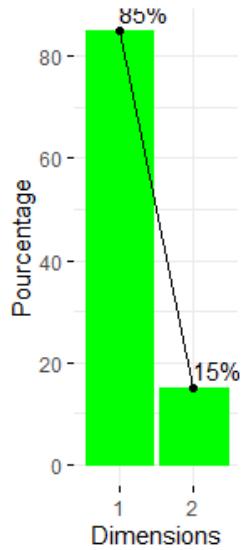


FIGURE 3.5 – Graphique des valeurs propres

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Les dimensions 1 et 2 expliquent environ 85% et 15% de l'inertie totale, respectivement. Cela correspond à un total cumulé de 100% de l'inertie totale retenue par les deux dimensions.

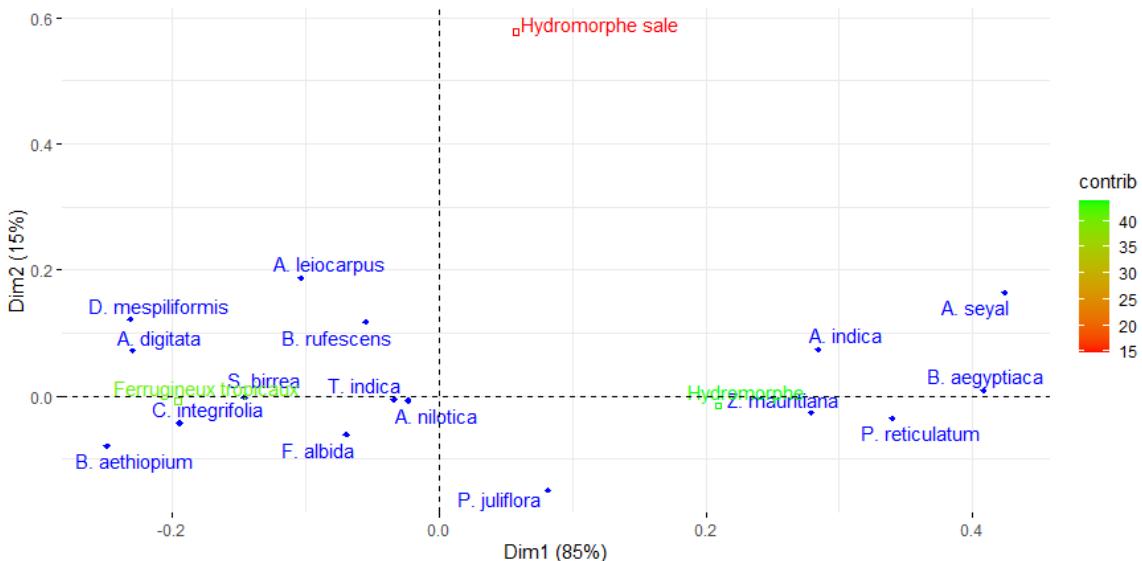


FIGURE 3.6 – Biplot symétrique et colonnes colorées selon leur contribution

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Dans le graphique ci-dessus, les lignes sont représentées par des points bleus et des colonnes par des carrés. La distance entre les points lignes ou entre les points colonnes donne une mesure de leur similitude (ou dissemblance). Les points lignes avec un profil similaire sont proches sur

le graphique. Il en va de même pour les points colonnes.

Ce graphique montre que :

- * les espèces *Acacia seyal*, *Azadirachta indica*, *Balanites aegyptiaca*, *Piliostigma reticulatum*, *Prosopis juliflora* et *Ziziphus mauritiana* sont associées le plus aux sols hydro-morphes ;
- * les espèces *Acacia nilotica*, *Adansonia digitata*, *Anogeissus leiocarpus*, *Bauhinia rufescens*, *Borassus aethiopum*, *Celtis integrifolia*, *Diospiros mespiliformis*, *Faidherbia albida*, *Sclerocarya birrea* et *Tamarindus indica* sont associées le plus aux sols ferrigineux tropicaux.

L'étape suivante, pour l'interprétation, est de déterminer les points lignes et colonnes qui contribuent le plus à la définition des différents axes principaux retenus dans l'AFC.

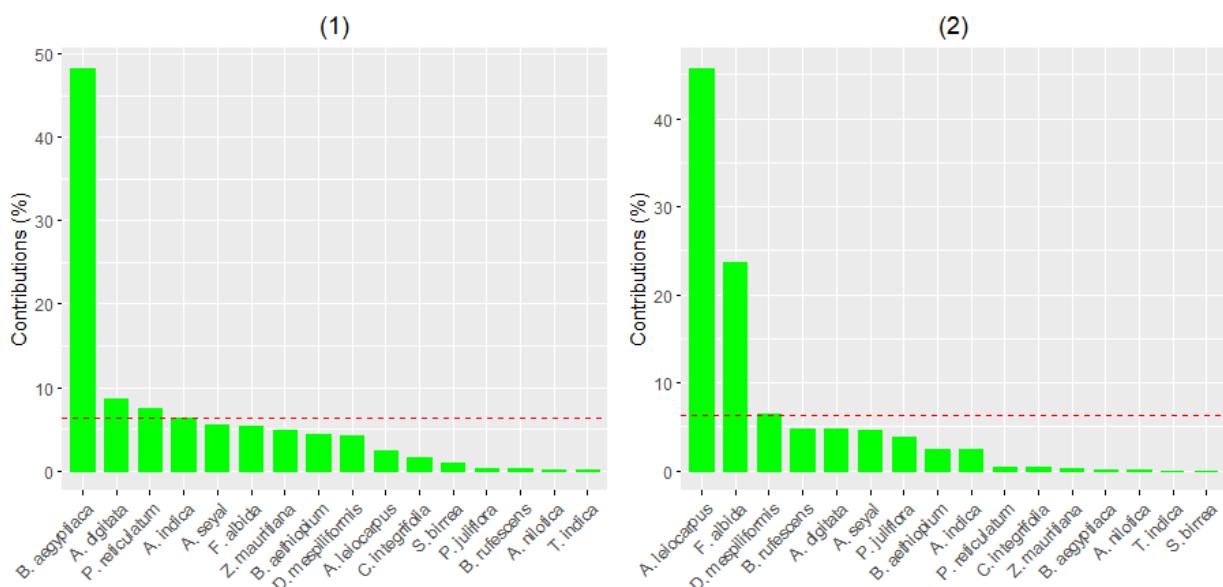


FIGURE 3.7 – Contributions en lignes

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

La droite en pointillée rouge, sur le graphique ci-dessus, indique la valeur moyenne attendue, si les contributions étaient uniformes.

On peut voir que :

- * les espèces *Balanites aegyptiaca*, *Adansonia digitata* et *Piliostigma reticulatum* sont les plus importantes dans la définition de la première dimension ;
- * les espèces *Faidherbia albida*, *Anogeissus leiocarpus* et *Diospiros mespiliformis* contribuent le plus à la deuxième dimension.

De même la droite en pointillée rouge, sur le graphique ci-dessous, indique la valeur moyenne attendue, si les contributions étaient uniformes.

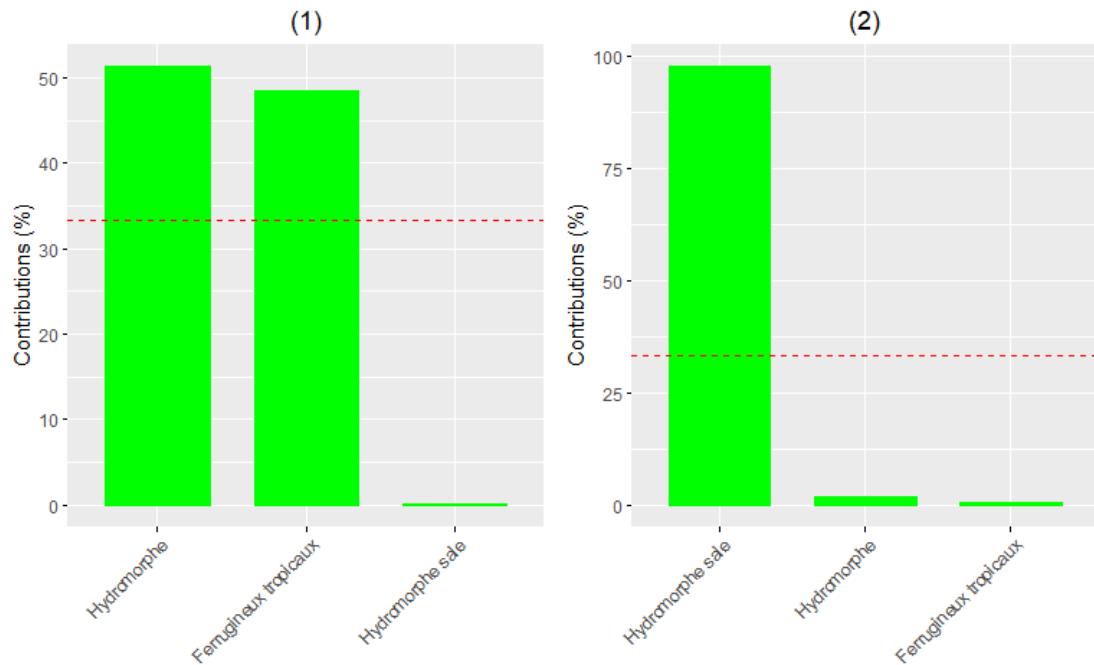


FIGURE 3.8 – Contributions en colonnes

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

On peut également voir que :

- * les types de sol hydromorphe et ferrigineux tropicaux contribuent le plus à la première dimension ;
- * le sol hydromorphe contribue le plus à la deuxième dimension.

Nous terminons par une classification des espèces suivant le type de sol.

- * la classe 1 régroupent les espèces qui préfèrent les sols hydromorphes ;
- * la deuxième classe, les espèces préférant les sols ferrigineux tropicaux ;
- * et la dernière, celles qui sont dans les sols hydromorphes salés.

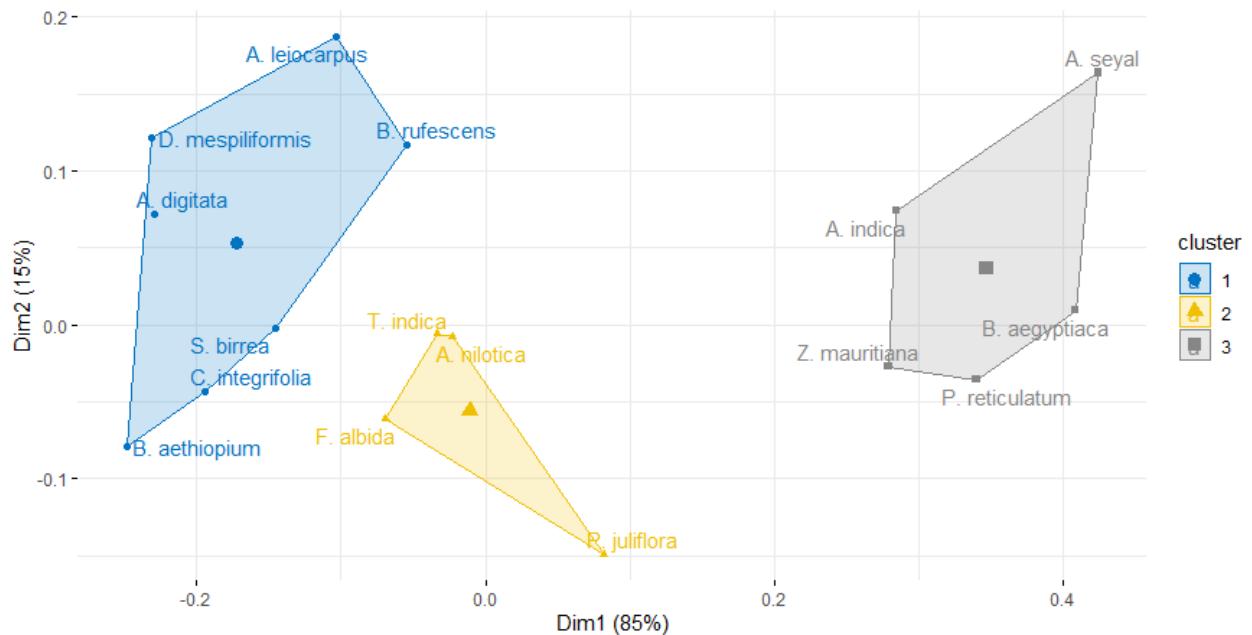


FIGURE 3.9 – Classification Hiérarchique sur Composantes Principales des espèces suivant le type de sol dans le parc à *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.2.2 Analyse factorielle des correspondances simples dans le parc à *Cordyla pinnata*

En utilisant la statistique de χ^2 pour examiner l'association entre les modalités des lignes et celles des colonnes, l'association est très significative (χ^2 : 317.99, $p\text{-value} = 0$). Le tableau 3.4 de profile ligne montre que :

- * les espèces *Anacardium occidentale*, *Anogeissus leiocarpus*, *Azadirachta indica*, *Eucalyptus camaldulensis*, *Prosopis africana* et *Ziziphus mauritiana* préfèrent les sols lithosols ;
- * les espèces *Adansonia digitata*, *Cordyla pinnata* et *Diospyros mespiliformis* préfèrent les sols ferrigineux tropicaux lessives ;
- * les espèces *Anogeissus leiocarpus*, *Borassus aethiopum*, *Ficus capensis*, *Prosopis africana*, *Terminalia macroptera* et *Ziziphus mauritiana* préfèrent les sols hydromorphes ;
- * les espèces *Anacardium occidentale*, *Borassus aethiopum*, *Diospyros mespiliformis*, *Faidherbia albida*, *Ficus capensis*, *Piliostigma reticulatum*, *Prosopis africana*, *Terminalia macroptera* et *Ziziphus mauritiana* préfèrent les sols ferrigineux tropicaux rubefiés.

TABLE 3.4 – Profile ligne des espèces dans le parc à *Cordyla pinnata*

Espèce	lithosols	tropicaux lessives	hydromorphes	tropicaux rubefies	Total
<i>Adansonia digitata</i>	15.28	24.07	0.00	60.65	100.00
<i>Anacardium occidentale</i>	18.99	0.00	0.00	81.01	100.00
<i>Anogeissus leiocarpus</i>	20.57	17.02	1.42	60.99	100.00
<i>Azadirachta indica</i>	21.13	19.27	0.16	59.44	100.00
<i>Borassus aethiopum</i>	11.81	20.14	1.39	66.67	100.00
<i>Cordyla pinnata</i>	13.99	25.37	0.07	60.57	100.00
<i>Diospyros mespiliformis</i>	4.07	29.07	0.00	66.86	100.00
<i>Eucalyptus camaldulensis</i>	43.51	0.76	0.00	55.73	100.00
<i>Faidherbia albida</i>	11.33	11.33	0.00	77.34	100.00
<i>Ficus capensis</i>	11.83	13.61	1.78	72.78	100.00
<i>Piliostigma reticulatum</i>	14.42	8.65	0.00	76.92	100.00
<i>Prosopis africana</i>	18.48	15.25	1.47	64.81	100.00
<i>Terminalia macroptera</i>	14.94	16.09	1.15	67.82	100.00
<i>Ziziphus mauritiana</i>	24.42	9.30	1.16	65.12	100.00
Total	16.26	20.67	0.34	62.73	100.00

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Les dimensions 1 et 2 expliquent environ 59.5% et 30% de l'inertie totale, respectivement. Cela correspond à un total cumulé de 89.5% de l'inertie totale retenue par les deux dimensions.

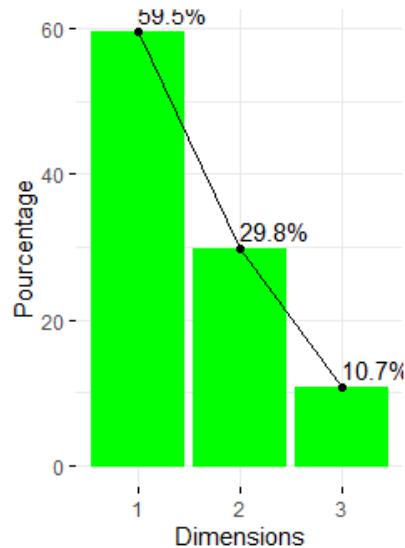


FIGURE 3.10 – Graphique des valeurs propres

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

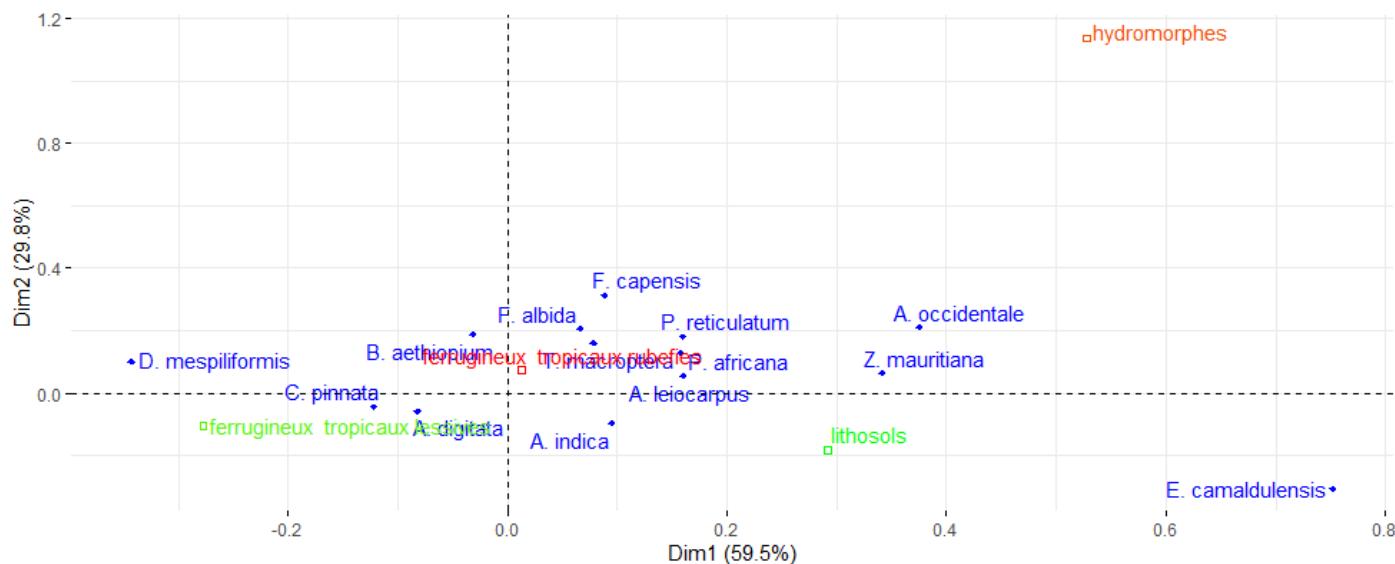


FIGURE 3.11 – Biplot symétrique et colonnes colorées selon leur contribution

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

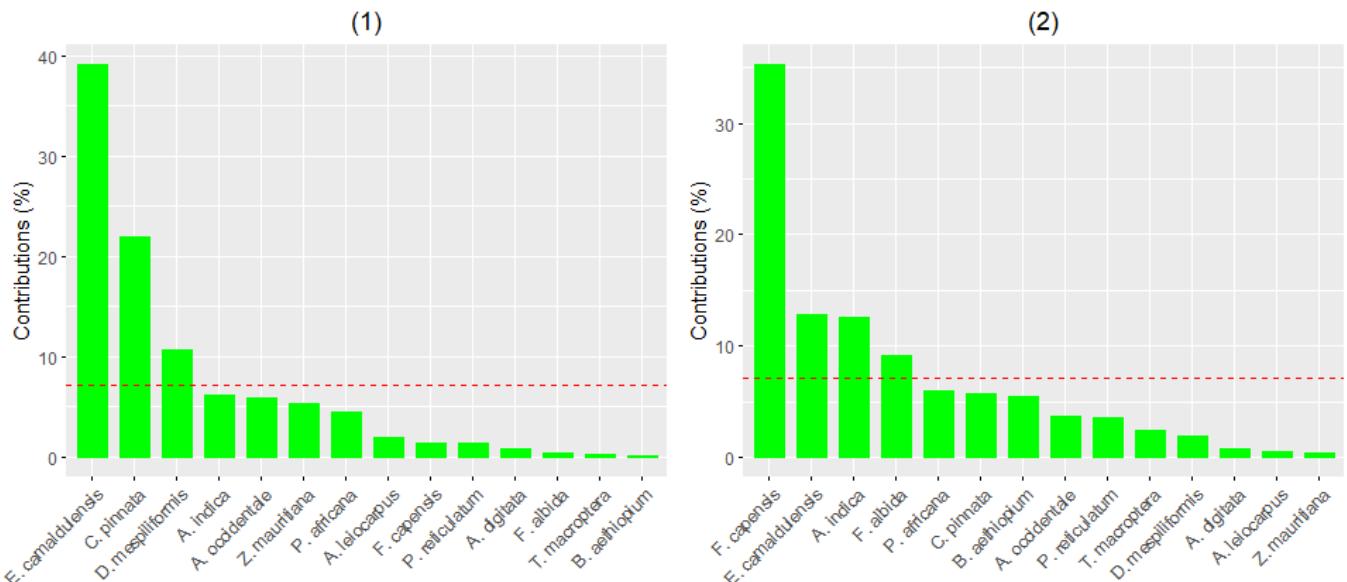


FIGURE 3.12 – Contributions en lignes

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

On peut voir que :

- ★ les espèces *Eucalyptus camaldulensis*, *Cordyla pinnata* et *Diospyros mespiliformis* sont les plus importantes dans la définition de la première dimension ;
- ★ les espèces *Ficus capensis*, *Eucalyptus camaldulensis*, *Azadirachta indica* et *Faidherbia albida* contribuent le plus à la deuxième dimension.

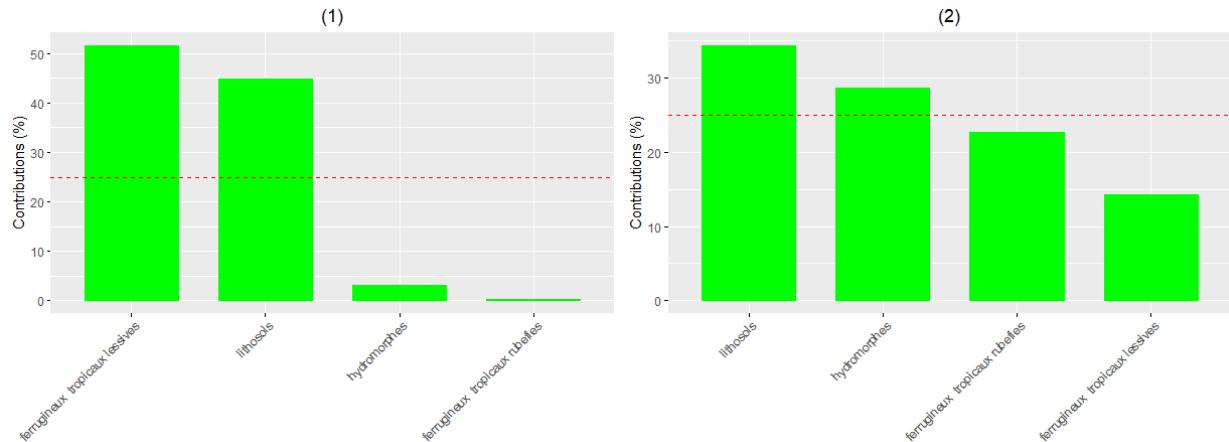


FIGURE 3.13 – Contributions en colonnes

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

On peut également voir que :

- * les types de sol lithosols et ferrigineux tropicaux lessivés contribuent le plus à la première dimension ;
- * le sols hydromorphe et lithosols contribuent le plus à la deuxième dimension.

Nous terminons par une classification des espèces suivant le type de sol.

- * la classe 1 régroupent les espèces qui préfèrent les sols ferrigineux tropicaux lessives ;
- * la deuxième classe, les espèces préférant les les sols hydromorphes ;
- * la troisième, celles qui sont dans les sols ferrigineux tropicaux rubefiés ;
- * et la dernière classe, les espèces qui sont dans les lithosols.

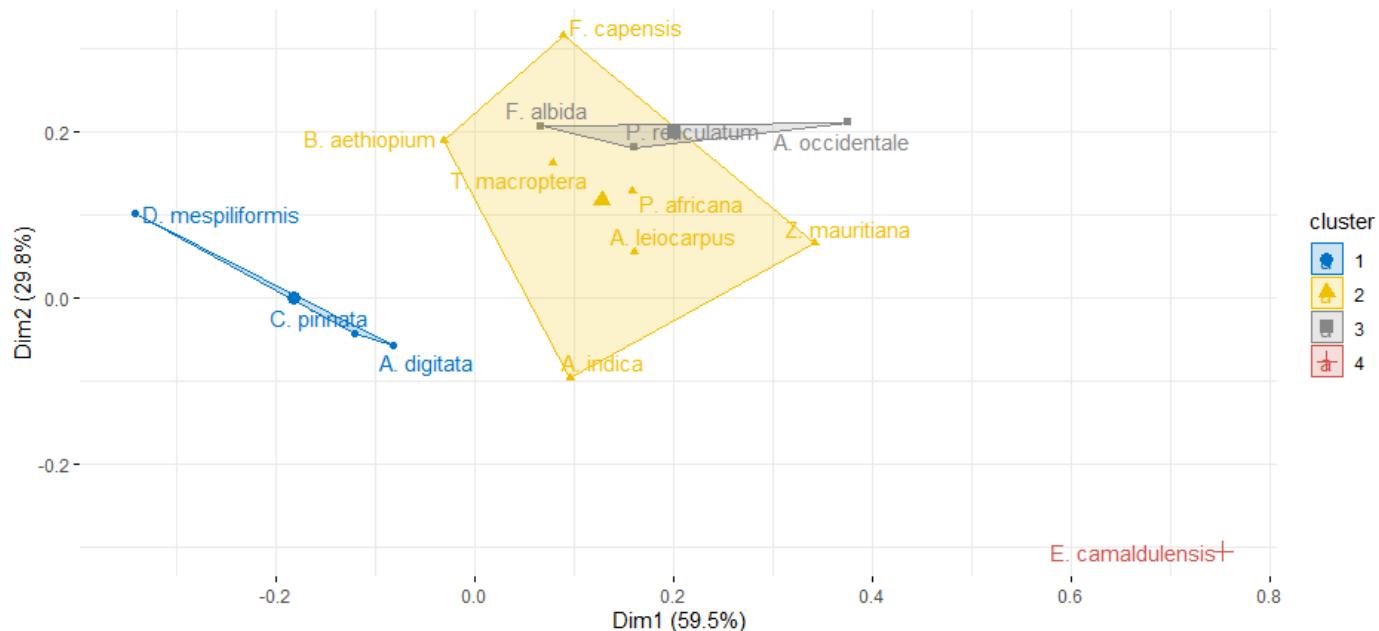


FIGURE 3.14 – Classification Hiérarchique sur Composantes Principales des espèces suivant le type de sol dans le parc à *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.2.3 Analyse factorielle de la niche écologique (ENFA) de la distribution des espèces étudiées

Les résultats de l'analyse ENFA pour nos neuf espèces ligneuses sont donnés en Annexe A et les figures 2.3 et 2.4. Le polygone gris clair montre les conditions environnementales globales disponibles dans la zone d'étude, le polygone gris foncé montre les conditions environnementales où des espèces ligneuses ont été observées (représentation de la niche observée), et le petit cercle blanc correspond au barycentre de sa distribution. Toutes les variables qui se trouvent dans le polygone gris foncé expliquent de manière significative la présence de l'espèce. Ces variables sont illustrées dans la figure 3.15 et dont la description est la suivante :

- ★ les variables explicatives bioclimatiques sont en orange ;
- ★ les variables explicatives de propriétés du sol sont en vert ;
- ★ les variables explicatives de métriques phénologiques sont en jaune ;
- ★ les variables explicatives sur la productivité de l'eau sont en bleu ;
- ★ les variables explicatives topographiques sont en noir.

Zone d'étude	Espèce	Variables explicatives
Niakhar (Parc à <i>Faidherbia albida</i>)	<i>Faidherbia albida</i>	bio5, CLYPPT, bio18, PHIHOX, Elev, SINT, bio7, Dep1, bio4, ORCDRC, bio13, Csl, bio9, bio16, bio15, NTO, bio1, SNDPPT, SOS, AETI, bio17, bio8, bio11, Len1, Wid1, NBWP, P, SLTPPT, Slo1, bio12, bio10
	<i>Balanites aegyptiaca</i>	bio18, bio15, AETI, bio13, NTO, CLYPPT, NBWP, bio4, bio9, Csl, P, Elev, PHIHOX, bio8, bio17, bio11, bio10, bio1, Wid1, Slo1, SOS, SNDPPT, SLTPPT, SINT, ORCDRC, Len1, Dep1
	<i>Anogeissus leiocarpus</i>	bio13, SNDPPT, Dep1, AETI, SOS, bio8, bio4, NBWP, Len1, CLYPPT, Slo1, bio18, bio1, bio15, Wid1, SLTPPT, bio9, ORCDRC, SINT, bio17, NTO, Elev, bio10, PHIHOX, P, Csl, bio11
	<i>Adansonia digitata</i>	AETI, SINT, NBWP, Elev, NTO, CLYPPT, Len1, bio4, SLTPPT, Slo1, P, bio8, Wid1, ORCDRC, bio18, bio17, Dep1
	<i>Acacia nilotica</i>	PHIHOX, ORCDRC, bio6, bio4, Dep1, Slo1, Len1, Elev, AETI, NBWP, SINT, NTO, bio11, CLYPPT, Wid1, SOS, bio9, bio8, bio1, bio18, SNDPPT, SLTPPT, Csl, bio17, P
Nioro (Parc à <i>Cordyla pinnata</i>)	<i>Cordyla pinnata</i>	AETI, NTO, Csl, bio1, bio15, bio11, Slo1, bio4, SOS, bio12, Len1, bio6, SNDPPT, CLYPPT, bio8, bio18, NBWP, bio10, P, SINT, bio13, PHIHOX, SLTPPT, ORCDRC, Dep1, Elev, Wid1, bio16
	<i>Azadirachta indica</i>	bio19, P, bio9, Len1, bio8, bio15, PHIHOX, bio12, Csl, SNDPPT, NBWP, bio11, bio18, SINT, Dep1, SLTPPT, Slo1, SOS, Elev, bio13, ORCDRC, NTO, AETI, Wid1
	<i>Prosopis africana</i>	ORCDRC, NTO, P, CLYPPT, SINT, NBWP, bio9, Elev, bio1, bio11, AETI, Len1, bio19, bio8, Csl, SLTPPT, bio13, Wid1, SNDPPT, Dep, Slo1, SOS
	<i>Ficus capensis</i>	bio1, bio10, NTO, P, SINT, Len1, Slo1, bio19, Elev, bio18, bio11, Dep1, ORCDRC, SNDPPT, Csl, SOS, AETI, bio8, bio4

FIGURE 3.15 – Variables explicatives de la distribution des espèces étudiées

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.3 Modélisation de la distribution des espèces

Dans cette partie, il s'agit de trouver le meilleur modèle permettant de mieux prédire les zones favorables à l'apparition des différentes espèces étudiées suivant les caractéristiques bioclimatiques et écologique de la zone d'étude.

3.1.3.1 Les modèles

La modélisation des espèces étudiées avec les sept modèles est illustrée en Annexe A (section 1.0). En se basant sur l'AUC pour évaluer le meilleur modèle, le tableau 3.5 nous permet de résumer l'information sur le meilleur modèle.

TABLE 3.5 – Meilleur modèle selon l'AUC

Espèce	Type	Ajout de variables à moyenne résolution	Choix	AUC
<i>Faidherbia albida</i>	MaxEnt	Oui	Avec blocs	0.72
<i>Acacia nilotica</i>	MaxEnt	Oui	Avec blocs	0.77
<i>Anogeissus leiocarpus</i>	MaxEnt	Non	Avec blocs	0.77
<i>Ficus capensis</i>	MaxEnt	Non	Avec blocs	0.78
<i>Adansonia digitata</i>	MaxEnt	Oui	Avec blocs	0.84
<i>Prosopis africana</i>	MaxEnt	Non	Avec blocs	0.85
<i>Balanites aegyptiaca</i>	MaxEnt	Oui	Avec blocs	0.89
<i>Azadirachta indica</i>	MaxEnt	Non	Avec blocs	0.93
<i>Cordyla pinnata</i>	MaxEnt	Non	Avec blocs	0.93

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Du tableau 3.5, MaxEnt reste le meilleur algorithme utilisé pour la modélisation de la distribution des espèces étudiées dans les deux parcs. Cependant, MaxEnt reste meilleur lorsque nous choisissons de l'appliquer avec les blocs spatiaux.

L'algorithme fournit une discrimination acceptable(AUC compris entre 0.7 et 0.8) pour les espèces *Faidherbia albida*, *Acacia nilotica*, *Anogeissus leiocarpus* et *Ficus capensis*, une bonne discrimination(AUC compris entre 0.8 et 0.9) pour les espèces *Adansonia digitata*, *Prosopis africana* et *Balanites aegyptiaca*, et une excellente discrimination(AUC compris entre 0.9 et 1) pour les espèces *Azadirachta indica* et *Cordyla pinnata*.

Rappelons que notre première hypothèse de recherche est que l'ajout de variables à haute résolution spatiale améliore les SDM. En effet, cela dépend de l'espèce. Par exemple l'hypothèse est confirmée pour les espèces *Faidherbia albida*, *Acacia nilotica*, *Adansonia digitata* et *Balanites aegyptiaca* et infirmée pour les autres espèces étudiées.

Par contre pour les espèces que lesquelles notre première hypothèse est infirmée, nous ferrons leur prédiction en ajoutant les variables à moyenne résolution spatiale car les AUC prouvent qu'il y a également une bonne discrimination comme l'illustre le tableau ci-dessous.

TABLE 3.6 – Meilleur modèle selon l'AUC et avec l'ajout de variables à moyenne résolution spatiale

Espèce	AUC	Ajout de variables à moyenne résolution	Type	Choix
<i>Faidherbia albida</i>	0.72	Oui	MaxEnt	Avec blocs
<i>Anogeissus leiocarpus</i>	0.73	Oui	MaxEnt	Avec blocs
<i>Ficus capensis</i>	0.74	Oui	MaxEnt	Avec blocs
<i>Acacia nilotica</i>	0.77	Oui	MaxEnt	Avec blocs
<i>Prosopis africana</i>	0.82	Oui	MaxEnt	Avec blocs
<i>Adansonia digitata</i>	0.84	Oui	MaxEnt	Avec blocs
<i>Cordyla pinnata</i>	0.86	Oui	MaxEnt	Avec blocs
<i>Azadirachta indica</i>	0.87	Oui	MaxEnt	Avec blocs
<i>Balanites aegyptiaca</i>	0.89	Oui	MaxEnt	Avec blocs

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Les cartes de probabilités de niches écologiques des espèces étudiées dans le parcs à *Faidherbia albida* issues du modèle Maximum d'Entropie (MaxEnt) sont représentées sur la figure ci-après. Elles sont différentes d'une espèce à l'autre.

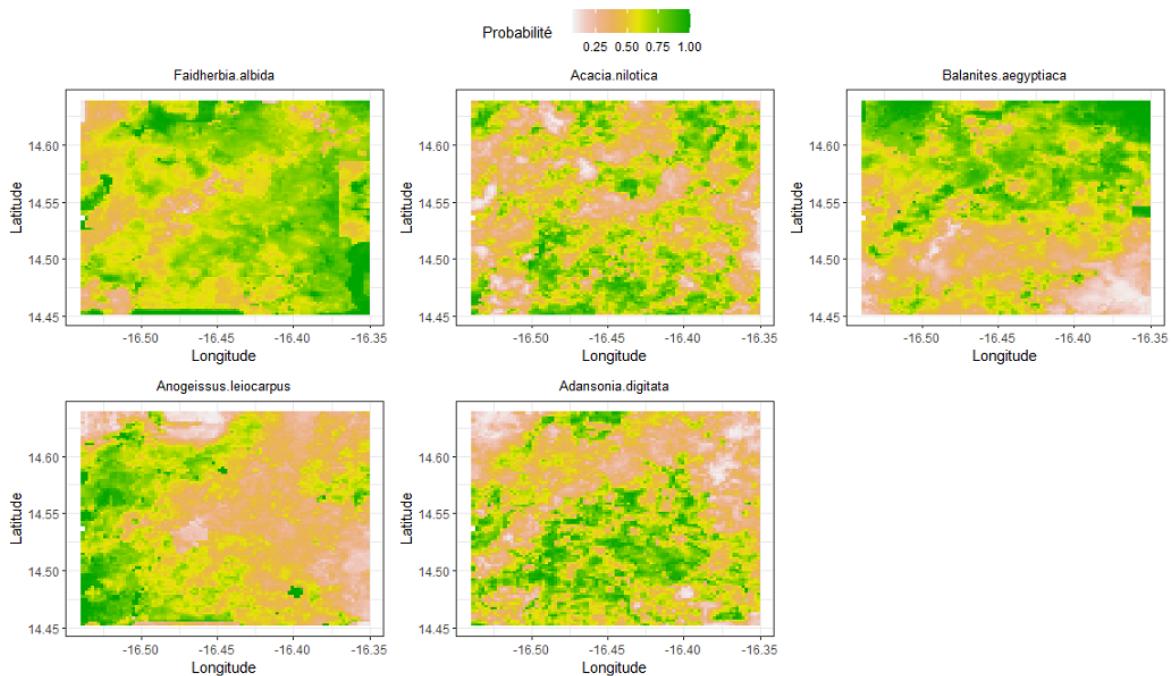


FIGURE 3.16 – Cartes de probabilité des espèces étudiées dans le parc à *Faidherbia albida* issues du modèle MaxEnt

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Celles du parc à *Cordyla pinnata* sont représentées sur la figure ci-dessous.

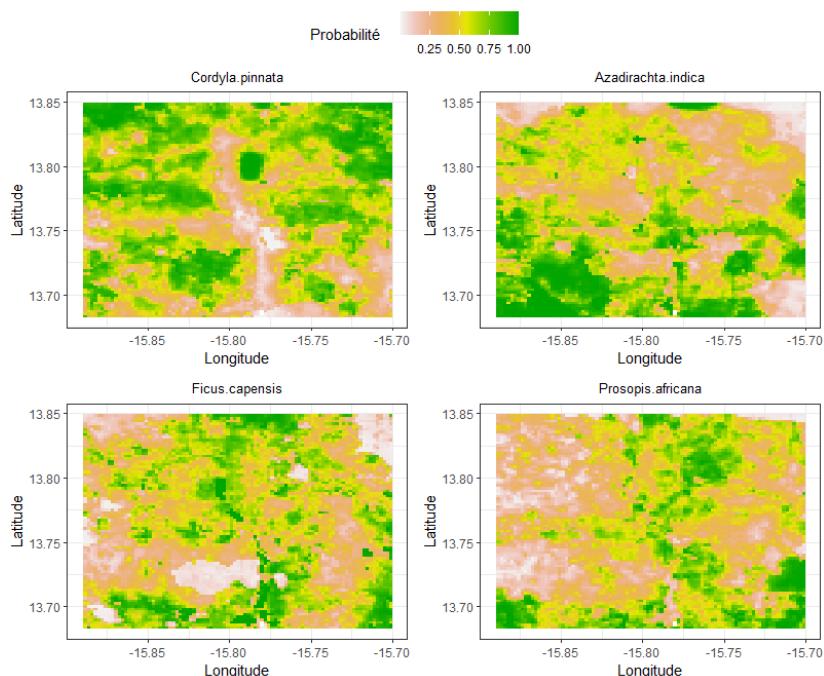


FIGURE 3.17 – Cartes de probabilité des espèces étudiées dans le parc à *Cordyla pinnata* issues du modèle MaxEnt

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Les cartes ci-dessus, sont les cartes de probabilités prédictées des espèces étudiées. Elles sont différentes d'une espèce à l'autre. Notons que ce sont des probabilités de présence et d'absence qui sont prédictées dans chaque pixel de résolution 250 mètres.

Cependant, pour savoir dans quels pixels se trouvent réellement une espèce nous fixons un seuil appelé seuil de positivité (cut-of). A l'intérieur de chaque pixel, si la probabilité prédictée (présence ou absence) est supérieure au seuil alors l'espèce est présente dans le dit pixel.

TABLE 3.7 – Le seuil de positivité fixé par le modèle des espèces étudiées

Espèce	cut-of ou seuil de positivité
<i>Faidherbia albida</i>	0.5965889
<i>Anogeissus leiocarpus</i>	0.4864574
<i>Ficus capensis</i>	0.4780085
<i>Acacia nilotica</i>	0.4858213
<i>Prosopis africana</i>	0.467388
<i>Adansonia digitata</i>	0.7080605
<i>Cordyla pinnata</i>	0.6292833
<i>Azadirachta indica</i>	0.6239142
<i>Balanites aegyptiaca</i>	0.5391948

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

A partir du seuil de positivité (cut-off) nous transformons les prédictions du modèle (probabilités) en un score binaire (présence ou absence) illustré dans les figures 3.18 et 3.19.

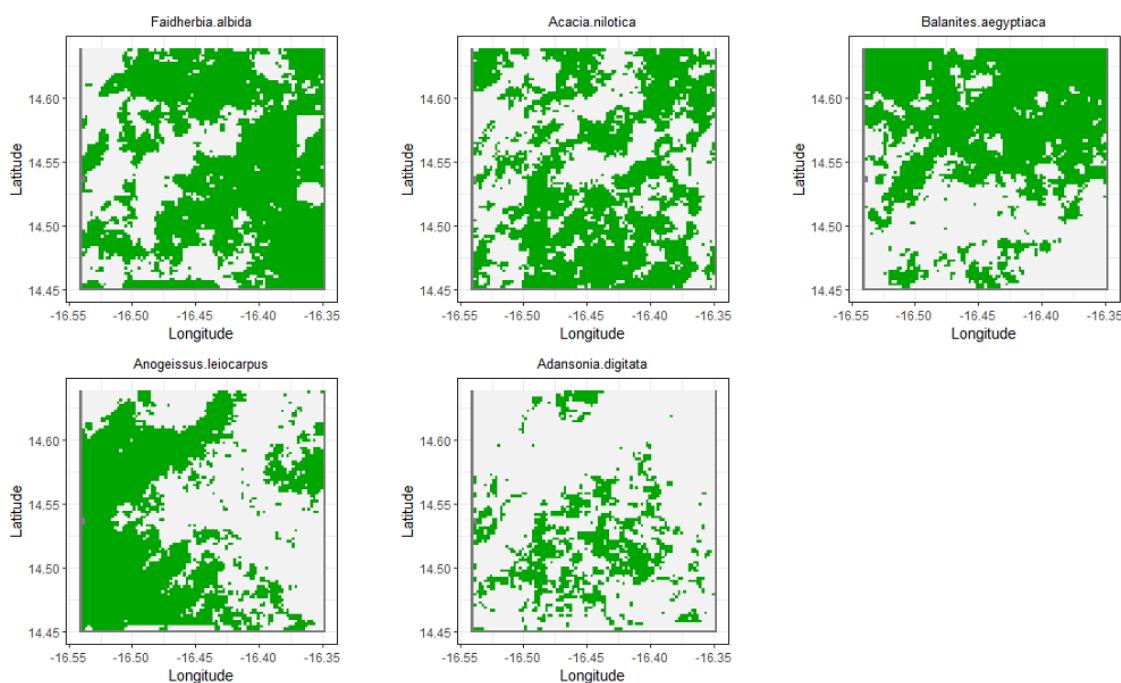


FIGURE 3.18 – Cartes de présence des espèces étudiées dans le parc à *Faidherbia albida* issues du modèle MaxEnt

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

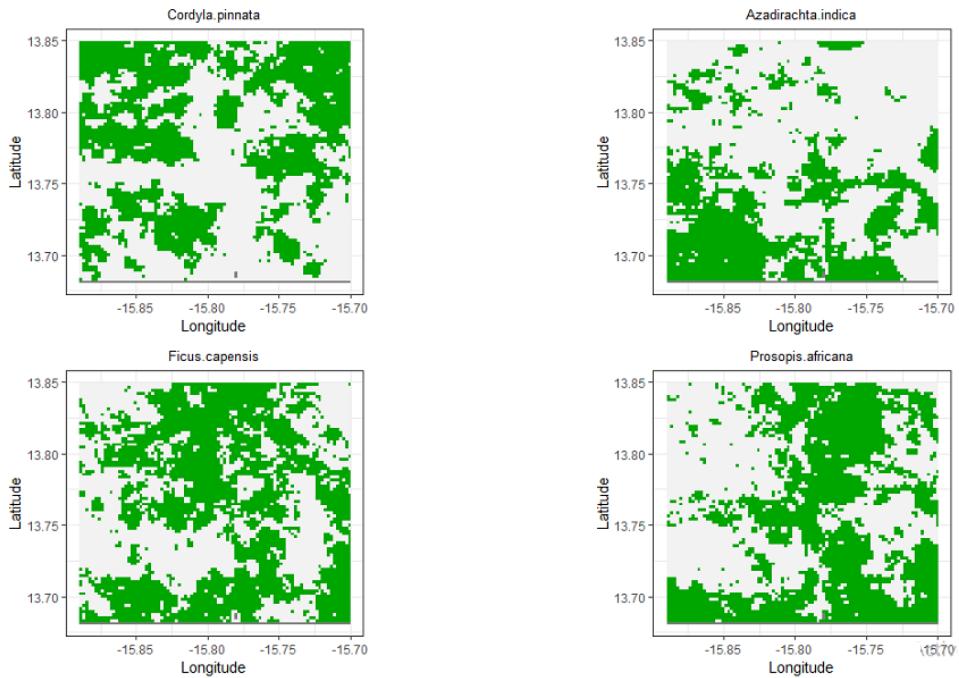


FIGURE 3.19 – Cartes de présence des espèces étudiées dans le parc à *Cordyla pinnata* issues du modèle MaxEnt

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4 Réponse écologique des espèces étudiées

Le modèle MaxEnt a été mis à contribution pour apprécier la réponse écologique des espèces étudiées vis-à-vis des variables climatique et écologique.

A partir de l'importance des variables (annexe) issues du modèle MaxEnt, nous présentons la réponse écologique des espèces étudiées avec les trois meilleures variables selon leur contribution.

3.1.4.1 Réponse écologique de *Faidherbia albida*

Les résultats du modèle MaxEnt illustrés sur les graphes de la figures 3.20 indiquent que *Faidherbia albida* a une abondance dans les biotopes où la température maximale du mois le plus chaud (bio5) est comprise entre 25 et 35°C, la proportion d'argile comprise entre 0 et 10 en terre fine et une précipitation du trimestre le plus chaud (bio18) supérieure à 140mm. Cette abondance n'est plus le cas lorsque la température maximale du mois le plus chaud (bio5) et la proportion d'argile deviennent plus grandes, et la précipitation du trimestre le plus chaud (bio18) est faible.

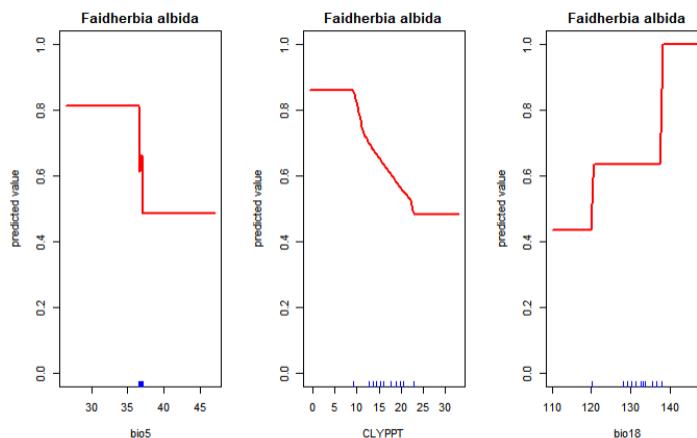


FIGURE 3.20 – Réponse écologique de *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.2 Réponse écologique de *Balanites aegyptiaca*

Balanites aegyptiaca a une faible abondance (figure 3.21) dans les biotopes où la précipitation du trimestre le plus chaud (bio18) est comprise entre 110 et 128 mm, la saisonnalité des précipitations (bio15) supérieur à 155 mm et une abondance fixe quelque soit le niveau d'EvapoTranspiration et Interception réelles (AETI). Son abondance est très faible lorsque la précipitation du trimestre le plus chaud (bio18) est élevée, et la saisonnalité des précipitations (bio15) est faible.

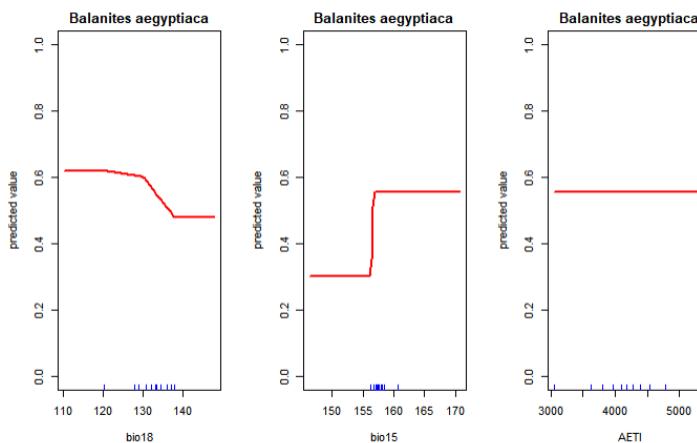


FIGURE 3.21 – Réponse écologique de *Balanites aegyptiaca*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.3 Réponse écologique d'*Anogeissus leiocarpus*

Anogeissus leiocarpus est abondante (figure 3.22) dans les biotopes où la précipitation du mois le plus humide (bio13) est supérieur à 192 mm, la proportion de sable (SNDPPT) plus élevée et une abondance faible pour Dep1.

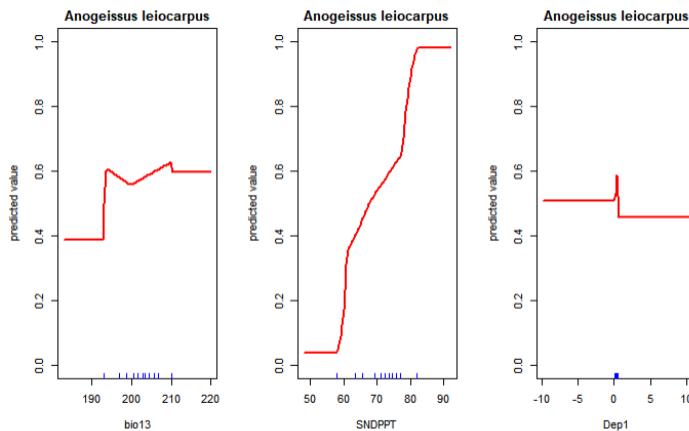


FIGURE 3.22 – Réponse écologique d'*Anogeissus leiocarpus*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.4 Réponse écologique d'*Adansonia digitata*

Adansonia digitata est abondante (figure 3.23) dans les biotopes où l'EvapoTranspiration et Interception réelles (AETI) est élevée. Cette abondance est aussi remarquée lorsque la productivité de la végétation sur la saison des cultures (SINT) est plus en plus élevée mais diminue lorsque cette dernière est supérieure à 23000 terre fine. L'abondance d'*Adansonia digitata* est importante lorsque la productivité nette de l'eau de biomasse (NBWP) est comprise entre 1100 et 1200 terre fine. Elle baisse lorsque cette dernière est supérieure à 1300 terre fine.

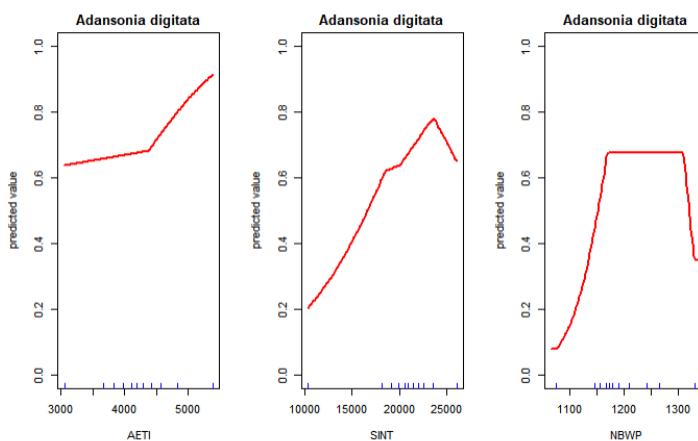


FIGURE 3.23 – Réponse écologique d'*Adansonia digitata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.5 Réponse écologique d'*Acacia nilotica*

Acacia nilotica est abondante(figure 3.24) dans les biotopes dont le pH du sol (PHIHDX) est inférieur à 7 c'est dans les biotopes acides , le carbone organique (ORDRC) compris entre 5 et 8 terre fine et la température minimale du mois le plus froid (bio6) supérieure à 16°C. Cette

abondance devient faible lorsque le milieu est basique, le carbone organique inférieur à 5 terre fine et supérieur à 8 terre fine, et la température minimale du mois le plus froid inférieure à 16°C.

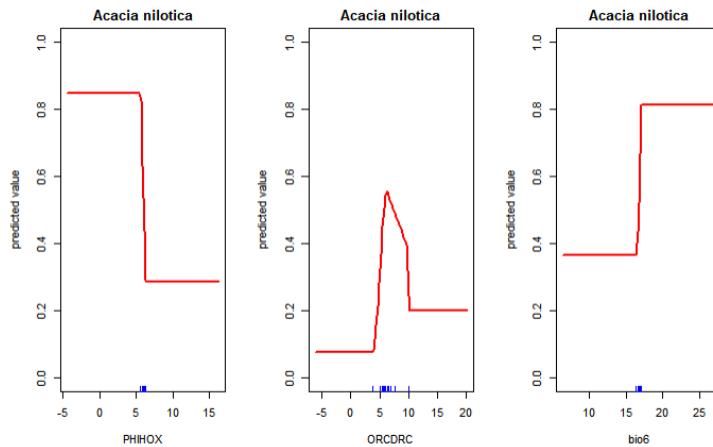


FIGURE 3.24 – Réponse écologique d'*Acacia nilotica*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.6 Réponse écologique de *Cordyla pinnata*

Selon les résultats du modèle MaxEnt (figure 3.25), l'espèce *Cordyla pinnata* est rencontrée dans le parc ayant des caractéristiques écologique et topographique suivantes : l'EvapoTranspiration et Interception réelles (AETI) inférieur à 6000, la quantité d'Azote inférieure à 450 terre fine et Csl positif. *Cordyla pinnata* devient rare lorsque l'EvapoTranspiration et Interception réelles et la quantité d'Azote sont élevées, et Csl négatif.

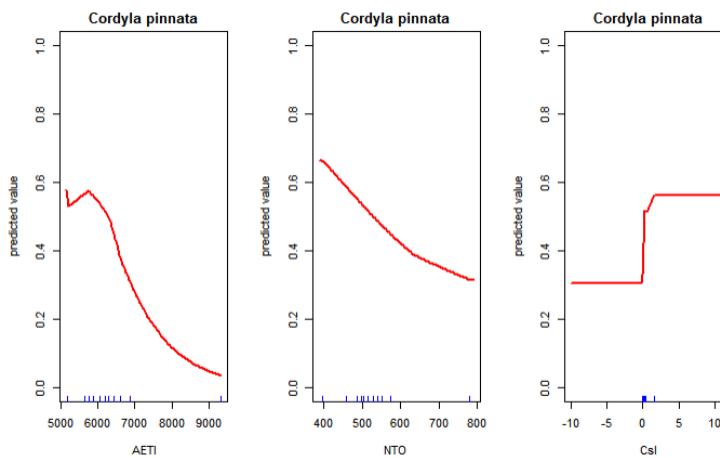


FIGURE 3.25 – Réponse écologique de *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.7 Réponse écologique d'*Azadirachta indica*

Une interprétation des graphes de la figure 3.26 illustrant la distribution potentielle des présences de l'espèce *Azadirachta indica* a montré que l'espèce est abondante dans les régions où la précipitation du trimestre le plus froid (bio19) est faible, la quantité de Phosphore comprise entre 200 et 250 terre fine, et la température moyenne du trimestre le plus sec (bio9) supérieure à $28^{\circ}C$.

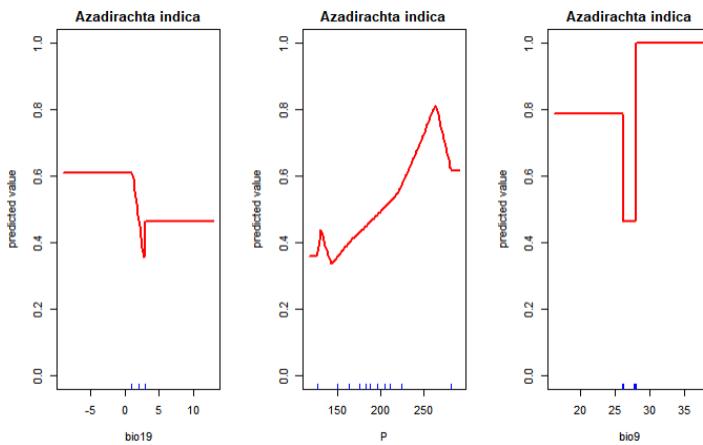


FIGURE 3.26 – Réponse écologique d'*Azadirachta indica*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.8 Réponse écologique de *Ficus capensis*

Le modèle MaxEnt (figure 3.27) a indiqué que l'espèce *Ficus capensis* est retrouvée dans les biotopes où la température moyenne annuelle (bio1) est supérieure à $27^{\circ}C$, la température moyenne du trimestre le plus chaud(bio10) est comprise entre $29^{\circ}C$ et $31^{\circ}C$, et lorsque la quantité d'Azote est supérieure à 450 terre fine. *Ficus capensis* devient rare lorsque $27^{\circ}C$, la température moyenne du trimestre le plus chaud est inférieur à $29^{\circ}C$ et supérieur à $31^{\circ}C$, et la quantité d'Azote est supérieure à 700 terre fine.

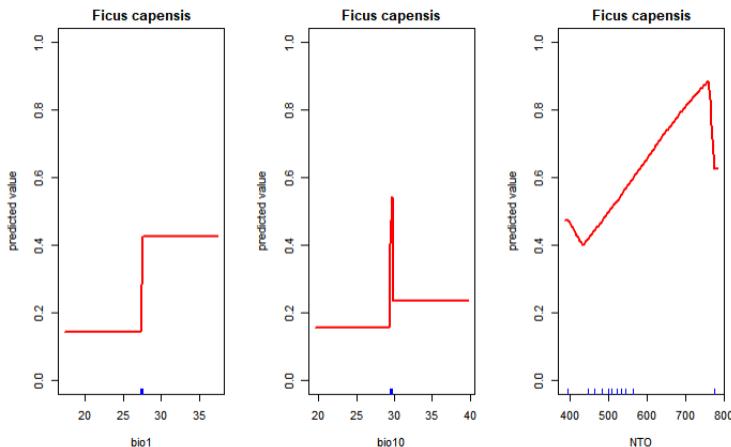


FIGURE 3.27 – Réponse écologique de *Ficus capensis*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.1.4.9 Réponse écologique de *Prosopis africana*

Selon les résultats du modèle MaxEnt (figure 3.28), l'espèce *Prosopis africana* est rencontrée dans le parc ayant des caractéristiques écologiques suivantes : le carbone organique (ORDRC) comprise entre 5 et 8 terre fine, la quantité d'Azote supérieure à 400 terre fine, et la quantité de Phosphore inférieure à 150 terre fine.

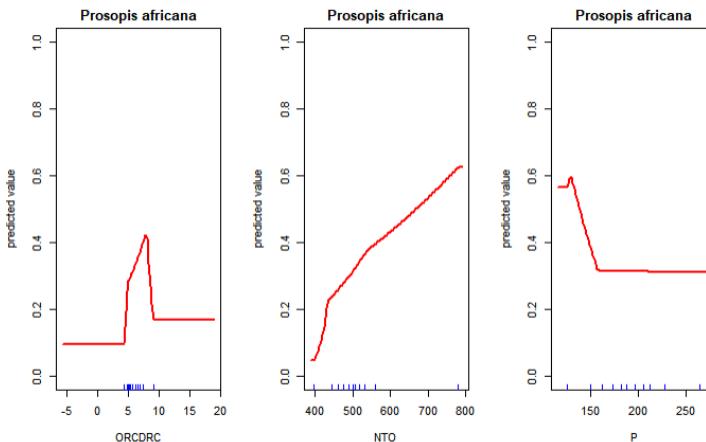


FIGURE 3.28 – Réponse écologique de *Prosopis africana*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

3.2 DISCUSSION

En modélisation de niche, il est crucial de définir en amont du processus la taille de l'aire de background à considérer lorsqu'il s'agit d'un modèle utilisant uniquement les données de présence (Chefaoui et Lobo, 2008). Il importe donc de connaître la distance à laquelle il faut échantillonner les données de pseudo-absences par rapport à la position géographique connue des données de présence (Van der Wal et al., 2009a). Il ressort que les modèles de niches sont sensibles à l'étendue de la zone d'étude. Van der Wal et al. (2009a), en étudiant essentiellement des espèces animales, ont conclu que les modèles statiques ont de faibles performances lorsque les points de pseudo-absences sont choisis dans les zones de présence de l'espèce, ou lorsque l'échelle de distribution est petite. Notre travail a donc l'originalité de confirmer ces résultats sur des espèces ligneuses des deux parcs agroforestiers du bassin arachidier sénégalais. Nos résultats démontrent que pour l'ensemble des sept modèles utilisés, les pseudo-absences doivent être choisis dans les zones de présence des espèces afin d'obtenir un modèle parcimonieux de bonne discrimination.

Lors de l'évaluation des performances des SDM, il est courant de tester dans quelle mesure les prédictions correspondent aux observations à un ensemble d'emplacements (Franklin, 2010; Guisan et al., 2017; Peterson et al., 2011). Alors que les premières applications des SDM avaient tendance à se concentrer sur les mesures statistiques de l'ajustement du modèle aux données utilisées pour former le modèle, l'attention s'est progressivement tournée vers les tests sur des données indépendantes (Elith et Leathwick, 2009). Étant donné que les données totalement indépendantes (telles que les nouvelles enquêtes) sont rarement disponibles, une approche courante consiste à sous-échantillonner les données disponibles pour la modélisation. En écologie, cela implique généralement de diviser les données en sous-ensembles pour la formation et les tests (également appelés respectivement étalonnage et validation) (Franklin, 2010 ; Radosavljevic et Anderson, 2014). Les données de formation sont utilisées pour ajuster le modèle et les données de test pour évaluer les performances. C'est ce qu'on appelle la validation du modèle (Hastie, Tibshirani et Friedman, 2009), avec des variantes basées sur des stratégies de fractionnement et de test. Dans la présente étude, nous avons procédé à la modélisation avec la séparation des données de tests et d'entraînement(constitution des blocs spatiaux) d'une part, et d'autre part, nous avons choisi 80% de données d'entraînement et 20% de données de test. L'un des défis de l'utilisation de blocs spatiaux est de choisir la taille optimale des blocs ou des tampons (Trachsel et Telford, 2016). La plage d'autocorrélation spatiale dans les résidus du modèle a été utilisée pour définir la distance de séparation optimale entre les ensembles d'entraînement et d'essai (Roberts et al., 2017; Telford et Birks, 2009; Trachsel et Telford, 2016). Elle est 6481 mètres dans le parc à *Cordyla pinnata* et de 8393 mètres pour le parc à *Faidherbia albida*. Dans notre travail, nous avons montré que généralement la modélisation avec les modèles spatiaux améliorent les modèles de distribution des espèces. Ce qui va dans le même sens

que les travaux de Roberts et al., 2017.

L'analyse écologique des facteurs de niche a été développée par Hirzel et al. (2002) pour analyser la position de la niche dans l'espace écologique. La méthode utilisée (ENFA) est spécifique pour l'identification des niches écologiques et permet de sélectionner les variables contribuant à la niche écologique des espèces.

L'ENFA fait ressortir pour une espèce donnée, le sous-espace de l'espace globale qui lui est favorable ainsi que les variables qui expliquent de manière significative la présence de celle-ci dans ce sous-espace. Pour chaque espèce dans les deux parcs, nous avons utilisés ENFA pour la sélection des variables qui expliquent de manière significatives sa présence. Cette méthode est largement utilisée dans les études de recherche pour la sélection des variables dont la plus récente est celle de Fallou NIAKH (2019).

En 2004, Guan et Krone on publié un travail qui a prouvé l'importance de l'utilisation des analyses de régression logistique dans les études d'écologie des espèces. Manel et al. (2001) soulignent d'ailleurs que le modèle logistique est la méthode de référence pour obtenir les réponses écologiques des espèces.

Nos résultats montrent que le modèle logistique utilisé est moins performant que MaxEnt, ce qui va dans le même sens que les conclusions de Segurado et Araujo (2004) ou de Wisz et al. (2008), de Fallou NIAKH (2019) qui a prouvé que MaxEnt reste le meilleur algorithme des modèles de distribution des espèces animales notamment les *Culicoïdes*. Toutefois, ce résultat ne confirme pas ceux de Wisz et Guisan (2009) qui ont suggéré que les modèles qui utilisent à la fois les données de présence et d'absence ou des pseudo-absences sont plus robustes que ceux utilisant uniquement les données de présence.

Le modèle logistique est limité par la nécessité de disposer de données d'absence qui semble pouvoir être correctement remplacées par des pseudo-absences judicieusement distribuées selon la méthode proposée par Van der Wal et al. (2009a). L'autre limite du modèle logistique est l'effectif de l'échantillon. En effet, la procédure de calcul du modèle logistique (Modèle général linéaire/non-linéaire) est basée sur la méthode du maximum de vraisemblance qui réclame de 30 à 50 observations par effet à inclure dans le modèle (Hart, non-publié), ce qui limite évidemment le nombre de variables qui peuvent être prises en compte en particulier lorsque l'on veut tester ou inclure des effets quadratiques et des interactions. Maxent semble beaucoup moins souffrir de ce problème puisqu'il s'est révélé le plus efficace particulièrement lorsque le nombre de données était limité (Wisz et al., 2008).

Le modèle MaxEnt a été mis à contribution pour apprécier la réponse écologique des espèces étudiées vis-à-vis des variables climatique et écologique.

Pour l'espèce *Faidherbia albida* la température maximale du mois le plus chaud, la proportion d'argile et les précipitations du trimestre le plus chaud ont été identifiées comme variables prin-

ciales. Le modèle met en évidence que l'espèce peut supporter une gamme de température maximale du mois le plus chaud comprise entre 25 et 35°C, proportion d'argile comprise entre 0 et 10 en terre fine et des précipitations du trimestre le plus chaud supérieure à 140 mm. Pour *Balanites aegyptiaca*, le modèle met en évidence la précipitation du trimestre le plus chaud, la saisonnalité des précipitations et l'EvapoTranspiration et Interception réelles. *Balanites aegyptiaca* peut supporter une gamme de précipitations du trimestre le plus chaud comprise entre 110 et 128 mm, la saisonnalité des précipitations supérieur à 155 mm et une gamme constante quelque soit le niveau d'EvapoTranspiration et Interception réelles. Les résultats obtenus pour *Anogeissus leiocarpus* ont montré que l'espèce est plus sensible aux précipitations du mois le plus humide, la proportion de sable et le niveau du bassin versant. *Anogeissus leiocarpus* est abondante dans les biotopes où les précipitations du mois le plus humide est supérieur à 192 mm, la proportion de sable plus élevée et une abondance faible pour le niveau du bassin versant. Pour *Adansonia digitata* l'EvapoTranspiration et Interception réelles, la productivité de la végétation sur la saison des cultures et la productivité nette de l'eau de biomasse ont été identifiées comme variables principales. *Adansonia digitata* est abondante dans les biotopes où l'EvapoTranspiration et Interception réelles est élevée, lorsque la productivité de la végétation sur la saison des cultures est de plus en plus élevée mais diminue lorsque cette dernière est supérieure à 23000 terre fine et abondante lorsque la productivité nette de l'eau de biomasse est comprise entre 1100 et 1200 terre fine. *Acacia nilotica* est abondante dans les biotopes dont le pH du sol est inférieur à 7 c'est-à-dire dans les biotopes acides , le carbone organique compris entre 5 et 8 terre fine et la température minimale du mois le plus froid supérieure à 16°C. Selon les résultats du modèle MaxEnt, l'espèce *Cordyla pinnata* est rencontrée dans le parc ayant des caractéristiques écologique et topographique suivantes : l'EvapoTranspiration et Interception réelles inférieur à 6000, la quantité d'Azote inférieure à 450 terre fine et une pente de bassin versant élevé. La distribution potentielle des présences de l'espèce *Azadirachta indica* issue de MaxEnt montre que l'espèce est abondante dans les régions où la précipitation du trimestre le plus froid est faible, la quantité de phosphore comprise entre 200 et 250 terre fine, et la température moyenne du trimestre le plus sec supérieure à 28°C. Le modèle MaxEnt a indiqué que l'espèce *Ficus capensis* est retrouvée dans les biotopes où la température moyenne annuelle est supérieure à 27°C, la température moyenne du trimestre le plus chaud est comprise entre 29°C et 31°C, et lorsque la quantité d'Azote est supérieure à 450 terre fine. *Prosopis africana* est rencontrée dans le parc ayant des caractéristiques écologiques suivantes : le carbone organique comprise entre 5 et 8 terre fine, la quantité d'Azote supérieure à 400 terre fine, et la quantité de Phosphore inférieure à 150 terre fine.

Conclusion et Perspectives

La modélisation constitue un outil précieux au service de la préservation de la biodiversité de notre planète. La présente étude avait pour objectif de modéliser la distribution des principales espèces ligneuses dans deux parcs agroforestiers sénégalaïs (le parc à *Faidherbia albida* dans la région de Niakhar et le parc à *Cordyla pinnata* dans la région de Nioro du Rip). Il s’agissait de déterminer les facteurs explicatifs de la distribution spatiale des espèces étudiées, d’évaluer la capacité de prédiction des différents algorithmes utilisés dans les modèles de distribution des espèces ligneuses, d’élaborer des cartes de probabilité de présence des espèces et de mettre en évidence au moyen du meilleur modèle la réponse écologique de chacune des espèces étudiées. A cet effet, deux bases de données géoréférencées sur les espèces ligneuses, à savoir 9258 arbres à Niakhar et 6611 arbres à Nioro, ont été utilisées. Egalement, différents types de variables environnementales essentiellement dérivées de données et techniques géospatiales ont été utilisés.

Au moyen de l’analyse factorielle de la niche écologique (ENFA), il a été possible de sélectionner les variables explicatives de la distribution des espèces étudiées.

Pour la modélisation des niches écologiques, nous avons utilisé deux approches : la modélisation avec les blocs spatiaux et la modélisation sans les blocs spatiaux. Pour chaque approche de modélisation, sept modèles ont été testés.

Cette étude révèle que la modélisation avec les blocs spatiaux améliorent de manière générale les modèles de distribution des espèces. Le meilleur modèle parmi les sept testés est MaxEnt. Le modèle MaxEnt a été mis à contribution pour apprécier la réponse écologique des espèces étudiées vis-à-vis des variables climatiques et écologiques. Les résultats de notre étude montrent que les espèces ligneuses répondent différemment aux conditions environnementales et climatiques, ce qui permet de confirmer la deuxième hypothèse de notre étude. Nous avons aussi montré que l’ajout des variables à moyenne résolution spatiale n’améliorent pas toujours les modèles de distribution des espèces, ce qui permet également d’infirmer la première hypothèse de la présente étude.

Toutefois, notre recherche n’a porté que sur deux parcs, le parc à *Faidherbia albida* et le parc à *Cordyla pinnata*. Les résultats de cette étude pourraient être tout autre dans une zone d’étude avec des conditions climatiques et écologiques différentes.

Cette étude pourrait être complétée et poursuivi sous différents aspects. Il serait pertinent d’étendre cette étude au niveau national. Le développement d’un nouveau package sur R serait également

pertinent pour faciliter une étude comparative entre plusieurs espèces.

Bibliographie

- [1] Araújo M. B., Pearson R.G., Thuiller W. et Erhard M., 2005, *Validation of species-climate impact models under climate change*, Global Change Biol. 11 : 1504-1513.
- [2] Baddeley A., Turner R., Møller J. and Hazelton M., 2005, *Residual analysis for spatial point processes (with discussion)*, Journal of the Royal Statistical Society : Series B (Statistical Methodology), 67, 617– 666, en ligne disponible à l'adresse <https://doi.org/10.1111/j.1467-9868.2005.00519.x>
- [3] Bahn V., and McGill B. J., 2012, *Testing the predictive performance of distribution models*, Oikos, 122, 321– 331.
- [4] Barbault R., 2005, *International Conference : Biodiversity : science and governance* , Déclaration de Paris sur la biodiversité, Paris, 24-28.
- [5] Bio A. M., De Becker P., De Bie E., Huybrechts W. and Wassen M., 2002, *Prediction of plant species distribution in lowland river valleys in Belgium : Modelling species response to site conditions*, Biodiversity and Conservation, 11, 2189– 2216, en ligne disponible à l'adresse <https://doi.org/10.1023/A:1021346712677>
- [6] Boko M., Niang I., Nyong A., Vogel C., Githeko A., Medany M., Osman-Lasha B., Tabo R. et Yanda, 2007, *Climate change 2007 : impacts, adaptation and vulnerability*, In : Contribution of Working Group II to the Fourth Change (Eds M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. Van Der Linden and C.E. Hanson).
- [7] Bradshaw CJA, Sodhi NS, and Brook BW, 2009, *Tropical turmoil : a biodiversity tragedy in progress*, Front Ecol Environ 7 :79–87.
- [8] Corsi F., Dupre, E. et Boitani, L., 1999, *A large-scale model of wolf distribution in Italy for conservation planning*, Conservation Biology. 13, 150–159.
- [9] Dunn P. K. and Smyth G. K., 1996, *Randomized quantile residuals*, Journal of Computational and Graphical Statistics, 5, 236–244.
- [10] Elith J., Graham C.H., Anderson R.P., Dudik M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li, J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.Mc.C., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberon J., Williams S., Wisz M.S. et Zimmermann N.E., 2006, *Novel methods improve prediction of species distributions from occurrence data*, Ecography 29 :129–151.

- [11] **Elith J., Leathwick J. R.**, 2009, *Species distribution models : Ecological explanation and prediction across space and time*, Annual Review of Ecology, Evolution, and Systematics, 40, 677– 697, en ligne disponible à l'adresse <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- [12] **Franklin, J.** , 2010, *Mapping species distributions : Spatial inference and prediction*, Cambridge, UK : Cambridge University Press, en ligne disponible à l'adresse <https://doi.org/10.1017/CBO9780511810602>
- [13] **Garrity, D. P., Akinnifesi, F. K., Ajayi, O. C., Weldezemayat, S. G., Mowo, J. G., Kalinganire, A. Bayala, J.** , 2010, *Evergreen agriculture : A robust approach to sustainable food security in Africa*, Food Security, 2 :197–214.
- [14] **Gaston K.J.**, 1996, *Species richness : measure and measurement. Biodiversity : a biology of numbers and difference* (ed. K.J. Gaston), Blackwell Science, Oxford pp.77–113.
- [15] **Guillera-Arroita G., Lahoz-Monfort J. J., Elith J., Gordon A., Kujala H., Lentini P. E., and Wintle B. A.**, 2015, *Is my species distribution model fit for purpose ? Matching data and models to applications*, Global Ecology and Biogeography, 24, 276– 292, en ligne disponible à l'adresse <https://doi.org/10.1111/geb.12268>
- [16] **Guisan A. et Zimmermann N.E**, 2000, *Predictive habitat distribution models in ecology*, Ecology Modelling 135 : 147–186.
- [17] **Guisan A., Zimmermann N.E., Elith J., Graham C.H., Phillips S.J., et Peterson A.T.**, 2007, *What matters for predicting the occurrences of trees : techniques, data, or species characteristics ?*, Ecological Monographs 77 : 615-630.
- [18] **Guisan A., Thuiller W., Zimmermann N. E.** , 2017, *Habitat suitability and distribution models : With applications in R*, Cambridge, UK : Cambridge University Press, en ligne disponible à l'adresse <https://doi.org/10.1017/9781139028271>
- [19] **Hartigan J. A. and Wong M. A.**, 1979, *A k-means clustering algorithm*, Journal of the Royal Statistical Society Series C (Applied Statistics), 28, 100– 108.
- [20] **Hastie T., Tibshirani R. and Friedman J.**, 2009, *The elements of statistical learning : Data mining, inference, and prediction* (2nd ed., Vol. 1), New York, NY : Springer series in statistics New York, en ligne disponible à l'adresse <https://doi.org/10.1007/978-0-387-84858-7>
- [21] **Hilton-Taylor C.**, 2000, *IUCN red list of threatened species*, IUCN, Switzerland and Cambridge, UK, 61pp.
- [22] **Jane Elith and John R. Leathwick**, 2009, *Species Distribution Models : Ecological Explanation and Prediction Across Space and Time*, Annual Review of Ecology, Evolution, and Systematics 40 :677-697, en ligne disponible à l'adresse <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

- [23] **Le Rest K., Pinaud D., Monestiez P., Chadoeuf J., and Bretagnolle V.**, 2014, *Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation*, Global Ecology and Biogeography, 23, 811– 820, en ligne disponible à l'adresse <https://doi.org/10.1111/geb.12161>
- [24] **Maxted N., Ford-Lloyd B.V., Jury S., Kell SP et Scholten M.A.**, 2006, *Towards a definition of a crop wild relatives*, Biodiversity and Conservation, 2673–2685.
- [25] **Mbatudde M., Mwanjololo M., Kakudidi K. E. et Dalitz H.**, 2012, *Modelling the potential distribution of endangered Prunus africana (Hook.f.) Kalkm. in East Africa*, African Journal of Ecology 50 (4) : 393–403.
- [26] **McClean C.J., Lovett J.C., Kuper W., Hannah L., Sommer J.H., Barthlott W., Termansen M., Smith G.E., Tokamine S. et Taplin J.R.D.**, 2005, *African Plant Diversity and Climate Change*, Annals of the Missouri Botanical Garden 92 :139–152.
- [27] **Miguel B. Araújo and Antoine Guisan**, 2006, *Five (or so) challenges for species distribution modelling*, Journal of Biogeography, Pages 1677-1688, en ligne disponible à l'adresse <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- [28] **Miguel Ângelo Marini, Morgane Barbet-Massin, Jaime Martinez, Nêmora P. Prestes, Frédéric Jiguet**, 2010, *Applying ecological niche modelling to plan conservation actions for the Red-spectacled Amazon (Amazona pretrei)*, Biological Conservation, Pages 102-112, en ligne disponible à l'adresse <https://doi.org/10.1016/j.biocon.2009.09.009>
- [29] **Morgan W. Tingley, William B. Monahan, Steven R. Beissinger, and Craig Moritz**, 2009, *Birds track their Grinnellian niche through a century of climate change*, University of California, Berkeley, CA, 19637-19643, en ligne disponible à l'adresse <https://doi.org/10.1073/pnas.0901562106>
- [30] **Neuenschwander P., Sinsin B. et Goergen G.**, 2011, *Protection de la Nature en Afrique de l'Ouest : Une liste Rouge pour le Bénin*, Ibadan, Nigeria, 365p.
- [31] **O'Sullivan D. and Unwin D. J.**, 2010, *Geographic information analysis*, 2nd ed. Hoboken, NJ : John Wiley and Sons, en ligne disponible à l'adresse <https://doi.org/10.1002/9780470549094>
- [32] **Peterson A. T., Soberón J., Pearson R. G., Anderson R. P., Martínez-Meyer E., Nakamura, M., Araújo, M. B.**, 2011, *Ecological niches and geographic distributions*, (Vol. 56). Princeton, NJ : Princeton University Press.
- [33] **Pohjankukka J., Pahikkala T., Nevalainen P. and Heikkonen J.**, 2017, *Estimating the prediction performance of spatial models via spatial k-fold cross validation*, International Journal of Geographical Information Science, 31, 2001– 2019, en ligne disponible à l'adresse <https://doi.org/10.1080/13658816.2017.1346255>
- [34] **Radosavljevic A. and Anderson R. P.**, 2014, *Making better Maxent models of species distributions : Complexity, overfitting and evaluation*, Journal of Biogeography, 41, 629–643, en ligne disponible à l'adresse <https://doi.org/10.1111/jbi.12227>

- [35] **Renner I. W., Elith J., Baddeley A., Fithian W., Hastie T., Phillips S. J. and War-ton D. I.**, 2015, *Point process models for presence-only analysis*, Methods in Ecology and Evolution, 6, 366– 379, en ligne disponible à l'adresse <https://doi.org/10.1111/2041-210X.12352>
- [36] **Roberts D. R., Bahn V., Ciuti S., Boyce M. S., Elith J., Guillera-Arroita G., and War-ton D. I.**, 2017, *Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure*, Ecography, 40, 913– 929, en ligne disponible à l'adresse <https://doi.org/10.1111/ecog.02881>
- [37] **Salinger M.J., Sivakumar M.V.K. et Motha R.**, 2005, *Reducing vulnerability of agriculture and forestry to climate variability and change : workshop summary and recommendations*, Climate Change 70 : 341–362.
- [38] **Schröter, D., Polksy, C. and Patt, A.G.**, 2005, *Assessing vulnerabilities to the effects of global change : an eight step approach*, Mitig Adapt Strat Glob Change, 10 :573–595, en ligne disponible à l'adresse <https://doi.org/10.1007/s11027-005-6135-9>
- [39] **Steven J. Phillips, Robert P. Anderson, Robert E. Schapire**, 2006, *Maximum entropy modeling of species geographic distributions*, Ecological Modelling, pages 231-259, en ligne disponible à l'adresse <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- [40] **Telford R. J. and Birks H. J. B.**, 2009, *Evaluation of transfer functions in spatially structured environments*, Quaternary Science Reviews, 28, 1309– 1316, en ligne disponible à l'adresse <https://doi.org/10.1016/j.quascirev.2008.12.020>
- [41] **Trachsse M. and Telford R. J.**, 2016, *Estimating unbiased transfer-function performances in spatially structured environments*, Climate of the Past, 12, 1215– 1223, en ligne disponible à l'adresse <https://doi.org/10.5194/cp-12-1215-2016>
- [42] **Welk E., Schubert K. et Hoffmann M.H.**, 2002, *Present and potential distribution of invasive mustard (*Alliara petiolata*) in North America*, Divers. Distrib. 8, 219–233. 150. Wiser S., Peet R.K., White P.S. (1998) Prediction of rare-plant occurrence : a southern Appalachian example. Applied Ecology 8 :909–920.
- [43] **Wenger S. J., and Olden J. D.**, 2012, *Assessing transferability of ecological models : An underappreciated aspect of statistical validation*, Methods in Ecology and Evolution, 3, 260– 267, en ligne disponible à l'adresse <https://doi.org/10.1111/j.2041-210X.2011.00170.x>
- [44] **Wisz M.S. et Guisan A.**, 2009, *Do pseudo-absence selection strategies influence species distribution models and their predictions ? An information-theoretic approach based on simulated data*, BMC Ecology 9(8) : 1472-6785.

Modélisation de la distribution des espèces

1.0 SÉLECTION DES VARIABLES : ENFA

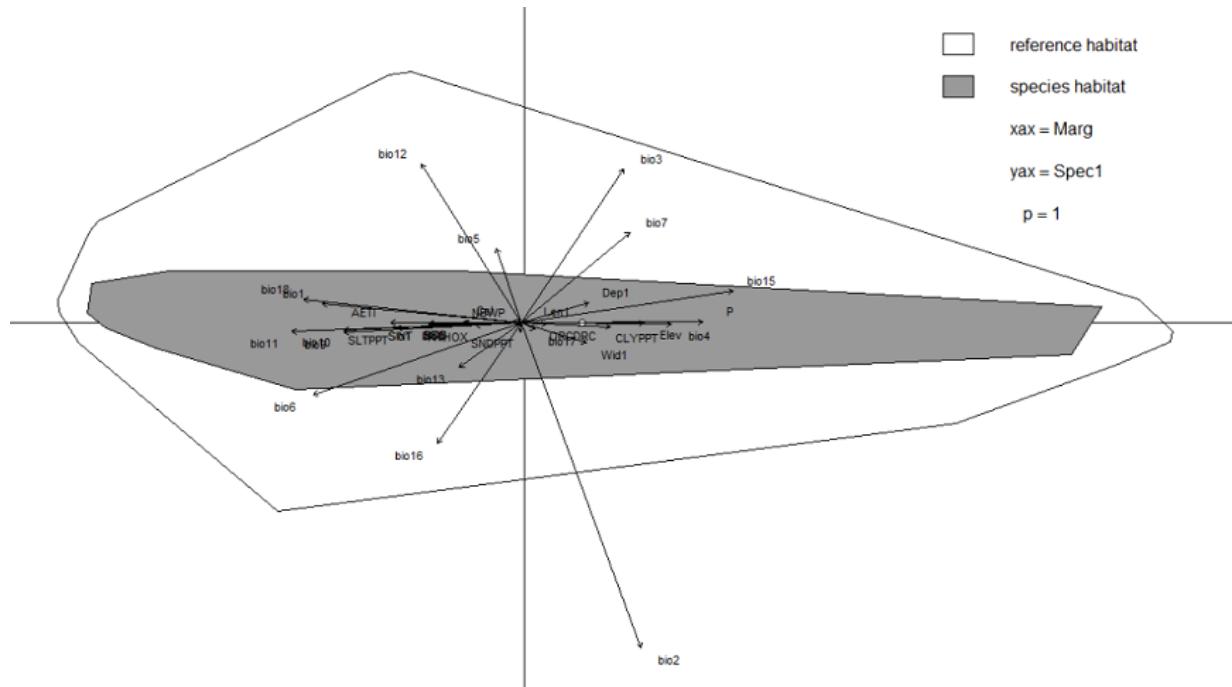


FIGURE A.1 – Analyse factorielle de la niche écologique (ENFA) de la distribution de *Balanites aegyptiaca*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Modélisation de la distribution des principales espèces ligneuses dans deux parcs agroforestiers du bassin arachider sénégalais.

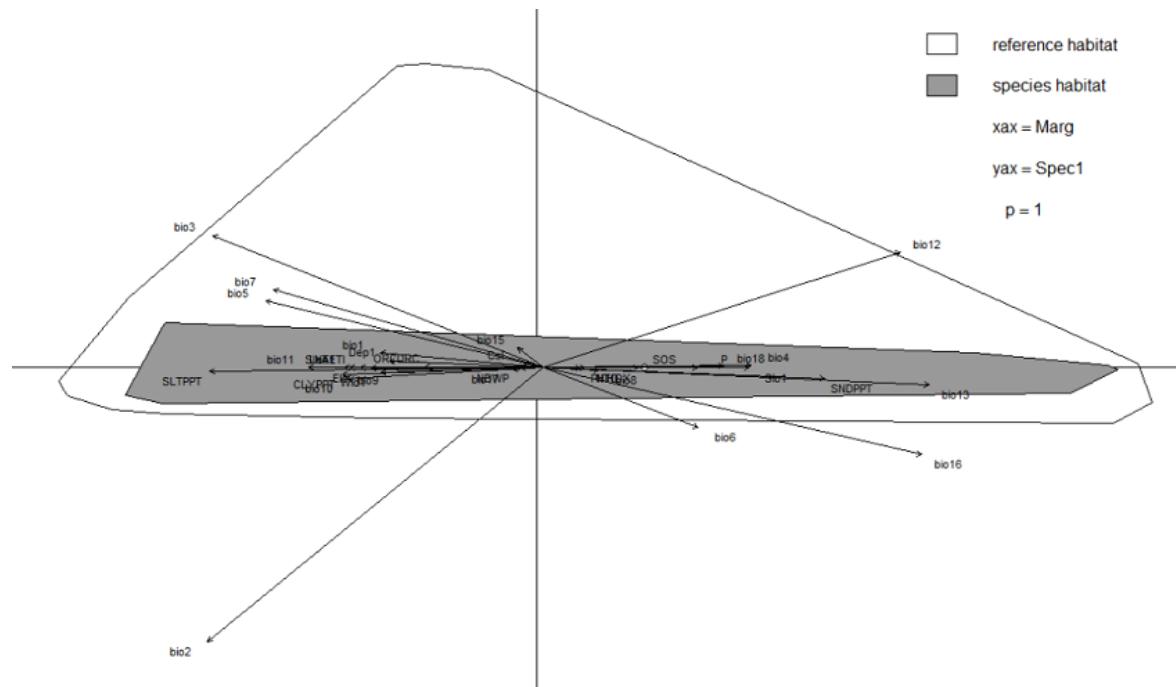


FIGURE A.2 – Analyse factorielle de la niche écologique (ENFA) de la distribution d' *Anogeissus leiocarpus*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

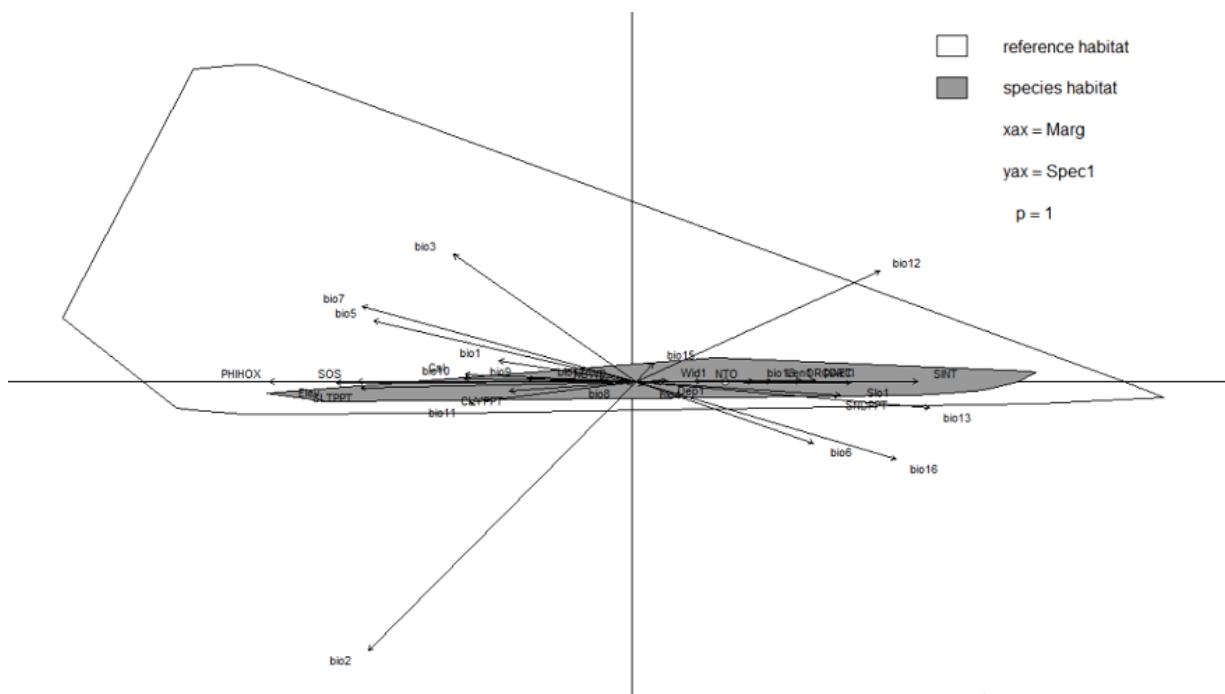


FIGURE A.3 – Analyse factorielle de la niche écologique (ENFA) de la distribution d' *Adansonia digitata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

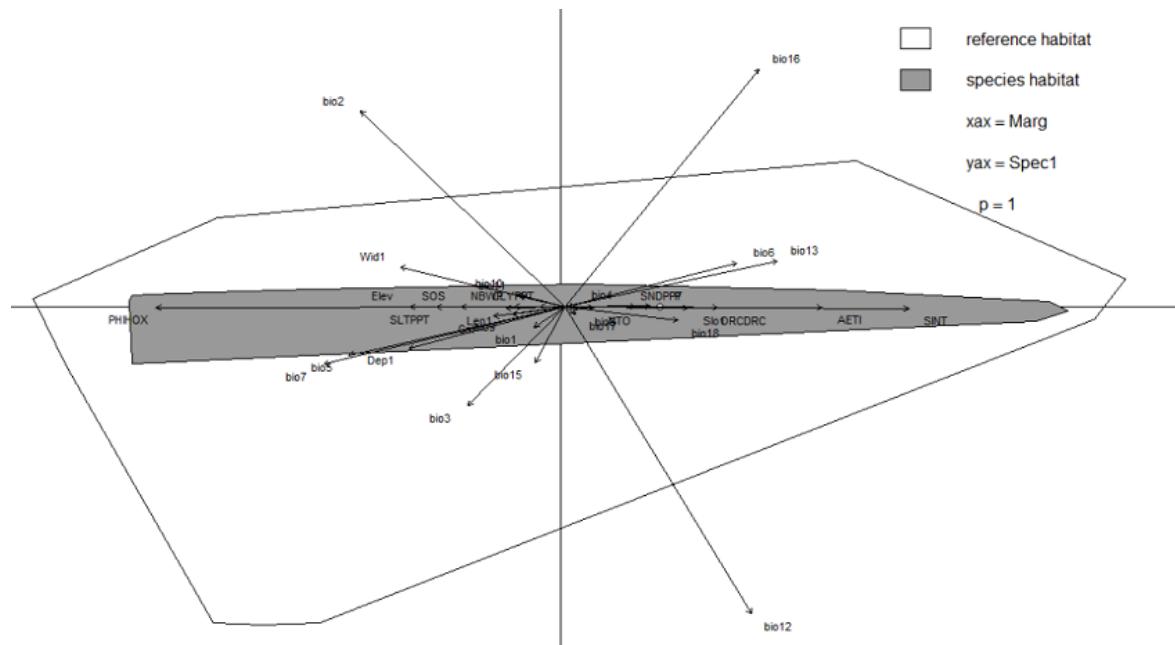


FIGURE A.4 – Analyse factorielle de la niche écologique (ENFA) de la distribution d' *Acacia nilotica*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

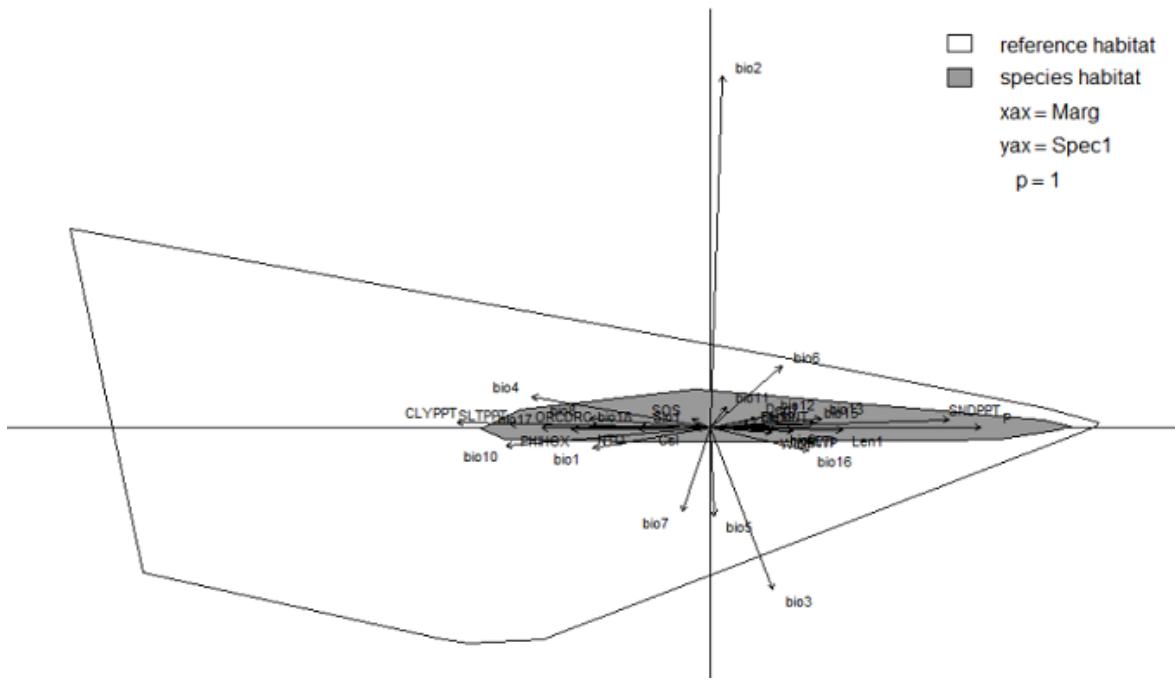


FIGURE A.5 – Analyse factorielle de la niche écologique (ENFA) de la distribution de *Azadirachta indica*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

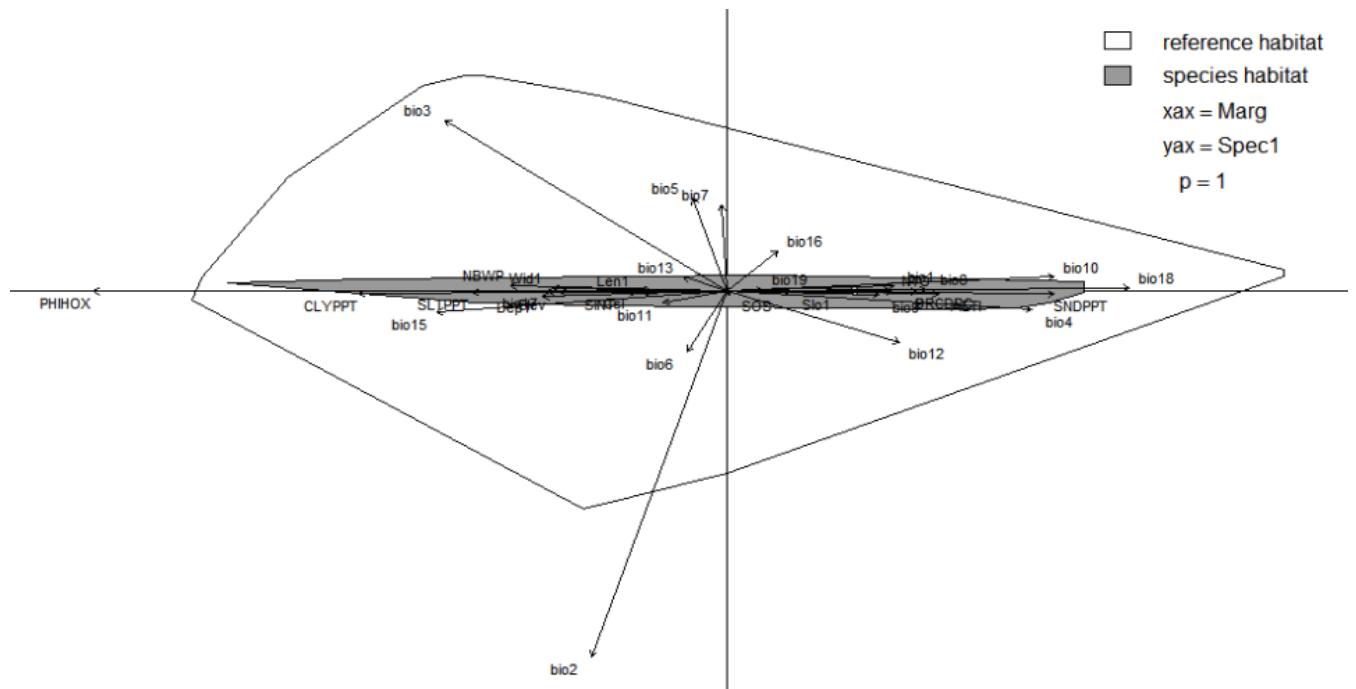


FIGURE A.6 – Analyse factorielle de la niche écologique (ENFA) de la distribution de *Prosopis africana*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

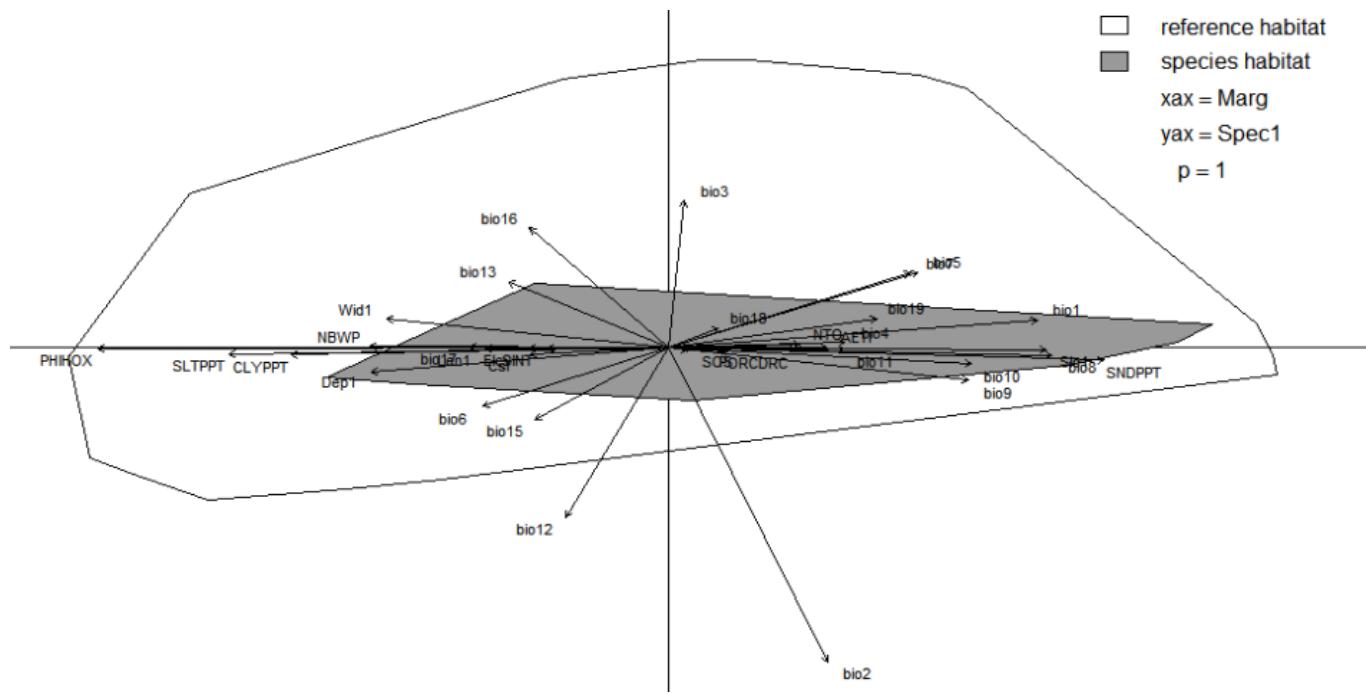


FIGURE A.7 – Analyse factorielle de la niche écologique (ENFA) de la distribution de *Ficus capensis*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

1.0 RÉSULTATS DES MODÈLES

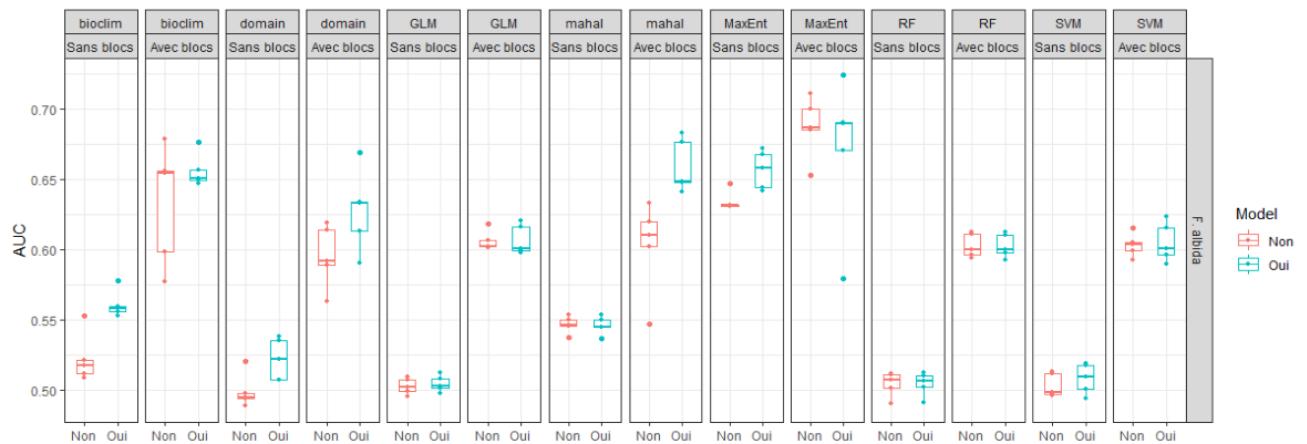


FIGURE A.8 – Comparaison des modèles pour *Faidherbia albida*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

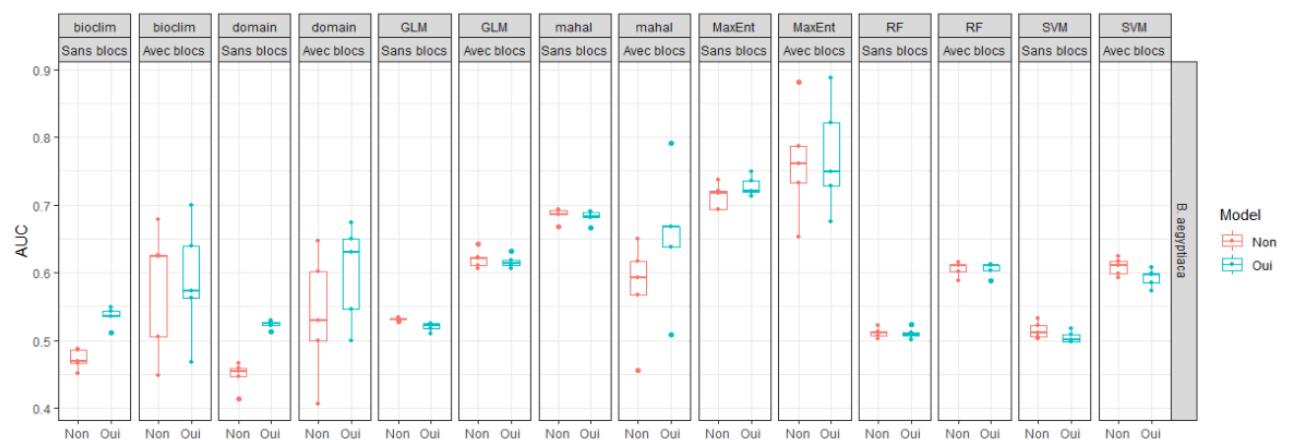


FIGURE A.9 – Comparaison des modèles pour *Balanites aegyptiaca*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Modélisation de la distribution des principales espèces ligneuses dans deux parcs agroforestiers du bassin arachider sénégalaïs.

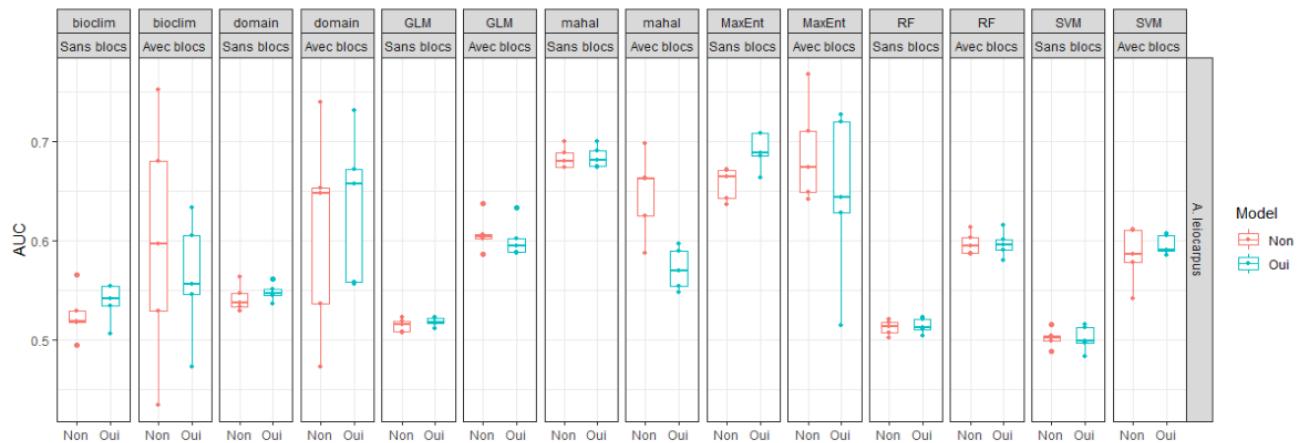


FIGURE A.10 – Comparaison des modèles pour *Anogeissus leiocarpus*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

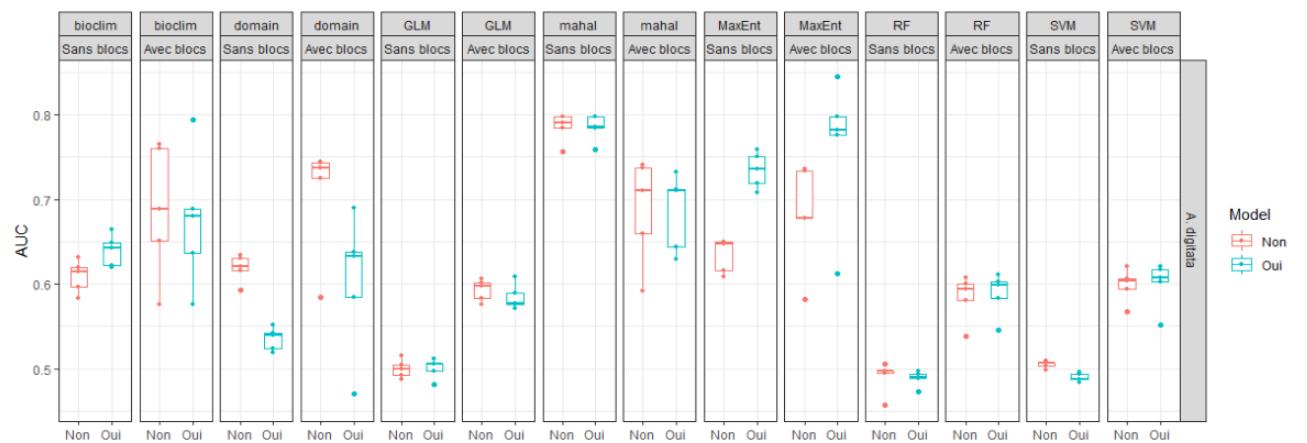


FIGURE A.11 – Comparaison des modèles pour *Adansonia digitata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

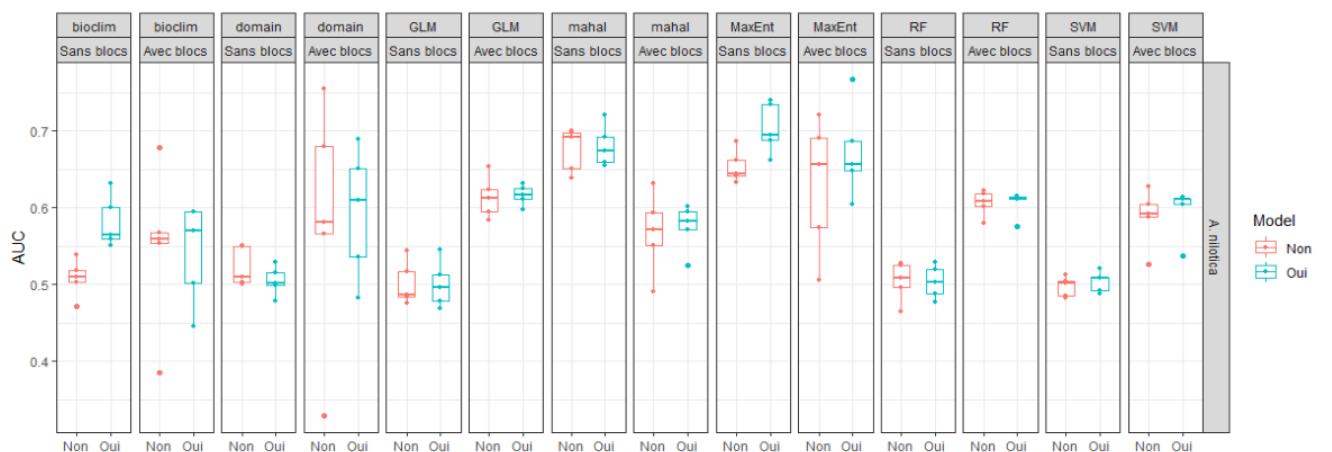


FIGURE A.12 – Comparaison des modèles pour *Acacia nilotica*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Modélisation de la distribution des principales espèces ligneuses dans deux parcs agroforestiers du bassin arachider sénégalaïs.

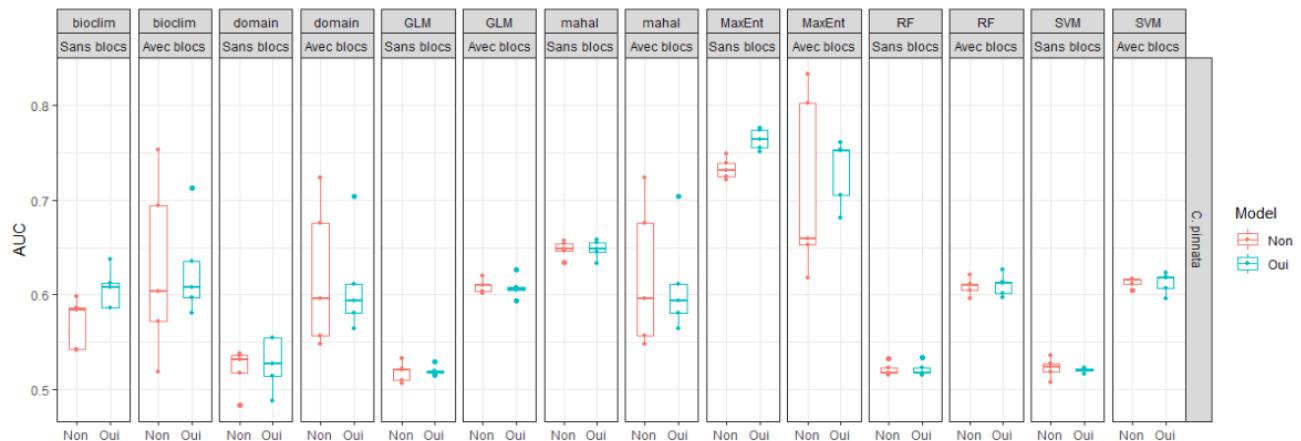


FIGURE A.13 – Comparaison des modèles pour *Cordyla pinnata*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

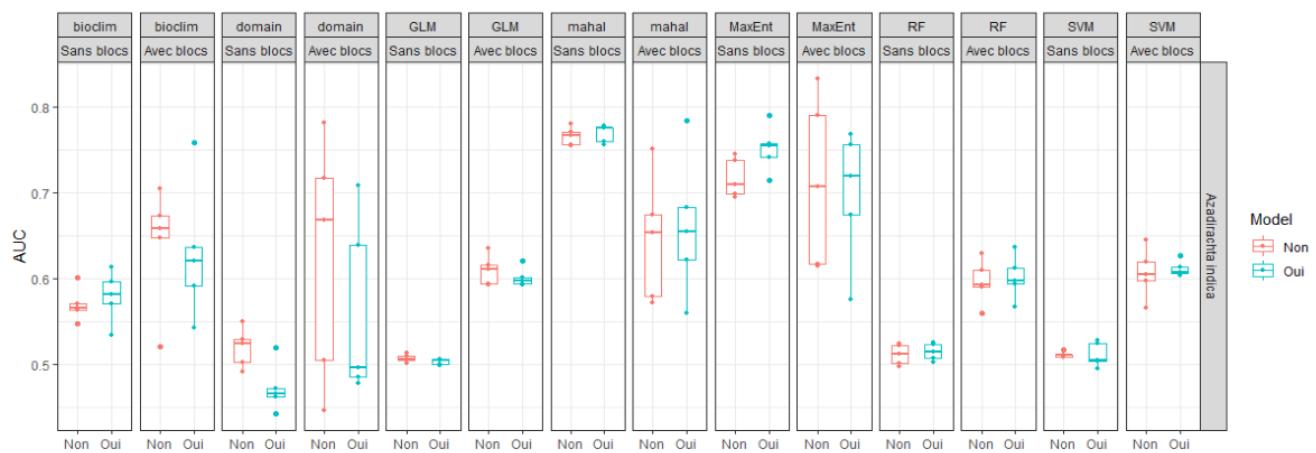


FIGURE A.14 – Comparaison des modèles pour *Azadirachta indica*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

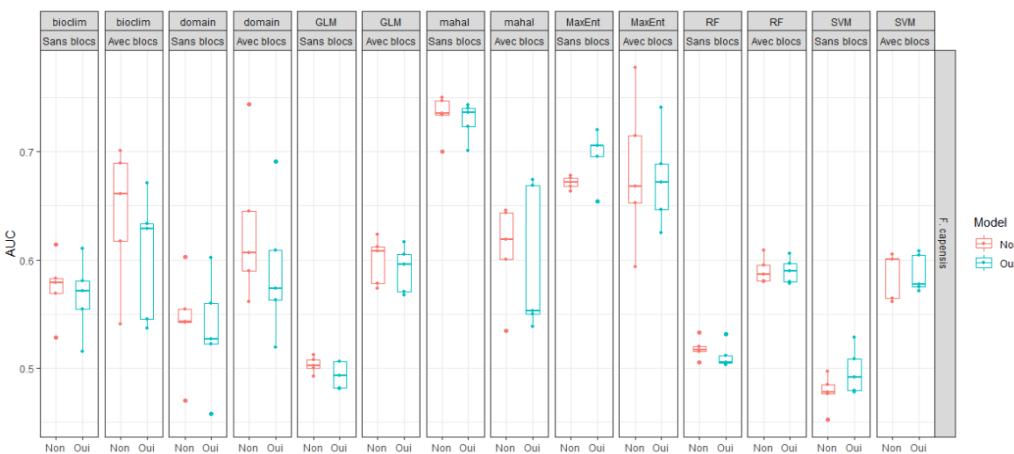


FIGURE A.15 – Comparaison des modèles pour *Ficus capensis*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

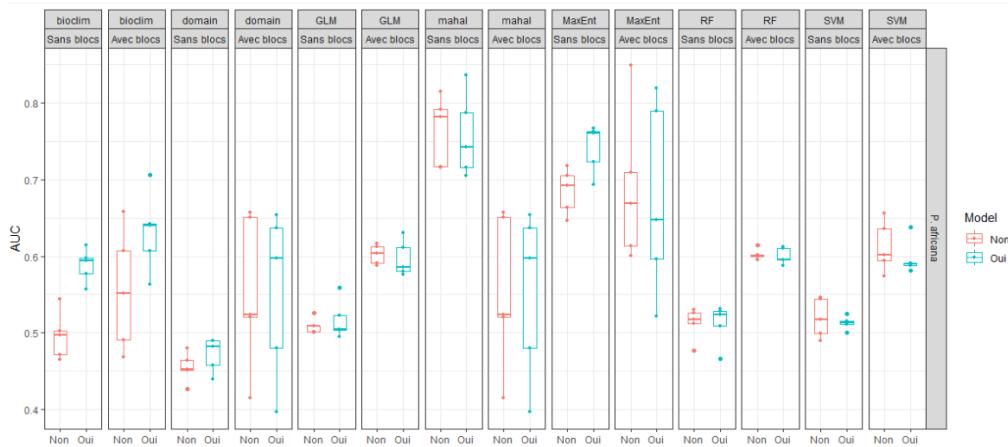


FIGURE A.16 – Comparaison des modèles pour *Prosopis africana*

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Analyse spatiale descriptive

2.0 RÉPRÉSENTATION GRAPHIQUE DES PRÉDICTEURS

2.0.0 Parc à *Faidherbia albida*

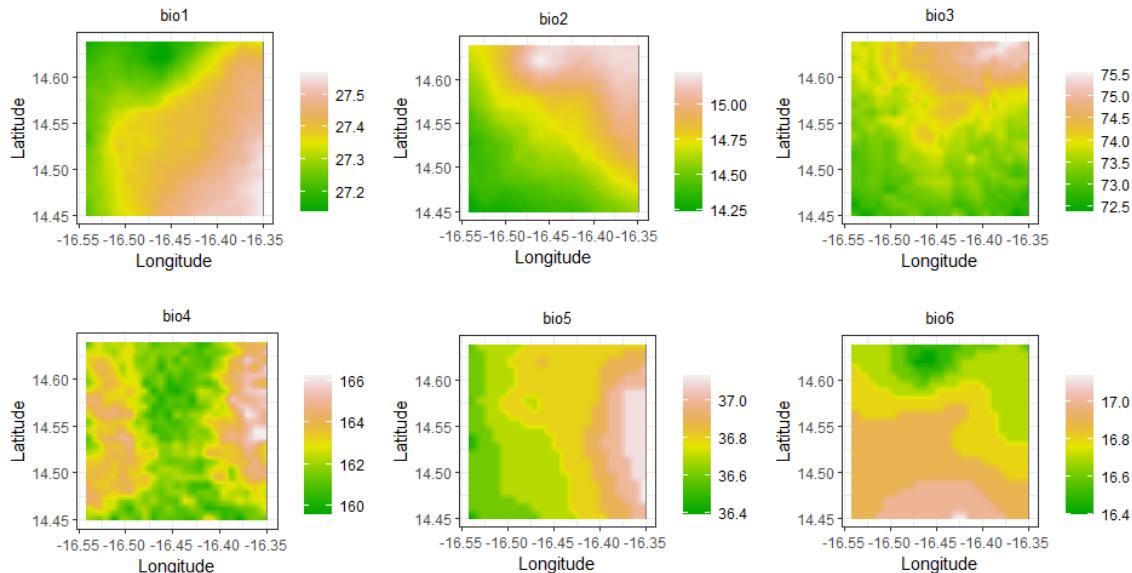


FIGURE B.1 – Réprésentation graphique des prédicteurs dans le parc à *Faidherbia albida* (bio1-bio6)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

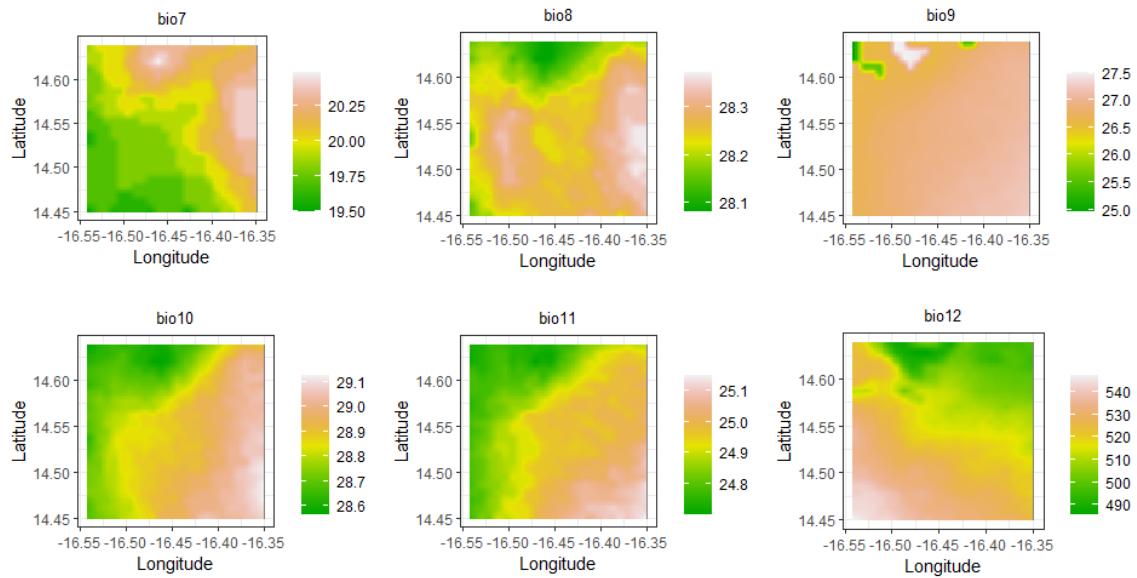


FIGURE B.2 – Réprésentation graphique des prédicteurs dans le parc à *Faidherbia albida* (bio7-bio12)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

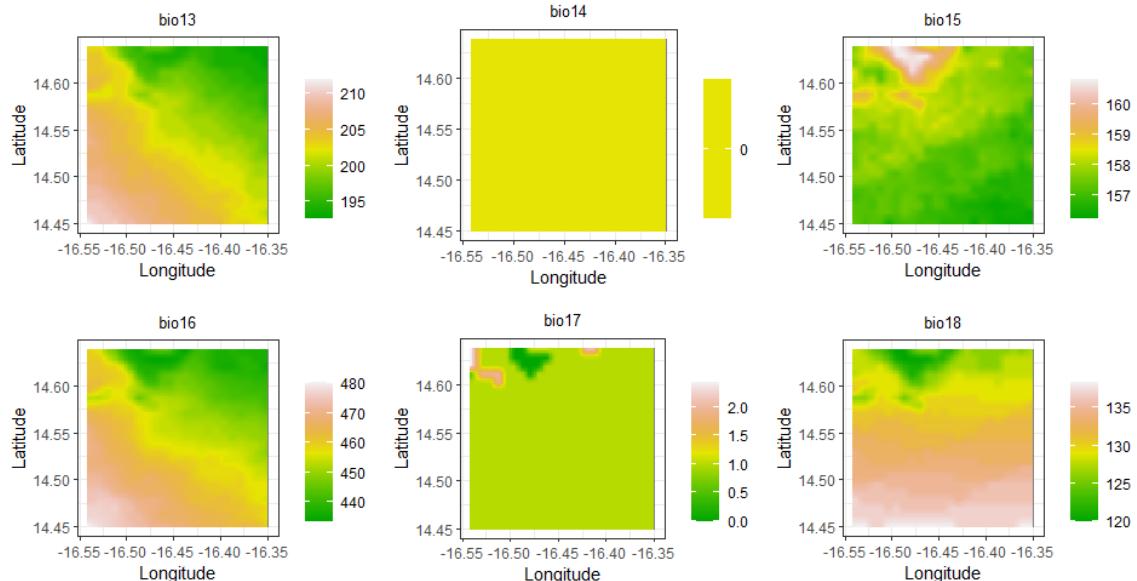


FIGURE B.3 – Réprésentation graphique des prédicteurs dans le parc à *Faidherbia albida* (bio13-bio18)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

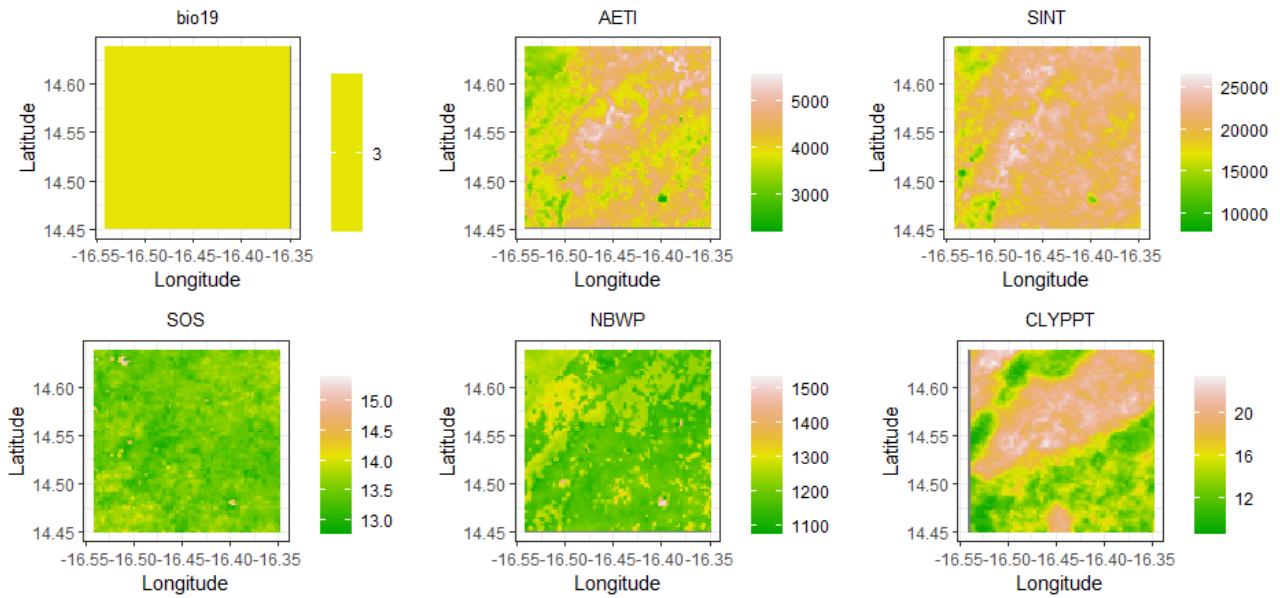


FIGURE B.4 – Réprésentation graphique des prédicteurs dans le parc à *Faidherbia albida* (bio19-CLYPPT)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

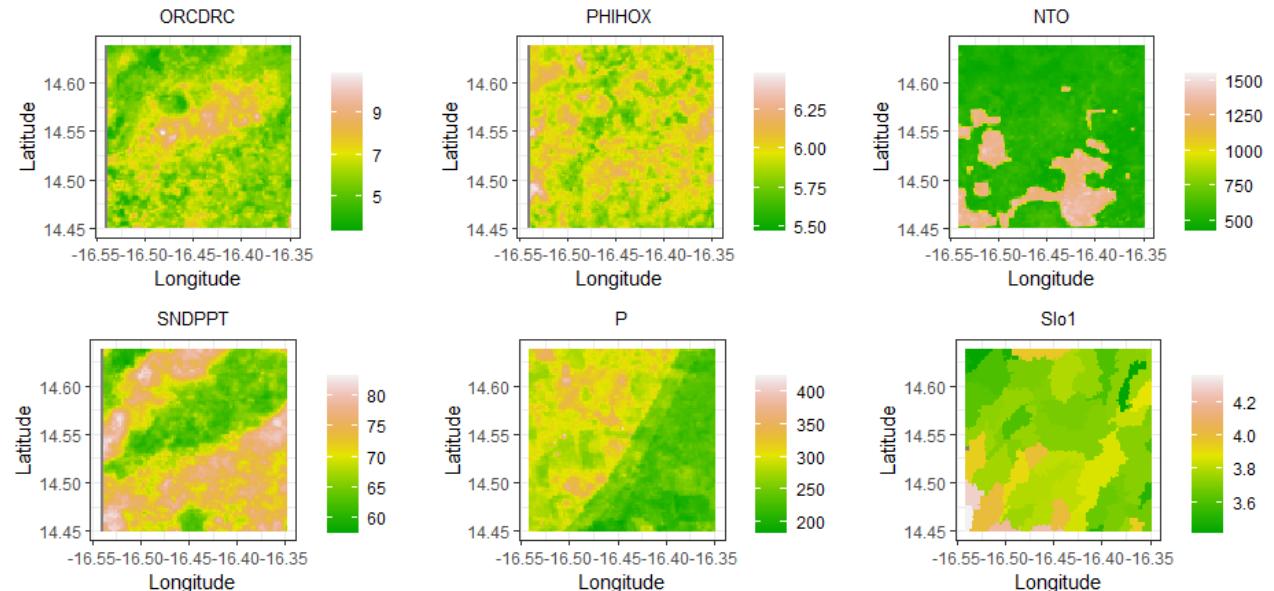


FIGURE B.5 – Réprésentation graphique des prédicteurs dans le parc à *Faidherbia albida* (ORCDRC-Slo1)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

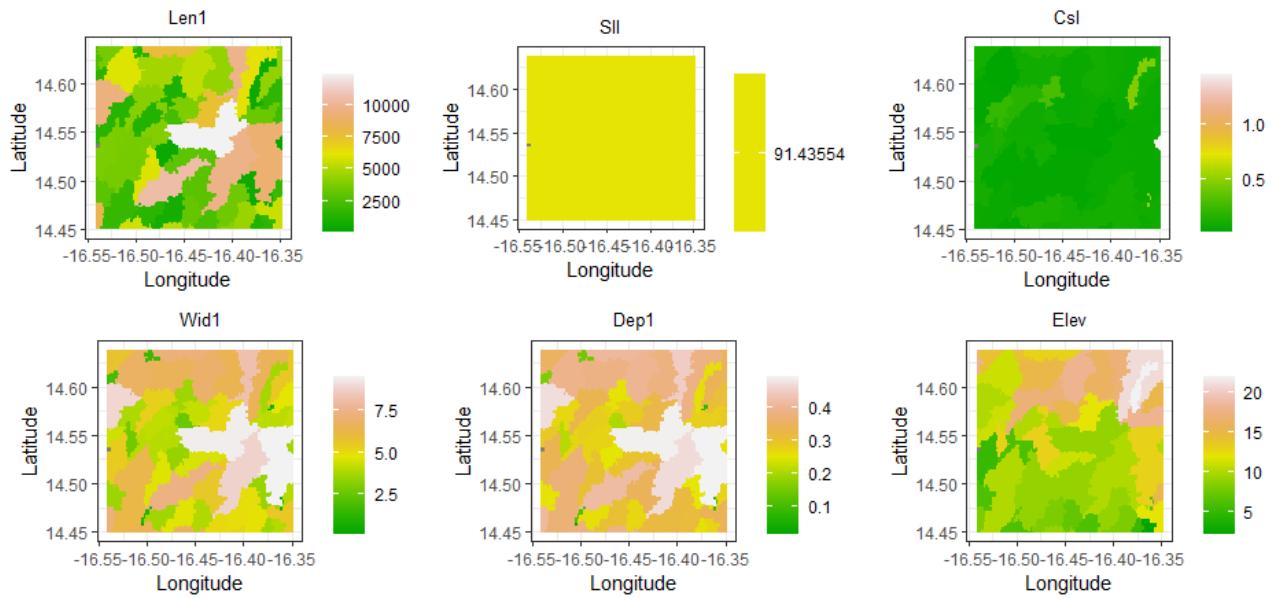


FIGURE B.6 – Réprésentation graphique des prédicteurs dans le parc à *Faidherbia albida* (Len1-Elev)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

2.0.0 Parc à *Cordyla pinnata*

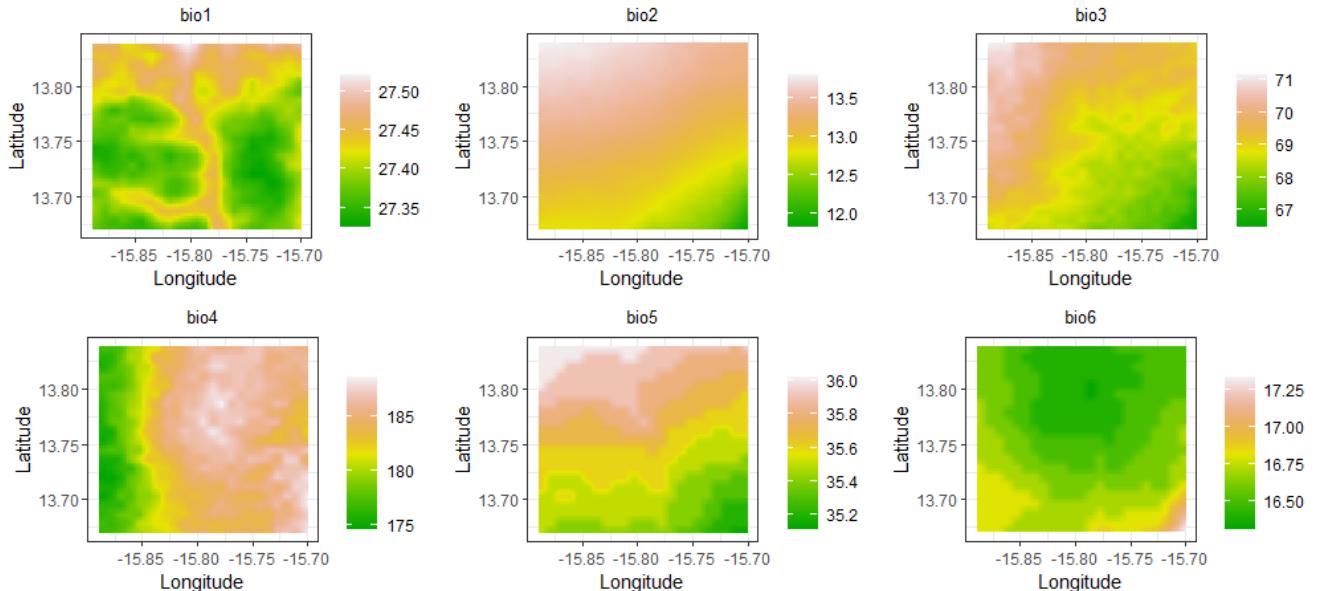


FIGURE B.7 – Réprésentation graphique des prédicteurs dans le parc à *Cordyla pinnata* (bio1-bio6)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

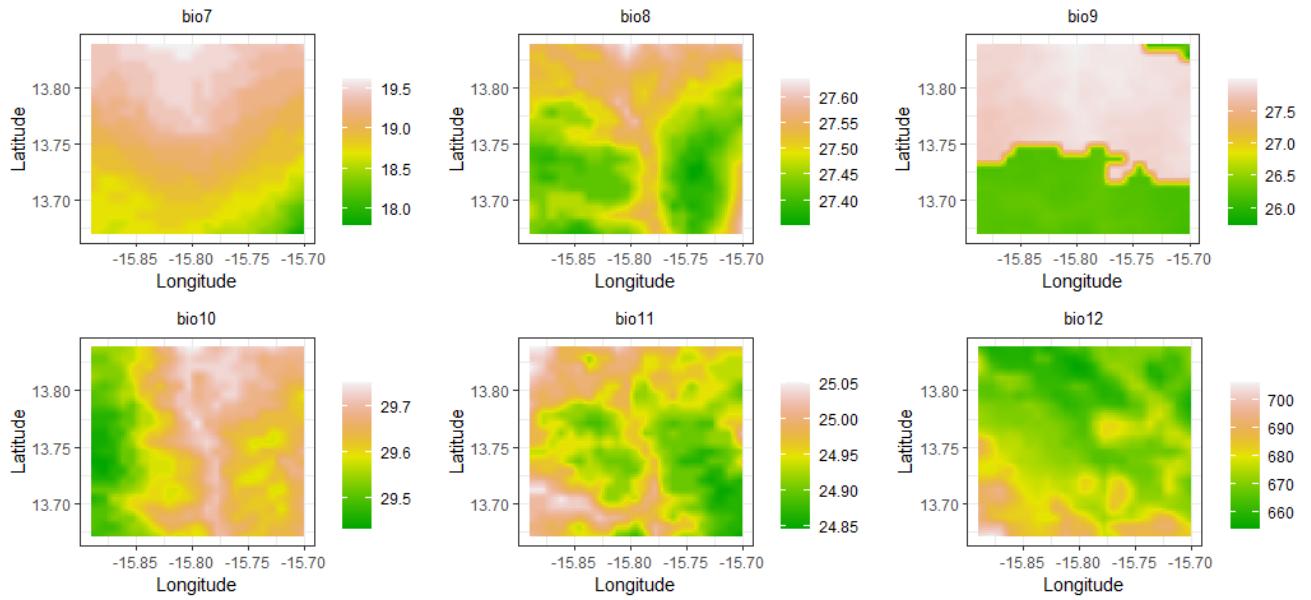


FIGURE B.8 – Réprésentation graphique des prédicteurs dans le parc à *Cordyla pinnata* (bio7-bio12)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

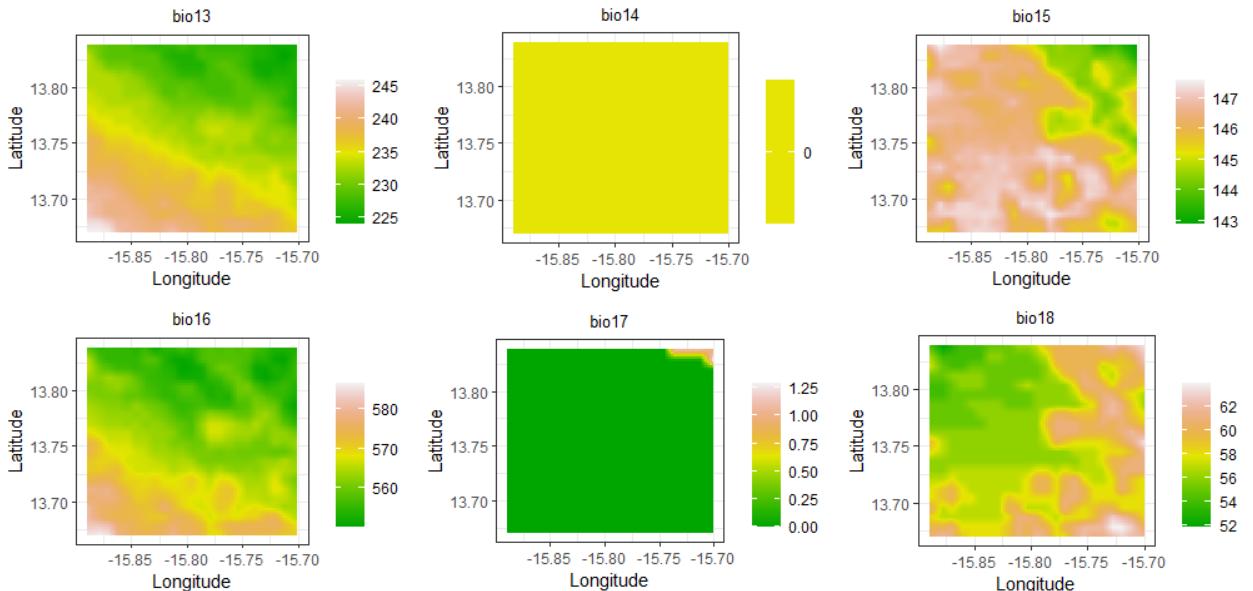


FIGURE B.9 – Réprésentation graphique des prédicteurs dans le parc à *Cordyla pinnata* (bio13-bio18)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

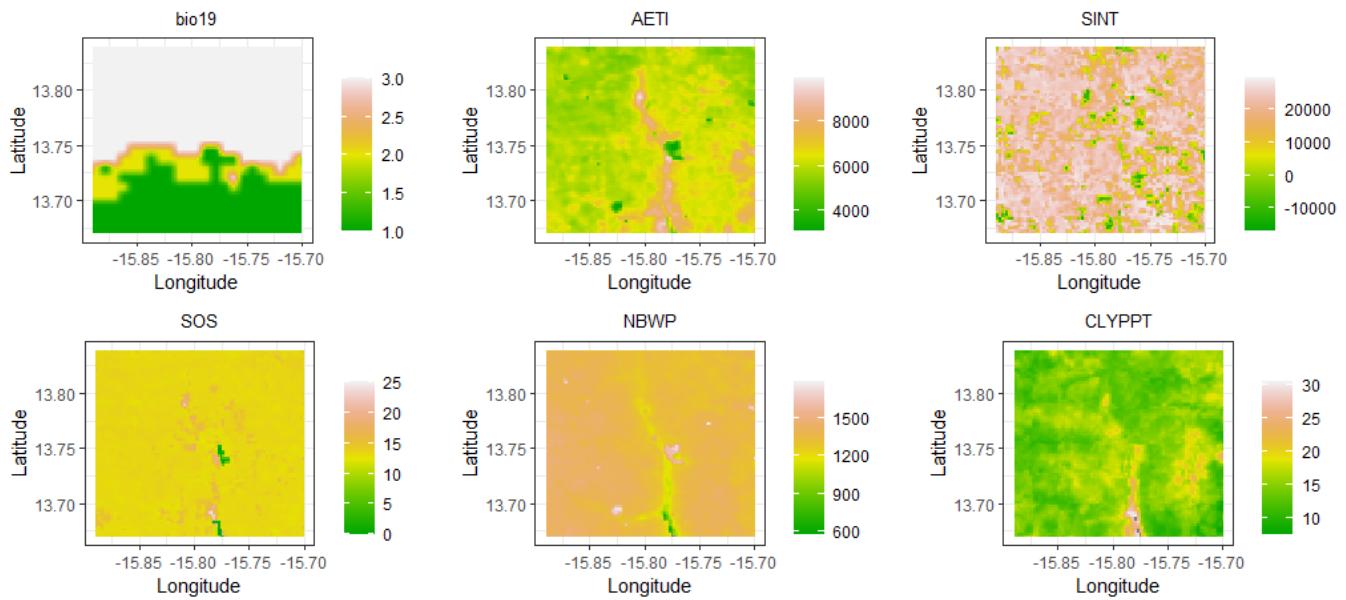


FIGURE B.10 – Réprésentation graphique des prédicteurs dans le parc à *Cordyla pinnata* (bio19-CLYPPTC)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

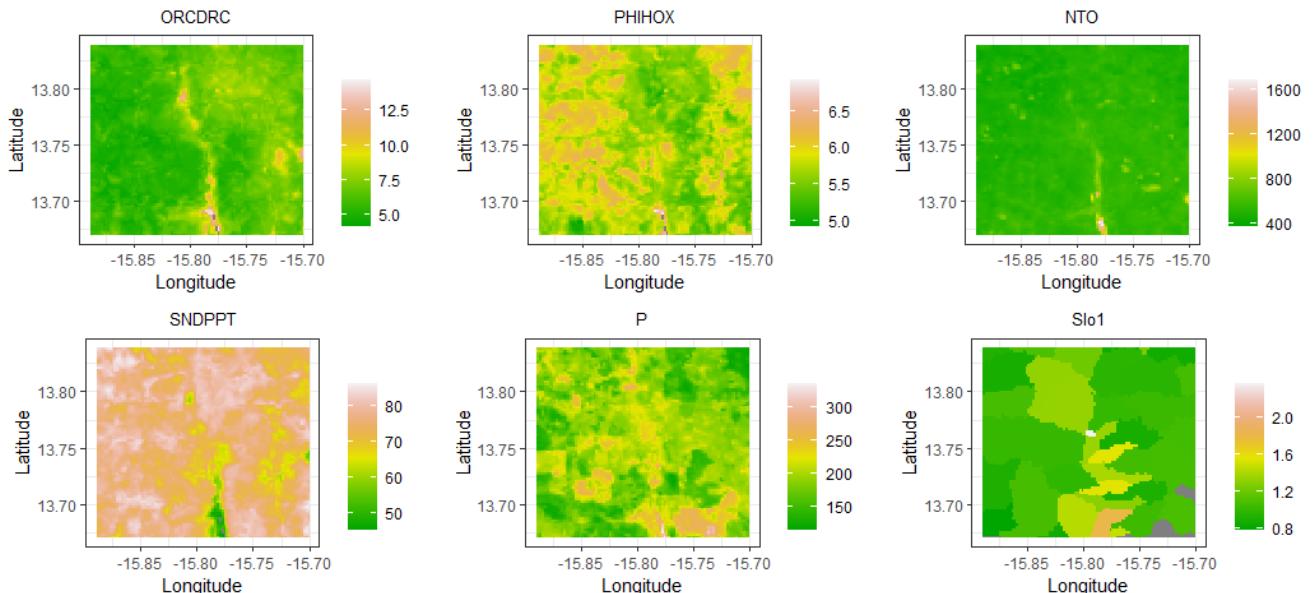


FIGURE B.11 – Réprésentation graphique des prédicteurs dans le parc à *Cordyla pinnata* (ORCDRC-Slo1)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

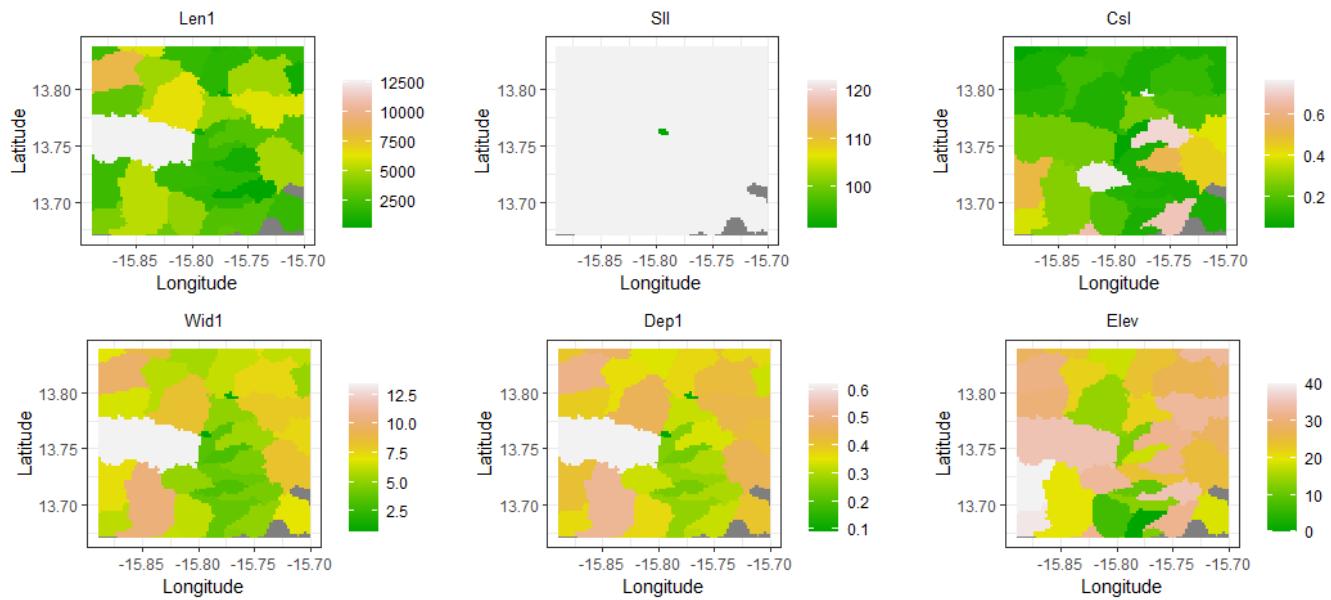


FIGURE B.12 – Réprésentation graphique des prédicteurs dans le parc à *Cordyla pinnata* (Len1-Elev)

Source : Projets SERENA et LYSA (2017-2019), Calcul de l'auteur

Table des matières

Remerciements	iv
Sigles et Abbréviations	x
Résumé	xi
Abstract	xii
Introduction Générale	1
1 Revue de la littérature	6
1.1 Modélisation de la distribution des espèces (SDM)	6
1.1.1 Concepts	6
1.1.2 Hypothèses	7
1.2 Méthodes de modélisation	7
1.2.1 Méthodes de profils	7
1.2.2 Méthodes de régression	7
1.2.3 Méthodes d'apprentissage automatique	7
1.3 Techniques de Modélisation	8
1.3.1 Les différentes techniques de blockages	9
1.3.2 Choix de la technique de blockage	11
1.3.3 Choix de la taille du bloc	12
1.4 Contexte biophysique du Sénégal	12
1.4.1 Les parcs agroforestiers du sénégal	13
2 Méthodologie	15
2.1 Zone d'étude	15
2.2 Données utilisées	16
2.2.1 Données de terrain	16
2.2.2 Données environnementales	16
2.3 Traitement des données et modélisation de la distribution des espèces	19
2.3.1 Prétraitement	19
2.3.2 Préparations des données	19
2.3.3 Choix des espèces étudiées	20

2.3.4	Autocorrélation spatiale	22
2.3.5	Sélection des variables par la méthode ENFA	24
2.3.6	Technique de modélisation	26
2.4	Validation des modèles	28
3	Résultats et Discussion	29
3.1	Résultats	29
3.1.1	Analyse descriptive	29
3.1.2	Analyses factorielles	34
3.1.3	Modélisation de la distribution des espèces	44
3.1.4	Réponse écologique des espèces étudiées	48
3.2	Discussion	54
Conclusion et Perspectives		57
Annexes		i
A Modélisation de la distribution des espèces		ii
B Analyse spatiale descriptive		x