



HAL
open science

Modélisation de la dynamique de population d'une plante native (palmier babaçu) dans le cadre d'un projet de gestion durable au Brésil

Nikolay Sirakov

► **To cite this version:**

Nikolay Sirakov. Modélisation de la dynamique de population d'une plante native (palmier babaçu) dans le cadre d'un projet de gestion durable au Brésil. Modélisation et simulation. Université Montpellier, 2016. Français. NNT : 2016MONTT302 . tel-01807940

HAL Id: tel-01807940

<https://theses.hal.science/tel-01807940>

Submitted on 5 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

Pour obtenir le grade de
Docteur

Délivré par l'Université de Montpellier

Préparée au sein de l'école doctorale **166 I2S**
Et de l'unité de recherche **MISTEA**

Spécialité: **Informatique**

Présentée par **Nikolay SIRAKOV**

**Modélisation de la
dynamique de population
d'une plante native (palmier
babaçu) dans le cadre d'un
projet de gestion durable au
Brésil.**

Soutenue le 13/12/2016 devant le jury composé de

Mme. Anne DOUCET	Professeur	UPMC	Rapporteur
M. Pascal MONESTIEZ	Dir. de Recherches	INRA	Rapporteur
Mme. Florence SÈDES	Professeur	UPS	Examinateur
M. Nicolas PICARD	Ingénieur en Chef	FAO	Examinateur
Mme. Thérèse LIBOUREL	Professeur émérite	UM	Directeur
M. Patrice LOISEL	Chargé de Recherches	INRA	Co-encadrant
Mme. Danielle MITJA	Chargé de Recherches	IRD	Co-encadrant
Mme. Bénédicte FONTEZ	Maître de Conférences	SupAgro	Co-encadrant



Remerciements

Je tiens tout d'abord à remercier vivement ma directrice de thèse Thérèse Libourel. Tu m'as soutenu depuis mon arrivée sur Montpellier et mes années de Master. Ton expérience et tes précieux conseils, critiques et remarques m'ont été indispensables tout au long de ce travail. Tu as toujours su trouver le juste mot, rebondir d'une situation difficile et me rassurer quant à l'avenir. Je ne pense pas dire faux en avançant que ton investissement personnel et ta générosité envers moi ont trouvé leur source dans un puits sans fond ! Je t'en suis extrêmement reconnaissant et t'exprime toute ma gratitude.

Je remercie très sincèrement mes trois co-encadrants Bénédicte Fontez, Patrice Loisel et Danielle Mitja. Le temps que vous avez consacré à répondre à mes questions ainsi que l'efficacité de l'encadrement que vous m'avez fourni sont inestimables.

Merci Bénédicte pour les heures passées à m'instruire à l'analyse bayésienne. Tu as su sans difficulté transmettre tes connaissances et répondre à mes questions.

Merci Patrice d'avoir partagé avec moi ton esprit critique et analytique. Ton incontestable capacité à discerner tous les problèmes a contribué au perfectionnement de ce travail.

Merci Dany d'avoir eu le courage de m'éclaircir et réexpliquer de maintes fois la biologie du babaçu. Ton savoir dans ce domaine est inestimable à mes yeux.

Mon grand respect va aussi à Anne Doucet et Pascal Monestiez qui, malgré le caractère profondément pluridisciplinaire de ce travail, n'ont pas hésité à accepter de le rapporter.

Je remercie également Florence Sèdes et Nicolas Picard qui ont accepté d'examiner le présent travail.

Cette thèse a fait l'objet d'un suivi annuel inscrit dans le dispositif d'un Comité de Suivi de Thèse composé en plus de ma directrice de thèse et de mes encadrants, d'Annie Chateau en qualité de représentant de l'École Doctorale I2S, d'Éric Delaître et d'Étienne Rivot en qualité de membres extérieurs. Je tiens donc à vous remercier d'avoir accepté de participer aux séances annuelles et d'avoir contribué à l'amélioration du présent travail grâce à vos avis critiques, vos orientations et vos propositions constructives.

La proposition que je fais en matière de modélisation de la dynamique de population du babaçu s'est appuyée sur des données de terrain qui m'ont été mises à disposition par Alessio dos Santos dans le cadre de la coopération franco-brésilienne

définie par le projet Open Science, je te remercie pour ta confiance.

Ce travail n'aurait pas été possible sans le support financier de la Fondation Agropolis via le projet Open Science au sein duquel cette thèse était inscrite. Les contributions financières de l'INRA et de SupAgro ont complété ce financement et m'ont ainsi permis de l'achever dans les délais impartis.

Ce manuscrit a fait l'objet de plusieurs relectures de la part de Bénédicte Fontez, Thérèse Libourel, Patrice Loisel et Danielle Mitja. Je vous suis très reconnaissant pour le temps que vous avez consacré à lire et relire les mêmes pages, les mêmes textes afin de les enrichir de vos corrections.

Un merci tout particulier aussi à Marie Anselmo pour le temps que tu as consacré à la relecture intégrale de cet écrit et la transformation de toute coquille en virgule et de toute virgule en coquille ! Tes conseils et tes remarques m'ont tant fait avancer dans les durs moments de manque d'inspiration et de concentration !

Je remercie aussi Anne Bisson, car j'ai bénéficié de sa part d'une aide d'urgence les jours précédant ma soutenance de thèse.

Je tiens à remercier aussi Pascal Neveu qui m'a accueilli au sein de l'UMR MISTEA. Merci pour ton implication dans mon « combat » face aux mécanismes administratifs bien souvent à caractère peu coopérants.

Merci à tous les membres de l'UMR MISTEA : dès mon arrivée, je me suis senti mis à l'aise, écouté et aidé.

Malika, merci pour ton support et aide face aux nombreux problèmes techniques que j'ai pu rencontrer durant les années passées au sein de l'unité. Véro et Maria, merci d'avoir toujours plaidé en ma faveur durant les longues procédures administratives auxquelles un doctorant doit faire face.

J'ai passé ma dernière année de thèse dans le bâtiment 21 dans une ambiance qui se prêtait bien à une fin de thèse!!! Anne, c'était vraiment super de partager le bureau avec toi ! Céline, je n'ai pas réussi à tuer tes plantes, mais je suis toujours prêt à réessayer. Bertrand bois un café, calme toi et fais nous un joli gâteau multi strates !

Благодаря на приятелите от София : Антон, Барона, Бобата, Василчо, Димо, Емо, Казака, Михата, Пацо, Пецата, Скикса, Телето, Черпака и всички останали : никога няма да забравя невероятните моменти прекарани заедно в България.

Depuis 2001, soit plus de quinze années passées en France j'ai fait de très belles rencontres à Bordeaux, Toulouse, Paris, Marseille, Lille, Lyon, Montpellier, Pompéi et leurs territoires adjacents. En écrivant ces lignes, j'ai bien évidemment, une pensée pour les agréables moments passés en compagnie d'Estelle et François, Sosso et Will, Laure et Eugène, Zou et Mathieu, La Rouge et Mike, Domi, Anne-Laure, M. Otto, Деница et François, Fanny, Bachel, Framy, M. Val, Foni, Наталия et Manu, Læti

et Sylvain, Ony et Jon, Bénito, Lucille, Noémie, Vaï, Vincent, Martin et Élo, Axoul et Cécile.

Je n'oublie pas les copains avec lesquels j'ai rendez-vous en 2024 au Cap Vert : Armando, Brenda, Camille, Chloé, Émi, Eugenio, Jesus, Jibi, Mathias, Nawelle, Ophé, Ponchis, Popo, Romain, Ross, Simon et Teresita. J'espère que vous y serez !

Une petite pensée pour les pompéiens Clarisse, Marie-K, Pascal-Michel, Marie-Jo, Maman Jackie, Olga et vos encouragements lors de la rédaction de ma thèse.

Merci aussi à Jean-Luc Guadelli pour tout son appui sans réserve durant mes années bordelaises et même avant. C'est certainement en partie grâce à toi que j'ai pu trouver ma place au sein de la République !

Ça a toujours été plaisir de partager un moment avec Philippe à Montpellier, Lyon, Marseille, София, Козарника . . .

L'archéologie n'est définitivement pas ma raison de vivre (j'espère que j'ai fini par convaincre ceux qui en doutent encore !), mais je suis né dedans (au sens figuré, mais aussi au sens propre du terme) et j'y prête toujours beaucoup d'attention et d'affection. J'ai toujours passé de très bons moments en la compagnie des archéos sur les sites de Темната дупка, Козарника, Le Piage, Les Rochers de Villeneuve, La Chapelle-aux-Saints, Sains-en-Gohelle, Pompeï . . . Je vous en suis reconnaissant !

J'ai également une pensée toute particulière pour Ценка, Росен et Jean-Gui et la longue amitié dont vous avez fait preuve. Je guette toujours avec impatience les prochains moments qu'on passera ensemble !!!

Je suis très reconnaissant envers toutes mes familles (valeurs ajoutées comprises bien évidemment) qui ont su m'écouter et répondre présent durant les moments difficiles de la vie. J'ai bénéficié de l'entière confiance des Anselmo, des Nossen, des Сиракови, de Росен et d'Isha, je vous en suis très reconnaissant.

Благодаря така също на моите български семейства Куманови, Конаклиеви, Чолакови и Доцеви, в които винаги съм се чувствал като у дома !

Je remercie également les Cagnin, Couteau, Langhart, Niarquin et Poletti : les fêtes de famille c'est toujours quelque chose avec vous !

Un très très grand merci à Danielle Anselmo pour son dévouement personnel à la préparation de la très grande majorité des délices pour le pot d'après-soutenance.

Un merci tout particulier à mes parents Боди и Кольо qui m'ont toujours encouragé et qui ont su éveiller la soif de connaissances en moi. Merci aussi à mes deux sœurs Деси и Дея, je suis toujours impatient de vous voir.

Marie, merci parce que tu as toujours été à mes cotés et que tu as toujours su

m'encourager, m'inspirer et me supporter. Tu as rendu ma fin de thèse splendide grâce à ton oreille « bien vaillante ». Je te dois notre être...

Sommaire

Remerciements	i
Sommaire	v
Liste des tableaux	vii
Table des figures	ix
Abréviations et sigles	xi
Préambule	xiii
I Introduction	1
1 Contexte général	3
2 Zone de l'étude	7
II Problématique	13
3 Développement durable en Amazonie	15
4 Autour de l'espèce babaçu	23
5 Projet Open Science, thèse et questionnements	33
III État de l'art	39
6 Données et connaissances	41
7 Dynamique de population	45
IV Proposition	53
8 Introduction	55
9 Acquisition et pérennisation des données	59

10 Modélisation mathématique	81
V Résultats, analyses et discussions	105
11 Données empiriques : premier regard analytique	107
12 Analyse de la dynamique de population	117
VI Conclusion et perspectives	133
13 Conclusion	135
14 Perspectives	137
Annexes	139
A Graphiques densités	141
B Scripts	143
C Convergence des chaînes de Markov	161
D Validation croisée - tableaux	167
Bibliographie	173
Index	183
Table des matières	185
Résumé	191
Abstract	193

Liste des tableaux

7.1	Définition des stades d'(Anderson, 1983, page 117)	50
7.2	Taux empiriques d'(Anderson, 1983, page 147)	50
9.1	Les stades biologiques du babaçu	61
9.2	Structure des parcelles	67
9.3	Données brutes par transect (pâturages 2013)	69
9.4	Densités par transect (pâturages 2013)	70
9.5	Données brutes par transect (pâturages 2014)	71
9.6	Données brutes par transect (pâturages 2015)	72
9.7	Données brutes par transect (babaçuals 2013)	73
9.8	Données brutes par transect (babaçuals 2014)	73
9.9	Données brutes par transect (babaçuals 2015)	73
9.10	Données brutes par transect (forêt tropicale humide 2013)	74
9.11	Données brutes par transect (forêt tropicale humide 2014)	74
9.12	Données brutes par transect (forêt tropicale humide 2015)	74
10.1	Notations se référant à la modélisation	82
11.1	Stade 3 entre les observations 2013 et 2014	111
11.2	Stade 3 entre les observations 2014 et 2015	112
11.3	Fruits par adulte de type 6a	113
11.4	Passage de fruits à nouveau palmier de stade 1	114
12.1	Analyse de sensibilité du modèle à l' <i>a priori</i> qui peut être non-informatif (ni) ou informatif (in) : les valeurs de l' <i>a posteriori</i> sont des valeurs moyennes de toutes les valeurs estimées par le modèle	118
12.2	Valeurs des <i>posterior predictive p-value</i> pour le modèle utilisant un <i>a</i> <i>priori</i> informatif	120
12.3	Nombre de transects utilisés pour l'estimation et pour la validation croisée	121
12.4	Validation croisée	122
A.1	Densités par transect (pâturages 2014)	141
A.2	Densités par transect (pâturages 2015)	142
D.1	Liste des transects utilisés pour le stade 1	167
D.2	Liste des transects utilisés pour le stade 2	168
D.3	Liste des transects utilisés pour le stade 3	169
D.4	Liste des transects utilisés pour le stade 4	170
D.5	Liste des transects utilisés pour le stade 5	171
D.6	Liste des transects utilisés pour le stade 6	172

Table des figures

2.1	<i>Projeto de Assentamento</i> (PA) Benfica au Brésil	8
2.2	PA Benfica au sein de son contexte territorial	9
2.3	PA Benfica : zonages	10
3.1	Déforestation du bassin amazonien	19
3.2	Le babaçu au sein des pâturages	20
4.1	Biologie du babaçu (formalisme <i>Unified Modelling Language</i> (UML), in Mitja <i>et al.</i> (sous presse))	26
4.2	Fruit du babaçu (adapté de Teixeira (2005))	27
4.3	Produits et Acteurs au sein du système babaçu (formalisme UML, adapté de Mitja <i>et al.</i> (sous presse))	30
4.4	<i>Quebradeiras</i> ramassant les noix du babaçu	31
4.5	<i>Quebradeiras</i> cassant les noix du babaçu	32
5.1	Interdisciplinarité et connaissance (formalisme UML)	34
5.2	Vision intégrée du système babaçu (formalisme UML, adapté de Si- mon <i>et al.</i> (2014))	35
5.3	Déroulement des activités du projet Open Science de l'Agropolis Fon- dation pour l'acquisition des données de terrain en écologie et la modé- lisation de la dynamique de population (la temporalité est représentée par « T0 » à « T3 »)	38
6.1	Cycle d'actions relatif au transfert et échanges de connaissances (d'après Grimshaw <i>et al.</i> (2012))	42
6.2	L'ontologie OBOE (formalisme UML)	43
7.1	La municipalité de Lago Verde au Brésil	51
8.1	Structuration des modèles (formalisme UML)	57
8.2	Les divers modèles utilisés (formalisme UML)	58
9.1	Stades biologiques du babaçu (clichés : Danielle Mitja)	60
9.2	Cycle de vie (diagramme d'états UML, in Mitja <i>et al.</i> (sous presse))	62
9.3	Modèle d'observation (formalisme UML)	64
9.4	PA Benfica - transects	65
9.5	Modèle conceptuel de données (formalisme UML)	76
9.6	Modèle relationnel (logique) de données	78
10.1	Cycle de vie (adapté de Caswell (2001))	83
10.2	Forme matricielle (adapté de Caswell (2001))	84

10.3	Niveaux hiérarchiques : les transects (premier niveau) au sein de parcelles (second niveau) ont une occupation du sol caractérisant des différences dans la dynamique de l'espèce (troisième niveau - forêt tropicale, pâturage ou babaçal).	87
10.4	Convergence des chaînes de Markov	90
10.5	Test de Gelman et Rubin (1992)	91
10.6	Cycle de vie du babaçu (adapté de Mitja <i>et al.</i> (sous presse))	94
10.7	La modélisation dans son ensemble (formalisme UML)	98
10.8	DAG (<i>Directed Acyclic Graph</i>) du modèle multinomial (basé sur le formalisme du logiciel BUGS - Lunn <i>et al.</i> (2000))	103
11.1	Données empiriques pour la mortalité	108
11.2	Données empiriques de passage de stade	109
11.3	Données empiriques pour rester dans le même stade	110
11.4	Densité-dépendance entre le nombre de fruits (n) et le nombre de nouveaux palmiers de stade 1 ($n + 1$)	115
12.1	Probabilités estimées pour le stade 1	123
12.2	Probabilités estimées pour les palmiers « Jeunes »	124
12.3	Probabilités estimées pour le stade 6	126
12.4	Estimation de la probabilité de donner des fruits	127
12.5	Estimation du nombre de fruits par adulte de type 6a	128
12.6	Estimation de la probabilité de nouveaux stade 1	129
C.1	Convergence des chaînes de Markov - stade 1	161
C.2	Convergence des chaînes de Markov - stade 2	161
C.3	Convergence des chaînes de Markov - stade 3	162
C.4	Convergence des chaînes de Markov - stade 4	162
C.5	Convergence des chaînes de Markov - stade 5	162
C.6	Convergence des chaînes de Markov - stade 6	163
C.7	Convergence des chaînes de Markov - recrutement	163
C.8	Test de Gelman-Rubin - stade 1	163
C.9	Test de Gelman-Rubin - stade 2	164
C.10	Test de Gelman-Rubin - stade 3	164
C.11	Test de Gelman-Rubin - stade 4	164
C.12	Test de Gelman-Rubin - stade 5	165
C.13	Test de Gelman-Rubin - stade 6	165
C.14	Test de Gelman-Rubin - recrutement	165

Abréviations et sigles

ASSEMA	<i>Associação em áreas de assentamento no estado do Maranhão</i> ..	32
BUGS	« <i>Bayesian inference Using Gibbs Sampling</i> »	93
Capes	<i>Coordenação de Aperfeiçoamento de Pessoal de Nível Superior</i>	37
CNPq	<i>Conselho Nacional de Desenvolvimento Científico e Tecnológico</i>	37
Embrapa	<i>Empresa Brasileira de Pesquisa Agropecuária</i>	
Espace-Dev	<i>Espace pour le Développement</i>	3
GiST	<i>Generalized Search Tree</i>	79
IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>	28
IRD	<i>Institut de Recherche pour le Développement</i>	3
INRA	<i>Institut National de Recherche en Agronomie</i>	4
ISO	<i>International Organization for Standardization</i>	
JAGS	« <i>Just Another Gibbs Sampler</i> »	93
LAMETA	<i>Laboratoire Montpellierain d'Économie Théorique et Appliquée</i> ..	4
LOV	<i>Linked Open Vocabularies</i>	137
MAMO	<i>Mathematical Modelling Ontology</i>	44
MCMC	<i>Markov Chain Monte Carlo</i>	90

MISTEA <i>Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie</i>	4
MIQCB <i>Movimento interestadual das quebradeiras de coco babaçu</i>	32
MOCAS <i>Modélisation, Optimisation, Commande pour les Agro-écoSystèmes</i>	6
OBOE <i>The Extensible Observation Ontology</i>	
OGC <i>Open Geospatial Consortium</i>	44
OntoMODEL <i>Ontological mathematical modeling knowledge management</i> .	44
PA <i>Projeto de Assentamento</i>	ix
PPP <i>posterior predictive p-value</i>	92
SQL <i>Structured Query Language</i>	78
UFRA <i>Universidade Federal Rural da Amazônia</i>	4
UML <i>Unified Modelling Language</i>	ix
UMR <i>Unité Mixte de Recherche</i>	4

Préambule

Nous souhaitons, en guise de préambule présenter la structure générale de ce mémoire.

L'Introduction (première partie) présente le contexte général dans lequel les recherches se sont déroulées, souligne le caractère pluridisciplinaire de celles-ci et précise les diverses caractéristiques de la zone d'étude concernée.

Par la suite, nos travaux sont présentés en plusieurs parties :

Problématique Cette partie situe la problématique globale autour du paradigme du développement durable en Amazonie. L'émergence d'espèces extractivistes comme le palmier babaçu au sein d'une large problématique de déforestation au Brésil aboutit à la présentation plus détaillée de celle-ci et des enjeux existant autour de cette espèce. Cela nous permet de revenir sur les questions majeures auxquelles les travaux vont s'intéresser et à dégager les verrous essentiels auxquels nous allons nous consacrer.

État de l'art Cette partie fournit une synthèse et une analyse des recherches existantes en rapport avec notre problématique : autour du partage des connaissances pluridisciplinaires et des données de terrain, autour de la dynamique de population des espèces végétales et autour de la gestion durable d'espèces endémiques évoluant dans un contexte anthropisé.

Proposition La proposition est conçue à partir d'une méthodologie qui relève d'une démarche itérative mixant induction et déduction adaptée au contexte sciences expérimentales. Nous détaillons la phase acquisition et pérennisation des données avant d'aborder la phase de modélisation mathématique.

Résultats, analyses et discussions Après avoir souligné l'apport de la pérennisation des données, nous examinerons plus en détail les résultats issus de la modélisation mathématique.

Conclusion et perspectives Nous concluons sur nos propos et développerons les perspectives qui s'ouvrent à la suite de ces travaux.

Première partie

Introduction

1	Contexte général	3
2	Zone de l'étude	7

CHAPITRE 1

Contexte général

LE TRAVAIL de cette thèse sera tout d’abord contextualisé dans ce chapitre. Nous précisons le fait que la collaboration pluridisciplinaire entre chercheurs puis entre différentes équipes a débuté en amont du projet dans lequel s’inscrit le présent travail. Cette collaboration a débouché sur la proposition d’un projet englobant le sujet de thèse.

Sommaire

1.1	Prémices : collaboration initiale dans le cadre d’un stage de master	3
1.1.1	Le stage de master	3
1.1.2	Partenariat recherche : rapprochement des équipes	4
1.2	Le projet Open science	4
1.3	La thèse	5

1.1 Prémices : collaboration initiale dans le cadre d’un stage de master

1.1.1 Le stage de master

Fin 2011, dans le cadre du Master 2 Géomatique (Université Montpellier 2 et 3) j’ai réalisé mon stage au sein de l’UMR Espace pour le Développement (*Espace-Dev*). Plusieurs sujets de recherche portant sur des problématiques et régions diverses et variées (Afrique, Asie et Amérique du Sud) étaient possibles.

Parmi eux, les problématiques autour de la modélisation de la dynamique de population du palmier babaçu ont tout particulièrement attiré mon attention. Sur les recommandations du Professeur Thérèse Libourel, j’ai choisi le sujet proposé par Danielle Mitja (CR - IRD) et Éric Delaître (IR - IRD) spécialisés dans les domaines respectivement de la biologie et du traitement d’images satellitaires. Le sujet proposé s’intitulait « Gestion durable du palmier babaçu (*Attalea speciosa* Mart. ex Spreng.) : modélisation mathématique et spatialisation de la disponibilité théorique d’une espèce native, Amazonie, Brésil ».

Danielle Mitja avait déjà fait partie de différents projets de recherche sur le babaçu comme par exemple le projet *Embrapa*/Institut de Recherche pour le Développement (IRD) (de 2001 à 2007) et le projet TOSCA (CIC-TOOB : Chaîne

d'intégration des capteurs de télédétection pour le suivi de l'occupation du sol et l'extraction automatique d'objet ressource : le cas du palmier babaçu en Amazonie, 2013-2014). Ces projets suivaient la politique générale de l'IRD en matière de collaboration internationale et étaient structurés autour d'un partenariat bi-parties réunissant des partenaires brésiliens.

Ma directrice de mémoire pour les quatre mois de stage a été Madame le professeur Thérèse Libourel, qui était à l'époque, co-responsable du Master Géomatique de l'Université Montpellier 2 au sein duquel j'étais inscrit.

C'est sous l'encadrement de ces trois personnes que j'ai initié cet exercice de recherche autour de la modélisation de la dynamique de population du palmier babaçu.

1.1.2 Partenariat recherche : rapprochement des équipes

Lors de la recherche effectuée dans ce stage, le rapprochement des équipes des Unité Mixte de Recherche (UMR) Espace-Dev et Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie (MISTEA) s'est naturellement imposé. En effet, nous nous sommes aperçus de la nécessité d'approfondir nos connaissances en matière de modélisation de dynamique de populations végétales. Nous nous sommes donc tournés vers Pascal Neveu de l'UMR MISTEA (Institut National de Recherche en Agronomie (INRA) - SupAgro)¹ afin de créer un premier partenariat et avons immédiatement entamé une étroite collaboration avec Patrice Loisel (INRA). Celui-ci est devenu très rapidement par la suite mon quatrième encadrant de stage.

Ce partenariat réussi a très vite débouché sur la volonté de poursuivre en déposant un projet de recherche plus vaste auprès du programme Open Science d'*Agropolis Foundation*.

1.2 Le projet Open science

Le projet de recherche déposé, s'intitulait « Méthodes et outils de décision pour la gestion durable des palmiers babaçu dans les pâturages ». Son périmètre englobait des chercheurs montpelliérains, outre ceux précédemment cités, relevant de sciences humaines et sociales (économie théorique et appliquée UMR Laboratoire Montpellierain d'Économie Théorique et Appliquée (LAMETA)) et un partenariat international (Izildinha Miranda, *Universidade Federal Rural da Amazônia (UFRA)* - Brésil).

Nous détaillons un peu ce projet afin de dégager les éléments essentiels de la problématique et des questions scientifiques abordées.

1. Le contexte du labex Numev avait permis des contacts pluridisciplinaires

Au Brésil, l'Amazonie légale comporte des espaces encore non défrichés et de grandes étendues gagnées sur la forêt dans lesquelles pâturages plantés et cultures coexistent. Parmi les espèces de ligneux ou de palmiers originaires des formations végétales initiales, qui peuvent être épargnées ou non au moment du défrichement, le palmier *Attalea speciosa* Mart. ex Spreng. ou babaçu, est une espèce utile qui fait l'objet d'une activité économique importante à l'échelle nationale. Elle a sa place dans des agrosystèmes spécifiques (dits « agrosystèmes à babaçu ») de zones rurales caractérisées par la petite agriculture familiale diversifiée (culture vivrière en début de succession culturale puis pâturages semés) et les grandes fermes orientées vers la production de bétail. Le projet porte sur un territoire sur lequel plusieurs points de vue et usages différents cohabitent autour de la gestion durable de cette espèce native extractiviste : le babaçu.

Cependant des connaissances sur la dynamique de population du palmier font défaut pour organiser une gestion durable de l'espèce dont la reproductibilité peut être compromise si elle est soumise à des prélèvements exagérés.

La dimension socio-économique du projet vise à appréhender les usages, les représentations des acteurs et la dynamique de la filière. Les travaux sur l'évolution des pratiques vis-à-vis de l'environnement et de la biodiversité témoignent de l'importance des représentations des agriculteurs vis-à-vis de l'environnement, notamment en identifiant des logiques altruistes, territoriales ou professionnelles qui favorisent le développement d'innovation dans les pratiques en faveur de la protection de l'environnement (Lémery, 2003; Michel-Guillou, 2006; Weiss *et al.*, 2006). L'identification des pratiques des acteurs et de la dynamique de filière s'inscrit dans le cadre des approches traditionnelles d'étude filière (Caron, 1998) et d'ingénierie territoriale (Lardon *et al.*, 2009; Rey-Valette, 2010).

Le projet comportait un sujet de doctorat (travail présenté dans ce mémoire)².

1.3 La thèse

Le sujet proposé se positionne au croisement de la plupart des domaines scientifiques représentés au sein de ce projet : informatique, mathématique, statistique, biologie, économie, géomatique. Intitulée « Modélisation de la dynamique de population d'une plante native (palmier babaçu) dans le cadre d'un projet de gestion durable au Brésil », la thèse comporte un volet modélisation mathématique important. Ce constat, ainsi que certaines restrictions administratives liées à la gestion du projet Open Science et auxquelles nous avons dû faire face ont eu pour incidence que la thèse a été réalisée au sein de l'UMR MISTEA, entre février 2013 et décembre 2016³. L'équipe d'encadrement est constituée de Patrice Loisel (responsable de l'équipe

2. Le partenariat brésilien proposait un autre sujet de doctorat plus centré biologie-écologie de l'espèce.

3. Cette UMR regroupe des enseignant-chercheurs de SupAgro Montpellier et des chercheurs INRA Montpellier

Modélisation, Optimisation, Commande pour les Agro-écoSystèmes (MOCAS)) et Bénédicte Fontez (Maître de conférences en statistiques à SupAgro), membres de MISTEA et de Danielle Mitja et Thérèse Libourel membres de l'UMR Espace-Dev.

De plus, afin d'assurer le bon avancement du présent travail, nous avons organisé de multiples réunions de travail et d'échanges entre les différents membres du projet Open Science (UMR MISTEA, UMR Espace-Dev, UMR LAMETA, UFRA).

CHAPITRE 2

Zone de l'étude

CE CHAPITRE se centre autour de la zone de l'étude. Les aspects géographiques, historiques et naturels seront examinés. Une section est également consacrée à la contextualisation nationale du territoire sur lequel les travaux de terrain ont été menés.

Sommaire

2.1	Territoire de Benfica	7
2.2	Anthropisation et colonisation	8
2.3	Cadre naturel	11
2.3.1	Hydrologie	11
2.3.2	Pédologie	11
2.3.3	Climat	11
2.3.4	Couvert végétal	12

2.1 Territoire de Benfica

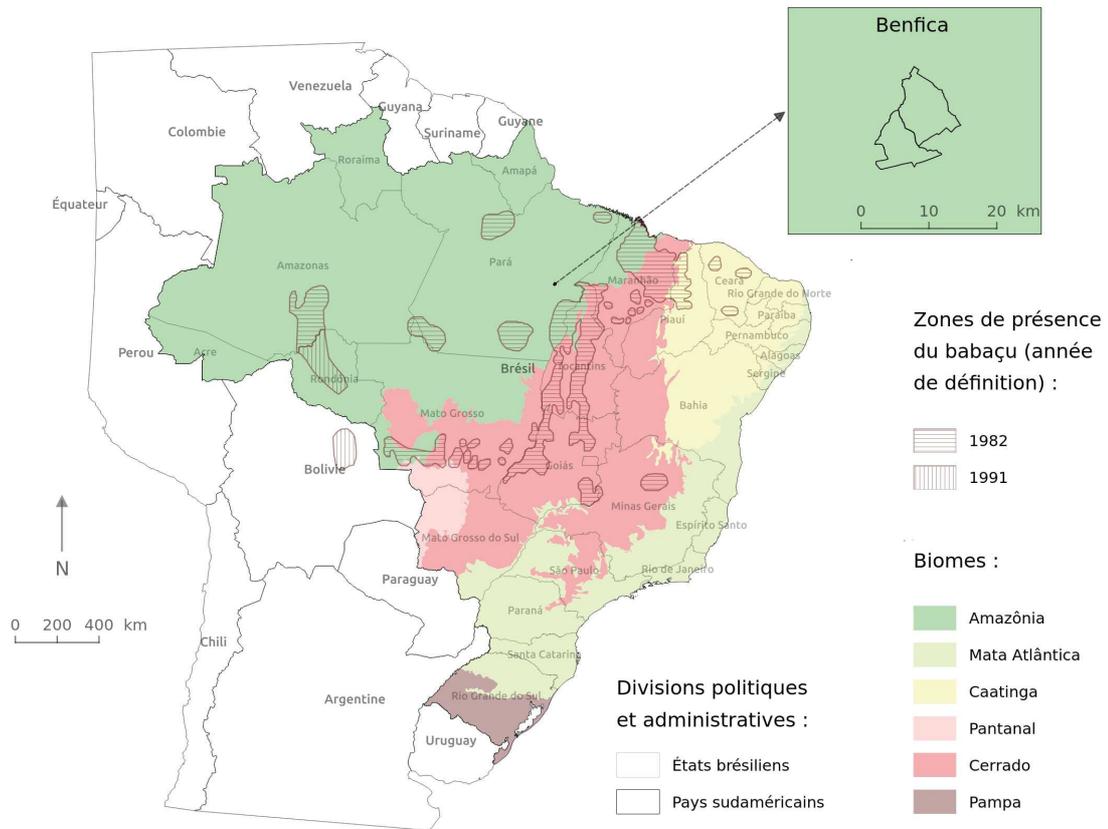
L'étude de terrain s'est effectuée au sein du territoire du *Projeto de Assentamento* (PA) de Benfica. Un PA désigne au Brésil un projet d'installation d'agriculteurs. Un tel projet se développe au sein d'une municipalité donnée - à grande échelle. Le PA Benfica se situe dans l'État brésilien de Pará (région Norte) (cf. FIGURE 2.1). Le Pará est le deuxième État du Brésil en terme de superficie (1 248 042km²). Sa population dépasse actuellement les six millions d'habitants, avec des concentrations de la population essentiellement dans sa partie orientale et dans une seconde mesure méridionale (Droulers, 2004).

Ce *Projeto de Assentamento* a été établi sur le territoire de la municipalité d'Itupiranga et occupe une superficie de 10 026 hectares. La forme de ce territoire est irrégulière et s'inscrit au sein d'un quadrilatère formé par la connexion en ligne droite des coordonnées suivantes :

- 05°12'20" de latitude sud,
- 05°20'40" de latitude sud,
- 49°48'00" de longitude ouest,
- 49°56'40" de longitude ouest.

Ce projet d'installation est à 70 kilomètres de la route Transamazonienne (BR-230). La plus proche ville nommée Marabá se situe à une distance d'environ 120

FIGURE 2.1 – PA Benfica au Brésil



N. Sirakov, 2016

Source : IBGE, May et al. (1985)

kilomètres au sud-est de Benfica (Sampaio, 2008) qui est accessible uniquement par piste. Le trajet prend entre 2,5 et 4 heures de route en fonction de l'état de celle-ci, car elle est goudronnée uniquement entre Marabá et Itupiranga (cf. FIGURE 2.2).

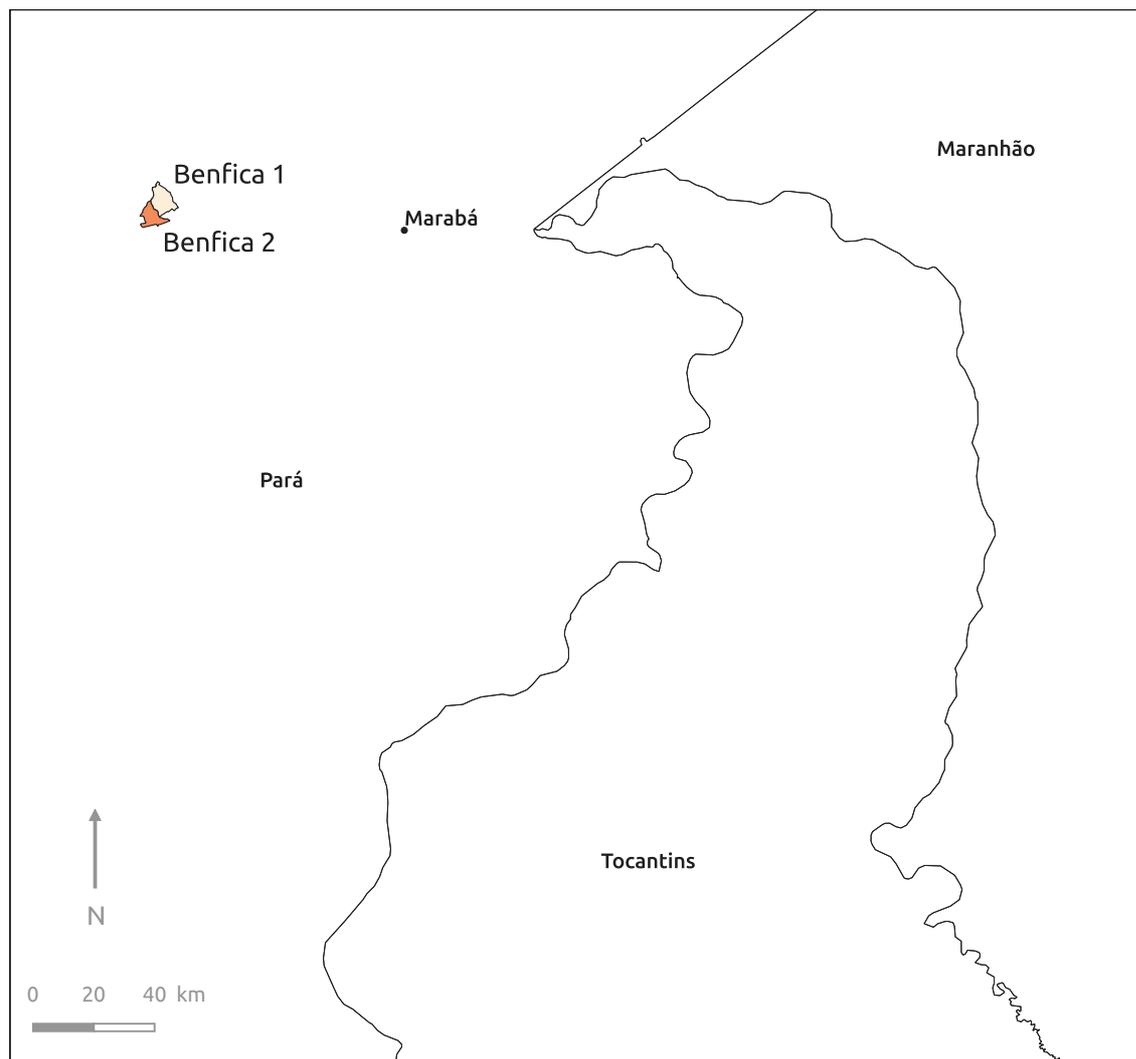
2.2 Anthropisation et colonisation

Le PA Benfica est occupé depuis 1994 par une population avoisinant les 1 000 habitants (Sampaio, 2008). Il se divise en deux sous ensembles séparés par un cours d'eau (cf. FIGURE 2.2) :

- Benfica 1 - occupant la partie septentrionale de ce territoire,
- Benfica 2 - situé sur la partie méridionale.

Le territoire de Benfica est divisé en 183 lots délimitant les propriétés agricoles (cf. FIGURE 2.3). La distinction entre les deux sous-ensembles se base sur l'ancienneté de l'occupation du sol et sur le mode d'accès à la terre agricole (Sampaio, 2008). Le PA Benfica était initialement occupé par des grands propriétaires (fazendas).

FIGURE 2.2 – PA Benfica au sein de son contexte territorial

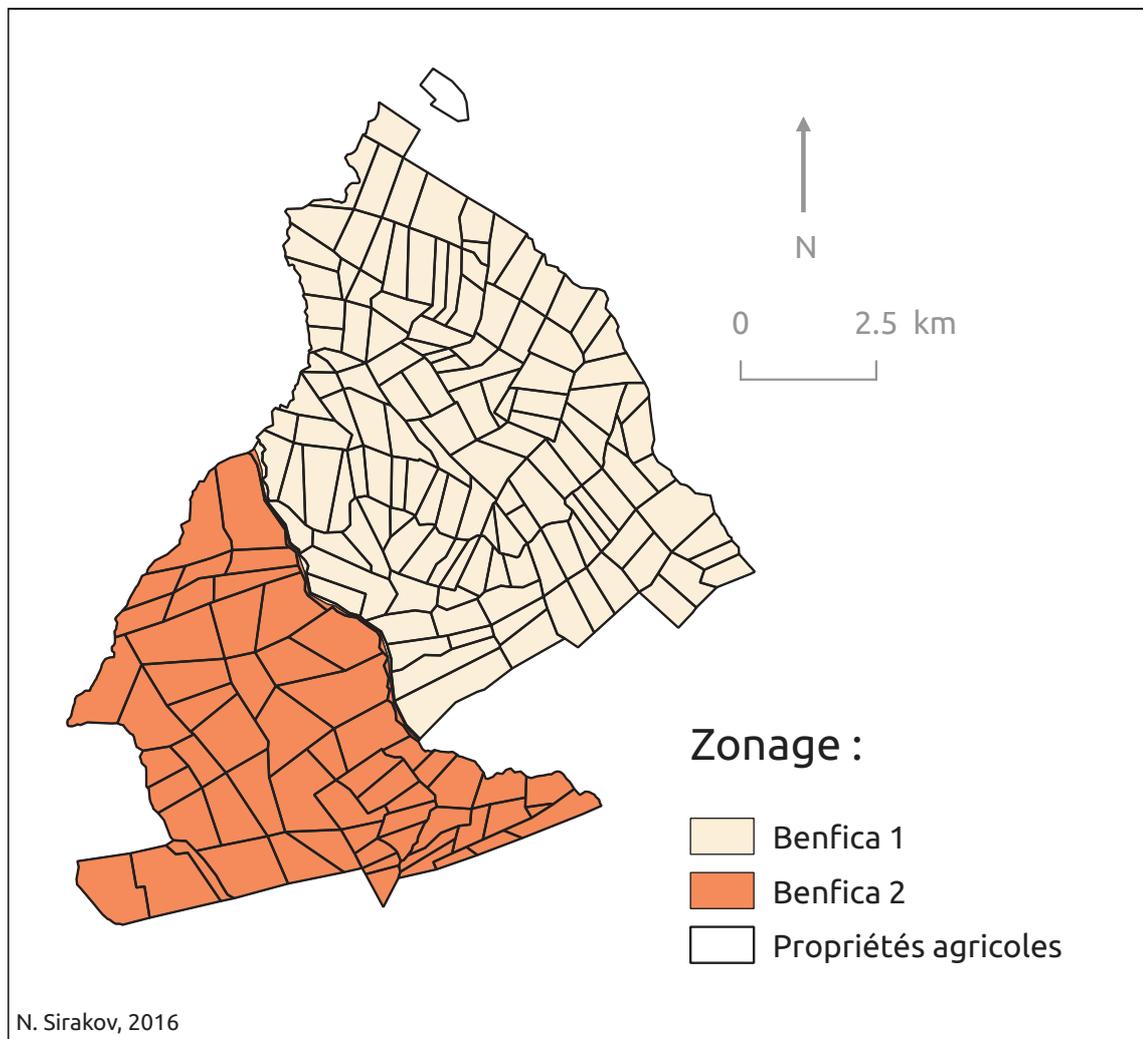


Actuellement Benfica 1, appelé aussi « *aréa dos colonos* », est peuplée par des colons dont les propriétés ont une superficie moyenne de 50 hectares (Laques *et al.*, 2012). Il y a 134 familles installées sur les 126 lots de cette zone. Ces lots, occupés pour certains par deux à trois familles, étaient composés lors de leur mise à disposition en moyenne de 37% de forêt tropicale humide, 15% de forêt secondaire, 7% d'agriculture et 41% de pâturages (Sampaio, 2008). Chaque lot possède donc une importante ressource en bois, mais aussi des cultures temporaires comme le riz, le manioc et le maïs qui constituent la principale source vivrière pour chaque famille. L'installation des activités agricoles est précédée par la déforestation de la forêt tropicale humide en place. Ces activités agricoles consistent soit en une agriculture après brûlis, soit en un élevage extensif de bovins sur des pâturages (Sampaio, 2008).

Benfica 2 est occupé par des petits fermiers dont les lots s'étalent en moyenne

sur une superficie de 150 hectares (Laques *et al.*, 2012). Les activités agricoles dans cette zone ont vu le jour pour 34 parmi ces 57 lots au début des années 1990 suite à leur achat par des colons. Les 23 lots restants, situés dans la partie sud-est des Benfica 2 (ancienne zone de protection de l'environnement), ont une superficie allant jusqu'à 50 hectares. Ces lots abritent des exploitations familiales spécialisées depuis des décennies dans l'élevage. Les activités agricoles sont assurées principalement par les membres de la famille. Le recours à une main d'œuvre extérieure employée en permanence reste rare. Une mauvaise gestion des pâturages fait que ceux-ci peuvent rapidement être dominés par des espèces endémiques autres que les cultures (comme le babaçu) et se transformer à terme en forêts secondaires de babaçu appelés *babaçuals*. Pour certains agriculteurs, le seul moyen de retrouver, dans ces conditions, des terres agricoles est alors désormais la déforestation au sein de ses propres réserves de forêt (Sampaio, 2008).

FIGURE 2.3 – PA Benfica : zonages



2.3 Cadre naturel

2.3.1 Hydrologie

Le PA Benfica s'insère à l'intérieur du bassin versant de la rivière Cajazeiras. Cette rivière qui s'écoule du sud-ouest vers le nord-est sur le territoire d'Itupiranga est un affluent du fleuve Tocantins qui est le fleuve majeur, entièrement brésilien, s'écoulant du sud au nord sur plus de 2 000 kilomètres. Le réseau hydrographique de Benfica est composé de quelques rivières permanentes, comme Rio da Esquerda, Igarapés Benfica, Santa Isabel et Palmeira, ainsi que de nombreux cours d'eau de plus petite taille. En moyenne, environ 70% des lots du PA Benfica ont accès à une source d'eau pérenne tout au long de l'année (Sampaio, 2008).

Dans Benfica 1 les cours d'eau sont en général étroits (larges de moins d'un mètre) et se faufilent entre les collines. Sur le territoire de Benfica 2 les cours d'eau, appelés *igarapés*, sont plus importants (dépassant un mètre de large), mais moins nombreux. Benfica 2 se démarque aussi par la présence de surfaces d'eau stagnantes (Sampaio, 2008).

2.3.2 Pédologie

Sur le territoire de Benfica les sols varient suivant les formes du relief :

- les plateaux sont pré-dominés par des sols ferrallitiques (latosols),
- les parties moyennes des versants sont couvertes de podzol,
- les parties basses des versants sont couvertes de cambisols.

Les sols les plus représentatifs sont les latosols jaunes d'une texture argileuse avec des nodules ferrallitiques dans leurs parties supérieures. Ils sont caractérisés par leur drainage relativement facile, ainsi que par leur épaisseur. Les cambisols sont moins épais et le drainage s'opère plutôt dans le sens latéral que vertical (Sampaio, 2008).

2.3.3 Climat

Il n'existe pas de station météorologique sur le site même de Benfica. Les suppositions présentées ci-dessous sont basées sur des données récoltées entre 1973 et 1990 et provenant de la station la plus proche qui est située dans la ville de Marabá. Le climat se caractérise comme un climat tropical humide. Malgré des indices pluviométriques annuels élevés, la présence d'une saison sèche distincte est enregistrée. La température moyenne annuelle est de 26,1°C avec un maximum de 31,7°C et un minimum de 22,1°C. Le taux mensuel moyen de l'humidité relative de l'air varie entre 76% et 86%. Sa moyenne annuelle se rapproche de 82% (Watrin *et al.*, 2003). La pluviométrie s'élève à un total annuel moyen de 2087,5 mm. Les plus hauts taux de pluviométrie ont été enregistrés durant les mois de février (357mm), mars (386,8mm) et avril (298,8mm), alors que les mois les plus secs sont juin (34,4mm), juillet (20,6mm) et août (56,1mm) (Sampaio *et al.*, 2000).

2.3.4 Couvert végétal

La végétation initiale est une forêt dense humide caractérisée par la présence de lianes et de palmiers parmi lesquels nous rencontrons le babaçu. Par le passé cette forêt couvrait la totalité du PA Benfica. Dans les années 1970 les premiers défrichements ne concernent que quelques hectares. A partir de la fin des années 1980 un processus de déforestation plus rapide lié à l'installation de petits fermiers et colons se met en place. Actuellement les surfaces encore occupées par la forêt sont de plus en plus rares.

Deuxième partie

Problématique

3	Développement durable en Amazonie	15
4	Autour de l'espèce babaçu	23
5	Projet Open Science, thèse et questionnements	33

Développement durable en Amazonie

AVANT de rentrer dans les détails de la problématique du projet Open Science et de la thèse elle-même, nous consacrons ce chapitre à une vision plus large de l'interaction Homme-Nature et de la notion de développement durable.

Depuis le siècle dernier, la question du caractère limité ou illimité des ressources naturelles a provoqué de nombreux débats et notamment introduit le concept de co-viabilité issu du programme « Coviabilité des Systèmes écologiques et économiques : approche modélisatrice en milieu marin » développé au sein de l'IRD (Fur *et al.*, 1999). Le cadre du projet Open Science nous amène à réfléchir sur la gestion durable du babaçu, c'est-à-dire à la co-viabilité de son écosystème avec les socio-systèmes dans lesquels il est présent. Le chapitre fait l'objet d'un large exposé centré autour de l'évolution de l'interaction Homme-Nature en s'appuyant sur une rétrospective générale allant du néolithique (période de changements majeurs dans la relation société-nature) jusqu'à l'époque charnière entre la fin du XIX^e siècle et le début du XX^e siècle (période d'importants questionnements autour de cette relation). Par la suite nous nous interrogerons plus particulièrement sur le construit territorial amazonien au sein duquel le palmier babaçu est amené désormais à évoluer.

Sommaire

3.1	Appropriation de la nature	15
3.2	La démesure humaine	17
3.3	Appropriation de l'Amazonie	17
3.3.1	Déforestation de la forêt amazonienne	18
3.3.2	Mécanismes du front pionnier	19

3.1 Appropriation de la nature

L'homme moderne (*Homo Sapiens*), se distingue de ses prédécesseurs par le fait qu'il a réussi à atteindre un niveau d'évolution lui permettant d'influencer l'évolution des mécanismes bio-climatiques globaux. Nous agissons tellement sur l'évolution de notre planète que des scientifiques s'unissent de plus en plus autour de la définition d'un nouveau concept reflétant la création d'une nouvelle époque géologique : l'« Anthropocène » (Lewis et Maslin, 2015).

Cependant, nous devons chercher les processus qui ont permis cette évolution, en nous tournant vers le passé. Un des plus grands faits marquants de la civilisation de l'homme moderne est le changement définitif, durant le néolithique, de l'organisation du système socio-économique dans lequel il évoluait. Il s'agit là de l'acceptation progressive de la sédentarisation au sein de la société. Elle consiste en l'abandon croissant des activités nomades des chasseurs-cueilleurs au profit de celles relevant du domaine des agriculteurs. La transition du paléolithique au néolithique se marque donc par l'évolution du fonctionnement en petites tribus vers des sociétés érigeant des villages et des villes. Ces nouvelles sociétés deviennent alors capables d'imaginer la domestication et l'élevage et de fonder les bases de la culture : sélection de graines, semailles, irrigation, stockage. . .

L'objectif de ces sociétés se tourne désormais vers la production et le stockage de surplus afin de réduire l'incertitude alimentaire et ainsi d'assurer leur pérennité à long terme. Le mode de pensée de ces peuples néolithiques a radicalement changé : sans le savoir, l'Homme du néolithique était au centre d'un des plus remarquables changements culturels qui se soient jamais produits au sein de l'humanité. La vision que les sociétés ont de la relation entre la Nature et l'Homme a été modifiée. L'Homme qui était dépendant des ressources naturelles assimile le fait qu'il peut les gérer pour assurer sa survie.

Côté scientifique, au cours des siècles, l'essor des sciences sociales porte sur des études exhaustives relatives à des sujets précis et limités (des monographies). Les sciences géographiques s'intéressent aux particularités de chaque région (en termes de localisation des ressources et des populations), alors que les anthropologues et les ethnologues se préoccupent de celles de chaque culture (usages et pratiques). Les questionnements se tournent vers une interrogation des relations de dépendance qui peuvent exister entre l'Homme et son environnement naturel (Deneux, 2006).

La vision déterministe des relations Homme-milieu est dominante à la fin du XIX^e siècle dans les interrogations autour de cette relation. Certains géographes comme Friedrich Ratzel, Emmanuel de Margerie et Albert de Lapparent, sont de fervents défenseurs du déterminisme en sciences géographiques. D'après eux, la localisation des implantations humaines est subordonnée au déterminisme naturel (Deneux, 2006). Le tournant dans la manière d'analyser cette relation se fait au début du XX^e siècle. C'est notamment à travers la notion de « genre de vie » que Vidal de la Blache ouvre la voie à ce que Lucien Febvre nommera en 1922 le « possibilisme » (Claval, 2001; Deneux, 2006). Vidal considère que les sociétés humaines s'adaptent et contournent les contraintes du milieu. Lucien Febvre fournit une interprétation de la pensée vidalienne qui peut être synthétisée par la phrase suivante : « la nature propose, l'Homme dispose » (Claval, 2001).

Désormais l'Homme est considéré en tant que libre arbitre de son sort. C'est exactement ce qui se produit en Amazonie brésilienne durant le XX^e siècle. Suite à la grandissante pression démographique, dans cet immense pays-continent, les gouvernements successifs vont entreprendre des politiques de colonisation du bassin amazonien (Droulers, 2004). Cette colonisation va de pair avec l'essor de l'urbani-

sation et de l'industrialisation du pays, ainsi qu'avec celui du développement du réseau routier à l'intérieur du bassin amazonien. On est loin de l'idée pionnière du géographe Élisée Reclus pour lequel il est nécessaire que l'Homme vive en harmonie avec la Nature et qui condamnait « la brutale violence avec laquelle la plupart des nations traitaient la terre nourricière » (Reclus, 1990).

3.2 La démesure humaine

Plus de dix mille ans après la révolution néolithique, notre société moderne capable plus que jamais de choisir son propre sort, demeure pourtant toujours dirigée par ce même dispositif transformé en mécanisme de rentabilité économique engendrant des profits financiers : produire de plus en plus à des coûts de moins en moins élevés. Ce mécanisme est devenu un leitmotiv pour la plupart des entreprises. Le cumul de richesses et de profits les incite à tomber dans cette spirale sans fin. Akerlof et Shiller (2016) deux prix Nobel d'économie s'efforcent de dénoncer dans leur ouvrage intitulé « Marchés de dupes » les pratiques mensongères, manipulatoires et malhonnêtes des entreprises qui nous incitent « à acheter, bien trop cher, des produits dont nous n'avons nul besoin. . . ».

Malgré l'émergence de la prise de conscience sociétale en termes de production raisonnée respectueuse de la nature et de la société, basée sur une réflexion et construction collective durable, nous vivons dans un monde où prédomine la consommation et la rentabilité. Dans ce monde assujetti à la puissance de l'industrie agro-alimentaire les conséquences en termes de pollution et les besoins alimentaires de la société importent peu. Il existe un net décalage entre les besoins alimentaires, les envies alimentaires créés par ce même secteur et largement soutenues et véhiculées par les médias et les niveaux de productions assurés. L'élevage se multiplie et s'intensifie sur tous les continents terrestres (exception faite bien sûr pour l'Antarctique), alors qu'il représente la majeure source de pollution et d'émission de gaz à effet de serre (Andersen, 2015). La démarche observée au Brésil ne fait pas exception et se soumet à la même règle. La coupe de la forêt tropicale native s'intensifie au Brésil depuis les années 1960. Cette coupe s'opère dans un objectif de création de terres agricoles (cultures ou pâturages) (Droulers, 2004).

3.3 Appropriation de l'Amazonie

L'anthropisation, certes lente, mais incontestable du bassin amazonien ne fait aujourd'hui aucun doute. Cette évolution se manifeste au Brésil au travers du remplacement progressif de la forêt tropicale humide amazonienne par des espaces anthropisés ouverts voués à l'installation d'activités agricoles : pâturages pour l'élevage bovin et diverses cultures des sols comme le soja.

3.3.1 Déforestation de la forêt amazonienne

La forêt tropicale humide amazonienne occupe une superficie d'environ 5,5 millions de km². Elle s'étale sur neuf pays en Amérique du Sud dont le Brésil qui abrite environ deux tiers de la superficie du bassin amazonien ($\approx 58\%$) (Droulers, 2004). C'est le plus grand bassin hydrographique sur la Terre (20% des terres immergées) et la plus grande réserve de forêts tropicales au monde (33% des forêts tropicales) (Droulers, 2004).

La déforestation brésilienne de grande ampleur de la forêt amazonienne a débuté dans les années 1930 (Droulers, 2004). Elle s'est intensifiée durant la deuxième moitié du xx^e siècle et surtout depuis les années 1970 au travers de pratiques de défrichement très destructrices (Droulers et Broggio, 2013). Les promoteurs de la colonisation amazonienne des années 1970 ont eu pour leitmotiv la phrase suivante : « une terre sans hommes pour des hommes sans terre » (Arnauld de Sartre, 2006, p. 19). Mais ils ont en quelques sortes nié le fait que le bassin amazonien était déjà un territoire occupé et appartenant aux populations indigènes.

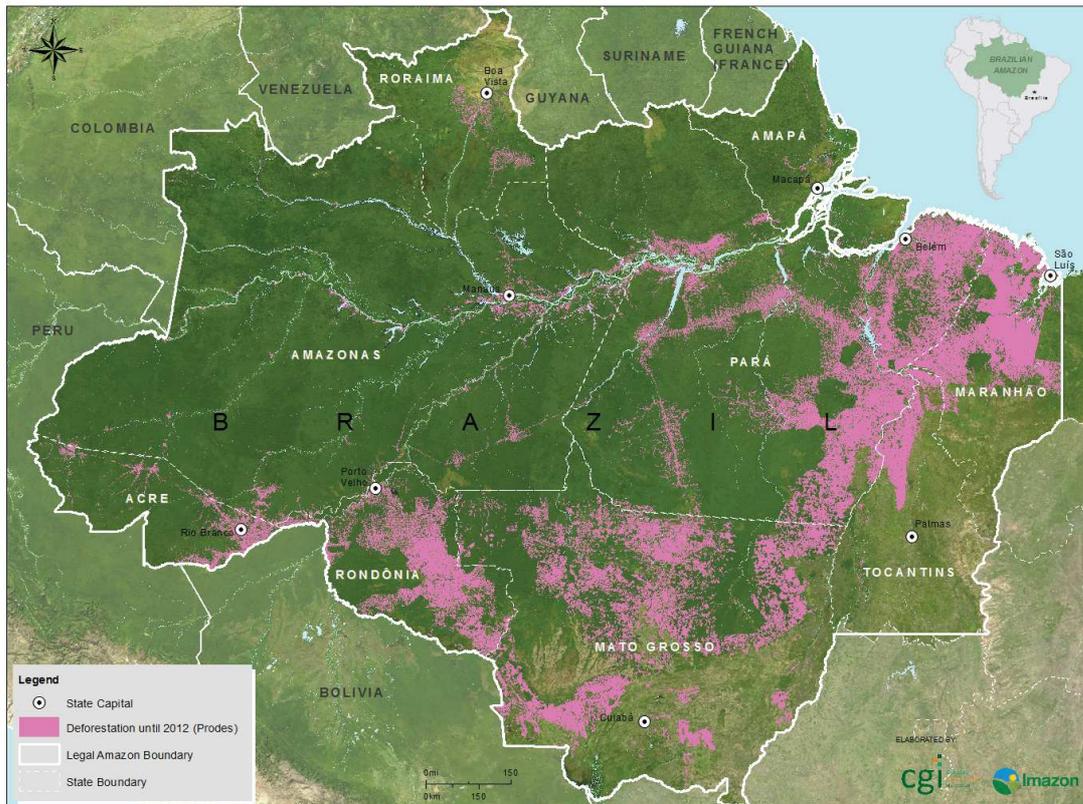
Les tendances actuelles dans ce pays, depuis le début du XXI^e siècle, sont à la stabilisation des niveaux de déforestation liés à l'engagement d'une politique de contrôle et de préservation de la forêt amazonienne à long terme de la part du gouvernement brésilien. La population de la partie brésilienne de l'Amazonie¹ est passée de 5 millions de personnes en 1970 à 23 millions de personnes 40 ans après (Droulers et Broggio, 2013).

La FIGURE 3.1 démontre l'ampleur actuelle de l'anthropisation de la plus grande forêt tropicale humide sur le globe terrestre. Sur cette image apparaissent clairement les zones d'engagement de la déforestation au sein de la forêt. D'après des estimations, en moyenne, ont été défrichés 20 000km²/an de son territoire entre 1985 et 2005.

La tendance actuelle est à la baisse avec par exemple une superficie défrichée de 6280km² pour la période août 2010 - juillet 2011. Le cumul de la superficie défrichée a atteint dans les années 2010 16% du territoire du bassin amazonien brésilien (Droulers et Broggio, 2013). Les deux tiers de la superficie défrichée sont occupés par des pâturages destinés à l'élevage bovin avec en moyenne un bovin par hectare et trois bovins par habitant. L'élevage est une activité bénéficiant d'abattoirs modernes fournissant de la viande à des prix très compétitifs, ce qui la rend plus attrayante financièrement que celle des cultures (Droulers et Broggio, 2013). Les pâturages abritent une autre particularité : durant le processus de défrichement, les hommes laissent systématiquement des palmiers dont le babaçu au sein de la zone défrichée, puisque ce palmier assure la protection nécessaire au bétail pour s'abriter du soleil (cf. FIGURE 3.2). C'est ainsi que ce palmier est une des rares espèces endémiques qui a l'opportunité d'évoluer et de s'adapter aux milieux ouverts anthropisés. D'autant plus que ce mécanisme facilite l'accès, aux populations locales

1. Dans le sens large de l'Amazonie légale incluant une grande partie du Maranhão et le Mato Grosso

FIGURE 3.1 – Déforestation du bassin amazonien



Source : <http://imazon.org.br/slide/desmatamento/>

défavorisées, à une des ressources extractivistes du Brésil qu'est le babaçu.

L'évolution de la déforestation au Brésil durant la deuxième moitié du XX^e siècle peut être associée aux grandes étapes de l'évolution socio-politique de ce pays (Nasuti, 2010). On passe peu à peu d'un système d'accès à la terre libre et non contrôlé centré sur l'exploitation du babaçu à un système étatique d'intégration des territoires et de modernisation conservatrice accompagnée par l'essor de l'élevage bovin et l'arrivée en masse de la culture de soja. Enfin depuis les années 1995 la politique nationale brésilienne s'oriente vers le soutien aux petits agriculteurs, le développement des PA, et plus généralement vers une réflexion à long terme sur la gestion et l'exploitation de ses ressources (Nasuti, 2010). La déforestation a amené de profondes transformations dans la relation entre le palmier babaçu et les populations locales l'utilisant comme ressource extractiviste (*caboclo* et colons).

3.3.2 Mécanismes du front pionnier

L'interaction entre les ressources naturelles et les populations indigènes a fortement été bouleversée par l'avancement du front pionnier au Brésil. La nouveauté de

FIGURE 3.2 – Le babaçu au sein des pâturages



Cliché : D. Mitja

cette interaction est qu'elle émerge autour d'un territoire frontalier de transition.

Il est important ici de donner une définition du concept de frontière de transition. Dans la plupart des cas, l'objet d'étude qu'est la frontière est assimilé à une rupture, une coupure, une séparation matérialisée par des marqueurs physiques comme des barrières, murs, bornes...

L'utilisation que nous faisons ici du concept de frontière de transition est toute autre. Nous faisons référence plutôt à un construit social assurant la fonction de transition entre ses deux côtés. Le front pionnier est une création de la société dont les valeurs, les fonctions et les significations sont régies par celle-ci (Picouet et Renard, 2007). C'est une zone qui accomplit la fonction d'interface entre des discontinuités spatiales (Wackermann, 2003). Le front pionnier agricole au Brésil se caractérise par la mise en relation des espaces situés de ses deux côtés :

- un milieu marqué par son fort caractère naturel - la forêt tropicale humide,
- un territoire déjà structuré marqué par son fort degré d'anthropisation - la côte atlantique urbanisée du Brésil.

Dans ce sens le front pionnier agricole apparaît comme une zone tampon de transition. Son avancement au Brésil s'est fait à partir du sud et de l'est du bassin

amazonien. Dans certains cas cet avancement a été facilité depuis les années 1970 par la construction des nouveaux axes routiers amazoniens (cf. FIGURE 3.1). Les éleveurs et agriculteurs sont parmi les premiers à exploiter ces percées au sein de la forêt tropicale humide (Droulers et Broggio, 2013). L'anthropisation lente mais sûre de l'Amazonie consiste dans le remplacement de la forêt par des pâturages ou cultures agricoles. Cette progression lente du front pionnier sous-entend que les différents états brésiliens ont vu leurs territoires se transformer à des périodes temporelles différentes : tout d'abord ont été concernés les états orientaux et méridionaux du bassin amazonien, puis ceux situés plus à l'ouest et plus au cœur du bassin amazonien.

Le mécanisme d'avancement du front pionnier est un processus complexe qui se manifeste plusieurs étapes consécutives d'anthropisation (Nasuti, 2010) :

1. la première étape ouvre de nouveaux territoires,
2. durant la seconde étape les territoires ouverts sont colonisés et structurés physiquement et socialement (certains acquièrent même le statut de villes urbanisées),
3. lors de la troisième étape ces territoires consolidés servent d'assise pour le départ de nouveaux colons vers une nouvelle première étape au sein du bassin amazonien

Nous avons vu dans la section 2.1 (page 7) que la zone de notre étude se situe dans la partie orientale de l'État de Pará. La situation de l'État du Pará est intéressante actuellement pour le projet Open Science : sa zone orientale (incluant notre zone d'étude sur le terrain) fait partie de l'arc de déforestation et a été donc fortement bouleversée durant le dernier quart du XX^e siècle, mais sa partie occidentale reste toujours très majoritairement occupée par la forêt amazonienne et fait l'objet de récentes poussées du front pionnier vers l'ouest (Droulers, 2004). Cet État effectivement a fait l'objet d'une colonisation plus tardive que celle de l'État voisin du Maranhão (situé à l'Est du Pará). La proximité de ces deux États sera utile sur la partie analyse socio-économique du projet Open Science. Nous pouvons donc croire, que les pratiques dans l'État du Pará vont à terme converger vers une situation proche de celles actuellement observées dans l'État du Maranhão. Ceci est d'autant plus vrai que le PA de Benfica se situe à moins de 150km de la partie occidentale de l'État de Maranhão.

Autour de l'espèce babaçu

LE BABAÇU est une espèce native des forêts tropicales humides de l'Amérique de Sud qui évolue désormais dans un contexte de milieux anthropisés. Ce chapitre retracera la présence du palmier babaçu en Amérique du Sud et plus particulièrement au Brésil. Il se centre ensuite sur les questionnements soulevés par la définition biologique de l'espèce. Enfin, nous examinerons également les problèmes soulevés par la complexité du système socio-économique au sein duquel cette espèce évolue.

Sommaire

4.1	Babaçu : importante présence au Brésil	23
4.2	Espèce étudiée et convoitée	24
4.2.1	Taxonomie complexe	25
4.2.2	Biologie du babaçu	25
4.2.3	Babaçu : espèce adaptative	27
4.3	Contexte socio-économique autour du babaçu	28
4.3.1	Le babaçu : une ressource extractiviste disputée	28

4.1 Babaçu : importante présence au Brésil

Le palmier babaçu est largement représenté au Brésil ; d'après des estimations datant de 1982 les surfaces de présence de ce palmier couvraient à cette période une superficie totale d'environ 196 370km². Ces surfaces de présence sont réparties inégalement sur le territoire brésilien. Seulement trois États regroupent 78% de la superficie totale des surfaces de présence : Maranhão (53%), Goiás (15%) et Piauí (10%). Les 22% restants de la superficie des surfaces de présence du babaçu se répartissent entre les États d'Amazonas, du Pará, du Mato Grosso, du Minas Gerais et du Ceara (May *et al.*, 1985). Quelques années plus tard, Anderson et Balick (1988) (d'après Anonyme (1981)) soulignent cette fois-ci que les surfaces de forte présence de la ressource sont localisées sur le territoire des États brésiliens d'Amazonas, Ceará, Goiás, Maranhão, Mato Grosso, Minas Gerais, Pará et Piauí (cf. FIGURE 2.1).

Nous remarquons un net décalage entre les deux sources d'information. En observant la FIGURE 2.1 nous pouvons constater par exemple que l'état de Maranhão ne concentre pas plus que la moitié des zones à babaçu répertoriées. D'après notre connaissance, il n'existe à ce jour aucune carte actualisée montrant la localisation

de la ressource. Nous nous sommes donc basés sur des informations provenant de plusieurs articles (May *et al.*, 1985; Anderson et Balick, 1988; Anderson *et al.*, 1991) pour construire notre carte de présence du babaçu (cf. FIGURE 2.1).

La Bolivie et l'État brésilien de Rondônia ont été définis plus tardivement comme des zones de présence de ce palmier (Anderson *et al.*, 1991). Des études menées plus récemment signalent également la présence du babaçu au Mexique (Teixeira, 2003).

D'après la FIGURE 2.1, le palmier babaçu peut se développer au sein de plusieurs biomes¹. Sa présence est déterminée au sein du biome Cerrado apparenté à une savane. Ce biome se caractérise par des formations végétales basses et est dominé par des plantes herbacées. Le Cerrado s'étale sur plus de 2 millions de km² et occupe un peu moins de la moitié de la partie orientale du pays (23,92% du territoire brésilien (IBGE, 2004), cf. FIGURE 2.1). Dans ces zones, le babaçu pousse exclusivement au sein des forêts galeries qui se forment le long des cours d'eau. Ainsi, il profite d'une ressource en eau pérenne tout en restant à l'abri de la sécheresse prédominante de ces régions.

Le palmier est également présent dans certaines régions du biome de Caatinga qui occupe près de 10% du territoire brésilien. Ce biome, situé dans la partie nord-est du pays se caractérise par son climat semi-aride (IBGE, 2004).

L'habitat par excellence du babaçu demeure toutefois celui des grandes forêts tropicales du biome Amazônia. Il s'agit de loin du plus étendu biome brésilien avec ses 4,2 millions de km², couvrant ainsi la moitié du territoire du Brésil.

Un des atouts majeurs de cette espèce est qu'elle se développe aussi bien dans les zones des forêts tropicales humides que dans celles des *cerrados* brésiliens (Anderson *et al.*, 1991). De par sa présence au sein de plusieurs biomes, le babaçu dispose d'un potentiel d'adaptabilité au milieu environnant, tout en soulignant que dans les régions semi-arides et de savane il se limitera aux alentours des ressources en eau.

4.2 Espèce étudiée et convoitée

Cette section fait le tour des définitions en terme de biologie de l'espèce. Nous exposerons d'une part la nomenclature complexe autour de la taxonomie de l'espèce et d'autre part nous fournirons une présentation détaillée des éléments biologiques définissant cette espèce. Pour cela nous utiliserons le formalisme UML. Ce formalisme, qui s'inscrit dans l'approche « objet », a l'avantage de disposer d'une notation graphique aisément assimilable par tout acteur et de permettre ainsi de définir des représentations consensuelles et aisément partageables.

1. « Un biome appelé aussi macro-écosystème est un ensemble d'écosystèmes caractéristiques d'une aire bio-géographique et nommée à partir de la végétation et des espèces animales qui prédominent et y sont adaptés » Wikipédia

4.2.1 Taxonomie complexe

Au XIX^e siècle et au début du XX^e siècle, lors des travaux de prospection du continent sud-américain, le babaçu se voit attribuer différents noms scientifiques (Glassman, 1977). Vingt-neuf espèces différentes ont été assimilées au babaçu par les scientifiques (Glassman, 1977, 1978). Une réorganisation taxonomique entreprise par Glassman à la fin des années 1970 afin de pouvoir mieux structurer et donc connaître cette espèce, aboutit à la mise en évidence d'au moins six synonymes utilisés pour la même espèce : *Orbignya martiana* Barb. Rodr., *Orbignya speciosa* (Mart.) Barb. Rodr., *Orbignya barbosiana* Burret, *Orbignya macropetata* Burret, *O. Dammeriana* Barb. Rodr. et *Orbignya phalerata* Mart (Glassman, 1977; Anderson et Balick, 1988; Anderson *et al.*, 1991). Actuellement le palmier babaçu est habituellement² reconnu sous le nom scientifique d'*Attalea speciosa* Mart. ex Spreng. (Henderson, 1995) mais est aussi nommé *Orbignya phalerata* Mart. par certains auteurs (Lorenzi *et al.*, 2004).

4.2.2 Biologie du babaçu

Le palmier babaçu (*Attalea speciosa*) est une espèce endémique de l'Amérique du Sud. Cette espèce appartient à la famille Arecaceae, qui regroupe plus de 2 800 espèces différentes (Tomlinson, 1979).

Un individu babaçu est caractérisé par plusieurs composants (cf. FIGURE 4.1). Comme tous les palmiers, le babaçu, se différencie des arbres par la présence d'un **stipe** assurant la fonction de tronc. Le stipe se matérialise comme une tige très robuste supportant le poids du palmier.

Les **feuilles** poussent directement sur le stipe et restent accrochées à celui-ci tout au long de leur existence. À terme elles finissent par se détacher et tomber par terre. Le nombre de feuilles augmente avec l'âge de l'individu et leur division se fait au fur et à mesure de la croissance du palmier.

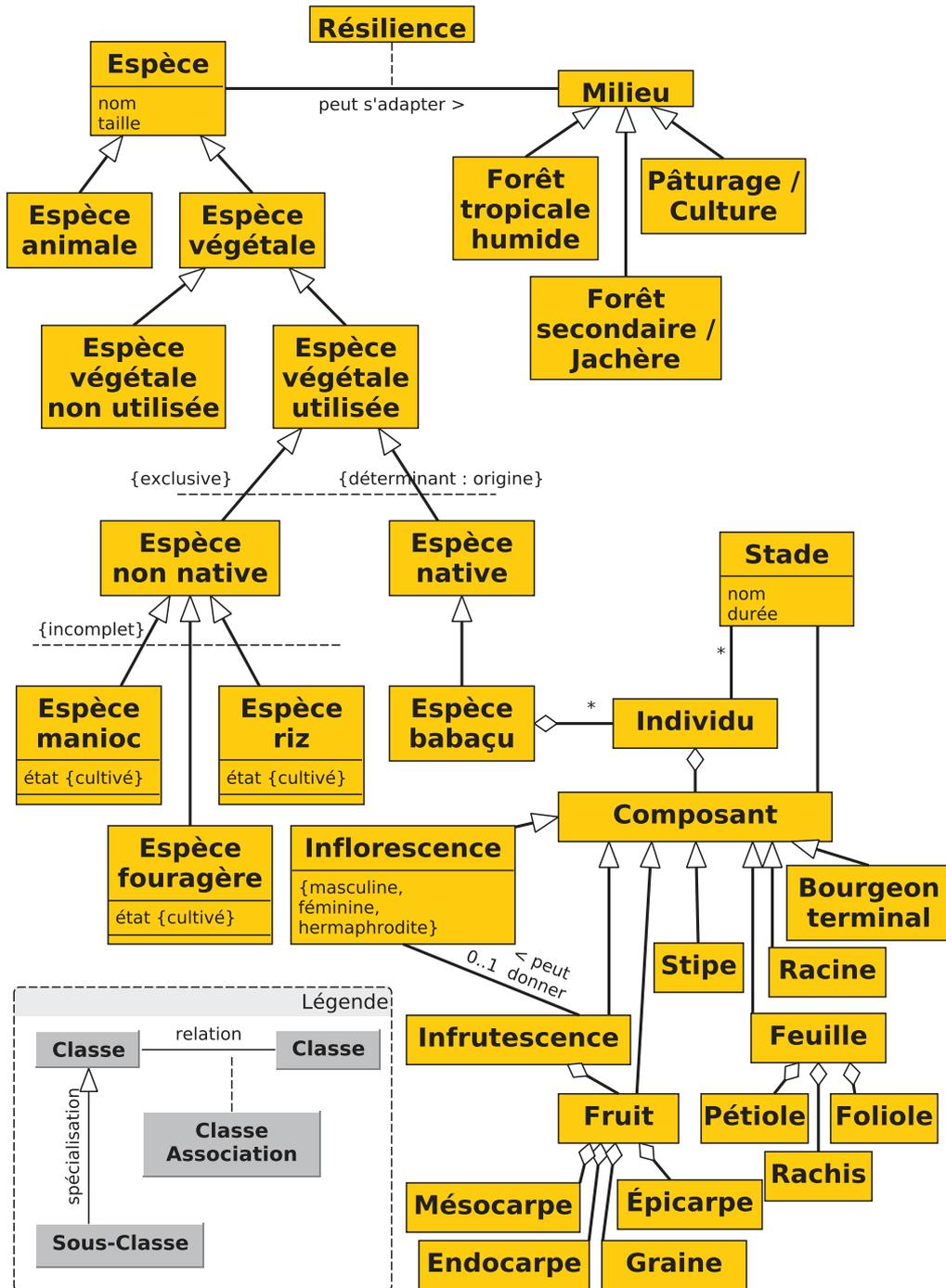
Les feuilles sont composées d'une **pétiole** transformée dans sa partie supérieure en **rachis** soutenant les **folioles**. Dans le cas d'un limbe divisé (la division succède à la non division) le limbe se compose de plusieurs folioles (cf. FIGURE 4.1).

La **racine** du babaçu est souterraine. Son bourgeon, matérialisé comme chez les autres palmiers en tant que **bourgeon terminal**, est caractérisé par une ascension spécifique. Ce bourgeon terminal est souterrain au début de la croissance du palmier. Puis au fur et à mesure que l'individu grandit, le bourgeon terminal devient aérien.

Certains individu adultes sont aussi caractérisés par la présence d'une ou plusieurs **inflorescence(s)** qui peuvent être féminines, masculines ou hermaphrodites. Les inflorescences féminines et hermaphrodites peuvent se transformer en **infrutescences** (cf. FIGURE 4.1).

Le **fruit** est accroché sur une infrutescence (cf. FIGURE 4.1). Il se caractérise par des zones concentriques bien délimitées : la semence est incorporée au sein du très

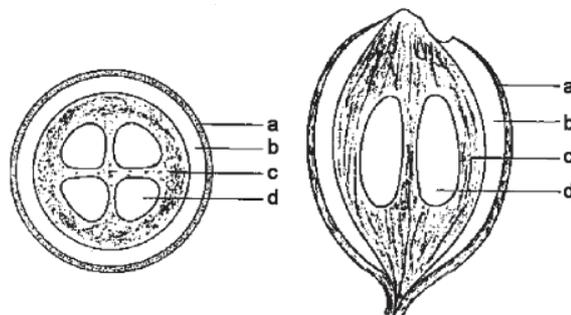
2. La communauté scientifique n'est, cependant pas, actuellement d'accord sur l'utilisation du terme *Attalea speciosa* pour désigner cette espèce de palmier.

FIGURE 4.1 – Biologie du babaçu (formalisme UML, in Mitja *et al.* (sous presse))

dur endocarpe qui est enveloppé par un mésocarpe farineux, lui même couvert par

le très résistant épicarpe (cf. FIGURE 4.2).

FIGURE 4.2 – Fruit du babaçu (adapté de Teixeira (2005))



a - épicarpe, b - mésocarpe, c - endocarpe, d - semence

4.2.3 Babaçu : espèce adaptative

Les espèces, comme le babaçu, nativement présentes au sein de la forêt tropicale humide sont amenées de plus en plus à subir les nouvelles conditions de vie dictées par les situations de forte présence humaine comme le front pionnier et le post-front pionnier. Le babaçu est reconnu pour sa bonne adaptabilité à ces conditions de vie anthropisées (Mitja et Ferraz, 2001). Il se distingue par sa résistance et sa persistance face aux feux annuels dans les champs (May *et al.*, 1985). Ces qualités vont aider le babaçu à dominer le paysage de post-déforestation (Nasuti, 2010).

Le babaçu est une plante colonisatrice et adventice³. Dans le cas d'un « abandon » des pâturages par les agriculteurs et de leur transformation en jachères, le babaçu développe le comportement d'une plante invasive et peut coloniser totalement ces milieux. En cas de poursuite de l'abandon des terres, celles-ci peuvent être monopolisées par le palmier et se transforment en forêts secondaires monospécifiques peuplées entièrement de babaçu. Ces forêts secondaires particulières sont connues au Brésil sous le nom de *babaçuais*. Le babaçu possède donc un très fort degré d'adaptabilité aux conditions bio-climatiques.

De plus cette espèce se distingue par son ajustement au contexte anthropisé et aux nouvelles conditions de vie dictées par des occupations du sol résultantes de l'activité anthropique (Anderson, 1983).

3. Utilisé en agriculture, le terme adventice désigne une plante qui pousse spontanément dans une culture et dont la présence est plus ou moins nocive à celle-ci.

4.3 Contexte socio-économique autour du babaçu

Le contexte socio-économique, au sein duquel le babaçu émerge comme une ressource extractiviste⁴, est un construit complexe associant des acteurs agissant à des échelons différents : du niveau local au niveau international. Il est fort de constater que les acteurs sont divers et que les relations qu'ils entretiennent entre eux ne sont pas toujours régies par des sentiments positifs. C'est pourquoi à présent nous souhaitons examiner plus en détail les problématiques régies par un contexte socio-économique bouleversé.

4.3.1 Le babaçu : une ressource extractiviste disputée

4.3.1.1 Historique

Le contexte socio-politique du Brésil a largement contribué à l'utilisation massive de ce palmier comme une ressource extractiviste. Depuis la moitié du XX^e siècle à nos jours nous distinguons trois grandes étapes dans la construction intellectuelle autour de cette espèce (Nasuti, 2010).

Les années 1950 jusqu'à mi-1970 sont le symbole d'un essor de la ressource largement exploitée par les *caboclo* et les petites unités familiales. Durant cette période le babaçu symbolisait la grandeur de la nature au Brésil et véhiculait le message de fierté nationale. Cette période est marquée par la création d'un parc industriel d'huileries destiné au marché national (Nasuti, 2010).

Les années mi-1970 à mi-1990 sont marquées par le développement de l'élevage bovin en Amazonie. Les activités autour du palmier babaçu constituent, petit à petit, une gêne pour les éleveurs et suscitent les prémices de ce qui deviendra, au fur et à mesure, un important conflit d'usages d'ampleur nationale autour de la ressource babaçu. Cette période marque le déclin du marché de l'huile de babaçu qui est fortement concurrencé par l'introduction et l'essor de la mise en culture du soja. Néanmoins durant cette période est enregistré un fort essor (pic) dans la production d'amandes issues du babaçu (Nasuti, 2010). D'après les estimations, dans les années 1980, l'extractivisme concernait 450 000 familles, pour un total de 2 millions de personnes⁵ parmi les plus démunies du pays⁶ (May *et al.*, 1985).

La dernière période s'étale de mi-1990 à nos jours. Cette période s'inscrit dans le contexte national de re-démocratisation et de volonté nationale de soutien à l'agriculture familiale (Nasuti, 2010). Durant ces années le babaçu est très peu concurrentiel sur le marché des huiles lauriques. Le fait marquant de cette période est l'essor d'une toute nouvelle organisation autour de l'extractivisme avec l'émergence de pe-

4. En portugais « extrativismo » signifie au Brésil une activité de cueillette suivie de commercialisation de produits non ligneux de la forêt (Pinton et Empereire, 1992). Ce terme donne la priorité à l'aspect marchand.

5. La population totale du Brésil était de 191 millions de personnes en 2010 (source : *Instituto Brasileiro de Geografia e Estatística* (IBGE))

6. Ces populations sont dans la très grande majorité des populations sans terres

tits producteurs et surtout des « Quebradeiras de coco babaçu⁷ ». Ces populations majoritairement féminines sont de plus en plus confrontées à des problèmes d'ordres socio-économique et géo-politique (Almeida *et al.*, 2005).

4.3.1.2 Acteurs et produits

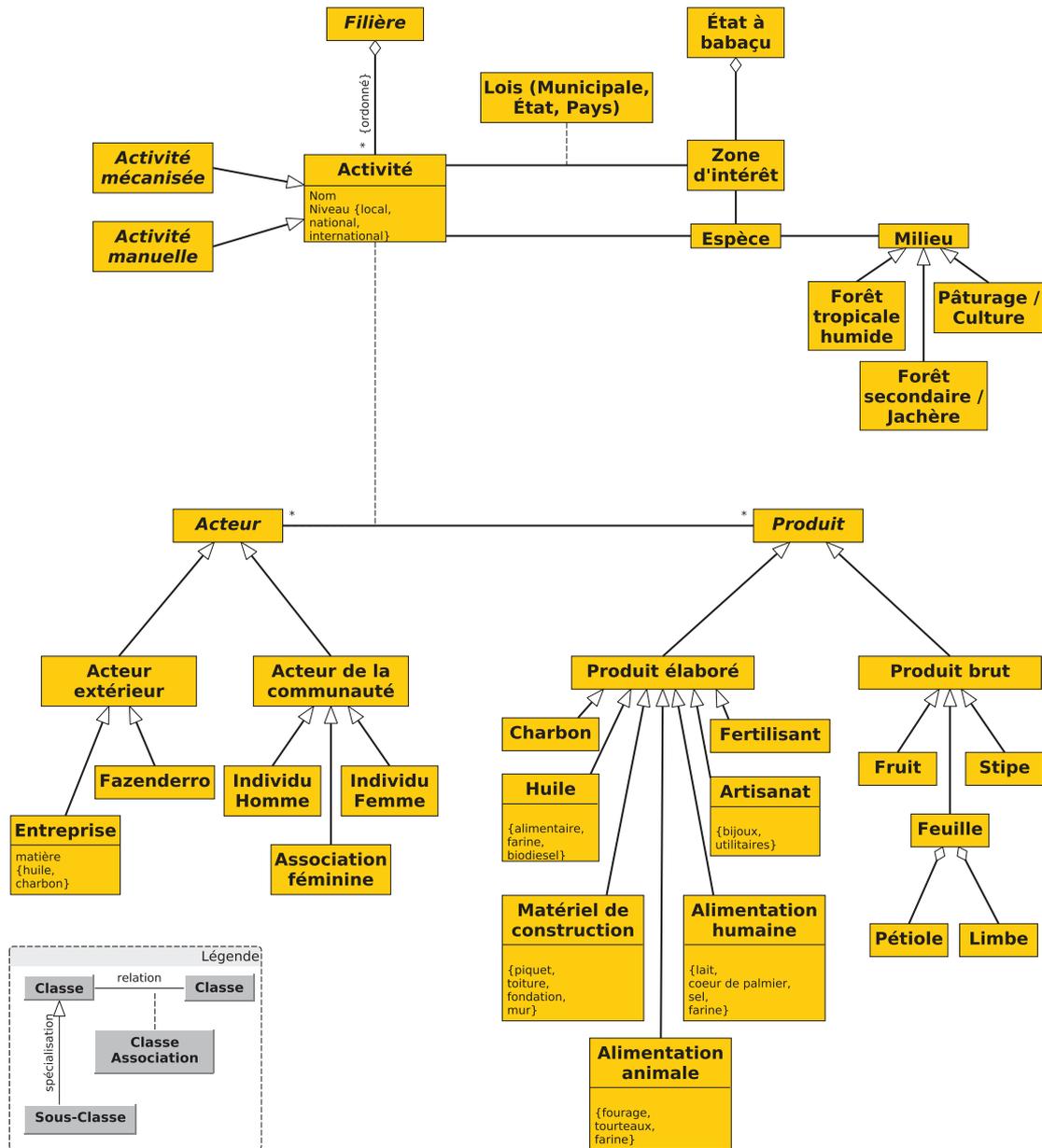
Le palmier babaçu est une espèce utilisée de la racine jusqu'à l'extrémité des feuilles avec plus de 60 utilités répertoriées (Araújo et Lopes, 2012). May *et al.* (1985) en citent une quarantaine dont les principales ont été représentées sur la FIGURE 4.3 et sont détaillées dans les paragraphes suivants. Certaines utilités de la filière babaçu découlent soit de collectes directes (produit brut) soit de transformations plus ou moins élaborées (produits dérivés).

Les individus adultes du palmier babaçu font l'objet de la plupart des collectes pour les diverses utilisations, seules les feuilles des individus les plus jeunes sont utilisées. Le tressage des folioles des feuilles permet de réaliser divers objets d'artisanat comme des paniers pour le transport et le stockage, des portes, fenêtres et tapis, des éventails, des pièges pour la chasse, des nids pour les poules. Les feuilles entières, limbe et pétiole servent de matériel de construction pour la réalisation de toits, de murs, ou d'encadrement de portes et de fenêtres. Les feuilles ont aussi un usage agricole ; leur ombre dans les pâturages est appréciée du bétail aux heures les plus chaudes de la journée et une fois séchées et décomposées, elles servent de fertilisant. Les jeunes feuilles des individus acaules servent de fourrage pour le bétail et les chevaux. Chez les indiens *Kayapó* les feuilles sont utilisées pour la confection de masques (*máscaras*) qui représentent des animaux comme les singes (*macacos*) pour différentes fêtes traditionnelles et aussi pour s'asseoir au cours de ces fêtes, ou dans la vie courante lors de déplacements dans la forêt et dans la savane. Les pétioles des feuilles permettent de construire des barrières pour séparer les parcelles ou isoler le jardin de case du reste de l'exploitation ; les indiens *Kayapó* les utilisent aussi pour barrer les petits cours d'eau lors d'une pêche traditionnelle. Le pétiole est pourvu d'une sève antiseptique utilisée pour soigner de petites blessures (Mitja *et al.*, sous presse).

Le stipe ou faux tronc sert pour réaliser les fondations des maisons, les ponts et certains meubles comme des bancs. Le cœur de palmier de 2 à 3 kg est comestible et est utilisé tant pour l'alimentation humaine que pour l'alimentation animale. Ce cœur de palmier produit également un gaz qui favorise la maturation des bananes. La sève du stipe peut être transformée en une boisson fermentée et elle est attractive pour des insectes dont les larves sont comestibles. Le stipe en décomposition fournit également un engrais très apprécié pour les jardins suspendus. Certaines tribus indigènes comme les *Kayapó* brûlent le stipe et utilisent cette cendre comme succédané du sel (Posey, 1987).

Les spathes qui protègent les infrutescences peuvent servir de corbeille à fruits et certaines d'entre elles spécialement grandes sont utilisées par les enfants comme

7. Ceci signifie : les femmes qui cassent les fruits du babaçu

FIGURE 4.3 – Produits et Acteurs au sein du système babaçu (formalisme UML, adapté de Mitja *et al.* (sous presse))

luge sur pâturage. Les semences des fruits sont comestibles et on en extrait un lait qui est utilisé dans la préparation de recettes locales. On peut également extraire de l'huile qui sert non seulement dans l'alimentation, et dans la cosmétique (savons, savonnettes), mais qui est également testée comme biodiesel. Les tourteaux de résidus obtenus après extraction de l'huile servent à l'alimentation animale, mais

peuvent aussi être utilisés comme appât pour la pêche aux crevettes et comme substitut du café. Les amandes sont insérées dans un endocarpe ligneux qui une fois brûlé se transforme en un charbon de très bonne qualité. La fumée résultant de ce procédé est insecticide. Cet endocarpe peut également produire des gaz condensés anesthésiants pour le mal de dents. Une fois découpé transversalement cet endocarpe peut être utilisé pour divers artisanats (bijoux, coupelles à fruit...). Le mésocarpe farineux peut être transformé en farine alimentaire qui sert à faire des gâteaux et qui est aussi un remède médicamenteux contre les problèmes gastro-intestinaux. Il peut être également utilisé pour l'alimentation animale. Il est, de plus, attractif pour les animaux sauvages : *agouti* et *paca* (Smith, 1974, 2015) et dans ce cas favorise la chasse. Le fruit entier (épicarpe, mésocarpe, endocarpe et semences) peut également être utilisé pour la production de charbon par des entreprises qui envoient des collecteurs de fruits dans les campagnes entrant alors en concurrence avec les femmes « casseuses de fruit » pour la ressource babaçu. La production de charbon à partir du fruit entier a aussi été enregistrée chez les *Kayapós* du petit village « Las Casas » (González-Pérez *et al.*, 2012). Le fruit entier réduit en poudre et additionné à quelques nutriments (sel et poudre de maïs) est utilisé comme aliment pour le bétail (Macedo, 2015). Parfois les fruits sont parasités par des insectes comme le *Pachymerus nuclearum* (Fabr.) ou le *Carybruchus lipismatus* (Bridwel), Coléoptères Bruchidae, appelés « gongos » (Anderson *et al.*, 1991), qui sont consommés par les populations.

FIGURE 4.4 – *Quebradeiras* ramassant les noix du babaçu



Cliché : Peter Caton/ISPN ; Source : <http://www.cerratinga.org.br/populacoes/quebradeiras>

En ce qui concerne les activités, les femmes ont la responsabilité de celles liées aux fruits (même si les hommes interviennent parfois dans leur transport). La collecte des feuilles est effectuée par les hommes et femmes, même si la construction des toits est plutôt réservée aux hommes, et peut parfois donner lieu à un commerce local.

Alors que la plupart des activités locales sont non commerciales et liées aux usages de la vie courante en milieu rural, l'extraction des graines peut donner lieu à une commercialisation initialement locale par les femmes qui ont comme profession de casser les fruits de babaçu (« *Quebradeiras de coco babaçu* »). Elles peuvent alors produire elles mêmes cette huile de manière artisanale, ou même produire des savons et savonnets qu'elles vendront sur les marchés ou dans des magasins spécialisés. Ces dernières peuvent aussi vendre les graines à des intermédiaires et/ou à des entreprises privées qui en extraient la précieuse huile. Certaines d'entre elles exercent cette activité de manière isolée et ont leurs propres circuits de commercialisation avec une clientèle locale plus ou moins fidélisée. D'autres sont regroupées en associations, comme l'*Associação em áreas de assentamento no estado do Maranhão (ASSEMA)* ou le *Movimento interestadual das quebradeiras de coco babaçu (MIQCB)*, qui leur offrent de nombreux avantages tels que la possibilité de bénéficier de matériel pour moulinier les semences ou pour extraire l'huile, et de tirer parti de circuits de commercialisation plus régionaux voire internationaux.

FIGURE 4.5 – *Quebradeiras* cassant les noix du babaçu



Cliché : D. Mitja

Le travail des *Quebradeiras* se structure principalement autour du ramassage des noix du babaçu (cf. FIGURE 4.4), suivi par la casse de celles-ci et l'extraction des amandes (cf. FIGURE 4.5) destinées à la filière huile-amande du babaçu (Nasuti, 2010).

Projet Open Science, thèse et questionnements

LE DÉBUT de ce travail de thèse a consisté à éclaircir la problématique de celle-ci et ceci à la lumière de la problématique du projet Open Science. Ce chapitre va synthétiser la problématique interdisciplinaire assez large du projet Open Science laquelle, abordée sous une approche systémique autour des interactions Homme-Nature, nous a amené à nous rapprocher des écrits relatifs aux notions de viabilité des écosystèmes et par la suite à participer à un ouvrage dédié au paradigme de coviabilité (Mitja *et al.*, sous presse).

La problématique, plus restreinte, dévolue à la thèse en découlera. Elle explicitera les connaissances autour de l'espèce étudiée afin d'étayer la modélisation ultérieure de la dynamique de population du babaçu. Dans le contexte de la coviabilité, l'explicitation des connaissances relatives aux usages de cette espèce révélera la taxonomie des acteurs et produits dérivés et nous amènera à une première réflexion et aux questionnements relatifs à la durabilité de sa gestion. Les divers verrous seront ensuite dégagés après la présentation de l'articulation et la chronologie des travaux effectués au cours des trois années.

Sommaire

5.1	Problématique Open Science et thèse	33
5.1.1	Open Science	33
5.1.2	Thèse	35
5.2	Questionnements et verrous	36
5.3	Chronologie	37

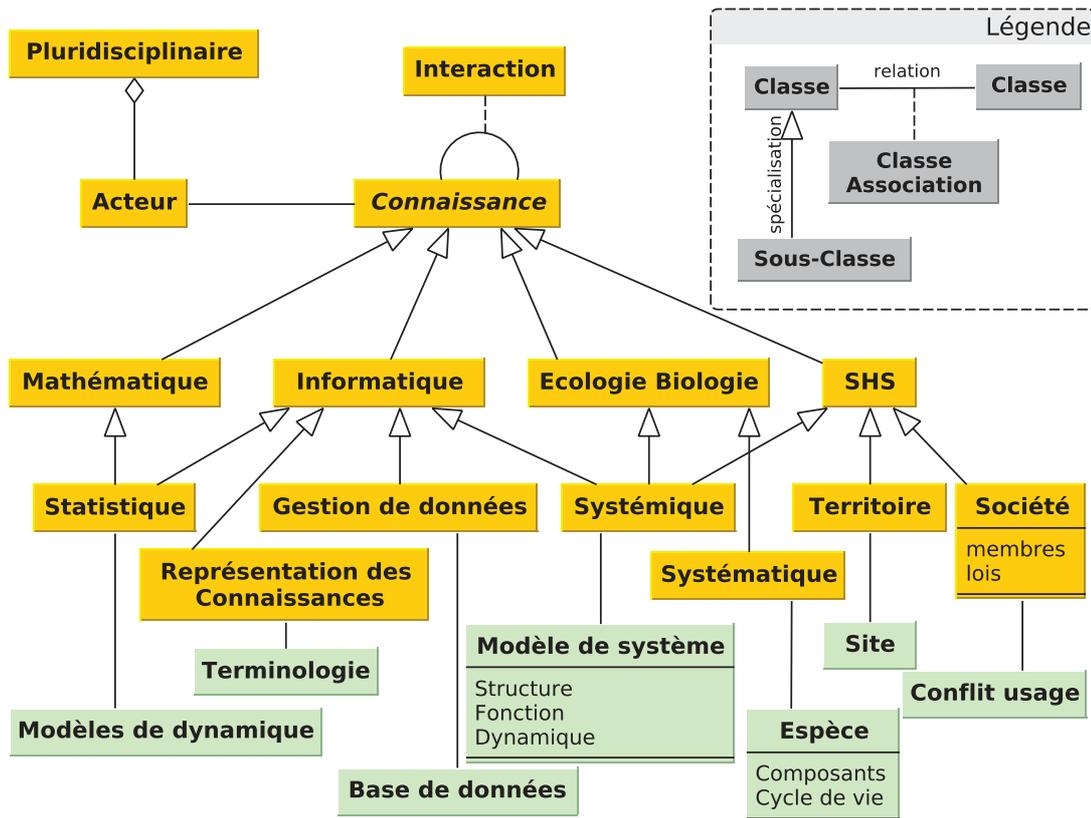
5.1 Problématique Open Science et thèse

5.1.1 Open Science

Le cœur du projet Open Science repose sur l'espèce native babaçu : une attention particulière doit être apportée aux liens entre les acteurs en prise directe avec celle-ci en vue d'orienter les politiques publiques vers une gestion durable. Le travail sera réalisé sur un site déjà connu de nos équipes lors de travaux antérieurs (projet CNPQ/IRD 2001-2006). Il s'agit de celui de la Communauté de Benfica située sur un front pionnier amazonien dans la région de Marabá, Sud-Est de l'État

du Pará, Brésil. Le projet met en œuvre des recherches diverses effectuées dans ce travail et des travaux de plusieurs masters. Les questions abordées se rapportent à des connaissances spatio-temporelles relatives à la zone d'étude (approche terrain et observation satellitale), à la biologie et écologie de l'espèce étudiée, ainsi qu'à l'économie liée aux pratiques des populations sur cette espèce.

FIGURE 5.1 – Interdisciplinarité et connaissance (formalisme UML)

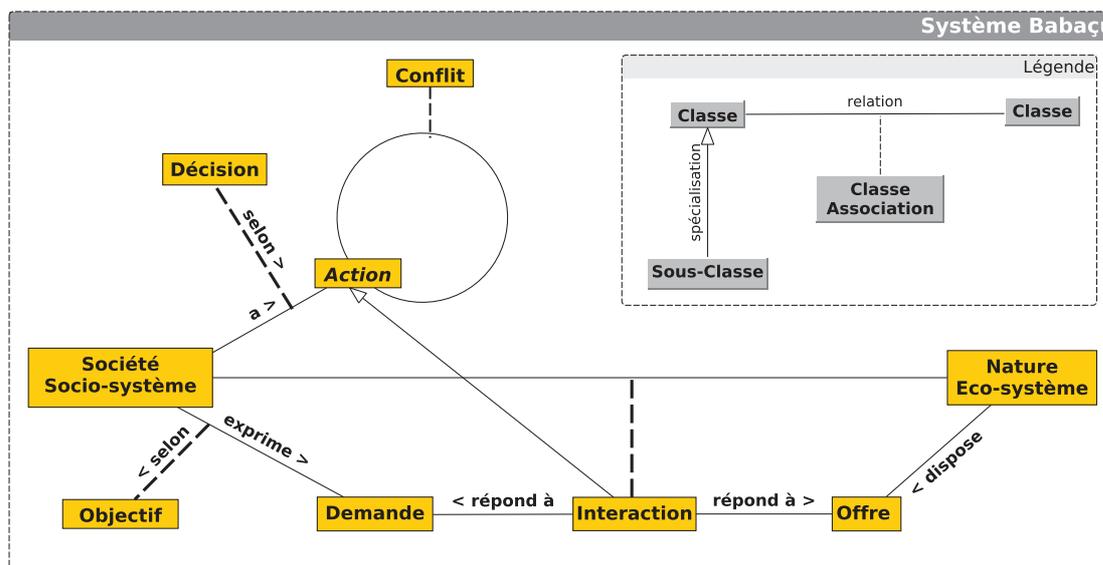


La diversité des problèmes autant biologiques qu'anthropiques soulevés autour du palmier babaçu nécessitait organisation et rigueur, d'où la création d'un groupe de réflexion autour de ce projet.

Ce groupe pluridisciplinaire inclut aussi bien des partenaires issus des Sciences Humaines et Sociales que des partenaires issus des Sciences dites « exactes » (cf. FIGURE 5.1). Les acteurs ont des connaissances propres qui proviennent de différentes disciplines (Mathématique, Informatique, Écologie/Biologie et SHS), l'enrichissement provient de l'interaction entre celles-ci. Les concepts et méthodes de chacune de ces disciplines sont ensuite utilisés et appliqués dans un domaine analytique. Par exemple la création et gestion d'une base de données fera appel à des méthodes et outils de gestion de données provenant du domaine de l'Informatique, alors que la

construction du modèle de dynamique de population sera basé sur des méthodes statistiques demandant des connaissances des domaines Mathématique et Informatique.

FIGURE 5.2 – Vision intégrée du système babaçu (formalisme UML, adapté de Simon *et al.* (2014))



Le système babaçu est un système complexe constitué de composants « nature » (éco-systèmes) et des composants « société » (socio-systèmes) en interactions. « Les interactions peuvent se résumer au schéma suivant : la société exprime des demandes basées sur les offres potentielles de la nature. Les demandes sociétales correspondent à des objectifs et après décision donnent lieu à des actions qui se concrétisent en interactions. Ceci permet de mettre en évidence le fait que les acteurs, selon leur perception de l'espèce babaçu, engagent des processus, qui sont source de conflits entre eux » (cf. FIGURE 5.2).

Pour finir, à partir de la réflexion menée autour de la coviabilité du système babaçu, nous souhaitons dégager une méthodologie de gestion durable, si possible générique, pour les plantes natives utiles, appartenant à des biomes divers.

5.1.2 Thèse

Le travail de thèse est pionnier. Les contextes biologique et socio-économique étayés dans le précédent chapitre démontrent bien la multiplicité des relations et liens qui sont tissés au sein de ce bio-éco-socio système. Notre recherche s'ancre sur deux points cruciaux, pour cette thèse, relatifs à l'étude du fonctionnement de cette espèce végétale : la connaissance imparfaite de son cycle de vie et l'absence

de modélisation de sa dynamique de population. Malgré la multitude d’usages de la ressource, le cycle de vie du babaçu reste peu connu, bien qu’il existe d’innombrables études menées au niveau de la biologie du babaçu. De même, à notre connaissance, aucun modèle de dynamique de population de cette espèce n’a été proposé. Il existe des travaux sur la structure de la population, mais il n’existe pas d’étude portant sur sa dynamique de population.

Pour aider à la gestion durable du babaçu, il existe une demande croissante, de la part des différents acteurs, de mieux connaître son fonctionnement au sein des espaces anthropisés. Les *Quebradeiras*, notamment, ont besoin de pouvoir estimer le niveau de disponibilité de la ressource afin de pouvoir assurer le maintien en équilibre de cette espèce à long terme.

5.2 Questionnements et verrous

Nous pouvons donc formuler une suite d’interrogations relatives toutes à l’évolution de cette espèce. Quel est le comportement de la plante sur le long terme ? Quelle est sa dynamique ? Dans quelle proportion cette dynamique dépend-elle du milieu environnant et notamment de l’interaction avec la société ? En effet, deux processus contradictoires peuvent en effet co-exister : d’une part le processus invasif (l’espèce reprend ses « droits » dans les pâturages) et d’autre part le processus de disparition (l’espèce s’éteint par prélèvement intensif et/ou par substitution de cultures). L’éco-socio-système babaçu subit aussi l’influence de systèmes plus larges. Dans ce contexte, est-il possible d’adapter la gestion de l’espèce pour les populations locales à travers le prélèvement de fruits afin qu’elle soit durable ? Quelle quantité de fruits pouvons-nous prélever au maximum ? Quel est donc le seuil de non reproduction ?

Les verrous auxquels nous allons tenter de répondre peuvent se résumer comme suit :

- ceux relatifs à l’explicitation des connaissances. En effet, l’approche pluridisciplinaire demande un travail d’harmonisation et d’explicitation de connaissances communes.
- ceux relatifs aux données et observations. Il n’existait pas de données ni de protocole d’observation adéquats sur la zone concernée. La prise en compte de la spatialisation de ces données est obligatoire ; leur conservation est aussi nécessaire pour des analyses ultérieures.
- ceux relatifs à la modélisation mathématique. Plusieurs modèles potentiels existent et il nous faut rapidement analyser les données recueillies afin d’effectuer le meilleur choix possible. La confrontation avec les données est obligatoire pour la validation des hypothèses formulées.

Les défis que nous nous lançons doivent répondre à ces verrous. De plus, en ce qui concerne le modèle de dynamique, une difficulté complémentaire a fait surface : en effet nous ne disposons pas de données au démarrage des recherches et le protocole de collecte n’a pu être conçu et amendé qu’en partenariat avec nos collègues brésiliens.

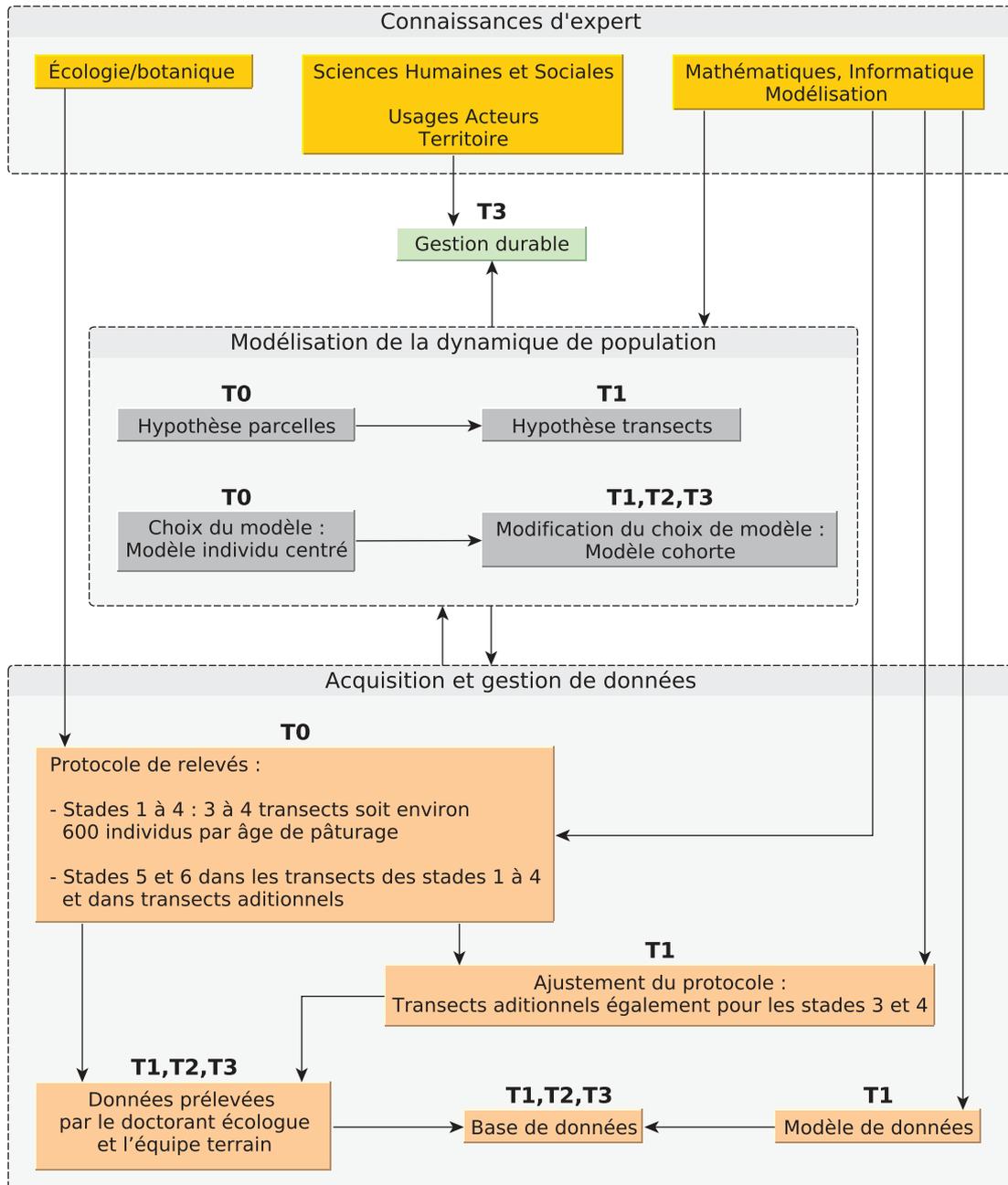
Une difficulté de cette recherche réside dans son côté pluridisciplinaire. La première contrainte était de construire une connaissance partagée sur le système babaçu.

5.3 Chronologie

La chronologie du travail effectué est présentée dans la FIGURE 5.3. La notation T0 correspond aux travaux amorcés durant l'année du master et les notations T1, T2 et T3 aux travaux des trois années de la thèse proprement dite.

L'approche retenue a consisté, à expliciter au sein de divers modèles, les connaissances d'experts partagées sur l'espèce, les modèles mathématiques potentiels et le contexte sociétal (usage-acteurs et territoires). Ceci, nous a permis de proposer un modèle d'observation conforme aux préconisations d'OBOE (cf. Chapitre 6 - Données et connaissances) permettant la mise en place du protocole d'acquisition des données. L'acquisition *in situ* a été réalisée par le partenariat franco-brésilien entre 2013 et 2016 au sein du territoire de la communauté de Benfica (Pará, Brésil). Pour cela, deux organismes gouvernementaux brésiliens offrant des bourses (la *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes)* et le *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)*), ont financé le doctorat d'Alessio Moreira dos Santos qui a débuté dès 2013 à l'UFRA. La pérennisation des données s'est faite après conception, au sein d'une base de données spatiales. La modélisation de la dynamique repose sur une approche statistique et le modèle résultant sera validé par confrontation avec les données de terrain.

FIGURE 5.3 – Déroulement des activités du projet Open Science de l'Agropolis Fondation pour l'acquisition des données de terrain en écologie et la modélisation de la dynamique de population (la temporalité est représentée par « T0 » à « T3 »)



Troisième partie

État de l'art

6	Données et connaissances	41
7	Dynamique de population	45

Données et connaissances

Sommaire

6.1	Capitalisation de connaissances et de données	41
6.2	Importance de la terminologie	42
6.3	Ontologie en modélisation mathématique	44

Les questionnements, nous l'avons vu précédemment portent sur la capitalisation des connaissances et des données ainsi que sur la modélisation de la dynamique de l'espèce babaçu.

6.1 Capitalisation de connaissances et de données

En ce qui concerne la capitalisation des connaissances, il est clair qu'au-delà de la thèse, si nous souhaitons recourir à ce que l'on désigne, dans le secteur de la santé¹, par « transfert et échanges de connaissances » (Choi, 2005, « *Knowledge translation* ») c'est-à-dire au processus complet d'échanges entre chercheurs et politiques décideurs, et ce à des fins d'amélioration des prises de décision, nous devons mettre en œuvre une méthodologie d'acquisition et de diffusion de celles-ci.

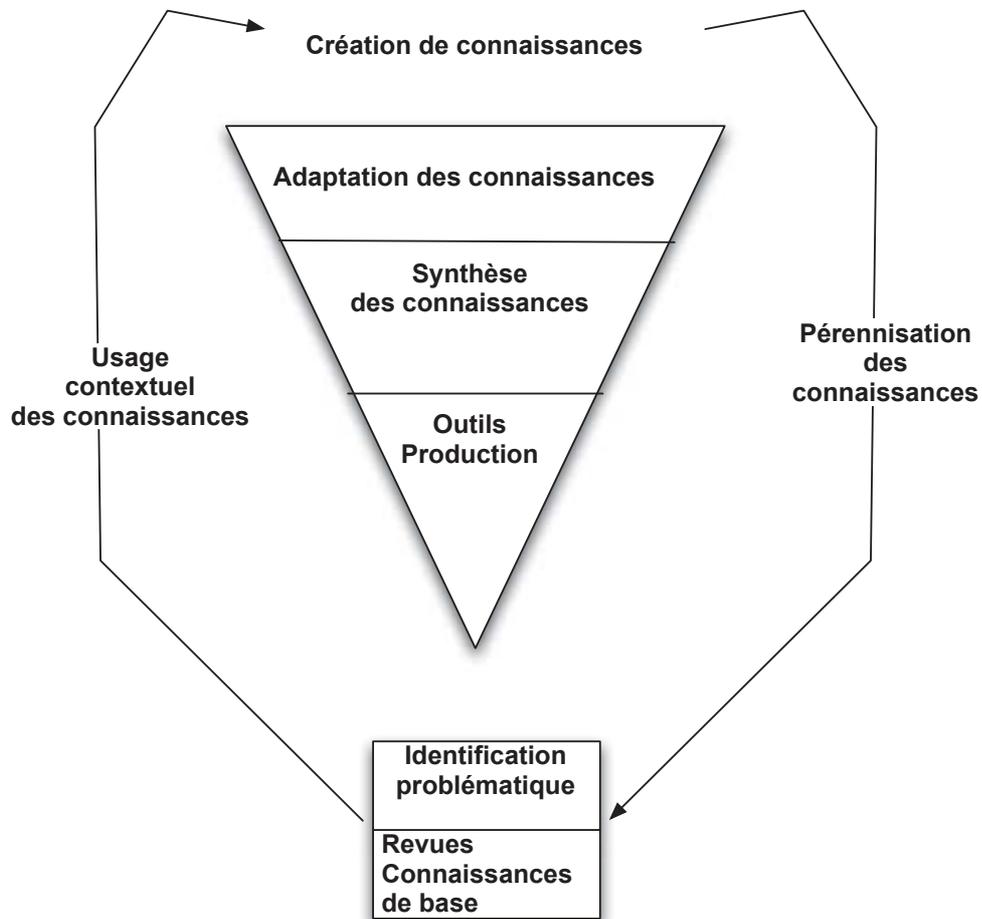
Le cycle d'actions liées à cette approche peut être résumé dans la FIGURE 6.1.

Dans tout projet de recherche, après identification de la problématique effectuée à partir des connaissances de base et de la littérature, le cycle d'actions se focalise sur la création de connaissances (celles-ci sont adaptées au contexte d'usage). La création de connaissances s'appuie sur l'échange entre scientifiques et le plus souvent aussi sur les données associées ou collectées au sein du projet. Les outils de production mis en œuvre les synthétisent et le cycle passe par la phase de pérennisation des connaissances produites (nécessitant leur appropriation notamment par les acteurs partenaires et décideurs).

Compte-tenu de la diversité des domaines scientifiques en présence, bien que des méthodes issues de l'Intelligence Artificielle existent et aient été largement utilisées (Charlet *et al.*, 2000), nous avons opté pour une approche plus simple, d'explication de connaissances, incrémentale, basée sur l'utilisation du formalisme UML (Rumbaugh *et al.*, 2004). Ce formalisme a été choisi, entre autres, pour son aspect graphique qui permet de manipuler et représenter aisément les concepts partagés (classes, relations, objets notamment).

1. Qui peut être largement évoqué dans de nombreux domaines.

FIGURE 6.1 – Cycle d’actions relatif au transfert et échanges de connaissances (d’après Grimshaw *et al.* (2012))



6.2 Importance de la terminologie

Comme souligné dans les divers chapitres précédents, la notion de « système babaçu » a émergé très rapidement au cours des discussions de notre groupe de travail. Nous avons alors entrepris une étude systémique : le formalisme UML bien qu’issu d’une approche objet nous a rendu celle-ci plus aisée. Nous souhaitons organiser notre analyse autour de plusieurs concepts : modèle, système, structure, dynamique... Cependant, l’utilisation de différents termes et notions au sein du groupe devait être aussi claire que possible et surtout ne pas prêter à confusion. Dès lors, il a été impératif de créer un référentiel terminologique et de faire appel aux outils et méthodes utilisés en matière de représentation des connaissances. Cet exercice nécessite une réflexion préalable autour de la définition des mots « concept », « notion » et « terme ». Comment peut-on les différencier, alors que leurs définitions divergent selon les disciplines et les théories ? Lequel des trois est le plus générique ?

Dans quel cas, est-il préférable d'utiliser l'un et dans quel autre cas un autre ?

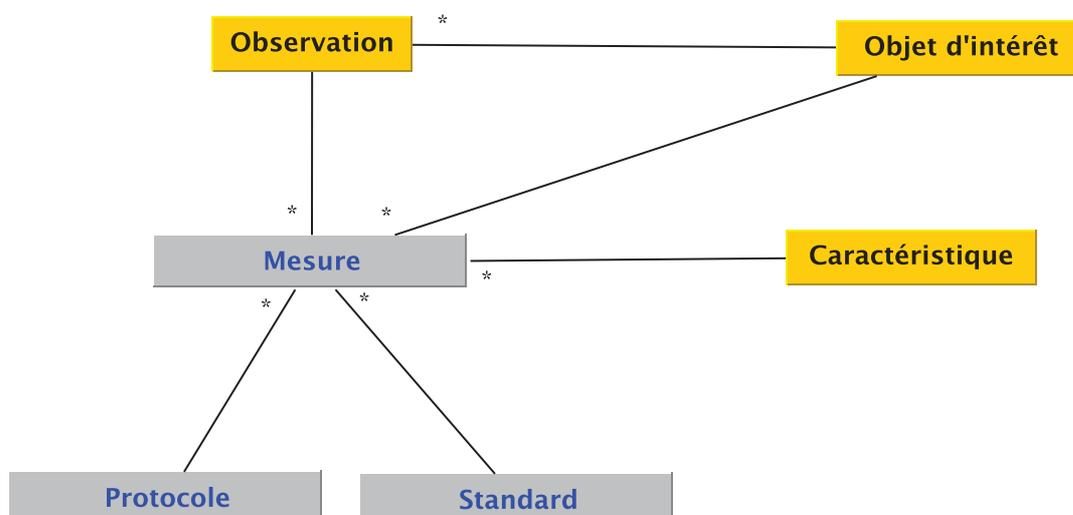
La définition proposée par Depecker et Roche (2007) et basée sur le travail du Comité technique 37 de l'ISO (élaboration et normalisation en matière de terminologie) nous paraît la plus juste. Ainsi le « concept » succède à la « notion ». Un concept exprime un niveau générique d'abstraction, il ne dépend pas du langage, mais est assujéti aux pratiques socio-culturelles qui vont l'ériger.

Les auteurs soulignent également que la terminologie et les ontologies sont deux champs disciplinaires qui se ressemblent de par leur « approche logique, au centre de laquelle figure la problématique du concept » (Depecker et Roche, 2007, p. 113).

Nous pouvons donc qualifier notre approche d'ontologique. Si nous nous cantonnons à la définition donnée par (Gruber, 1995) l'ontologie est une spécification d'une conceptualisation². Les concepts de base du « système babaçu » vont de manière naturelle (car facilement appropriables) être décrits au travers de divers modèles descriptifs, via le formalisme UML dans les diagrammes structurels (classes, relations objets) ainsi que dans les diagrammes d'états (état, transition) .

De nombreux chercheurs ont aussi opté pour cette approche dans les disciplines relevant de l'écologie et des sciences humaines et sociales (Pages *et al.*, 2000; Pointet, 2007; Sibertin-Blanc *et al.*, 2011; Mazzega *et al.*, 2013).

FIGURE 6.2 – L'ontologie OBOE (formalisme UML)



Nous pouvons aussi relever que, par exemple, la proposition de Madin *et al.* (2007) (cf. Figure 6.2) a été directement réutilisée lors de la définition du protocole de collecte des données de terrain qui a précédé la modélisation conceptuelle de la base.

2. Cependant nous reviendrons sur cela au niveau de nos perspectives.

La sémantique de cette proposition se veut générique et relative à toute observation scientifique et mesures associées à un objet d'intérêt et à une ou plusieurs de ses caractéristiques. Le modèle peut être utilisé pour caractériser le contexte d'une observation (par exemple, l'espace et le temps), clarifier le protocole et la description des unités de mesure utilisées.

La dimension spatiale a ainsi pu être introduite aisément dans le contexte de la modélisation conceptuelle de la base de données réceptacle des données (cf. Chapitre 9 - Acquisition et pérennisation des données). En ce qui concerne l'existant relatif aux systèmes de gestion des données spatiales, à l'heure actuelle, nombreux sont les systèmes obéissant aux standards établis par l'*Open Geospatial Consortium* (OGC) (logiciels propriétaires ou Open Source).

6.3 Ontologie en modélisation mathématique

En ce qui concerne les modèles mathématiques, depuis le milieu des années 2000, l'explicitation des connaissances a abouti à une approche basée sur les ontologies. Les travaux dans ce sens se sont intensifiés depuis les années 2010. Des ontologies génériques relatives à la classification des modèles mathématiques ont vu le jour : *Mathematical Modelling Ontology* (MAMO) (Zhukova *et al.*, 2014).

Des ontologies plus spécifiques de type applicative ont également été introduites, se référant à la gestion d'un modèle (définition des fonctions, équations, variables...) telle que : *Ontological mathematical modeling knowledge management* (OntoMODEL) dédiée à la gestion de modèles pour le développement de produits pharmaceutiques (Suresh *et al.*, 2008). La modélisation mathématique est un élément essentiel de la démarche scientifique mais elle doit aussi, nous semble-t-il, s'ouvrir à l'explicitation des éléments constitutifs des modèles utilisés afin de faciliter leur compréhension.

Dynamique de population

CE CHAPITRE dresse, tout d'abord l'état de l'art en matière d'étude de dynamique de population des espèces végétales. Dans un second temps, il se centre sur des recherches existantes en matière de dynamique de population du palmier babaçu.

Sommaire

7.1	Dynamique de population des espèces végétales	45
7.1.1	La démarche générale	45
7.1.2	A propos des palmiers	46
7.2	Recherches antérieures sur la dynamique de population du babaçu	49
7.2.1	Processus de croissance du babaçu	49
7.2.2	Le recrutement	51

7.1 Dynamique de population des espèces végétales

Nous aborderons le thème de dynamique de population des végétaux par des aspects plus généraux s'appliquant à toutes les espèces. Puis, nous indiquerons une liste non exhaustive des démarches appliquées en matière de dynamique de population des palmiers.

7.1.1 La démarche générale

Une population végétale peut être caractérisée par ses structures. Nous parlons de structures au pluriel, puisqu'une population peut potentiellement être ordonnée selon différentes variables la caractérisant. Par exemple, elle peut être structurée en fonction de ces caractéristiques génétiques, de sa localisation spatiale, de sa taille ou de l'âge des individus (Silvertown et Lovett Doust, 1993). Le champ disciplinaire chargé de l'étude de cette population sera donc défini en fonction des caractéristiques qui rentrent en compte pour sa structuration. Par exemple la caractérisation des changements génétiques au sein d'une population fera appel à des techniques de l'écologie de l'évolution. Alors que les changements dans le nombre d'individus se

référeront plutôt à l'étude de la dynamique de population.

Pour cette recherche, nous nous intéresserons à l'étude de l'évolution du nombre d'individus dans le temps : la dynamique temporelle de la population. Nous n'aborderons pas ici les préoccupations concernant les changements génétiques. Ces problèmes relevant plutôt du champs disciplinaire évolutionniste, nous les laisserons de côté.

Toute étude de dynamique de population se base sur l'observation et l'analyse de ce qui est communément appelé « paramètres démographiques d'une population végétale » (Skalski *et al.*, 2005). Ces paramètres démographiques sont construits pour chaque population à partir du cycle de vie de cette même population. Dans le cas le plus générique nous nous intéressons aux deux paramètres suivants :

- le nombre de nouveaux individus (R) entre deux relevés consécutifs,
- le nombre d'individus morts (M) entre deux relevés consécutifs.

$$N(t + 1) = N(t) + R - M \quad (7.1)$$

Ainsi, comme le démontre l'équation 7.1, le nombre d'individus au temps $t + 1$ ($N(t + 1)$) est égal au nombre d'individus au temps t ($N(t)$) augmenté de nouveaux individus et diminué des individus morts entre t et $t + 1$.

Le ratio $N(t + 1)/N(t)$ est aussi appelé taux annuel d'accroissement (noté généralement par λ). La valeur de ce ratio peut être située dans un des trois cas de figures possibles (Crawley, 2005) :

1. $\lambda > 1$ - la population croît ($R > M$)
2. $\lambda = 1$ - la population reste stable ($R = M$)
3. $\lambda < 1$ - la population décroît ($R < M$)

7.1.2 A propos des palmiers

Chez les *Arecaceae* la dynamique de population est étudiée, comme chez d'autres espèces végétales, grâce à l'étude de deux processus biologiques majeurs, la croissance des palmiers depuis les plantules jusqu'aux palmiers adultes et le recrutement qui inclut tous les sous-processus biologiques allant de la floraison à la germination des graines.

La croissance est un processus central dans l'étude de la dynamique de population des palmiers. Plusieurs méthodes ont été appliquées dans des études différentes.

Guarín et del Valle (2014) proposent par exemple de déduire l'âge du stipe du palmier *Oenocarpus bataua* à partir d'une combinaison entre l'estimation de la croissance annuelle de celui-ci et l'estimation du nombre moyen de feuilles produites par le palmier durant la période de référence (une année). Cette méthode permet l'estimation de l'âge des individus.

Pinero *et al.* (1984) ont exploité auparavant une approche semblable à celle de Guarín et del Valle (2014) mais ils se sont efforcés de calculer la croissance du palmier *Astrocaryum mexicanum* à partir de la formation de feuilles pour une partie de la population (plantules et une partie des jeunes) et à partir de l'augmentation de la taille pour le reste de la population (partie complémentaire de jeunes et les adultes). Le recrutement est un processus souvent complexe chez les palmiers dont la modélisation peut être traitée de différentes manières. Pinero *et al.* (1984) ont répertorié toutes les infrutescences et les fruits sur les infrutescences pour tous les individus adultes dans le but d'en déduire le taux de fécondité comme un produit de la probabilité de reproduction et le nombre de fruits par individu capable de se reproduire.

Une autre méthode a été choisie par Bullock (1980) lors de sa modélisation de la dynamique de population du palmier *Podococcus barteri* au Cameroun. L'auteur stipule la séparation des adultes en plusieurs stades en fonction de leur comportement reproductif. Pour faire ceci il a étudié la dynamique de population de ce palmier en s'appuyant sur un modèle matriciel de transition avec une séparation de la population en stades basé sur les caractéristiques biologiques de l'espèce. Cette méthode implique d'exactes connaissances *a priori* sur le mécanisme du comportement reproductif du palmier. C'est pourquoi, l'auteur insiste sur la nécessité d'une étude plus approfondie centrée sur la pollinisation et le fleurissement permettant d'aboutir à des estimations des inflorescences mâles et femelles par individu adulte.

Une autre étude menée par Pinard (1993) relative à la modélisation de la dynamique de population d'*Iriarteia deltoidea* (Arecaceae) sous l'influence de la récolte des tiges au Brésil, stipule également la séparation des palmiers adultes (capables d'assurer la reproduction) en plusieurs stades. Un cadre rigoureux du calcul du nombre de plantules produites par palmier adulte et par classe (F_i) est fourni par l'auteur :

$$F_i = (N_{ir}/N_r) \cdot N_0/N_i \quad (7.2)$$

N_{ir} est le nombre de palmiers de la classe i capables de se reproduire. N_r est le nombre total de palmiers capables de se reproduire, toutes classes confondues. N_i est le nombre total de palmiers dans la classe i . N_0 est le nombre total de plantules recrutées au stade 1 durant chaque intervalle de temps. Pinard (1993) souligne cependant que la validité de cette construction se base sur deux conditions majeures :

- la proportion de palmiers capables de se reproduire au moment de la collecte de données sur le terrain reste un indicateur valide qu'un palmier sera désormais capable de se reproduire,
- le nombre de plantules est relatif réellement aux palmiers adultes capables de se reproduire situés au sein de la zone d'étude.

Une approche différente a été réalisée par Holm *et al.* (2008) lors de leur étude de la fécondité du palmier *Mauritia flexuosa* en Amazonie. Ils stipulent que l'estimation du paramètre de fécondité (f_{0i}) peut se faire de la manière suivante :

$$f_{0i} = (p_0 \cdot n_0(t+1)/2)/N_i \quad (7.3)$$

où f_{0i} est le nombre annuel de plantules par palmier femelle adulte pour le stade i , p_0 est la probabilité de survie des plantules, $n_0(t + 1)$ est le nombre moyen de plantules établies durant l'intervalle de temps suivant. Enfin, N_i est le nombre de palmiers femelles capables de se reproduire dans le stade i .

Une recherche a été menée également au niveau de la dynamique de population du palmier *Rhopalostylis sapida* (Wendl. et Drude) en Nouvelle-Zélande (Enright et Watson, 1992). Les auteurs présentent une approche déterministe basée sur les matrices de transition et la caractérisation de la population en stades distingués à partir de la taille des palmiers. Le processus de croissance est alors modélisé en tant que passage d'un stade biologique à un autre. Ils modélisent le processus de reproduction d'une manière plus « classique » : en intégrant directement au sein de la matrice de transition la valeur moyenne du nombre de plantules par palmier adulte par an.

Olmsted et Alvarez-Buylla (1995) proposent une solution similaire quant à la modélisation du recrutement chez deux espèces de palmiers (*Thrinax radiata* et *Coccothrinax readii*) de la forêt tropicale sèche au Mexique. Les auteurs proposent de pondérer le nombre de fruits par stade par la probabilité de germination et d'utiliser cette valeur directement en entrée au sein de la matrice de transition. De plus, la caractérisation de la population en stades permet de définir les probabilités de passage d'un stade à un autre et celles de rester dans le même stade durant la période temporelle (un an). Ces stades sont définis pour les palmiers sans stipe en fonction de la présence/absence de feuilles et de la taille de celles-là si elles sont présentes. Quant aux palmiers ayant un stipe, ils sont répartis dans les différents stades en fonction de la taille du stipe du palmier. Les caractéristiques biologiques des deux espèces ne semblent pas avoir une influence prédominante sur le choix des stades et la définition de leur nombre si ce n'est que pour allouer chaque stade au sein d'un des trois groupes plus génériques : plantules, jeunes et adultes.

Cette méthode a été également utilisée par Barot *et al.* (2000) au sein de leur modèle de dynamique de population du palmier *Borassus aethiopum* en Côte d'Ivoire.

Comme nous l'avons vu il existe de nombreuses études portant sur la dynamique de population des palmiers dont nous venons de fournir ci-dessus un aperçu non exhaustif. Les populations étudiées sont systématiquement divisées en stades dont la définition est basée sur des caractéristiques comme la biologie, la taille et/ou le nombre de feuilles de chaque palmier. Les approches d'estimation des taux de passage et de recrutement choisies par les auteurs se basent sur ces stades qui ont été définis en fonction des choix faits par chacun des auteurs. Cependant, la plupart des travaux cités ci-dessus puisent leur source le plus souvent dans l'ouvrage de référence en matière de modélisation matricielle écrit par Caswell (2001) et appliquent donc des méthodes de modélisation basées sur ce type de modèles.

7.2 Recherches antérieures sur la dynamique de population du babaçu

De nombreux travaux portent sur l'étude de différentes composantes ou dérivées du babaçu utilisées comme des matières premières destinées aux industries. Nous pouvons en citer quelques-unes :

- utilisation de l'huile du babaçu pour les biocarburants (Teixeira, 2008; Teixeira *et al.*, 2011),
- utilisation du charbon obtenu à partir des noix entières du babaçu pour la fabrication d'acier (Emmerich et Luengo, 1994, 1996; Protásio *et al.*, 2014),
- utilisation de l'amidon des noix pour la fabrication d'éthanol (Baruque Filho *et al.*, 1998),
- ...

Cette liste non exhaustive peut être très longue. Nous en fournissons un simple aperçu, mais le plus important est qu'à notre connaissance et à ce jour aucun travail de recherche valorisé par une publication scientifique n'a été fourni en matière de proposition d'un modèle de dynamique de population du palmier babaçu. Nous sommes en mesure de citer uniquement le travail de recherche en master de Ferreira (1999) relatif à la modélisation de l'évolution de l'abondance du babaçu au sein de la forêt tropicale humide.

Dans cette section nous présentons les recherches relatives aux processus de croissance et de recrutement du palmier babaçu. Anderson (1983) a effectué une étude sur le cycle de vie du palmier babaçu dont nous présenterons les aspects en lien avec nos travaux. Nous exposerons également les travaux de Barot *et al.* (2005) relatifs à une première approche dans l'étude de la structure de population du babaçu.

7.2.1 Processus de croissance du babaçu

Anderson (1983) propose une structuration des palmiers babaçu à l'aide de douze stades de développement biologique. La définition de ces stades est résumée dans le TABLEAU 7.1. La distinction au niveau des stades 2 à 4 se base sur le degré de développement (division) des feuilles. Pour les autres stades la distinction est basée uniquement sur la hauteur du palmier.

L'auteur propose une étude de la dynamique de population du palmier babaçu basée sur la définition des stades du palmier du TABLEAU 7.1. Le terrain d'étude du travail de recherche d'Anderson (1983) se situe sur le site du Lago Verde dans l'État brésilien du Maranhão (cf. FIGURE 7.1). Dans cette étude, l'auteur livre une analyse comparative de la structure et de la dynamique de population du babaçu en examinant trois occupations du sol bien distinctes : la forêt tropicale humide, les

1. Stade incluant des palmiers qui n'ont pas encore fleuri et ceux qui ne sont pas matures de point de vue de la reproduction.

TABLEAU 7.1 – Définition des stades d'(Anderson, 1983, page 117)

Stade		Début
Description	Numéro	
Fruit (noix)	1	Abscission du fruit
Plantule	2	Apparition de la première feuille
Juvéniles pré-installés	3	Première division du limbe des feuilles
Juvéniles installés	4	Folioles > 100 par côté
	5	Hauteur ≥ 100 cm
	6	Hauteur ≥ 200 cm
Palmiers matures	7 ¹	Hauteur ≥ 500 cm
	8	Hauteur ≥ 1000 cm
	9	Hauteur ≥ 1500 cm
	10	Hauteur ≥ 2000 cm
	11	Hauteur ≥ 2500 cm
	12	Hauteur ≥ 3000 cm

pâturages et la forêt secondaire. Ses résultats relatifs à la dynamique de population du babaçu dans les pâturages permettent de déduire des taux empiriques de mortalité et de croissance par stade (cf. TABLEAU 7.2). Nous remarquons que les taux empiriques de croissance pour les stades 2 à 4 sont extrêmement faibles ($< 5\%$). Les taux empiriques de mortalité pour les stades 2 à 6 sont supérieurs à 10%.

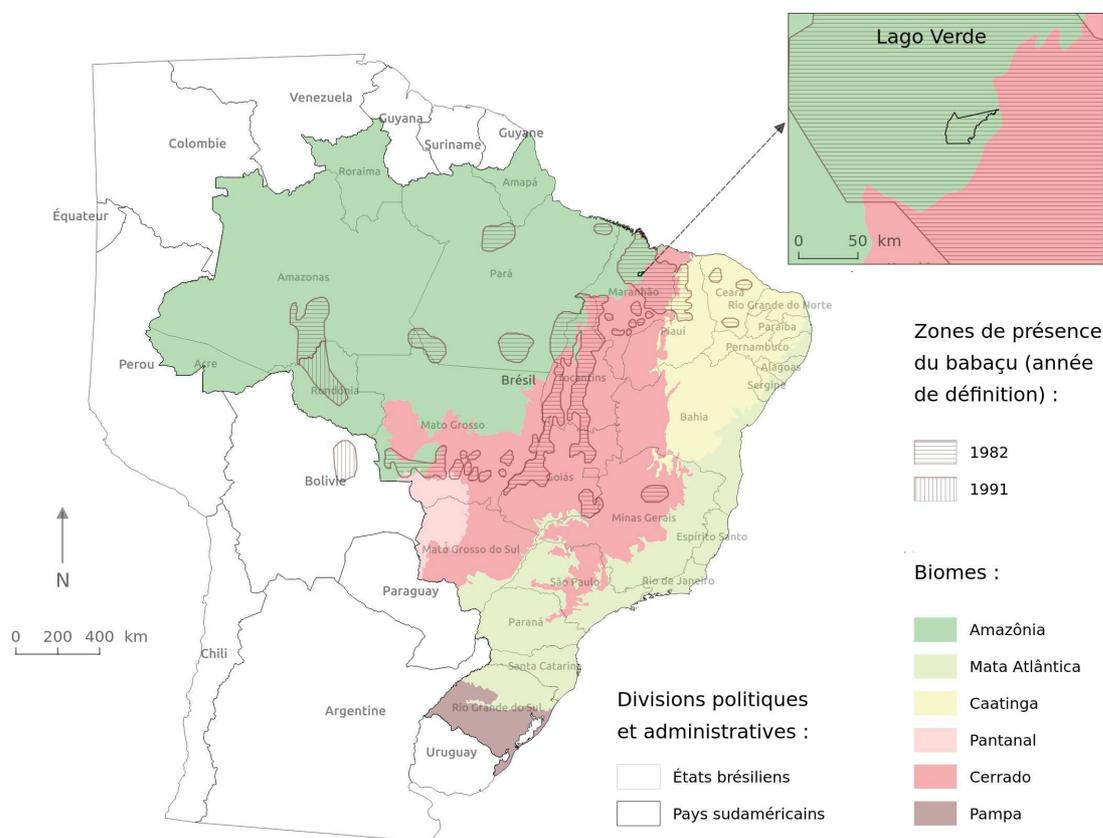
TABLEAU 7.2 – Taux empiriques d'(Anderson, 1983, page 147)

Stade		Taux de croissance (%)	Taux de mortalité (%)
Description	Numéro		
Fruit (noix)	1	-	-
Plantule	2	1,6	38,4
Juvéniles pré-installés	3	0,7	11,6
Juvéniles installés	4	2,7	10,8
	5	15,6	15,6
	6	13,3	20
Palmiers matures	7	18	0
	8	0	0
	9	0	0
	10	-	-
	11	-	-
	12	-	-

Cependant, les chiffres bruts annoncés par Anderson sont à relativiser puisqu'il s'agit d'extrapolations : le dénombrement des palmiers appartenant à des stades différents a été fait sur des superficies de taille différente (corrélation positive entre

superficie et stade). Ces chiffres sont basés sur trois mesures prises à deux intervalles de temps de 6 mois entre 1981 et 1982.

FIGURE 7.1 – La municipalité de Lago Verde au Brésil



N. Sirakov. 2016

Source : IBGE, May *et al.* (1985)

7.2.2 Le recrutement

Le babaçu est une plante pérenne dont le cycle de floraison/fructification est annuel. La production de fruits s'étale sur neuf mois, de juin à février ; ils sont gros et ont un poids variable selon les auteurs et les régions. D'après Barot *et al.* (2005) les fruits pèsent entre 200 g et 350 g dans le Pará ; d'après Anderson *et al.* (1991), leur poids est en moyenne 250 g dans le Maranhão. Les fruits entament leur phase de germination à partir du mois de janvier, durant la saison des pluies (Anderson, 1983; Anderson *et al.*, 1991). Cependant, Anderson *et al.* (1991) souligne qu'au sein de la forêt tropicale humide et des babaçuais, le babaçu peut passer par une longue période de latence entre la germination des fruits et le développement du stipe.

L'auteur annonce une moyenne de 38 ans pour la forêt tropicale et de 29 ans pour le babaçal.

Le palmier babaçu se caractérise, comme tous les palmiers, par sa longévité. D'après les estimations effectuées par Anderson (1983), un palmier adulte peut atteindre l'âge de 185 ans dans des conditions « naturelles » au sein de la forêt tropicale humide.

Barot *et al.* (2005) ont constaté que dans les pâturages et les babaçals les inflorescences masculines sont plus régulièrement observables chez les palmiers adultes de petite taille, contrairement à ce qui est constaté en forêt. Les auteurs proposent trois interprétations possibles :

- l'hypothèse que les inflorescences féminines sont plus « coûteuses » que les inflorescences masculines,
- le palmier fait l'allocation des ressources pour les inflorescences mâles ou femelles suivant le modèle des courbes d'aptitude-gain.
- l'existence d'une relation d'environnement-dépendance dans l'allocation du sexe de l'inflorescence par le palmier adulte basée sur la taille de celui-ci.

Quatrième partie

Proposition

8	Introduction	55
9	Acquisition et pérennisation des données	59
10	Modélisation mathématique	81

CHAPITRE 8

Introduction

LA méthodologie adoptée relève d'une démarche itérative mixant induction et déduction et ceci compte tenu du contexte sciences expérimentales. Par rapport à la question initiale posée dans le projet Open Science « la gestion durable du babaçu est-elle possible ? » notre réflexion s'articulera autour de plusieurs modèles issus d'interactions entre connaissances disciplinaires diverses. Le présent chapitre sera l'objet d'une introduction précisant ce que nous entendons par modélisation et la logique méthodologique mise en place. Les deux chapitres suivants seront dédiés à l'acquisition et à la pérennisation des données, puis à la modélisation de la dynamique de population du babaçu.

Sommaire

8.1	Autour de la modélisation	55
8.1.1	Modélisation : un processus d'itération	55
8.1.2	Classification des modèles	56
8.2	Méthodologie générale	56

8.1 Autour de la modélisation

Le concept de modélisation est utilisé dans pratiquement toutes les disciplines et en conséquence peut avoir autant de définitions que de champs d'applications. C'est pour cette raison que nous souhaitons le définir afin d'éviter toute confusion possible par la suite.

8.1.1 Modélisation : un processus d'itération

Quelque soit le champ d'application, le scientifique est confronté à l'obtention d'un modèle basé sur une hypothèse initiale. En reprenant les dires de Minsky (1965), nous pouvons admettre que pour un observateur donné, un objet « A^* » est un modèle de l'objet « A » dans la mesure où l'observateur peut utiliser « A^* » pour répondre aux questions qui l'intéressent sur « A ». Ceci est à rapprocher de la définition du modèle proposée par Nouvel et Lecourt : le modèle est une « formalisation qui permet un traitement théorique du phénomène étudié » (Nouvel et Lecourt, 2002, p. 192). L'obtention du modèle correspond à un exercice de traduction du réel selon un formalisme adapté. Ce modèle se base sur la réalité afin de la retranscrire tout

en la simplifiant (du réel vers le virtuel). Comme le soulignent [Nouvel et Lecourt \(2002\)](#) la quête du scientifique lors de la définition du modèle, est une simplification de la réalité, mais n'est pas forcément une réduction de celle-ci. De plus, [Brunet](#) souligne qu'un modèle est non seulement la simplification de la réalité, mais reflète surtout la vision que l'on a de cette réalité ([Brunet, 1980](#), p. 254). Toute la difficulté et l'ambiguïté de la démarche modélisatrice réside dans le fait qu'elle est fortement liée au relativisme sémantique du scientifique.

Pour le scientifique le modèle et l'hypothèse initiale doivent être confrontés à la réalité dans le but de les valider, les améliorer ou les rejeter (du virtuel vers le réel). La robustesse et la validité d'un modèle ne sont donc pas jugées uniquement en fonction de ses composantes (variables, paramètres...), mais plutôt en fonction de la force avec laquelle le modèle arrive à donner des prédictions justes, au plus proche de la réalité ([Badiou, 2007](#)). La construction du modèle et sa confrontation à la réalité impliquent un mécanisme itératif que nous allons retrouver dans la démarche adoptée.

8.1.2 Classification des modèles

En sciences le modèle peut prendre des formes variées et multiples : modèle mathématique, modèle de données, modèle d'entreprise, modèle économique...

Notre réflexion s'est construite autour de la discrimination de nos modèles en trois ensembles :

- modèle d'observation pour le cadre de collecte des données sur le terrain
- modèle de données pour la structuration et la sauvegarde de ces données
- modèle mathématique pour l'analyse de la dynamique de population du ba-baçu

Les connaissances explicites autour de ces modèles permettent au scientifique de déterminer le formalisme de sa modélisation avec d'autant plus de pertinence.

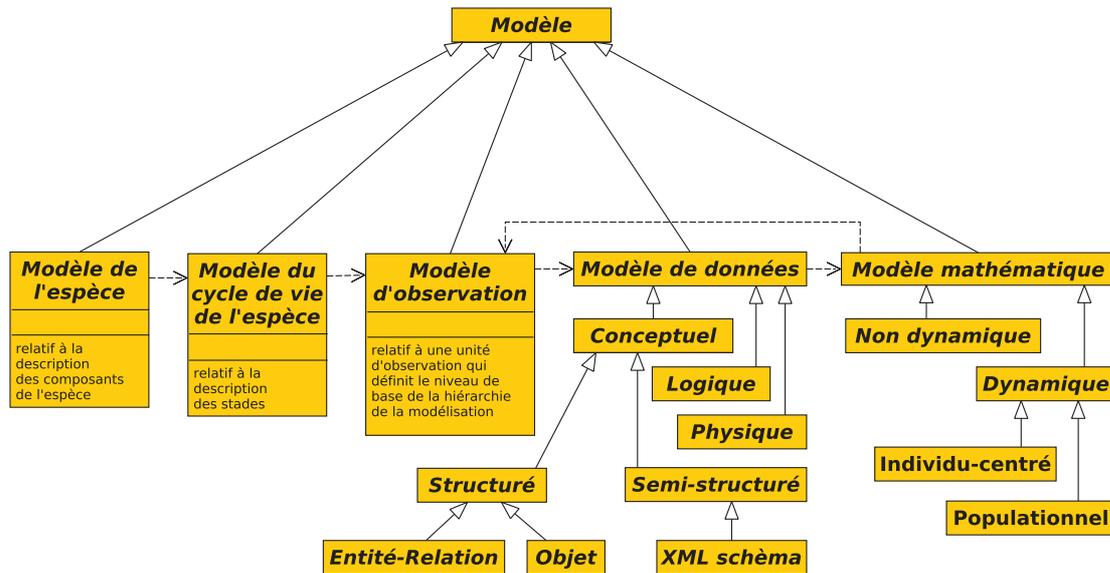
La [FIGURE 8.1](#) reflète cette réflexion autour de la construction et de la proposition d'une structuration regroupant ces trois ensembles. Cette structuration explicite les connaissances utilisées dans notre démarche. Elle nous a amené à construire un vocabulaire partagé autour de ces différents modèles qui peut servir de base à une ontologie opérationnelle pour des projets analogues (dynamique de population d'espèces végétales).

8.2 Méthodologie générale

Dans cette section nous présenterons la logique qui nous a guidé dans la réalisation de cette thèse.

La complexité de l'analyse nous a contraint à proposer un schéma regroupant les types de modèles de la [FIGURE 8.1](#) et à préciser l'articulation entre ceux-ci (représentée par les flèches pointillées).

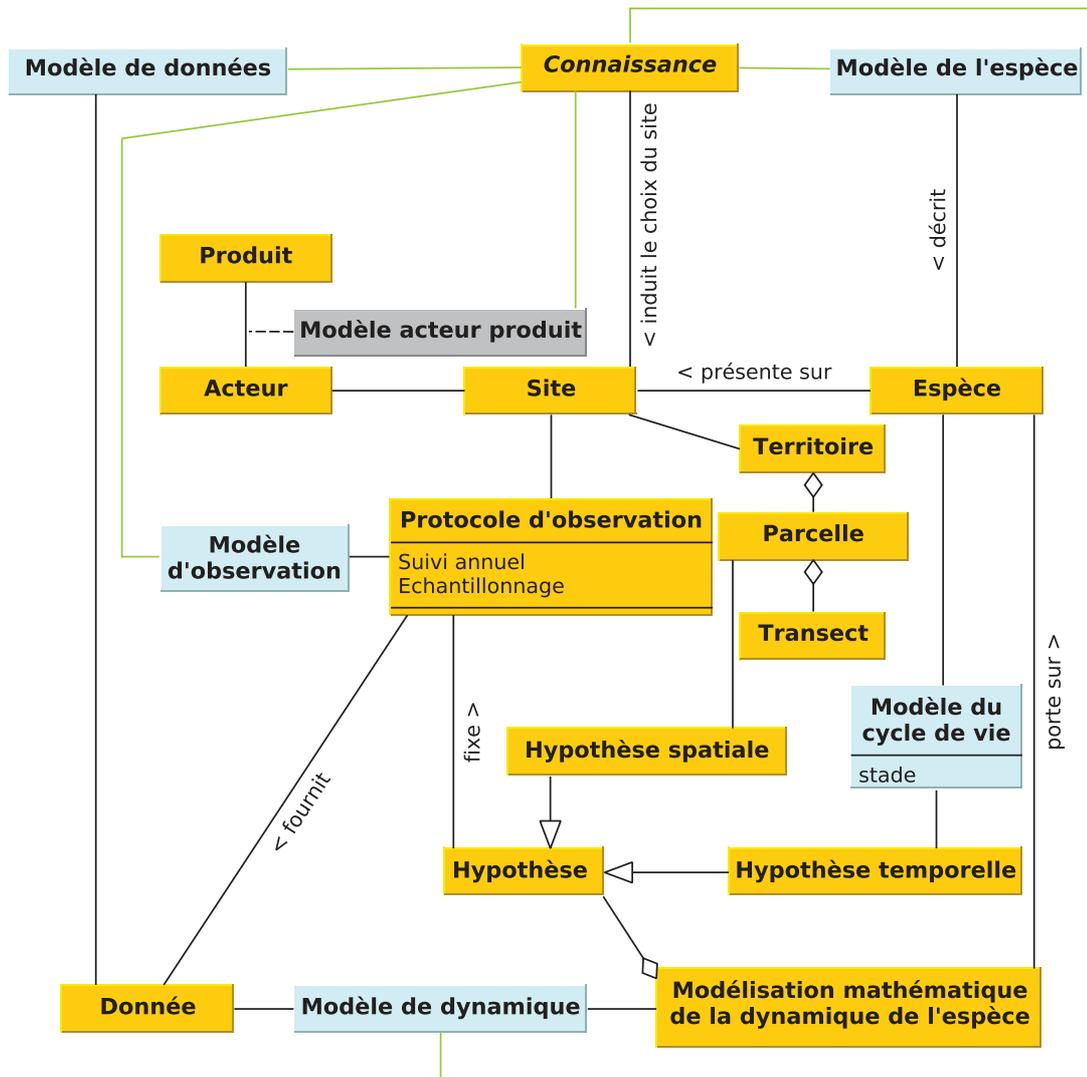
FIGURE 8.1 – Structuration des modèles (formalisme UML)



Notre objectif est de construire un modèle de dynamique de population du babaçu en nous basant sur des connaissances de différentes disciplines scientifiques. Les biologistes et écologues nous ont permis de construire le modèle de l'espèce (description de ses composants). À partir de ce modèle nous avons conçu le modèle du cycle de vie du babaçu et ainsi nous avons pu émettre des hypothèses (temporelles sur ce cycle de vie et spatiales sur la répartition de l'espèce sur le territoire) utiles au modèle de dynamique. De ces hypothèses et du modèle d'observation, dérivé du standard OBOE (Ontologie pour modéliser et représenter les observations scientifiques (Madin *et al.*, 2007)) a découlé le protocole d'acquisition des données. Les données ont été structurées selon un modèle de données qui nous a permis leur pérennisation. Le modèle de dynamique a utilisé les connaissances et données pour sa validation (cf. FIGURE 8.2).

La FIGURE 8.2 fournit un aperçu de l'articulation de ces différents modèles cités précédemment. De plus, nous avons déjà présenté le modèle acteurs-produits représentant les interactions entre les acteurs *in situ* et les produits que le babaçu fournit.

FIGURE 8.2 – Les divers modèles utilisés (formalisme UML)



Acquisition et pérennisation des données

CE CHAPITRE est dévolu à la suite logique « modèle d'espèce », « modèle de cycle de vie », « modèle d'observation » et « modèle de données ». Dans le Chapitre 5 nous avons vu que le cycle de vie du babaçu est peu étudié et reste mal connu, c'est d'ailleurs un des points bloquants dans l'analyse biologique de cette espèce. Nous souhaitons porter tout d'abord une attention particulière à la définition des stades biologiques du babaçu. Ensuite, nous nous consacrerons à l'acquisition des données issues des observations *in situ*, puis nous aborderons la structuration et la pérennisation de ces données.

Sommaire

9.1	Définition des stades biologiques	59
9.2	Observations	62
9.2.1	Choix du site	63
9.2.2	Protocole : préparation de l'observation	63
9.3	Collecte des données	66
9.4	Données résultantes	68
9.4.1	Données des pâturages	68
9.4.2	Données des babaçuals	72
9.4.3	Données de la forêt tropicale humide	73
9.5	Structuration et pérennisation des données	74
9.5.1	Phase conceptuelle	75
9.5.2	Phase de passage au modèle logique	75
9.5.3	Phase d'implémentation physique	77

9.1 Définition des stades biologiques

Avant de débiter tout travail sur le terrain, nous avons considéré comme extrêmement important de mener une réflexion préalable sur la définition rigoureuse des stades de développement du palmier. Cette réflexion s'est très largement basée sur la description des composants de l'espèce (cf. Chapitre 4 - Autour de l'espèce babaçu) et sur des observations antérieures du fonctionnement biologique de l'espèce réalisées par Mitja et Ferraz (2001); Barot *et al.* (2005); Santos et Mitja (2011). Elle a

ainsi abouti à une définition scientifique des stades reposant sur les caractéristiques biologiques du palmier. Ce qui nous a permis de proposer une définition possible du cycle de vie du babaçu.

Les écologues de l'équipe ont défini six stades de développement biologique du palmier babaçu présentés dans la FIGURE 9.1. Ce choix s'est largement inspiré de l'expérience *in situ* de Danielle Mitja et Alessio Moreira dos Santos, qui ont effectué de nombreuses missions de terrain dans le cadre de l'étude biologique du palmier babaçu au sein de la forêt primaire, des jachères et des pâturages (Mitja et Ferraz, 2001; Barot *et al.*, 2005; Santos et Mitja, 2011). La liste des caractéristiques distinctives de chaque stade est dressée ci-dessous.

FIGURE 9.1 – Stades biologiques du babaçu (clichés : Danielle Mitja)



Le stade 1 est celui de la plantule dont le limbe de toutes les feuilles n'est pas divisé (cf. FIGURE 9.1a).

Les 4 stades suivants (2 à 5) sont relatifs aux palmiers jeunes, mais présentant chacun des caractéristiques différentes.

Le stade 2 se distingue par au moins une feuille divisée ou en cours de division. Aucun pétiole ne dépassant les 50 cm (cf. FIGURE 9.1b).

Le stade 3 se caractérise par le limbe des feuilles divisé ou en cours de division. Au moins un pétiole d'une des feuilles dépasse les 50 cm. Le bourgeon terminal reste enterré. (cf. FIGURE 9.1c).

Le stade 4 est marqué par un bourgeon terminal qui est au-dessus de la surface du sol (les gaines des feuilles sont encore bien visibles). Le palmier de ce stade a une circonférence du stipe qui ne dépasse pas 1,5 m (cf. FIGURE 9.1d).

Le stade 5 regroupe les palmiers dont le stipe, couvert ou non par les gaines des feuilles, est aérien. La circonférence du stipe, lorsqu'il est couvert par les gaines des feuilles, est supérieure à 1,5 m. Aucune trace d'organes reproducteurs féminins ou masculins n'est décelée (cf. FIGURE 9.1e).

Le stade 6 se réfère aux palmiers adultes dont le stipe, couvert ou non par les gaines des feuilles, est aérien. Les organes reproducteurs féminins ou masculins sont présents (cf. FIGURE 9.1f).

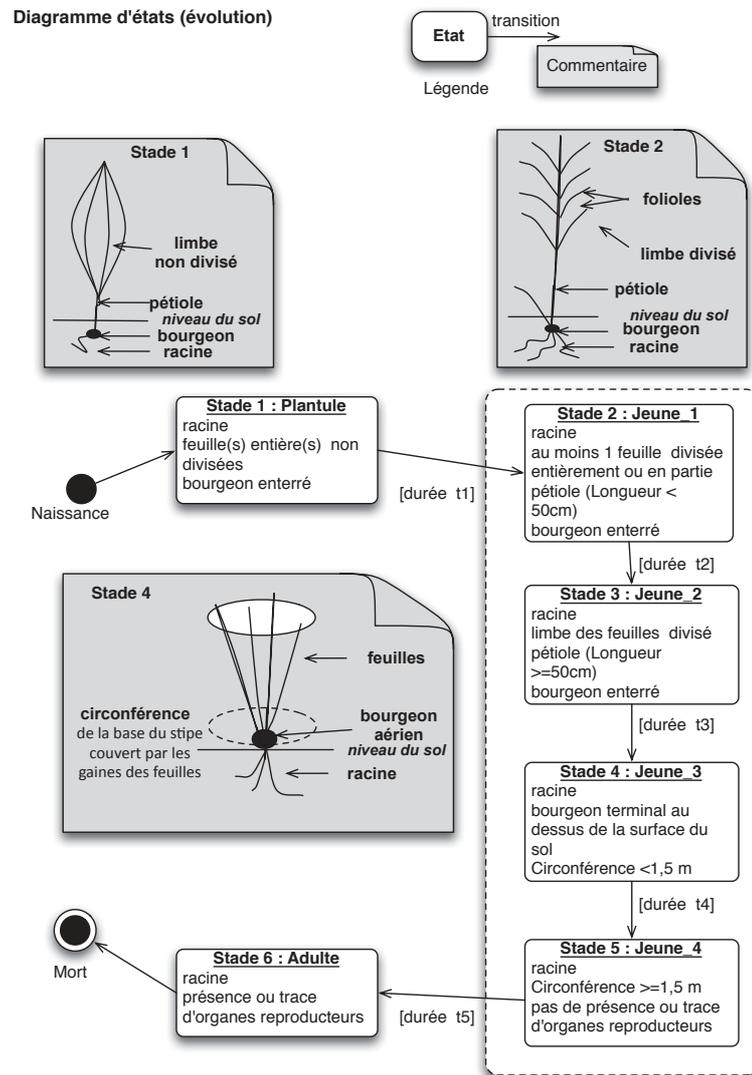
Ci-dessous, nous proposons sous forme de diagramme d'états UML le cycle de vie du babaçu (cf. FIGURE 9.2), ainsi qu'un résumé de ces informations sous la forme d'un tableau synthétique (cf. TABLEAU 9.1).

TABLEAU 9.1 – Les stades biologiques du babaçu

Stade	Nom	Abréviation	Caractéristiques
1	Plantule	P	Uniquement des feuilles entières
2	Jeune 1	J1	Division au moins partielle de la (des) feuille(s)
3	Jeune 2	J2	Le limbe des feuilles divisé ou en cours de division. Au moins un pétiole d'une des feuilles dépasse les 50 cm. Le bourgeon terminal reste enterré.
4	Jeune 3	J3	Bourgeon terminal est à la surface du sol
5	Jeune 4	J4	Présence d'un stipe
6	Adulte	A	Organes reproducteurs

Les caractéristiques ainsi définies de chacun de ces six stades ont pour objectif de minimiser la confusion possible entre deux stades successifs. Ces définitions ont été établies afin qu'un public averti (formé à la reconnaissance des stades) puisse identifier clairement et sans ambiguïté le stade de n'importe quel palmier.

Le choix du nombre de stades et de leur définition biologique découle des connaissances de la littérature et de nos collègues écologues. Cependant, notons qu'une information cruciale pour le futur modèle de dynamique de population nous échappe :

FIGURE 9.2 – Cycle de vie (diagramme d'états UML, in Mitja *et al.* (sous presse))

combien de temps dure chaque stade ?

Il n'existe, à notre connaissance, aucune étude de ce type menée à ce jour relative à cette espèce.

9.2 Observations

Cette section décrit le protocole d'observation et la méthode d'acquisition des données *in situ*.

9.2.1 Choix du site

Le choix du site, consacré à l'étude de terrain, est le fruit d'une réflexion collective. Un des critères retenus était que le site choisi ait déjà été étudié par certains membres de notre projet. Nous avons pour ce faire choisi le territoire du PA de Benfica (cf. FIGURE 2.1 - page 8).

Plusieurs études ont été déjà menées par certains membres de notre équipe sur le territoire de Benfica. Barot *et al.* (2005) ont travaillé sur la problématique du processus de reproduction du palmier babaçu sur ce même site. Santos et Mitja (2011) ont fait un considérable travail d'inventaire des espèces d'arbres et de palmiers présents au sein des pâturages de Benfica.

Nous avons estimé qu'utiliser l'infrastructure et les liens déjà tissés sur place ne peut qu'être un avantage pour cette nouvelle étude de terrain. En effet, Danielle Mitja et Éric Delaître ont également effectué des recherches en matière de déforestation à partir d'images satellitales portant sur ce même territoire (Delaître *et al.*, 2012; Laques *et al.*, 2012).

9.2.2 Protocole : préparation de l'observation

Le protocole de relevés de terrain a été avancé par les écologues de l'équipe en suivant la proposition OBOE (cf. Chapitre 6 - Données et connaissances). Les observations portent sur des objets d'intérêt (soit transects, soit individus - palmiers babaçu) et les mesures récoltées seront de nature numérique et qualitative. Pour les transects c'est le nombre de fruits isolés sur ceux-ci qui nous intéresse. Pour les palmiers babaçu, les mesures seront plus complexes : en premier lieu le stade et ensuite en fonction de celui-ci des comptages relatifs aux inflorescences, infrutescences et fruits au sol (sous le palmier).

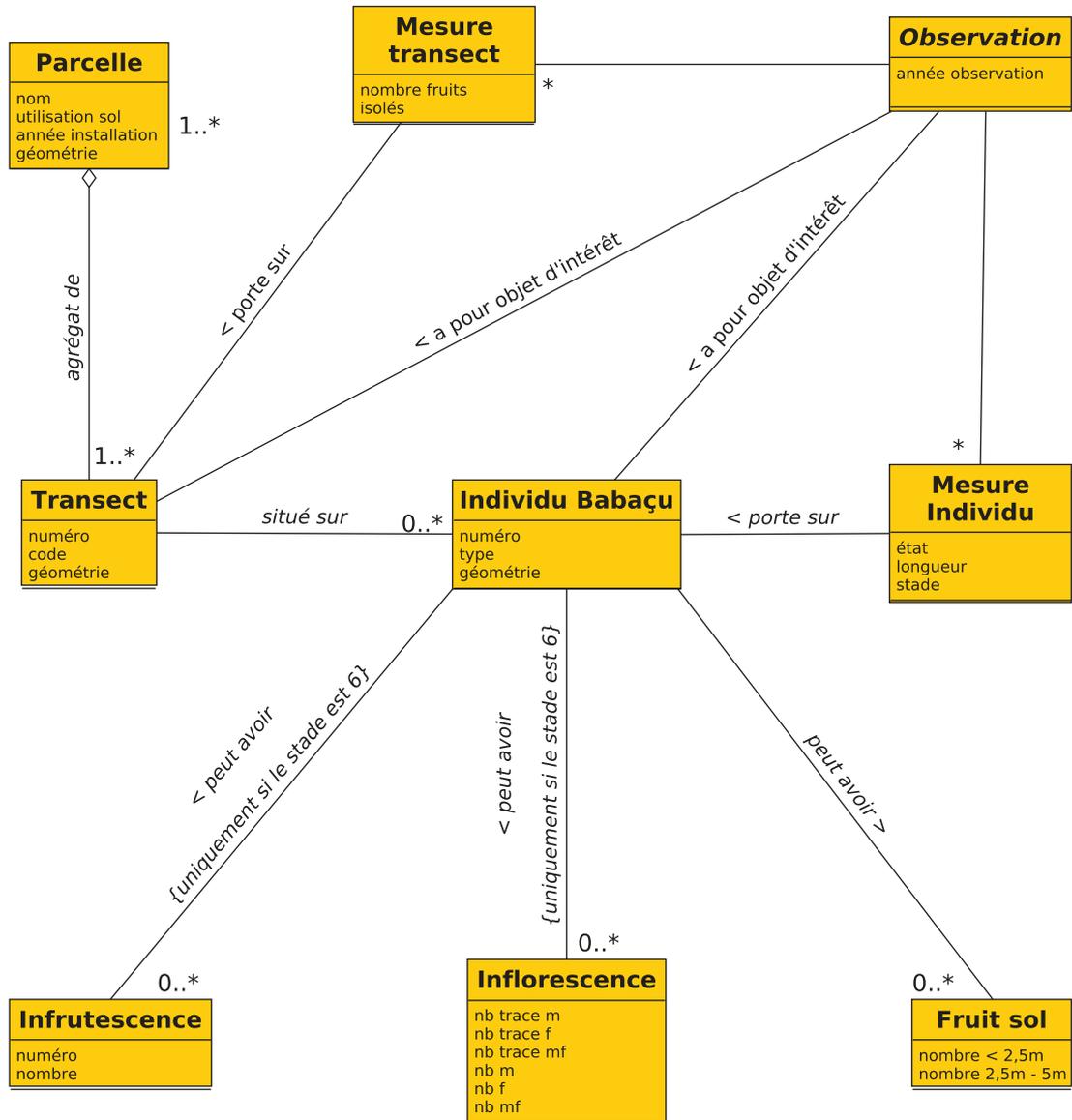
En ce qui concerne l'observation sur les palmiers babaçu, le protocole a été élaboré à partir d'expériences antérieures provenant de la littérature et adapté grâce à leurs propres connaissances du terrain. Une première ébauche de celui-ci a été proposée pendant nos travaux de recherche du master (Sirakov, 2012). Par la suite, ce protocole a été affiné par les différents membres du projet « Gestion durable du babaçu ». Lors de ces réunions nous avons décidé à la demande des modélisateurs mathématiciens qu'il y aurait un marquage exhaustif des palmiers de chaque stade (le comptage se faisant à partir du marquage) au sein de portions de parcelles que nous appellerons désormais « transects ».

Le marquage de chaque palmier a été fait à l'aide d'étiquettes rectangulaires en acier inoxydable de taille 50 mm par 20 mm, chaque étiquette comportant un numéro unique. Afin de s'assurer d'une déperdition minimale des étiquettes, elles ont été attachées aux palmiers grâce à un fil métallique ou clouées sur le stipe.

Ce marquage a été également défini à partir d'hypothèses spatio-temporelles relatives à l'âge des pâturages et émises par les modélisateurs mathématiciens.

Sur le plan temporel, la période des relevés de terrain est définie en fonction

FIGURE 9.3 – Modèle d'observation (formalisme UML)



du cycle annuel de production de fruits du babaçu. Compte tenu des informations suivantes :

- La période de floraison des inflorescences féminines du palmier durent 9 mois (d'octobre à juin), avec un pic en février et mars,
- La période de production de fruits est décalée par rapport à la floraison et dure 9 mois. Elle commence en juin pour se terminer en février, avec d'importants pics en octobre et novembre,
- La germination des fruits commence en moyenne trois mois après les pics de la production de fruits et coïncide avec le milieu de la saison des pluies

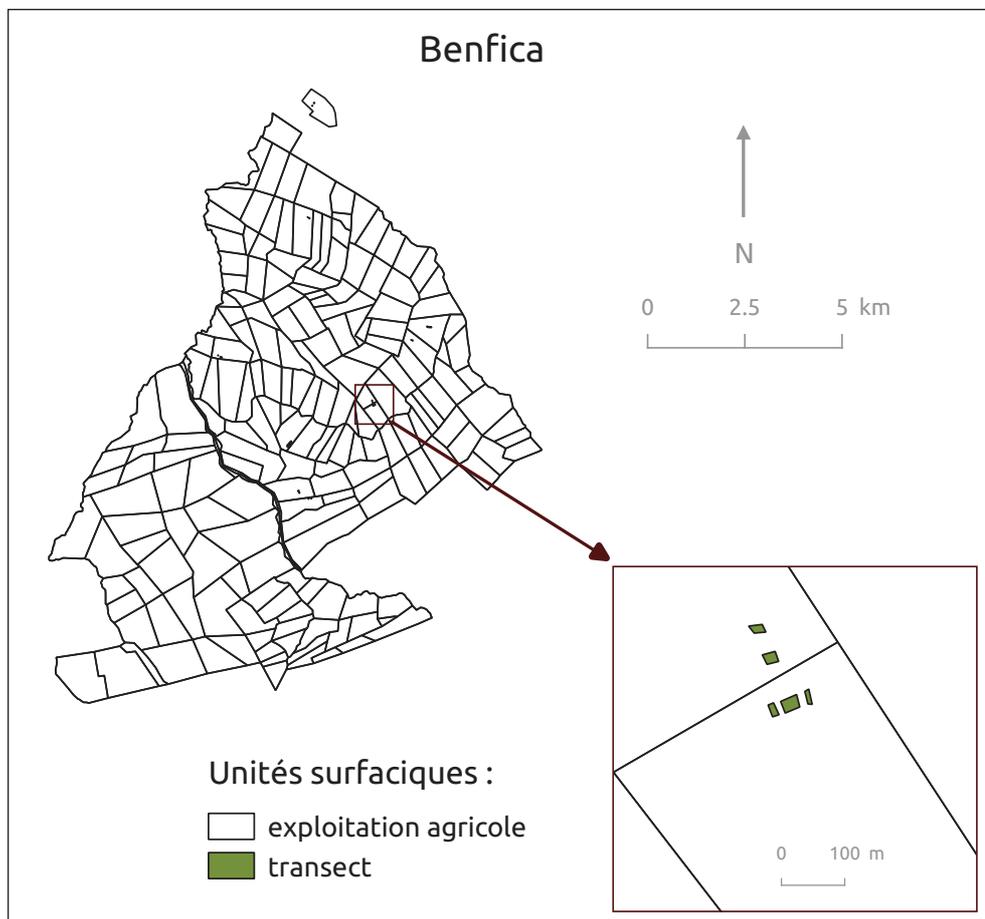
(Anderson *et al.*, 1991).

nous avons retenu l'idée d'effectuer les relevés durant les mois de juillet-août et entre 2013 et 2016.

Côté spatialisation, dans un premier temps nous avons effectué les mesures dans des transects (définis au sein des parcelles pour exploiter la variabilité au sein de celles-ci). Les parcelles ne doivent pas être éloignées les unes des autres : surveillance du respect quant à l'uniformité des caractéristiques bio-chimiques des sols, du climat, des pratiques anthropiques.

Ces transects ont été présélectionnés *a priori* aléatoirement au sein de chaque parcelle. La FIGURE 9.4 montre un exemple comportant des parcelles et des transects sur le territoire de Benfica.

FIGURE 9.4 – PA Benfica - transects



Le choix des parcelles concernées par l'étude de terrain s'est fait après concertation avec les agriculteurs locaux. Elles ont été sélectionnées en fonction de l'occupation du sol : forêt tropicale humide, pâturage ou babaçal. Une des contraintes

majeures résidait dans le fait de trouver des parcelles qui conserveraient le même type d'occupation du sol tout au long de l'étude de terrain. Nous devions nous assurer que les propriétaires n'envisageaient pas de modifier leur fonctionnement habituel, que le marquage resterait sur place et que nous disposerions d'un accès à la parcelle durant les quatre années de relevés de terrain. Un des critères importants pour la sélection des parcelles de type pâturages concerne l'âge de celle-ci (c'est-à-dire année de déforestation). Cette discrimination en âge des différentes parcelles avait pour but de faciliter l'appréhension de l'évolution du système babaçu dans le temps. Discriminer des comportements de l'espèce en fonction de la durée de l'ouverture du milieu depuis le défrichement.

Au total nous avons sélectionné sept parcelles de type « pâturage » regroupant un total de vingt-quatre transects. Parmi ces transects, dix-sept ont été répartis à hauteur de 2 ou 3 par pâturage (notés « 1 », « 2 » ou « 3 » dans le TABLEAU 9.2). Ces transects-là ont servi pour le marquage/comptage des palmiers issus de tous les six stades biologiques. Sept autres transects (notés « suppl » dans le TABLEAU 9.2) ont servi au marquage/comptage uniquement des palmiers des stades 5 et 6. Les lettres « a, b, . . . , h » au sein de ce tableau servent pour la correspondance avec les données présentées dans le TABLEAU 11.4 (page 114).

Deux parcelles sur une occupation du sol de type forêt secondaire mono-spécifique de babaçu (« babaçal ») et deux sur une occupation du sol de type forêt tropicale ont été aussi sélectionnées. Chacune de ces parcelles comporte deux transects réunissant tous les stades (cf. TABLEAU 9.2)

Notons que conformément à la notion d'itération présentée dans le Chapitre 8, certains aspects techniques et analytiques du protocole ont été revus et réadaptés. C'est un processus incontournable, puisqu'il existe toujours un degré d'aléa entre la théorie et son application sur le terrain. Les modifications du protocole ont été effectuées, suite à l'avancée de nos travaux, durant l'année 2015 et n'ont pu donc être appliquées que durant les collectes 2015 et 2016. Malheureusement l'arrivée de ces nouvelles mesures ne pourra être prise en compte qu'au delà de la soutenance de ce travail.

9.3 Collecte des données

Avant tout, il nous semble important de préciser que la phase de collecte des données sur le terrain a été effectuée par le partenariat brésilien : Alessio Moreira dos Santos (doctorant brésilien travaillant sur l'écologie du babaçu et ayant en charge cette collecte) et Deurival da Costa Carvalho (technicien de terrain) et certains membres de notre équipe : Danielle Mitja et Éric Delaître. Le partenariat avec l'UFRA dans le cadre du projet Open Science a défini un accord bilatéral avantageant les deux doctorants : cette thèse bénéficie de l'usage des données de terrain collectées et la thèse d'Alessio Moreira dos Santos disposera des avancées réalisées en matière de modélisation de la dynamique de population du babaçu.

TABLEAU 9.2 – Structure des parcelles

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
forêt tropicale	1	DE_flo	-	1
				2
				suppl
	2	ME_flo	-	1
				2
				suppl
pâturage	1	DEpasto_12	2012	1 (a)
				2
				3
				suppl
	2	CLpasto_10	2010	1
				2
				suppl
	3	CHpasto_10	2010	1
				2
				suppl
	4	RVpasto_08	2008	1 (b)
				2
				3
				suppl
	5	AMpasto_06	2006	1 (c)
				2
				suppl
	6	ARpasto_06	2006	1 (d)
				2 (e)
				suppl
	7	AMpasto_04	2004	1 (f)
				2 (g)
				3 (h)
				suppl
babaçal	1	RL_bab	-	1
				2
	2	ZM_bab	-	1
				2

Le marquage de tous les palmiers de tous les stades a été réalisé en 2013. Tous les nouveaux palmiers de stade 1 (apparus entre deux passages consécutifs) ont également été marqués et rajoutés au fur et à mesure au sein des données.

En 2014, 2015 et 2016 le suivi des palmiers marqués auparavant a été effectué. En parallèle du marquage, la collecte des données issues des observations a été effectuée. Les palmiers ont été inscrits au sein de l'inventaire sous une des trois différentes

formes possibles.

- le palmier était retrouvé, signalé comme « vivant » et ses caractéristiques enregistrées
- seule l'étiquette du palmier était retrouvée, mais le palmier n'était plus présent. Ce palmier était enregistré comme « mort »
- ni l'étiquette, ni le palmier n'étaient retrouvés. Le palmier était alors enregistré comme « perdu ». Un palmier qui a disparu l'année $n + 1$ et a réapparu l'année $n + 2$ est réintégré dans l'inventaire à partir de l'année $n + 1$.

En ce qui concerne le comptage des fruits attachés au palmier, celui-ci prend en compte le nombre de fruits comptés directement sur les infrutescences des palmiers adultes, ainsi que le nombre de fruits au sol sous le palmier dans un rayon de 5 m autour du stipe. Ce comptage n'a pas été réalisé au sein des transects supplémentaires marqués comme « suppl » dans les TABLEAUX 9.3, 9.4, 9.5 et 9.6.

L'observation portant sur les infrutescences s'effectue à distance (à partir du sol) et est dépendante de l'observateur (degré de connaissance et d'évaluation rapide). Nous ne disposons d'aucun moyen pour estimer l'erreur de comptage associée à chaque observateur. De ce fait, le protocole d'observation a été ajusté pour les collectes des données de 2014, 2015 et 2016 : le comptage a été accompagné par la prise de photographies de chaque infrutescence.

Remarque :

La collecte des données s'effectue en parallèle de cette thèse : les données de 2016 ont été collectées sur le terrain durant le mois d'août 2016 et ne nous sont pas encore parvenues à ce jour (le 15 octobre 2016). De ce fait, malheureusement nous n'avons pas pu les utiliser dans le cadre de cette thèse.

9.4 Données résultantes

Le nombre total de palmiers marqués et enregistrés s'élève à 4 979 en 2013 auxquels ont été rajoutés 639 nouveaux palmiers en 2014 et 703 autres en 2015. Les palmiers sont répartis inégalement au sein des six stades.

9.4.1 Données des pâturages

Le TABLEAU 9.3 résume les données récoltées en 2013 au sein des pâturages.

Le nombre de palmiers par transect et par stade y est indiqué. Notre logique était basée sur l'équité du nombre de palmiers enregistrés en fonction de l'année d'ouverture du pâturage. L'emprise de chaque transect a été donc définie en fonction du nombre de palmiers le composant. La superficie des transects peut en conséquence être variable d'un transect à un autre. L'unique critère impactant sur la taille du

TABLEAU 9.3 – Données brutes par transect (pâturages 2013)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
DEpasto_12	1	8	46	114	36	8	8	5	217
	2	5,2	110	65	16	4	3	0	198
	3	5	70	71	38	1	4	1	185
	suppl	8,4					4	12	16
CLpasto_10	1	33,6	28	97	16	0	1	0	142
	2	33,6	29	90	19	3	1	0	141
	suppl	137,38					7	3	10
CHpasto_10	1	4,4	39	107	0	0	0	0	146
	2	2	145	19	0	0	0	0	164
	suppl	183,14					1	14	15
RVpasto_08	1	6	35	116	31	1	1	3	187
	2	6	28	138	30	0	0	0	196
	3	3,4	6	155	36	0	0	0	197
	suppl	135,93					4	22	26
AMpasto_06	1	3,2	107	84	17	0	0	2	210
	2	2,8	0	46	30	0	0	0	76
	suppl	217,3					1	15	16
ARpasto_06	1	3,2	29	86	18	0	0	3	136
	2	5,4	9	110	26	1	2	6	154
	suppl	15,34					9	3	12
AMpasto_04	1	1,8	13	150	12	0	0	1	176
	2	5,6	37	132	14	0	0	1	184
	3	2	5	157	1	1	0	1	165
	suppl	93,38					6	11	17

transect était le nombre de palmiers répertoriés en fonction de l'âge de la parcelle. Ce nombre a été fixé à environ 600 palmiers.

Prenons, pour illustrer nos propos, l'exemple d'une parcelle unique comportant trois transects plus un transect dit supplémentaire, c'est-à-dire les parcelles « DEpasto_12 », « RVpasto_08 » et « AMpasto_04 », ouvertes respectivement en 2012, 2008 et 2004. Le nombre de palmiers par année d'ouverture (600) a été divisé par 3 (nombre de transects) et ainsi chaque transect a comporté environ 200 palmiers (cf. TABLEAU 9.3).

Les quatre autres parcelles (« CLpasto_10 », « CHpasto_10 », « AMpasto_06 » et « ARpasto_06 »), composées de deux transects plus un transect supplémentaire, ont été ouvertes respectivement en 2010 et 2006. Nous disposons donc de deux parcelles (quatre transects en excluant les transects supplémentaires) par année d'ouverture. Les transects, pour une même année d'ouverture (par exemple 2010), ont été délimités de manière que chacun comporte environ 150 palmiers (cf. TABLEAU 9.3).

TABLEAU 9.4 – Densités par transect (pâturages 2013)

Code parcelle	Transect	Superficie (ares)	Densité (palmiers/are)					
			Stades					
			1	2	3	4	5	6
DEpasto_12	1	8	5,75	14,25	4,5	1	1	0,63
	2	5,2	21,15	12,5	3,08	0,77	0,58	0
	3	5	13,8	14,4	7,6	0,2	0,8	0,2
	suppl	8,4					0,48	1,43
CLpasto_10	1	33,6	0,83	2,89	0,48	0	0,03	0
	2	33,6	0,86	2,71	0,54	0,09	0,03	0
	suppl	137,38					0,05	0,02
CHpasto_10	1	4,4	8,86	24,32	0	0	0	0
	2	2	72,5	9,5	0	0	0	0
	suppl	183,14					0,01	0,08
RVpasto_08	1	6	5,83	19,33	5,17	0,17	0,17	0,5
	2	6	4,67	23,17	4,83	0	0	0
	3	3,4	1,76	45,59	10,59	0	0	0
	suppl	135,93					0,03	0,16
AMpasto_06	1	3,2	33,75	25,94	5,31	0	0	0,63
	2	2,8	0	16,43	10,71	0	0	0
	suppl	217,3					0	0,07
ARpasto_06	1	3,2	9,06	26,88	5,63	0	0	0,94
	2	5,4	1,67	20,37	4,81	0,19	0,37	1,11
	suppl	15,34					0,59	0,2
AMpasto_04	1	1,8	7,22	81,67	8,33	0	0	0,56
	2	5,6	6,79	23,21	2,68	0	0	0,18
	3	2	2,5	78	0,5	0,5	0	0,5
	suppl	93,38					0,06	0,12

Cette fois-ci nous avons divisé 600 (nombre de palmiers par année d'ouverture) par 4 (nombre de transects par année d'ouverture).

Du fait de l'hétérogénéité de la superficie des transects, les données brutes ne permettent pas la comparaison entre les transects. Nous avons donc opté pour une transformation de ces données brutes en rapportant le nombre de palmiers par transect à la superficie de chaque transect. Ainsi nous avons obtenu des données comparables exprimées en nombre de palmiers par unité de surface. Le TABLEAU 9.4 résume cette information pour les données de 2013 (les TABLEAUX A.1 et A.2 relatifs respectivement aux densités des années 2015 et 2016 ont été placés en Annexe A - page 141) . Sa lecture nous renseigne quant à l'importante hétérogénéité et variabilité qui existent entre les transects : même deux transects du même pâturage peuvent avoir un écart en densité de babaçus plus important que des transects de pâturages différents. Prenons exemple des palmiers du stade 1 du pâturage « DEpasto_12 » : l'écart entre le transect un (5,75 individus par are) et le transect deux

TABLEAU 9.5 – Données brutes par transect (pâturages 2014)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
DEpasto_12	1	8	37	84	65	10	9	5	210
	2	5,2	61	74	41	3	4	0	183
	3	5	44	63	61	1	4	1	174
	suppl	8,4					4	12	16
CLpasto_10	1	33,6	12	96	30	3	1	0	142
	2	33,6	20	80	42	3	2	0	147
	suppl	137,38					6	4	10
CHpasto_10	1	4,4	25	122	2	0	0	0	149
	2	2	138	26	0	0	0	0	164
	suppl	183,14					1	13	14
RVpasto_08	1	6	44	108	47	1	1	3	204
	2	6	17	106	72	0	0	0	195
	3	3,4	4	88	105	0	0	0	197
	suppl	135,93					2	24	26
AMpasto_06	1	3,2	366	101	25	0	0	2	494
	2	2,8	0	40	36	0	0	0	76
	suppl	217,3					1	15	16
ARpasto_06	1	3,2	28	82	25	0	0	3	138
	2	5,4	5	103	34	0	3	6	151
	suppl	15,34					3	9	12
AMpasto_04	1	1,8	44	136	26	0	0	1	207
	2	5,6	88	103	22	0	0	1	214
	3	2	3	129	1	1	0	1	135
	suppl	93,38					6	11	17

(21, 15 palmiers par are) est de l'ordre du rapport un à quatre. La comparaison de la densité de ce même transect un (5, 75 individus par are) avec les densités des transects un et deux du pâturage « RVpasto_08 » (respectivement 5, 83 et 4, 67 palmiers par are) nous indique des valeurs proches. Il en résulte donc pour le stade un que les densités de ces trois transects sont plus proches que les densités au sein du pâturage « DEpasto_12 ».

La structure de ces données collectées en 2014 et 2015 est identique à celle des données collectées en 2013. Ces données sont présentées respectivement dans les TABLEAUX 9.5 et 9.6. Nous notons une relative stabilité dans le nombre total de palmiers par transect entre 2013 et 2014. Cette stabilité se confirme également entre 2014 et 2015. Certaines valeurs, comme par exemple le transect 1 du pâturage « AMpasto_06 », sont marquées par une nette évolution du nombre de palmiers par transect : passage de 210 palmiers par transect en 2013 à 494 palmiers par transect en 2014 (cf. TABLEAUX 9.3 et 9.5). Cette évolution est due à la variabilité du processus

de recrutement et s'explique par une nette augmentation des palmiers répertoriés de stade 1.

TABLEAU 9.6 – Données brutes par transect (pâturages 2015)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
DEpasto_12	1	8	49	74	83	9	9	5	229
	2	5,2	41	86	47	2	5	0	181
	3	5	49	49	69	1	4	2	174
	suppl	8,4					4	12	16
CLpasto_10	1	33,6	2	75	44	10	2	0	133
	2	33,6	2	62	56	5	4	0	129
	suppl	137,38					5	5	10
CHpasto_10	1	4,4	17	125	3	0	0	0	145
	2	2	107	31	0	0	0	0	138
	suppl	183,14					1	13	14
RVpasto_08	1	6	228	118	53	2	1	3	405
	2	6	10	98	85	0	0	0	193
	3	3,4	3	83	111	0	0	0	197
	suppl	135,93					1	24	25
AMpasto_06	1	3,2	292	106	29	0	0	2	429
	2	2,8	0	29	47	0	0	0	76
	suppl	217,3					0	16	16
ARpasto_06	1	3,2	35	85	27	0	0	3	150
	2	5,4	17	94	34	0	2	7	154
	suppl	15,34					3	9	12
AMpasto_04	1	1,8	87	124	43	0	0	1	255
	2	5,6	80	90	39	0	0	1	210
	3	2	5	122	5	1	0	1	134
	suppl	93,38					5	12	17

9.4.2 Données des babaçuals

Le TABLEAU 9.7 résume les données récoltées en 2013 au sein des babaçuals. Pour ce milieu, notre logique avait changé, puisque nous n'étions pas en mesure de connaître l'année exacte de transformation du milieu en babaçual. Pour ces milieux, nous nous sommes basés donc sur le principe d'équité du nombre de palmiers enregistrés par parcelle et non plus sur une équité par année d'ouverture. Nous disposions de deux parcelles de babaçual (« RI_bab » et « ZM_bab ») composées chacune de deux transects. Le nombre total de palmiers par parcelle a été défini à hauteur de 500 palmiers. En conséquence, les transects étaient délimités dès le dénombrement d'environ 250 palmiers au sein de chacun d'eux.

TABLEAU 9.7 – Données brutes par transect (babaçuals 2013)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
RL_bab	1	2,6	108	94	22	2	1	4	231
	2	4,6	89	120	24	1	10	7	251
ZM_bab	1	5,2	138	91	30	1	6	5	271
	2	6,8	97	69	51	8	8	3	236

Ci-dessous, nous avons présenté les données de 2014 et 2015 respectivement dans les TABLEAUX 9.8 et 9.9.

TABLEAU 9.8 – Données brutes par transect (babaçuals 2014)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
RL_bab	1	2,6	107	95	24	2	1	4	233
	2	4,6	82	124	25	3	10	7	251
ZM_bab	1	5,2	119	103	34	1	6	5	268
	2	6,8	67	88	58	7	7	5	232

TABLEAU 9.9 – Données brutes par transect (babaçuals 2015)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
RL_bab	1	2,6	92	96	16	7	1	4	216
	2	4,6	72	113	37	9	7	7	245
ZM_bab	1	5,2	67	102	38	2	1	1	211
	2	6,8	33	95	55	8	2	0	193

9.4.3 Données de la forêt tropicale humide

Le TABLEAU 9.10 résume les données récoltées en 2013 au sein de la forêt primaire. Pour ce milieu, notre logique était sensiblement la même que celle utilisée pour les babaçuals. Nous disposions de deux parcelles de forêt primaire (« DE_flo » et « ME_flo ») composées chacune de deux transects. Comme pour les babaçuals, le nombre total de palmiers par parcelle a été défini à hauteur de 500 palmiers et les transects étaient délimités dès le dénombrement d'environ 250 palmiers au sein de chacun d'eux.

TABLEAU 9.10 – Données brutes par transect (forêt tropicale humide 2013)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
DE_flo	1	5	229	11	4	0	1	3	248
	2	2,6	220	15	4	0	0	0	239
	suppl	115,78					4	12	16
ME_flo	1	2	166	56	17	1	2	1	243
	2	3,6	162	45	31	2	1	6	247
	suppl	2,76					4	6	10

Les deux TABLEAUX suivants (9.11 et 9.12) font référence respectivement des données de 2014 et 2015.

TABLEAU 9.11 – Données brutes par transect (forêt tropicale humide 2014)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
DE_flo	1	5	212	14	4	0	0	4	234
	2	2,6	192	22	4	0	0	0	218
	suppl	115,78					3	13	16
ME_flo	1	2	150	56	22	1	2	1	232
	2	3,6	150	46	32	2	1	6	237
	suppl	2,76					4	6	10

TABLEAU 9.12 – Données brutes par transect (forêt tropicale humide 2015)

Code parcelle	Transect	Superficie (ares)	Nombre de palmiers						Total
			Stades						
			1	2	3	4	5	6	
DE_flo	1	5	200	18	5	0	0	4	227
	2	2,6	180	23	4	0	0	0	207
	suppl	115,78					3	13	16
ME_flo	1	2	209	54	28	1	2	1	295
	2	3,6	141	44	32	2	1	6	226
	suppl	2,76					4	6	10

9.5 Structuration et pérennisation des données

En première approche pour pérenniser les données, nous aurions pu utiliser des fichiers tableur. Cependant, il nous a semblé judicieux de constituer un outil ras-

semblant plus d'informations compte tenu du partenariat élargi au sein du projet et exploitant les dimensions spatio-temporelles. Nous avons donc opté pour la mise en place d'une base de données spatialisée.

De manière classique cette mise en place s'est articulée autour de trois étapes consécutives dont la finalité consiste dans l'élaboration d'un cadre cohérent et performant pour l'analyse et la sauvegarde des données. Nous fournirons ici une présentation détaillée de ces trois étapes :

- la phase de conceptualisation,
- la phase de passage au modèle logique,
- la phase d'implémentation physique au sein d'un logiciel dédié.

9.5.1 Phase conceptuelle

La modélisation conceptuelle est la première étape de cette démarche. Il s'agit de modéliser la sémantique des informations d'une façon compréhensible pour tout utilisateur de la future base de données. Plusieurs méthodes existent pour cet exercice, nous nous sommes appuyés sur une méthode de modélisation orientée objet utilisant le langage de modélisation unifié UML. Le schéma conceptuel obtenu (diagramme de classes) définit des classes d'objets concrets ou abstraits caractérisées par leur(s) attribut(s) et reliées entre elles à l'aide de relations. Ces relations sont nommées et leur sémantique est renforcée par les multiplicités (0, 1 ou * etc.) qui précisent le nombre d'objets de chaque classe concernés par la relation.

La FIGURE 9.5 présente le modèle conceptuel obtenu. Dans ce modèle figurent des informations plus générales comme « Pays » ou « Biomes » qui ont été adjointes aux informations introduites précédemment par le protocole de collecte comme « Transects » et « Individu babaçu ». Le modèle présente clairement les structures administratives et spatiales des données : le pays est un agrégat de régions, la région est un agrégat d'états, ..., la parcelle est un agrégat de transects¹. Les individus sont situés sur les transects et les observations portent soit sur les individus soit sur les transects.

9.5.2 Phase de passage au modèle logique

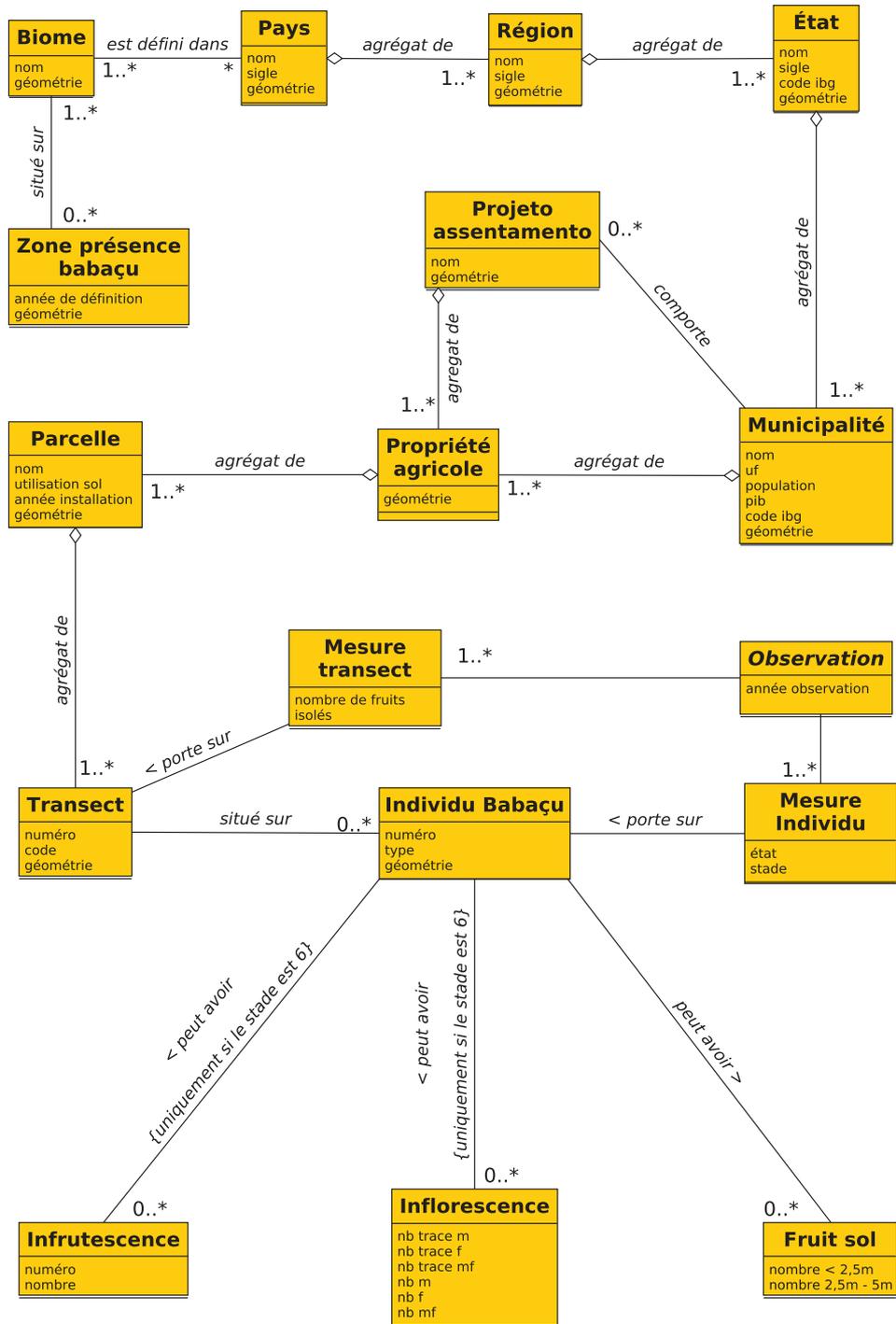
Le logiciel SGBD envisagé correspond à la catégorie des logiciels relationnels, nous devons donc transformer le modèle conceptuel en modèle dit logique² correspondant au formalisme du modèle relationnel.

Le modèle relationnel, introduit par Codd (1970) est issu de la théorie mathématique des ensembles. Dans cette théorie une relation est une partie d'un produit cartésien d'une liste de domaines. Un domaine est un ensemble de valeurs. Une relation R n -aire sera une partie du produit cartésien $D_1 \times D_2 \times \dots \times D_n$. Les éléments de la relation

1. Le schéma relatif au protocole a été simplifié : les relations liant la classe abstraite « *Observation* » objets d'intérêt individu et transect ont été supprimés.

2. Terminologie dérivée de la méthode Merise

FIGURE 9.5 – Modèle conceptuel de données (formalisme UML)



sont des n-uplets qui sont notés (d_1, d_2, \dots, d_n) . Dans ce modèle, l'ensemble des données est perçu comme des tableaux. Chaque tableau a un nom et est caractérisé par

plusieurs colonnes ayant chacune un nom, par exemple, un tableau « PERSONNE » est décrit par les colonnes « NO_PERS », « NOM », « ADRESSE », on est alors dans l'intention ou schéma de la relation. La notation « PERSONNE(NO_PERS, NOM, ADRESSE) » est une simplification, chaque constituant est associé à un domaine de valeur implicite. La relation « PERSONNE » contient un ensemble de lignes (ici triplets) qui correspond à une des extensions possibles du schéma.

Des propriétés relatives au schéma des relations complètent leur sémantique, il s'agit de contraintes d'intégrité telles le concept de clé lié à celui de dépendance fonctionnelle, ou celui de clé étrangère lié à celui de dépendance de référence.

Ce modèle a été étendu, dans le cas des données spatiales, afin de prendre en compte un domaine complexe relié dans notre schéma au constituant géométrie³.

Des règles spécifiques ont été respectées pour le passage du modèle conceptuel au modèle logique :

Règle 1 toute classe UML donne lieu à un schéma de relation intentionnel dont le nom est celui de la classe et dont l'ensemble des attributs correspond à l'ensemble des propriétés de la classe. La définition de la clé primaire du schéma de relation demande qu'il existe une propriété ou un ensemble de propriétés dit(s) *Unique* en UML, c'est-à-dire correspondant à une valeur unique lors de la réalisation de l'extension du schéma de relation. Si aucune propriété ayant cette caractéristique n'existe, on crée un attribut dit identifiant complémentaire dans le schéma relationnel.

Règle 2 deux cas possibles :

Cas 1 toute relation entre classes dont les multiplicités sont de type *,* donne lieu à un schéma de relation intentionnel dont les attributs correspondent aux attributs clés primaires extraits des deux classes impliquées. La clé primaire est l'union de ces attributs

Cas 2 Si une des multiplicités est égale à 1, la relation n'engendre pas de schéma relationnel, on se contente de faire migrer la clé primaire du schéma de relation correspondant à la classe liée à la multiplicité 1 comme attribut additionnel au sein du schéma de relation correspondant à la deuxième classe. De plus, cet attribut additionnel devient clé étrangère référençant l'attribut clé primaire de la classe origine.

La FIGURE 9.6 expose le modèle relationnel de la base obtenu à partir du modèle conceptuel présenté précédemment (cf. FIGURE 9.5).

9.5.3 Phase d'implémentation physique

Dans cette phase restait à choisir le système de gestion de base de données. Compte tenu de l'aspect spatial, présence d'attributs dénommés *the_geom* dont le domaine correspond à un type géométrique dans de nombreux schémas de relation,

3. Il s'agit d'un modèle dit objet-relationnel autorisant les domaines complexes.

FIGURE 9.6 – Modèle relationnel (logique) de données

Zone_presence_babaçu(Id_z, annee_def, the_geom)
 Biome(Nom, the_geom)
 BiomePays(#Nom_biome, #Nom_pays)
 Pays(Nom, sigle, the_geom)
 Region(Nom, sigle, the_geom, #Nom_pays)
 Etat(Nom, sigle, code_ibg, the_geom, #Nom_region)
 Municipalite(Nom, uf, pop, pib, code_ibg, the_geom, #Nom_etat)
 Projeto_assentamento(Nom, the_geom, #Nom_municipalite)
 Propriete_agricole(Id_pa, the_geom, #Nom_municipalite, #Nom_projetto_ass)
 Parcelle(Nom, utilisation_sol, annee_installation, the_geom, #Nom_prop_agri)
 Transect(Numero, code, the_geom, #Nom_parcelle)
 Individu_babaçu(Numero, type, the_geom, #Numero_transect)
 Mesure_transect(Id_mt, nombre, #annee_observation, #Numero_transect)
 Mesure_individu(Stade, #annee_observation, #Numero_individu, etat)
 Mesure_individu_fruit_sol(Id_mif, nombre1, nombre2, #Stade, #annee_observation,
#Numero_individu)
 Mesure_individu_Infrutescence(Numero, #annee_observation, #Numero_individu,
#Stade, nombre, annee)
 Mesure_individu_Inflorescence(#Stade, #annee_observation, #Numero_individu,
 nb_trace_m, nb_trace_f, nb_m, nb_f)

notre choix c'est porté sur le SGBD Postgres/SQL doté de l'extension spatiale PostGIS.

Ce choix est justifié par les nombreux avantages que ce logiciel offre. Il s'inscrit intégralement dans une démarche de logiciels Open source et libres et est disponible sous la licence « PostgreSQL License ». PostgreSQL/PostGIS se base sur un puissant langage d'interrogation de base de données, le *Structured Query Language (SQL)* augmenté des opérateurs spatiaux spécifiques métriques et topologiques. De plus le logiciel est disponible dans divers environnements (Linux et Windows).

Notre avons donc créé une base de données PostgreSQL/PostGIS sous le nom de « babassu » et l'avons alimentée avec des données provenant de sources différentes :

- données de terrain (collectées *in situ* et correspondant à celles présentées dans la section 9.3 (Collecte des données),
- données vectorielles de l'IBGE (biomes (IBGE, 2004), pays, régions, états et municipalités (IBGE, 2013)),
- données vectorielles de May *et al.* (1985); Anderson et Balick (1988); Anderson *et al.* (1991) (zone présence babaçu)

La plupart de ces données ont pu être incorporées au sein de la base via des

scripts regroupant des commandes de terminal et des requêtes SQL (cf. Annexe B). Nous avons veillé à l'automatisation maximale de tous ces traitements. De plus, les contraintes d'intégrité définies précédemment (clé primaire et clé étrangère) ont permis la détection et la correction d'erreurs de saisie survenues lors de la collecte. Ceci a assuré la fluidité méthodologique dans le traitement des données au cours des diverses années. Les données à caractère spatial ont été indexées selon la méthode d'indexation *Generalized Search Tree* (GiST) pour réduire le temps l'accès aux données lors du requêtage. Un des atouts majeurs de PostgreSQL/PostGIS est qu'il assure l'interopérabilité avec les autres logiciels dont nous nous sommes servis :

- Qgis pour la visualisation spatiale. Outre l'assurance de la pérennité des données collectées, un des avantages et non des moindres de la base, est qu'elle constitue le support qui nous a permis d'élaborer l'ensemble des représentations cartographiques de ce mémoire.
- R pour le traitement mathématique et statistique des données.

Modélisation mathématique

EN SUIVANT la méthodologie définie dans le Chapitre 8, nous pouvons nous consacrer à la modélisation de la dynamique de population du babaçu. Nous souhaitons modéliser des processus interagissant au sein du « cycle de vie ». Ce chapitre expose d'abord les bases théoriques relatives à la modélisation de la dynamique de population avant d'exposer la construction et mise en œuvre du modèle utilisé pour le babaçu.

Sommaire

10.1	Modèle théorique	83
10.1.1	Modélisation de la dynamique de populations végétales	83
10.1.2	Inférence statistique en modélisation	86
10.2	Mise en œuvre	93
10.2.1	Méthodes d'analyse des données	93
10.2.2	Construction du modèle babaçu	96

Au sein de ce chapitre nous ferons apparaître progressivement un vocabulaire statistique méritant, de notre point de vue, d'être introduit et expliqué afin de nous éviter une quelconque confusion par la suite. Nous nous arrêterons notamment sur les définitions de variable aléatoire, variable latente, variable d'état, paramètre et hyper-paramètre.

La **variable aléatoire** est une variable dont les valeurs sont tirées de manière aléatoire (stochastique) à partir d'une loi de probabilité.

La *variable latente* est une variable aléatoire qui n'est pas observée directement, mais qui peut être modélisée afin d'améliorer le modèle.

La **variable d'état** est quant à elle une variable aléatoire particulière qui sert à décrire l'état du système.

Le **paramètre** est considéré au sein d'un modèle comme une valeur qui accomplit la fonction d'entrée du modèle. Les paramètres peuvent prendre la forme d'une valeur fixe tout au long du processus de modélisation mathématique ou des valeurs aléatoires issues d'une loi de probabilité.

L'**hyper-paramètre** désigne le(s) paramètre(s) fixe(s) de la loi de probabilité utilisée pour modéliser un paramètre aléatoire.

Ci-dessous nous présentons en guise de préambule un tableau regroupant les notations utilisées dans ce chapitre et qui ont été présentées dans la partie état de l'art sur la dynamique de population.

TABLEAU 10.1 – Notations se référant à la modélisation

Modèle théorique (Caswell, 2001)	
N_i	effectifs au stade i
R_i	recrutement au stade i
p_{ij}	probabilités de passage du stade i au stade j
pm_i	probabilité de mortalité au stade i
Notre modélisation	
Effectifs	
N_s	nombre de palmiers par stade s
$N_{l,s}$	nombre de palmiers par transect l et par stade s
$N_{l,s,m}$	nombre de palmiers morts par transect l et par stade s
$N_{l,s,0}$	nombre de palmiers restés dans le même stade par transect l et par stade s
$N_{l,s,1}$	nombre de palmiers passés au stade suivant par transect l et par stade s
Nf_i	nombre de fruits produits par palmier adulte i
Nft_l	nombre de fruits produits dans le transect l
$N_{l,1}^{new}$	nombre de nouveaux palmiers de stade 1 par transect l
N_{6a}	nombre de palmiers adultes (de stade 6) produisant des fruits
Probabilités	
$p_{s,m}$	probabilité de mortalité par stade s
$p_{s,0}$	probabilité de rester dans le même stade par stade s
$p_{s,1}$	probabilité de passage par stade s
$pl_{s,m}$	probabilité de mortalité par transect l et par stade s
$pl_{s,0}$	probabilité de rester dans le même stade par transect l et par stade s
$pl_{s,1}$	probabilité de passage par transect l et par stade s
p_{6a}	probabilité pour un palmier adulte de produire des fruits
pf	probabilité pour qu'un fruit se transforme en palmier de stade 1
Processus	
F_6	production de fruits
μ	nombre moyen de fruits par palmier adulte
σ^2	variance autour de μ
γ	espérance fixe
ω^2	variance autour de γ

Nous y avons fait figurer aussi bien des notations issues de la présentation plus générale de la modélisation de la dynamique de population que des notations rela-

tives au modèle que nous avons construit et dont nous ferons la présentation dans la seconde section de ce chapitre.

10.1 Modèle théorique

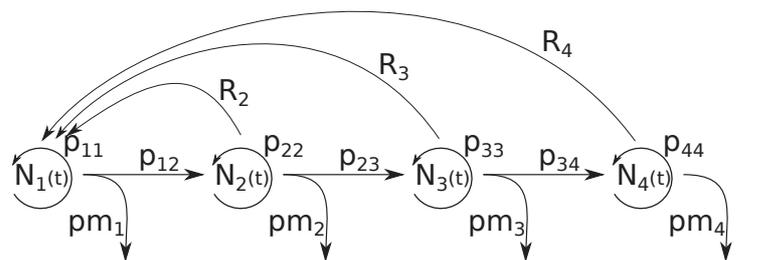
Dans cette section nous faisons une présentation rétrospective non exhaustive des approches théoriques en matière de modélisation de la dynamique de population. Nous souhaitons ainsi donner des clés de lecture pour la deuxième section consacrée à la présentation de notre modèle.

10.1.1 Modélisation de la dynamique de populations végétales

10.1.1.1 Basée sur le graphique du cycle de vie

Les caractéristiques biologiques des individus permettent leur attribution au sein d'une classe d'appartenance à un stade dans un graphique de cycle de vie. La FIGURE 10.1 présente un graphique du cycle de vie en quatre stades.

FIGURE 10.1 – Cycle de vie (adapté de Caswell (2001))



Pour le stade i et $j = 1$ à 4
 $N_i(t)$: effectifs au temps t
 R_i : recrutement

pm_i : probabilité de mortalité
 p_{ij} : probabilité de croissance

Les classes (stades) sont représentées par des nœuds et à chaque nœud correspond une **variable d'état** (N_i) au temps t .

Les passages entre ces différentes classes sont illustrées à l'aide d'une flèche reliant deux nœuds (changement de stade), alors que les individus qui restent dans le même stade sont représentés par des flèches faisant le tour de chaque nœud. Cela transcrit biologiquement un processus de croissance avec une probabilité (p_{ij}) entre le temps t et le temps $t + 1$.

Les mortalités correspondent à des individus qui sortent du système avec une probabilité (pm_i) entre le temps t et le temps $t + 1$ et sont stylisées sur le graphique à l'aide de flèches partant de chaque nœud vers le vide.

Enfin le recrutement (R_i), exprimé en nombre d'individus, est un processus généralement plus complexe qui exprime la capacité des individus du stade i de produire de nouveaux individus de stade 1 (Caswell, 2001). Ce processus peut impliquer, en fonction des particularités biologiques de chaque espèce, des individus soit d'un seul stade, soit de plusieurs stades. Sur ce graphique nous avons représenté un recrutement opéré à partir des stades 2, 3 et 4 (cf. FIGURE 10.1).

Le graphique de cycle de vie peut être utilisé comme point de départ pour la construction du modèle mathématique de dynamique de population. Nous allons donc présenter par la suite cette utilisation du graphique en modèle de dynamique.

10.1.1.2 Du modèle graphique au modèle matriciel

Modéliser la dynamique de population d'une espèce équivaut à vouloir connaître le nombre d'individus ($N_i(t+1)$) au pas de temps $t+1$ et pour chaque classe (stade) i . Pour faire ceci, nous nous appuyons sur des modèles mathématiques retranscrivant les processus biologiques de la FIGURE 10.1 explicités ci-dessus (cf. 10.1.1.1). Nous émettons l'hypothèse que ce vecteur d'état ($N_1(t+1), \dots, N_4(t+1)$) au temps $t+1$ est égal au vecteur d'état ($N_1(t), \dots, N_4(t)$) au temps t multiplié par une matrice de transition (A) qui intègre toutes les informations et probabilités présentées dans FIGURE 10.1. Ainsi nous obtenons l'équation 10.1.

$$N(t+1) = A * N(t) \quad (10.1)$$

Pour illustrer ceci nous présentons dans la FIGURE 10.2 la matrice de transition correspondant au graphique du cycle de vie présenté dans la FIGURE 10.1.

FIGURE 10.2 – Forme matricielle (adapté de Caswell (2001))

$$\begin{pmatrix} N_1(t+1) \\ N_2(t+1) \\ N_3(t+1) \\ N_4(t+1) \end{pmatrix} = \begin{pmatrix} p_{11} & R_2 & R_3 & R_4 \\ p_{12} & p_{22} & 0 & 0 \\ 0 & p_{23} & p_{33} & 0 \\ 0 & 0 & p_{34} & p_{44} \end{pmatrix} * \begin{pmatrix} N_1(t) \\ N_2(t) \\ N_3(t) \\ N_4(t) \end{pmatrix}$$

L'objectif en modélisation est de correctement renseigner les valeurs de la matrice de transition (A). Au sein de cette matrice carrée, dont la dimension est toujours égale au nombre de stades, la diagonale principale intègre les valeurs des probabilités pour que les individus croissent mais restent dans le même stade (p_{ij} avec $i = j$) entre t et $t+1$. La sous-diagonale comprend celles relatives au passage entre deux stades consécutifs (p_{ij} avec $i \neq j$). La première ligne de la matrice nous renseigne les valeurs relatives au recrutement (R_i).

Le pas de temps t à $t+1$ peut être journalier, hebdomadaire, mensuel, annuel... Ce pas de temps est défini par le modélisateur et est au minimum égal au laps de temps écoulé entre deux observations successives sur le terrain.

Cette représentation matricielle a donné le nom des modèles mathématiques matriciels de dynamique de population. Ces modèles ont été beaucoup utilisés pour la modélisation de populations végétales. Connus aussi sous le nom de modèles matriciels de projection de population, ils ont été introduits par Leslie (1945) et Lefkovich (1965). Ils sont utilisés pour la modélisation de populations animales à partir des années 1960, puis végétales à partir des années 1970 (Bierzychudek, 1999). L'avantage de ces modèles est qu'ils peuvent être mis en place sans avoir recours aux outils informatiques puisqu'ils modélisent la dynamique de population en se basant sur une approche analytique de calcul en espérances ils sont déterministes malgré des dynamiques stochastiques.

La dynamique d'une population définie par un modèle déterministe peut être estimée simplement et de plus nous pouvons en déduire analytiquement des propriétés asymptotiques (quand le temps tend vers l'infini). Ainsi, durant les trente dernières années du XX^e siècle, la modélisation de la dynamique de population d'espèces végétales basée sur les modèles déterministes s'est démocratisée au sein de la communauté scientifique internationale (Bierzychudek, 1999).

Caswell (2001) indique que les analyses, basées sur les matrices de transition, doivent être menées comme des situations traduisant ce qui **pourrait se produire si** les paramètres restaient inchangés (caractère stationnaire) et non comme des situations montrant ce qui **va se produire** (Bierzychudek, 1999; Caswell, 2001). Sous cette hypothèse une finalité des modèles matriciels de transition est de calculer un taux de croissance de la population λ (cf. section 7.1 - page 45) utilisant les données de la matrice (Bierzychudek, 1999; Caswell, 2001). Une limite de cette approche de calcul du λ est qu'il ne varie pas dans le temps, ce qui est rarement le cas dans la nature. Ceci indique que ce type de modèles sont efficaces qu'en situation de stabilité temporelle des paramètres. La manière dont l'information est synthétisée par ces modèles suppose que le système soit stationnaire, le même état étant observé à tout instant t .

Les paramètres sont souvent estimés à partir des observations limitées en nombre (Bierzychudek, 1999). Le plus souvent, il s'agit d'observations de moins de 3 années consécutives (Bernal, 1998; Bullock, 1980; Olmsted et Alvarez-Buylla, 1995). D'après Bierzychudek (1999) ce laps de temps est insuffisant pour pouvoir estimer s'il y a des variations temporelles, même au sein d'environnements stables.

10.1.1.3 Révolution informatique

Le progrès des outils informatiques a grandement facilité et ouvert la voie aux modèles stochastiques où les paramètres de la matrice de transition sont aléatoires et où les trajectoires (dynamiques) ne sont plus calculées en espérance mais simulées.

Le développement de modèles matriciels stochastiques a permis l'introduction de l'aléa démographique (régé par des facteurs internes : densité-dépendance) et/ou environnemental (régé par des facteurs externes : climat) au sein du modèle (Kaye

et Pyke, 2003). Par exemple, si l'utilisation d'au moins deux matrices de passage comportant des données distinctes consécutives de trois observations : un premier passage de t à $t + 1$ et un second passage de $t + 1$ à $t + 2$ (t étant une unité de mesure du temps : jour, mois, année...) amène a deux valeurs estimées distinctes pour chaque paramètre. Le modélisateur peut en conséquence émettre l'hypothèse que ceci est dû aux variations induites par l'aléa environnemental (Fieberg et Ellner, 2001). Les paramètres doivent en conséquence être considérés comme aléatoires ou dépendant de variables latentes.

La variété de ces modèles est très grande : Fieberg et Ellner (2001), lors de l'inventaire des modèles matriciels stochastiques, illustrent cette diversité à l'aide d'un tableau récapitulatif dénombrant 28 modèles proposés en biologie de la conservation autour d'espèces différentes.

Suite à l'étude de nos données brutes de 2013 (cf. TABLEAU 9.4 - page 70), nous avons constaté qu'il existait une forte fluctuation des variables d'état (d'importants écarts entre les différents transects) et en avons conclu à un environnement pas stable. Ceci nous a conduit à écarter la possibilité de travailler en espérances (modèle déterministe). Nous nous sommes orienté vers une modélisation stochastique en choisissant d'introduire de l'aléa au sein de notre modèle : les paramètres ont été considérés comme aléatoires. Par contre, nous avons traité les observations du vecteur d'état comme étant sans erreur de mesure.

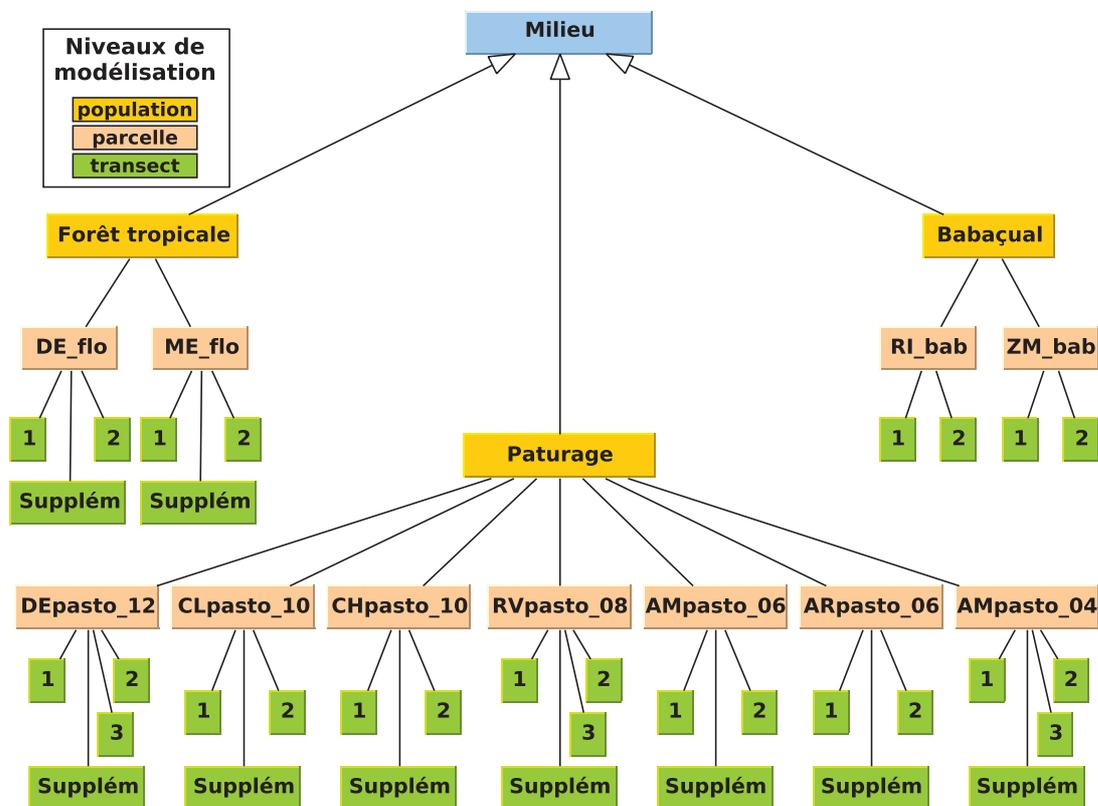
10.1.1.4 Modèle hiérarchique

Les technologies ont aussi ouvert la voie à l'exploration de modèles multi-strates appelés modèles hiérarchiques. Ils sont construits dans l'objectif de résoudre des problèmes complexes multi-dimensionnels en les décomposant en séries de structures (composantes) plus simplifiées reliées par des estimations de probabilités (Parent et Rivot, 2012). Nos données sont par définition organisées au sein d'une structure hiérarchique. La FIGURE 10.3 illustre les trois niveaux hiérarchiques (la population (ensemble de l'échantillon), la parcelle et le transect) que nous avons décrit lors de la présentation des données dans la section 9.3 (page 66) et résumés dans le TABLEAU 9.2 (page 67). En fait, cette hiérarchie permet le transfert de l'information entre les différents transects lors du processus d'estimation. Ainsi, d'après la FIGURE 10.3 nous constatons que l'information provenant d'un transect par le biais hiérarchique est utilisée pour l'estimation au sein de n'importe quel autre transect grâce aux liens qui les relient à travers la strate des parcelles et celle de la population.

10.1.2 Inférence statistique en modélisation

Il nous semble important d'attirer l'attention du lecteur sur la notion de statistique inférentielle. À la différence de la statistique descriptive, qui comme son nom l'indique se restreint à la description de phénomènes et « dont le but est de décrire et

FIGURE 10.3 – Niveaux hiérarchiques : les transects (premier niveau) au sein de parcelles (second niveau) ont une occupation du sol caractérisant des différences dans la dynamique de l'espèce (troisième niveau - forêt tropicale, pâturage ou babaçal).



de résumer de façon formalisée » (Py, 1996, p. 10), l'objectif de la statistique inférentielle est d'aboutir à une analyse en induisant « les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population » (Lethielleux, 1998, p. 1). L'inférence statistique est nécessaire pour estimer les paramètres inconnus de la dynamique de population.

La probabilité est une notion qui trouve une place centrale au sein de l'inférence statistique. Le terme de « probabilité » est très largement utilisé aussi bien dans le milieu scientifique que lors d'une discussion tout à fait anodine. Cependant il existe une importante confusion entre probabilité et fréquence. Nous souhaitons donc les définir clairement afin de pouvoir les différencier par la suite dans ce manuscrit. La fréquence exprime un rapport, c'est une proportion d'observations qui peut être mesurée dans des unités de mesure très variables (unités de quantité, de temps, de superficie...). La probabilité est la mesure d'une incertitude sur un événement. Elle est mesurée sur une échelle allant de 0 (événement improbable) à 1 (événement absolument sûr). La fréquence devient une mesure d'incertitude (probabilité) si et

seulement si nous procédons à un tirage basé sur cette fréquence et suivant une loi de probabilités (Rouanet, 2008).

Il existe en statistiques deux approches complémentaires mais différentes d'estimation des paramètres d'un modèle (les probabilités de passage dans notre contexte).

La démarche fréquentiste est construite, autour de l'estimation d'une probabilité à partir de la fréquence d'apparition d'un événement (fréquences observées lors de répétitions consécutives). Ce cadre analytique est basé sur l'estimation des probabilités uniquement à partir des données.

Le formalisme bayésien est construit autour d'une conception différente de l'estimation. Toutes les inconnues du modèle (paramètres, variables latentes issues de la structure hiérarchique) peuvent prendre la forme de variables aléatoires suivant des lois de distribution. Cette méthode impose de définir une distribution *a priori* pour les inconnues. Cette loi *a priori* doit traduire la connaissance sur chaque inconnue. Elle permet un cadre mathématique plus rigoureux (Rouanet *et al.*, 1998). Pour les travaux de cette thèse nous avons privilégié cette approche bayésienne, puisqu'en nous basant sur la nature et la structure de nos données nous avons considéré qu'elle pourrait fournir un cadre d'estimation qui permet l'apprentissage au cours du temps (Parent et Rivot, 2012) grâce aux données de terrain acquises successivement entre 2013 et 2016. Dans la définition de l'*a priori* nous pourrions intégrer l'expertise (acquise au cours du temps) ou à défaut, utiliser un *a priori* non informatif (toutes les valeurs ont la même chance d'être la bonne).

L'estimation dans un cadre bayésien utilise comme principe le théorème de Bayes que nous développons ci-après.

Le révérend Bayes est le fondateur du théorème portant son nom. Il a introduit la notion de probabilités conditionnelles durant la deuxième moitié du XIX^e siècle. Son idée, très répandue depuis, est basée sur le principe de dépendance relationnelle entre des probabilités connexes. En 1774 Laplace aboutit à la même formalisation des probabilités que Bayes. Il publie ses travaux sous le titre d'*Essai philosophique sur les probabilités*. Lethielleux (1998) propose une définition du fonctionnement des probabilités conditionnelles, résumant la logique de ces deux auteurs.

La probabilité pour que l'événement A soit réalisé sachant que l'événement B est réalisé (noté $p(A|B)$) est égale à la probabilité que l'événement A se réalise (noté $p(A)$) multipliée par la probabilité que l'événement B soit réalisé sachant que l'événement A est réalisé (noté $p(B|A)$) le tout divisé par la probabilité que l'événement B soit réalisé (noté $p(B)$) (cf. équation 10.2).

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)} \quad (10.2)$$

Le fonctionnement des probabilités conditionnelles est appliqué par analogie dans le cadre d'une analyse statistique par inférence (Gelman *et al.*, 2003). Au sein de l'analyse bayésienne nous souhaitons obtenir *a posteriori* une distribution de densité

de probabilité des inconnues¹ (θ) sachant les données (y) (notée $p(\theta|y)$) à partir de la distribution de densité de probabilité *a priori* (θ) et de la distribution de l'échantillon ou vraisemblance ($y|\theta$). Ci-dessous nous présentons les différentes étapes de l'analyse bayésienne :

1. *a priori* (θ) : connaissances - expertises (informatif/non-informatif)
2. données (y)
3. mettre à jour l'*a priori* grâce aux données : *a posteriori* ($\theta|y$)
4. analyser : distribution *a posteriori* (moyenne, médiane, écart-type...)
5. utiliser éventuellement la loi *a posteriori* comme loi *a priori* sur les données

Autrement dit, le posterior est proportionnel (\propto) au prior augmenté (\times) de la vraisemblance :

$$\text{posterior} \propto \text{prior} \times \text{vraisemblance} \quad (10.3)$$

Retranscrit en termes probabilistes, cette équation prends donc la forme suivante (Gelman *et al.*, 2003) :

$$p(\theta|y) \propto p(\theta) \times p(y|\theta) \quad (10.4)$$

L'un des intérêts de la démarche bayésienne par rapport à la démarche fréquentiste est notamment de pouvoir introduire au sein du modèle des connaissances antérieures déjà acquises. Cependant, dans le cas d'une inférence bayésienne, la plus grande attention doit être portée quant au choix de la distribution initiale pour ces croyances *a priori*.

Si nous ne disposons d'aucun savoir déjà acquis, nous devons veiller à ce qu'un *a priori* non-informatif soit mis en place. Dans ce cas, la difficulté consiste dans le fait de choisir une distribution de probabilités « non-informative » reflétant ce manque de connaissances. Ainsi la distribution *a posteriori* n'exprimera que l'apport des données et non l'apport des données augmenté de l'apport de nos croyances *a priori* (Rouanet, 2008). La proposition de l'utilisation d'une distribution *a priori* non-informative est attribuée à Laplace et développée par Jeffreys (1961) pour indiquer un « état d'ignorance » concernant les paramètres du modèle (Rouanet *et al.*, 1998).

La quantité de l'information contenue dans les données doit également être prise en compte. Par exemple, un *a priori* informatif exercera davantage d'influence, sur le résultat *a posteriori*, en présence de données peu conséquentes qu'en présence de données conséquentes (Congdon, 2001).

Très souvent les lois *a posteriori* sont connues analytiquement mais à une constante près (la constante d'intégration, qui est souvent non calculable). Metropolis *et al.* (1953) proposent une première version d'un algorithme éponyme capable de simuler une loi à la constante près (pas besoin de connaître la valeur de la constante). Presque vingt ans après, Hastings (1970) publie l'extension de cet algorithme qui est

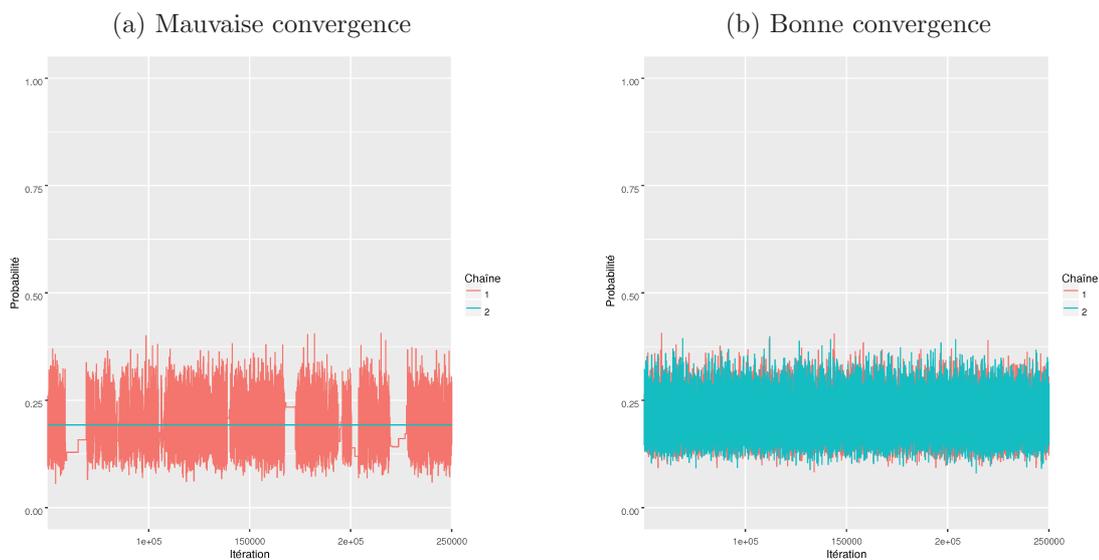
1. variables latentes, paramètres aléatoires

aujourd'hui communément appelé Metropolis-Hastings. Ces travaux ont été utilisés en inférence bayésienne pour simuler des valeurs selon la loi *a posteriori* (même si la constante d'intégration est incalculable/inconnue). Par exemple, Geman et Geman (1984) ont développé l'échantillonneur de Gibbs basé sur l'algorithme de Metropolis-Hastings. Depuis cette période, l'utilisation de l'algorithme appelé *Markov Chain Monte Carlo* (MCMC)² se généralise au domaine de la statistique et de l'analyse du signal pour être utilisé au sein du cadre analytique bayésien.

L'algorithme proposé converge vers la vraie loi *a posteriori* au bout d'un temps de chauffe. La convergence des chaînes de Markov est un important indicateur du bon déroulement du processus d'estimation. En fait, la convergence d'une chaîne de Markov est nécessaire afin de nous assurer que l'estimation est faite sur des données simulées selon la bonne loi *a posteriori*.

Nous devons nous assurer que cette convergence est faite indépendamment des conditions initiales. Autrement dit, nous devons veiller à ce que la condition dite d'ergodicité du processus d'estimation soit accomplie (Robert, 1996). Un processus est dit être ergodique si et seulement si « . . . toutes les transitions d'un état vers un autre sont possibles, c'est-à-dire, tous les états peuvent être atteints de n'importe où en un seul pas. . . » (Solaiman, 2006, p. 175).

FIGURE 10.4 – Convergence des chaînes de Markov



Il existe des diagnostics de la convergence pouvant nous renseigner si celle-ci a été atteinte au bout d'un nombre d'itérations fixé par le modélisateur. Ce diagnostic est

2. L'algorithme de simulation repose sur le principe des chaînes de Markov qui simulent une nouvelle valeur uniquement à partir de la valeur simulée au pas précédent (pas de mémoire du processus).

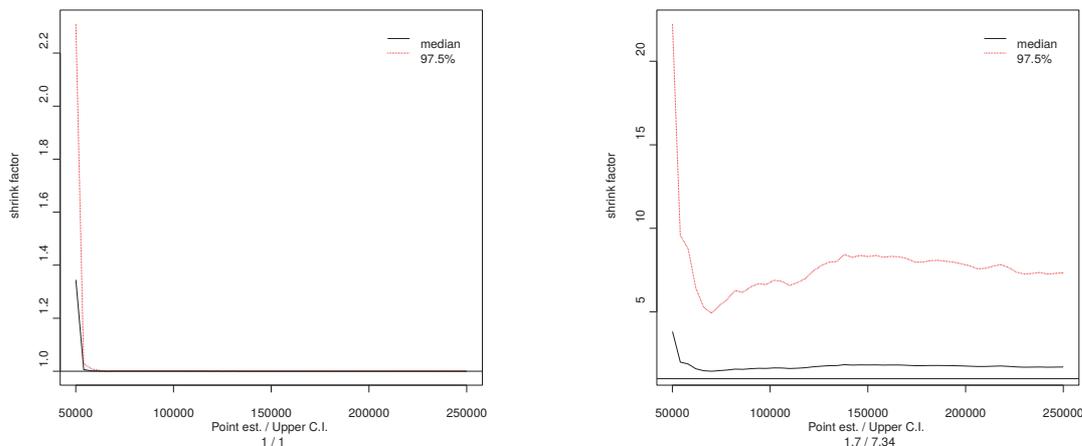
primordial avant toute simulation ultérieure. En analyse bayésienne il est difficile de juger si une chaîne a réellement atteint l'état de convergence. Par contre, il est très facile de dire si une chaîne ne l'a pas atteint. Pour faire ceci nous nous basons sur des graphiques illustrant la trace des valeurs prises par la variable aléatoire durant les simulations. Une trace plate ou une trace effectuant des « sauts » signifie que la chaîne n'a pas atteint sa convergence.

Sur la FIGURE 10.4a nous avons fait apparaître deux chaînes de Markov qui n'ont pas atteint la convergence. Nous pouvons constater que la chaîne numéro 2 (en bleu) n'a exploité qu'une seule valeur de l'espace probabiliste lors de simulation. La chaîne numéro 1 (en rouge), quant à elle, a davantage exploré cet espace, mais a effectué des « sauts » en se positionnant sur une seule valeur unique durant certaines itérations (il s'agit de plusieurs séquences de simulation de valeurs répétées). Il existe aussi des cas dans lesquels les chaînes de Markov n'exploitent jamais le même espace et apparaissent donc sur le graphique comme deux lignes distinctes (Lunn *et al.*, 2013). La FIGURE 10.4b illustre au contraire un exemple où nous sommes amenés à croire que la convergence a été atteinte. Les deux chaînes ont exploré l'espace probabiliste et se recouvrent mutuellement.

FIGURE 10.5 – Test de Gelman et Rubin (1992)

(a) Bon test

(b) Mauvais test



Une autre méthode de vérification d'une convergence est l'utilisation d'un test proposé par Gelman *et al.* (2003). Les auteurs proposent de tester la convergence en se basant sur au moins deux chaînes de Markov exécutées en parallèle. Chaque chaîne est bien évidemment initiée avec des valeurs différentes tout en respectant la vraie distribution *a posteriori*. Cette méthode, proposée initialement par Gelman et Rubin (1992), repose sur le rapport des variances intra-chaîne et entre-chaînes et fait

figurer les valeurs associées à la médiane et celles associées à l'intervalle de crédibilité de 97.5%. Les auteurs sont partis du postulat que si les chaînes ont convergé, alors leurs comportements doivent être similaires. L'indication qu'une chaîne a convergé est retranscrite par une valeur du test se rapprochant de 1 (dans tous les cas une valeur $< 1, 1$).

La FIGURE 10.5 illustre deux cas différents du test de Gelman-Rubin. Dans le premier cas, la convergence est atteinte : nous pouvons observer que la médiane s'est rapprochée de 1 dès le début des estimations ; l'intervalle de crédibilité correspondant à 97.5% est très proche de la médiane (FIGURE 10.5a). Dans le second cas, nous observons que la médiane n'a jamais atteint la valeur de référence (FIGURE 10.5b).

Nous nous sommes assurés de la convergence des chaînes de Markov de notre modèle. Les graphiques présentant la convergence des chaînes de Markov utilisées pour l'estimation, ainsi que le test de Gelman-Rubin de chacun de nos paramètres sont consultables dans l'Annexe C (Convergence des chaînes de Markov).

10.1.2.1 Évaluation du modèle

La *posterior predictive p-value* (PPP) est un indicateur de fiabilité d'un modèle. L'idée centrale derrière la PPP est de comparer de combien les valeurs simulées à partir du modèle s'écartent des données et dans quel sens. Cet indicateur sert aussi à la comparaison de modèles entre eux et démontre leur niveau d'efficacité à fournir des prédictions fiables et réalistes. La comparaison s'opère entre les données observées et la distribution prédictive *a posteriori*. La *p-value* est définie comme la probabilité de combien les données répliquées (y^{rep}) s'éloignent des données observées (y^{obs}) (Gelman *et al.*, 1996).

$$ppp = Pr(D(y^{rep}, \theta|y^{obs}) \geq D(y^{obs}, \theta|y^{obs})) \quad (10.5)$$

Dans cette équation D est une mesure de l'erreur moyenne³ calculée séparément pour les données répliquées et ensuite pour les données observées. Ce calcul est réalisé en utilisant les valeurs *a posteriori* de θ . Les données répliquées y^{rep} sont obtenues à partir de la distribution prédictive $\int \pi(y^{rep}|\theta)\pi(\theta|y^{obs})d\theta$ (Steinbakk et Storvik, 2009).

Une seconde méthode d'évaluation de la qualité prédictive du modèle est la méthode dite de validation croisée. Elle consiste dans le fait de faire fonctionner le modèle sur une partie des données (généralement au moins égale à la moitié) et d'utiliser les valeurs estimées pour prédire les valeurs des variables d'état de la partie non utilisée. Enfin, une comparaison entre les données prédites et les données observées est effectuée, afin de détecter les écarts.

3. la moyenne des écarts au carré entre les données répliquées ou observées et leur moyenne (c'est le calcul du carré moyen de l'erreur). Cette moyenne est ensuite divisée par la variance pour standardiser.

10.2 Mise en œuvre

10.2.1 Méthodes d'analyse des données

Ces travaux de recherche s'inscrivent dans le cadre d'une étude expérimentale. Ils s'intègrent dans le champ scientifique de l'induction bayésienne, au sein de laquelle tous les choix sont explicites : nettement et complètement formulés, sans aucune ambiguïté et sans aucun doute possible (Lecoutre, 2006).

Afin de pouvoir modéliser la dynamique de population du palmier, nous avons tout d'abord mobilisé un certain nombre d'outils d'analyse.

10.2.1.1 Graphique du cycle de vie du babaçu

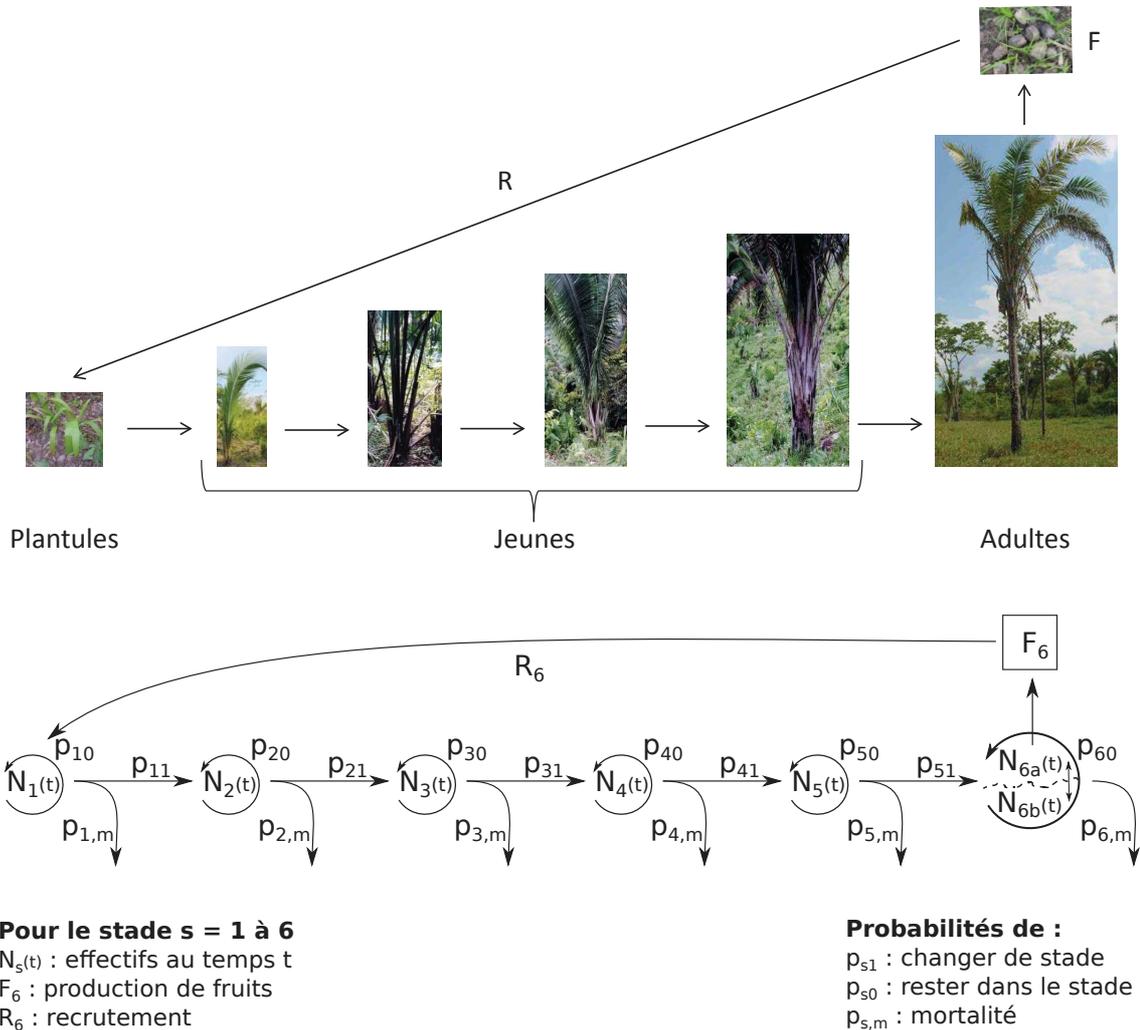
Nous avons élaboré et proposons ici le diagramme du cycle de vie du palmier babaçu basé sur les définitions des stades biologiques de celui-ci définies dans la section 9.1 (page 59). La FIGURE 10.6 donne une image schématisée de celui-ci. De surcroît, cette figure présente un premier aspect de la dynamique de population de l'espèce puisqu'elle met aussi en évidence l'existence de certains mécanismes biologiques.

Notamment, les trois processus majeurs biologiques ressortent à la lecture de cette figure. La croissance est stylisée soit par le passage du stade s au stade $s + 1$ (p_{s1}), soit par le fait que les palmiers restent dans le même stade (p_{s0}) l'année n à l'année $n + 1$. La mortalité, quant à elle, concerne l'intégralité des stades, les palmiers du stade s peuvent mourir avec une probabilité $p_{s,m}$. Le recrutement (R_6) s'effectue par l'intermédiaire de la production de fruits (F_6). Il ne concerne que certains palmiers du stade 6. Ce sont en effet les palmiers disposant uniquement d'inflorescences féminines et donc d'infrutescences par la suite. En fonction de l'état de reproduction du babaçu, les palmiers adultes vont avoir des inflorescences masculines et/ou féminines. Dans le cas d'un palmier adulte n'ayant que des inflorescences masculines, ce palmier ne fournira pas de fruits durant l'année. Le processus de production d'inflorescences est très complexe, car le babaçu est une plante monoïque, c'est-à-dire présentant les deux sexes sur le même palmier. De plus, les inflorescences sont parfois hermaphrodites lorsqu'elles comportent des fleurs féminines et masculines. Il arrive aussi qu'elles soient exclusivement composées de fleurs masculines.

Comme illustré dans la FIGURE 10.6 les palmiers adultes peuvent basculer de producteur de fruits (N_{6a}) à non producteurs (N_{6b}) et vice-versa d'une saison de reproduction à une autre. Dès lors, nous avons considéré qu'uniquement les palmiers appartenant au sous groupe N_{6a} fourniront des fruits.

10.2.1.2 Outils mis en œuvre

La modélisation et l'analyse des données se sont appuyées sur différents outils de l'analyse bayésienne. Nous avons utilisé les logiciels « *Bayesian inference Using Gibbs Sampling* » (BUGS) et « *Just Another Gibbs Sampler* » (JAGS) pour cette

FIGURE 10.6 – Cycle de vie du babaçu (adapté de Mitja *et al.* (sous presse))

Uniquement une partie des palmiers adultes (N_{6a}) donne des fruits et la partie complémentaire des adultes (N_{6b}) n'en donne pas. Cependant, d'une année à une autre, le statut d'un palmier peut varier.

analyse. Nous avons choisi ces outils puisqu'ils s'inscrivent, tout comme PostgreSQL/PostGIS, dans une démarche de « code source ouvert » (open source). Ce sont également des outils multi-plateformes interopérables avec le logiciel PostgreSQL/PostGIS. Nous avons assuré l'extraction des données de la base de données avec le logiciel R. Ce même logiciel, via le package R2OpenBUGS, nous a aussi servi à l'exportation des données dans un format conforme pour la modélisation bayésienne (logiciels BUGS et JAGS).

Il existe plusieurs packages dans R qui proposent pratiquement les mêmes fonctionnalités de l'analyse de la convergence des chaînes. Nous avons appuyé notre analyse sur le package Coda ou encore ggcmc.

L'étape suivante a consisté dans la définition de nos choix de modélisation.

10.2.1.3 Échelon d'analyse : la cohorte

Nous disposons de données au niveau du palmier (cf. section 9.3 - page 66). Ce type de données permet, via des modèles individu-centrés, d'étudier l'impact que le comportement d'un seul individu peut avoir sur l'évolution de la population totale. Les modèles individu-centrés se focalisent sur l'analyse de la variabilité des caractéristiques de chaque individu (dynamique individuelle). Un modèle individu-centré est surtout intéressant quand les individus sont en interaction. Or, dans notre contexte d'étude nous supposons qu'il n'y a pas d'interaction entre individus.

Aussi, nous avons considéré qu'une approche de modélisation basée sur le suivi pluriannuel des mêmes individus était mieux adaptée à l'étude de la dynamique de population du babaçu. Cette approche a l'avantage de nous fournir des informations non plus au niveau du comportement individuel, mais au niveau populationnel. Ce type de modélisation peut être basé directement sur le cycle de vie de la plante et permet de dégager la tendance centrale et la variabilité au sein de chaque compartiment. Le cycle de vie du babaçu est défini autour de ses six stades biologiques (cf. 10.2.1.1). Par souci d'uniformité de notre approche, nous avons décidé que chaque compartiment du modèle correspondrait à un stade biologique. Nous avons donc basé notre analyse sur un modèle à six compartiments.

Ainsi l'estimation des paramètres du modèle est faite pour chacun des stades biologiques.

10.2.1.4 Variables d'état

Nous avons considéré au sein de notre modèle un vecteur de variables d'état qui sont des variables aléatoires décrivant l'évolution temporelle de la dynamique de population du babaçu. Ces variables assurent la relation entre les processus biologiques et les processus statistiques modélisés.

Aussi, nous avons défini comme variables d'état de notre modèle le nombre total de palmiers par stade (N_1, \dots, N_6) et le nombre de fruits (Nft). Ces variables d'état sont définies au niveau de la population : on parle de variables de p -état (avec p venant de population) Caswell (2001). Cette définition s'appuie sur la présomption préalable que tous les palmiers sont égaux au sein d'un même stade biologique et que leurs comportements sont indépendants les uns des autres.

10.2.1.5 Variables d'intérêt

Nous avons défini les variables du modèle en fonction de nos objectifs : l'analyse des trois processus biologiques (mortalité, croissance et recrutement) régissant la dynamique de population du palmier babaçu. Nous nous sommes donc intéressés aux paramètres aléatoires relatifs à ces processus : le taux de mortalité par stade

biologique, le taux de croissance (passage d'un stade à un autre) et le taux de recrutement.

En amont de ces analyses, nous avons dû poser des hypothèses sur la biologie de l'espèce et sur la modélisation.

10.2.1.6 Hypothèses sur la biologie

Nous avons considéré que le cycle de vie du babaçu pouvait être structuré à l'aide d'une définition (au sens biologique) de ses stades de développement. Cette définition a été étayée dans la section 9.1 (page 59). Nous avons donc proposé l'hypothèse qu'*a priori* les palmiers ont un comportement indépendant au sein d'un même stade et qu'il n'existe aucun processus de densité-dépendance au sein de la population des palmiers babaçu. Autrement dit, que la mort ou le passage du stade s au stade $s + 1$ d'un palmier du stade s n'influence aucunement la mortalité ou le passage du stade s au stade $s + 1$ d'un autre palmier du même stade s . Nous avons considéré qu'au sein d'un même stade s tous les palmiers ont la même probabilité de mortalité et de passage du stade s au stade $s + 1$.

Nous n'avons pas considéré de processus lié à la compétition au sein des pâturages entre le babaçu et les autres espèces.

Nous avons exploité la variabilité spatiale à travers la localisation spatiale des transects. Les transects se différencient en fonction de caractéristiques telles que la nature du couvert végétal, la productivité liée à la qualité du sol, les pratiques agricoles de chaque propriétaire. . .

La variabilité temporelle (effet année) n'est pas explicitement modélisée, puisque nous ne disposons pas de variables climatiques. Cependant, ce processus est implicitement introduit au sein du modèle grâce à la prise en compte de deux passages consécutifs (2013 - 2014 et 2014 - 2015) lors de l'estimation.

10.2.1.7 Hypothèses de modélisation

Ces hypothèses se rapportent à la démarche de transcription des processus biologiques en processus de modélisation. Afin de faciliter la modélisation de la dynamique de population, nous avons assimilé les palmiers enregistrés comme « perdus » à des palmiers enregistrés comme « morts ». Ainsi le nombre de palmiers morts (réellement constatés morts sur le terrain) a été augmenté du nombre de palmiers dits « perdus ».

10.2.2 Construction du modèle babaçu

La classification générale des modèles a été abordée en introduction de cette partie dans la section 8.1.2 (cf. FIGURE 8.1 - page 57). Ici nous approfondissons cette approche à travers la présentation plus détaillée de notre modèle mathématique de dynamique de population du babaçu, tout en l'inscrivant au sein du contexte des

modèles mathématiques/statistiques plus génériques. Cette approche présente à nos yeux un double avantage :

- recenser et classer les modèles mathématiques/statistiques que nous pouvons développer,
- positionner notre modèle de dynamique de population au sein de cette classification.

La FIGURE 10.7 reflète cette réflexion tout en résumant et structurant notre approche. Comme nous pouvons l’observer sur cette figure, notre modèle est inscrit au sein de la famille des modèles stochastiques. L’aspect hiérarchique de nos données (transect, parcelle, population) et de leur fréquence annuelle d’obtention nous ont orienté vers un modèle hiérarchique à temps discret.

Notre réflexion s’est ensuite orientée vers la sélection des lois statistiques les mieux adaptées à notre modèle. Nous avons opté pour des lois appartenant à la famille des lois conjuguées. Le principe des lois conjuguées est largement appliqué au sein de la démarche bayésienne. Il s’agit de l’utilisation conjointe de deux lois statistiques dont les domaines de densité sont compatibles : la première est utilisée pour l’estimation de l’*a priori*, alors que la deuxième est appliquée au niveau de la vraisemblance. Pour la construction de notre modèle nous avons eu recours à trois de ces couples de lois : *Dirichlet-Multinomiale*, *Beta-Binomiale* et *Normale-Normale*.

Les choix possibles au sein d’une loi *Multinomiale* sont exhaustivement complémentaires. C’est-à-dire que la somme des probabilités de cette loi (cas possibles) doit être égale à un. Donc pour faire fonctionner cette loi de probabilités nous avons besoin de calculer des probabilités dont la somme est égale à un. Nous avons eu recours à la loi de probabilités de *Dirichlet* puisqu’elle satisfait à cette contrainte. En analyse bayésienne la difficulté provient souvent du calcul de la loi *a posteriori*. Il est courant dans ce cas de faire recours aux lois conjuguées, car dans ce cas la loi *a priori* et la loi *a posteriori* ont la même forme. Donc si nous connaissons la forme de la loi *a priori* nous pouvons très facilement déduire la forme de la loi *a posteriori*. La loi de *Dirichlet* est une généralisation de la distribution *Beta* qui est pour sa part conjuguée avec la loi *Binomiale*.

10.2.2.1 Choix du niveau hiérarchique : transect *vs* parcelle

Notre modèle s’inscrit au sein du cadre analytique des modèles hiérarchiques. Ceci signifie que nous avons proposé une structure hiérarchique du modèle en nous basant sur la hiérarchie de nos données. Nos données ont été collectées au niveau de la parcelle : l’échantillonnage était réalisé en fonction de l’âge de celle-ci (cf. TABLEAU 9.2 - page 67). Chacune des parcelles appartient et est gérée par un seul propriétaire. De ce fait, nous avons émis l’hypothèse de départ que chaque parcelle est censée refléter les spécificités relatives à la gouvernance de son propriétaire.

Cependant chaque parcelle est composée de plusieurs transects (2 ou 3) qui *a priori* assuraient le rôle de répétition au niveau de la parcelle. Nous disposions donc dès le début de ce travail de recherche, du choix de la définition du niveau hiérar-

avons estimé qu'une analyse portée à l'échelon du transect nous permettra de mieux exploiter la variabilité contenue dans l'information, alors qu'une analyse au niveau de la parcelle sous-estimerait cette variabilité.

Le choix de ne pas analyser les données à l'échelon de la parcelle, et donc de ne pas intégrer ce niveau dans le modèle hiérarchique, pourra être remis en cause dans des études ultérieures. Ceci dépendra de la présence d'éventuelles informations indiquant une différenciation par parcelle.

Nous avons orienté notre réflexion qui s'inscrit dans une démarche systémique vers la construction de plusieurs modèles, et donc la comparaison des avantages et inconvénients de chacun d'entre eux.

10.2.2.2 Modèle multinomial de croissance

Le modèle multinomial permet un tirage parmi un nombre variable de possibilités. Notre modèle est construit pour un tirage parmi trois possibilités : mourir (m), rester dans le même stade (0), passer au stade suivant (1). Connaissant pour chaque transect l et stade s le nombre de palmiers $N_{l,s}$ et le vecteur des probabilités de mourir $p_{l,s,m}$, de rester dans le même stade $p_{l,s,0}$ et de passer au stade suivant $p_{l,s,1}$, nous répartissons le nombre total de palmiers $N_{l,s}$ dans trois groupes de palmiers correspondants à chaque possibilité $N_{l,s,m}$, $N_{l,s,0}$ et $N_{l,s,1}$:

$$\begin{aligned} & \text{pour } l = 1, \dots, n_l \text{ et } s = 1, \dots, 5 \\ (N_{l,s,m}, N_{l,s,0}, N_{l,s,1}) & \sim \mathcal{M}(N_{l,s}, (p_{l,s,m}, p_{l,s,0}, p_{l,s,1})) \end{aligned} \quad (10.6)$$

$$(N_{l,6,m}, N_{l,6,0}) \sim \mathcal{M}(N_{l,6}, (p_{l,6,m}, p_{l,6,0})) \text{ pour } l = 1, \dots, n_l \quad (10.7)$$

Les probabilités par transect l et par stade s : $p_{l,s,m}$, $p_{l,s,0}$ et $p_{l,s,1}$ sont générées à partir des probabilités par stade de mourir $p_{s,m}$, de rester dans le stade $p_{s,0}$ et passer au stade suivant $p_{s,1}$. Pour cela nous utilisons une loi de Dirichlet de paramètres $p_{s,m} \cdot n_s$, $p_{s,0} \cdot n_s$ et $p_{s,1} \cdot n_s$:

$$\begin{aligned} & \text{pour } l = 1, \dots, n_l \text{ et } s = 1, \dots, 5 \\ (p_{l,s,m}, p_{l,s,0}, p_{l,s,1}) & \sim \text{Dir}(p_{s,m} \cdot n_s, p_{s,0} \cdot n_s, p_{s,1} \cdot n_s) \end{aligned} \quad (10.8)$$

$$(p_{l,6,m}, p_{l,6,0}) \sim \text{Dir}(p_{6,m} \cdot n_6, p_{6,0} \cdot n_6) \text{ pour } l = 1, \dots, n_l \quad (10.9)$$

La loi de *Dirichlet* utilise un paramètre de taille n_s qui traduit la « confiance » que nous avons dans les estimations des probabilités $p_{s,m}$, $p_{s,0}$ et $p_{s,1}$.

Recrutement

La totalité de la complexité du processus de recrutement n'a pas été modélisée. Le processus de passage entre inflorescences féminines/masculines et infrutescences

est pertinent quant à l'étude biologique du babaçu. Malgré ceci, ce processus très complexe n'a pas été pris en compte dans le modèle : nous avons considéré que son intégration représentait un risque de sur-paramétrisation avec des variables aléatoires entachées d'erreurs de mesures. Nous avons estimé que sa modélisation introduirait plus de bruit dans le modèle de recrutement qu'elle n'apporterait de réponses. Nous avons donc décidé de ne pas l'introduire au sein de la modélisation du recrutement. Ainsi, le processus de recrutement est modélisé à l'aide de deux processus enchaînés : la production de fruits par les palmiers adultes puis la transformation d'un fruit en un palmier de stade 1.

Cependant, comme nous l'avons vu dans la section [Graphique du cycle de vie du babaçu \(10.2.1.1\)](#) tous les palmiers adultes ne donnent pas des infrutescences et donc ne sont pas susceptibles de produire des fruits : uniquement les palmiers de type $6a$ l'année n peuvent produire des fruits entre les années n et $n + 1$. La variable aléatoire décrivant le nombre de palmiers adultes producteurs de fruit (N_{6a}) suit une loi de probabilité binomiale de paramètres : le nombre total de palmier adultes N_6 et la probabilité pour un palmier adulte de produire des fruits p_{6a} :

$$N_{6a} \sim \mathcal{B}(N_6, p_{6a}) \quad (10.10)$$

p_{6a} est donc un paramètre que nous devons estimer.

Recrutement : processus 1

Nous proposons de modéliser le nombre de fruits par palmier adulte producteur de fruits (Nf_i) comme une variable aléatoire suivant la loi de probabilité *Log-Normale* de paramètres μ et σ .

$$Nf_i \sim \text{Log-}\mathcal{N}(\mu, \sigma^2) \text{ pour } i = 1, \dots, N_{6a} \quad (10.11)$$

La variable aléatoire μ que nous cherchons à estimer retranscrit le nombre moyen de fruits par individu adulte, alors que σ^2 est un paramètre fixe qui retranscrit la variabilité autour de ce nombre moyen - la variance.

Recrutement : processus 2

Connaissant la production de fruits pour chaque palmier d'un même transect nous en déduisons la production de fruits par transect comme étant la somme des nombres de fruits des palmiers du transect :

$$Nft_l = \sum_{i \in \text{transect } l} Nf_i \text{ pour } l = 1, \dots, n_l \quad (10.12)$$

Connaissant la production de fruits pour chaque transect, nous avons modélisé la transformation des fruits en nouveaux palmiers de stade 1 $N_{l,1}^{new}$ en utilisant une loi *Binomiale* de paramètres le nombre de fruits par transect l (Nft_l) et la probabilité de transformation d'un fruit en nouveau palmier l'année suivante pf :

$$N_{l,1}^{new} \sim \mathcal{B}(Nft_l, pf) \text{ pour } l = 1, \dots, n_l \quad (10.13)$$

Pour ceci nous avons émis l'hypothèse que le nombre de nouveaux palmiers de stade 1 pour chaque transect est tiré au sort parmi le nombre total de fruits présents sur ce même transect. Nous avons écarté l'hypothèse de densité-dépendance des fruits lors de ce processus. De plus, en nous basant sur nos connaissances quant à la répartition des fruits après être tombés au sol, nous avons aussi écarté l'hypothèse de « mobilité » entre transects des fruits par des facteurs exogènes.

Comme nous venons de le voir, le recrutement a été modélisé en distinguant deux sous processus biologiques. Le premier processus permettra d'estimer la valeur moyenne et le taux de variabilité du nombre de fruits produits par palmier adulte disposant d'inflorescences. Ce nombre moyen marqué μ est issu d'un tirage au sort dans une loi *Normale* de paramètres fixes γ et ω^2 .

$$\mu \sim \mathcal{N}(\gamma, \omega^2) \quad (10.14)$$

Les probabilités qu'un fruit devienne un palmier de stade 1 (pf) et qu'un palmier produise des fruits (p_{6a}) sont tirés selon une loi *Beta* non-informative de paramètres $a = b = 1, 5$.

$$pf \sim \text{Beta}(a, b) \quad (10.15)$$

$$p_{6a} \sim \text{Beta}(a, b) \quad (10.16)$$

10.2.2.3 Niveau hiérarchique

L'application d'un modèle multinomial (cf. equation 10.6) permet l'estimation en parallèle des probabilités de mortalité ($p_{l,s,m}$), de rester dans le même stade ($p_{l,s,0}$) et de passage au stade suivant ($p_{l,s,1}$) pour chaque **transect** et pour chaque **stade**.

Nous avons privilégié une structure hiérarchique en deux niveaux : **transect** et **population** (cf. FIGURE 10.3). Nous avons écarté le niveau de la parcelle dans l'estimation des paramètres puisque nous n'avons pas obtenu des résultats convergeant vers une possible différenciation à ce niveau. Nous avons donc révisé notre hypothèse de départ et avons stipulé que la différenciation s'opère au niveau du transect.

Cette structure nous a conduit à la sélection de multiples lois de probabilité et à la définition des paramètres qui leur sont associés pour chacun des niveaux hiérarchiques du modèle. L'équation 10.8 correspond au deuxième niveau hiérarchique,

puisque l'estimation des trois probabilités pour chaque localisation et pour chaque stade (partie gauche de l'équation) se base sur les trois probabilités pour chaque stade (partie droite de l'équation). En effet, l'information provenant de chaque transect (l) est transférée au niveau de la population pour l'estimation d'une probabilité par stade résumant l'information de tous les transects.

La probabilité de mortalité par stade ($p_{s,m}$), la probabilité de rester dans le même stade ($p_{s,0}$) et la probabilité de passer au stade suivant ($p_{s,1}$) (partie droite de l'équation 10.8) constituent le dernier niveau hiérarchique, car elles-mêmes suivent une loi de probabilité de *Dirichlet* avec des paramètres a_s , b_s et c_s , avec $a = b = c = 1/3$ pour $s = 1, \dots, 5$ et $a = b = 1/2$ pour $s = 6$.

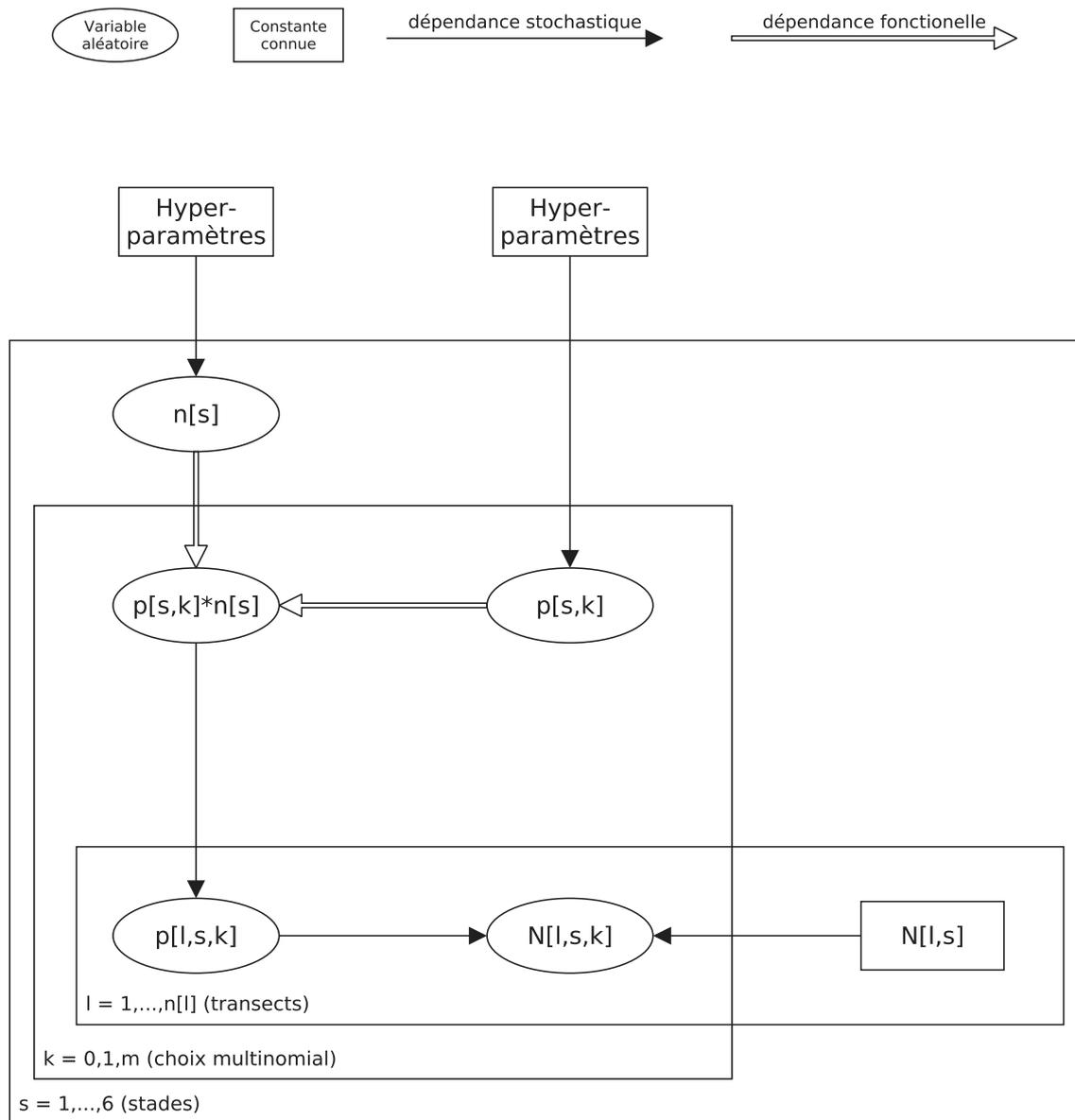
$$(p_{s,m}, p_{s,0}, p_{s,1}) \sim \text{Dirichlet}(a_s, b_s, c_s) \text{ pour } s = 1, \dots, 5 \quad (10.17)$$

$$(p_{6,m}, p_{6,0}) \sim \text{Dirichlet}(a_6, b_6) \quad (10.18)$$

Afin de pouvoir utiliser ces probabilités comme paramètre d'entrée de la loi de Dirichlet dans l'équation (10.8), nous avons pondéré ces trois probabilités ($p_{s,m}$, $p_{s,0}$ et $p_{s,1}$) par un paramètre de taille (d'échelle) pour chaque stade qui s'exprime en tant que nombre n_s issu d'un tirage au sort entre 1 et 200 d'après une loi uniforme.

$$n_s \sim \mathcal{U}(1, 200) \text{ pour } s = 1, \dots, 6 \quad (10.19)$$

FIGURE 10.8 – DAG (*Directed Acyclic Graph*) du modèle multinomial (basé sur le formalisme du logiciel BUGS - Lunn *et al.* (2000))



Cinquième partie

Résultats, analyses et discussions

11 Données empiriques : premier regard analytique	107
12 Analyse de la dynamique de population	117

Données empiriques : premier regard analytique

CETTE dernière partie est consacrée à la présentation de nos résultats et les analyses effectuées. Le premier résultat tangible du travail est la constitution de l'ensemble des connaissances partagées c'est-à-dire le catalogue des modèles descriptifs élaborés consensuellement et la constitution de la base de données qui va recueillir les données au delà de l'année en cours et qui pourra aussi s'enrichir des données socio-économiques. Bien évidemment, c'est surtout, autour de la présentation des données observées et de leur analyse, que nous allons nous centrer. Tout d'abord nous exposerons celles relatives à la croissance du babaçu, puis dans un second temps nous nous intéresserons à celles relevant du recrutement.

Sommaire

11.1 Processus de croissance	108
11.1.1 Fréquence de mortalité	108
11.1.2 Fréquence de passage	109
11.1.3 Fréquence de stabilité dans le même stade	110
11.2 Recrutement	112
11.2.1 Adultes - infrutescences	112
11.2.2 Adultes - fruits	113
11.2.3 Fruits - nouveaux stade 1	113

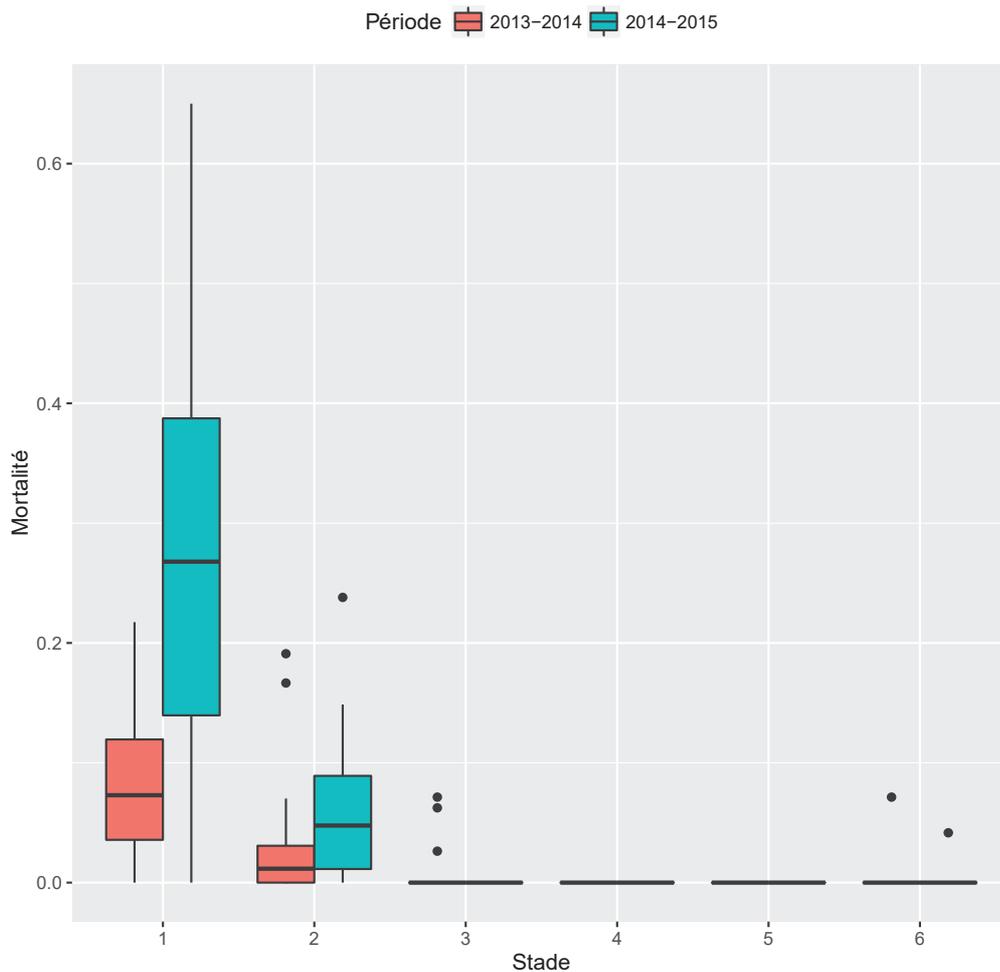
Nous entamerons notre analyse des résultats par la présentation de quelques indicateurs de centralité (moyenne et médiane) et de dispersion (écart-type) des données observées. Pour illustrer ceci, nous avons fait appel à des représentations de type « boîtes à moustaches ». Ces représentations forment un résumé de la distribution des données : elles font apparaître un indicateur de centralité représenté par la médiane (barre centrale dans la boîte), ainsi qu'un indicateur de dispersion premier et troisième quartiles (limites de la boîte). Cette boîte enferme ainsi 50% des valeurs observées. Les données en dehors des limites des moustaches sont considérées comme hors-normes et sont représentées par des cercles noirs.

11.1 Processus de croissance

11.1.1 Fréquence de mortalité

La FIGURE 11.1 présente les fréquences relatives à la mortalité au sein des pâturages. La tendance générale observée pour la première période d'observations (2013-2014) est une diminution de la mortalité du stade 1 au stade 6.

FIGURE 11.1 – Données empiriques pour la mortalité



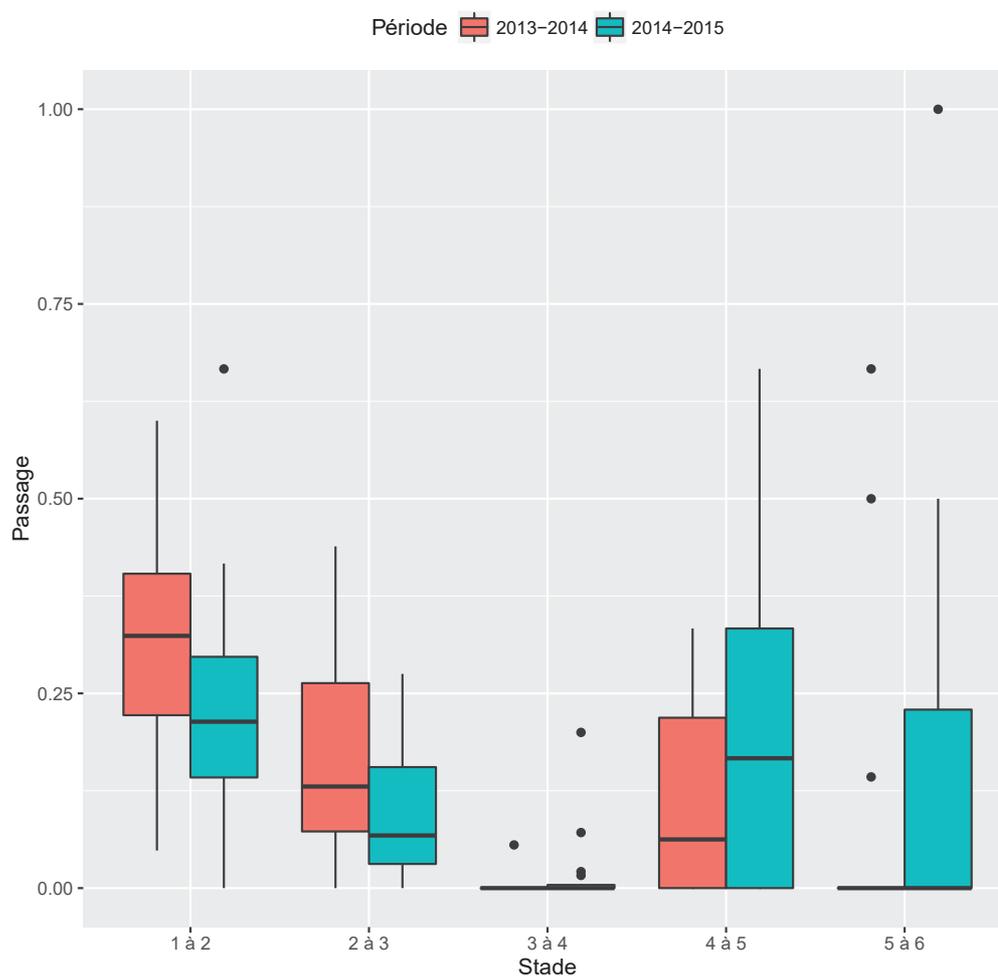
Cette tendance est confirmée par les données résultantes de la seconde période d'observations (2014-2015). Le point marquant est lié aux stades quatre et cinq : aucun palmier n'est mort durant cette période. Ce premier résultat nous a fait réviser notre protocole d'observations et ajuster les observations en cours (cf. section 9.2.2 - page 63). La seconde information qui ressort de la lecture de la FIGURE 11.1 consiste dans les considérables écarts enregistrés entre les deux périodes pour les stades 1 et

2. En effet, les moyennes basées sur les fréquences empiriques de mortalité pour la période 2013-2014 sont de 0,089 pour le stade 1 et de 0,036 pour le stade 2. Pour la seconde période (2014-2015), ces moyennes sont sensiblement différentes et passent respectivement à 0.259 et 0.064.

11.1.2 Fréquence de passage

La FIGURE 11.2 présente les taux empiriques de croissance pour les passages 2013-2014 et 2014-2015.

FIGURE 11.2 – Données empiriques de passage de stade



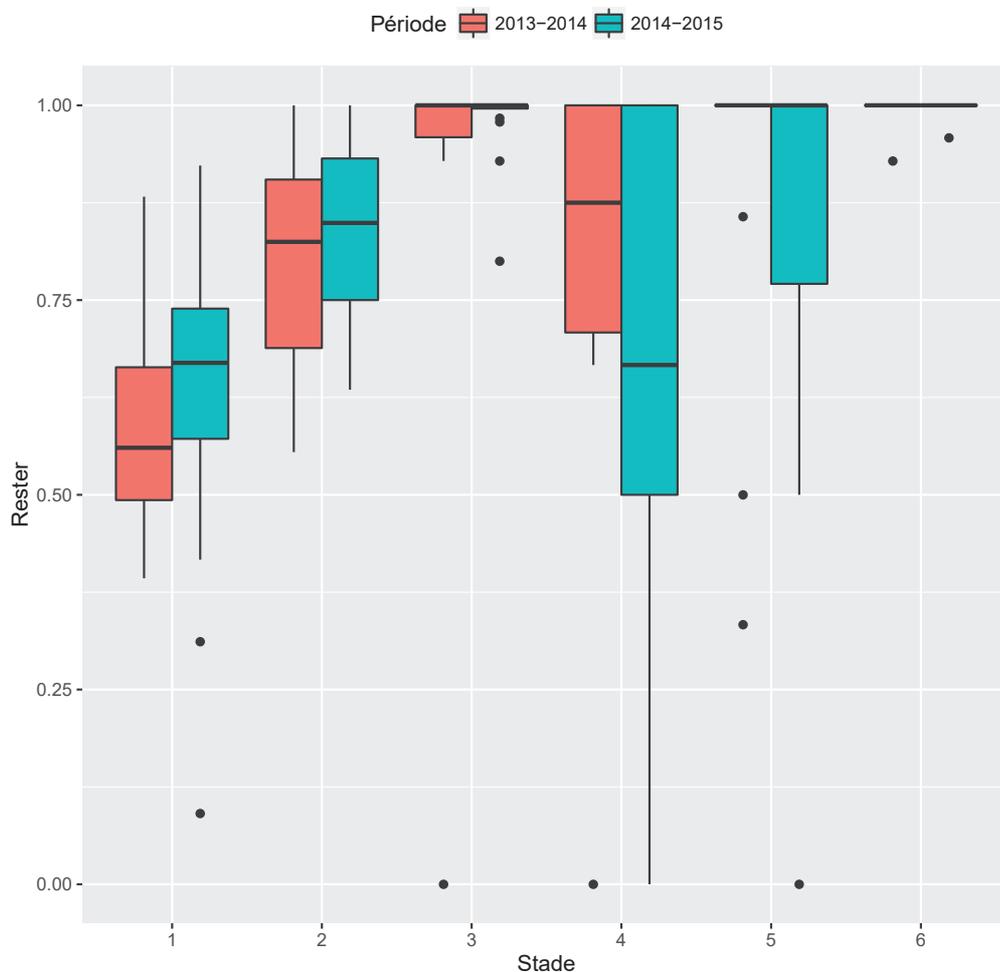
De notables différences sont observables pour tous les stades. Le fait marquant que nous n'avons su qu'à partir de la réception des données de terrain en septembre/octobre 2014 est le très faible passage entre les stade 3 et stade 4. Nous avons

constaté que les palmiers du stade 3 se retrouvent dans un goulot d'étranglement et ne passent pas au stade suivant : le taux moyen de passage pour ce stade est de l'ordre de 2% pour la période 2013-2014 et de 5% pour la période 2014-2015. Nous avons essayé durant le processus de modélisation d'agréger les données provenant des stades 3 et 4, puis celles des stades 4 et 5, afin d'asseoir la modélisation du processus de passage d'un stade à un autre sur un plus important volume de données.

11.1.3 Fréquence de stabilité dans le même stade

Comme nous pouvons l'observer sur la FIGURE 11.3, la fréquence de stabilité dans le même stade est supérieure à 0,5 pour tous les stades. Ceci nous donne une indication : la durée des stades est pluriannuelle.

FIGURE 11.3 – Données empiriques pour rester dans le même stade



Cette fréquence est de pratiquement 1 pour les stades 3 et 6. Le stade 6 est un stade de sortie : les palmiers restent dans ce stade et continuent de grandir. Le plus intéressant cas de figure est celui du stade 3, puisque d'après les estimations empiriques, les palmiers restent dans ce stade à une très grande majorité (proche de 1). Ceci nous a amené à deux déductions non exclusives :

- la durée de ce stade est beaucoup plus longue que celle des stades 1 à 5,
- nous ne disposons pas d'assez d'informations pour capturer le processus de passage au sein de ce stade.

Afin de pouvoir mieux analyser le comportement de l'espèce au sein de ce stade, nous avons modifié le protocole d'observation dans le but d'obtenir un échantillon plus conséquent. Pour ceci, durant les années 2015 et 2016, il a été décidé de marquer davantage de palmiers de stades 3 et 4. Malheureusement, comme nous l'avons souligné dans la remarque de la section 9.3 (page 66), nous ne sommes pas en mesure d'inclure ces données au sein du présent travail de thèse.

Nous avons aussi souhaité mieux comprendre la source de ce goulot d'étranglement du stade 3. Nous avons émis l'hypothèse d'une possible influence des pratiques anthropiques sur le passage entre les stades 3 et 4. Nous disposons d'informations concernant chaque palmier relatives à s'il était coupé ou non durant la période d'intérêt. Nous avons donc différencié les palmiers du stade 3 en deux catégories : « coupé » et « pas coupé ». Ensuite nous avons vérifié, pour les deux périodes (2013-2014 et 2014-2015), l'influence anthropique (palmiers coupés) sur le passage des palmiers du stade 3 au stade 4. Les TABLEAUX 11.1 et 11.2 résument cette information.

TABLEAU 11.1 – Stade 3 entre les observations 2013 et 2014

Observation		Passent		Restent		Total
2013	2014	%	nombre	%	nombre	
coupé	coupé	≈0	2	≈100	316	318
coupé	pas coupé	0	0	100	1	1
pas coupé	coupé	0	0	100	2	2
pas coupé	pas coupé	0	0	100	15	15
Total		-	2	-	334	336

Dans ces tableaux nous avons vérifié la dépendance du passage des palmiers à l'intervention anthropique sur ceux-là (coupé ou non) en début de période et en fin de période.

La première période est marquée par une prédominance des palmiers coupés en début et en fin de période (318) sur les palmiers se retrouvant dans les trois autres situations :

- coupé en début de période et pas coupé en fin de période - 1 palmier,
- pas coupé en début de période et coupé en fin de période - 2 palmiers,
- pas coupé en début de période et pas coupé en fin de période - 15 palmiers

pour un total de 336 palmiers (cf. TABLEAU 11.1). Uniquement 2, parmi les 318 palmiers qui ont été coupés en début et en fin de période, passent au stade suivant durant cette période ; ce qui représente 0,6% du total (2/336).

Malgré un nombre de palmiers de stade 3 presque doublé, passant donc de 318 à 638, sur la seconde période, le nombre de palmiers passant au stade 4 reste très faible. Seulement 11 palmiers parmi 638 passent au stade suivant, représentant 1,7% (cf. TABLEAU 11.2).

TABLEAU 11.2 – Stade 3 entre les observations 2014 et 2015

Observation		Passent		Restent		Total
2014	2015	%	nombre	%	nombre	
coupé	coupé	2	5	98	218	223
coupé	pas coupé	2	6	98	337	343
pas coupé	coupé	0	0	100	26	26
pas coupé	pas coupé	0	0	100	46	46
Total		-	11	-	627	638

Même si le nombre de palmiers passant au stade suivant durant les deux périodes n'est pas conséquent, nous remarquons que la totalité de ces palmiers ont été coupés en début de période. Aucun palmier non coupé en début de période, n'est passé au stade 4.

11.2 Recrutement

11.2.1 Adultes - infrutescences

Ce processus concerne uniquement les palmiers adultes (stade 6). Nous rappelons que nous avons distingué ceux-là en deux catégories : ceux donnant des infrutescences l'année n donc de type $6a$ et ceux de type $6b$ ne produisant pas d'infrutescences l'année n (cf. section 10.2.1.1 - page 93). Les données récoltées en 2013 nous indiquent un total de 103 palmiers adultes au sein des pâturages (cf. TABLEAU 9.3 - page 69) et 111 palmiers adultes en 2014. Cependant, les données relatives aux inflorescences/infrutescences n'ont pas été relevées au sein des transects additionnels marqués comme « suppl » dans ce même tableau. En conséquence, ces observations ont porté en 2013 et 2014 sur un total de 23 palmiers adultes.

Suite à la collecte d'informations relatives aux infrutescences ou traces de celles-là sur les palmiers adultes, nous avons déduit que 12 palmiers adultes étaient de type $6a$ en 2013 et 13 en 2014.

Ces observations fournissent une estimation empirique de la fréquence des palmiers de type $6a$ de 0,52 entre 2013 et 2014 et de 0,57 entre 2014 et 2015.

11.2.2 Adultes - fruits

La production de fruits par les palmiers adultes de type 6a en 2013 concernait 12 palmiers, leur nombre a évolué à 13 en 2014. Ce changement est le résultat de la distribution aléatoire des palmiers adultes (stade 6) en deux sous-catégories expliqué dans la section 10.2.1.1 (page 93).

Les données observées en 2013 et 2014 indiquent que ces palmiers ont produit en moyenne respectivement 313 et 320 fruits. Au sein de notre modèle, nous avons souhaité estimer la valeur de cette moyenne (μ). Les valeurs minimales et maximales observées sont très hétérogènes, puisque d'un palmier à un autre la production de fruits passe de 62 à 882 fruits, ce qui représente un rapport de dix (cf. TABLEAU 11.3). Le constat de cette hétérogénéité, nous a poussé à nous interroger sur les méthodes d'observations. En effet, nous avons constaté que plusieurs personnes effectuaient le comptage de fruits accrochés sur les infrutescences et qu'en fonction de leur niveau d'expertise, les résultats pouvaient varier. Ce fait nous a incité à changer le protocole d'observation des fruits et d'y introduire la prise en photographie systématique de chaque infrutescence pour assurer la vérification ultérieure du comptage sur le terrain.

TABLEAU 11.3 – Fruits par adulte de type 6a

Année	Min	Max	Moyenne	Écart-type
2013	88	845	313	244
2014	62	882	320	277

11.2.3 Fruits - nouveaux stade 1

Enfin, nous nous sommes intéressés au processus de germination : passage du fruit à la plantule (nouveau palmier de stade 1). Les observations relatives aux palmiers de stade 1 ont été réalisées au niveau de chaque transect. Nous avons décidé de modéliser ce processus en nous positionnant sur le même échelon. Pour ceci, nous avons additionné les fruits présents sur un même transect : les fruits provenant de plusieurs palmiers adultes ont été additionnés si ceux-là se situaient sur le même transect. Notre raisonnement s'est basé sur une estimation de la probabilité qu'un fruit devienne plantule (pf) à partir des nombres observés de fruits (l'année n) et du nombre de nouvelles plantules l'année $n + 1$.

Les observations relatives aux fruits, ont été effectuées seulement au sein des principaux transects (numérotés de 1 à 3) au sein des pâturages. Aucune observation n'a été enregistrée au sein des transects dits « suppl » (cf. TABLEAU 9.2 - page 67). Cependant, la présence de palmiers adultes a été observée uniquement au sein de 9 transects parmi les 17 mentionnés ci-dessus (cf. TABLEAUX 9.3 et 9.5 - respectivement pages 69 et 71). Les relevés nous ont indiqués qu'au moment des missions sur le terrain en 2013 et 2014, des fruits n'ont été observés que sur 8 parmi ces 9 transects.

Le TABLEAU 11.4 présente pour les deux périodes (2013-2014 et 2014-2015) ces 8 transects¹, avec pour chacun d'eux, le nombre de fruits observés l'année n et le nombre de nouveaux palmiers de stade 1 l'année $n + 1$.

TABLEAU 11.4 – Passage de fruits à nouveau palmier de stade 1

Période	Nombre	Transects							
		a	b	c	d	e	f	g	h
2013-2014	Fruits (n)	150	794	718	220	578	845	228	221
	Stade 1 ($n + 1$)	11	24	296	4	0	35	58	1
2014-2015	Fruits (n)	403	1117	1377	161	230	350	398	443
	Stade 1 ($n + 1$)	23	200	78	20	15	61	31	4

Afin d'apporter une analyse plus approfondie suite à la lecture de ce tableau, nous avons voulu vérifier notre hypothèse de départ relative à l'absence d'une densité-dépendance entre le nombre de fruits et le nombre de plantules (palmiers de stade 1) émise dans la section 10.2.1.6 (page 96).

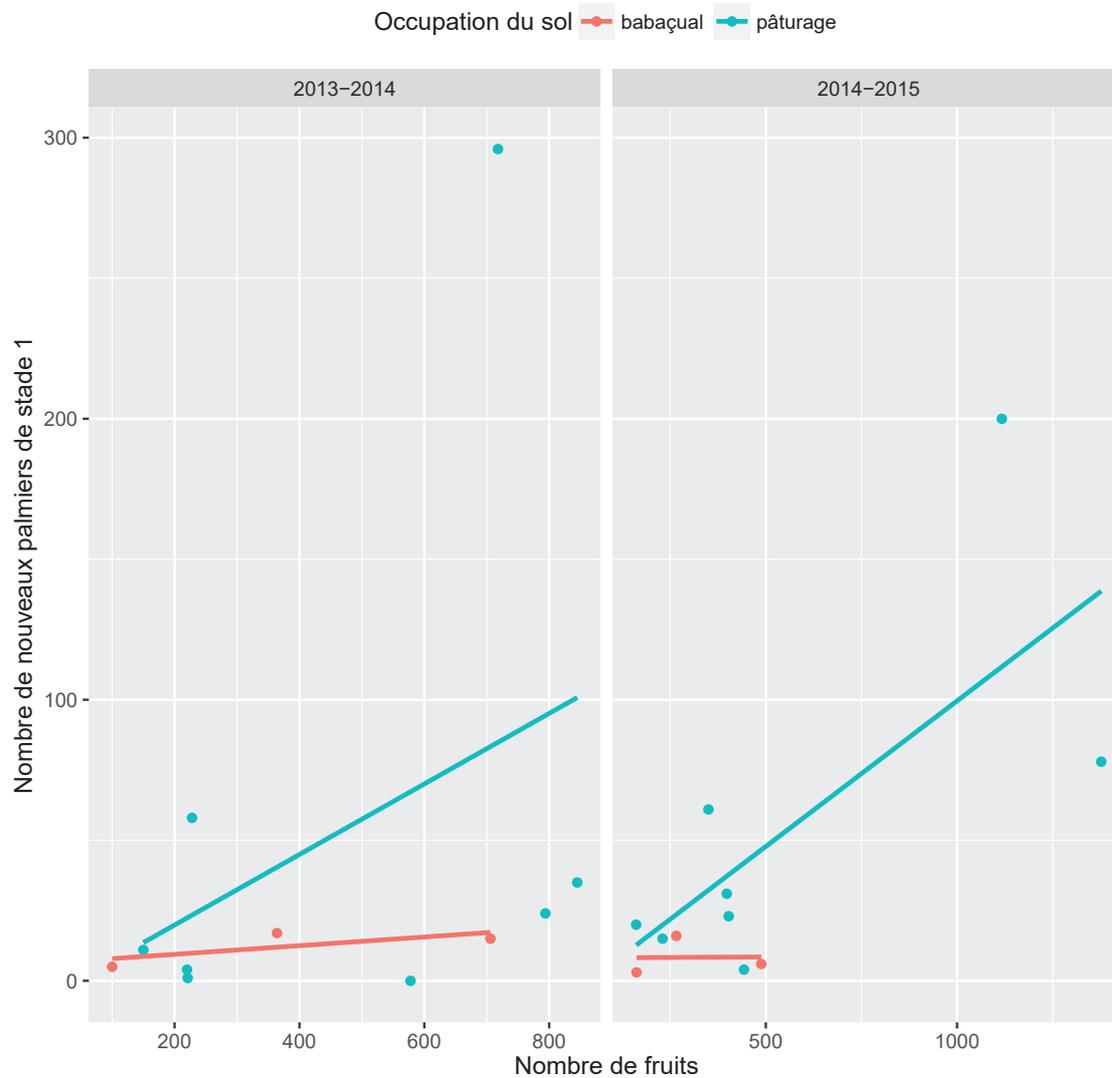
Comme nous l'avons vu dans le Chapitre 2), au sein du PA de Benfica, il existe trois principaux types d'occupation du sol dont la succession temporelle se présente comme suit : la forêt tropicale humide est remplacée par des pâturages (déforestation anthropique) qui peuvent être transformés en babaçuals (abandon). En plus de la vérification de notre hypothèse de départ, nous avons saisi l'occasion afin de comparer les densités-dépendances au sein des deux milieux anthropisés : les pâturages et les babaçuals. Les seconds se caractérisent par définition par une plus grande densité des palmiers babaçu. Nous avons donc considéré intéressant de comparer leurs densités-dépendances dans l'optique de la construction ultérieure d'un « méta-modèle » incluant la modélisation de transition d'une occupation du sol de type pâturage vers une occupation du sol de type babaçual.

La FIGURE 11.4 présente cette information. Nous avons fait figurer le nombre de fruits en abscisse et le nombre de nouveaux palmiers de stade 1 en ordonnée. La couleur bleu correspond aux observations relatives aux pâturages, alors que la couleur rouge est dédiée à celles relatives aux babaçuals.

Le premier fait marquant suite à l'observation de ce graphique est la différence entre la densité-dépendance au sein des pâturages et celle au sein des babaçuals. Les pâturages sont marqués par une importante variabilité des valeurs (les points bleus), mais la pente des droites de régression (lignes bleus) garde le même aspect pour les deux périodes. La situation est différente au sein des babaçuals. Certes, la pente des droites de régression (lignes rouges) pour les deux périodes est sensiblement la même, ainsi que les valeurs observées (points rouges).

Cependant, les droites de régression sont très proches de l'horizontale et nous indiquent ainsi une forte densité-dépendance : les palmiers des stades plus avancés

1. La correspondance des transects entre le TABLEAU 11.4 et le TABLEAU 9.2 (page 67) s'opère grâce aux lettres « a, b, ... ,h ».

FIGURE 11.4 – Densité-dépendance entre le nombre de fruits (n) et le nombre de nouveaux palmiers de stade 1 ($n + 1$)

influent sur le processus de germination. Du fait de manque d'observations plus prolongées, nous ne pouvons que supposer que les fruits, suite à la germination, passent plusieurs années en état de latence, avant de saisir une « opportunité » pour entamer leur développement du stipe. Ceci nous permet de faire le parallèle avec le comportement de l'espèce au sein de la forêt tropicale humide (cf. section 7.2.2 - page 51) : un laps de temps pluriannuel peut être observé entre la germination du fruit et l'apparition à la surface du stipe (Anderson *et al.*, 1991). La situation observée au sein des babaçuals et les résultats obtenus viennent donc, *a priori*, en appui de

ceux annoncés par *Anderson et al.* (1991) sur le site de Lago Verde concernant la forêt tropicale humide et les babaçuals. Dans les deux cas de figure (forêt tropicale humide et babaçual), il s'agit bien d'une forêt caractérisée par un plus haut degré de densité que les pâturages. La grande différence entre la forêt tropicale humide et le babaçual réside dans le fait que la seconde est une forêt mono-spécifique qui est le résultat de l'action anthropique exercée sur la forêt tropicale humide.

Ce résultat trouve tout son intérêt dans l'optique de la gestion à plus long terme de l'espèce. En effet, du simple fait de l'avancement de la déforestation ouvrant la voie à la délivrance du comportement de plante colonisatrice et adventice du babaçu, la tendance à l'avenir serait de retrouver les plus importantes densités du palmier babaçu de plus en plus au sein d'occupations du sol de type babaçual et de moins en moins au sein de la forêt tropicale humide. Les problèmes auxquels les populations locales devront donc faire face sont complexes. Ces populations devront trouver le juste milieu entre maintien de l'espèce au sein des pâturages et lutte contre la colonisation de ceux-là par l'espèce.

Analyse de la dynamique de population

CE CHAPITRE est consacré à la présentation et à l'analyse de nos résultats de modélisation. Nous débuterons par une section consacrée à l'analyse de sensibilité du modèle. Par la suite, nous nous intéresserons à la validation du modèle à l'aide des techniques présentées précédemment (Chapitre 10).

Cette présentation débouchera sur une analyse plus approfondie des estimations des probabilités de mortalité, de rester dans le même stade et de passage, ainsi que celles liées au recrutement, en appliquant les trois familles de loi conjuguées (*Dirichlet-Multinomiale*, *Beta-Binomiale* et *Normale-Normale*). Nous présenterons pour chacun des paramètres la valeur estimée à partir des données de 2013-2014 avec un *a priori* non-informatif, puis les estimations à partir des données 2014-2015 avec deux *a priori* distincts (un non-informatif et un informatif basé sur la loi *a posteriori* de 2013-2014).

Enfin, nous fournirons une analyse centrée sur les simulations d'évolution du palmier babaçu dans le temps pour des pas de temps distincts.

Sommaire

12.1	Analyse de sensibilité	118
12.2	Validation du modèle	119
12.2.1	<i>Posterior predictive p-values</i>	120
12.2.2	Validation croisée	121
12.3	Estimation des probabilités	122
12.3.1	Modèle de croissance	122
12.3.2	Recrutement	125
12.4	Modèle déterministe Caswell	128

Le modèle que nous avons proposé tient compte des processus régissant la dynamique de population du palmier babaçu. Nous avons ainsi obtenu des résultats pour ces trois processus que sont la croissance, la mortalité et le recrutement. L'aspect peu consolidé de certaines de nos données, notamment celles liées au recrutement nous a orienté vers une validation du modèle basée sur des données résultantes des observations de la croissance et de la mortalité.

12.1 Analyse de sensibilité

Nous avons testé la sensibilité de notre modèle en substituant l'*a priori* uniforme pour les probabilités par stade dans les équations 10.17 et 10.18 (page 102) par un *a priori* suivant une loi empirique. Nous avons donné de l'information *a priori* basée sur nos données pour observer son influence sur les estimations *a posteriori* et comparer les résultats aux estimations obtenues avec l'*a priori* non informatif (loi uniforme).

Les résultats de l'analyse de sensibilité sont présentées dans le TABLEAU 12.1. Ce tableau indique, au niveau de la population, les moyennes des estimations des probabilités de mourir ($p_{s,m}$), de rester dans le même stade ($p_{s,0}$) et de passer au stade suivant ($p_{s,1}$) obtenues en fonction d'un *a priori* non informatif pour la période 2013-2014 et d'un *a priori* non informatif et un autre informatif pour la période 2014-2015.

TABLEAU 12.1 – Analyse de sensibilité du modèle à l'*a priori* qui peut être non-informatif (ni) ou informatif (in) : les valeurs de l'*a posteriori* sont des valeurs moyennes de toutes les valeurs estimées par le modèle

Stade	Entrée-sortie	2013-2014 (ni)			2014-2015 (ni)			2014-2015 (in)		
		$p_{s,m}$	$p_{s,0}$	$p_{s,1}$	$p_{s,m}$	$p_{s,0}$	$p_{s,1}$	$p_{s,m}$	$p_{s,0}$	$p_{s,1}$
1	<i>a priori</i>	0,33	0,33	0,33	0,33	0,33	0,33	0,11	0,63	0,26
	<i>a posteriori</i>	0,12	0,6	0,28	0,27	0,52	0,21	0,21	0,57	0,22
2	<i>a priori</i>	0,33	0,33	0,33	0,33	0,33	0,33	0,04	0,79	0,17
	<i>a posteriori</i>	0,05	0,79	0,16	0,07	0,83	0,1	0,06	0,81	0,13
3	<i>a priori</i>	0,33	0,33	0,33	0,33	0,33	0,33	0,01	0,98	0,01
	<i>a posteriori</i>	0,005	0,993	0,002	0,001	0,994	0,005	0,005	0,989	0,006
4	<i>a priori</i>	0,33	0,33	0,33	0,33	0,33	0,33	0	0,79	0,21
	<i>a posteriori</i>	0,02	0,77	0,21	0,02	0,75	0,23	0,01	0,77	0,22
5	<i>a priori</i>	0,33	0,33	0,33	0,33	0,33	0,33	0	0,83	0,17
	<i>a posteriori</i>	0,01	0,85	0,14	0,01	0,87	0,12	0	0,85	0,15
6	<i>a priori</i>	0,5	0,5	-	0,5	0,5	-	0,01	0,99	-
	<i>a posteriori</i>	0,01	0,99	-	0,01	0,99	-	0,01	0,99	-

La lecture de ce tableau nous indique pour les stades 1 et 2 nous avons obtenus trois estimations moyennes différentes des probabilités de mourir ($p_{s,m}$), de rester dans le même stade ($p_{s,0}$) ou de passer au stade suivant ($p_{s,1}$) en fonction de l'*a priori* utilisé. Dans les deux cas d'une utilisation d'un *a priori* non-informatif avec les deux périodes d'observations, les estimation *a posteriori* divergent. Si nous prenons le cas de la probabilité de mourir ($p_{s,m}$), nous pouvons observer que la valeur moyenne obtenue *a posteriori* passe de 0,12 (observations 2013-2014) à 0,27 (observations 2014-2015). Malgré un *a priori* uniforme (non informatif avec $p = 0,33$) identique, les estimations obtenues diffèrent entre les deux années et la différence provient de l'information contenue dans les données. On constate que l'estimation de la même probabilité pour 2014-2015 avec un *a priori* informatif est légèrement plus basse (égale à 0,21). En effet, l'estimation a été obtenue tout en prenant compte d'une

part l'*a priori* informatif (basé sur l'expertise antérieure - observations de 2013-2014) et d'autre part les observations de 2014-2015. En fait, la valeur estimée (0,21) résulte du « compromis » entre l'*a priori* (centré sur 0,12 en informatif au lieu de 0,33 en non informatif) et les observations (centrées sur 0,259).

Nous observons que cette sensibilité par rapport au prior utilisé est en accord avec la grande variabilité observée entre les deux périodes et nous engage à penser que nous devons redéfinir l'*a priori* informatif grâce à de plus longues périodes d'observations ou par l'introduction d'un effet annuel lié aux conditions bio-climatiques.

Cette analyse peut être appliquée et peut expliquer par analogie les moyennes estimées pour les probabilités de rester dans le même stade ($p_{s,0}$) et de passer dans le stade suivant ($p_{s,1}$). De plus, les moyennes estimées pour le stade 2, peuvent être analysées suivant exactement la même logique.

Les estimations relatives aux stades 3, 4, 5 et 6 ne semblent pas être autant sensibles à l'*a priori* utilisé que les estimations pour les stades 1 et 2. Les estimations *a posteriori* dans les trois cas de figures (*a priori* non-informatifs avec les observations de 2013-2014 et 2014-2015 et *a priori* informatif avec les observations 2014-2015) pour ces trois stades sont très proches. Ceci résulte du fait de la faible variabilité dans les valeurs observées entre les deux périodes.

La sensibilité du modèle à l'*a priori* utilisé est caractéristique des deux premiers stades de développement du babaçu. Nous avons fourni ici une analyse de la sensibilité au niveau de la population, mais en analyse bayésienne cette sensibilité est aussi transmise aux autres niveaux hiérarchiques inférieurs. Dans notre cas de figure le niveau le plus bas dans la hiérarchie correspond aux transects (premier niveau). La différence entre les niveaux hiérarchique en matière d'analyse de la sensibilité réside dans le fait que plus nous sommes proches des observations (premier niveau), moins l'estimation sera sensible aux valeurs de l'*a priori*. Aussi, la sensibilité de notre modèle est moins importante au niveau de chaque transect.

12.2 Validation du modèle

La validation de notre modèle s'appuie sur les probabilités estimées du modèle de croissance. En fait, la non connaissance de l'erreur de mesure lors de l'observation portant sur les fruits, ainsi que le changement du protocole d'observation en cours de thèse, nous ont incité à concentrer l'analyse de la validité du modèle sur la partie croissance où les données sont fiables, mieux structurées et documentées.

Nous avons utilisé deux approches différentes pour mesurer la validité de notre modèle. Tout d'abord nous exposerons les résultats issus de celle relative aux calculs de probabilités prédictives *a posteriori* (*posterior predictive p-value*).

Puis, nous analyserons ceux provenant de l'approche dite de validation croisée.

12.2.1 *Posterior predictive p-values*

Les valeurs relatives aux *posterior predictive p-value* sont visibles dans le TABLEAU 12.2. Ces valeurs ont été calculées pour tous les stades et pour les trois probabilités d'intérêt au sein du modèle de croissance ($p_{s,m}$, $p_{s,0}$ et $p_{s,1}$). Elles concernent les estimations *a posteriori* pour 2014-2015 du modèle utilisant l'*a priori* informatif¹.

TABLEAU 12.2 – Valeurs des *posterior predictive p-value* pour le modèle utilisant un *a priori* informatif

Stade	Mortalité	Rester dans le stade	Changer de stade
1	0	0,5	0,22
2	0,62	0,1	0,04
3	0,39	0,14	0,24
4	0,15	0,64	0,76
5	0,14	0,43	0,57
6	0,71	0,36	-

Certaines des valeurs présentées, proches de 0,5, montrent une bonne adéquation entre les valeurs simulées par le modèle et celles observées sur le terrain. Quand la *posterior predictive p-value* est plus petite, alors les données simulées sont moins dispersées (autour de la moyenne) que les données observées. Par exemple, c'est le cas de la mortalité du stade 1. Ceci s'explique par la très grande variabilité en terme de mortalité entre les deux périodes de mesures (2013-2014 et 2014-2015, cf. FIGURE 11.1). En effet, nous avons estimé les probabilités de mortalité par transect pour la période 2014-2015 en utilisant les informations de 2013-2014 (l'*a priori* du modèle est construit sur les résultats de la première période). Or, en 2013-2014, les mortalités variaient moins d'un transect à l'autre, ce qui a influencé les résultats du modèle avec sous estimation de la variabilité des mortalités entre transects. Nous constatons que le modèle proposé est sensible au choix de l'*a priori* (cf. section 12.1) et nous recommandons par conséquent :

- une plus longue période d'observations pour construire un *a priori* pertinent. Cette déduction s'inscrit dans la pensée de Bierzychudek (1999) présentée dans la section 10.1.1.2 (page 84),
- la modification éventuelle du modèle, suite à l'analyse plus approfondie de l'effet année grâce aux observations de 2016, en vue d'intégration de celui-ci au sein du modèle.

1. Nous rappelons que la valeur de référence pour l'évaluation du modèle à l'aide des *posterior predictive p-value* est 0.5 (cf. section 10.1.2.1 - page 92)

12.2.2 Validation croisée

Pour la validation croisée nous avons suivi la procédure décrite dans la section 10.1.2.1 (page - 92). Ainsi, pour chaque stade, nous avons séparé notre échantillon de données observées (y^{obs}) de **2014 et 2015** en deux sous-ensembles. L'ensemble le plus conséquent de données ($\approx 75\%$) a été utilisé pour l'estimation et l'obtention de données prédites. Le second ensemble regroupant les données restantes a été utilisé pour la validation. Nous nous sommes basés sur la distinction de nos données, opérée lors de l'observation, et avons donc pris comme unité de distinction pour la validation croisée le transect. Nous avons utilisé une partie de ceux-là pour l'estimation des probabilités. Les données provenant du reste des transects ont été en conséquence utilisées pour la validation. Le TABLEAU 12.3 présente le nombre de transects en fonction de leur répartition au sein de ces deux ensembles².

TABLEAU 12.3 – Nombre de transects utilisés pour l'estimation et pour la validation croisée

Nombre	Stade 1	Stade 2	Stade 3	Stade 4	Stade 5	Stade 6
estimation	12	12	12	4	11	12
validation	4	5	4	3	3	4
Total	16	17	16	7	14	16

Le TABLEAU 12.4 expose, pour chacun des 6 stade du babaçu, les valeurs observées et pour les valeurs prédites la moyenne et l'intervalle de crédibilité (2,5% - 97,5%). Le nombre total par stade (dernière colonne de ce tableau), est obtenu par la somme des palmiers au sein des transects (utilisés pour la validation) correspondant à ce stade dans le TABLEAU 12.3.

Pour le **stade 1**, 75 palmiers qui croissent mais restent dans le même stade ont été observés. Le modèle prédit une valeur proche de celle-là : 81 palmiers. Par contre, il existe un décalage entre les nombres empiriques et prédits de palmiers morts ou passants dans le stade suivant. Le premier nombre est sous-estimé (27 prédits contre 46 observés), alors que le second est surestimé (32 prédits contre 19 observés).

Concernant le **stade 2**, le nombre observé de palmiers qui restent dans le même stade s'élève à 481, alors que celui prédit par le modèle est de 443 donc sous-estimé. Au contraire, le nombre de palmiers passants au stade suivant est surestimé. Il s'élève à 79 palmiers pour la prédiction, contre seulement 39 palmiers réellement observés. Ainsi, les prédictions des palmiers restant dans le même stade et ceux passant au stade suivant sont décalées par rapport aux valeurs observées. En effet, les valeurs observées se situent en dehors des fourchettes données par l'intervalle de crédibilité 2,5%-97,5%.

2. La liste complète des transects utilisés pour la validation croisée est visible dans l'Annexe D (page - 167), pour les stades 1 à 6 respectivement dans les TABLEAUX D.1, D.2, D.3, D.4, D.5 et D.6.

TABLEAU 12.4 – Validation croisée

Stade	Nombre	Morts			Restants dans le même stade			Passant au stade suivant			Total
		2,5%	50%	97,5%	2,5%	50%	97,5%	2,5%	50%	97,5%	
1	observation	-	46	-	-	75	-	-	19	-	140
	prédit	15	27	40	65	81	96	20	32	46	
2	observation	-	33	-	-	481	-	-	39	-	553
	prédit	16	31	51	410	443	471	74	79	107	
3	observation	-	0	-	-	83	-	-	0	-	83
	prédit	0	0	2	79	82	83	0	1	3	
4	observation	-	0	-	-	3	-	-	2	-	5
	prédit	0	0	1	2	4	5	0	1	3	
5	observation	-	0	-	-	8	-	-	2	-	10
	prédit	0	0	1	6	9	10	0	1	4	
6	observation	-	1	-	-	58	-	-	-	-	59
	prédit	0	1	3	56	58	58	-	-	-	

La prédiction du nombre de palmiers morts (31 palmiers), est en concordance avec les valeurs observées (33 palmiers), même si l'intervalle de crédibilité nous indique une fourchette située entre 16 et 51 palmiers.

Nous pensons que le décalage entre les données observées et les données prédites pour les stades 1 et 2 peut provenir du grand écart dans les observations de 2013-2014 et celles de 2014-2015. Comme nous l'avons fait remarquer dans le chapitre précédent, les écarts dans les observations des données relatives à la mortalité, au passage au stade suivant et au fait de rester dans le même stade entre les deux périodes sont considérables.

Les valeurs prédites pour les **stades 3, 4, 5 et 6** sont en concordance avec les valeurs observées. De plus, pour la plupart d'entre elles, la coïncidence à un palmier près avec les valeurs observées est un signe de la forte qualité prédictive du modèle concernant ces quatre stades.

12.3 Estimation des probabilités

Nous présentons dans cette section les résultats obtenus à partir des estimations relatives d'une part au modèle de croissance et d'autre part aux processus du recrutement.

12.3.1 Modèle de croissance

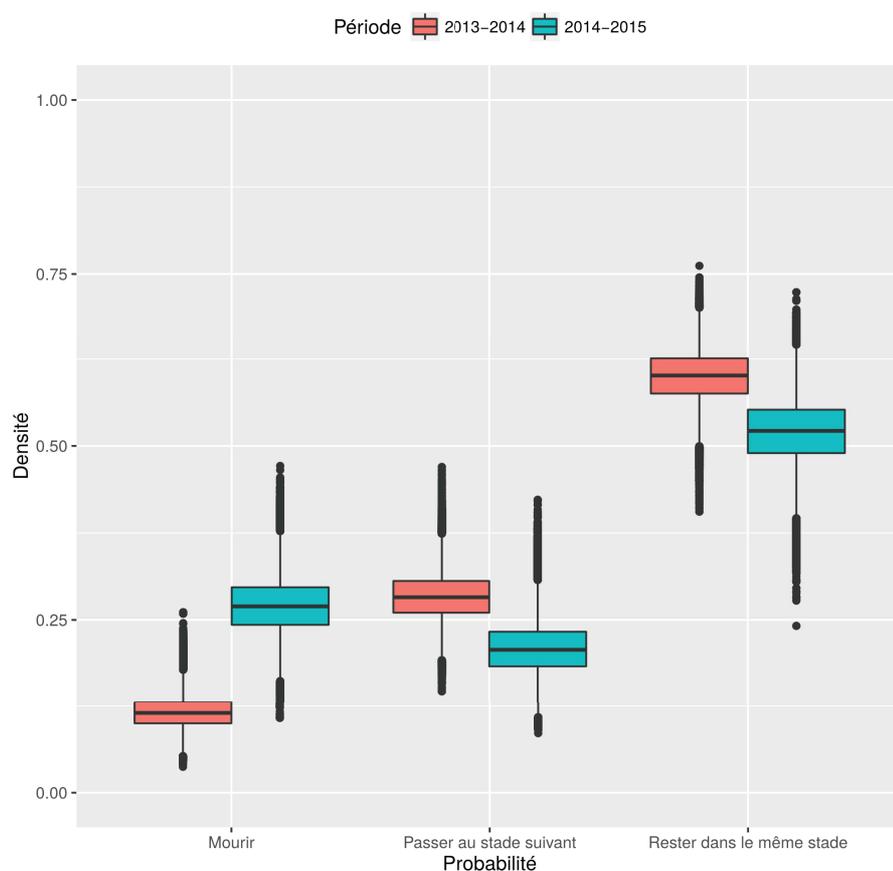
L'estimation des probabilités de mourir ($p_{s,m}$) et de rester dans le même stade ($p_{s,0}$) a été faite pour les six stades ; celle de passer au stade suivant ($p_{s,1}$) a été faite pour les stades 1 à 5. Nous avons accordé la présentation de nos résultats en nous basant sur la distinction des stades proposée précédemment en plantule, jeunes (1, 2, 3 et 4) et adulte (cf. TABLEAU 9.1 et FIGURE 10.6 - respectivement pages 61 et 94).

Les résultats seront présentés à l'aide de diagrammes à moustaches pour chacun de ces trois paramètres. Chaque diagramme fera la comparaison des estimations basées sur les données des deux périodes (2013-2014 et 2014-2015) et un *a priori* non-informatif.

La plantule - le stade 1

Sur la FIGURE 12.1 nous avons fait apparaître les estimations des probabilités pour les plantules.

FIGURE 12.1 – Probabilités estimées pour le stade 1

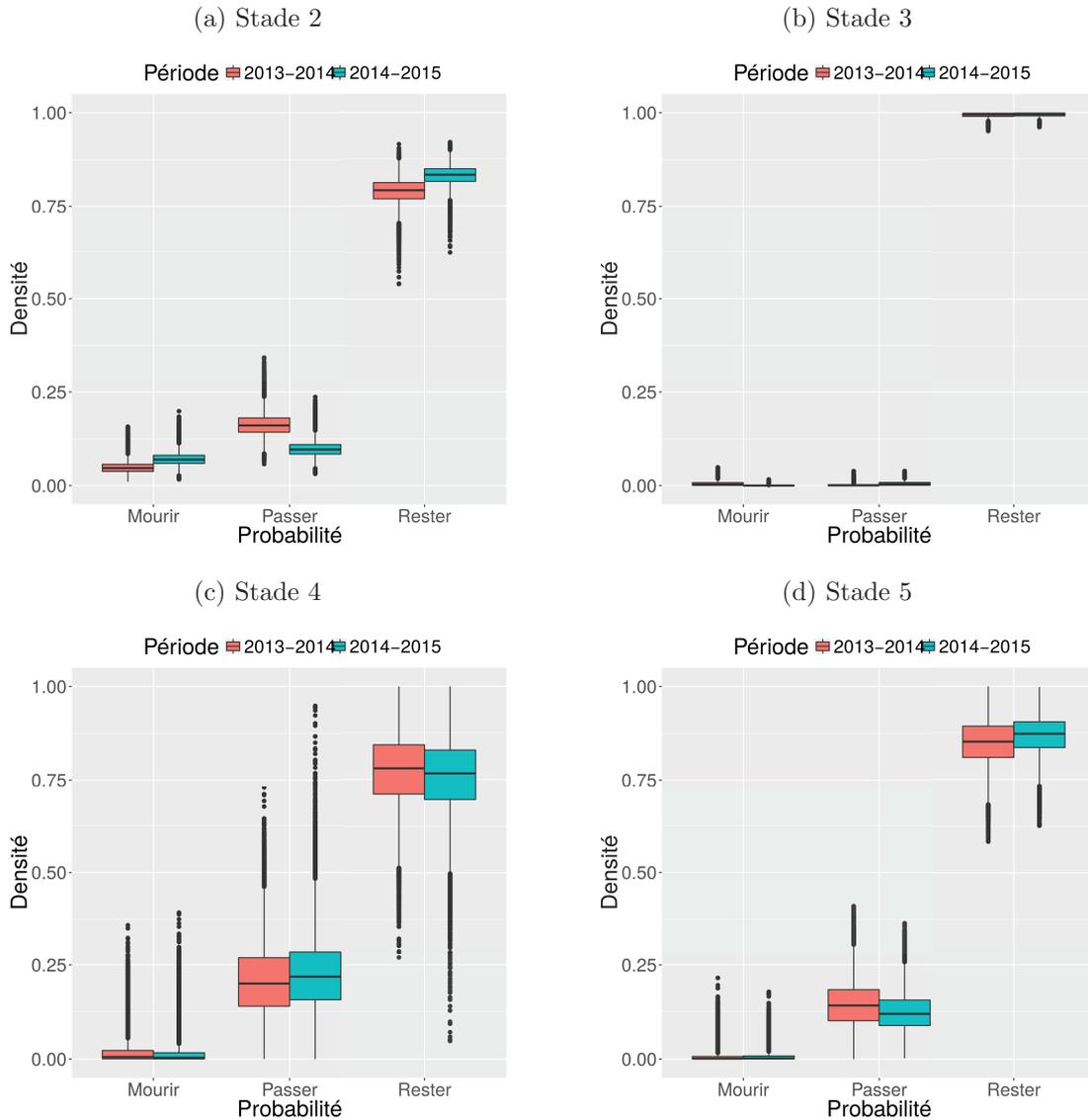


Ce stade est marqué par une forte variabilité au niveau de la mortalité entre les deux périodes. Nous soupçonnons l'existence d'un véritable effet année liée aux conditions bioclimatiques. Même si ce stade est marqué par une forte variabilité de l'estimation de la moyenne de la probabilité de mortalité, passant de 0,12 pour la première période à 0,27 pour la seconde période, nous n'avons pas obtenu des estimations en accord avec les résultats avancés par Anderson (1983) et des taux de mortalité se situant aux environs de 0,38.

Les Jeunes (stades 2 à 5)

Les stades définis comme « Jeunes » sont marqués par des mortalités très faibles. Comme nous pouvons l'observer sur les FIGURES 12.2a à 12.2d, nous n'avons pas décelé un notable effet année entre les deux périodes pour l'ensemble des stades.

FIGURE 12.2 – Probabilités estimées pour les palmiers « Jeunes »



Les profils de ces quatre stades se ressemblent à l'exception du stade 3 qui, comme nous l'avons déjà vu, est marqué par de très faibles probabilités de mortalité et de passage au stade suivant. Nous pouvons rappeler ici, que la particularité du

stade 3 (faibles mortalités et passages au stade suivant observées) nous ont incité à procéder à plusieurs essais de regroupement de stades dans l'objectif d'une recherche d'amélioration des estimations des probabilités. Pour ceci, nous avons d'abord essayé une estimation des probabilités en créant un stade regroupant les données observées des stades 3 et 4. Cette expérience n'a pas abouti à des conclusions satisfaisantes, puisque nous n'avons pas pu déceler une amélioration dans la qualité de nos estimations. C'est pourquoi, dans un second temps nous avons regroupé les stades 4 et 5 au sein d'un même stade. Les résultats obtenus suite aux estimations des probabilités pour ce stade n'ont pas été concluants.

Ces deux essais de recherche d'amélioration des estimations obtenues s'inscrivent dans notre démarche pionnière de prospections et d'interrogations autour de la proposition de ce premier modèle de dynamique de population du palmier babaçu. À travers de cette prospection, nous n'avons pas remis en question les définitions des stades biologiques (faites par nos collègues écologues). Nous nous sommes interrogés sur le sens plus large de ces stades en termes de modélisation mathématique des processus biologiques.

La démarche modélisatrice est, comme nous l'avons déjà souligné, un processus itératif (cf. 8.1.1 - page 55). Lors de la modélisation nous avons recherché le meilleur cas de figure possible pour retranscrire statistiquement des processus naturels. C'est pourquoi nous avons entrepris ces deux essais. À travers de ces actions, nous n'avons pas réellement remis en question la définition des stades biologiques. Nous nous sommes plutôt interrogés sur le sens que ces définitions avaient en matière de modélisation. Nous avons recherché le sens statistique du découpage de la population opéré en fonction de caractéristiques biologiques de l'espèce.

Ainsi, certaines de nos estimations vont s'accorder avec les définitions des stades, mais d'autres ne vont pas le faire, puisqu'il y a eu des choix de définition biologique des stades (basée sur l'aspect visuel du palmier, ses capacités reproductives. . .) faits *a priori*.

Les adultes (stade 6)

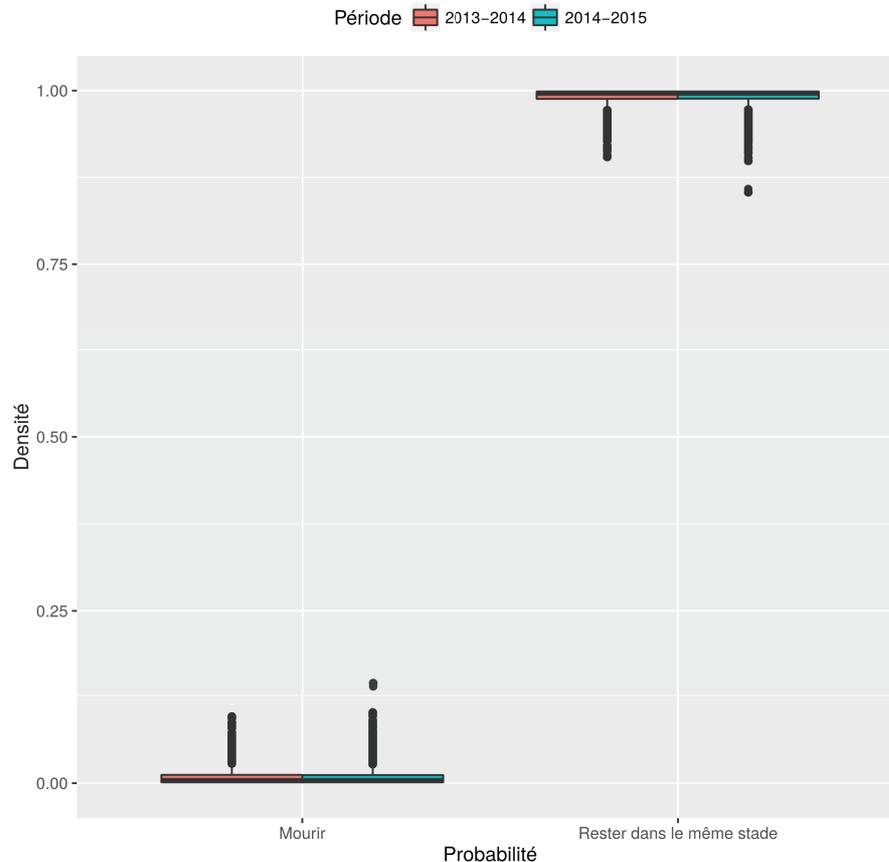
Le stade 6 est par définition celui des palmiers babaçu adultes, c'est-à-dire des palmiers capables de produire des fruits et assurer le processus de recrutement de l'espèce. La spécificité biologique de ce stade réside dans sa grande longévité. C'est, en effet, le stade dans lequel le palmier passera la plus grande partie de son existence. Biologiquement, ce stade correspond également aux palmiers qui se retrouvent désormais dans un état biologique plus pérenne comparé aux autres stades.

12.3.2 Recrutement

12.3.2.1 Séparation des palmiers adultes en deux catégories : 6a et 6b

L'estimation de la probabilité pour qu'un palmier adulte produise de fruits (p_{6a}) a abouti à une valeur moyenne de 0,54 avec un écart-type de 0,07. Cette estimation

FIGURE 12.3 – Probabilités estimées pour le stade 6

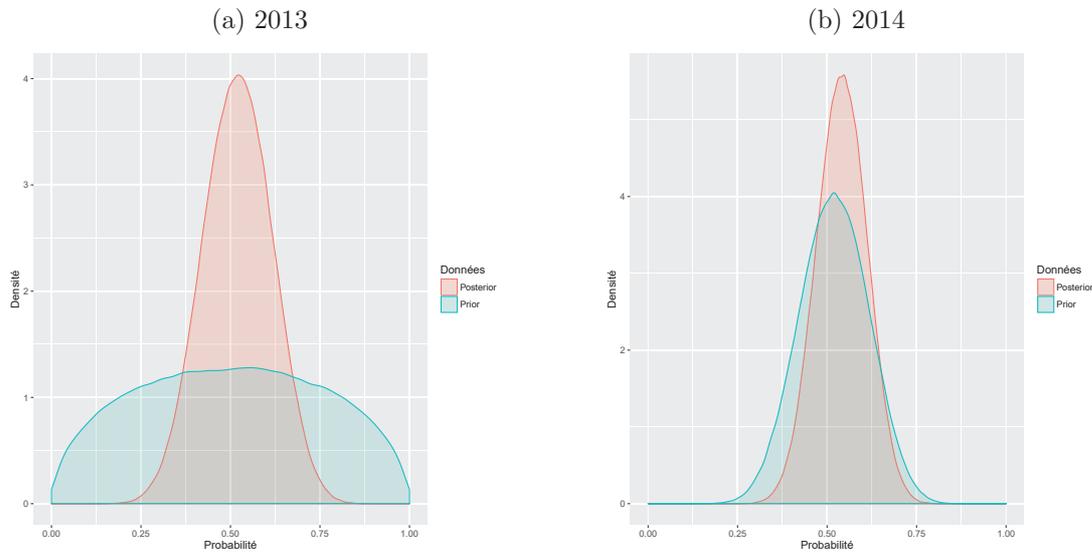


s'est construite en deux phases : dans une premier temps elle était basée sur les données de 2013 avec un *a priori* non-informatif, puis sur les données de 2014 avec un *a priori* intégrant la connaissance acquise.

La FIGURE 12.4a illustre en rouge les résultats issus de l'estimation de cette probabilité avec les données de 2013 (première phase). En termes de biologie de l'espèce, cette valeur indique qu'annuellement un peu plus de la moitié des palmiers adultes ont des inflorescences féminines donnant lieu à des infrutescences. Sur cette figure nous avons illustré en bleu la distribution *a priori* non-informative utilisée (loi *Beta* avec des paramètres fixes $a = b = 1,5$). Nous avons utilisé ces résultats afin de guider la loi *Beta a priori* à privilégier des valeurs de probabilité, lors du processus d'estimation, en fonction de la loi *a posteriori* obtenue à partir des données de 2013.

La FIGURE 12.4b montre toujours en rouge l'estimation de la même probabilité à partir des données de terrain de 2014 et d'un *a priori* informatif (loi *Beta* avec paramètres $a = 31,5$ et $b = 74,5$ à partir des données de 2013) illustré en bleu (seconde phase).

FIGURE 12.4 – Estimation de la probabilité de donner des fruits



Nous avons pu obtenir les valeurs de ces paramètres, grâce à l'utilisation des lois de probabilité *Beta-Binomiale* dites conjuguées (cf. section 10.2.2 - page 96).

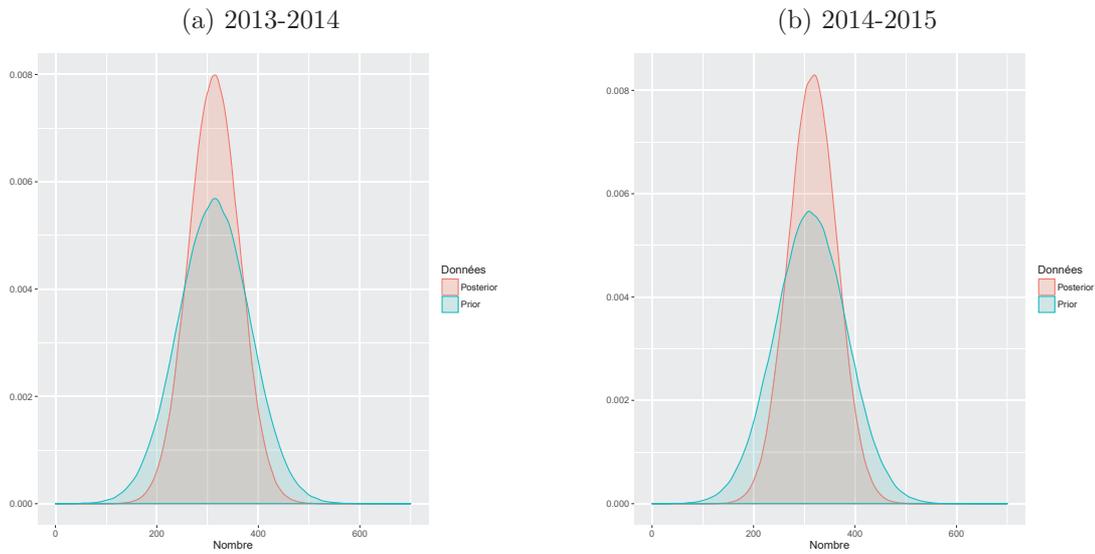
Une fois que nous avons obtenu l'estimation de cette probabilité, nous nous sommes tournés vers l'estimation des paramètres du recrutement proprement dit.

12.3.2.2 Le nombre moyen de fruits par palmier adulte de type 6a

Le premier sous-processus dans le processus de recrutement est la fructification. Afin de mieux l'appréhender, nous avons estimé le nombre de fruits (μ) par palmier adulte en mesure de donner des infrutescences : $\approx 54\%$ de la totalité des palmiers adultes. La moyenne du nombre de fruits estimée à partir des données est de 317 fruits par adulte, avec un écart-type de 48 fruits.

La FIGURE 12.5a illustre les valeurs issues de l'estimation *a posteriori* (en rouge) et l'information *a priori* centrée sur 313 fruits par palmier adulte (en bleu). Comme précédemment, nous avons utilisé l'information *a posteriori* ainsi obtenue pour l'estimation de ce nombre à partir des données de 2014 et 2015. La FIGURE 12.5b reflète cette estimation, dont l'*a priori* (en bleu) est cette fois-ci centré sur la valeur moyenne estimée précédemment. Ici aussi la variabilité dans la production de fruits par palmier est conséquente. Les variations observées ne s'expriment pas cette fois-ci en fonction des périodes d'observations, mais sont enregistrées au sein d'une même période d'observations.

FIGURE 12.5 – Estimation du nombre de fruits par adulte de type 6a



12.3.2.3 Le processus de passage du fruit au nouveau palmier de stade 1

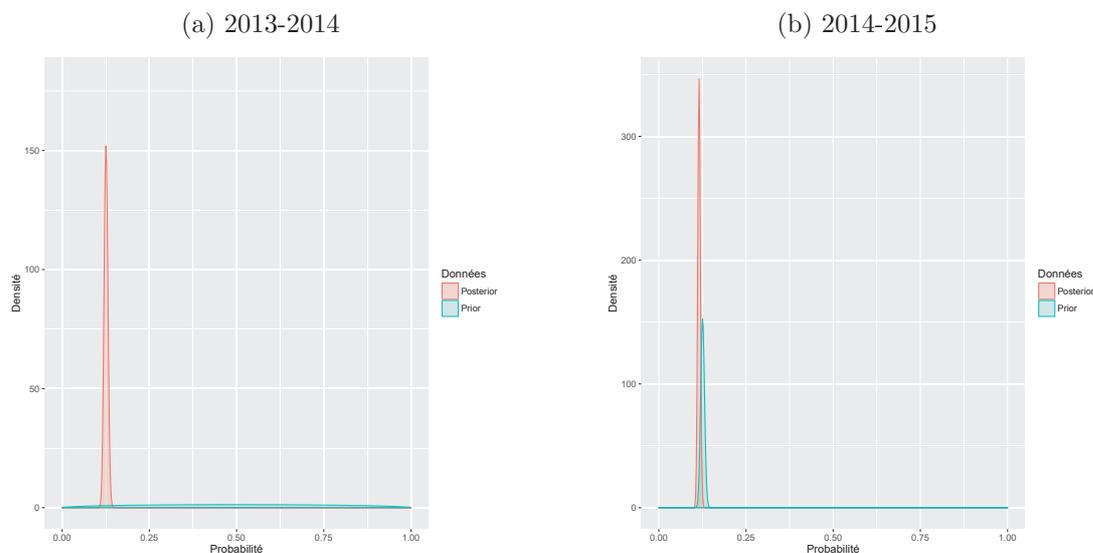
Le second sous-processus lié au processus de recrutement est le passage d'un fruit observé l'année n à un nouveau palmier de stade 1 l'année $n + 1$. Nous avons modélisé ce processus en tant qu'estimation d'une probabilité reflétant ce passage (pf).

La FIGURE 12.6a démontre, comme précédemment, l'estimation de cette probabilité en rouge et l'information de l'*a priori* non informatif en bleu. Cette probabilité se retranscrit biologiquement par la germination des fruits et leur transformation en plantules. Pour l'estimation de la valeur de cette probabilité pour la seconde période (2014-2015), nous nous sommes servis d'un *a priori* informatif basé sur les données de la première période (2013-2014) et l'*a priori* non-informatif utilisé pour cette première période. La valeur moyenne ainsi obtenue se situe à $\approx 0,12$ (cf. FIGURE 12.6b).

12.4 Modèle déterministe Caswell

Nous analyserons ici les résultats obtenus grâce au développement d'une première approche basée sur des simulations. Pour faire ceci, nous avons créé un modèle matriciel déterministe. La matrice utilise en entrée les valeurs estimées grâce au modèle multinomial hiérarchique bayésien proposé. Cette matrice était élaborée en accord avec les formulations avancées par Caswell (2001) et présentées dans la section 10.1.1.2 (page 84). Nous rappelons que ce modèle fournit des estimations

FIGURE 12.6 – Estimation de la probabilité de nouveaux stade 1



relatives au vecteur d'état pour l'année $n + 1$ en se basant sur les données de ce même vecteur d'état l'année n et une matrice de transition. La diagonale principale au sein de cette matrice recueille les valeurs relatives aux probabilités pour que les palmiers restent dans le même stade. La sous-diagonale, quant à elle, est renseignée avec les valeurs de probabilités pour que les palmiers passent au stade suivant. Enfin, les paramètres relatifs au processus de recrutement sont renseignés sur la première ligne de la matrice.

Les équations liées à ce modèle déterministe s'écrivent comme suit :

$$N_1(n + 1) = R_1 \cdot N_1(n) + Nft(n) \cdot pf \quad (12.1)$$

$$N_2(n + 1) = R_2 \cdot N_2(n) + C_1 \cdot N_1(n) \quad (12.2)$$

$$N_3(n + 1) = R_3 \cdot N_3(n) + C_2 \cdot N_2(n) \quad (12.3)$$

$$N_4(n + 1) = R_4 \cdot N_4(n) + C_3 \cdot N_3(n) \quad (12.4)$$

$$N_5(n + 1) = R_5 \cdot N_5(n) + C_4 \cdot N_4(n) \quad (12.5)$$

$$N_6(n + 1) = R_6 \cdot N_6(n) + C_5 \cdot N_5(n) \quad (12.6)$$

$$N_{6a}(n + 1) = p_{6a} \cdot N_6(n + 1) \quad (12.7)$$

$$Nft(n + 1) = \mu \cdot N_{6a}(n + 1) \quad (12.8)$$

Des équations (12.6) et (12.7) écrites en années $n + 1$:

$$N_{6a}(n + 1) = p_{6a} \cdot N_6(n + 1) = p_{6a} \cdot R_6 \cdot N_6(n) + p_{6a} \cdot C_5 \cdot N_5(n) \quad (12.9)$$

De l'équation (12.7) écrite en années n nous déduisons :

$$p_{6a} \cdot N_{6a}(n+1) = R_6 \cdot N_{6a}(n) + p_{6a} \cdot C_5 \cdot N_5(n) \quad (12.10)$$

Des équations (12.1) et (12.8), nous déduisons :

$$N_1(n+1) = R_1 \cdot N_1(n) + pf \cdot \mu \cdot N_{6a}(n) \quad (12.11)$$

L'ensemble des équations ci-dessus nous permet d'obtenir la forme matricielle suivante :

$$\begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \\ N_{6a} \end{pmatrix} (n+1) = \begin{pmatrix} R_1 & 0 & 0 & 0 & 0 & pf \cdot \mu \\ C_1 & R_2 & 0 & 0 & 0 & 0 \\ 0 & C_2 & R_3 & 0 & 0 & 0 \\ 0 & 0 & C_3 & R_4 & 0 & 0 \\ 0 & 0 & 0 & C_4 & R_5 & 0 \\ 0 & 0 & 0 & 0 & p_{6a} \cdot C_5 & R_6 \end{pmatrix} \cdot \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \\ N_{6a} \end{pmatrix} (n) \quad (12.12)$$

L'équation 12.12 fournit un aperçu du modèle matriciel déterministe que nous proposons. Nous retrouvons donc dans ce modèle notre variable d'état composée d'un vecteur d'état (N_1, \dots, N_6) au sein duquel une seule modification relative à la co-existence de palmiers adultes de types 6a et 6b a été apportée. Le nombre de palmiers du stade 6 (N_6) a été remplacé par le nombre de palmiers pouvant fournir des fruits (N_{6a}). La matrice de transition est remplie grâce aux probabilités que nous avons estimés : les probabilités par stade de rester dans le même stade ($p_{s,0}$), les probabilités par stade de passer au stade suivant ($p_{s,1}$) et les paramètres relatifs au recrutement (pf et μ).

La valeur asymptotique du taux de croissance de la population (λ) associée à cette matrice est 1,1. Cette valeur est supérieure à 1 et indique donc, comme nous l'avons souligné dans la section 7.1.1 (page 45) que la population a une tendance à croître. Nous souhaitons connaître le niveau maximal de la pression anthropique exercée sur la ressource. C'est-à-dire le niveau pour lequel λ serait égal à 1. Nous savons que cette pression anthropique se manifeste par les prélèvements de fruits. Appelons donc ρ le coefficient obtenu par la multiplication des deux paramètres relatifs au recrutement pf et μ ($pf \cdot \mu = 0,12316 = 37,92$). En conséquence nous savons que pour une valeur de ρ égale à 37,92, la valeur de λ obtenue d'après le modèle déterministe équivaut à 1,1. Une population en état d'équilibre (ni croissance, ni décroissance) révèle une valeur de λ égale à 1. Donc, connaître les niveaux de prélèvements maximum (n'impliquant pas de décroissance de la population), implique de connaître la valeur minimale de ρ (ρ_{min}) pour laquelle λ sera égal à 1. Toute valeur propre λ est solution de : $\det(A - \lambda I) = 0$ où I est la matrice identité.

Pour faire ceci, il faut résoudre l'équation suivante :

$$\prod_{i=1}^6 (R_i - \lambda) + (-1)^5 \cdot \rho \cdot \prod_{i=1}^5 C_i \cdot p_{6a} = 0 \text{ pour } \lambda = 1 \quad (12.13)$$

Nous obtenons à partir de l'équation (12.13), l'équation suivante :

$$\rho_{min} = \frac{\prod_{i=1}^6 (R_i - 1)}{\prod_{i=1}^5 C_i \cdot p_{6a}} = 2,23 \quad (12.14)$$

Sixième partie

Conclusion et perspectives

13 Conclusion	135
14 Perspectives	137

CHAPITRE 13

Conclusion

NOUS synthétiserons l'ensemble des conclusions que nous pouvons dégager de cette expérience. Ces conclusions porteront tout d'abord sur le contexte général des recherches effectuées, puis aborderont le contexte plus spécifique de l'étude de la dynamique de population du babaçu.

Sommaire

13.1 Autour du contexte général	135
13.2 Autour de la dynamique du babaçu	135

13.1 Autour du contexte général

Le contexte général était, nous l'avons bien énoncé, assez ambitieux, les partenaires étant de disciplines très variées. Malgré cela, nous pouvons être satisfaits de la manière dont les échanges ont pu se réaliser et, ceci, nous le pensons, grâce à notre proposition de méthodologie générale. Celle-ci semble avoir répondu à nos attentes et peut servir de cadre générique à tout projet interdisciplinaire relevant de l'étude de systèmes éco-socio-logiques.

L'aspect pluridisciplinaire de cette étude, nous a poussé vers une méthode de travail nécessitant de nombreux échanges volontaires dans le domaine du partage de connaissances. Pour ce faire, nous avons concentré nos efforts autour de la création d'un vocabulaire commun consensuel pour tous les partenaires.

Nous souhaitions également pérenniser les données issues du modèle d'observation. Pour ceci, nous les avons structurées et les avons intégrées au sein d'une base de données dont l'emprise était réfléchie et pensée plus large que le périmètre de l'étude de la dynamique de population du babaçu. Nous souhaitions y ajouter également des données provenant de sources diverses et pouvant être utilisées dans le cadre plus élargi du projet Open Science de l'Agropolis Fondation. La conceptualisation de cette base de données a été faite de manière que celle-ci puisse répondre aux besoins futurs en matière d'intégration et gestion des données issues des observations.

13.2 Autour de la dynamique du babaçu

Ce travail pionnier autour d'une espèce native était un vrai challenge, d'autant que, comme souligné, nous ne disposions au départ d'aucunes données de terrain.

Nous avons su surmonter cette première difficulté, et nous avons, même, su en tirer profit, puisque aussitôt que nous nous rendions compte de certaines incohérences et/ou insuffisances dans le protocole d'observation, nous apportions des modifications sur celui-ci et ceci dans l'objectif d'obtenir de meilleurs résultats pour les observations des années à venir.

Malgré cette incertitude liée au caractère d'obtention des données au fil du travail, nous avons su porter cette analyse jusqu'à la fin. Nous avons donc pu analyser la dynamique de population du palmier babaçu et en proposer un modèle intégrant au mieux les caractéristiques spécifiques de cette espèce.

Nous avons inscrit cette recherche autour du modèle de dynamique de population dans la lignée des modèles hiérarchiques. Cette décision a été guidée, d'une part par la structure de nos données, dont le caractère d'acquisition suivait sensiblement une hiérarchie (transect, parcelle, population), et d'autre part par le souhait d'effectuer une analyse permettant le transfert de l'information très variable à travers les différentes composantes du premier niveau hiérarchique (niveau plus proche des observations).

Le constat de cette grande variabilité d'un transect à un autre, nous a aussi incité, notamment, à orienter notre recherche vers le champ des modèles stochastiques permettant le traitement de celle-ci.

Ainsi, nous avons construit et proposé un modèle aléatoire hiérarchique inscrit dans le formalisme bayésien au sein duquel les paramètres du modèle sont traités à l'instar de variables aléatoires suivant des lois de probabilité.

CHAPITRE 14

Perspectives

CE DERNIER chapitre exposera un ensemble d'aspects se référant aux travaux en perspective qui s'ouvrent suite aux travaux avancés dans cette thèse. Deux principaux domaines, que nous examinerons successivement, sont concernés par les avancées proposées : le domaine de la capitalisation des données et connaissances et celui de la modélisation de dynamiques de populations végétales.

Sommaire

14.1 Autour de la capitalisation des données et connaissances	. 137
14.2 Autour du modèle de dynamique 137

14.1 Autour de la capitalisation des données et connaissances

Sur les perspectives ouvertes par la méthodologie nous retiendrons :

- Le fait que la production de connaissances devrait être capitalisée, par exemple, par la création d'un vocabulaire dédié à partir du catalogue des modèles descriptifs UML. La réflexion pourrait porter sur la traduction des termes de celui-ci en portugais (voire en anglais) qui pourrait déboucher sur un vocabulaire inscrit par exemple dans le cadre de l'initiative *Linked Open Vocabularies* (LOV) qui se positionne comme un observatoire de vocabulaires réutilisables permettant la valorisation de données ouvertes.
- De même manière l'analyse pré-ontologique des modèles de données, de connaissances et mathématiques pourrait être suivie d'une approche ontologique plus opérationnelle. La réflexion peut être poursuivie et partagée via des outils comme Webprotégé, ou permettre l'enrichissement du site AgroPortal (<http://agroportal.lirmm.fr/ontologies>)

14.2 Autour du modèle de dynamique

Grâce aux données de terrain de 2016, que nous n'avons malheureusement pas pu exploiter lors de ce travail de thèse, nous espérons pouvoir approfondir notre démarche de modélisation en janvier 2017. Nous avons prévu une session de travail

avec le doctorant brésilien travaillant sur la biologie de l'espèce Alessio Moreira dos Santos dans l'objectif de traiter ces données et d'enrichir notre modèle.

Au-delà du modèle mathématique, le projet Open Science a pour objectif l'aide à la gestion durable de cette espèce auprès des populations de la zone d'étude.

Nous pensons que le modèle proposé pourra aussi être transposé à d'autres espèces de palmiers, voire à d'autres domaines. Il nous semble qu'il peut aussi être adapté, dans le cadre d'études à une plus petite échelle (niveau des États Brésiliens, voire niveau national) pour mener des simulations à partir des données issues des techniques de la télédétection. Il nous paraît judicieux de prendre exemple sur des cas de figure proches du nôtre qui ont engendré de profondes modifications socio-culturelles au sein de populations natives dans d'autres régions planétaires. En nous plaçant à l'échelon d'analyse globale, nous pouvons citer l'exemple du développement du Karité et du passage d'une ressource d'extraction d'importance locale à un bien d'utilisation mondiale, posé sur des pratiques d'accès à la ressource plus que douteuses et très peu équitables. Nous nous interrogeons donc sur une possible évolution du système babaçu, vers un contexte de monopolisation du secteur par des sociétés multinationales provenant de l'industrie pharmaceutique et cosmétique, et donc à la place laissée aux populations locales et notamment aux *Quebradeiras* au sein d'une telle évolution...

Annexes

ANNEXE A

Graphiques densités

A.1 Pâturages

De même que pour les données de 2013, nous souhaitons pouvoir comparer les densités par transects pour les données de 2014 et 2015. Ainsi le TABLEAU A.1 fournit un résumé des densités observées par transect et par stade en 2014. Nous constatons que la variabilité intra-pâturage des valeurs obtenues n'est pas inférieure à la variabilité inter-pâturages.

TABLEAU A.1 – Densités par transect (pâturages 2014)

Code parcelles	Transect	Superficie (ares)	Densité (palmiers/are)					
			Stades					
			1	2	3	4	5	6
DEpasto_12	1	8	4,63	10,5	8,13	1,25	1,13	0,63
	2	5,2	11,73	14,23	7,88	0,58	0,77	0
	3	5	8,8	12,6	12,2	0,2	0,8	0,2
	suppl	8,4					0,48	1,43
CLpasto_10	1	33,6	0,36	2,86	0,89	0,09	0,03	0
	2	33,6	0,6	2,38	1,25	0,09	0,06	0
	suppl	137,38					0,04	0,03
CHpasto_10	1	4,4	5,68	27,73	0,45	0	0	0
	2	2	69	13	0	0	0	0
	suppl	183,14					0,01	0,07
RVpasto_08	1	6	7,33	18	7,83	0,17	0,17	0,5
	2	6	2,83	17,67	12	0	0	0
	3	3,4	1,18	25,88	30,88	0	0	0
	suppl	135,93					0,01	0,18
AMpasto_06	1	3,2	114,38	31,56	7,81	0	0	0,63
	2	2,8	0	14,29	12,86	0	0	0
	suppl	217,3					0	0,07
ARpasto_06	1	3,2	8,75	25,63	7,81	0	0	0,94
	2	5,4	0,93	19,07	6,3	0	0,56	1,11
	suppl	15,34					0,2	0,59
AMpasto_04	1	1,8	24,44	75,56	14,44	0	0	0,56
	2	5,6	15,71	18,39	3,93	0	0	0,18
	3	2	1,5	64,5	0,5	0,5	0	0,5
	suppl	93,38					0,06	0,12

TABLEAU A.2 – Densités par transect (pâturages 2015)

Code parcelle	Transect	Superficie (ares)	Densité (palmiers/are)					
			Stades					
			1	2	3	4	5	6
DEpasto_12	1	8	6,13	9,25	10,38	1,13	1,13	0,63
	2	5,2	7,88	16,54	9,04	0,58	0,96	0
	3	5	9,8	9,8	13,8	0,2	0,8	0,4
	suppl	8,4					0,48	1,43
CLpasto_10	1	33,6	0,06	2,23	1,31	0,3	0,06	0
	2	33,6	0,06	1,85	1,67	0,15	0,12	0
	suppl	137,38					0,04	0,04
CHpasto_10	1	4,4	3,86	28,41	0,68	0	0	0
	2	2	53,5	15,5	0	0	0	0
	suppl	183,14					0,01	0,07
RVpasto_08	1	6	38	19,67	8,83	0,33	0,17	0,5
	2	6	1,67	16,33	14,17	0	0	0
	3	3,4	0,88	24,41	32,65	0	0	0
	suppl	135,93					0,01	0,18
AMpasto_06	1	3,2	91,25	33,13	9,06	0	0	0,63
	2	2,8	0	10,36	16,79	0	0	0
	suppl	217,3					0	0,07
ARpasto_06	1	3,2	10,94	26,56	8,44	0	0	0,94
	2	5,4	3,15	17,41	6,3	0	0,37	1,3
	suppl	15,34					0,2	0,59
AMpasto_04	1	1,8	48,33	68,89	23,89	0	0	0,56
	2	5,6	14,29	16,07	6,96	0	0	0,18
	3	2	2,5	61	2,5	0,5	0	0,5
	suppl	93,38					0,05	0,13

B.1 Commandes génériques

```
- Installer PostgreSQL
$ sudo apt-get install postgresql-9.3
- Installer PostGIS
$ sudo apt-get install -y postgis postgresql-9.3-postgis-2.1
- Configurer le fichier pour ne pas demander de mot de passe pour "postgres" (rem-
placer Md5 ou peer par "trust")
$ cd /etc/postgresql/9.3/main
$ sudo gedit pg_hba.conf
- Se connecter en tant que postgres
$ sudo su postgres
$ psql postgres
- Création du nouveau role
$ CREATE USER sirakov WITH PASSWORD 'X' CREATEDB CREATEUSER;
- Création de la BD
$ CREATE DATABASE babassu OWNER sirakov;
- Dans un terminal exécution du script
$ cd /home/sirakov/Doctorat/Data/babassu_db
$ sh 2_commandes_terminal.sh
```

B.2 Script bash « 2_commandes_terminal.sh »

```
#!/bin/bash
```

```
Rscript 3_mise_forme_donnees.R
```

```
psql -d babassu -h localhost -U sirakov -c 'CREATE EXTENSION postgis;'
```

```
psql -d babassu -h localhost -U sirakov -f /home/sirakov/Doctorat
```

```
/Data/babassu_db/4_creation.sql
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil/pays  
/pays_SIRGAS2000.shp public.pays | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil/regions  
/regions_SIRGAS2000.shp public.regions | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil/etats  
/etats_SIRGAS2000.shp public.etats | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil  
/municipios_2010/municipios_2010_UTF8.shp  
public.municipalites_2010 | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil/Benfica  
/Limites/limites_benfica1_2.shp  
public.projeto_assentamento | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil/parcelles  
/parcelles.shp public.parcelles | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil  
/occurence_babassu/occurence_babassu_SIRGAS2000.shp  
public.occurence_babassu | psql -d babassu
```

```
shp2pgsql -a -s 4674 /home/sirakov/Doctorat/SIG_Brasil/biomes  
/biomes.shp public.biomes | psql -d babassu
```

```
psql -d babassu -h localhost -U sirakov -f /home/sirakov  
/Doctorat/Data/babassu_db/5_insertion.sql
```

```
psql -d babassu -h localhost -U sirakov -f /home/sirakov  
/Doctorat/Data/babassu_db/6_verification.sql
```

```
psql -d babassu -h localhost -U sirakov -f /home/sirakov  
/Doctorat/Data/babassu_db/7_requetes_creation.sql
```

```
shp2pgsql -s 4674 /home/sirakov/Doctorat/SIG_Brasil/transects  
/transects.shp temp_s | psql -d babassu
```

```
shp2pgsql -s 4674 /home/sirakov/Doctorat/SIG_Brasil/transects  
/transects_p.shp transect_p | psql -d babassu
```

```
psql -d babassu -h localhost -U sirakov -f /home/sirakov/Doctorat/Data/babassu_db/8_colonnes_geom.sql
```

```
Rscript 9_multinom_hierar_JAGS.R
```

B.3 Script R « 3_mise_forme_donnees.R »

```
require(xlsx)
donnees_2013 <- data.frame(read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Data_babassu_2013_prepare.xls", sheetName = "total"))

donnees_2014 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Data_babassu_2014_prepare.xls", sheetName = "Donnees2014")

donnees_2015 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Data_babassu_2015_prepare.xls", sheetName = "Donnees2015")

fruits_sol_2013 <- data.frame(read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Data_fruits_2013_prepare.xls", sheetName = "Fruits_sol"))

fruits_seul_2013 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Data_fruits_2013_prepare.xls", sheetName = "Fruits_seuls")

fruits_infr_2013 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Data_fruits_2013_prepare.xls", sheetName = "Infrutescence")

fruits_sol_2014 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Donnees2014_fruits.xls", sheetName = "Fruits_sol")

fruits_seul_2014 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Donnees2014_fruits.xls", sheetName = "Fruits_seuls")

fruits_infr_2014 <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Donnees2014_fruits.xls", sheetName = "Infrutescence")

fruits_infl <- read.xlsx("/home/sirakov/Doctorat/Data/data_brut/Inflorescence_prepare.xls", sheetName = "Feuil2")

babassu_geo <- data.frame(read.xlsx("/home/sirakov/Doctorat/Data/data_brut/PontosBenficacomestagios.xls", sheetName = "adult_export"))

transect_geo <- data.frame(read.xlsx("/home/sirakov/Doctorat/Data
```

```
"/data_brut/PontosBenficacomestagios.xls", sheetName = "transect_export"
```

```
write.table(donnees_2013, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_babassu_2013_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(donnees_2014, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_babassu_2014_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(donnees_2015, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_babassu_2015_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_sol_2013, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_fruits_sol_2013_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_seul_2013, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_fruits_seul_2013_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_infr_2013, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_fruits_infr_2013_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_sol_2014, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_fruits_sol_2014_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_seul_2014, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_fruits_seul_2014_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_infr_2014, file = "/home/sirakov/Doctorat/Data  
/data_brut/Data_fruits_infr_2014_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(fruits_infl, file = "/home/sirakov/Doctorat/Data  
/data_brut/Infl_prepare.csv", row.names=FALSE,  
na="", col.names=TRUE, quote=FALSE, sep=",")
```

```
write.table(babassu_geo, file = "/home/sirakov/Doctorat/Data
```

```
/data_brut/babassu_geo.csv",row.names=FALSE,
na="",col.names=TRUE,quote=FALSE,sep=",")
```

```
write.table(transect_geo, file = "/home/sirakov/Doctorat/Data
/data_brut/transect_geo.csv",row.names=FALSE,
na="",col.names=TRUE,quote=FALSE,sep=",")
```

B.4 Script SQL « 4_creation.sql »

2013

Supprimer les tables

```
DROP TABLE IF EXISTS individus CASCADE;
DROP TABLE IF EXISTS parcelles CASCADE;
DROP TABLE IF EXISTS stade CASCADE;
DROP TABLE IF EXISTS fruits_sol;
DROP TABLE IF EXISTS fruits_seuls;
DROP TABLE IF EXISTS infrutescence;
DROP TABLE IF EXISTS inflorescence;
DROP TABLE IF EXISTS etats CASCADE;
DROP TABLE IF EXISTS municipalites_2010 CASCADE;
DROP TABLE IF EXISTS regions CASCADE;
DROP TABLE IF EXISTS pays CASCADE;
DROP TABLE IF EXISTS projeto_assentamento CASCADE;
DROP TABLE IF EXISTS occurrence_babassu CASCADE;
DROP TABLE IF EXISTS biomes CASCADE;
```

Creation des tables

```
CREATE TABLE public.individus (
num INTEGER PRIMARY KEY,
parcelle VARCHAR(15),
transect_code VARCHAR(10),
transect INTEGER,
stade_2013 INTEGER,
stade_2014 INTEGER,
stade_2015 INTEGER,
stade_2016 INTEGER,
age_2013 INTEGER,
age_2014 INTEGER,
age_2015 INTEGER,
age_2016 INTEGER,
utilisation_sol VARCHAR(20),
annee_installation VARCHAR(4),
```

```
releve_2013 DATE,  
releve_2014 DATE,  
releve_2015 DATE,  
releve_2016 DATE,  
coupe_2013 VARCHAR(1),  
coupe_2014 VARCHAR(1),  
coupe_2015 VARCHAR(1),  
coupe_2016 VARCHAR(1)  
);
```

```
CREATE TABLE public.stade (  
id SERIAL PRIMARY KEY,  
num_stade INTEGER,  
nom_stade VARCHAR (20),  
nom_raccourci VARCHAR (2)  
);
```

```
CREATE TABLE public.fruits_sol (  
num_adulte INTEGER PRIMARY KEY,  
Nb_fr_sol_anc_2_5_m INTEGER,  
Nb_fr_sol_2013_2_5_m INTEGER,  
Nb_fr_sol_2014_2_5_m INTEGER,  
Nb_fr_sol_2015_2_5_m INTEGER,  
Nb_fr_sol_2016_2_5_m INTEGER,  
Nb_fr_sol_anc_5_m INTEGER,  
Nb_fr_sol_2013_5_m INTEGER,  
Nb_fr_sol_2014_5_m INTEGER,  
Nb_fr_sol_2015_5_m INTEGER,  
Nb_fr_sol_2016_5_m INTEGER,  
FOREIGN KEY (num_adulte) references individus(num)  
);
```

```
CREATE TABLE public.fruits_seuls (  
Nb_fr_seul_anc INTEGER,  
Nb_fr_seul_2013 INTEGER,  
Nb_fr_seul_2014 INTEGER,  
Nb_fr_seul_2015 INTEGER,  
Nb_fr_seul_2016 INTEGER,  
num_tr INTEGER,  
transect_fr VARCHAR(15),  
parcelle_fr VARCHAR(10)  
);
```

```
CREATE TABLE public.infrutescence (  
  num_infr INTEGER PRIMARY KEY,  
  num_adulte INTEGER,  
  annee_releve VARCHAR(4),  
  nb_fr_infr INTEGER,  
FOREIGN KEY (num_adulte) references individus(num)  
);
```

```
CREATE TABLE public.inflorescence (  
  num_adulte INTEGER,  
  annee_releve VARCHAR(4),  
  nb_trace_m INTEGER,  
  nb_infl_m INTEGER,  
  nb_trace_f INTEGER,  
  nb_infl_f INTEGER,  
CONSTRAINT num_annee PRIMARY KEY(num_adulte, annee_releve),  
FOREIGN KEY (num_adulte) references individus(num)  
);
```

```
CREATE TABLE public.pays (  
  nom VARCHAR(254) PRIMARY KEY,  
  sigle VARCHAR(2),  
  nom_en VARCHAR(254),  
  geom geometry,  
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),  
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =  
  'MULTIPOLYGON'::text OR geom IS NULL),  
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674)  
);
```

```
CREATE TABLE public.regions (  
  nom VARCHAR(254) PRIMARY KEY,  
  sigle VARCHAR(2),  
  pays_nom VARCHAR(254),  
  geom geometry,  
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),  
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =  
  'MULTIPOLYGON'::text OR geom IS NULL),  
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),  
FOREIGN KEY (pays_nom) references pays(nom)  
);
```

```
CREATE TABLE public.etats (  
  nom VARCHAR(254) PRIMARY KEY,  
  sigle VARCHAR(2),  
  pays_nom VARCHAR(254),  
  geom geometry,  
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),  
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =  
  'MULTIPOLYGON'::text OR geom IS NULL),  
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),  
FOREIGN KEY (pays_nom) references pays(nom)  
);
```

```

id INTEGER PRIMARY KEY,
nom VARCHAR(254),
sigle VARCHAR(2),
region_nom VARCHAR(254),
code_ibg INTEGER,
geom geometry,
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =
'MULTIPOLYGON'::text OR geom IS NULL),
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),
FOREIGN KEY (region_nom) references regions(nom)
);

```

```

CREATE TABLE public.municipalites_2010 (
id INTEGER PRIMARY KEY,
nome VARCHAR(254),
uf VARCHAR(2),
populacao INTEGER,
pib INTEGER,
estado_id INTEGER,
codigo_ibg INTEGER,
geom geometry,
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =
'MULTIPOLYGON'::text OR geom IS NULL),
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),
FOREIGN KEY (estado_id) references etats(id)
);

```

```

CREATE TABLE public.projeto_assentamento (
id INTEGER PRIMARY KEY,
nom VARCHAR(254),
muni_id INTEGER,
geom geometry,
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =
'MULTIPOLYGON'::text OR geom IS NULL),
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),
FOREIGN KEY (muni_id) references municipalites_2010(id)
);

```

```

CREATE TABLE public.parcelles (
nom VARCHAR(15) PRIMARY KEY,

```

```

benfica_id INTEGER,
muni_id INTEGER,
geom geometry,
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =
'MULTIPOLYGON'::text OR geom IS NULL),
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),
FOREIGN KEY (benfica_id) references projeto_assentamento(id),
FOREIGN KEY (muni_id) references municipalites_2010(id)
);

```

```

CREATE TABLE public.occurrence_babassu (
id INTEGER PRIMARY KEY,
def_year VARCHAR(4),
pays_nom VARCHAR(254),
geom geometry,
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =
'MULTIPOLYGON'::text OR geom IS NULL),
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),
FOREIGN KEY (pays_nom) references pays(nom)
);

```

```

CREATE TABLE public.biomes (
id INTEGER PRIMARY KEY,
nom VARCHAR(20),
pays_nom VARCHAR(254),
geom geometry,
CONSTRAINT enforce_dims_geom CHECK (st_ndims(geom) = 2),
CONSTRAINT enforce_geotype_geom CHECK (geometrytype(geom) =
'MULTIPOLYGON'::text OR geom IS NULL),
CONSTRAINT enforce_srid_geom CHECK (st_srid(geom) = 4674),
FOREIGN KEY (pays_nom) references pays(nom)
);

```

2014

Supprimer les tables

```

—DROP TABLE IF EXISTS fruits_sol_14;
—DROP TABLE IF EXISTS fruits_seuls_14;

```

Creation des tables

```

CREATE TABLE public.individus_14 (

```

```

num_13 INTEGER PRIMARY KEY,
num_14 INTEGER,
parcelle_14 VARCHAR(15),
transect_14 VARCHAR(10),
stade_14 VARCHAR(10),
utilisation_sol_14 VARCHAR(20),
annee_installation_14 VARCHAR(4),
releve_14 DATE,
coupe_14 VARCHAR(2)
);

```

```

CREATE TABLE public.fruits_sol_14 (
num_adulte INTEGER PRIMARY KEY,
Nb_fr_sol_2014_2_5_m INTEGER,
Nb_fr_sol_2014_5_m INTEGER,
FOREIGN KEY (num_adulte) references individus(num)
);

```

```

CREATE TABLE public.fruits_seuls_14 (
Nb_fr_seul_2014 INTEGER,
num_tr INTEGER,
transect_fr VARCHAR(15),
parcelle_fr VARCHAR(10)
);

```

2015

————— *Creation des tables* —————

```

CREATE TABLE public.individus_15 (
num_13 INTEGER PRIMARY KEY,
num_14 INTEGER,
num_15 INTEGER,
parcelle_15 VARCHAR(15),
transect_15 VARCHAR(10),
stade_15 VARCHAR(10),
utilisation_sol_15 VARCHAR(20),
annee_installation_15 VARCHAR(4),
releve_15 DATE,
coupe_15 VARCHAR(2)
);

```

B.5 Script SQL « 5_insertion.sql »

```

----- Renommer les colonnes -----
ALTER TABLE pays RENAME geom TO the_geom;

ALTER TABLE regions RENAME geom TO the_geom;

ALTER TABLE etats RENAME geom TO the_geom;

ALTER TABLE municipalites_2010 RENAME nome TO nom;
ALTER TABLE municipalites_2010 RENAME populacao TO population;
ALTER TABLE municipalites_2010 RENAME estado_id TO etat_id;
ALTER TABLE municipalites_2010 RENAME codigo_ibg TO code_ibg;
ALTER TABLE municipalites_2010 RENAME geom TO the_geom;

ALTER TABLE projeto_assentamento RENAME muni_id TO municipalite_id;
ALTER TABLE projeto_assentamento RENAME geom TO the_geom;

ALTER TABLE parcelles RENAME benefica_id TO projeto_id;
ALTER TABLE parcelles RENAME muni_id TO municipalite_id;
ALTER TABLE parcelles RENAME geom TO the_geom;
ALTER TABLE parcelles ADD COLUMN utilisation_sol VARCHAR(20);
ALTER TABLE parcelles ADD COLUMN annee_installation VARCHAR(4);

ALTER TABLE occurrence_babassu RENAME geom TO the_geom;

ALTER TABLE biomes RENAME geom TO the_geom;

----- 2013 -----

----- Import des donnees a partir du fichier .csv -----
COPY individus FROM '/home/sirakov/Doctorat/Data/data_brut/Data_babassu_2013_prepare.csv' CSV HEADER;
COPY fruits_sol FROM '/home/sirakov/Doctorat/Data/data_brut/Data_fruits_sol_2013_prepare.csv' CSV HEADER;
COPY fruits_seuls FROM '/home/sirakov/Doctorat/Data/data_brut/Data_fruits_seul_2013_prepare.csv' CSV HEADER;
COPY infrutescence FROM '/home/sirakov/Doctorat/Data/data_brut/Data_fruits_infr_2013_prepare.csv' CSV HEADER;
COPY inflorescence FROM '/home/sirakov/Doctorat/Data/data_brut/Infl_prepare.csv' CSV HEADER;

----- Mise a jour des colonnes -----
UPDATE individus SET parcelle = 'AMpasto_04' WHERE parcelle = '_AMpasto_04';
UPDATE individus SET parcelle = 'DEpasto_12' WHERE parcelle = '_DEpasto_12';
UPDATE individus SET parcelle = 'DEpasto_12' WHERE parcelle = 'DEpasto_12_';
UPDATE individus
  SET transect =
  (CASE
    WHEN parcelle = 'DEpasto_12' AND transect_code = '1' THEN 1
    WHEN parcelle = 'DEpasto_12' AND transect_code = '2' THEN 2
    WHEN parcelle = 'DEpasto_12' AND transect_code = '3' THEN 3
    WHEN parcelle = 'DEpasto_12' AND transect_code = 'supplem.' THEN 18
    WHEN parcelle = 'CLpasto_10' AND transect_code = '1' THEN 4
    WHEN parcelle = 'CLpasto_10' AND transect_code = '2' THEN 5
    WHEN parcelle = 'CLpasto_10' AND transect_code = 'supplem.' THEN 19
    WHEN parcelle = 'CHpasto_10' AND transect_code = '1' THEN 6
    WHEN parcelle = 'CHpasto_10' AND transect_code = '2' THEN 7
    WHEN parcelle = 'CHpasto_10' AND transect_code = 'supplem.' THEN 20
    WHEN parcelle = 'RVpasto_08' AND transect_code = '1' THEN 8
    WHEN parcelle = 'RVpasto_08' AND transect_code = '2' THEN 9
    WHEN parcelle = 'RVpasto_08' AND transect_code = '3' THEN 10
    WHEN parcelle = 'RVpasto_08' AND transect_code = 'supplem.' THEN 21
    WHEN parcelle = 'AMpasto_06' AND transect_code = '1' THEN 11
  )

```

```

WHEN parcelle = 'AMpasto_06' AND transect_code = '2' THEN 12
WHEN parcelle = 'AMpasto_06' AND transect_code = 'supplem.' THEN 22
WHEN parcelle = 'ARpasto_06' AND transect_code = '1' THEN 13
WHEN parcelle = 'ARpasto_06' AND transect_code = '2' THEN 14
WHEN parcelle = 'ARpasto_06' AND transect_code = 'supplem.' THEN 23
WHEN parcelle = 'AMpasto_04' AND transect_code = '1' THEN 15
WHEN parcelle = 'AMpasto_04' AND transect_code = '2' THEN 16
WHEN parcelle = 'AMpasto_04' AND transect_code = '3' THEN 17
WHEN parcelle = 'AMpasto_04' AND transect_code = 'supplem.' THEN 24
WHEN parcelle = 'RI_bab' AND transect_code = '1' THEN 25
WHEN parcelle = 'RI_bab' AND transect_code = '2' THEN 26
WHEN parcelle = 'ZM_bab' AND transect_code = '1' THEN 27
WHEN parcelle = 'ZM_bab' AND transect_code = '2' THEN 28
WHEN parcelle = 'DE_flo' AND transect_code = '1' THEN 29
WHEN parcelle = 'DE_flo' AND transect_code = '2' THEN 30
WHEN parcelle = 'DE_flo' AND transect_code = 'supplem.' THEN 31
WHEN parcelle = 'ME_flo' AND transect_code = '1' THEN 32
WHEN parcelle = 'ME_flo' AND transect_code = '2' THEN 33
WHEN parcelle = 'ME_flo' AND transect_code = 'supplem.' THEN 34
ELSE 1000
END);
UPDATE individus SET age_2013 = -1;
UPDATE fruits_seuls SET parcelle_fr = 'DE_flo' WHERE parcelle_fr = '_DE_flo';
UPDATE fruits_seuls SET num_tr =
(CASE
WHEN parcelle_fr = 'DEpasto_12' AND transect_fr = '1' THEN 1
WHEN parcelle_fr = 'DEpasto_12' AND transect_fr = '2' THEN 2
WHEN parcelle_fr = 'DEpasto_12' AND transect_fr = '3' THEN 3
WHEN parcelle_fr = 'DEpasto_12' AND transect_fr = 'supplem.' THEN 18
WHEN parcelle_fr = 'CLpasto_10' AND transect_fr = '1' THEN 4
WHEN parcelle_fr = 'CLpasto_10' AND transect_fr = '2' THEN 5
WHEN parcelle_fr = 'CLpasto_10' AND transect_fr = 'supplem.' THEN 19
WHEN parcelle_fr = 'CHpasto_10' AND transect_fr = '1' THEN 6
WHEN parcelle_fr = 'CHpasto_10' AND transect_fr = '2' THEN 7
WHEN parcelle_fr = 'CHpasto_10' AND transect_fr = 'supplem.' THEN 20
WHEN parcelle_fr = 'RVpasto_08' AND transect_fr = '1' THEN 8
WHEN parcelle_fr = 'RVpasto_08' AND transect_fr = '2' THEN 9
WHEN parcelle_fr = 'RVpasto_08' AND transect_fr = '3' THEN 10
WHEN parcelle_fr = 'RVpasto_08' AND transect_fr = 'supplem.' THEN 21
WHEN parcelle_fr = 'AMpasto_06' AND transect_fr = '1' THEN 11
WHEN parcelle_fr = 'AMpasto_06' AND transect_fr = '2' THEN 12
WHEN parcelle_fr = 'AMpasto_06' AND transect_fr = 'supplem.' THEN 22
WHEN parcelle_fr = 'ARpasto_06' AND transect_fr = '1' THEN 13
WHEN parcelle_fr = 'ARpasto_06' AND transect_fr = '2' THEN 14
WHEN parcelle_fr = 'ARpasto_06' AND transect_fr = 'supplem.' THEN 23
WHEN parcelle_fr = 'AMpasto_04' AND transect_fr = '1' THEN 15
WHEN parcelle_fr = 'AMpasto_04' AND transect_fr = '2' THEN 16
WHEN parcelle_fr = 'AMpasto_04' AND transect_fr = '3' THEN 17
WHEN parcelle_fr = 'AMpasto_04' AND transect_fr = 'supplem.' THEN 24
WHEN parcelle_fr = 'RI_bab' AND transect_fr = '1' THEN 25
WHEN parcelle_fr = 'RI_bab' AND transect_fr = '2' THEN 26
WHEN parcelle_fr = 'ZM_bab' AND transect_fr = '1' THEN 27
WHEN parcelle_fr = 'ZM_bab' AND transect_fr = '2' THEN 28
WHEN parcelle_fr = 'DE_flo' AND transect_fr = '1' THEN 29
WHEN parcelle_fr = 'DE_flo' AND transect_fr = '2' THEN 30
WHEN parcelle_fr = 'DE_flo' AND transect_fr = 'supplem.' THEN 31
WHEN parcelle_fr = 'ME_flo' AND transect_fr = '1' THEN 32
WHEN parcelle_fr = 'ME_flo' AND transect_fr = '2' THEN 33

```

```

        WHEN parcelle_fr = 'ME_flo' AND transect_fr = 'supplem.' THEN 34
        ELSE 1000
    END);
ALTER TABLE fruits_seuls ADD PRIMARY KEY (num_tr);

```

----- 2014 -----

----- Import des donnees a partir du fichier .csv -----

```

COPY individus_14 FROM '/home/sirakov/Doctorat/Data/data_brut/Data_babassu-2014_prepare.csv' CSV HEADER;
COPY fruits_sol_14 FROM '/home/sirakov/Doctorat/Data/data_brut/Data_fruits_sol_2014_prepare.csv' CSV HEADER;
COPY fruits_seuls_14 FROM '/home/sirakov/Doctorat/Data/data_brut/Data_fruits_seul_2014_prepare.csv' CSV HEADER;
COPY infrutescence FROM '/home/sirakov/Doctorat/Data/data_brut/Data_fruits_infr_2014_prepare.csv' CSV HEADER;

```

----- Mise a jour des colonnes -----

```

UPDATE individus_14 SET parcelle_14 = 'ARpasto.06' WHERE parcelle_14 = 'Arpasto.06';
UPDATE individus_14 SET parcelle_14 = 'DEpasto.12' WHERE parcelle_14 = 'Depasto.12';
UPDATE individus_14 SET parcelle_14 = 'DEpasto.12' WHERE parcelle_14 = 'DEpasto.12_';
UPDATE individus_14 SET parcelle_14 = 'ME_flo' WHERE parcelle_14 = '_ME_flo';
UPDATE individus_14 SET parcelle_14 = 'RI_bab' WHERE parcelle_14 = '_RI_bab';
UPDATE individus_14 SET coupe_14 = 'M' WHERE stade_14 = 'MORTO';
UPDATE individus_14 SET coupe_14 = 'P' WHERE stade_14 = 'NE';
UPDATE individus_14 SET coupe_14 = 'A' WHERE stade_14 = 'INAJA';
UPDATE individus_14 SET coupe_14 = 'I' WHERE coupe_14 = '12' OR coupe_14 = '13' OR coupe_14 = '14' OR coupe_14 = '15' OR coupe_14 = '17' OR coupe_14 = '18';
UPDATE individus_14 SET coupe_14 = 'C' WHERE coupe_14 = 'S' OR coupe_14 = 'S';
UPDATE individus_14 SET stade_14 = NULL WHERE stade_14 = 'MORTO';
UPDATE individus_14 SET stade_14 = NULL WHERE stade_14 = 'NE';
UPDATE individus_14 SET stade_14 = NULL WHERE stade_14 = 'INAJA';
ALTER TABLE individus_14 ALTER COLUMN stade_14 TYPE integer USING (stade_14::integer);

```

```

UPDATE fruits_seuls_14
    SET num_tr =
    (CASE
        WHEN parcelle_fr = 'DEpasto.12' AND transect_fr = '1' THEN 1
        WHEN parcelle_fr = 'DEpasto.12' AND transect_fr = '2' THEN 2
        WHEN parcelle_fr = 'DEpasto.12' AND transect_fr = '3' THEN 3
        WHEN parcelle_fr = 'DEpasto.12' AND transect_fr = 'supplem.' THEN 18
        WHEN parcelle_fr = 'CLpasto.10' AND transect_fr = '1' THEN 4
        WHEN parcelle_fr = 'CLpasto.10' AND transect_fr = '2' THEN 5
        WHEN parcelle_fr = 'CLpasto.10' AND transect_fr = 'supplem.' THEN 19
        WHEN parcelle_fr = 'CHpasto.10' AND transect_fr = '1' THEN 6
        WHEN parcelle_fr = 'CHpasto.10' AND transect_fr = '2' THEN 7
        WHEN parcelle_fr = 'CHpasto.10' AND transect_fr = 'supplem.' THEN 20
        WHEN parcelle_fr = 'RVpasto.08' AND transect_fr = '1' THEN 8
        WHEN parcelle_fr = 'RVpasto.08' AND transect_fr = '2' THEN 9
        WHEN parcelle_fr = 'RVpasto.08' AND transect_fr = '3' THEN 10
        WHEN parcelle_fr = 'RVpasto.08' AND transect_fr = 'supplem.' THEN 21
        WHEN parcelle_fr = 'AMPasto.06' AND transect_fr = '1' THEN 11
        WHEN parcelle_fr = 'AMPasto.06' AND transect_fr = '2' THEN 12
        WHEN parcelle_fr = 'AMPasto.06' AND transect_fr = 'supplem.' THEN 22
        WHEN parcelle_fr = 'ARpasto.06' AND transect_fr = '1' THEN 13
        WHEN parcelle_fr = 'ARpasto.06' AND transect_fr = '2' THEN 14
        WHEN parcelle_fr = 'ARpasto.06' AND transect_fr = 'supplem.' THEN 23
        WHEN parcelle_fr = 'AMPasto.04' AND transect_fr = '1' THEN 15
        WHEN parcelle_fr = 'AMPasto.04' AND transect_fr = '2' THEN 16
        WHEN parcelle_fr = 'AMPasto.04' AND transect_fr = '3' THEN 17
        WHEN parcelle_fr = 'AMPasto.04' AND transect_fr = 'supplem.' THEN 24
        WHEN parcelle_fr = 'RI_bab' AND transect_fr = '1' THEN 25
        WHEN parcelle_fr = 'RI_bab' AND transect_fr = '2' THEN 26
        WHEN parcelle_fr = 'ZM_bab' AND transect_fr = '1' THEN 27
        WHEN parcelle_fr = 'ZM_bab' AND transect_fr = '2' THEN 28
    )

```

```

WHEN parcelle_fr = 'DE_flo' AND transect_fr = '1' THEN 29
WHEN parcelle_fr = 'DE_flo' AND transect_fr = '2' THEN 30
WHEN parcelle_fr = 'DE_flo' AND transect_fr = 'supplem.' THEN 31
WHEN parcelle_fr = 'ME_flo' AND transect_fr = '1' THEN 32
WHEN parcelle_fr = 'ME_flo' AND transect_fr = '2' THEN 33
WHEN parcelle_fr = 'ME_flo' AND transect_fr = 'supplem.' THEN 34
ELSE 1000
END);
ALTER TABLE fruits_seuls_14 ADD PRIMARY KEY (num_tr);

----- Inserer les nouvelles donnees au sein de la table "individus" -----
INSERT INTO individus (SELECT num_13 FROM individus_14 WHERE num_13 > 5000);

----- Charger donnees de la table "individus_14" vers la table "individus" -----
UPDATE individus i SET coupe_2014 = (SELECT i_14.coupe_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num);
UPDATE individus i SET releve_2014 = (SELECT i_14.releve_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num);
UPDATE individus i SET stade_2014 = (SELECT i_14.stade_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num);
UPDATE individus i SET parcelle = (SELECT i_14.parcelle_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) WHERE num > 5000;
UPDATE individus i SET transect_code = (SELECT i_14.transect_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) WHERE num > 5000;
UPDATE individus i SET utilisation_sol = (SELECT i_14.utilisation_sol_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) WHERE num > 5000;
UPDATE individus i SET annee_installation = (SELECT i_14.annee_installation_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) WHERE num > 5000;

----- Charger donnees de la table "fruits_seuls_14" vers la table "fruits_seuls" -----
UPDATE fruits_seuls fs SET nb_fr_seul_2014 = (SELECT fs_14.nb_fr_seul_2014 FROM fruits_seuls_14 fs_14 WHERE fs_14.num_tr = fs.num_tr);
ALTER TABLE fruits_seuls DROP COLUMN transect_fr;
ALTER TABLE fruits_seuls DROP COLUMN parcelle_fr;
DROP TABLE IF EXISTS fruits_seuls_14;

----- Inserer les nouvelles donnees au sein de la table "fruits_sol" -----
INSERT INTO fruits_sol (SELECT num_adulte FROM fruits_sol_14 WHERE num_adulte = 1809 OR num_adulte = 1812);

----- Charger donnees de la table "fruits_sol_14" vers la table "fruits_sol" -----
UPDATE fruits_sol fsol SET nb_fr_sol_2014_2_5_m = (SELECT fsol_14.nb_fr_sol_2014_2_5_m FROM fruits_sol_14 fsol_14 WHERE fsol_14.num_adulte = fsol.num_adulte);
UPDATE fruits_sol fsol SET nb_fr_sol_2014_5_m = (SELECT fsol_14.nb_fr_sol_2014_5_m FROM fruits_sol_14 fsol_14 WHERE fsol_14.num_adulte = fsol.num_adulte);
DROP TABLE IF EXISTS fruits_sol_14;

----- Mise a jour de la colonne "transect_num" -----
UPDATE individus
  SET transect =
    (CASE
      WHEN parcelle = 'DEpasto_12' AND transect_code = '1' THEN 1
      WHEN parcelle = 'DEpasto_12' AND transect_code = '2' THEN 2
      WHEN parcelle = 'DEpasto_12' AND transect_code = '3' THEN 3
      WHEN parcelle = 'DEpasto_12' AND transect_code = 'supplem.' THEN 18
      WHEN parcelle = 'CLpasto_10' AND transect_code = '1' THEN 4
      WHEN parcelle = 'CLpasto_10' AND transect_code = '2' THEN 5
      WHEN parcelle = 'CLpasto_10' AND transect_code = 'supplem.' THEN 19
      WHEN parcelle = 'CHpasto_10' AND transect_code = '1' THEN 6
      WHEN parcelle = 'CHpasto_10' AND transect_code = '2' THEN 7
      WHEN parcelle = 'CHpasto_10' AND transect_code = 'supplem.' THEN 20
      WHEN parcelle = 'RVpasto_08' AND transect_code = '1' THEN 8
      WHEN parcelle = 'RVpasto_08' AND transect_code = '2' THEN 9
      WHEN parcelle = 'RVpasto_08' AND transect_code = '3' THEN 10
      WHEN parcelle = 'RVpasto_08' AND transect_code = 'supplem.' THEN 21
      WHEN parcelle = 'AMPasto_06' AND transect_code = '1' THEN 11
      WHEN parcelle = 'AMPasto_06' AND transect_code = '2' THEN 12
      WHEN parcelle = 'AMPasto_06' AND transect_code = 'supplem.' THEN 22
      WHEN parcelle = 'ARPasto_06' AND transect_code = '1' THEN 13
      WHEN parcelle = 'ARPasto_06' AND transect_code = '2' THEN 14
      WHEN parcelle = 'ARPasto_06' AND transect_code = 'supplem.' THEN 23
    )

```

```

WHEN parcelle = 'AMpasto_04' AND transect_code = '1' THEN 15
WHEN parcelle = 'AMpasto_04' AND transect_code = '2' THEN 16
WHEN parcelle = 'AMpasto_04' AND transect_code = '3' THEN 17
WHEN parcelle = 'AMpasto_04' AND transect_code = 'supplem.' THEN 24
WHEN parcelle = 'RI_bab' AND transect_code = '1' THEN 25
WHEN parcelle = 'RI_bab' AND transect_code = '2' THEN 26
WHEN parcelle = 'ZM_bab' AND transect_code = '1' THEN 27
WHEN parcelle = 'ZM_bab' AND transect_code = '2' THEN 28
WHEN parcelle = 'ZM_bab' AND transect_code = '2' THEN 28
WHEN parcelle = 'DE_flo' AND transect_code = '1' THEN 29
WHEN parcelle = 'DE_flo' AND transect_code = '2' THEN 30
WHEN parcelle = 'DE_flo' AND transect_code = 'supplem.' THEN 31
WHEN parcelle = 'ME_flo' AND transect_code = '1' THEN 32
WHEN parcelle = 'ME_flo' AND transect_code = '2' THEN 33
WHEN parcelle = 'ME_flo' AND transect_code = 'supplem.' THEN 34
ELSE 1000
END) WHERE num > 5000;
----- Mise a jour de la colonne "age_2014" -----
UPDATE individus
  SET age_2014 =
    (CASE
      WHEN stade_2013 = '1' AND stade_2014 = '1' THEN 1
      WHEN stade_2013 = '2' AND stade_2014 = '2' THEN 1
      WHEN stade_2013 = '3' AND stade_2014 = '3' THEN 1
      WHEN stade_2013 = '4' AND stade_2014 = '4' THEN 1
      WHEN stade_2013 = '5' AND stade_2014 = '5' THEN 1
      WHEN stade_2013 = '6' AND stade_2014 = '6' THEN 1
      WHEN stade_2013 = '1' AND stade_2014 = '2' THEN 0
      WHEN stade_2013 = '2' AND stade_2014 = '3' THEN 0
      WHEN stade_2013 = '3' AND stade_2014 = '4' THEN 0
      WHEN stade_2013 = '4' AND stade_2014 = '5' THEN 0
      WHEN stade_2013 = '5' AND stade_2014 = '6' THEN 0
      WHEN stade_2013 = '1' AND stade_2014 = '3' THEN 0
      WHEN stade_2013 = '2' AND stade_2014 = '4' THEN 0
      WHEN stade_2013 = '3' AND stade_2014 = '5' THEN 0
      WHEN stade_2013 = '4' AND stade_2014 = '6' THEN 0
      WHEN num > 5000 THEN -1
      ELSE NULL
    END);

INSERT INTO stade (num_stade) SELECT DISTINCT stade_2013 FROM individus ORDER BY stade_2013;
--UPDATE stade SET num_stade = -1 WHERE num_stade IS NULL;

--UPDATE individus SET stade_2014 = -1 WHERE stade_2014 IS NULL;

----- 2015 -----
----- Import des donnees a partir du fichier .csv -----
COPY individus_15 FROM '/home/sirakov/Doctorat/Data/data_brut/Data_babassu_2015_prepare.csv' CSV HEADER;

----- Mise a jour des colonnes -----
UPDATE individus_15 SET parcelle_15 = 'DEpasto_12' WHERE parcelle_15 = 'Depasto_12';
UPDATE individus_15 SET parcelle_15 = 'DEpasto_12' WHERE parcelle_15 = 'DEpasto_12_';
UPDATE individus_15 SET parcelle_15 = 'RVpasto_08' WHERE parcelle_15 = 'RVpasto_08';
UPDATE individus_15 SET coupe_15 = 'M' WHERE stade_15 = 'MORTO';
UPDATE individus_15 SET coupe_15 = 'P' WHERE stade_15 = 'NE';
UPDATE individus_15 SET coupe_15 = 'C' WHERE coupe_15 = 'S';
UPDATE individus_15 SET stade_15 = NULL WHERE stade_15 = 'MORTO';

```

```
UPDATE individus_15 SET stade_15 = NULL WHERE stade_15 = 'NE';
ALTER TABLE individus_15 ALTER COLUMN stade_15 TYPE integer USING (stade_15::integer);
```

```
----- Inserer les nouvelles donnees au sein de la table "individus" -----
INSERT INTO individus (SELECT num_13 FROM individus_15 WHERE num_13 > 5687 AND (num_13 < 5802 OR num_13 > 5810));
```

```
----- Charger donnees de la table "individus_15" vers la table "individus" -----
```

```
UPDATE individus i SET coupe_2015 = (SELECT i_15.coupe_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num);
UPDATE individus i SET releve_2015 = (SELECT i_15.releve_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num);
UPDATE individus SET releve_2015 = '2015-07-07' WHERE releve_2015 IS NULL;
UPDATE individus i SET stade_2015 = (SELECT i_15.stade_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num);
UPDATE individus i SET parcelle = (SELECT i_15.parcelle_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) WHERE num > 5687 AND (num < 5802 OR num > 5810);
UPDATE individus i SET transect_code = (SELECT i_15.transect_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) WHERE num > 5687 AND (num < 5802 OR num > 5810);
UPDATE individus i SET utilisation_sol = (SELECT i_15.utilisation_sol_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) WHERE num > 5687 AND (num < 5802 OR num > 5810);
UPDATE individus i SET annee_installation = (SELECT i_15.annee_installation_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) WHERE num > 5687 AND (num < 5802 OR num > 5810);
```

```
----- Mise a jour de la colonne "transect_num" -----
```

```
UPDATE individus
  SET transect =
  (CASE
    WHEN parcelle = 'DEpasto_12' AND transect_code = '1' THEN 1
    WHEN parcelle = 'DEpasto_12' AND transect_code = '2' THEN 2
    WHEN parcelle = 'DEpasto_12' AND transect_code = '3' THEN 3
    WHEN parcelle = 'DEpasto_12' AND transect_code = 'supplem.' THEN 18
    WHEN parcelle = 'CLpasto_10' AND transect_code = '1' THEN 4
    WHEN parcelle = 'CLpasto_10' AND transect_code = '2' THEN 5
    WHEN parcelle = 'CLpasto_10' AND transect_code = 'supplem.' THEN 19
    WHEN parcelle = 'CHpasto_10' AND transect_code = '1' THEN 6
    WHEN parcelle = 'CHpasto_10' AND transect_code = '2' THEN 7
    WHEN parcelle = 'CHpasto_10' AND transect_code = 'supplem.' THEN 20
    WHEN parcelle = 'RVpasto_08' AND transect_code = '1' THEN 8
    WHEN parcelle = 'RVpasto_08' AND transect_code = '2' THEN 9
    WHEN parcelle = 'RVpasto_08' AND transect_code = '3' THEN 10
    WHEN parcelle = 'RVpasto_08' AND transect_code = 'supplem.' THEN 21
    WHEN parcelle = 'AMPasto_06' AND transect_code = '1' THEN 11
    WHEN parcelle = 'AMPasto_06' AND transect_code = '2' THEN 12
    WHEN parcelle = 'AMPasto_06' AND transect_code = 'supplem.' THEN 22
    WHEN parcelle = 'ARpasto_06' AND transect_code = '1' THEN 13
    WHEN parcelle = 'ARpasto_06' AND transect_code = '2' THEN 14
    WHEN parcelle = 'ARpasto_06' AND transect_code = 'supplem.' THEN 23
    WHEN parcelle = 'AMPasto_04' AND transect_code = '1' THEN 15
    WHEN parcelle = 'AMPasto_04' AND transect_code = '2' THEN 16
    WHEN parcelle = 'AMPasto_04' AND transect_code = '3' THEN 17
    WHEN parcelle = 'AMPasto_04' AND transect_code = 'supplem.' THEN 24
    WHEN parcelle = 'RI_bab' AND transect_code = '1' THEN 25
    WHEN parcelle = 'RI_bab' AND transect_code = '2' THEN 26
    WHEN parcelle = 'ZM_bab' AND transect_code = '1' THEN 27
    WHEN parcelle = 'ZM_bab' AND transect_code = '2' THEN 28
    WHEN parcelle = 'ZM_bab' AND transect_code = '2' THEN 28
    WHEN parcelle = 'DE_flo' AND transect_code = '1' THEN 29
    WHEN parcelle = 'DE_flo' AND transect_code = '2' THEN 30
    WHEN parcelle = 'DE_flo' AND transect_code = 'supplem.' THEN 31
    WHEN parcelle = 'ME_flo' AND transect_code = '1' THEN 32
    WHEN parcelle = 'ME_flo' AND transect_code = '2' THEN 33
    WHEN parcelle = 'ME_flo' AND transect_code = 'supplem.' THEN 34
    ELSE 1000
  )
  END) WHERE num > 5687 AND (num < 5802 OR num > 5810);
```

```
----- Mise a jour de la colonne "age_2015" -----
UPDATE individus
```

```

SET age_2015 =
(CASE
WHEN stade_2013 = '1' AND stade_2014 = '1' AND stade_2015 = '1' THEN 2
WHEN stade_2013 = '2' AND stade_2014 = '2' AND stade_2015 = '2' THEN 2
WHEN stade_2013 = '3' AND stade_2014 = '3' AND stade_2015 = '3' THEN 2
WHEN stade_2013 = '4' AND stade_2014 = '4' AND stade_2015 = '4' THEN 2
WHEN stade_2013 = '5' AND stade_2014 = '5' AND stade_2015 = '5' THEN 2
WHEN stade_2013 = '6' AND stade_2014 = '6' AND stade_2015 = '6' THEN 2
WHEN stade_2013 IS NULL AND stade_2014 = '1' AND stade_2015 = '1' THEN 1
WHEN (stade_2013 = '1' OR stade_2013 IS NULL) AND stade_2014 = '2' AND stade_2015 = '2' THEN 1
WHEN (stade_2013 = '2' OR stade_2013 IS NULL) AND stade_2014 = '3' AND stade_2015 = '3' THEN 1
WHEN (stade_2013 = '3' OR stade_2013 IS NULL) AND stade_2014 = '4' AND stade_2015 = '4' THEN 1
WHEN (stade_2013 = '4' OR stade_2013 IS NULL) AND stade_2014 = '5' AND stade_2015 = '5' THEN 1
WHEN (stade_2013 = '5' OR stade_2013 IS NULL) AND stade_2014 = '6' AND stade_2015 = '6' THEN 1
WHEN stade_2014 = '1' AND stade_2015 = '2' THEN 0
WHEN stade_2014 = '2' AND stade_2015 = '3' THEN 0
WHEN stade_2014 = '3' AND stade_2015 = '4' THEN 0
WHEN stade_2014 = '4' AND stade_2015 = '5' THEN 0
WHEN stade_2014 = '5' AND stade_2015 = '6' THEN 0
WHEN stade_2014 = '1' AND stade_2015 = '3' THEN 0
WHEN stade_2014 = '2' AND stade_2015 = '4' THEN 0
WHEN stade_2014 = '3' AND stade_2015 = '5' THEN 0
WHEN stade_2014 = '4' AND stade_2015 = '6' THEN 0
WHEN num > 5687 AND (num < 5802 OR num > 5810) THEN -1
ELSE NULL
END);

```

B.6 Script SQL « 6_verification.sql »

```

----- "individus" -----
SELECT DISTINCT parcelle FROM individus ORDER BY parcelle;
SELECT DISTINCT transect_code FROM individus ORDER BY transect_code;
SELECT DISTINCT stade_2013 FROM individus ORDER BY stade_2013;
SELECT DISTINCT utilisation_sol FROM individus ORDER BY utilisation_sol;
SELECT DISTINCT annee_installation FROM individus ORDER BY annee_installation;
SELECT DISTINCT coupe_2013 FROM individus ORDER BY coupe_2013;

```

```

----- "individus_14" -----
SELECT DISTINCT transect_14 FROM individus_14 ORDER BY transect_14;
SELECT DISTINCT stade_14 FROM individus_14 ORDER BY stade_14;
SELECT DISTINCT utilisation_sol_14 FROM individus_14 ORDER BY utilisation_sol_14;
SELECT DISTINCT annee_installation_14 FROM individus_14 ORDER BY annee_installation_14;
SELECT DISTINCT coupe_14 FROM individus_14 ORDER BY coupe_14;

```

```

----- "individus_15" -----
SELECT DISTINCT transect_15 FROM individus_15 ORDER BY transect_15;
SELECT DISTINCT stade_15 FROM individus_15 ORDER BY stade_15;
SELECT DISTINCT utilisation_sol_15 FROM individus_15 ORDER BY utilisation_sol_15;
SELECT DISTINCT annee_installation_15 FROM individus_15 ORDER BY annee_installation_15;
SELECT DISTINCT coupe_15 FROM individus_15 ORDER BY coupe_15;

```

```

----- Coherence des donnees entre 2013 et 2014 -----
SELECT num FROM individus i WHERE parcelle <> (SELECT i_14.parcelle_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) ORDER BY num;
SELECT num FROM individus i WHERE transect_code <> (SELECT i_14.transect_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) ORDER BY num;
SELECT num FROM individus i WHERE annee_installation <> (SELECT i_14.annee_installation_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) ORDER BY num;
SELECT * FROM individus i WHERE stade_2013 > (SELECT i_14.stade_14 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) ORDER BY num;

```

```
----- Coherence des donnees entre 2013 et 2015 -----  
SELECT num FROM individus i WHERE parcelle <> (SELECT i_15.parcelle_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) ORDER BY num;  
SELECT num FROM individus i WHERE transect_code <> (SELECT i_15.transect_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) ORDER BY num;  
SELECT num FROM individus i WHERE annee_installation <> (SELECT i_15.annee_installation_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) ORDER BY num;  
SELECT * FROM individus i WHERE stade_2013 > (SELECT i_15.stade_15 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) ORDER BY num;
```

```
----- Correspondance entre les PK -----  
SELECT * FROM individus_14 i_14 WHERE num_13 NOT IN (SELECT i.num FROM individus i WHERE i.num = i_14.num_13) ORDER BY num_13;  
SELECT * FROM individus i WHERE num NOT IN (SELECT i_14.num_13 FROM individus_14 i_14 WHERE i_14.num_13 = i.num) ORDER BY num;  
  
SELECT * FROM individus_15 i_15 WHERE num_13 NOT IN (SELECT i.num FROM individus i WHERE i.num = i_15.num_13) ORDER BY num_13;  
SELECT * FROM individus i WHERE num NOT IN (SELECT i_15.num_13 FROM individus_15 i_15 WHERE i_15.num_13 = i.num) ORDER BY num;  
  
DROP TABLE IF EXISTS individus_14;  
DROP TABLE IF EXISTS individus_15;
```

ANNEXE C

Convergence des chaînes de Markov

FIGURE C.1 – Convergence des chaînes de Markov - stade 1

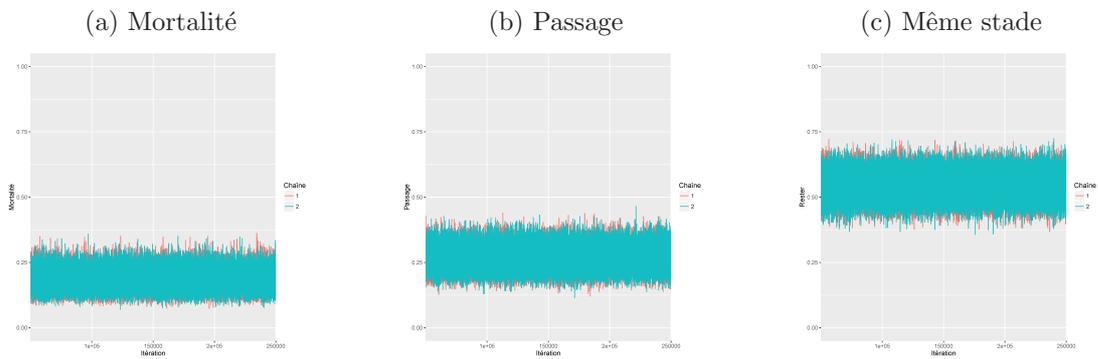


FIGURE C.2 – Convergence des chaînes de Markov - stade 2

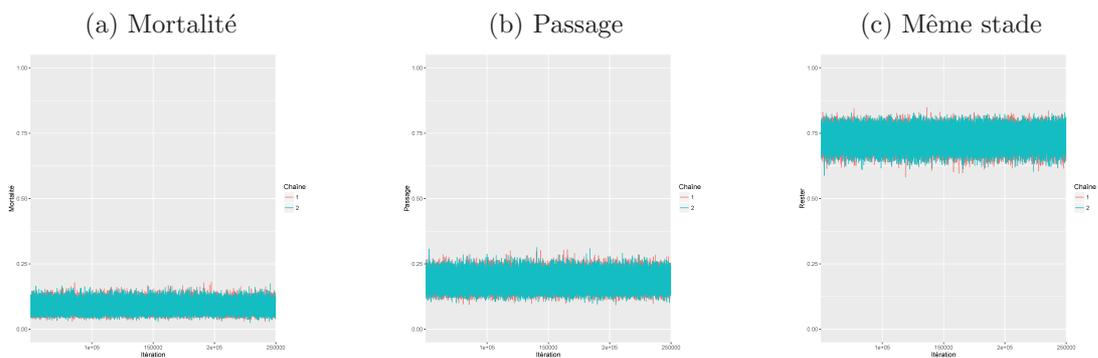


FIGURE C.3 – Convergence des chaînes de Markov - stade 3

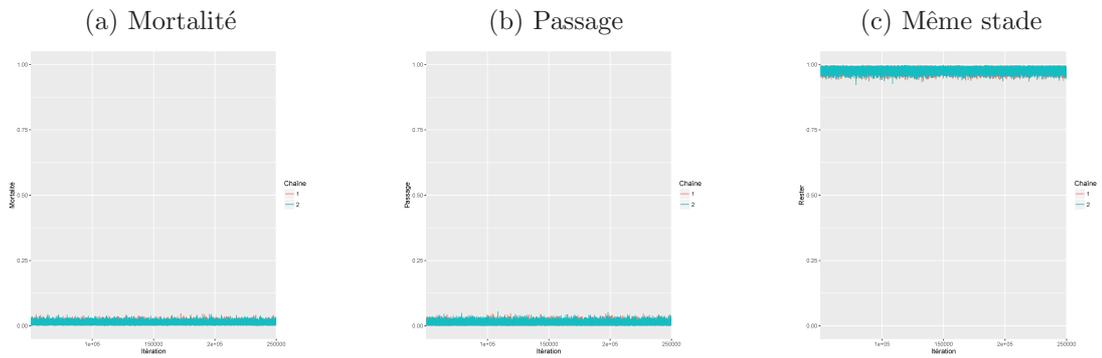


FIGURE C.4 – Convergence des chaînes de Markov - stade 4

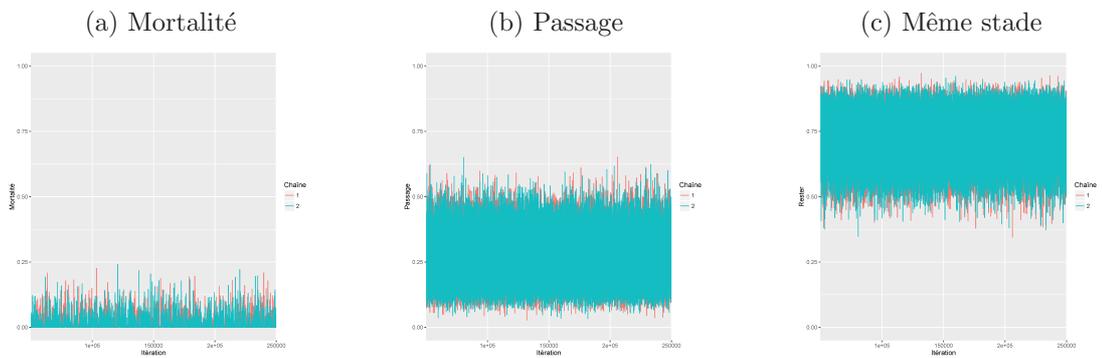


FIGURE C.5 – Convergence des chaînes de Markov - stade 5

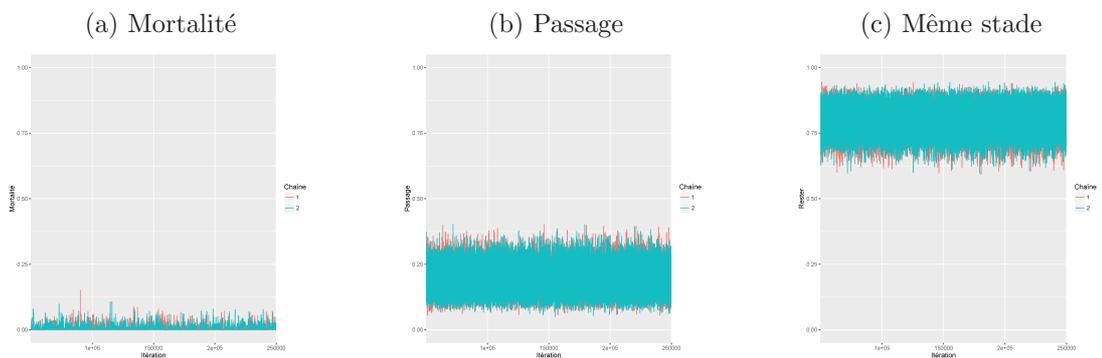


FIGURE C.6 – Convergence des chaînes de Markov - stade 6

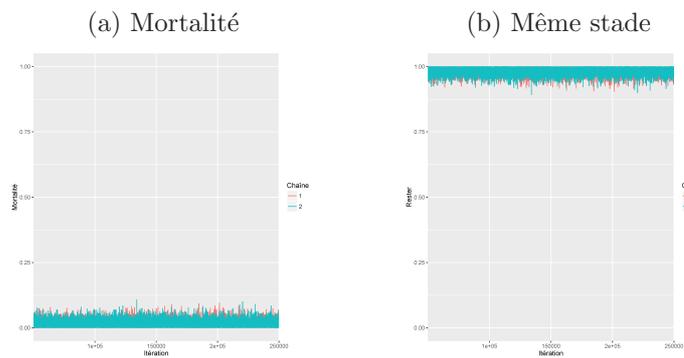


FIGURE C.7 – Convergence des chaînes de Markov - recrutement

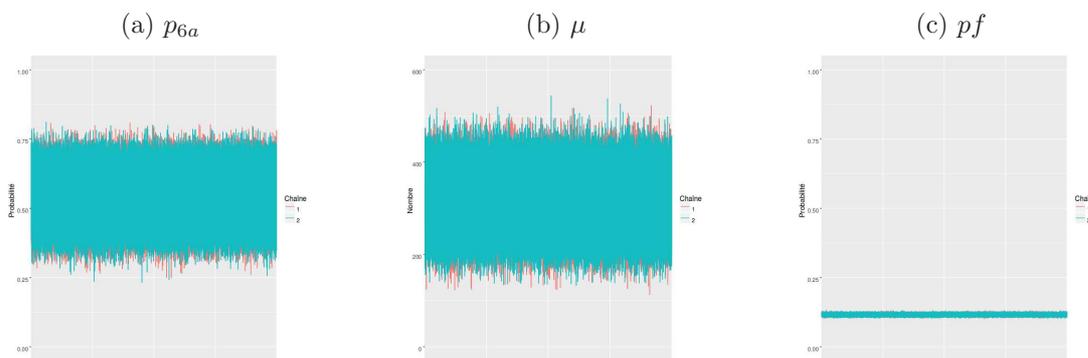


FIGURE C.8 – Test de Gelman-Rubin - stade 1

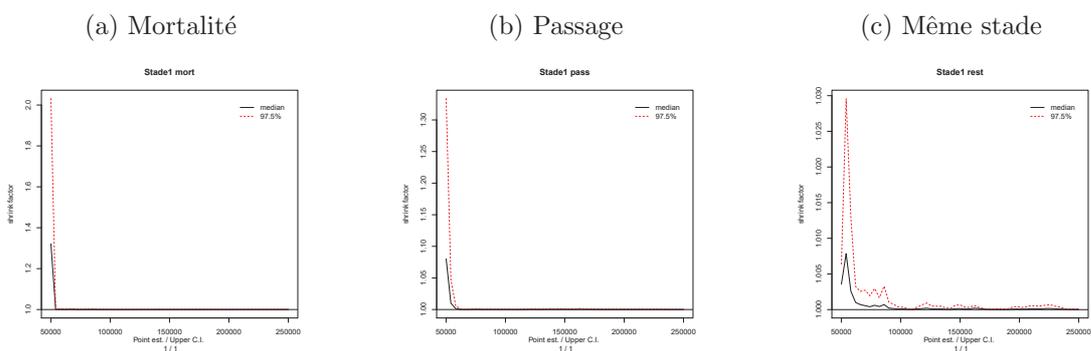
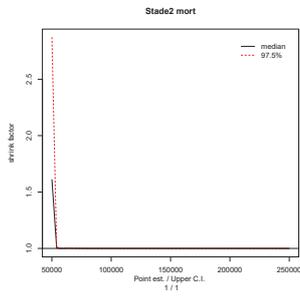
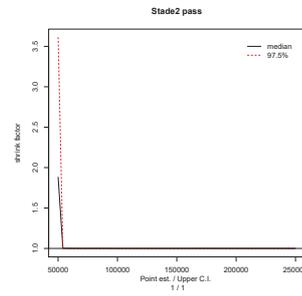


FIGURE C.9 – Test de Gelman-Rubin - stade 2

(a) Mortalité



(b) Passage



(c) Même stade

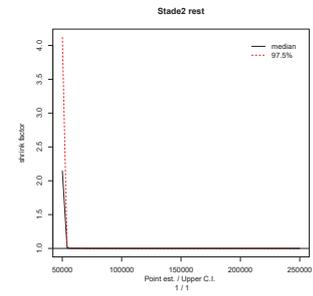
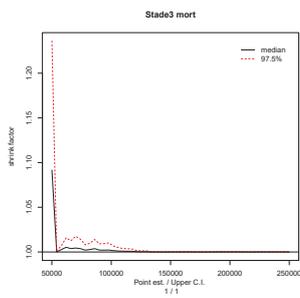
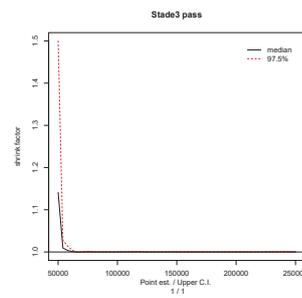


FIGURE C.10 – Test de Gelman-Rubin - stade 3

(a) Mortalité



(b) Passage



(c) Même stade

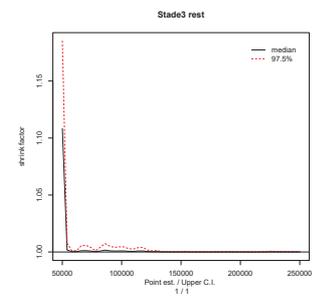
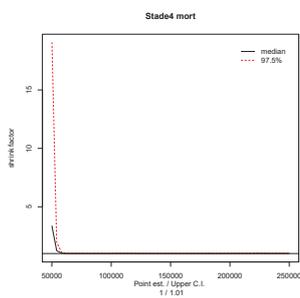
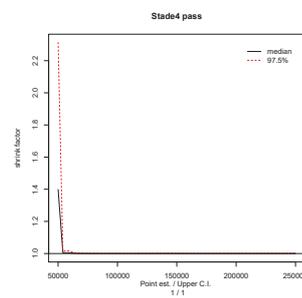


FIGURE C.11 – Test de Gelman-Rubin - stade 4

(a) Mortalité



(b) Passage



(c) Même stade

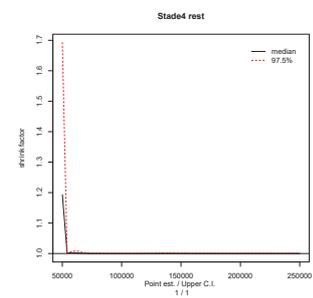
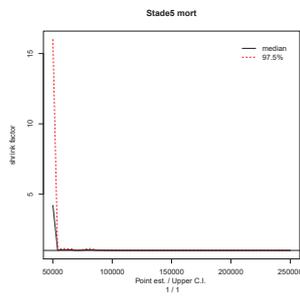
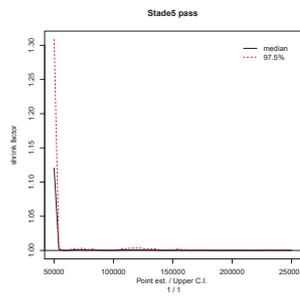


FIGURE C.12 – Test de Gelman-Rubin - stade 5

(a) Mortalité



(b) Passage



(c) Même stade

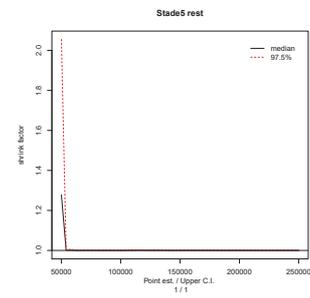
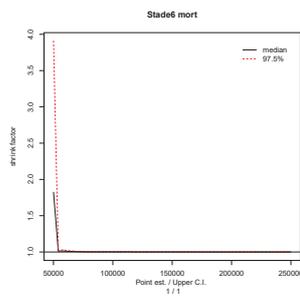


FIGURE C.13 – Test de Gelman-Rubin - stade 6

(a) Mortalité



(b) Même stade

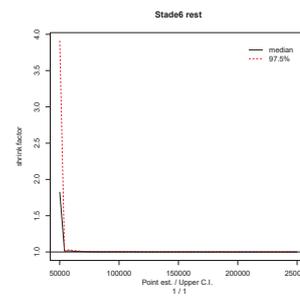
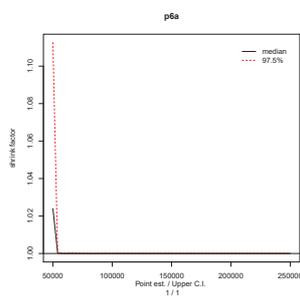
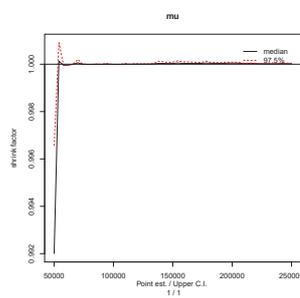
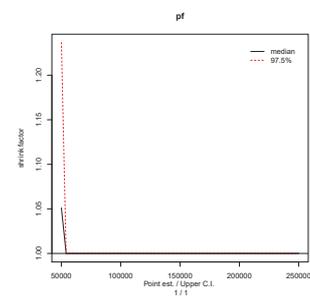


FIGURE C.14 – Test de Gelman-Rubin - recrutement

(a) p_{6a} (b) μ (c) pf 

Validation croisée - tableaux

TABLEAU D.1 – Liste des transects utilisés pour le stade 1

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
pâturage	1	DEpasto_12	2012	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	2	CLpasto_10	2010	1 (estimation)
				2 (estimation)
				suppl
	3	CHpasto_10	2010	1 (estimation)
				2 (estimation)
				suppl
	4	RVpasto_08	2008	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	5	AMpasto_06	2006	1 (estimation)
				2 (estimation)
				suppl
	6	ARpasto_06	2006	1 (pas servi)
				2 (validation)
				suppl
	7	AMpasto_04	2004	1 (validation)
2 (validation)				
3 (validation)				
suppl				

TABLEAU D.2 – Liste des transects utilisés pour le stade 2

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
pâturage	1	DEpasto_12	2012	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	2	CLpasto_10	2010	1 (estimation)
				2 (estimation)
				suppl
	3	CHpasto_10	2010	1 (estimation)
				2 (estimation)
				suppl
	4	RVpasto_08	2008	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	5	AMpasto_06	2006	1 (estimation)
				2 (estimation)
				suppl
	6	ARpasto_06	2006	1 (validation)
				2 (validation)
				suppl
	7	AMpasto_04	2004	1 (validation)
2 (validation)				
3 (validation)				
suppl				

TABLEAU D.3 – Liste des transects utilisés pour le stade 3

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
pâturage	1	DEpasto_12	2012	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	2	CLpasto_10	2010	1 (estimation)
				2 (estimation)
				suppl
	3	CHpasto_10	2010	1 (estimation)
				2 (pas servi)
				suppl
	4	RVpasto_08	2008	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	5	AMpasto_06	2006	1 (estimation)
				2 (estimation)
				suppl
	6	ARpasto_06	2006	1 (estimation)
				2 (validation)
				suppl
	7	AMpasto_04	2004	1 (validation)
2 (validation)				
3 (validation)				
suppl				

TABLEAU D.4 – Liste des transects utilisés pour le stade 4

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
pâturage	1	DEpasto_12	2012	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl
	2	CLpasto_10	2010	1 (estimation)
				2 (validation)
				suppl
	3	CHpasto_10	2010	1 (pas servi)
				2 (pas servi)
				suppl
	4	RVpasto_08	2008	1 (validation)
				2 (pas servi)
				3 (pas servi)
				suppl
	5	AMpasto_06	2006	1 (pas servi)
				2 (pas servi)
				suppl
	6	ARpasto_06	2006	1 (pas servi)
				2 (pas servi)
				suppl
	7	AMpasto_04	2004	1 (pas servi)
2 (pas servi)				
3 (validation)				
suppl				

TABLEAU D.5 – Liste des transects utilisés pour le stade 5

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
pâturage	1	DEpasto_12	2012	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl (estimation)
	2	CLpasto_10	2010	1 (estimation)
				2 (estimation)
				suppl (estimation)
	3	CHpasto_10	2010	1 (pas servi)
				2 (pas servi)
				suppl (estimation)
	4	RVpasto_08	2008	1 (estimation)
				2 (pas servi)
				3 (pas servi)
				suppl (estimation)
	5	AMpasto_06	2006	1 (pas servi)
				2 (pas servi)
				suppl (validation)
	6	ARpasto_06	2006	1 (pas servi)
				2 (estimation)
				suppl (validation)
	7	AMpasto_04	2004	1 (pas servi)
				2 (pas servi)
				3 (pas servi)
				suppl (validation)

TABLEAU D.6 – Liste des transects utilisés pour le stade 6

Occupation du sol	Numéro	Code parcelle	Année supposée d'ouverture	Transect
pâturage	1	DEpasto_12	2012	1 (estimation)
				2 (pas servi)
				3 (estimation)
				suppl (estimation)
	2	CLpasto_10	2010	1 (pas servi)
				2 (pas servi)
				suppl (estimation)
	3	CHpasto_10	2010	1 (pas servi)
				2 (pas servi)
				suppl (estimation)
	4	RVpasto_08	2008	1 (estimation)
				2 (pas servi)
				3 (pas servi)
				suppl (validation)
	5	AMpasto_06	2006	1 (estimation)
				2 (pas servi)
				suppl (validation)
	6	ARpasto_06	2006	1 (estimation)
				2 (estimation)
				suppl (validation)
	7	AMpasto_04	2004	1 (estimation)
				2 (estimation)
				3 (estimation)
				suppl (validation)

Bibliographie

- AKERLOF, G. et SHILLER, R. (2016). *Marchés de dupes*. économie. Odile Jacob, Paris.
- ALMEIDA, A. W. B., SHIRAIISHI NETO, J. et MARTINS, C. C. (2005). *Guerra ecológica Nos Babaçuais (o processo de devastação dos palmeiras, a elevação do preço de commodities e o aquecimento do mercado de terras na Amazônia)*. MIQCB/-Balaios typographia, São Luís-MA.
- ANDERSEN, K. (2015). Cowspiracy : The sustainability secret. <https://culturedocumentee.wordpress.com/2015/01/31/cowspiracy-the-sustainability-secret/>. Consulté le 26.05.2016.
- ANDERSON, A. B. (1983). *The biology of Orbignya Martiana (Palmae), a tropical dry forest dominant in Brazil*. Ph. D., University of Florida.
- ANDERSON, A. B. et BALICK, M. J. (1988). Taxonomy of the babassu complex (Orbignya spp. palmae). *Systematic Botany*, 13(1):32–50.
- ANDERSON, A. B., MAY, P. H. et BALICK, M. J. (1991). *The Subsidy from Nature : Palm Forests, Peasantry, and Development on an Amazon Frontier*. Columbia University Press, New York.
- ANONYME (1981). Mapeamento das ocorrências e prospecção do potencial atual do babaçu no Estado do Maranhão. Rapport technique, Companhia de Pesquisa e Aproveitamento de Recursos Naturais (COPENAT) e Fundação Instituto Estadual do Babaçu (INEB), São Luís.
- ARAÚJO, F. R. et LOPES, M. A. (2012). Diversity of use and local knowledge of palms (Arecaceae) in eastern Amazonia. *Biodiversity and Conservation*, 21(2):487–501.
- Arnauld de SARTRE, X. (2006). *Fronts pionniers d'Amazonie. Les dynamiques paysannes au Brésil*. Espaces et milieux. CNRS Éditions, Paris.
- BADIOU, A. (2007). *Le concept de modèle : introduction à une épistémologie matérialiste des mathématiques*. Ouvertures. Fayard, Paris.
- BAROT, S., GIGNOUX, J., VUATTOUX, R. et LEGENDRE, S. (2000). Demography of a Savanna Palm Tree in Ivory Coast (Lamto) : Population Persistence and Life-History. *Journal of Tropical Ecology*, 16(5):637–655.
- BAROT, S., MITJA, D., MIRANDA, I., MEIJA, G. et GRIMALDI, M. (2005). Reproductive plasticity in an Amazonian palm. *Evolutionary Ecology Research*, 7(7):1051–1065.

- BARUQUE FILHO, E. A., BARUQUE, M. d. G. A., FREIRE, D. M. et SANT'ANNA, G. L. (1998). Ethanol from babassu coconut starch. *Applied biochemistry and biotechnology*, 70(1):877–886.
- BERNAL, R. (1998). Demography of the Vegetable Ivory Palm *Phytelephas seemanii* in Colombia, and the Impact of Seed Harvesting. *Journal of Applied Ecology*, 35(1):64–74.
- BIERZYCHUDEK, P. (1999). Looking Backwards : Assessing the Projections of a Transition Matrix Model. *Ecological Applications*, 9(4):1278–1287.
- BRUNET, R. (1980). La composition des modèles dans l'analyse spatiale. *Espace géographique*, 9(4):253–265.
- BULLOCK, S. H. (1980). Demography of an Undergrowth Palm in Littoral Cameroon. *Biotropica*, 12(4):247–255.
- CARON, P. (1998). *Espaces, élevage et dynamique du changement : analyse, niveaux d'organisation et action*. Thèse de doctorat, Paris 10, France.
- CASWELL, H. (2001). *Matrix population models : construction, analysis, and interpretation*. Sinauer, Sunderland (Massachusetts).
- CHARLET, J., KASSEL, G., ZACKLAD, M. et BOURIGAULT, D. (2000). *Ingénierie des connaissances : évolutions récentes et nouveaux défis*. Eyrolles, Paris, France. Google-Books-ID : GJfSAAAACAAJ.
- CHOI, B. C. (2005). Understanding the basic principles of knowledge translation. *Journal of Epidemiology and Community Health*, 59(2):93–93.
- CLAVAL, P. (2001). *Épistémologie de la géographie*. Université. Nathan, Paris.
- CODD, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.
- CONGDON, P. (2001). *Bayesian statistical modelling*. Wiley, Chichester.
- CRAWLEY, M. J. (2005). *Plant ecology*. Blackwell Science, Oxford, 2 édition.
- DELAÎTRE, E., LAQUES, A.-E., MITJA, D., CORDEIRO THALES, M., de SOUZA MIRANDA, I., RODRIGUES COELHO, R. d. F. et NEIVA SAMPAIO, S. (2012). Des images satellites pour suivre l'influence des politiques publiques sur la biodiversité d'un front pionnier d'Amazonie brésilienne à deux échelles d'analyse : du lot d'exploitations à la communauté agricole. In *Cultural diversity and biological diversity for sustainable development : exploring the past to build up the future*, pages 10–11, Montpellier. ISE.

- DENEUX, J.-F. (2006). *Histoire de la pensée géographique*. Atouts géographique. Belin, Paris.
- DEPECKER, L. et ROCHE, C. (2007). Entre idée et concept : vers l'ontologie. *Langages*, 4(168):106–114.
- DROULERS, M. (2004). *L'Amazonie, vers un développement durable*. U. Armand Colin, Paris.
- DROULERS, M. et BROGGIO, C. (2013). *Le Brésil*. Numéro 628 de Que sais-je ? Presses Universitaires de France, Paris.
- EMMERICH, F. G. et LUENGO, C. A. (1994). Reduction of emissions from blast furnaces by using blends of coke and babassu charcoal. *Fuel*, 73(7):1235–1236.
- EMMERICH, F. G. et LUENGO, C. A. (1996). Babassu charcoal : A sulfurless renewable thermo-reducing feedstock for steelmaking. *Biomass and Bioenergy*, 10(1):41–44.
- ENRIGHT, N. et WATSON, A. D. (1992). Population dynamics of the nikau palm, *Rhopalostylis sapida* (Wendl. et Drude), in a temperate forest remnant near Auckland, New Zealand. *New Zealand Journal of Botany*, 30(1):29–43.
- FERREIRA, M. E. M. (1999). *Modelos Log-Normal e markoviano para estudo da evolução de abundância em uma floresta de babaçu*. Thèse de doctorat, Universidade Federal de Santa Catarina, Florianópolis.
- FIEBERG, J. et ELLNER, S. P. (2001). Stochastic matrix models for conservation and management : a comparative review of methods. *Ecology Letters*, 4(3):244–266.
- FUR, J. L., CURY, P., LALOË, F., DURAND, M.-H. et CHABOUD, C. (1999). Co-viabilité des systèmes halieutiques. *Natures Sciences Sociétés*, 07(2):19–32.
- GELMAN, A., CARLIN, J. B., STERN, H. S. et RUBIN, D. B. (2003). *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC, Boca Raton, Fla., 2 édition.
- GELMAN, A., MENG, X.-L. et STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 6(4):733–760.
- GELMAN, A. et RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.
- GEMAN, S. et GEMAN, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.
- GLASSMAN, S. F. (1977). Preliminary taxonomic studies in the palm genus *Orbignya* Mart. *Phytologia*, 36(2):89–115.

- GLASSMAN, S. F. (1978). Corrections and changes in recent palm articles published in *Phytologia*. *Phytologia*, 40(4):313–315.
- GONZÁLEZ-PÉREZ, S. E., COELHO-FERREIRA, M., de ROBERT, P. et GARCÉS, C. L. L. (2012). Knowledge and use of babassu (*Attalea speciosa* Mart. and *Attalea eichleri* (Drude) A.J. Hend.) among Mebengokre-kayapó from Las Casas Indigenous Land, Pará state, Brazil. *Acta Botanica Brasilica*, 26(2):295–308.
- GRIMSHAW, J. M., ECCLES, M. P., LAVIS, J. N., HILL, S. J. et SQUIRES, J. E. (2012). Knowledge translation of research findings. *Implementation science*, 7(1):1.
- GRUBER, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5):907–928.
- GUARÍN, J. R. et del VALLE, J. I. (2014). Modeling the stipe growth of the *Oenocarpus bataua* palm in the Central Cordillera of the Andes, Colombia. *Forest Ecology and Management*, 314:141–149.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- HENDERSON, A. (1995). *The palms of the Amazon*. Oxford University Press, New York.
- HOLM, J. A., MILLER, C. J. et CROPPER, W. P. (2008). Population Dynamics of the Dioecious Amazonian Palm *Mauritia flexuosa* : Simulation Analysis of Sustainable Harvesting. *Biotropica*, 40(5):550–558.
- IBGE (2004). Mapa de Biomas e de Vegetação. <http://ibge.gov.br/home/presidencia/noticias/21052004biomashtml.shtm>. Consulté le 24.05.2016.
- IBGE (2013). Geoserver. <http://www.geoservicos.ibge.gov.br/geoserver/web/>. Consulté le 16.04.2014.
- JEFFREYS, H. (1961). *Theory of probability*. The international series of monographs on physics. Clarendon Press, Oxford, 3 édition.
- KAYE, T. N. et PYKE, D. A. (2003). The effect of stochastic technique on estimates of population viability from transition matrix models. *Ecology*, 84(6):1464–1476.
- LAQUES, A.-E., MITJA, D., DELAÎTRE, E., THALES, M. C., MIRANDA, I. d. S., COELHO, R. d. F. R. et SAMPAIO, S. M. N. (2012). Spatialisation de la biodiversité en Amazonie brésilienne pour appréhender l'influence de la colonisation des terres et des politiques publiques. *VertigO - la revue électronique en sciences de l'environnement*, (Hors-série 14).

- LARDON, S., VOLLET, D., RIEUTORT, L., DEVÈS, C. et MAMDY, J.-F. (2009). Développement, attractivité et ingénierie des territoires : Des enjeux de recherche pour l'action et la formation. *Revue d'Auvergne*, 123(590-591):13–18.
- LECOUTRE, B. (2006). Training students and researchers in Bayesian methods. *Journal of Data Science*, 4:207–232.
- LEFKOVITCH, L. P. (1965). The Study of Population Growth in Organisms Grouped by Stages. *Biometrics*, 21(1):1–18.
- LESLIE, P. H. (1945). On the Use of Matrices in Certain Population Mathematics. *Biometrika*, 33(3):183–212.
- LETHIELLEUX, M. (1998). *Probabilités*. Dunod, Paris.
- LEWIS, S. L. et MASLIN, M. A. (2015). Defining the Anthropocene. *Nature*, 519(7542):171–180.
- LORENZI, H., Moreira de SOUZA, H., Coelho de CERQUEIRA, L. S., FERREIRA, E. et MADEIROS COSTA, J. T. d. (2004). *Palmeiras brasileiras e exóticas cultivadas*. Instituto Planetarum de Estudos Da Florida, Nova Odessa.
- LUNN, D., JACKSON, C., BEST, N., THOMAS, A. et SPIEGELHALTER, D. (2013). *The BUGS book : A practical introduction to Bayesian analysis*. CRC press.
- LUNN, D. J., THOMAS, A., BEST, N. et SPIEGELHALTER, D. (2000). WinBUGS - A Bayesian modelling framework : Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337.
- LÉMERY, B. (2003). Les agriculteurs dans la fabrique d'une nouvelle agriculture. *Sociologie du Travail*, 45(1):9–25.
- MACEDO, D. J. (2015). Palmier babaçu et société en Amazonie : enquêtes et cartographie par télédétection. Rapport technique, Université de Montpellier, Montpellier.
- MADIN, J., BOWERS, S., SCHILDHAUER, M., KRIVOV, S., PENNINGTON, D. et VILLA, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological informatics*, 2(3):279–296.
- MAY, P., ANDERSON, A., BALICK, M. et FRAZAO, J. (1985). Subsistence benefits from the babassu palm (Orbignya-martiana). *Economic Botany*, 39(2):113–129.
- MAZZEGA, P., SIBERTIN-BLANC, C., THÉRON, O., AMBLARD, F., ARCANGELI, J.-P., BALESTRAT, M., CHARRON-MOIREZ, M.-H., CONDAMINES, A., FÈVRE-PERNET, C., GAUDOU, B. et OTHERS (2013). Impact Assessment Modeling of Low-Water Management Policy. In *4th Workshop of Applied Computing for the*

- Management of the Environment and Natural Resources (WCAMA 2013)*, pages 1014–1024, Maceió, Alagoas, Brazil.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. et TELLER, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- MICHEL-GUILLOU, E. (2006). Représentations sociales et pratiques sociales : l'exemple de l'engagement pro-environnemental en agriculture. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 56(3):157–165.
- MINSKY, M. L. (1965). Matter, mind and models. In KALENISH, W. A., éditeur : *Proc. IFIP Congress Information Processing*, pages 45–49, New York City. Spartan Books, Washington (1965).
- MITJA, D. et FERRAZ, I. D. K. (2001). Establishment of Babassu in Pastures in Pará, Brazil. *PALMS*, 45(3):138–147.
- MITJA, D., SIRAKOV, N., SANTOS, A. M. d., GONZÁLEZ-PÉREZ, S. E., MACEDO, J. D., DELAÎTRE, E., DEMAGISTRI, L., LOISEL, P., MIRANDA, I. d. S., REY-VALETTE, H., ROCHA, M. R. T. d., FONTEZ, B. et LIBOUREL, T. (sous presse). Viability of the babassu palm eco-socio-system in Brazil : the challenges of a co-viability. In *Co-viability of Social and Ecological Systems : Reconnect Man to the Biosphere in an Era of Global Change*. Springer-Verlag.
- NASUTI, S.-C. (2010). *Une situation de ' post-front pionnier ' : le cas de la région rurale de Ciriaco [Maranhão - Brésil]*. Thèse de Doctorat, Université de la Sorbonne nouvelle - Paris III, Paris.
- NOUVEL, P. et LECOURT, D. (2002). *Enquête sur le concept de modèle*. Science, histoire et société. Presses universitaires de France, Paris.
- OLMSTED, I. et ALVAREZ-BUYLLA, E. R. (1995). Sustainable Harvesting of Tropical Trees : Demography and Matrix Models of Two Palm Species in Mexico. *Ecological Applications*, 5(2):484–500.
- PAGES, L., LESCOURRET, F., BIDEL, L. et PELLOUX, G. (2000). Utilisation de méthodes orientées objet pour accroître nos capacités à élaborer et utiliser des modèles de fonctionnement de plantes et de peuplements. In *Fonctionnement des peuplements végétaux sous contraintes environnementales, 20-21 janvier 1998*, Les Colloques de l'INRA (93), pages 217–231. P. Maillard, R. Bonhomme, Paris, France, inra édition.
- PARENT, E. et RIVOT, E. (2012). *Introduction to Hierarchical Bayesian Modeling for Ecological Data*. Chapman and Hall/CRC.

- PICOUET, P. et RENARD, J.-P. (2007). *Les frontières mondiales, origines et dynamiques*. Une géographie. Ed. du Temps, Nantes.
- PINARD, M. (1993). Impacts of Stem Harvesting on Populations of *Iriartea deltoidea* (Palmae) in an Extractive Reserve in Acre, Brazil. *Biotropica*, 25(1):2–14.
- PINERO, D., MARTINEZ-RAMOS, M. et SARUKHAN, J. (1984). A Population Model of *Astrocaryum Mexicanum* and a Sensitivity Analysis of its Finite Rate of Increase. *Journal of Ecology*, 72(3):977–991.
- PINTON, F. et EMPERAIRE, L. (1992). L'extractivisme en Amazonie brésilienne : un système en crise d'identité. *Cahiers des Sciences Humaines*, 28(4):685–703.
- POINTET, A. (2007). *Rencontre de la science de l'information géographique et de l'anthropologie culturelle : modélisation spatiale et représentation de phénomènes culturels*. Thèse de Doctorat, École polytechnique fédérale de Lausanne ; Faculté de l'environnement naturel, architectural et construit, Lausanne, Suisse.
- POSEY, D. A. (1987). Manejo da floresta secundária, capoeiras, campos e cerrados (Kayapó). In *Suma Etnológica Brasileira*, volume 1, pages 173–185. 2 édition.
- PROTÁSIO, T. d. P., TRUGILHO, P. F., MIRMEHDI, S. et SILVA, M. G. d. (2014). Quality and energetic evaluation of the charcoal made of babassu nut residues used in the steel industry. *Ciência e Agrotecnologia*, 38(5):435–444.
- PY, B. (1996). *Statistique descriptive : nouvelle méthode pour bien comprendre et réussir*. Economica, Paris, 4 édition.
- RECLUS, E. (1990). *L'homme et la terre : Histoire contemporaine*. Fayard, Paris.
- REY-VALETTE, H. (2010). L'intégration du développement durable dans l'ingénierie territoriale : enjeux et modalités. *ResearchGate*, pages 193–210.
- ROBERT, C. P. (1996). *Méthodes de Monte Carlo par chaînes de Markov*. Economica, Paris. OCLC : 36067374.
- ROUANET, H. (2008). Statistique et Probabilités. <http://www.math-info.univ-paris5.fr/~lerb/rouanet>. Consulté le 02.06.2016.
- ROUANET, H., BERNARD, J.-M., BERT, M.-C., LECOUTRE, B., LECOUTRE, M.-P. et LE ROUX, B. (1998). *New ways in statistical methodology*, volume 618 de *European University Studies*. Peter Lang, Bern.
- RUMBAUGH, J., JACOBSON, I. et BOOCH, G. (2004). *The Unified Modeling Language Reference Manual*. Addison Wesley, Boston, 2 édition.

- SAMPAIO, S. M. N. (2008). *Dinâmica e complexidade da paisagem do Projeto de Assentamento Benfica, Sudeste Paraense*. Ph. D., Universidade Rural da Amazônia, Belém, PA.
- SAMPAIO, S. M. N., WATRIN, O. D. S. et VENTURIERI, A. (2000). *Dinâmica da cobertura vegetal e do uso da terra do "polígono dos castanhais" no sudeste paraense*. Embrapa Amazônia Oriental. Documentos, 31, Belém, PA.
- SANTOS, A. M. d. et MITJA, D. (2011). Wooded Cattle Pasture in the Benfica Seetling Project in Itupiranga, Para, Brazil. *Revista Arvore*, 35(4):919–930.
- SIBERTIN-BLANC, C., THEROND, O., MONTEIL, C. et MAZEGA, P. (2011). Formal modeling of social-ecological systems. *In European Social Simulation Association*, 19-23 september, Montpellier.
- SILVERTOWN, J. W. et LOVETT DOUST, J. (1993). *Introduction to plant population biology*. Blackwell Scientific Publications, Oxford, 3 édition.
- SIMON, H., FARGETTE, M., LOIREAU, M. et LIBOUREL, T. (2014). Du spécifique au générique : ontologie des systèmes territoriaux. *In PHAN, D., éditeur : Ontologies et modélisation par SMA en SHS*, Recherches, Technologie, Applications : Informatique et Systèmes d'Information, pages 281–287. Hermès, Paris.
- SIRAKOV, N. (2012). Gestion durable du palmier babaçu (*Attalea speciosa* Mart. ex Spreng.) : modélisation mathématique et spatialisation de la disponibilité théorique d'une espèce native, Amazonie, Brésil. Rapport technique, Université Montpellier 2, Montpellier.
- SKALSKI, J. R., RYDING, K. E. et MILLSPAUGH, J. (2005). *Wildlife Demography : Analysis of Sex, Age, and Count Data*. Academic Press, Amsterdam ; Boston, 1 édition.
- SMITH, N. (1974). Agouti and Babassu. *Oryx*, 12(05):581–582.
- SMITH, N. (2015). *Palms and People in the Amazon*. Geobotany Studies. Springer, New York.
- SOLAIMAN, B. (2006). *Processus stochastiques pour l'ingénieur*. PPUR, Lausanne.
- STEINBAKK, G. H. et STORVIK, G. O. (2009). Posterior Predictive p-values in Bayesian Hierarchical Models. *Scandinavian Journal of Statistics*, 36(2):320–336.
- SURESH, P., JOGLEKAR, G., HSU, S., AKKISSETTY, P., HAILEMARIAM, L., JAIN, A., REKLAITIS, G. et VENKATASUBRAMANIAN, V. (2008). Onto MODEL : Ontological mathematical modeling knowledge management. *In 18th European Symposium on Computer Aided Process Engineering*, volume 25 de *Computer Aided Chemical Engineering*, pages 985–990. Elsevier.

- TEIXEIRA, G. A. A., MAIA, A. S., SANTOS, I. M. G., SOUZA, A. L., SOUZA, A. G. et QUEIROZ, N. (2011). Biodiesels from beef tallow/soybean oil/babassu oil blends : Correlation between fluid dynamic properties and TMDSC data. *Journal of Thermal Analysis and Calorimetry*, 106(2):563–567.
- TEIXEIRA, M. A. (2003). *Caracterização energética do babaçu e análise do potencial de cogeração*. Ph. D., Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica, Campinas, SP.
- TEIXEIRA, M. A. (2005). Heat and power demands in babassu palm oil extraction industry in Brazil. *Energy Conversion and Management*, 46(13):2068–2074.
- TEIXEIRA, M. A. (2008). Babassu - A new approach for an ancient Brazilian biomass. *Biomass and Bioenergy*, 32(9):857–864.
- TOMLINSON, P. B. (1979). Systematics and Ecology of the Palmae. *Annual Review of Ecology & Systematics*, 10:85–107.
- WACKERMANN, G. (2003). *Les frontières dans un monde en mouvement*. Carrefours. Ellipses, Paris.
- WATRIN, O. D. S., SHIMABUKURO, Y. E., CRUZ, C. B. M. et MIRANDA, I. d. S. (2003). Bandas sintéticas derivadas de modelo de mistura espectral na avaliação da dinâmica da paisagem em assentamento rural na fronteira agrícola Amazônica. *In Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, volume 11, pages 729–736, Belo Horizonte - MG, Brasil.
- WEISS, K., MOSER, G. et GERMANN, C. (2006). Perception de l'environnement, conceptions du métier et pratiques culturelles des agriculteurs face au développement durable. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 56(2):73–81.
- ZHUKOVA, A., WALTEMATH, D., SWAT, M. J., VIK, J. O., RODRIGUEZ, N. et LE NOVÈRE, N. (2014). Mathematical Modelling Ontology (MAMO). <https://sourceforge.net/projects/mamo-ontology/>. Consulté le 01.06.2016.

Index

- Échantillonneur de Gibbs, 90
- Aléa environnemental, 86
- Algorithme Metropolis-Hastings, 90
- Arecaceae, 25
- Babaçual, 27
- Bayésien, 88
- Biome, 24
- Boîte à moustaches, 107
- BUGS, 94
- Cerrado, 24
- Chaînes de Markov, 90
- Conditions « naturelles », 52
- Convergence, 90
- Cycle de vie, 83, 93
- Densité-dépendance, 96, 114
- Droites de régression, 114
- Effet année, 96
- Ergodique, 90
- Fréquence, 87
- Goulot d'étranglement, 110
- Hyper-paramètre, 81
- Induction bayésienne, 93
- Informatif, 89
- JAGS, 94
- Jeune, 61
- Lois conjuguées, 97
- Matrice, 85
- Matrice de transition, 84, 85
- Modélisation, 55
- Modèle, 56
- Modèle d'observation, 56
- Modèle de données, 56
- Modèle hiérarchique, 86
- Modèle individu-centré, 95
- Modèle mathématique, 56
- Modèle relationnel, 75
- Modèle stochastique, 85
- Non-informatif, 89
- Paramètre, 81
- Parcelle, 86, 97
- Plantule, 61
- PostgreSQL/PostGIS, 78, 94
- PPP, 92
- Probabilité, 87
- Probabilité conditionnelle, 88
- Projeto de assentamento, 7
- Protocole d'observation, 111
- Quebradeiras, 29, 32
- R, 94
- Schéma conceptuel, 75
- SGBD, 78
- SQL, 78
- Stade, 60
- Stationnaire, 85
- Taux de croissance de la population λ , 85
- Test de Gelman-Rubin, 92
- Transect, 63, 86, 98
- UML, 24, 41–43, 75
- Validation croisée, 92
- Variable aléatoire, 81
- Variable d'état, 81, 95
- Variable latente, 81

Table des matières

Remerciements	i
Sommaire	v
Liste des tableaux	vii
Table des figures	ix
Abréviations et sigles	xi
Préambule	xiii
I Introduction	1
1 Contexte général	3
1.1 Prémices : collaboration initiale dans le cadre d'un stage de master	3
1.1.1 Le stage de master	3
1.1.2 Partenariat recherche : rapprochement des équipes	4
1.2 Le projet Open science	4
1.3 La thèse	5
2 Zone de l'étude	7
2.1 Territoire de Benfica	7
2.2 Anthropisation et colonisation	8
2.3 Cadre naturel	11
2.3.1 Hydrologie	11
2.3.2 Pédologie	11
2.3.3 Climat	11
2.3.4 Couvert végétal	12
II Problématique	13
3 Développement durable en Amazonie	15
3.1 Appropriation de la nature	15
3.2 La démesure humaine	17
3.3 Appropriation de l'Amazonie	17
3.3.1 Déforestation de la forêt amazonienne	18
3.3.2 Mécanismes du front pionnier	19

4	Autour de l'espèce babaçu	23
4.1	Babaçu : importante présence au Brésil	23
4.2	Espèce étudiée et convoitée	24
4.2.1	Taxonomie complexe	25
4.2.2	Biologie du babaçu	25
4.2.3	Babaçu : espèce adaptative	27
4.3	Contexte socio-économique autour du babaçu	28
4.3.1	Le babaçu : une ressource extractiviste disputée	28
4.3.1.1	Historique	28
4.3.1.2	Acteurs et produits	29
5	Projet Open Science, thèse et questionnements	33
5.1	Problématique Open Science et thèse	33
5.1.1	Open Science	33
5.1.2	Thèse	35
5.2	Questionnements et verrous	36
5.3	Chronologie	37
III	État de l'art	39
6	Données et connaissances	41
6.1	Capitalisation de connaissances et de données	41
6.2	Importance de la terminologie	42
6.3	Ontologie en modélisation mathématique	44
7	Dynamique de population	45
7.1	Dynamique de population des espèces végétales	45
7.1.1	La démarche générale	45
7.1.2	A propos des palmiers	46
7.2	Recherches antérieures sur la dynamique de population du babaçu	49
7.2.1	Processus de croissance du babaçu	49
7.2.2	Le recrutement	51
IV	Proposition	53
8	Introduction	55
8.1	Autour de la modélisation	55
8.1.1	Modélisation : un processus d'itération	55
8.1.2	Classification des modèles	56
8.2	Méthodologie générale	56

9	Acquisition et pérennisation des données	59
9.1	Définition des stades biologiques	59
9.2	Observations	62
9.2.1	Choix du site	63
9.2.2	Protocole : préparation de l'observation	63
9.3	Collecte des données	66
9.4	Données résultantes	68
9.4.1	Données des pâturages	68
9.4.2	Données des babaçuals	72
9.4.3	Données de la forêt tropicale humide	73
9.5	Structuration et pérennisation des données	74
9.5.1	Phase conceptuelle	75
9.5.2	Phase de passage au modèle logique	75
9.5.3	Phase d'implémentation physique	77
10	Modélisation mathématique	81
10.1	Modèle théorique	83
10.1.1	Modélisation de la dynamique de populations végétales	83
10.1.1.1	Basée sur le graphique du cycle de vie	83
10.1.1.2	Du modèle graphique au modèle matriciel	84
10.1.1.3	Révolution informatique	85
10.1.1.4	Modèle hiérarchique	86
10.1.2	Inférence statistique en modélisation	86
10.1.2.1	Évaluation du modèle	92
10.2	Mise en œuvre	93
10.2.1	Méthodes d'analyse des données	93
10.2.1.1	Graphique du cycle de vie du babaçu	93
10.2.1.2	Outils mis en œuvre	93
10.2.1.3	Échelon d'analyse : la cohorte	95
10.2.1.4	Variables d'état	95
10.2.1.5	Variables d'intérêt	95
10.2.1.6	Hypothèses sur la biologie	96
10.2.1.7	Hypothèses de modélisation	96
10.2.2	Construction du modèle babaçu	96
10.2.2.1	Choix du niveau hiérarchique : transect <i>vs</i> parcelle	97
10.2.2.2	Modèle multinomial de croissance	99
10.2.2.3	Niveau hiérarchique	101
V	Résultats, analyses et discussions	105
11	Données empiriques : premier regard analytique	107
11.1	Processus de croissance	108

11.1.1	Fréquence de mortalité	108
11.1.2	Fréquence de passage	109
11.1.3	Fréquence de stabilité dans le même stade	110
11.2	Recrutement	112
11.2.1	Adultes - infrutescences	112
11.2.2	Adultes - fruits	113
11.2.3	Fruits - nouveaux stade 1	113
12	Analyse de la dynamique de population	117
12.1	Analyse de sensibilité	118
12.2	Validation du modèle	119
12.2.1	<i>Posterior predictive p-values</i>	120
12.2.2	Validation croisée	121
12.3	Estimation des probabilités	122
12.3.1	Modèle de croissance	122
12.3.2	Recrutement	125
12.3.2.1	Séparation des palmiers adultes en deux catégories : 6 <i>a</i> et 6 <i>b</i>	125
12.3.2.2	Le nombre moyen de fruits par palmier adulte de type 6 <i>a</i>	127
12.3.2.3	Le processus de passage du fruit au nouveau palmier de stade 1	128
12.4	Modèle déterministe Caswell	128
VI	Conclusion et perspectives	133
13	Conclusion	135
13.1	Autour du contexte général	135
13.2	Autour de la dynamique du babaçu	135
14	Perspectives	137
14.1	Autour de la capitalisation des données et connaissances	137
14.2	Autour du modèle de dynamique	137
	Annexes	139
A	Graphiques densités	141
A.1	Pâturages	141
B	Scripts	143
B.1	Commandes génériques	143
B.2	Script bash « 2_commandes_terminal.sh »	143

B.3	Script R « 3_mise_forme_donnees.R »	145
B.4	Script SQL « 4_creation.sql »	147
B.5	Script SQL « 5_insertion.sql »	153
B.6	Script SQL « 6_verification.sql »	159
C	Convergence des chaînes de Markov	161
D	Validation croisée - tableaux	167
	Bibliographie	173
	Index	183
	Table des matières	185
	Résumé	191
	Abstract	193

Résumé

LE PALMIER babaçu (*Attalea speciosa* Mart. ex Spreng.) est une plante native de la forêt amazonienne. La déforestation a pour impact qu'il apparaît dans des milieux ouverts désormais anthropisés (pâturages, et champs cultivés). Le babaçu fait partie des ressources « extractivistes » du Brésil : activité de cueillette suivie de commercialisation de produits non ligneux. Cette activité concerne des personnes parmi les plus démunies du pays d'où l'importance de gérer au mieux sa durabilité dans un contexte conflictuel. Force est de constater que les connaissances du fonctionnement durable de l'espèce au sein de ces milieux anthropisés font cruellement défaut : son cycle de vie est peu connu et sa dynamique de population non étudiée.

Le travail de thèse est donc pionnier et s'inscrit au sein d'un projet interdisciplinaire Open Science de la fondation Agropolis. L'objectif est double : il consiste à produire un modèle de la dynamique de population du babaçu validé par une analyse *in situ* et à capitaliser les connaissances issues de divers milieux scientifiques (biologistes, écologues, mathématiciens, informaticiens et économistes français et brésiliens). Le modèle, à plus long terme, doit étayer les recommandations relatives à la gestion durable de l'espèce.

L'approche retenue a consisté à expliciter au sein de divers modèles, les connaissances partagées sur l'espèce et les modèles mathématiques potentiels ainsi que sur le contexte sociétal. Ceci nous a permis de proposer un modèle d'observation conforme aux préconisations d'OBOE (Ontologie relative à la sémantique des observations scientifiques) permettant la mise en place du protocole d'acquisition des données. L'acquisition *in situ* a été réalisée par le partenariat franco-brésilien entre 2013 et 2016 au sein du territoire de la communauté de Benfica (Pará, Brésil). La pérennisation des données s'est faite après conception, au sein d'une base de données spatiales prenant en compte l'aspect socio-économique.

Enfin, nous proposons un modèle matriciel aléatoire ayant pour entrée des variables agrégées au niveau des stades biologiques de l'espèce. Les probabilités de la matrice de transition entre les stades sont modélisées selon un modèle hiérarchique Dirichlet-Multinomial. L'intégration des informations *a priori* a été formalisée grâce à une approche bayésienne. L'estimation et la validation du modèle ont été effectuées avec des critères bayésiens.

Des simulations basées sur les paramètres estimés constituent une première ébauche d'étude du comportement de l'espèce.

Mots-clés : modélisation données et connaissances, modélisation mathématique, base de données spatiales, statistique bayésienne, dynamique de population, palmier babaçu

Abstract

THE BABASSU palm tree (*Attalea speciosa* Mart. ex Spreng.) is an endemic species of the amazonian forests. The deforestation highlights henceforth this palm tree in the anthropogenic open areas (pastures and cultivated fields). The babassu is one of the "extractive" resources in Brazil : gathering activity followed by marketing of non-timber products. This activity involves people among the most disadvantaged in the country thus the importance to better manage its sustainability in a context of conflict. It is clear that knowledge of sustainable functioning of the species within these manmade environments is sorely lacking : its life cycle is not well known and its population dynamics unstudied.

This PhD work is pioneer and is a part of an Open Science interdisciplinary project of the Agropolis foundation. The objective is twofold it is to generate a model of population dynamics of the babassu palm tree validated by *in situ* analysis and capitalizing knowledge from various scientific communities (biologists, ecologists, mathematicians, computer scientists and economists from France and Brazil). In the long-term the model should support the recommendations for the sustainable management of the species.

Our approach explicit within different models the shared knowledge about the species and the potential mathematical models as well as the societal context. This has allowed us to provide an observation model compliant with the recommendations of OBOE (Ontology on a semantic scientific observations) for the development of data acquisition protocol. The acquisition was made *in situ* by the French-Brazilian partnership between 2013 and 2016 in the community of Benfica (Pará, Brazil). The perpetuation of the data was made after conception, within a spatial database taking into account the socio-economic aspect.

Finally, we proposed a random matrix model having as input aggregated variables based on the biological stages of the species. The probabilities of the between stages transition matrix are modeled using a hierarchical Dirichlet-Multinomial model. The integration of prior information was formulated through a Bayesian approach. The estimation and model validation were performed with Bayesian criteria.

Simulations based on the estimated parameters represent a first outline of the study of the species behavior.

Keywords : modeling, mathematical modelling, spatial data base, bayesian statistics, babassu palm tree, sustainable management

